Instructor's Manual

by Thomas H. Cormen

to Accompany

Introduction to Algorithms

Third Edition

by Thomas H. Cormen
Charles E. Leiserson
Ronald L. Rivest
Clifford Stein

The MIT Press
Cambridge, Massachusetts London, England

Instructor's Manual to Accompany *Introduction to Algorithms*, Third Edition
by Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein

Contents

Revision History

Revisions are listed by date rather than being numbered.

- 22 February 2014. Corrected an error in the solution to Exercise 4.3-7, courtesy of Dan Suthers. Corrected an error in the solution to Exercise 23.1-6, courtesy of Rachel Ginzberg. Updated the Preface.
- 3 January 2012. Added solutions to Chapter 27. Added an alternative solution to Exercise 2.3-7, courtesy of Viktor Korsun and Crystal Peng. Corrected a minor error in the Chapter 15 notes in the recurrence for $T(n)$ for the recursive CUT-ROD procedure. Updated the solution to Problem 24-3. Corrected an error in the proof about the Edmonds-Karp algorithm performing $O(VE)$ flow augmentations. The bodies of all pseudocode procedures are indented slightly.
- 28 January 2011. Corrected an error in the solution to Problem 2-4(c), and removed unnecessary code in the solution to Problem 2-4(d). Added a missing parameter to recursive calls of REC-MAT-MULT on page 4-7. Changed the pseudocode for HEAP-EXTRACT-MAX on page 6-8 and MAX-HEAP-INSERT on page 6-9 to assume that the parameter n is passed by reference.
- 7 May 2010. Changed the solutions to Exercises 22.2-3 and 22.3-4 because these exercises changed.
- 17 February 2010. Corrected a minor error in the solution to Exercise 4.3-7.
- 16 December 2009. Added an alternative solution to Exercise 6.3-3, courtesy of Eyal Mashiach.
- 7 December 2009. Added solutions to Exercises 16.3-1, 26.1-1, 26.1-3, 26.1-7, 26.2-1, 26.2-8, 26.2-9, 26.2-12, 26.2-13, and 26.4-1 and to Problem 26-3. Corrected spelling in the solution to Exercise 16.2-4. Several corrections to the solution to Exercise 16.4-3, courtesy of Zhixiang Zhu. Minor changes to the solutions to Exercises 24.3-3 and 24.4-7 and Problem 24-1.
- 7 August 2009. Initial release.

Preface

This document is an instructor's manual to accompany *Introduction to Algorithms*, Third Edition, by Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. It is intended for use in a course on algorithms. You might also find some of the material herein to be useful for a CS 2-style course in data structures.

Unlike the instructor's manual for the first edition of the text—which was organized around the undergraduate algorithms course taught by Charles Leiserson at MIT in Spring 1991—but like the instructor's manual for the second edition, we have chosen to organize the manual for the third edition according to chapters of the text. That is, for most chapters we have provided a set of lecture notes and a set of exercise and problem solutions pertaining to the chapter. This organization allows you to decide how to best use the material in the manual in your own course.

We have not included lecture notes and solutions for every chapter, nor have we included solutions for every exercise and problem within the chapters that we have selected. We felt that Chapter 1 is too nontechnical to include here, and Chapter 10 consists of background material that often falls outside algorithms and data-structures courses. We have also omitted the chapters that are not covered in the courses that we teach: Chapters 18–20 and 27–35 (though we do include some solutions for Chapter 27), as well as Appendices A–D; future editions of this manual may include some of these chapters. There are two reasons that we have not included solutions to all exercises and problems in the selected chapters. First, writing up all these solutions would take a long time, and we felt it more important to release this manual in as timely a fashion as possible. Second, if we were to include all solutions, this manual would be much longer than the text itself.

We have numbered the pages in this manual using the format *CC-PP*, where *CC* is a chapter number of the text and *PP* is the page number within that chapter's lecture notes and solutions. The *PP* numbers restart from 1 at the beginning of each chapter's lecture notes. We chose this form of page numbering so that if we add or change solutions to exercises and problems, the only pages whose numbering is affected are those for the solutions for that chapter. Moreover, if we add material for currently uncovered chapters, the numbers of the existing pages will remain unchanged.

The lecture notes

The lecture notes are based on three sources:

- Some are from the first-edition manual; they correspond to Charles Leiserson's lectures in MIT's undergraduate algorithms course, 6.046.
- Some are from Tom Cormen's lectures in Dartmouth College's undergraduate algorithms course, CS 25.
- Some are written just for this manual.

You will find that the lecture notes are more informal than the text, as is appropriate for a lecture situation. In some places, we have simplified the material for lecture presentation or even omitted certain considerations. Some sections of the text—usually starred—are omitted from the lecture notes. (We have included lecture notes for one starred section: 12.4, on randomly built binary search trees, which we covered in an optional CS 25 lecture.)

In several places in the lecture notes, we have included "asides" to the instructor. The asides are typeset in a slanted font and are enclosed in square brackets. *[Here is an aside.]* Some of the asides suggest leaving certain material on the board, since you will be coming back to it later. If you are projecting a presentation rather than writing on a blackboard or whiteboard, you might want to replicate slides containing this material so that you can easily reprise them later in the lecture.

We have chosen not to indicate how long it takes to cover material, as the time necessary to cover a topic depends on the instructor, the students, the class schedule, and other variables.

There are two differences in how we write pseudocode in the lecture notes and the text:

- Lines are not numbered in the lecture notes. We find them inconvenient to number when writing pseudocode on the board.
- We avoid using the *length* attribute of an array. Instead, we pass the array length as a parameter to the procedure. This change makes the pseudocode more concise, as well as matching better with the description of what it does.

We have also minimized the use of shading in figures within lecture notes, since drawing a figure with shading on a blackboard or whiteboard is difficult.

The solutions

The solutions are based on the same sources as the lecture notes. They are written a bit more formally than the lecture notes, though a bit less formally than the text. We do not number lines of pseudocode, but we do use the *length* attribute (on the assumption that you will want your students to write pseudocode as it appears in the text).

As of the third edition, we have publicly posted a few solutions on the book's website. These solutions also appear in this manual, with the notation "This solution is also posted publicly" after the exercise or problem number. The set of publicly posted solutions might increase over time, and so we encourage you to check whether a particular solution is posted on the website before you assign an exercise or problem to your students.

The index lists all the exercises and problems for which this manual provides solutions, along with the number of the page on which each solution starts.

Asides appear in a handful of places throughout the solutions. Also, we are less reluctant to use shading in figures within solutions, since these figures are more likely to be reproduced than to be drawn on a board.

Source files

For several reasons, we are unable to publish or transmit source files for this manual. We apologize for this inconvenience.

You can use the clrscode3e package for LaTeX 2_ε to typeset pseudocode in the same way that we do. You can find this package at http://www.cs.dartmouth.edu/~thc/clrscode/. That site also includes documentation. Make sure to use the clrscode3e package, not the clrscode package; clrscode is for the second edition of the book.

Reporting errors and suggestions

Undoubtedly, instructors will find errors in this manual. Please report errors by sending email to clrs-manual-bugs@mitpress.mit.edu.

If you have a suggestion for an improvement to this manual, please feel free to submit it via email to clrs-manual-suggestions@mitpress.mit.edu.

As usual, if you find an error in the text itself, please verify that it has not already been posted on the errata web page before you submit it. You can use the MIT Press web site for the text, http://mitpress.mit.edu/algorithms/, to locate the errata web page and to submit an error report.

We thank you in advance for your assistance in correcting errors in both this manual and the text.

How we produced this manual

Like the third edition of *Introduction to Algorithms*, this manual was produced in LaTeX 2_ε. We used the Times font with mathematics typeset using the MathTime Pro 2 fonts. As in all three editions of the textbook, we compiled the index using Windex, a C program that we wrote. We drew the illustrations using MacDraw Pro,[1] with some of the mathematical expressions in illustrations laid in with the psfrag package for LaTeX 2_ε. We created the PDF files for this manual on a MacBook Pro running OS 10.5.

Acknowledgments

This manual borrows heavily from the manuals for the first two editions. Julie Sussman, P.P.A., wrote the first-edition manual. Julie did such a superb job on the

[1] See our plea in the preface for the third edition to Apple, asking that they update MacDraw Pro for OS X.

first-edition manual, finding numerous errors in the first-edition text in the process, that we were thrilled to have her serve as technical copyeditor for both the second and third editions of the book. Charles Leiserson also put in large amounts of time working with Julie on the first-edition manual.

The manual for the second edition was written by Tom Cormen, Clara Lee, and Erica Lin. Clara and Erica were undergraduate computer science majors at Dartmouth at the time, and they did a superb job.

The other three *Introduction to Algorithms* authors—Charles Leiserson, Ron Rivest, and Cliff Stein—provided helpful comments and suggestions for solutions to exercises and problems. Some of the solutions are modifications of those written over the years by teaching assistants for algorithms courses at MIT and Dartmouth. At this point, we do not know which TAs wrote which solutions, and so we simply thank them collectively. Several of the solutions to new exercises and problems in the third edition were written by Sharath Gururaj of Columbia University; we thank Sharath for his fine work. The solutions for Chapter 27 were written by Priya Natarajan.

We also thank the MIT Press and our editors—Ada Brunstein, Jim DeWolf, and Marie Lee— for moral and financial support. Tim Tregubov and Wayne Cripps provided computer support at Dartmouth.

THOMAS H. CORMEN
Hanover, New Hampshire
August 2009

Lecture Notes for Chapter 2: Getting Started

Chapter 2 overview

Goals

- Start using frameworks for describing and analyzing algorithms.
- Examine two algorithms for sorting: insertion sort and merge sort.
- See how to describe algorithms in pseudocode.
- Begin using asymptotic notation to express running-time analysis.
- Learn the technique of "divide and conquer" in the context of merge sort.

Insertion sort

The sorting problem

Input: A sequence of n numbers $\langle a_1, a_2, \ldots, a_n \rangle$.

Output: A permutation (reordering) $\langle a'_1, a'_2, \ldots, a'_n \rangle$ of the input sequence such that $a'_1 \leq a'_2 \leq \cdots \leq a'_n$.

The sequences are typically stored in arrays.

We also refer to the numbers as ***keys***. Along with each key may be additional information, known as ***satellite data***. *[You might want to clarify that "satellite data" does not necessarily come from a satellite.]*

We will see several ways to solve the sorting problem. Each way will be expressed as an ***algorithm***: a well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output.

Expressing algorithms

We express algorithms in whatever way is the clearest and most concise.

English is sometimes the best way.

When issues of control need to be made perfectly clear, we often use ***pseudocode***.

- Pseudocode is similar to C, C++, Pascal, and Java. If you know any of these languages, you should be able to understand pseudocode.
- Pseudocode is designed for *expressing algorithms to humans*. Software engineering issues of data abstraction, modularity, and error handling are often ignored.
- We sometimes embed English statements into pseudocode. Therefore, unlike for "real" programming languages, we cannot create a compiler that translates pseudocode to machine code.

Insertion sort

A good algorithm for sorting a small number of elements.

It works the way you might sort a hand of playing cards:

- Start with an empty left hand and the cards face down on the table.
- Then remove one card at a time from the table, and insert it into the correct position in the left hand.
- To find the correct position for a card, compare it with each of the cards already in the hand, from right to left.
- At all times, the cards held in the left hand are sorted, and these cards were originally the top cards of the pile on the table.

Pseudocode

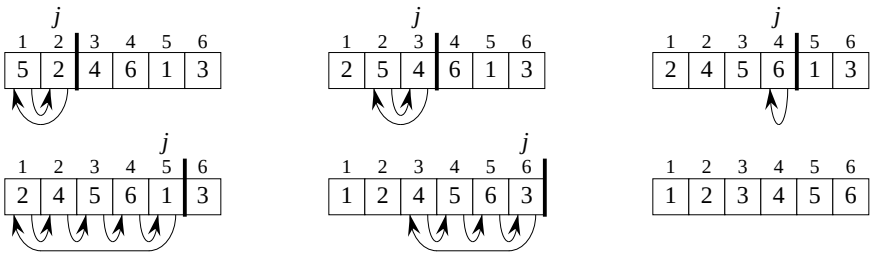
We use a procedure INSERTION-SORT.

- Takes as parameters an array $A[1 \,..\, n]$ and the length n of the array.
- As in Pascal, we use ". ." to denote a range within an array.
- *[We usually use 1-origin indexing, as we do here. There are a few places in later chapters where we use 0-origin indexing instead. If you are translating pseudocode to C, C++, or Java, which use 0-origin indexing, you need to be careful to get the indices right. One option is to adjust all index calculations in the C, C++, or Java code to compensate. An easier option is, when using an array $A[1 \,..\, n]$, to allocate the array to be one entry longer—$A[0 \,..\, n]$—and just don't use the entry at index 0.]*
- *[In the lecture notes, we indicate array lengths by parameters rather than by using the length attribute that is used in the book. That saves us a line of pseudocode each time. The solutions continue to use the length attribute.]*
- The array A is sorted ***in place***: the numbers are rearranged within the array, with at most a constant number outside the array at any time.

INSERTION-SORT(A, n)	*cost*	*times*
for $j = 2$ **to** n	c_1	n
$key = A[j]$	c_2	$n-1$
// Insert $A[j]$ into the sorted sequence $A[1..j-1]$.	0	$n-1$
$i = j-1$	c_4	$n-1$
while $i > 0$ and $A[i] > key$	c_5	$\sum_{j=2}^{n} t_j$
$A[i+1] = A[i]$	c_6	$\sum_{j=2}^{n}(t_j-1)$
$i = i-1$	c_7	$\sum_{j=2}^{n}(t_j-1)$
$A[i+1] = key$	c_8	$n-1$

[Leave this on the board, but show only the pseudocode for now. We'll put in the "cost" and "times" columns later.]

Example



[Read this figure row by row. Each part shows what happens for a particular iteration with the value of j indicated. j indexes the "current card" being inserted into the hand. Elements to the left of $A[j]$ that are greater than $A[j]$ move one position to the right, and $A[j]$ moves into the evacuated position. The heavy vertical lines separate the part of the array in which an iteration works—$A[1..j]$—from the part of the array that is unaffected by this iteration—$A[j+1..n]$. The last part of the figure shows the final sorted array.]

Correctness

We often use a ***loop invariant*** to help us understand why an algorithm gives the correct answer. Here's the loop invariant for INSERTION-SORT:

> **Loop invariant:** At the start of each iteration of the "outer" **for** loop—the loop indexed by j—the subarray $A[1..j-1]$ consists of the elements originally in $A[1..j-1]$ but in sorted order.

To use a loop invariant to prove correctness, we must show three things about it:

Initialization: It is true prior to the first iteration of the loop.

Maintenance: If it is true before an iteration of the loop, it remains true before the next iteration.

Termination: When the loop terminates, the invariant—usually along with the reason that the loop terminated—gives us a useful property that helps show that the algorithm is correct.

Using loop invariants is like mathematical induction:

- To prove that a property holds, you prove a base case and an inductive step.
- Showing that the invariant holds before the first iteration is like the base case.
- Showing that the invariant holds from iteration to iteration is like the inductive step.
- The termination part differs from the usual use of mathematical induction, in which the inductive step is used infinitely. We stop the "induction" when the loop terminates.
- We can show the three parts in any order.

For insertion sort

Initialization: Just before the first iteration, $j = 2$. The subarray $A[1..j-1]$ is the single element $A[1]$, which is the element originally in $A[1]$, and it is trivially sorted.

Maintenance: To be precise, we would need to state and prove a loop invariant for the "inner" **while** loop. Rather than getting bogged down in another loop invariant, we instead note that the body of the inner **while** loop works by moving $A[j-1]$, $A[j-2]$, $A[j-3]$, and so on, by one position to the right until the proper position for *key* (which has the value that started out in $A[j]$) is found. At that point, the value of *key* is placed into this position.

Termination: The outer **for** loop ends when $j > n$, which occurs when $j = n+1$. Therefore, $j - 1 = n$. Plugging n in for $j-1$ in the loop invariant, the subarray $A[1..n]$ consists of the elements originally in $A[1..n]$ but in sorted order. In other words, the entire array is sorted.

Pseudocode conventions

[Covering most, but not all, here. See book pages 20–22 for all conventions.]

- Indentation indicates block structure. Saves space and writing time.
- Looping constructs are like in C, C++, Pascal, and Java. We assume that the loop variable in a **for** loop is still defined when the loop exits (unlike in Pascal).
- // indicates that the remainder of the line is a comment.
- Variables are local, unless otherwise specified.
- We often use ***objects***, which have ***attributes***. For an attribute *attr* of object x, we write $x.attr$. (This notation matches $x.attr$ in Java and is equivalent to x->*attr* in C++.) Attributes can cascade, so that if $x.y$ is an object and this object has attribute *attr*, then $x.y.attr$ indicates this object's attribute. That is, $x.y.attr$ is implicitly parenthesized as $(x.y).attr$.
- Objects are treated as references, like in Java. If x and y denote objects, then the assignment $y = x$ makes x and y reference the same object. It does not cause attributes of one object to be copied to another.
- Parameters are passed by value, as in Java and C (and the default mechanism in Pascal and C++). When an object is passed by value, it is actually a reference (or pointer) that is passed; changes to the reference itself are not seen by the caller, but changes to the object's attributes are.

- The boolean operators "and" and "or" are ***short-circuiting***: if after evaluating the left-hand operand, we know the result of the expression, then we don't evaluate the right-hand operand. (If x is FALSE in "x and y" then we don't evaluate y. If x is TRUE in "x or y" then we don't evaluate y.)

Analyzing algorithms

We want to predict the resources that the algorithm requires. Usually, running time.

In order to predict resource requirements, we need a computational model.

Random-access machine (RAM) model

- Instructions are executed one after another. No concurrent operations.
- It's too tedious to define each of the instructions and their associated time costs.
- Instead, we recognize that we'll use instructions commonly found in real computers:
  - Arithmetic: add, subtract, multiply, divide, remainder, floor, ceiling). Also, shift left/shift right (good for multiplying/dividing by 2^k).
  - Data movement: load, store, copy.
  - Control: conditional/unconditional branch, subroutine call and return.

  Each of these instructions takes a constant amount of time.

The RAM model uses integer and floating-point types.

- We don't worry about precision, although it is crucial in certain numerical applications.
- There is a limit on the word size: when working with inputs of size n, assume that integers are represented by $c \lg n$ bits for some constant $c \geq 1$. ($\lg n$ is a very frequently used shorthand for $\log_2 n$.)
  - $c \geq 1 \Rightarrow$ we can hold the value of $n \Rightarrow$ we can index the individual elements.
  - c is a constant $\Rightarrow$ the word size cannot grow arbitrarily.

How do we analyze an algorithm's running time?

The time taken by an algorithm depends on the input.

- Sorting 1000 numbers takes longer than sorting 3 numbers.
- A given sorting algorithm may even take differing amounts of time on two inputs of the same size.
- For example, we'll see that insertion sort takes less time to sort n elements when they are already sorted than when they are in reverse sorted order.

Input size

Depends on the problem being studied.

- Usually, the number of items in the input. Like the size n of the array being sorted.
- But could be something else. If multiplying two integers, could be the total number of bits in the two integers.
- Could be described by more than one number. For example, graph algorithm running times are usually expressed in terms of the number of vertices and the number of edges in the input graph.

Running time

On a particular input, it is the number of primitive operations (steps) executed.

- Want to define steps to be machine-independent.
- Figure that each line of pseudocode requires a constant amount of time.
- One line may take a different amount of time than another, but each execution of line i takes the same amount of time c_i.
- This is assuming that the line consists only of primitive operations.
  - If the line is a subroutine call, then the actual call takes constant time, but the execution of the subroutine being called might not.
  - If the line specifies operations other than primitive ones, then it might take more than constant time. Example: "sort the points by x-coordinate."

Analysis of insertion sort

[Now add statement costs and number of times executed to INSERTION-SORT *pseudocode.]*

- Assume that the ith line takes time c_i, which is a constant. (Since the third line is a comment, it takes no time.)
- For $j = 2, 3, \ldots, n$, let t_j be the number of times that the **while** loop test is executed for that value of j.
- Note that when a **for** or **while** loop exits in the usual way—due to the test in the loop header—the test is executed one time more than the loop body.

The running time of the algorithm is

$$\sum_{\text{all statements}} (\text{cost of statement}) \cdot (\text{number of times statement is executed}) \ .$$

Let $T(n) =$ running time of INSERTION-SORT.

$$\begin{aligned} T(n) &= c_1 n + c_2(n-1) + c_4(n-1) + c_5 \sum_{j=2}^{n} t_j + c_6 \sum_{j=2}^{n} (t_j - 1) \\ &\quad + c_7 \sum_{j=2}^{n} (t_j - 1) + c_8(n-1) \ . \end{aligned}$$

The running time depends on the values of t_j. These vary according to the input.

Best case

The array is already sorted.

- Always find that $A[i] \le key$ upon the first time the **while** loop test is run (when $i = j - 1$).
- All t_j are 1.
- Running time is
$$\begin{aligned} T(n) &= c_1 n + c_2(n-1) + c_4(n-1) + c_5(n-1) + c_8(n-1) \\ &= (c_1 + c_2 + c_4 + c_5 + c_8)n - (c_2 + c_4 + c_5 + c_8) \ . \end{aligned}$$
- Can express $T(n)$ as $an + b$ for constants a and b (that depend on the statement costs c_i) $\Rightarrow$ $T(n)$ is a *linear function* of n.

Worst case

The array is in reverse sorted order.

- Always find that $A[i] > key$ in while loop test.
- Have to compare *key* with all elements to the left of the jth position $\Rightarrow$ compare with $j - 1$ elements.
- Since the while loop exits because i reaches 0, there's one additional test after the $j - 1$ tests $\Rightarrow$ $t_j = j$.
- $\displaystyle\sum_{j=2}^{n} t_j = \sum_{j=2}^{n} j$ and $\displaystyle\sum_{j=2}^{n} (t_j - 1) = \sum_{j=2}^{n} (j - 1)$.
- $\displaystyle\sum_{j=1}^{n} j$ is known as an ***arithmetic series***, and equation (A.1) shows that it equals $\dfrac{n(n+1)}{2}$.
- Since $\displaystyle\sum_{j=2}^{n} j = \left(\sum_{j=1}^{n} j\right) - 1$, it equals $\dfrac{n(n+1)}{2} - 1$.
[The parentheses around the summation are not strictly necessary. They are there for clarity, but it might be a good idea to remind the students that the meaning of the expression would be the same even without the parentheses.]
- Letting $k = j - 1$, we see that $\displaystyle\sum_{j=2}^{n} (j-1) = \sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$.
- Running time is
$$\begin{aligned} T(n) &= c_1 n + c_2(n-1) + c_4(n-1) + c_5\left(\frac{n(n+1)}{2} - 1\right) \\ &\quad + c_6\left(\frac{n(n-1)}{2}\right) + c_7\left(\frac{n(n-1)}{2}\right) + c_8(n-1) \\ &= \left(\frac{c_5}{2} + \frac{c_6}{2} + \frac{c_7}{2}\right)n^2 + \left(c_1 + c_2 + c_4 + \frac{c_5}{2} - \frac{c_6}{2} - \frac{c_7}{2} + c_8\right)n \\ &\quad - (c_2 + c_4 + c_5 + c_8) \ . \end{aligned}$$
- Can express $T(n)$ as $an^2 + bn + c$ for constants a, b, c (that again depend on statement costs) $\Rightarrow$ $T(n)$ is a *quadratic function* of n.

Worst-case and average-case analysis

We usually concentrate on finding the ***worst-case running time***: the longest running time for *any* input of size n.

Reasons

- The worst-case running time gives a guaranteed upper bound on the running time for any input.
- For some algorithms, the worst case occurs often. For example, when searching, the worst case often occurs when the item being searched for is not present, and searches for absent items may be frequent.
- Why not analyze the average case? Because it's often about as bad as the worst case.

  Example: Suppose that we randomly choose n numbers as the input to insertion sort.

  On average, the key in $A[j]$ is less than half the elements in $A[1\,..\,j-1]$ and it's greater than the other half.
  $\Rightarrow$ On average, the **while** loop has to look halfway through the sorted subarray $A[1\,..\,j-1]$ to decide where to drop *key*.
  $\Rightarrow t_j \approx j/2$.

  Although the average-case running time is approximately half of the worst-case running time, it's still a quadratic function of n.

Order of growth

Another abstraction to ease analysis and focus on the important features.

Look only at the leading term of the formula for running time.

- Drop lower-order terms.
- Ignore the constant coefficient in the leading term.

Example: For insertion sort, we already abstracted away the actual statement costs to conclude that the worst-case running time is $an^2 + bn + c$.
Drop lower-order terms $\Rightarrow an^2$.
Ignore constant coefficient $\Rightarrow n^2$.

But we cannot say that the worst-case running time $T(n)$ *equals* n^2.

It *grows like* n^2. But it doesn't *equal* n^2.

We say that the running time is $\Theta(n^2)$ to capture the notion that the *order of growth* is n^2.

We usually consider one algorithm to be more efficient than another if its worst-case running time has a smaller order of growth.

Designing algorithms

There are many ways to design algorithms.

For example, insertion sort is ***incremental***: having sorted $A[1 \,..\, j-1]$, place $A[j]$ correctly, so that $A[1 \,..\, j]$ is sorted.

Divide and conquer

Another common approach.

Divide the problem into a number of subproblems that are smaller instances of the same problem.

Conquer the subproblems by solving them recursively.
Base case: If the subproblems are small enough, just solve them by brute force.

[It would be a good idea to make sure that your students are comfortable with recursion. If they are not, then they will have a hard time understanding divide and conquer.]

Combine the subproblem solutions to give a solution to the original problem.

Merge sort

A sorting algorithm based on divide and conquer. Its worst-case running time has a lower order of growth than insertion sort.

Because we are dealing with subproblems, we state each subproblem as sorting a subarray $A[p \,..\, r]$. Initially, $p = 1$ and $r = n$, but these values change as we recurse through subproblems.

To sort $A[p \,..\, r]$:

Divide by splitting into two subarrays $A[p \,..\, q]$ and $A[q+1 \,..\, r]$, where q is the halfway point of $A[p \,..\, r]$.

Conquer by recursively sorting the two subarrays $A[p \,..\, q]$ and $A[q+1 \,..\, r]$.

Combine by merging the two sorted subarrays $A[p \,..\, q]$ and $A[q+1 \,..\, r]$ to produce a single sorted subarray $A[p \,..\, r]$. To accomplish this step, we'll define a procedure MERGE(A, p, q, r).

The recursion bottoms out when the subarray has just 1 element, so that it's trivially sorted.

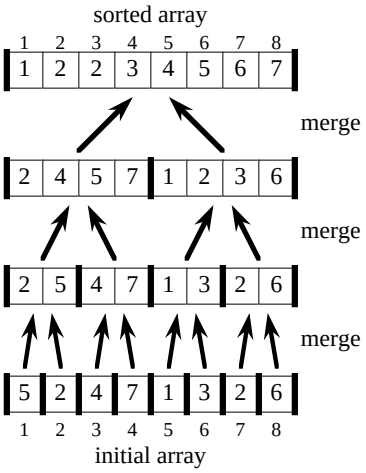
```
MERGE-SORT(A, p, r)
  if p < r                          // check for base case
     q = ⌊(p + r)/2⌋                // divide
     MERGE-SORT(A, p, q)            // conquer
     MERGE-SORT(A, q + 1, r)        // conquer
     MERGE(A, p, q, r)              // combine
```

Initial call: MERGE-SORT$(A, 1, n)$

[It is astounding how often students forget how easy it is to compute the halfway point of p and r as their average $(p + r)/2$. We of course have to take the floor to ensure that we get an integer index q. But it is common to see students perform calculations like $p + (r - p)/2$, or even more elaborate expressions, forgetting the easy way to compute an average.]

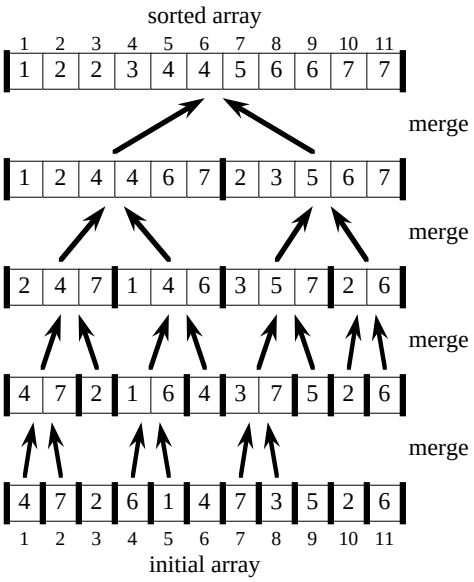
Example

Bottom-up view for $n = 8$: *[Heavy lines demarcate subarrays used in subproblems.]*



[Examples when n is a power of 2 are most straightforward, but students might also want an example when n is not a power of 2.]

Bottom-up view for $n = 11$:



[Here, at the next-to-last level of recursion, some of the subproblems have only 1 element. The recursion bottoms out on these single-element subproblems.]

Merging

What remains is the MERGE procedure.

Input: Array A and indices p, q, r such that

- $p \le q < r$.
- Subarray $A[p\,..\,q]$ is sorted and subarray $A[q+1\,..\,r]$ is sorted. By the restrictions on p, q, r, neither subarray is empty.

Output: The two subarrays are merged into a single sorted subarray in $A[p\,..\,r]$.

We implement it so that it takes $\Theta(n)$ time, where $n = r - p + 1 =$ the number of elements being merged.

What is n? Until now, n has stood for the size of the original problem. But now we're using it as the size of a subproblem. We will use this technique when we analyze recursive algorithms. Although we may denote the original problem size by n, in general n will be the size of a given subproblem.

Idea behind linear-time merging

Think of two piles of cards.

- Each pile is sorted and placed face-up on a table with the smallest cards on top.
- We will merge these into a single sorted pile, face-down on the table.
- A basic step:
  - Choose the smaller of the two top cards.
  - Remove it from its pile, thereby exposing a new top card.
  - Place the chosen card face-down onto the output pile.
- Repeatedly perform basic steps until one input pile is empty.
- Once one input pile empties, just take the remaining input pile and place it face-down onto the output pile.
- Each basic step should take constant time, since we check just the two top cards.
- There are $\le n$ basic steps, since each basic step removes one card from the input piles, and we started with n cards in the input piles.
- Therefore, this procedure should take $\Theta(n)$ time.

We don't actually need to check whether a pile is empty before each basic step.

- Put on the bottom of each input pile a special ***sentinel*** card.
- It contains a special value that we use to simplify the code.
- We use ∞, since that's guaranteed to "lose" to any other value.
- The only way that ∞ *cannot* lose is when both piles have ∞ exposed as their top cards.
- But when that happens, all the nonsentinel cards have already been placed into the output pile.
- We know in advance that there are exactly $r - p + 1$ nonsentinel cards $\Rightarrow$ stop once we have performed $r - p + 1$ basic steps. Never a need to check for sentinels, since they'll always lose.
- Rather than even counting basic steps, just fill up the output array from index p up through and including index r.

Pseudocode

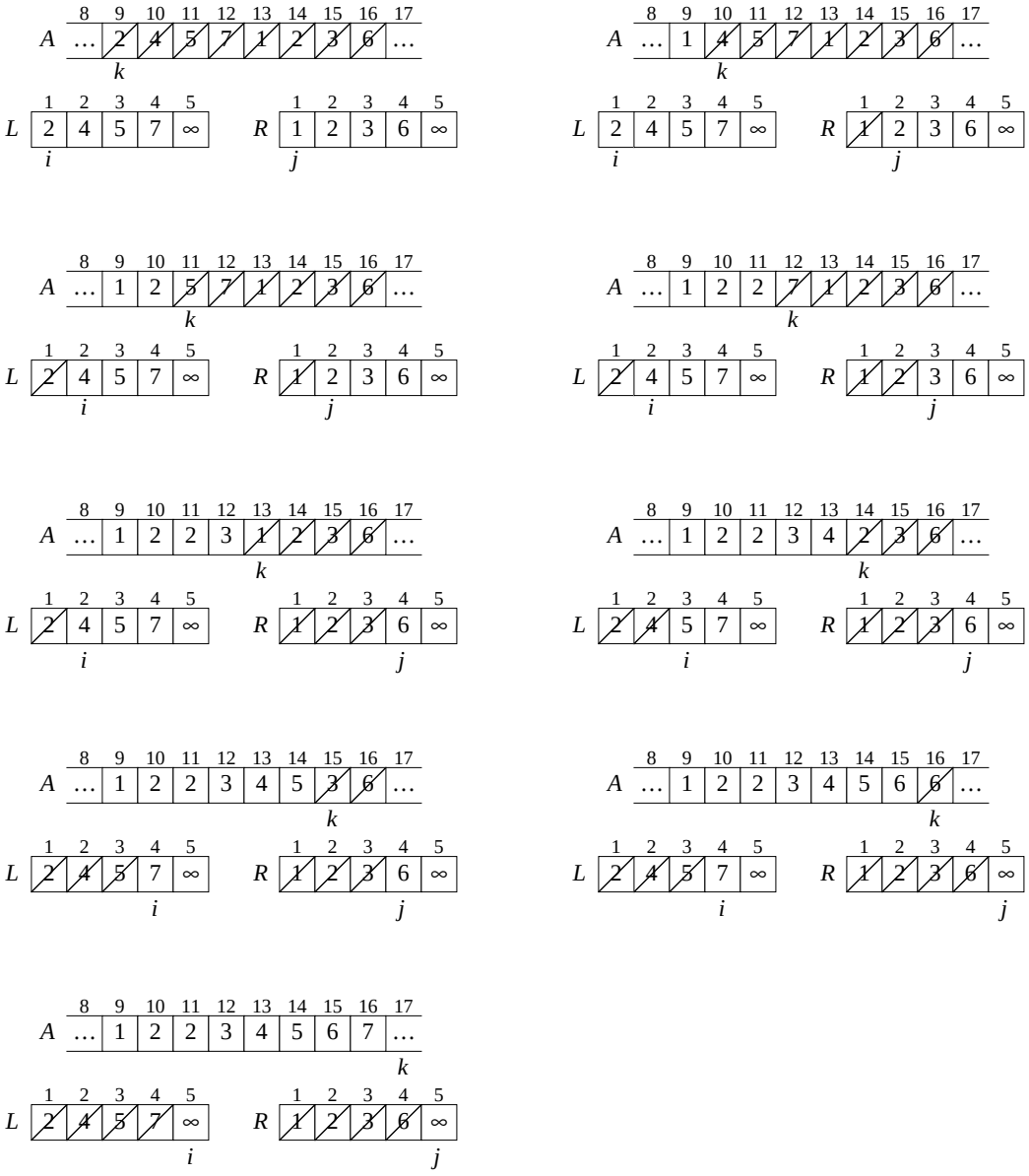
MERGE(A, p, q, r)

```
n_1 = q - p + 1
n_2 = r - q
let L[1..n_1 + 1] and R[1..n_2 + 1] be new arrays
for i = 1 to n_1
    L[i] = A[p + i - 1]
for j = 1 to n_2
    R[j] = A[q + j]
L[n_1 + 1] = ∞
R[n_2 + 1] = ∞
i = 1
j = 1
for k = p to r
    if L[i] ≤ R[j]
        A[k] = L[i]
        i = i + 1
    else A[k] = R[j]
        j = j + 1
```

[The book uses a loop invariant to establish that MERGE *works correctly. In a lecture situation, it is probably better to use an example to show that the procedure works correctly.]*

Example

A call of MERGE$(9, 12, 16)$



*[Read this figure row by row. The first part shows the arrays at the start of the "**for** $k = p$ **to** r" loop, where $A[p..q]$ is copied into $L[1..n_1]$ and $A[q+1..r]$ is copied into $R[1..n_2]$. Succeeding parts show the situation at the start of successive iterations. Entries in A with slashes have had their values copied to either L or R and have not had a value copied back in yet. Entries in L and R with slashes have been copied back into A. The last part shows that the subarrays are merged back into $A[p..r]$, which is now sorted, and that only the sentinels (∞) are exposed in the arrays L and R.]*

Running time

The first two **for** loops take $\Theta(n_1 + n_2) = \Theta(n)$ time. The last **for** loop makes n iterations, each taking constant time, for $\Theta(n)$ time.
Total time: $\Theta(n)$.

Analyzing divide-and-conquer algorithms

Use a ***recurrence equation*** (more commonly, a ***recurrence***) to describe the running time of a divide-and-conquer algorithm.

Let $T(n) =$ running time on a problem of size n.

- If the problem size is small enough (say, $n \le c$ for some constant c), we have a base case. The brute-force solution takes constant time: $\Theta(1)$.
- Otherwise, suppose that we divide into a subproblems, each $1/b$ the size of the original. (In merge sort, $a = b = 2$.)
- Let the time to divide a size-n problem be $D(n)$.
- Have a subproblems to solve, each of size $n/b \Rightarrow$ each subproblem takes $T(n/b)$ time to solve $\Rightarrow$ we spend $aT(n/b)$ time solving subproblems.
- Let the time to combine solutions be $C(n)$.
- We get the recurrence

$$T(n) = \begin{cases} \Theta(1) & \text{if } n \le c \ , \\ aT(n/b) + D(n) + C(n) & \text{otherwise} \ . \end{cases}$$

Analyzing merge sort

For simplicity, assume that n is a power of $2 \Rightarrow$ each divide step yields two subproblems, both of size exactly $n/2$.

The base case occurs when $n = 1$.

When $n \ge 2$, time for merge sort steps:

Divide: Just compute q as the average of p and $r \Rightarrow D(n) = \Theta(1)$.

Conquer: Recursively solve 2 subproblems, each of size $n/2 \Rightarrow 2T(n/2)$.

Combine: MERGE on an n-element subarray takes $\Theta(n)$ time $\Rightarrow C(n) = \Theta(n)$.

Since $D(n) = \Theta(1)$ and $C(n) = \Theta(n)$, summed together they give a function that is linear in n: $\Theta(n) \Rightarrow$ recurrence for merge sort running time is

$$T(n) = \begin{cases} \Theta(1) & \text{if } n = 1 \ , \\ 2T(n/2) + \Theta(n) & \text{if } n > 1 \ . \end{cases}$$

Solving the merge-sort recurrence

By the master theorem in Chapter 4, we can show that this recurrence has the solution $T(n) = \Theta(n \lg n)$. *[Reminder:* $\lg n$ *stands for* $\log_2 n$*.]*

Compared to insertion sort ($\Theta(n^2)$ worst-case time), merge sort is faster. Trading a factor of n for a factor of $\lg n$ is a good deal.

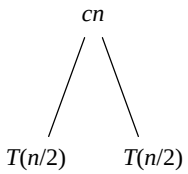
On small inputs, insertion sort may be faster. But for large enough inputs, merge sort will always be faster, because its running time grows more slowly than insertion sort's.

We can understand how to solve the merge-sort recurrence without the master theorem.

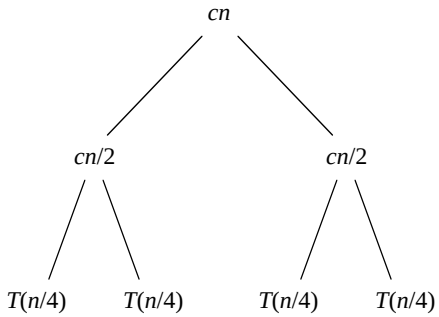
- Let c be a constant that describes the running time for the base case and also is the time per array element for the divide and conquer steps. *[Of course, we cannot necessarily use the same constant for both. It's not worth going into this detail at this point.]*
- We rewrite the recurrence as

$$T(n) = \begin{cases} c & \text{if } n = 1 \ , \\ 2T(n/2) + cn & \text{if } n > 1 \ . \end{cases}$$

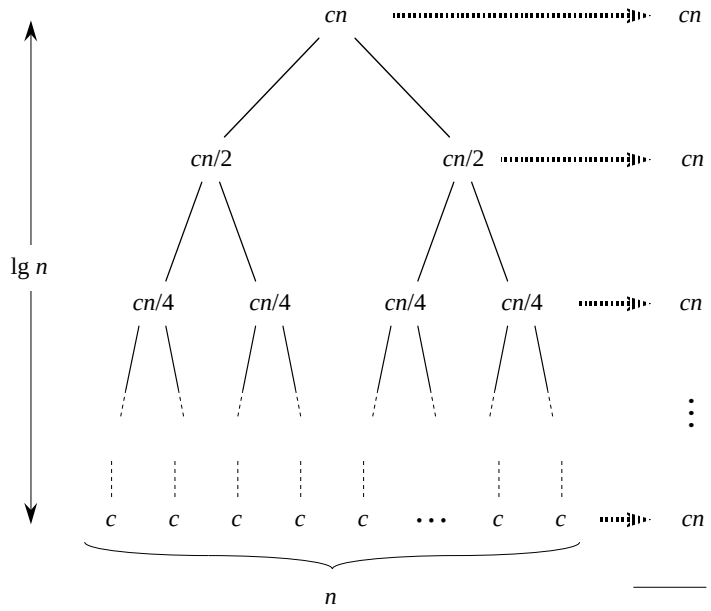
- Draw a ***recursion tree***, which shows successive expansions of the recurrence.
- For the original problem, we have a cost of cn, plus the two subproblems, each costing $T(n/2)$:



- For each of the size-$n/2$ subproblems, we have a cost of $cn/2$, plus two subproblems, each costing $T(n/4)$:



- Continue expanding until the problem sizes get down to 1:



- Each level has cost cn.
  - The top level has cost cn.
  - The next level down has 2 subproblems, each contributing cost $cn/2$.
  - The next level has 4 subproblems, each contributing cost $cn/4$.
  - Each time we go down one level, the number of subproblems doubles but the cost per subproblem halves $\Rightarrow$ cost per level stays the same.
- There are $\lg n + 1$ levels (height is $\lg n$).
  - Use induction.
  - Base case: $n = 1 \Rightarrow 1$ level, and $\lg 1 + 1 = 0 + 1 = 1$.
  - Inductive hypothesis is that a tree for a problem size of 2^i has $\lg 2^i + 1 = i + 1$ levels.
  - Because we assume that the problem size is a power of 2, the next problem size up after 2^i is 2^{i+1}.
  - A tree for a problem size of 2^{i+1} has one more level than the size-2^i tree $\Rightarrow$ $i + 2$ levels.
  - Since $\lg 2^{i+1} + 1 = i + 2$, we're done with the inductive argument.
- Total cost is sum of costs at each level. Have $\lg n + 1$ levels, each costing cn $\Rightarrow$ total cost is $cn \lg n + cn$.
- Ignore low-order term of cn and constant coefficient $c \Rightarrow \Theta(n \lg n)$.

Solutions for Chapter 2: Getting Started

Solution to Exercise 2.2-2

This solution is also posted publicly

```
SELECTION-SORT(A)
n = A.length
for j = 1 to n − 1
    smallest = j
    for i = j + 1 to n
        if A[i] < A[smallest]
            smallest = i
    exchange A[j] with A[smallest]
```

The algorithm maintains the loop invariant that at the start of each iteration of the outer **for** loop, the subarray $A[1..j-1]$ consists of the $j-1$ smallest elements in the array $A[1..n]$, and this subarray is in sorted order. After the first $n-1$ elements, the subarray $A[1..n-1]$ contains the smallest $n-1$ elements, sorted, and therefore element $A[n]$ must be the largest element.

The running time of the algorithm is $\Theta(n^2)$ for all cases.

Solution to Exercise 2.2-4

This solution is also posted publicly

Modify the algorithm so it tests whether the input satisfies some special-case condition and, if it does, output a pre-computed answer. The best-case running time is generally not a good measure of an algorithm.

Solution to Exercise 2.3-3

The base case is when $n = 2$, and we have $n \lg n = 2 \lg 2 = 2 \cdot 1 = 2$.

For the inductive step, our inductive hypothesis is that $T(n/2) = (n/2)\lg(n/2)$. Then

$$\begin{aligned} T(n) &= 2T(n/2) + n \\ &= 2(n/2)\lg(n/2) + n \\ &= n(\lg n - 1) + n \\ &= n\lg n - n + n \\ &= n\lg n\ , \end{aligned}$$

which completes the inductive proof for exact powers of 2.

Solution to Exercise 2.3-4

Since it takes $\Theta(n)$ time in the worst case to insert $A[n]$ into the sorted array $A[1\,..\,n-1]$, we get the recurrence

$$T(n) = \begin{cases} \Theta(1) & \text{if } n = 1\ , \\ T(n-1) + \Theta(n) & \text{if } n > 1\ . \end{cases}$$

Although the exercise does not ask you to solve this recurrence, its solution is $T(n) = \Theta(n^2)$.

Solution to Exercise 2.3-5

This solution is also posted publicly

Procedure BINARY-SEARCH takes a sorted array A, a value v, and a range $[low\,..\,high]$ of the array, in which we search for the value v. The procedure compares v to the array entry at the midpoint of the range and decides to eliminate half the range from further consideration. We give both iterative and recursive versions, each of which returns either an index i such that $A[i] = v$, or NIL if no entry of $A[low\,..\,high]$ contains the value v. The initial call to either version should have the parameters $A, v, 1, n$.

```
ITERATIVE-BINARY-SEARCH(A, v, low, high)
  while low ≤ high
      mid = ⌊(low + high)/2⌋
      if v == A[mid]
          return mid
      elseif v > A[mid]
          low = mid + 1
      else high = mid − 1
  return NIL
```

```
RECURSIVE-BINARY-SEARCH(A, v, low, high)
  if low > high
      return NIL
  mid = ⌊(low + high)/2⌋
  if v == A[mid]
      return mid
  elseif v > A[mid]
      return RECURSIVE-BINARY-SEARCH(A, v, mid + 1, high)
  else return RECURSIVE-BINARY-SEARCH(A, v, low, mid − 1)
```

Both procedures terminate the search unsuccessfully when the range is empty (i.e., $low > high$) and terminate it successfully if the value v has been found. Based on the comparison of v to the middle element in the searched range, the search continues with the range halved. The recurrence for these procedures is therefore $T(n) = T(n/2) + \Theta(1)$, whose solution is $T(n) = \Theta(\lg n)$.

Solution to Exercise 2.3-6

The **while** loop of lines 5–7 of procedure INSERTION-SORT scans backward through the sorted array $A[1 \mathinner{..} j-1]$ to find the appropriate place for $A[j]$. The hitch is that the loop not only searches for the proper place for $A[j]$, but that it also moves each of the array elements that are bigger than $A[j]$ one position to the right (line 6). These movements can take as much as $\Theta(j)$ time, which occurs when all the $j-1$ elements preceding $A[j]$ are larger than $A[j]$. We can use binary search to improve the running time of the search to $\Theta(\lg j)$, but binary search will have no effect on the running time of moving the elements. Therefore, binary search alone cannot improve the worst-case running time of INSERTION-SORT to $\Theta(n \lg n)$.

Solution to Exercise 2.3-7

The following algorithm solves the problem:

1. Sort the elements in S.
2. Form the set $S' = \{z : z = x - y \text{ for some } y \in S\}$.
3. Sort the elements in S'.
4. Merge the two sorted sets S and S'.
5. There exist two elements in S whose sum is exactly x if and only if the same value appears in consecutive positions in the merged output.

To justify the claim in step 4, first observe that if any value appears twice in the merged output, it must appear in consecutive positions. Thus, we can restate the condition in step 5 as there exist two elements in S whose sum is exactly x if and only if the same value appears twice in the merged output.

Suppose that some value w appears twice. Then w appeared once in S and once in S'. Because w appeared in S', there exists some $y \in S$ such that $w = x - y$, or $x = w + y$. Since $w \in S$, the elements w and y are in S and sum to x.

Conversely, suppose that there are values $w, y \in S$ such that $w + y = x$. Then, since $x - y = w$, the value w appears in S'. Thus, w is in both S and S', and so it will appear twice in the merged output.

Steps 1 and 3 require $\Theta(n \lg n)$ steps. Steps 2, 4, 5, and 6 require $O(n)$ steps. Thus the overall running time is $O(n \lg n)$.

A reader submitted a simpler solution that also runs in $\Theta(n \lg n)$ time. First, sort the elements in S, taking $\Theta(n \lg n)$ time. Then, for each element y in S, perform a binary search in S for $x - y$. Each binary search takes $O(\lg n)$ time, and there are are most n of them, and so the time for all the binary searches is $O(n \lg n)$. The overall running time is $\Theta(n \lg n)$.

Another reader pointed out that since S is a set, if the value $x/2$ appears in S, it appears in S just once, and so $x/2$ cannot be a solution.

Solution to Problem 2-1

[It may be better to assign this problem after covering asymptotic notation in Section 3.1; otherwise part (c) may be too difficult.]

a. Insertion sort takes $\Theta(k^2)$ time per k-element list in the worst case. Therefore, sorting n/k lists of k elements each takes $\Theta(k^2 n/k) = \Theta(nk)$ worst-case time.

b. Just extending the 2-list merge to merge all the lists at once would take $\Theta(n \cdot (n/k)) = \Theta(n^2/k)$ time (n from copying each element once into the result list, n/k from examining n/k lists at each step to select next item for result list).

To achieve $\Theta(n \lg(n/k))$-time merging, we merge the lists pairwise, then merge the resulting lists pairwise, and so on, until there's just one list. The pairwise merging requires $\Theta(n)$ work at each level, since we are still working on n elements, even if they are partitioned among sublists. The number of levels, starting with n/k lists (with k elements each) and finishing with 1 list (with n elements), is $\lceil \lg(n/k) \rceil$. Therefore, the total running time for the merging is $\Theta(n \lg(n/k))$.

c. The modified algorithm has the same asymptotic running time as standard merge sort when $\Theta(nk + n \lg(n/k)) = \Theta(n \lg n)$. The largest asymptotic value of k as a function of n that satisfies this condition is $k = \Theta(\lg n)$.

To see why, first observe that k cannot be more than $\Theta(\lg n)$ (i.e., it can't have a higher-order term than $\lg n$), for otherwise the left-hand expression wouldn't be $\Theta(n \lg n)$ (because it would have a higher-order term than $n \lg n$). So all we need to do is verify that $k = \Theta(\lg n)$ works, which we can do by plugging $k = \lg n$ into $\Theta(nk + n \lg(n/k)) = \Theta(nk + n \lg n - n \lg k)$ to get

$\Theta(n \lg n + n \lg n - n \lg \lg n) = \Theta(2n \lg n - n \lg \lg n)$,

which, by taking just the high-order term and ignoring the constant coefficient, equals $\Theta(n \lg n)$.

d. In practice, k should be the largest list length on which insertion sort is faster than merge sort.

Solution to Problem 2-2

a. We need to show that the elements of A' form a permutation of the elements of A.

b. **Loop invariant:** At the start of each iteration of the **for** loop of lines 2–4, $A[j] = \min\{A[k] : j \le k \le n\}$ and the subarray $A[j\,..\,n]$ is a permutation of the values that were in $A[j\,..\,n]$ at the time that the loop started.

Initialization: Initially, $j = n$, and the subarray $A[j\,..\,n]$ consists of single element $A[n]$. The loop invariant trivially holds.

Maintenance: Consider an iteration for a given value of j. By the loop invariant, $A[j]$ is the smallest value in $A[j\,..\,n]$. Lines 3–4 exchange $A[j]$ and $A[j-1]$ if $A[j]$ is less than $A[j-1]$, and so $A[j-1]$ will be the smallest value in $A[j-1\,..\,n]$ afterward. Since the only change to the subarray $A[j-1\,..\,n]$ is this possible exchange, and the subarray $A[j\,..\,n]$ is a permutation of the values that were in $A[j\,..\,n]$ at the time that the loop started, we see that $A[j-1\,..\,n]$ is a permutation of the values that were in $A[j-1\,..\,n]$ at the time that the loop started. Decrementing j for the next iteration maintains the invariant.

Termination: The loop terminates when j reaches i. By the statement of the loop invariant, $A[i] = \min\{A[k] : i \le k \le n\}$ and $A[i\,..\,n]$ is a permutation of the values that were in $A[i\,..\,n]$ at the time that the loop started.

c. **Loop invariant:** At the start of each iteration of the **for** loop of lines 1–4, the subarray $A[1\,..\,i-1]$ consists of the $i-1$ smallest values originally in $A[1\,..\,n]$, in sorted order, and $A[i\,..\,n]$ consists of the $n-i+1$ remaining values originally in $A[1\,..\,n]$.

Initialization: Before the first iteration of the loop, $i = 1$. The subarray $A[1\,..\,i-1]$ is empty, and so the loop invariant vacuously holds.

Maintenance: Consider an iteration for a given value of i. By the loop invariant, $A[1\,..\,i-1]$ consists of the i smallest values in $A[1\,..\,n]$, in sorted order. Part (b) showed that after executing the **for** loop of lines 2–4, $A[i]$ is the smallest value in $A[i\,..\,n]$, and so $A[1\,..\,i]$ is now the i smallest values originally in $A[1\,..\,n]$, in sorted order. Moreover, since the **for** loop of lines 2–4 permutes $A[i\,..\,n]$, the subarray $A[i+1\,..\,n]$ consists of the $n-i$ remaining values originally in $A[1\,..\,n]$.

Termination: The **for** loop of lines 1–4 terminates when $i = n$, so that $i-1 = n-1$. By the statement of the loop invariant, $A[1\,..\,i-1]$ is the subarray

$A[1\,..\,n-1]$, and it consists of the $n-1$ smallest values originally in $A[1\,..\,n]$, in sorted order. The remaining element must be the largest value in $A[1\,..\,n]$, and it is in $A[n]$. Therefore, the entire array $A[1\,..\,n]$ is sorted.

Note: Tn the second edition, the **for** loop of lines 1–4 had an upper bound of $A.length$. The last iteration of the outer **for** loop would then result in no iterations of the inner **for** loop of lines 1–4, but the termination argument would simplify: $A[1\,..\,i-1]$ would be the entire array $A[1\,..\,n]$, which, by the loop invariant, is sorted.

d. The running time depends on the number of iterations of the **for** loop of lines 2–4. For a given value of i, this loop makes $n-i$ iterations, and i takes on the values $1, 2, \ldots, n-1$. The total number of iterations, therefore, is

$$\begin{aligned}\sum_{i=1}^{n-1}(n-i) &= \sum_{i=1}^{n-1} n - \sum_{i=1}^{n-1} i \\ &= n(n-1) - \frac{n(n-1)}{2} \\ &= \frac{n(n-1)}{2} \\ &= \frac{n^2}{2} - \frac{n}{2}\,.\end{aligned}$$

Thus, the running time of bubblesort is $\Theta(n^2)$ in all cases. The worst-case running time is the same as that of insertion sort.

Solution to Problem 2-4

This solution is also posted publicly

a. The inversions are $(1,5), (2,5), (3,4), (3,5), (4,5)$. (Remember that inversions are specified by indices rather than by the values in the array.)

b. The array with elements from $\{1, 2, \ldots, n\}$ with the most inversions is $\langle n, n-1, n-2, \ldots, 2, 1\rangle$. For all $1 \le i < j \le n$, there is an inversion (i, j). The number of such inversions is $\binom{n}{2} = n(n-1)/2$.

c. Suppose that the array A starts out with an inversion (k, j). Then $k < j$ and $A[k] > A[j]$. At the time that the outer **for** loop of lines 1–8 sets $key = A[j]$, the value that started in $A[k]$ is still somewhere to the left of $A[j]$. That is, it's in $A[i]$, where $1 \le i < j$, and so the inversion has become (i, j). Some iteration of the **while** loop of lines 5–7 moves $A[i]$ one position to the right. Line 8 will eventually drop *key* to the left of this element, thus eliminating the inversion. Because line 5 moves only elements that are greater than *key*, it moves only elements that correspond to inversions. In other words, each iteration of the **while** loop of lines 5–7 corresponds to the elimination of one inversion.

d. We follow the hint and modify merge sort to count the number of inversions in $\Theta(n \lg n)$ time.

To start, let us define a ***merge-inversion*** as a situation within the execution of merge sort in which the MERGE procedure, after copying $A[p..q]$ to L and $A[q+1..r]$ to R, has values x in L and y in R such that $x > y$. Consider an inversion (i, j), and let $x = A[i]$ and $y = A[j]$, so that $i < j$ and $x > y$. We claim that if we were to run merge sort, there would be exactly one merge-inversion involving x and y. To see why, observe that the only way in which array elements change their positions is within the MERGE procedure. Moreover, since MERGE keeps elements within L in the same relative order to each other, and correspondingly for R, the only way in which two elements can change their ordering relative to each other is for the greater one to appear in L and the lesser one to appear in R. Thus, there is at least one merge-inversion involving x and y. To see that there is exactly one such merge-inversion, observe that after any call of MERGE that involves both x and y, they are in the same sorted subarray and will therefore both appear in L or both appear in R in any given call thereafter. Thus, we have proven the claim.

We have shown that every inversion implies one merge-inversion. In fact, the correspondence between inversions and merge-inversions is one-to-one. Suppose we have a merge-inversion involving values x and y, where x originally was $A[i]$ and y was originally $A[j]$. Since we have a merge-inversion, $x > y$. And since x is in L and y is in R, x must be within a subarray preceding the subarray containing y. Therefore x started out in a position i preceding y's original position j, and so (i, j) is an inversion.

Having shown a one-to-one correspondence between inversions and merge-inversions, it suffices for us to count merge-inversions.

Consider a merge-inversion involving y in R. Let z be the smallest value in L that is greater than y. At some point during the merging process, z and y will be the "exposed" values in L and R, i.e., we will have $z = L[i]$ and $y = R[j]$ in line 13 of MERGE. At that time, there will be merge-inversions involving y and $L[i], L[i+1], L[i+2], \ldots, L[n_1]$, and these $n_1 - i + 1$ merge-inversions will be the only ones involving y. Therefore, we need to detect the first time that z and y become exposed during the MERGE procedure and add the value of $n_1 - i + 1$ at that time to our total count of merge-inversions.

The following pseudocode, modeled on merge sort, works as we have just described. It also sorts the array A.

```
COUNT-INVERSIONS(A, p, r)
  inversions = 0
  if p < r
      q = ⌊(p + r)/2⌋
      inversions = inversions + COUNT-INVERSIONS(A, p, q)
      inversions = inversions + COUNT-INVERSIONS(A, q + 1, r)
      inversions = inversions + MERGE-INVERSIONS(A, p, q, r)
  return inversions
```

```
MERGE-INVERSIONS(A, p, q, r)
n1 = q − p + 1
n2 = r − q
let L[1 .. n1 + 1] and R[1 .. n2 + 1] be new arrays
for i = 1 to n1
    L[i] = A[p + i − 1]
for j = 1 to n2
    R[j] = A[q + j]
L[n1 + 1] = ∞
R[n2 + 1] = ∞
i = 1
j = 1
inversions = 0
for k = p to r
    if R[j] < L[i]
        inversions = inversions + n1 − i + 1
        A[k] = R[j]
        j = j + 1
    else A[k] = L[i]
        i = i + 1
return inversions
```

The initial call is COUNT-INVERSIONS$(A, 1, n)$.

In MERGE-INVERSIONS, whenever $R[j]$ is exposed and a value greater than $R[j]$ becomes exposed in the L array, we increase *inversions* by the number of remaining elements in L. Then because $R[j+1]$ becomes exposed, $R[j]$ can never be exposed again. We don't have to worry about merge-inversions involving the sentinel ∞ in R, since no value in L will be greater than ∞.

Since we have added only a constant amount of additional work to each procedure call and to each iteration of the last **for** loop of the merging procedure, the total running time of the above pseudocode is the same as for merge sort: $\Theta(n \lg n)$.

Lecture Notes for Chapter 3: Growth of Functions

Chapter 3 overview

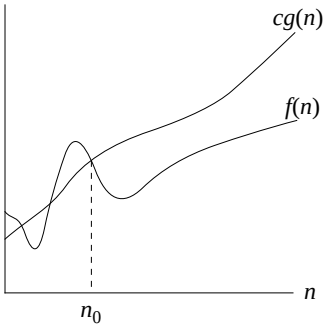
- A way to describe behavior of functions *in the limit*. We're studying ***asymptotic*** efficiency.
- Describe *growth* of functions.
- Focus on what's important by abstracting away low-order terms and constant factors.
- How we indicate running times of algorithms.
- A way to compare "sizes" of functions:

O	$\approx$	$\leq$
Ω	$\approx$	$\geq$
Θ	$\approx$	$=$
o	$\approx$	$<$
ω	$\approx$	$>$

Asymptotic notation

O-notation

$O(g(n)) = \{f(n)$: there exist positive constants c and n_0 such that $0 \leq f(n) \leq cg(n)$ for all $n \geq n_0\}$.



$g(n)$ is an ***asymptotic upper bound*** for $f(n)$.

If $f(n) \in O(g(n))$, we write $f(n) = O(g(n))$ (will precisely explain this soon).

Example

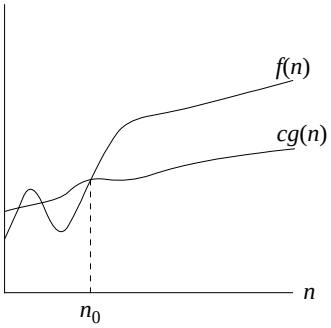
$2n^2 = O(n^3)$, with $c = 1$ and $n_0 = 2$.

Examples of functions in $O(n^2)$:

n^2
$n^2 + n$
$n^2 + 1000n$
$1000n^2 + 1000n$
Also,
n
$n/1000$
$n^{1.99999}$
$n^2/\lg\lg\lg n$

Ω-notation

$\Omega(g(n)) = \{f(n)$: there exist positive constants c and n_0 such that $0 \le cg(n) \le f(n)$ for all $n \ge n_0\}$.



$g(n)$ is an ***asymptotic lower bound*** for $f(n)$.

Example

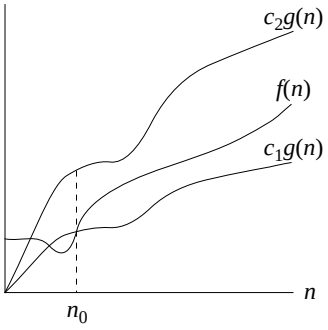
$\sqrt{n} = \Omega(\lg n)$, with $c = 1$ and $n_0 = 16$.

Examples of functions in $\Omega(n^2)$:

n^2
$n^2 + n$
$n^2 - n$
$1000n^2 + 1000n$
$1000n^2 - 1000n$
Also,
n^3
$n^{2.00001}$
$n^2 \lg\lg\lg n$
2^{2^n}

Θ-notation

$\Theta(g(n)) = \{f(n)$: there exist positive constants c_1, c_2, and n_0 such that
$0 \le c_1 g(n) \le f(n) \le c_2 g(n)$ for all $n \ge n_0\}$.



$g(n)$ is an ***asymptotically tight bound*** for $f(n)$.

Example

$n^2/2 - 2n = \Theta(n^2)$, with $c_1 = 1/4$, $c_2 = 1/2$, and $n_0 = 8$.

Theorem

$f(n) = \Theta(g(n))$ if and only if $f = O(g(n))$ and $f = \Omega(g(n))$.

Leading constants and low-order terms don't matter.

Asymptotic notation in equations

When on right-hand side

$O(n^2)$ stands for some anonymous function in the set $O(n^2)$.

$2n^2 + 3n + 1 = 2n^2 + \Theta(n)$ means $2n^2 + 3n + 1 = 2n^2 + f(n)$ *for some* $f(n) \in \Theta(n)$. In particular, $f(n) = 3n + 1$.

By the way, we interpret # of anonymous functions as = # of times the asymptotic notation appears:

$\sum_{i=1}^{n} O(i)$ OK: 1 anonymous function

$O(1) + O(2) + \cdots + O(n)$ not OK: n hidden constants
$\Rightarrow$ no clean interpretation

When on left-hand side

No matter how the anonymous functions are chosen on the left-hand side, there is a way to choose the anonymous functions on the right-hand side to make the equation valid.

Interpret $2n^2 + \Theta(n) = \Theta(n^2)$ as meaning *for all* functions $f(n) \in \Theta(n)$, there exists a function $g(n) \in \Theta(n^2)$ such that $2n^2 + f(n) = g(n)$.

Can chain together:

$$\begin{aligned} 2n^2 + 3n + 1 &= 2n^2 + \Theta(n) \\ &= \Theta(n^2) . \end{aligned}$$

Interpretation:

- First equation: There exists $f(n) \in \Theta(n)$ such that $2n^2 + 3n + 1 = 2n^2 + f(n)$.
- Second equation: For all $g(n) \in \Theta(n)$ (such as the $f(n)$ used to make the first equation hold), there exists $h(n) \in \Theta(n^2)$ such that $2n^2 + g(n) = h(n)$.

o-notation

$o(g(n)) = \{f(n) :$ for all constants $c > 0$, there exists a constant $n_0 > 0$ such that $0 \le f(n) < cg(n)$ for all $n \ge n_0\}$.

Another view, probably easier to use: $\lim_{n\to\infty} \dfrac{f(n)}{g(n)} = 0$.

$n^{1.9999} = o(n^2)$
$n^2/\lg n = o(n^2)$
$n^2 \ne o(n^2)$ (just like $2 \not< 2$)
$n^2/1000 \ne o(n^2)$

ω-notation

$\omega(g(n)) = \{f(n) :$ for all constants $c > 0$, there exists a constant $n_0 > 0$ such that $0 \le cg(n) < f(n)$ for all $n \ge n_0\}$.

Another view, again, probably easier to use: $\lim_{n\to\infty} \dfrac{f(n)}{g(n)} = \infty$.

$n^{2.0001} = \omega(n^2)$
$n^2 \lg n = \omega(n^2)$
$n^2 \ne \omega(n^2)$

Comparisons of functions

Relational properties:

Transitivity:
$f(n) = \Theta(g(n))$ and $g(n) = \Theta(h(n)) \Rightarrow f(n) = \Theta(h(n))$.
Same for O, Ω, o, and ω.

Reflexivity:
$f(n) = \Theta(f(n))$.
Same for O and Ω.

Symmetry:
$f(n) = \Theta(g(n))$ if and only if $g(n) = \Theta(f(n))$.

Transpose symmetry:
$f(n) = O(g(n))$ if and only if $g(n) = \Omega(f(n))$.
$f(n) = o(g(n))$ if and only if $g(n) = \omega(f(n))$.

Comparisons:

- $f(n)$ is ***asymptotically smaller*** than $g(n)$ if $f(n) = o(g(n))$.
- $f(n)$ is ***asymptotically larger*** than $g(n)$ if $f(n) = \omega(g(n))$.

No trichotomy. Although intuitively, we can liken O to $\leq$, Ω to $\geq$, etc., unlike real numbers, where $a < b$, $a = b$, or $a > b$, we might not be able to compare functions.

Example: $n^{1+\sin n}$ and n, since $1 + \sin n$ oscillates between 0 and 2.

Standard notations and common functions

[You probably do not want to use lecture time going over all the definitions and properties given in Section 3.2, but it might be worth spending a few minutes of lecture time on some of the following.]

Monotonicity

- $f(n)$ is ***monotonically increasing*** if $m \leq n \Rightarrow f(m) \leq f(n)$.
- $f(n)$ is ***monotonically decreasing*** if $m \geq n \Rightarrow f(m) \geq f(n)$.
- $f(n)$ is ***strictly increasing*** if $m < n \Rightarrow f(m) < f(n)$.
- $f(n)$ is ***strictly decreasing*** if $m > n \Rightarrow f(m) > f(n)$.

Exponentials

Useful identities:

$$\begin{aligned} a^{-1} &= 1/a \ , \\ (a^m)^n &= a^{mn} \ , \\ a^m a^n &= a^{m+n} \ . \end{aligned}$$

Can relate rates of growth of polynomials and exponentials: for all real constants a and b such that $a > 1$,

$$\lim_{n \to \infty} \frac{n^b}{a^n} = 0 \ ,$$

which implies that $n^b = o(a^n)$.

A suprisingly useful inequality: for all real x,

$$e^x \geq 1 + x \ .$$

As x gets closer to 0, e^x gets closer to $1 + x$.

Logarithms

Notations:

$$\begin{aligned}
\lg n &= \log_2 n \quad \text{(binary logarithm)} ,\\
\ln n &= \log_e n \quad \text{(natural logarithm)} ,\\
\lg^k n &= (\lg n)^k \quad \text{(exponentiation)} ,\\
\lg\lg n &= \lg(\lg n) \quad \text{(composition)} .
\end{aligned}$$

Logarithm functions apply only to the next term in the formula, so that $\lg n + k$ means $(\lg n) + k$, and *not* $\lg(n + k)$.

In the expression $\log_b a$:

- If we hold b constant, then the expression is strictly increasing as a increases.
- If we hold a constant, then the expression is strictly decreasing as b increases.

Useful identities for all real $a > 0$, $b > 0$, $c > 0$, and n, and where logarithm bases are not 1:

$$\begin{aligned}
a &= b^{\log_b a} ,\\
\log_c(ab) &= \log_c a + \log_c b ,\\
\log_b a^n &= n \log_b a ,\\
\log_b a &= \frac{\log_c a}{\log_c b} ,\\
\log_b(1/a) &= -\log_b a ,\\
\log_b a &= \frac{1}{\log_a b} ,\\
a^{\log_b c} &= c^{\log_b a} .
\end{aligned}$$

Changing the base of a logarithm from one constant to another only changes the value by a constant factor, so we usually don't worry about logarithm bases in asymptotic notation. Convention is to use lg within asymptotic notation, unless the base actually matters.

Just as polynomials grow more slowly than exponentials, logarithms grow more slowly than polynomials. In $\lim_{n\to\infty} \frac{n^b}{a^n} = 0$, substitute $\lg n$ for n and 2^a for a:

$$\lim_{n\to\infty} \frac{\lg^b n}{(2^a)^{\lg n}} = \lim_{n\to\infty} \frac{\lg^b n}{n^a} = 0 ,$$

implying that $\lg^b n = o(n^a)$.

Factorials

$n! = 1 \cdot 2 \cdot 3 \cdot n$. Special case: $0! = 1$.

Can use ***Stirling's approximation***,

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \Theta\left(\frac{1}{n}\right)\right) ,$$

to derive that $\lg(n!) = \Theta(n \lg n)$.

Solutions for Chapter 3: Growth of Functions

Solution to Exercise 3.1-1

First, let's clarify what the function $\max(f(n), g(n))$ is. Let's define the function $h(n) = \max(f(n), g(n))$. Then

$$h(n) = \begin{cases} f(n) & \text{if } f(n) \geq g(n) , \\ g(n) & \text{if } f(n) < g(n) . \end{cases}$$

Since $f(n)$ and $g(n)$ are asymptotically nonnegative, there exists n_0 such that $f(n) \geq 0$ and $g(n) \geq 0$ for all $n \geq n_0$. Thus for $n \geq n_0$, $f(n) + g(n) \geq f(n) \geq 0$ and $f(n) + g(n) \geq g(n) \geq 0$. Since for any particular n, $h(n)$ is either $f(n)$ or $g(n)$, we have $f(n) + g(n) \geq h(n) \geq 0$, which shows that $h(n) = \max(f(n), g(n)) \leq c_2(f(n) + g(n))$ for all $n \geq n_0$ (with $c_2 = 1$ in the definition of Θ).

Similarly, since for any particular n, $h(n)$ is the larger of $f(n)$ and $g(n)$, we have for all $n \geq n_0$, $0 \leq f(n) \leq h(n)$ and $0 \leq g(n) \leq h(n)$. Adding these two inequalities yields $0 \leq f(n) + g(n) \leq 2h(n)$, or equivalently $0 \leq (f(n) + g(n))/2 \leq h(n)$, which shows that $h(n) = \max(f(n), g(n)) \geq c_1(f(n)+g(n))$ for all $n \geq n_0$ (with $c_1 = 1/2$ in the definition of Θ).

Solution to Exercise 3.1-2

This solution is also posted publicly

To show that $(n + a)^b = \Theta(n^b)$, we want to find constants $c_1, c_2, n_0 > 0$ such that $0 \leq c_1 n^b \leq (n + a)^b \leq c_2 n^b$ for all $n \geq n_0$.

Note that

$$\begin{aligned} n + a &\leq n + |a| \\ &\leq 2n \qquad \text{when } |a| \leq n , \end{aligned}$$

and

$$\begin{aligned} n + a &\geq n - |a| \\ &\geq \frac{1}{2}n \qquad \text{when } |a| \leq \tfrac{1}{2}n . \end{aligned}$$

Thus, when $n \geq 2|a|$,

$$0 \le \frac{1}{2}n \le n + a \le 2n \ .$$

Since $b > 0$, the inequality still holds when all parts are raised to the power b:

$$0 \le \left(\frac{1}{2}n\right)^b \le (n + a)^b \le (2n)^b \ ,$$

$$0 \le \left(\frac{1}{2}\right)^b n^b \le (n + a)^b \le 2^b n^b \ .$$

Thus, $c_1 = (1/2)^b$, $c_2 = 2^b$, and $n_0 = 2\,|a|$ satisfy the definition.

Solution to Exercise 3.1-3
This solution is also posted publicly

Let the running time be $T(n)$. $T(n) \ge O(n^2)$ means that $T(n) \ge f(n)$ for some function $f(n)$ in the set $O(n^2)$. This statement holds for any running time $T(n)$, since the function $g(n) = 0$ for all n is in $O(n^2)$, and running times are always nonnegative. Thus, the statement tells us nothing about the running time.

Solution to Exercise 3.1-4
This solution is also posted publicly

$2^{n+1} = O(2^n)$, but $2^{2n} \ne O(2^n)$.

To show that $2^{n+1} = O(2^n)$, we must find constants $c, n_0 > 0$ such that

$0 \le 2^{n+1} \le c \cdot 2^n$ for all $n \ge n_0$.

Since $2^{n+1} = 2 \cdot 2^n$ for all n, we can satisfy the definition with $c = 2$ and $n_0 = 1$.

To show that $2^{2n} \ne O(2^n)$, assume there exist constants $c, n_0 > 0$ such that

$0 \le 2^{2n} \le c \cdot 2^n$ for all $n \ge n_0$.

Then $2^{2n} = 2^n \cdot 2^n \le c \cdot 2^n \Rightarrow 2^n \le c$. But no constant is greater than all 2^n, and so the assumption leads to a contradiction.

Solution to Exercise 3.1-8

$\Omega(g(n,m)) = \{f(n,m)$: there exist positive constants c, n_0, and m_0
such that $0 \le cg(n,m) \le f(n,m)$
for all $n \ge n_0$ or $m \ge m_0\}$.

$\Theta(g(n,m)) = \{f(n,m)$: there exist positive constants c_1, c_2, n_0, and m_0
such that $0 \le c_1 g(n,m) \le f(n,m) \le c_2 g(n,m)$
for all $n \ge n_0$ or $m \ge m_0\}$.

Solution to Exercise 3.2-4

This solution is also posted publicly

$\lceil \lg n \rceil !$ is not polynomially bounded, but $\lceil \lg \lg n \rceil !$ is.

Proving that a function $f(n)$ is polynomially bounded is equivalent to proving that $\lg(f(n)) = O(\lg n)$ for the following reasons.

- If f is polynomially bounded, then there exist constants c, k, n_0 such that for all $n \geq n_0$, $f(n) \leq cn^k$. Hence, $\lg(f(n)) \leq kc \lg n$, which, since c and k are constants, means that $\lg(f(n)) = O(\lg n)$.
- Similarly, if $\lg(f(n)) = O(\lg n)$, then f is polynomially bounded.

In the following proofs, we will make use of the following two facts:

1. $\lg(n!) = \Theta(n \lg n)$ (by equation (3.19)).
2. $\lceil \lg n \rceil = \Theta(\lg n)$, because
   - $\lceil \lg n \rceil \geq \lg n$
   - $\lceil \lg n \rceil < \lg n + 1 \leq 2 \lg n$ for all $n \geq 2$

$$\begin{aligned} \lg(\lceil \lg n \rceil !) &= \Theta(\lceil \lg n \rceil \lg \lceil \lg n \rceil) \\ &= \Theta(\lg n \lg \lg n) \\ &= \omega(\lg n) \ . \end{aligned}$$

Therefore, $\lg(\lceil \lg n \rceil !) \neq O(\lg n)$, and so $\lceil \lg n \rceil !$ is not polynomially bounded.

$$\begin{aligned} \lg(\lceil \lg \lg n \rceil !) &= \Theta(\lceil \lg \lg n \rceil \lg \lceil \lg \lg n \rceil) \\ &= \Theta(\lg \lg n \lg \lg \lg n) \\ &= o((\lg \lg n)^2) \\ &= o(\lg^2(\lg n)) \\ &= o(\lg n) \ . \end{aligned}$$

The last step above follows from the property that any polylogarithmic function grows more slowly than any positive polynomial function, i.e., that for constants $a, b > 0$, we have $\lg^b n = o(n^a)$. Substitute $\lg n$ for n, 2 for b, and 1 for a, giving $\lg^2(\lg n) = o(\lg n)$.

Therefore, $\lg(\lceil \lg \lg n \rceil !) = O(\lg n)$, and so $\lceil \lg \lg n \rceil !$ is polynomially bounded.

Solution to Exercise 3.2-5

$\lg^*(\lg n)$ is asymptotically larger because $\lg^*(\lg n) = \lg^* n - 1$.

Solution to Exercise 3.2-6

Both ϕ^2 and $\phi + 1$ equal $(3 + \sqrt{5})/2$, and both $\hat{\phi}^2$ and $\hat{\phi} + 1$ equal $(3 - \sqrt{5})/2$.

Solution to Exercise 3.2-7

We have two base cases: $i = 0$ and $i = 1$. For $i = 0$, we have

$$\begin{aligned}\frac{\phi^0 - \hat{\phi}^0}{\sqrt{5}} &= \frac{1 - 1}{\sqrt{5}} \\ &= 0 \\ &= F_0 \ ,\end{aligned}$$

and for $i = 1$, we have

$$\begin{aligned}\frac{\phi^1 - \hat{\phi}^1}{\sqrt{5}} &= \frac{(1 + \sqrt{5}) - (1 - \sqrt{5})}{2\sqrt{5}} \\ &= \frac{2\sqrt{5}}{2\sqrt{5}} \\ &= 1 \\ &= F_1 \ .\end{aligned}$$

For the inductive case, the inductive hypothesis is that $F_{i-1} = (\phi^{i-1} - \hat{\phi}^{i-1})/\sqrt{5}$ and $F_{i-2} = (\phi^{i-2} - \hat{\phi}^{i-2})/\sqrt{5}$. We have

$$\begin{aligned}F_i &= F_{i-1} + F_{i-2} && \text{(equation (3.22))} \\ &= \frac{\phi^{i-1} - \hat{\phi}^{i-1}}{\sqrt{5}} + \frac{\phi^{i-2} - \hat{\phi}^{i-2}}{\sqrt{5}} && \text{(inductive hypothesis)} \\ &= \frac{\phi^{i-2}(\phi + 1) - \hat{\phi}^{i-2}(\hat{\phi} + 1)}{\sqrt{5}} \\ &= \frac{\phi^{i-2}\phi^2 - \hat{\phi}^{i-2}\hat{\phi}^2}{\sqrt{5}} && \text{(Exercise 3.2-6)} \\ &= \frac{\phi^i - \hat{\phi}^i}{\sqrt{5}} \ .\end{aligned}$$

Solution to Problem 3-3

a. Here is the ordering, where functions on the same line are in the same equivalence class, and those higher on the page are Ω of those below them:

$2^{2^{n+1}}$	
2^{2^n}	
$(n+1)!$	
$n!$	see justification 7
e^n	see justification 1
$n \cdot 2^n$	
2^n	
$(3/2)^n$	
$(\lg n)^{\lg n} = n^{\lg \lg n}$	see identity 1
$(\lg n)!$	see justifications 2, 8
n^3	
$n^2 = 4^{\lg n}$	see identity 2
$n \lg n$ and $\lg(n!)$	see justification 6
$n = 2^{\lg n}$	see identity 3
$(\sqrt{2})^{\lg n} (= \sqrt{n})$	see identity 6, justification 3
$2^{\sqrt{2 \lg n}}$	see identity 5, justification 4
$\lg^2 n$	
$\ln n$	
$\sqrt{\lg n}$	
$\ln \ln n$	see justification 5
$2^{\lg^* n}$	
$\lg^* n$ and $\lg^*(\lg n)$	see identity 7
$\lg(\lg^* n)$	
$n^{1/\lg n} (= 2)$ and 1	see identity 4

Much of the ranking is based on the following properties:

- Exponential functions grow faster than polynomial functions, which grow faster than polylogarithmic functions.
- The base of a logarithm doesn't matter asymptotically, but the base of an exponential and the degree of a polynomial do matter.

We have the following *identities*:

1. $(\lg n)^{\lg n} = n^{\lg \lg n}$ because $a^{\log_b c} = c^{\log_b a}$.
2. $4^{\lg n} = n^2$ because $a^{\log_b c} = c^{\log_b a}$.
3. $2^{\lg n} = n$.
4. $2 = n^{1/\lg n}$ by raising identity 3 to the power $1/\lg n$.
5. $2^{\sqrt{2 \lg n}} = n^{\sqrt{2/\lg n}}$ by raising identity 4 to the power $\sqrt{2 \lg n}$.
6. $\left(\sqrt{2}\right)^{\lg n} = \sqrt{n}$ because $\left(\sqrt{2}\right)^{\lg n} = 2^{(1/2)\lg n} = 2^{\lg \sqrt{n}} = \sqrt{n}$.
7. $\lg^*(\lg n) = (\lg^* n) - 1$.

The following *justifications* explain some of the rankings:

1. $e^n = 2^n (e/2)^n = \omega(n 2^n)$, since $(e/2)^n = \omega(n)$.
2. $(\lg n)! = \omega(n^3)$ by taking logs: $\lg(\lg n)! = \Theta(\lg n \lg \lg n)$ by Stirling's approximation, $\lg(n^3) = 3 \lg n$. $\lg \lg n = \omega(3)$.

3. $(\sqrt{2})^{\lg n} = \omega\left(2^{\sqrt{2\lg n}}\right)$ by taking logs: $\lg(\sqrt{2})^{\lg n} = (1/2)\lg n$, $\lg 2^{\sqrt{2\lg n}} = \sqrt{2\lg n}$. $(1/2)\lg n = \omega(\sqrt{2\lg n})$.
4. $2^{\sqrt{2\lg n}} = \omega(\lg^2 n)$ by taking logs: $\lg 2^{\sqrt{2\lg n}} = \sqrt{2\lg n}$, $\lg\lg^2 n = 2\lg\lg n$. $\sqrt{2\lg n} = \omega(2\lg\lg n)$.
5. $\ln\ln n = \omega(2^{\lg^* n})$ by taking logs: $\lg 2^{\lg^* n} = \lg^* n$. $\lg\ln\ln n = \omega(\lg^* n)$.
6. $\lg(n!) = \Theta(n\lg n)$ (equation (3.19)).
7. $n! = \Theta(n^{n+1/2}e^{-n})$ by dropping constants and low-order terms in equation (3.18).
8. $(\lg n)! = \Theta((\lg n)^{\lg n+1/2}e^{-\lg n})$ by substituting $\lg n$ for n in the previous justification. $(\lg n)! = \Theta((\lg n)^{\lg n+1/2}n^{-\lg e})$ because $a^{\log_b c} = c^{\log_b a}$.

b. The following $f(n)$ is nonnegative, and for all functions $g_i(n)$ in part (a), $f(n)$ is neither $O(g_i(n))$ nor $\Omega(g_i(n))$.

$$f(n) = \begin{cases} 2^{2^{n+2}} & \text{if } n \text{ is even}, \\ 0 & \text{if } n \text{ is odd}. \end{cases}$$

Lecture Notes for Chapter 4: Divide-and-Conquer

Chapter 4 overview

Recall the divide-and-conquer paradigm, which we used for merge sort:

Divide the problem into a number of subproblems that are smaller instances of the same problem.

Conquer the subproblems by solving them recursively.
Base case: If the subproblems are small enough, just solve them by brute force.

Combine the subproblem solutions to give a solution to the original problem.

We look at two more algorithms based on divide-and-conquer.

Analyzing divide-and-conquer algorithms

Use a recurrence to characterize the running time of a divide-and-conquer algorithm. Solving the recurrence gives us the asymptotic running time.

A ***recurrence*** is a function is defined in terms of

- one or more base cases, and
- itself, with smaller arguments.

Examples

- $T(n) = \begin{cases} 1 & \text{if } n = 1\,, \\ T(n-1) + 1 & \text{if } n > 1\,. \end{cases}$
  Solution: $T(n) = n$.
- $T(n) = \begin{cases} 1 & \text{if } n = 1\,, \\ 2T(n/2) + n & \text{if } n \geq 1\,. \end{cases}$
  Solution: $T(n) = n \lg n + n$.
- $T(n) = \begin{cases} 0 & \text{if } n = 2\,, \\ T(\sqrt{n}) + 1 & \text{if } n > 2\,. \end{cases}$
  Solution: $T(n) = \lg \lg n$.

- $T(n) = \begin{cases} 1 & \text{if } n = 1 , \\ T(n/3) + T(2n/3) + n & \text{if } n > 1 . \end{cases}$

  Solution: $T(n) = \Theta(n \lg n)$.

[The notes for this chapter are fairly brief because we teach recurrences in much greater detail in a separate discrete math course.]

Many technical issues:

- Floors and ceilings

  [Floors and ceilings can easily be removed and don't affect the solution to the recurrence. They are better left to a discrete math course.]
- Exact vs. asymptotic functions
- Boundary conditions

In algorithm analysis, we usually express both the recurrence and its solution using asymptotic notation.

- Example: $T(n) = 2T(n/2) + \Theta(n)$, with solution $T(n) = \Theta(n \lg n)$.
- The boundary conditions are usually expressed as "$T(n) = O(1)$ for sufficiently small n."
- When we desire an exact, rather than an asymptotic, solution, we need to deal with boundary conditions.
- In practice, we just use asymptotics most of the time, and we ignore boundary conditions.

[In my course, there are only two acceptable ways of solving recurrences: the substitution method and the master method. Unless the recursion tree is carefully accounted for, I do not accept it as a proof of a solution, though I certainly accept a recursion tree as a way to generate a guess for substitution method. You may choose to allow recursion trees as proofs in your course, in which case some of the substitution proofs in the solutions for this chapter become recursion trees.

I also never use the iteration method, which had appeared in the first edition of Introduction to Algorithms. I find that it is too easy to make an error in parenthesization, and that recursion trees give a better intuitive idea than iterating the recurrence of how the recurrence progresses.]

Maximum-subarray problem

Input: An array $A[1 \,..\, n]$ of numbers. *[Assume that some of the numbers are negative, because this problem is trivial when all numbers are nonnegative.]*

Output: Indices i and j such that $A[i \,..\, j]$ has the greatest sum of any nonempty, contiguous subarray of A, along with the sum of the values in $A[i \,..\, j]$.

Scenario

- You have the prices that a stock traded at over a period of n consecutive days.
- When should you have bought the stock? When should you have sold the stock?
- Even though it's in retrospect, you can yell at your stockbroker for not recommending these buy and sell dates.

To convert to a maximum-subarray problem, let

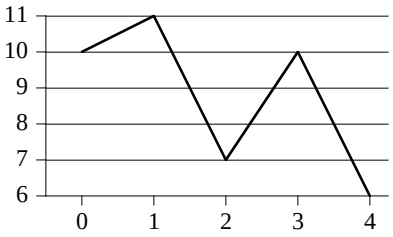
$A[i] =$ (price after day i) $-$ (price after day $(i-1)$) .

[Assuming that we start with a price after day 0, i.e., just before day 1.] Then the nonempty, contiguous subarray with the greatest sum brackets the days that you should have held the stock.

If the maximum subarray is $A[i \,..\, j]$, then should have bought just before day i (i.e., just after day $(i-1)$) and sold just after day j.

Why do we need to find the maximum subarray? Why not just "buy low, sell high"?

- Lowest price might occur *after* the highest price.
- But wouldn't the optimal strategy involve buying at the lowest price *or* selling at the highest price?
- Not necessarily:



Maximum profit is \$3 per share, from buying after day 2 and selling after day 3. Yet lowest price occurs after day 4 and highest occurs after day 1.

Can solve by brute force: check all $\binom{n}{2} = \Theta(n^2)$ subarrays. Can organize the computation so that each subarray $A[i \,..\, j]$ takes $O(1)$ time, given that you've computed $A[i \,..\, j-1]$, so that the brute-force solution takes $\Theta(n^2)$ time.

Solving by divide-and-conquer

Use divide-and-conquer to solve in $O(n \lg n)$ time.

[Maximum subarray might not be unique, though its value is, so we speak of a *maximum subarray, rather than* the *maximum subarray.]*

Subproblem: Find a maximum subarray of $A[low \,..\, high]$.
In original call, $low = 1$, $high = n$.

Divide the subarray into two subarrays of as equal size as possible. Find the midpoint *mid* of the subarrays, and consider the subarrays $A[low\,..\,mid]$ and $A[mid+1\,..\,high]$.

Conquer by finding a maximum subarrays of $A[low\,..\,mid]$ and $A[mid+1\,..\,high]$.

Combine by finding a maximum subarray that crosses the midpoint, and using the best solution out of the three (the subarray crossing the midpoint and the two solutions found in the conquer step).

This strategy works because any subarray must either lie entirely on one side of the midpoint or cross the midpoint.

Finding the maximum subarray that crosses the midpoint

Not a smaller instance of the original problem: has the added restriction that the subarray must cross the midpoint.

Again, could use brute force. If size of $A[low\,..\,high]$ is n, would have $n/2$ choices for left endpoint and $n/2$ choices right endpoint, so would have $\Theta(n^2)$ combinations altogether.

Can solve in linear time.

- Any subarray crossing the midpoint $A[mid]$ is made of two subarrays $A[i\,..\,mid]$ and $A[mid+1\,..\,j]$, where $low \le i \le mid$ and $mid < j \le high$.
- Find maximum subarrays of the form $A[i\,..\,mid]$ and $A[mid+1\,..\,j]$ and then combine them.

Procedure to take array A and indices *low*, *mid*, *high* and return a tuple giving indices of maximum subarray that crosses the midpoint, along with the sum in this maximum subarray:

```
FIND-MAX-CROSSING-SUBARRAY(A, low, mid, high)
  // Find a maximum subarray of the form A[i..mid].
  left-sum = −∞
  sum = 0
  for i = mid downto low
      sum = sum + A[i]
      if sum > left-sum
          left-sum = sum
          max-left = i
  // Find a maximum subarray of the form A[mid+1..j].
  right-sum = −∞
  sum = 0
  for j = mid + 1 to high
      sum = sum + A[j]
      if sum > right-sum
          right-sum = sum
          max-right = j
  // Return the indices and the sum of the two subarrays.
  return (max-left, max-right, left-sum + right-sum)
```

Time: The two loops together consider each index in the range $low, \ldots, high$ exactly once, and each iteration takes $\Theta(1)$ time $\Rightarrow$ procedure takes $\Theta(n)$ time.

Divide-and-conquer procedure for the maximum-subarray problem

FIND-MAXIMUM-SUBARRAY$(A, low, high)$

```
if high == low
    return (low, high, A[low])          // base case: only one element
else mid = ⌊(low + high)/2⌋
    (left-low, left-high, left-sum) =
        FIND-MAXIMUM-SUBARRAY(A, low, mid)
    (right-low, right-high, right-sum) =
        FIND-MAXIMUM-SUBARRAY(A, mid + 1, high)
    (cross-low, cross-high, cross-sum) =
        FIND-MAX-CROSSING-SUBARRAY(A, low, mid, high)
    if left-sum ≥ right-sum and left-sum ≥ cross-sum
        return (left-low, left-high, left-sum)
    elseif right-sum ≥ left-sum and right-sum ≥ cross-sum
        return (right-low, right-high, right-sum)
    else return (cross-low, cross-high, cross-sum)
```

Initial call: FIND-MAXIMUM-SUBARRAY$(A, 1, n)$

- Divide by computing *mid*.
- Conquer by the two recursive calls to FIND-MAXIMUM-SUBARRAY.
- Combine by calling FIND-MAX-CROSSING-SUBARRAY and then determining which of the three results gives the maximum sum.
- Base case is when the subarray has only 1 element.

Analysis

Simplifying assumption: Original problem size is a power of 2, so that all subproblem sizes are integer. *[We made the same simplifying assumption when we analyzed merge sort.]*

Let $T(n)$ denote the running time of FIND-MAXIMUM-SUBARRAY on a subarray of n elements.

Base case: Occurs when *high* equals *low*, so that $n = 1$. The procedure just returns $\Rightarrow T(n) = \Theta(1)$.

Recursive case: Occurs when $n > 1$.

- Dividing takes $\Theta(1)$ time.
- Conquering solves two subproblems, each on a subarray of $n/2$ elements. Takes $T(n/2)$ time for each subproblem $\Rightarrow 2T(n/2)$ time for conquering.
- Combining consists of calling FIND-MAX-CROSSING-SUBARRAY, which takes $\Theta(n)$ time, and a constant number of constant-time tests $\Rightarrow \Theta(n) + \Theta(1)$ time for combining.

Recurrence for recursive case becomes

$$\begin{aligned} T(n) &= \Theta(1) + 2T(n/2) + \Theta(n) + \Theta(1) \\ &= 2T(n/2) + \Theta(n) \qquad \text{(absorb } \Theta(1) \text{ terms into } \Theta(n)) \text{ .} \end{aligned}$$

The recurrence for all cases:

$$T(n) = \begin{cases} \Theta(1) & \text{if } n = 1 \text{ ,} \\ 2T(n/2) + \Theta(n) & \text{if } n > 1 \text{ .} \end{cases}$$

Same recurrence as for merge sort. Can use the master method to show that it has solution $T(n) = \Theta(n \lg n)$.

Thus, with divide-and-conquer, we have developed a $\Theta(n \lg n)$-time solution. Better than the $\Theta(n^2)$-time brute-force solution.

[Can actually solve this problem in $\Theta(n)$ time. See Exercise 4.1-5.]

Strassen's algorithm for matrix multiplication

Input: Two $n \times n$ (square) matrices, $A = (a_{ij})$ and $B = (b_{ij})$.
Output: $n \times n$ matrix $C = (c_{ij})$, where $C = A \cdot B$, i.e.,

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

for $i, j = 1, 2, \ldots, n$.

Need to compute n^2 entries of C. Each entry is the sum of n values.

Obvious method

[Using a shorter procedure name than in the book.]

```
SQUARE-MAT-MULT(A, B, n)
  let C be a new n × n matrix
  for i = 1 to n
      for j = 1 to n
          c_ij = 0
          for k = 1 to n
              c_ij = c_ij + a_ik · b_kj
  return C
```

Analysis: Three nested loops, each iterates n times, and innermost loop body takes constant time $\Rightarrow \Theta(n^3)$.

Is $\Theta(n^3)$ the best we can do? Can we multiply matrices in $o(n^3)$ time?

Seems like any algorithm to multiply matrices must take $\Omega(n^3)$ time:

- Must compute n^2 entries.
- Each entry is the sum of n terms.

But with Strassen's method, we can multiply matrices in $o(n^3)$ time.

- Strassen's algorithm runs in $\Theta(n^{\lg 7})$ time.
- $2.80 \leq \lg 7 \leq 2.81$.
- Hence, runs in $O(n^{2.81})$ time.

Simple divide-and-conquer method

As with the other divide-and-conquer algorithms, assume that n is a power of 2.

Partition each of A, B, C into four $n/2 \times n/2$ matrices:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}.$$

Rewrite $C = A \cdot B$ as

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

giving the four equations

$$\begin{aligned} C_{11} &= A_{11} \cdot B_{11} + A_{12} \cdot B_{21}, \\ C_{12} &= A_{11} \cdot B_{12} + A_{12} \cdot B_{22}, \\ C_{21} &= A_{21} \cdot B_{11} + A_{22} \cdot B_{21}, \\ C_{22} &= A_{21} \cdot B_{12} + A_{22} \cdot B_{22}. \end{aligned}$$

Each of these equations multiplies two $n/2 \times n/2$ matrices and then adds their $n/2 \times n/2$ products.

Use these equations to get a divide-and-conquer algorithm: *[Using a shorter procedure name than in the book.]*

```
REC-MAT-MULT(A, B, n)
  let C be a new n × n matrix
  if n == 1
      c11 = a11 · b11
  else partition A, B, and C into n/2 × n/2 submatrices
      C11 = REC-MAT-MULT(A11, B11, n/2) + REC-MAT-MULT(A12, B21, n/2)
      C12 = REC-MAT-MULT(A11, B12, n/2) + REC-MAT-MULT(A12, B22, n/2)
      C21 = REC-MAT-MULT(A21, B11, n/2) + REC-MAT-MULT(A22, B21, n/2)
      C22 = REC-MAT-MULT(A21, B12, n/2) + REC-MAT-MULT(A22, B22, n/2)
  return C
```

[The book briefly discusses the question of how to avoid copying entries when partitioning matrices. Can partition matrices without copying entries by instead using index calculations. Identify a submatrix by ranges of row and column matrices

from the original matrix. End up representing a submatrix differently from how we represent the original matrix. The advantage of avoiding copying is that partitioning would take only constant time, instead of $\Theta(n^2)$ time. The result of the asymptotic analysis won't change, but using index calculations to avoid copying gives better constant factors.]

Analysis

Let $T(n)$ be the time to multiply two $n/2 \times n/2$ matrices.

Base case: $n = 1$. Perform one scalar multiplication: $\Theta(1)$.

Recursive case: $n > 1$.

- Dividing takes $\Theta(1)$ time, using index calculations. *[Otherwise, $\Theta(n^2)$ time.]*
- Conquering makes 8 recursive calls, each multiplying $n/2 \times n/2$ matrices $\Rightarrow$ $8T(n/2)$.
- Combining takes $\Theta(n^2)$ time to add $n/2 \times n/2$ matrices four times. *[Doesn't even matter asymptotically whether we use index calculations or copy: would be $\Theta(n^2)$ either way.]*

Recurrence is

$$T(n) = \begin{cases} \Theta(1) & \text{if } n = 1 \ , \\ 8T(n/2) + \Theta(n^2) & \text{if } n > 1 \ . \end{cases}$$

Can use master method to show that it has solution $T(n) = \Theta(n^3)$.
Asymptotically, no better than the obvious method.

Constant factors and recurrences: When setting up recurrences, can absorb constant factors into asymptotic notation, but cannot absorb a constant number of subprobems. Although we absorb the 4 additions of $n/2 \times n/2$ matrices into the $\Theta(n^2)$ time, we cannot lose the 8 in front of the $T(n/2)$ term. If we absorb the constant number of subproblems, then the recursion tree would not be "bushy" and would instead just be a linear chain.

Strassen's method

Idea: Make the recursion tree less bushy. Perform only 7 recursive multiplications of $n/2 \times n/2$ matrices, rather than 8. Will cost several additions of $n/2 \times n/2$ matrices, but just a constant number more $\Rightarrow$ can still absorb the constant factor for matrix additions into the $\Theta(n/2)$ term.

The algorithm:

1. As in the recursive method, partition each of the matrices into four $n/2 \times n/2$ submatrices. Time: $\Theta(1)$.
2. Create 10 matrices $S_1, S_2, \ldots, S_{10}$. Each is $n/2 \times n/2$ and is the sum or difference of two matrices created in previous step. Time: $\Theta(n^2)$ to create all 10 matrices.
3. Recursively compute 7 matrix products $P_1, P_2, \ldots, P_7$, each $n/2 \times n/2$.
4. Compute $n/2 \times n/2$ submatrices of C by adding and subtracting various combinations of the P_i. Time: $\Theta(n^2)$.

Analysis

Recurrence will be

$$T(n) = \begin{cases} \Theta(1) & \text{if } n = 1 \text{ ,} \\ 7T(n/2) + \Theta(n^2) & \text{if } n > 1 \text{ .} \end{cases}$$

By the master method, solution is $T(n) = \Theta(n^{\lg 7})$.

Details

Step 2: Create the 10 matrices

$$\begin{aligned} S_1 &= B_{12} - B_{22} \text{ ,} \\ S_2 &= A_{11} + A_{12} \text{ ,} \\ S_3 &= A_{21} + A_{22} \text{ ,} \\ S_4 &= B_{21} - B_{11} \text{ ,} \\ S_5 &= A_{11} + A_{22} \text{ ,} \\ S_6 &= B_{11} + B_{22} \text{ ,} \\ S_7 &= A_{12} - A_{22} \text{ ,} \\ S_8 &= B_{21} + B_{22} \text{ ,} \\ S_9 &= A_{11} - A_{21} \text{ ,} \\ S_{10} &= B_{11} + B_{12} \text{ .} \end{aligned}$$

Add or subtract $n/2 \times n/2$ matrices 10 times $\Rightarrow$ time is $\Theta(n/2)$.

Step 3: Create the 7 matrices

$$\begin{aligned} P_1 &= A_{11} \cdot S_1 &&= A_{11} \cdot B_{12} - A_{11} \cdot B_{22} \text{ ,} \\ P_2 &= S_2 \cdot B_{22} &&= A_{11} \cdot B_{22} + A_{12} \cdot B_{22} \text{ ,} \\ P_3 &= S_3 \cdot B_{11} &&= A_{21} \cdot B_{11} + A_{22} \cdot B_{11} \text{ ,} \\ P_4 &= A_{22} \cdot S_4 &&= A_{22} \cdot B_{21} - A_{22} \cdot B_{11} \text{ ,} \\ P_5 &= S_5 \cdot S_6 &&= A_{11} \cdot B_{11} + A_{11} \cdot B_{22} + A_{22} \cdot B_{11} + A_{22} \cdot B_{22} \text{ ,} \\ P_6 &= S_7 \cdot S_8 &&= A_{12} \cdot B_{21} + A_{12} \cdot B_{22} - A_{22} \cdot B_{21} - A_{22} \cdot B_{22} \text{ ,} \\ P_7 &= S_9 \cdot S_{10} &&= A_{11} \cdot B_{11} + A_{11} \cdot B_{12} - A_{21} \cdot B_{11} - A_{21} \cdot B_{12} \text{ .} \end{aligned}$$

The only multiplications needed are in the middle column; right-hand column just shows the products in terms of the original submatrices of A and B.

Step 4: Add and subtract the P_i to construct submatrices of C:

$$\begin{aligned} C_{11} &= P_5 + P_4 - P_2 + P_6 \text{ ,} \\ C_{12} &= P_1 + P_2 \text{ ,} \\ C_{21} &= P_3 + P_4 \text{ ,} \\ C_{22} &= P_5 + P_1 - P_3 - P_7 \text{ .} \end{aligned}$$

To see how these computations work, expand each right-hand side, replacing each P_i with the submatrices of A and B that form it, and cancel terms: *[We expand out all four right-hand sides here. You might want to do just one or two of them, to convince students that it works.]*

$$
\begin{array}{llll}
A_{11}\cdot B_{11} + A_{11}\cdot B_{22} + A_{22}\cdot B_{11} + A_{22}\cdot B_{22} & & & \\
\quad - A_{22}\cdot B_{11} & + A_{22}\cdot B_{21} & & \\
- A_{11}\cdot B_{22} & & - A_{12}\cdot B_{22} & \\
& - A_{22}\cdot B_{22} - A_{22}\cdot B_{21} + A_{12}\cdot B_{22} + A_{12}\cdot B_{21} & & \\
\hline
A_{11}\cdot B_{11} & & & + A_{12}\cdot B_{21}
\end{array}
$$

$$
\begin{array}{ll}
A_{11}\cdot B_{12} - A_{11}\cdot B_{22} & \\
\quad + A_{11}\cdot B_{22} + A_{12}\cdot B_{22} & \\
\hline
A_{11}\cdot B_{12} & + A_{12}\cdot B_{22}
\end{array}
$$

$$
\begin{array}{ll}
A_{21}\cdot B_{11} + A_{22}\cdot B_{11} & \\
\quad - A_{22}\cdot B_{11} + A_{22}\cdot B_{21} & \\
\hline
A_{21}\cdot B_{11} & + A_{22}\cdot B_{21}
\end{array}
$$

$$
\begin{array}{llll}
A_{11}\cdot B_{11} + A_{11}\cdot B_{22} + A_{22}\cdot B_{11} + A_{22}\cdot B_{22} & & & \\
\quad - A_{11}\cdot B_{22} & + A_{11}\cdot B_{12} & & \\
& - A_{22}\cdot B_{11} & - A_{21}\cdot B_{11} & \\
- A_{11}\cdot B_{11} & & - A_{11}\cdot B_{12} + A_{21}\cdot B_{11} + A_{21}\cdot B_{12} & \\
\hline
& A_{22}\cdot B_{22} & & + A_{21}\cdot B_{12}
\end{array}
$$

Theoretical and practical notes

Strassen's algorithm was the first to beat $\Theta(n^3)$ time, but it's not the asymptotically fastest known. A method by Coppersmith and Winograd runs in $O(n^{2.376})$ time.

Practical issues against Strassen's algorithm:

- Higher constant factor than the obvious $\Theta(n^3)$-time method.
- Not good for sparse matrices.
- Not numerically stable: larger errors accumulate than in the obvious method.
- Submatrices consume space, especially if copying.

Numerical stability problem is not as bad as previously thought. And can use index calculations to reduce space requirement.

Various researchers have tried to find the crossover point, where Strassen's algorithm runs faster than the obvious $\Theta(n^3)$-time method. Analyses (that ignore caches and hardware pipelines) have produced crossover points as low as $n = 8$, and experiments have found crossover points as low as $n = 400$.

Substitution method

1. Guess the solution.
2. Use induction to find the constants and show that the solution works.

Example

$$T(n) = \begin{cases} 1 & \text{if } n = 1 \,, \\ 2T(n/2) + n & \text{if } n > 1 \,. \end{cases}$$

1. *Guess:* $T(n) = n \lg n + n$. *[Here, we have a recurrence with an exact function, rather than asymptotic notation, and the solution is also exact rather than asymptotic. We'll have to check boundary conditions and the base case.]*
2. *Induction:*

   Basis: $n = 1 \Rightarrow n \lg n + n = 1 = T(n)$

   Inductive step: Inductive hypothesis is that $T(k) = k \lg k + k$ for all $k < n$. We'll use this inductive hypothesis for $T(n/2)$.

$$\begin{aligned} T(n) &= 2T\left(\frac{n}{2}\right) + n \\ &= 2\left(\frac{n}{2}\lg\frac{n}{2} + \frac{n}{2}\right) + n \qquad \text{(by inductive hypothesis)} \\ &= n \lg \frac{n}{2} + n + n \\ &= n(\lg n - \lg 2) + n + n \\ &= n \lg n - n + n + n \\ &= n \lg n + n \,. \end{aligned}$$

■

Generally, we use asymptotic notation:

- We would write $T(n) = 2T(n/2) + \Theta(n)$.
- We assume $T(n) = O(1)$ for sufficiently small n.
- We express the solution by asymptotic notation: $T(n) = \Theta(n \lg n)$.
- We don't worry about boundary cases, nor do we show base cases in the substitution proof.
  - $T(n)$ is always constant for any constant n.
  - Since we are ultimately interested in an asymptotic solution to a recurrence, it will always be possible to choose base cases that work.
  - When we want an asymptotic solution to a recurrence, we don't worry about the base cases in our proofs.
  - When we want an exact solution, then we have to deal with base cases.

For the substitution method:

- Name the constant in the additive term.
- Show the upper (O) and lower (Ω) bounds separately. Might need to use different constants for each.

Example

$T(n) = 2T(n/2) + \Theta(n)$. If we want to show an upper bound of $T(n) = 2T(n/2) + O(n)$, we write $T(n) \le 2T(n/2) + cn$ for some positive constant c.

1. ***Upper bound:***

   Guess: $T(n) \le dn \lg n$ for some positive constant d. We are given c in the recurrence, and we get to choose d as any positive constant. It's OK for d to depend on c.

   Substitution:

$$\begin{aligned} T(n) &\le 2T(n/2) + cn \\ &= 2\left(d\frac{n}{2}\lg\frac{n}{2}\right) + cn \\ &= dn\lg\frac{n}{2} + cn \\ &= dn\lg n - dn + cn \\ &\le dn\lg n \qquad \text{if } -dn + cn \le 0, \\ &\qquad\qquad\qquad\qquad d \ge c \end{aligned}$$

   Therefore, $T(n) = O(n \lg n)$.

2. ***Lower bound:*** Write $T(n) \ge 2T(n/2) + cn$ for some positive constant c.

   Guess: $T(n) \ge dn \lg n$ for some positive constant d.

   Substitution:

$$\begin{aligned} T(n) &\ge 2T(n/2) + cn \\ &= 2\left(d\frac{n}{2}\lg\frac{n}{2}\right) + cn \\ &= dn\lg\frac{n}{2} + cn \\ &= dn\lg n - dn + cn \\ &\ge dn\lg n \qquad \text{if } -dn + cn \ge 0, \\ &\qquad\qquad\qquad\qquad d \le c \end{aligned}$$

   Therefore, $T(n) = \Omega(n \lg n)$.

Therefore, $T(n) = \Theta(n \lg n)$. *[For this particular recurrence, we can use $d = c$ for both the upper-bound and lower-bound proofs. That won't always be the case.]* ■

Make sure you show the same *exact* form when doing a substitution proof.

Consider the recurrence

$$T(n) = 8T(n/2) + \Theta(n^2) .$$

For an upper bound:

$$T(n) \le 8T(n/2) + cn^2 .$$

Guess: $T(n) \le dn^3$.

$$\begin{aligned} T(n) &\le 8d(n/2)^3 + cn^2 \\ &= 8d(n^3/8) + cn^2 \\ &= dn^3 + cn^2 \\ &\not\le dn^3 \qquad \text{doesn't work!} \end{aligned}$$

Remedy: *Subtract off* a lower-order term.

Guess: $T(n) \le dn^3 - d'n^2$.

$$\begin{aligned} T(n) &\le 8(d(n/2)^3 - d'(n/2)^2) + cn^2 \\ &= 8d(n^3/8) - 8d'(n^2/4) + cn^2 \\ &= dn^3 - 2d'n^2 + cn^2 \\ &= dn^3 - d'n^2 - d'n^2 + cn^2 \\ &\le dn^3 - d'n^2 \qquad \text{if } -d'n^2 + cn^2 \le 0\,, \\ &\qquad\qquad\qquad\qquad\qquad\quad d' \ge c \end{aligned}$$

Be careful when using asymptotic notation.

The false proof for the recurrence $T(n) = 4T(n/4) + n$, that $T(n) = O(n)$:

$$\begin{aligned} T(n) &\le 4(c(n/4)) + n \\ &\le cn + n \\ &= O(n) \qquad \text{wrong!} \end{aligned}$$

Because we haven't proven the *exact form* of our inductive hypothesis (which is that $T(n) \le cn$), this proof is false.

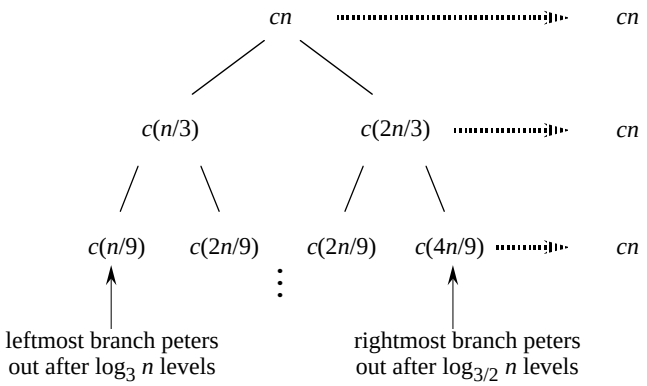
Recursion trees

Use to generate a guess. Then verify by substitution method.

Example

$T(n) = T(n/3) + T(2n/3) + \Theta(n)$.
For upper bound, rewrite as $T(n) \le T(n/3) + T(2n/3) + cn$; for lower bound, as $T(n) \ge T(n/3) + T(2n/3) + cn$.

By summing across each level, the recursion tree shows the cost at each level of recursion (minus the costs of recursive calls, which appear in subtrees):



- There are $\log_3 n$ full levels, and after $\log_{3/2} n$ levels, the problem size is down to 1.
- Each level contributes $\le cn$.
- Lower bound guess: $\ge dn \log_3 n = \Omega(n \lg n)$ for some positive constant d.

- Upper bound guess: $\le dn \log_{3/2} n = O(n \lg n)$ for some positive constant d.
- Then *prove* by substitution.

1. ***Upper bound:***

   Guess: $T(n) \le dn \lg n$.

   Substitution:

$$\begin{aligned} T(n) &\le T(n/3) + T(2n/3) + cn \\ &\le d(n/3)\lg(n/3) + d(2n/3)\lg(2n/3) + cn \\ &= (d(n/3)\lg n - d(n/3)\lg 3) \\ &\qquad + (d(2n/3)\lg n - d(2n/3)\lg(3/2)) + cn \\ &= dn\lg n - d((n/3)\lg 3 + (2n/3)\lg(3/2)) + cn \\ &= dn\lg n - d((n/3)\lg 3 + (2n/3)\lg 3 - (2n/3)\lg 2) + cn \\ &= dn\lg n - dn(\lg 3 - 2/3) + cn \\ &\le dn\lg n \qquad \text{if } -dn(\lg 3 - 2/3) + cn \le 0 , \\ &\qquad\qquad\qquad\qquad\qquad\qquad d \ge \frac{c}{\lg 3 - 2/3} . \end{aligned}$$

   Therefore, $T(n) = O(n \lg n)$.

   Note: Make sure that the symbolic constants used in the recurrence (e.g., c) and the guess (e.g., d) are different.

2. ***Lower bound:***

   Guess: $T(n) \ge dn \lg n$.

   Substitution: Same as for the upper bound, but replacing $\le$ by $\ge$. End up needing

$$0 < d \le \frac{c}{\lg 3 - 2/3} .$$

   Therefore, $T(n) = \Omega(n \lg n)$.

Since $T(n) = O(n \lg n)$ and $T(n) = \Omega(n \lg n)$, we conclude that $T(n) = \Theta(n \lg n)$. ■

Master method

Used for many divide-and-conquer recurrences of the form

$$T(n) = aT(n/b) + f(n) ,$$

where $a \ge 1$, $b > 1$, and $f(n) > 0$.

Based on the ***master theorem*** (Theorem 4.1).

Compare $n^{\log_b a}$ vs. $f(n)$:

Case 1: $f(n) = O(n^{\log_b a - \epsilon})$ for some constant $\epsilon > 0$.
($f(n)$ is polynomially smaller than $n^{\log_b a}$.)
Solution: $T(n) = \Theta(n^{\log_b a})$.
(Intuitively: cost is dominated by leaves.)

Case 2: $f(n) = \Theta(n^{\log_b a} \lg^k n)$, where $k \geq 0$.
[This formulation of Case 2 is more general than in Theorem 4.1, and it is given in Exercise 4.6-2.]
($f(n)$ is within a polylog factor of $n^{\log_b a}$, but not smaller.)
Solution: $T(n) = \Theta(n^{\log_b a} \lg^{k+1} n)$.
(Intuitively: cost is $n^{\log_b a} \lg^k n$ at each level, and there are $\Theta(\lg n)$ levels.)
Simple case: $k = 0 \Rightarrow f(n) = \Theta(n^{\log_b a}) \Rightarrow T(n) = \Theta(n^{\log_b a} \lg n)$.

Case 3: $f(n) = \Omega(n^{\log_b a + \epsilon})$ for some constant $\epsilon > 0$ and $f(n)$ satisfies the regularity condition $af(n/b) \leq cf(n)$ for some constant $c < 1$ and all sufficiently large n.
($f(n)$ is polynomially greater than $n^{\log_b a}$.)
Solution: $T(n) = \Theta(f(n))$.
(Intuitively: cost is dominated by root.)

What's with the Case 3 regularity condition?

- Generally not a problem.
- It always holds whenever $f(n) = n^k$ and $f(n) = \Omega(n^{\log_b a + \epsilon})$ for constant $\epsilon > 0$. *[Proving this makes a nice homework exercise. See below.]* So you don't need to check it when $f(n)$ is a polynomial.

[Here's a proof that the regularity condition holds when $f(n) = n^k$ and $f(n) = \Omega(n^{\log_b a + \epsilon})$ for constant $\epsilon > 0$.

Since $f(n) = \Omega(n^{\log_b a + \epsilon})$ and $f(n) = n^k$, we have that $k > \log_b a$. Using a base of b and treating both sides as exponents, we have $b^k > b^{\log_b a} = a$, and so $a/b^k < 1$. Since a, b, and k are constants, if we let $c = a/b^k$, then c is a constant strictly less than 1. We have that $af(n/b) = a(n/b)^k = (a/b^k)n^k = cf(n)$, and so the regularity condition is satisfied.]

Examples

- $T(n) = 5T(n/2) + \Theta(n^2)$
  $n^{\log_2 5}$ vs. n^2
  Since $\log_2 5 - \epsilon = 2$ for some constant $\epsilon > 0$, use Case 1 $\Rightarrow T(n) = \Theta(n^{\lg 5})$

- $T(n) = 27T(n/3) + \Theta(n^3 \lg n)$
  $n^{\log_3 27} = n^3$ vs. $n^3 \lg n$
  Use Case 2 with $k = 1 \Rightarrow T(n) = \Theta(n^3 \lg^2 n)$

- $T(n) = 5T(n/2) + \Theta(n^3)$
  $n^{\log_2 5}$ vs. n^3
  Now $\lg 5 + \epsilon = 3$ for some constant $\epsilon > 0$
  Check regularity condition (don't really need to since $f(n)$ is a polynomial):
  $af(n/b) = 5(n/2)^3 = 5n^3/8 \leq cn^3$ for $c = 5/8 < 1$
  Use Case 3 $\Rightarrow T(n) = \Theta(n^3)$

- $T(n) = 27T(n/3) + \Theta(n^3/\lg n)$
  $n^{\log_3 27} = n^3$ vs. $n^3/\lg n = n^3 \lg^{-1} n \neq \Theta(n^3 \lg^k n)$ for any $k \geq 0$.
  Cannot use the master method.

[We don't prove the master theorem in our algorithms course. We sometimes prove a simplified version for recurrences of the form $T(n) = aT(n/b) + n^c$. Section 4.6 of the text has the full proof of the master theorem.]

Solutions for Chapter 4: Divide-and-Conquer

Solution to Exercise 4.1-1

If the index of the greatest element of A is i, it returns $(i, i, A[i])$.

Solution to Exercise 4.1-2

```
MAX-SUBARRAY-BRUTE-FORCE(A)
  n = A.length
  max-so-far = −∞
  for l = 1 to n
      sum = 0
      for h = l to n
          sum = sum + A[h]
          if sum > max-so-far
              max-so-far = sum
              low = l
              high = h
  return (low, high)
```

Solution to Exercise 4.1-4

If the algorithm returns a negative sum, toss out the answer and use an empty subarray instead.

Solution to Exercise 4.1-5

```
MAX-SUBARRAY-LINEAR(A)
  n = A.length
  max-sum = −∞
  ending-here-sum = −∞
  for j = 1 to n
      ending-here-high = j
      if ending-here-sum > 0
          ending-here-sum = ending-here-sum + A[j]
      else ending-here-low = j
          ending-here-sum = A[j]
      if ending-here-sum > max-sum
          max-sum = ending-here-sum
          low = ending-here-low
          high = ending-here-high
  return (low, high, max-sum)
```

The variables are intended as follows:

- *low* and *high* demarcate a maximum subarray found so far.
- *max-sum* gives the sum of the values in a maximum subarray found so far.
- *ending-here-low* and *ending-here-high* demarcate a maximum subarray ending at index j. Since the high end of any subarray ending at index j must be j, every iteration of the **for** loop automatically sets *ending-here-high* $= j$.
- *ending-here-sum* gives the sum of the values in a maximum subarray ending at index j.

The first test within the **for** loop determines whether a maximum subarray ending at index j contains just $A[j]$. As we enter an iteration of the loop, *ending-here-sum* has the sum of the values in a maximum subarray ending at $j-1$. If *ending-here-sum* $+ A[j] > A[j]$, then we extend the maximum subarray ending at index $j-1$ to include index j. (The test in the **if** statement just subtracts out $A[j]$ from both sides.) Otherwise, we start a new subarray at index j, so both its low and high ends have the value j and its sum is $A[j]$. Once we know the maximum subarray ending at index j, we test to see whether it has a greater sum than the maximum subarray found so far, ending at any position less than or equal to j. If it does, then we update *low*, *high*, and *max-sum* appropriately.

Since each iteration of the **for** loop takes constant time, and the loop makes n iterations, the running time of MAX-SUBARRAY-LINEAR is $\Theta(n)$.

Solution to Exercise 4.2-2

STRASSEN(A, B)

$n = A.rows$
let C be a new $n \times n$ matrix
if $n == 1$
    $c_{11} = a_{11} \cdot b_{11}$
else partition A and B in equations (4.9)
    let C_{11}, C_{12}, C_{21}, and C_{22} be $n/2 \times n/2$ matrices
    create $n/2 \times n/2$ matrices $S_1, S_2, \ldots, S_{10}$ and $P_1, P_2, \ldots, P_7$
    $S_1 = B_{12} - B_{22}$
    $S_2 = A_{11} + A_{12}$
    $S_3 = A_{12} + A_{22}$
    $S_4 = B_{21} - B_{11}$
    $S_5 = A_{11} + A_{22}$
    $S_6 = B_{11} + B_{22}$
    $S_7 = A_{12} - A_{22}$
    $S_8 = B_{21} + B_{22}$
    $S_9 = A_{11} - A_{21}$
    $S_{10} = B_{11} + B_{12}$
    $P_1 =$ STRASSEN(A_{11}, S_1)
    $P_2 =$ STRASSEN(S_2, B_{22})
    $P_3 =$ STRASSEN(S_3, B_{11})
    $P_4 =$ STRASSEN(A_{22}, S_4)
    $P_5 =$ STRASSEN(S_5, S_6)
    $P_6 =$ STRASSEN(S_7, S_8)
    $P_7 =$ STRASSEN(S_9, S_{10})
    $C_{11} = P_5 + P_4 - P_2 + P_6$
    $C_{12} = P_1 + P_2$
    $C_{21} = P_3 + P_4$
    $C_{22} = P_5 + P_1 - P_3 - P_7$
    combine C_{11}, C_{12}, C_{21}, and C_{22} into C
return C

Solution to Exercise 4.2-4

This solution is also posted publicly

If you can multiply 3×3 matrices using k multiplications, then you can multiply $n \times n$ matrices by recursively multiplying $n/3 \times n/3$ matrices, in time $T(n) = kT(n/3) + \Theta(n^2)$.

Using the master method to solve this recurrence, consider the ratio of $n^{\log_3 k}$ and n^2:

- If $\log_3 k = 2$, case 2 applies and $T(n) = \Theta(n^2 \lg n)$. In this case, $k = 9$ and $T(n) = o(n^{\lg 7})$.

- If $\log_3 k < 2$, case 3 applies and $T(n) = \Theta(n^2)$. In this case, $k < 9$ and $T(n) = o(n^{\lg 7})$.
- If $\log_3 k > 2$, case 1 applies and $T(n) = \Theta(n^{\log_3 k})$. In this case, $k > 9$. $T(n) = o(n^{\lg 7})$ when $\log_3 k < \lg 7$, i.e., when $k < 3^{\lg 7} \approx 21.85$. The largest such integer k is 21.

Thus, $k = 21$ and the running time is $\Theta(n^{\log_3 k}) = \Theta(n^{\log_3 21}) = O(n^{2.80})$ (since $\log_3 21 \approx 2.77$).

Solution to Exercise 4.3-1

We guess that $T(n) \le cn^2$ for some constant $c > 0$. We have

$$\begin{aligned} T(n) &= T(n-1) + n \\ &\le c(n-1)^2 + n \\ &= cn^2 - 2cn + c + n \\ &= cn^2 + c(1-2n) + n \ . \end{aligned}$$

This last quantity is less than or equal to cn^2 if $c(1-2n) + n \le 0$ or, equivalently, $c \ge n/(2n-1)$. This last condition holds for all $n \ge 1$ and $c \ge 1$.

For the boundary condition, we set $T(1) = 1$, and so $T(1) = 1 \le c \cdot 1^2$. Thus, we can choose $n_0 = 1$ and $c = 1$.

Solution to Exercise 4.3-7

If we were to try a straight substitution proof, assuming that $T(n) \le cn^{\log_3 4}$, we would get stuck:

$$\begin{aligned} T(n) &\le 4(c(n/3)^{\log_3 4}) + n \\ &= 4c\left(\frac{n^{\log_3 4}}{4}\right) + n \\ &= cn^{\log_3 4} + n \ , \end{aligned}$$

which is greater than $cn^{\log_3 4}$. Instead, we subtract off a lower-order term and assume that $T(n) \le cn^{\log_3 4} - dn$. Now we have

$$\begin{aligned} T(n) &\le 4(c(n/3)^{\log_3 4} - dn/3) + n \\ &= 4\left(\frac{cn^{\log_3 4}}{4} - \frac{dn}{3}\right) + n \\ &= cn^{\log_3 4} - \frac{4}{3}dn + n \ , \end{aligned}$$

which is less than or equal to $cn^{\log_3 4} - dn$ if $d \ge 3$.

Solution to Exercise 4.4-6

This solution is also posted publicly

The shortest path from the root to a leaf in the recursion tree is $n \to (1/3)n \to (1/3)^2 n \to \cdots \to 1$. Since $(1/3)^k n = 1$ when $k = \log_3 n$, the height of the part of the tree in which every node has two children is $\log_3 n$. Since the values at each of these levels of the tree add up to cn, the solution to the recurrence is at least $cn \log_3 n = \Omega(n \lg n)$.

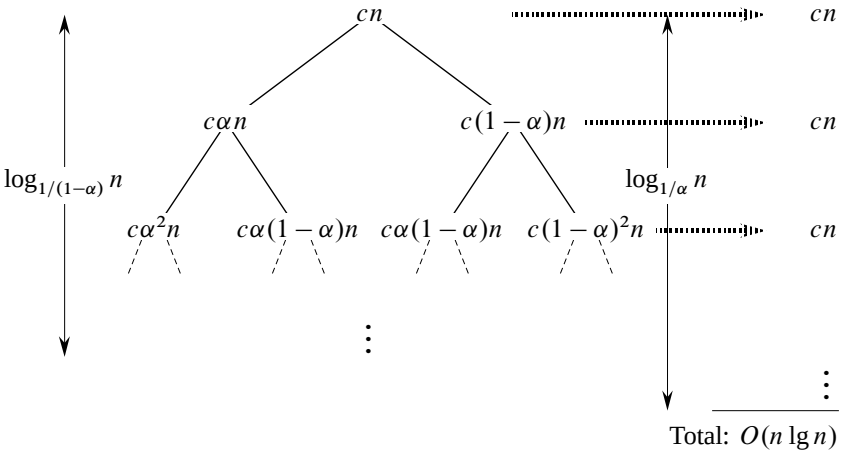
Solution to Exercise 4.4-9

This solution is also posted publicly

$T(n) = T(\alpha n) + T((1-\alpha)n) + cn$

We saw the solution to the recurrence $T(n) = T(n/3) + T(2n/3) + cn$ in the text. This recurrence can be similarly solved.

Without loss of generality, let $\alpha \geq 1-\alpha$, so that $0 < 1-\alpha \leq 1/2$ and $1/2 \leq \alpha < 1$.



The recursion tree is full for $\log_{1/(1-\alpha)} n$ levels, each contributing cn, so we guess $\Omega(n \log_{1/(1-\alpha)} n) = \Omega(n \lg n)$. It has $\log_{1/\alpha} n$ levels, each contributing $\leq cn$, so we guess $O(n \log_{1/\alpha} n) = O(n \lg n)$.

Now we show that $T(n) = \Theta(n \lg n)$ by substitution. To prove the upper bound, we need to show that $T(n) \leq dn \lg n$ for a suitable constant $d > 0$.

$$\begin{aligned}
T(n) &= T(\alpha n) + T((1-\alpha)n) + cn \\
&\leq d\alpha n \lg(\alpha n) + d(1-\alpha)n \lg((1-\alpha)n) + cn \\
&= d\alpha n \lg \alpha + d\alpha n \lg n + d(1-\alpha)n \lg(1-\alpha) + d(1-\alpha)n \lg n + cn \\
&= dn \lg n + dn(\alpha \lg \alpha + (1-\alpha)\lg(1-\alpha)) + cn \\
&\leq dn \lg n \ ,
\end{aligned}$$

if $dn(\alpha \lg \alpha + (1-\alpha)\lg(1-\alpha)) + cn \leq 0$. This condition is equivalent to

$d(\alpha \lg \alpha + (1-\alpha)\lg(1-\alpha)) \leq -c$.

Since $1/2 \le \alpha < 1$ and $0 < 1-\alpha \le 1/2$, we have that $\lg\alpha < 0$ and $\lg(1-\alpha) < 0$. Thus, $\alpha\lg\alpha + (1-\alpha)\lg(1-\alpha) < 0$, so that when we multiply both sides of the inequality by this factor, we need to reverse the inequality:

$$d \ge \frac{-c}{\alpha\lg\alpha + (1-\alpha)\lg(1-\alpha)}$$

or

$$d \ge \frac{c}{-\alpha\lg\alpha + -(1-\alpha)\lg(1-\alpha)} .$$

The fraction on the right-hand side is a positive constant, and so it suffices to pick any value of d that is greater than or equal to this fraction.

To prove the lower bound, we need to show that $T(n) \ge dn\lg n$ for a suitable constant $d > 0$. We can use the same proof as for the upper bound, substituting $\ge$ for $\le$, and we get the requirement that

$$0 < d \le \frac{c}{-\alpha\lg\alpha - (1-\alpha)\lg(1-\alpha)} .$$

Therefore, $T(n) = \Theta(n\lg n)$.

Solution to Exercise 4.5-2

We need to find the largest integer a such that $\log_4 a < \lg 7$. The answer is $a = 48$.

Solution to Problem 4-1

Note: In parts (a), (b), and (d) below, we are applying case 3 of the master theorem, which requires the regularity condition that $af(n/b) \le cf(n)$ for some constant $c < 1$. In each of these parts, $f(n)$ has the form n^k. The regularity condition is satisfied because $af(n/b) = an^k/b^k = (a/b^k)n^k = (a/b^k)f(n)$, and in each of the cases below, a/b^k is a constant strictly less than 1.

a. $T(n) = 2T(n/2) + n^3 = \Theta(n^3)$. This is a divide-and-conquer recurrence with $a = 2$, $b = 2$, $f(n) = n^3$, and $n^{\log_b a} = n^{\log_2 2} = n$. Since $n^3 = \Omega(n^{\log_2 2+2})$ and $a/b^k = 2/2^3 = 1/4 < 1$, case 3 of the master theorem applies, and $T(n) = \Theta(n^3)$.

b. $T(n) = T(9n/10) + n = \Theta(n)$. This is a divide-and-conquer recurrence with $a = 1$, $b = 10/9$, $f(n) = n$, and $n^{\log_b a} = n^{\log_{10/9} 1} = n^0 = 1$. Since $n = \Omega(n^{\log_{10/9} 1+1})$ and $a/b^k = 1/(10/9)^1 = 9/10 < 1$, case 3 of the master theorem applies, and $T(n) = \Theta(n)$.

c. $T(n) = 16T(n/4) + n^2 = \Theta(n^2\lg n)$. This is another divide-and-conquer recurrence with $a = 16$, $b = 4$, $f(n) = n^2$, and $n^{\log_b a} = n^{\log_4 16} = n^2$. Since $n^2 = \Theta(n^{\log_4 16})$, case 2 of the master theorem applies, and $T(n) = \Theta(n^2\lg n)$.

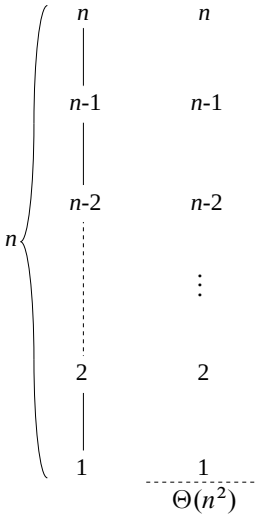
d. $T(n) = 7T(n/3) + n^2 = \Theta(n^2)$. This is a divide-and-conquer recurrence with $a = 7$, $b = 3$, $f(n) = n^2$, and $n^{\log_b a} = n^{\log_3 7}$. Since $1 < \log_3 7 < 2$, we have that $n^2 = \Omega(n^{\log_3 7 + \epsilon})$ for some constant $\epsilon > 0$. We also have $a/b^k = 7/3^2 = 7/9 < 1$, so that case 3 of the master theorem applies, and $T(n) = \Theta(n^2)$.

e. $T(n) = 7T(n/2) + n^2 = O(n^{\lg 7})$. This is a divide-and-conquer recurrence with $a = 7$, $b = 2$, $f(n) = n^2$, and $n^{\log_b a} = n^{\log_2 7}$. Since $2 < \lg 7 < 3$, we have that $n^2 = O(n^{\log_2 7 - \epsilon})$ for some constant $\epsilon > 0$. Thus, case 1 of the master theorem applies, and $T(n) = \Theta(n^{\lg 7})$.

f. $T(n) = 2T(n/4) + \sqrt{n} = \Theta(\sqrt{n} \lg n)$. This is another divide-and-conquer recurrence with $a = 2$, $b = 4$, $f(n) = \sqrt{n}$, and $n^{\log_b a} = n^{\log_4 2} = \sqrt{n}$. Since $\sqrt{n} = \Theta(n^{\log_4 2})$, case 2 of the master theorem applies, and $T(n) = \Theta(\sqrt{n} \lg n)$.

g. $T(n) = T(n-1) + n$

Using the recursion tree shown below, we get a guess of $T(n) = \Theta(n^2)$.



First, we prove the $T(n) = \Omega(n^2)$ part by induction. The inductive hypothesis is $T(n) \geq cn^2$ for some constant $c > 0$.

$$\begin{aligned} T(n) &= T(n-1) + n \\ &\geq c(n-1)^2 + n \\ &= cn^2 - 2cn + c + n \\ &\geq cn^2 \end{aligned}$$

if $-2cn + n + c \geq 0$ or, equivalently, $n(1-2c) + c \geq 0$. This condition holds when $n \geq 0$ and $0 < c \leq 1/2$.

For the upper bound, $T(n) = O(n^2)$, we use the inductive hypothesis that $T(n) \leq cn^2$ for some constant $c > 0$. By a similar derivation, we get that $T(n) \leq cn^2$ if $-2cn + n + c \leq 0$ or, equivalently, $n(1-2c) + c \leq 0$. This condition holds for $c = 1$ and $n \geq 1$.

Thus, $T(n) = \Omega(n^2)$ and $T(n) = O(n^2)$, so we conclude that $T(n) = \Theta(n^2)$.

h. $T(n) = T(\sqrt{n}) + 1$

The easy way to do this is with a change of variables, as on page 86 of the text. Let $m = \lg n$ and $S(m) = T(2^m)$. $T(2^m) = T(2^{m/2}) + 1$, so $S(m) = S(m/2) + 1$. Using the master theorem, $n^{\log_b a} = n^{\log_2 1} = n^0 = 1$ and $f(n) = 1$. Since $1 = \Theta(1)$, case 2 applies and $S(m) = \Theta(\lg m)$. Therefore, $T(n) = \Theta(\lg \lg n)$.

Solution to Problem 4-3

[This problem is solved only for parts a, c, e, f, g, h, and i.]

a. $T(n) = 3T(n/2) + n \lg n$

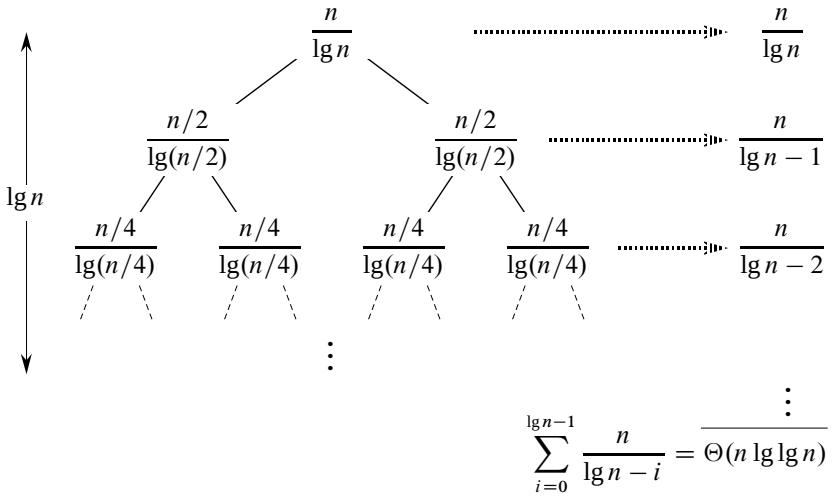
We have $f(n) = n \lg n$ and $n^{\log_b a} = n^{\lg 3} \approx n^{1.585}$. Since $n \lg n = O(n^{\lg 3 - \epsilon})$ for any $0 < \epsilon \leq 0.58$, by case 1 of the master theorem, we have $T(n) = \Theta(n^{\lg 3})$.

c. $T(n) = 4T(n/2) + n^2\sqrt{n}$

We have $f(n) = n^2\sqrt{n} = n^{5/2}$ and $n^{\log_b a} = n^{\log_2 4} = n^2$. Since $n^{5/2} = \Omega(n^{2+\epsilon})$ for $\epsilon = 1/2$, we look at the regularity condition in case 3 of the master theorem. We have $af(n/b) = 4(n/2)^2\sqrt{n/2} = n^{5/2}/\sqrt{2} \leq cn^{5/2}$ for $1/\sqrt{2} \leq c < 1$. Case 3 applies, and we have $T(n) = \Theta(n^2\sqrt{n})$.

e. $T(n) = 2T(n/2) + n/\lg n$

We can get a guess by means of a recursion tree:



We get the sum on each level by observing that at depth i, we have 2^i nodes, each with a numerator of $n/2^i$ and a denominator of $\lg(n/2^i) = \lg n - i$, so that the cost at depth i is

$$2^i \cdot \frac{n/2^i}{\lg n - i} = \frac{n}{\lg n - i} .$$

The sum for all levels is

$$\begin{aligned}
\sum_{i=0}^{\lg n-1} \frac{n}{\lg n - i} &= n \sum_{i=1}^{\lg n} \frac{n}{i} \\
&= n \sum_{i=1}^{\lg n} 1/i \\
&= n \cdot \Theta(\lg\lg n) \quad \text{(by equation (A.7), the harmonic series)} \\
&= \Theta(n \lg\lg n) \; .
\end{aligned}$$

We can use this analysis as a guess that $T(n) = \Theta(n \lg\lg n)$. If we were to do a straight substitution proof, it would be rather involved. Instead, we will show by substitution that $T(n) \le n(1 + H_{\lfloor \lg n \rfloor})$ and $T(n) \ge n \cdot H_{\lceil \lg n \rceil}$, where H_k is the kth harmonic number: $H_k = 1/1 + 1/2 + 1/3 + \cdots + 1/k$. We also define $H_0 = 0$. Since $H_k = \Theta(\lg k)$, we have that $H_{\lfloor \lg n \rfloor} = \Theta(\lg \lfloor \lg n \rfloor) = \Theta(\lg\lg n)$ and $H_{\lceil \lg n \rceil} = \Theta(\lg \lceil \lg n \rceil) = \Theta(\lg\lg n)$. Thus, we will have that $T(n) = \Theta(n \lg\lg n)$.

The base case for the proof is for $n = 1$, and we use $T(1) = 1$. Here, $\lg n = 0$, so that $\lg n = \lfloor \lg n \rfloor = \lceil \lg n \rceil$. Since $H_0 = 0$, we have $T(1) = 1 \le 1(1 + H_0)$ and $T(1) = 1 \ge 0 = 1 \cdot H_0$.

For the upper bound of $T(n) \le n(1 + H_{\lfloor \lg n \rfloor})$, we have

$$\begin{aligned}
T(n) &= 2T(n/2) + n/\lg n \\
&\le 2((n/2)(1 + H_{\lfloor \lg(n/2) \rfloor})) + n/\lg n \\
&= n(1 + H_{\lfloor \lg n - 1 \rfloor}) + n/\lg n \\
&= n(1 + H_{\lfloor \lg n \rfloor - 1} + 1/\lg n) \\
&\le n(1 + H_{\lfloor \lg n \rfloor - 1} + 1/\lfloor \lg n \rfloor) \\
&= n(1 + H_{\lfloor \lg n \rfloor}) \; ,
\end{aligned}$$

where the last line follows from the identity $H_k = H_{k-1} + 1/k$.

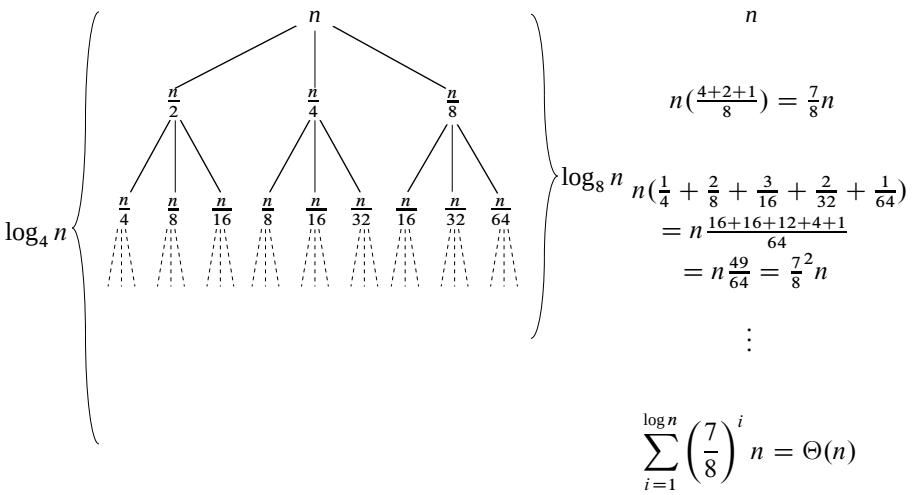
The upper bound of $T(n) \ge n \cdot H_{\lceil \lg n \rceil}$ is similar:

$$\begin{aligned}
T(n) &= 2T(n/2) + n/\lg n \\
&\ge 2((n/2) \cdot H_{\lceil \lg(n/2) \rceil}) + n/\lg n \\
&= n \cdot H_{\lceil \lg n - 1 \rceil} + n/\lg n \\
&= n \cdot (H_{\lceil \lg n \rceil - 1} + 1/\lg n) \\
&\ge n \cdot (H_{\lceil \lg n \rceil - 1} + 1/\lceil \lg n \rceil) \\
&= n \cdot H_{\lceil \lg n \rceil} \; .
\end{aligned}$$

Thus, $T(n) = \Theta(n \lg\lg n)$.

f. $T(n) = T(n/2) + T(n/4) + T(n/8) + n$

Using the recursion tree shown below, we get a guess of $T(n) = \Theta(n)$.



We use the substitution method to prove that $T(n) = O(n)$. Our inductive hypothesis is that $T(n) \le cn$ for some constant $c > 0$. We have

$$\begin{aligned} T(n) &= T(n/2) + T(n/4) + T(n/8) + n \\ &\le cn/2 + cn/4 + cn/8 + n \\ &= 7cn/8 + n \\ &= (1 + 7c/8)n \\ &\le cn \qquad \text{if } c \ge 8 . \end{aligned}$$

Therefore, $T(n) = O(n)$.

Showing that $T(n) = \Omega(n)$ is easy:

$$T(n) = T(n/2) + T(n/4) + T(n/8) + n \ge n .$$

Since $T(n) = O(n)$ and $T(n) = \Omega(n)$, we have that $T(n) = \Theta(n)$.

g. $T(n) = T(n-1) + 1/n$

This recurrence corresponds to the harmonic series, so that $T(n) = H_n$, where $H_n = 1/1+1/2+1/3+\cdots+1/n$. For the base case, we have $T(1) = 1 = H_1$. For the inductive step, we assume that $T(n-1) = H_{n-1}$, and we have

$$\begin{aligned} T(n) &= T(n-1) + 1/n \\ &= H_{n-1} + 1/n \\ &= H_n . \end{aligned}$$

Since $H_n = \Theta(\lg n)$ by equation (A.7), we have that $T(n) = \Theta(\lg n)$.

h. $T(n) = T(n-1) + \lg n$

We guess that $T(n) = \Theta(n \lg n)$. To prove the upper bound, we will show that $T(n) = O(n \lg n)$. Our inductive hypothesis is that $T(n) \le cn \lg n$ for some constant c. We have

$$\begin{aligned}
T(n) &= T(n-1) + \lg n \\
&\le c(n-1)\lg(n-1) + \lg n \\
&= cn\lg(n-1) - c\lg(n-1) + \lg n \\
&\le cn\lg(n-1) - c\lg(n/2) + \lg n \\
&\qquad \text{(since } \lg(n-1) \ge \lg(n/2) \text{ for } n \ge 2\text{)} \\
&= cn\lg(n-1) - c\lg n + c + \lg n \\
&< cn\lg n - c\lg n + c + \lg n \\
&\le cn\lg n \ ,
\end{aligned}$$

if $-c\lg n + c + \lg n \le 0$. Equivalently,

$$\begin{aligned}
-c\lg n + c + \lg n &\le 0 \\
c &\le (c-1)\lg n \\
\lg n &\ge c/(c-1) \ .
\end{aligned}$$

This works for $c = 2$ and all $n \ge 4$.

To prove the lower bound, we will show that $T(n) = \Omega(n\lg n)$. Our inductive hypothesis is that $T(n) \ge cn\lg n + dn$ for constants c and d. We have

$$\begin{aligned}
T(n) &= T(n-1) + \lg n \\
&\ge c(n-1)\lg(n-1) + d(n-1) + \lg n \\
&= cn\lg(n-1) - c\lg(n-1) + dn - d + \lg n \\
&\ge cn\lg(n/2) - c\lg(n-1) + dn - d + \lg n \\
&\qquad \text{(since } \lg(n-1) \ge \lg(n/2) \text{ for } n \ge 2\text{)} \\
&= cn\lg n - cn - c\lg(n-1) + dn - d + \lg n \\
&\ge cn\lg n \ ,
\end{aligned}$$

if $-cn - c\lg(n-1) + dn - d + \lg n \ge 0$. Since

$$-cn - c\lg(n-1) + dn - d + \lg n > -cn - c\lg(n-1) + dn - d + \lg(n-1) \ ,$$

it suffices to find conditions in which $-cn - c\lg(n-1) + dn - d + \lg(n-1) \ge 0$. Equivalently,

$$\begin{aligned}
-cn - c\lg(n-1) + dn - d + \lg(n-1) &\ge 0 \\
(d-c)n &\ge (c-1)\lg(n-1) + d \ .
\end{aligned}$$

This works for $c = 1$, $d = 2$, and all $n \ge 2$.

Since $T(n) = O(n\lg n)$ and $T(n) = \Omega(n\lg n)$, we conclude that $T(n) = \Theta(n\lg n)$.

i. $T(n) = T(n-2) + 2\lg n$

We guess that $T(n) = \Theta(n\lg n)$. We show the upper bound of $T(n) = O(n\lg n)$ by means of the inductive hypothesis $T(n) \le cn\lg n$ for some constant $c > 0$. We have

$$\begin{aligned}
T(n) &= T(n-2) + 2\lg n \\
&\le c(n-2)\lg(n-2) + 2\lg n \\
&\le c(n-2)\lg n + 2\lg n \\
&= (cn - 2c + 2)\lg n
\end{aligned}$$

$$\begin{aligned} &= cn\lg n + (2-2c)\lg n \\ &\le cn\lg n \qquad \text{if } c > 1\,. \end{aligned}$$

Therefore, $T(n) = O(n\lg n)$.

For the lower bound of $T(n) = \Omega(n\lg n)$, we'll show that $T(n) \ge cn\lg n + dn$, for constants $c, d > 0$ to be chosen. We assume that $n \ge 4$, which implies that

1. $\lg(n-2) \ge \lg(n/2)$,
2. $n/2 \ge \lg n$, and
3. $n/2 \ge 2$.

(We'll use these inequalities as we go along.) We have

$$\begin{aligned} T(n) &\ge c(n-2)\lg(n-2) + d(n-2) + 2\lg n \\ &= cn\lg(n-2) - 2c\lg(n-2) + dn - 2d + 2\lg n \\ &> cn\lg(n-2) - 2c\lg n + dn - 2d + 2\lg n \\ &\qquad \text{(since } -\lg n < -\lg(n-2)\text{)} \\ &= cn\lg(n-2) - 2(c-1)\lg n + dn - 2d \\ &\ge cn\lg(n/2) - 2(c-1)\lg n + dn - 2d \quad \text{(by inequality (1) above)} \\ &= cn\lg n - cn - 2(c-1)\lg n + dn - 2d \\ &\ge cn\lg n\,, \end{aligned}$$

if $-cn - 2(c-1)\lg n + dn - 2d \ge 0$ or, equivalently, $dn \ge cn + 2(c-1)\lg n + 2d$. Pick any constant $c > 1/2$, and then pick any constant d such that

$$d \ge 2(2c-1)\,.$$

(The requirement that $c > 1/2$ means that d is positive.) Then

$$d/2 \ge 2c - 1 = c + (c-1)\,,$$

and adding $d/2$ to both sides, we have

$$d \ge c + (c-1) + d/2\,.$$

Multiplying by n yields

$$dn \ge cn + (c-1)n + dn/2\,,$$

and then both multiplying and dividing the middle term by 2 gives

$$dn \ge cn + 2(c-1)n/2 + dn/2\,.$$

Using inequalities (2) and (3) above, we get

$$dn \ge cn + 2(c-1)\lg n + 2d\,,$$

which is what we needed to show. Thus $T(n) = \Omega(n\lg n)$. Since $T(n) = O(n\lg n)$ and $T(n) = \Omega(n\lg n)$, we conclude that $T(n) = \Theta(n\lg n)$.

Lecture Notes for Chapter 5: Probabilistic Analysis and Randomized Algorithms

[This chapter introduces probabilistic analysis and randomized algorithms. It assumes that the student is familiar with the basic probability material in Appendix C.

The primary goals of these notes are to

- *explain the difference between probabilistic analysis and randomized algorithms,*
- *present the technique of indicator random variables, and*
- *give another example of the analysis of a randomized algorithm (permuting an array in place).*

These notes omit the technique of permuting an array by sorting, and they omit the starred Section 5.4.]

The hiring problem

Scenario

- You are using an employment agency to hire a new office assistant.
- The agency sends you one candidate each day.
- You interview the candidate and must immediately decide whether or not to hire that person. But if you hire, you must also fire your current office assistant—even if it's someone you have recently hired.
- Cost to interview is c_i per candidate (interview fee paid to agency).
- Cost to hire is c_h per candidate (includes cost to fire current office assistant + hiring fee paid to agency).
- Assume that $c_h > c_i$.
- You are committed to having hired, at all times, the best candidate seen so far. Meaning that whenever you interview a candidate who is better than your current office assistant, you must fire the current office assistant and hire the candidate. Since you must have someone hired at all times, you will always hire the first candidate that you interview.

Goal

Determine what the price of this strategy will be.

Pseudocode to model this scenario

Assumes that the candidates are numbered 1 to n and that after interviewing each candidate, we can determine if it's better than the current office assistant. Uses a dummy candidate 0 that is worse than all others, so that the first candidate is always hired.

```
HIRE-ASSISTANT(n)
  best = 0          // candidate 0 is a least-qualified dummy candidate
  for i = 1 to n
      interview candidate i
      if candidate i is better than candidate best
          best = i
          hire candidate i
```

Cost

If n candidates, and we hire m of them, the cost is $O(nc_i + mc_h)$.

- Have to pay nc_i to interview, no matter how many we hire.
- So we focus on analyzing the hiring cost mc_h.
- mc_h varies with each run—it depends on the order in which we interview the candidates.
- This is a model of a common paradigm: we need to find the maximum or minimum in a sequence by examining each element and maintaining a current "winner." The variable m denotes how many times we change our notion of which element is currently winning.

Worst-case analysis

In the worst case, we hire all n candidates.

This happens if each one is better than all who came before. In other words, if the candidates appear in increasing order of quality.

If we hire all n, then the cost is $O(nc_i + nc_h) = O(nc_h)$ (since $c_h > c_i$).

Probabilistic analysis

In general, we have no control over the order in which candidates appear.

We could assume that they come in a random order:

- Assign a rank to each candidate: $rank(i)$ is a unique integer in the range 1 to n.
- The ordered list $\langle rank(1), rank(2), \ldots, rank(n)\rangle$ is a permutation of the candidate numbers $\langle 1, 2, \ldots, n\rangle$.
- The list of ranks is equally likely to be any one of the $n!$ permutations.
- Equivalently, the ranks form a ***uniform random permutation***: each of the possible $n!$ permutations appears with equal probability.

Essential idea of probabilistic analysis

We must use knowledge of, or make assumptions about, the distribution of inputs.

- The expectation is over this distribution.
- This technique requires that we can make a reasonable characterization of the input distribution.

Randomized algorithms

We might not know the distribution of inputs, or we might not be able to model it computationally.

Instead, we use randomization within the algorithm in order to impose a distribution on the inputs.

For the hiring problem

Change the scenario:

- The employment agency sends us a list of all n candidates in advance.
- On each day, we randomly choose a candidate from the list to interview (but considering only those we have not yet interviewed).
- Instead of relying on the candidates being presented to us in a random order, we take control of the process and enforce a random order.

What makes an algorithm randomized

An algorithm is ***randomized*** if its behavior is determined in part by values produced by a ***random-number generator***.

- RANDOM(a, b) returns an integer r, where $a \le r \le b$ and each of the $b - a + 1$ possible values of r is equally likely.
- In practice, RANDOM is implemented by a ***pseudorandom-number generator***, which is a deterministic method returning numbers that "look" random and pass statistical tests.

Indicator random variables

A simple yet powerful technique for computing the expected value of a random variable.

Helpful in situations in which there may be dependence.

Given a sample space and an event A, we define the ***indicator random variable***

$$\mathrm{I}\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs ,} \\ 0 & \text{if } A \text{ does not occur .} \end{cases}$$

Lemma

For an event A, let $X_A = \mathrm{I}\{A\}$. Then $\mathrm{E}[X_A] = \Pr\{A\}$.

Proof Letting $\overline{A}$ be the complement of A, we have

$$\begin{aligned} \mathrm{E}[X_A] &= \mathrm{E}[\mathrm{I}\{A\}] \\ &= 1 \cdot \Pr\{A\} + 0 \cdot \Pr\{\overline{A}\} \quad \text{(definition of expected value)} \\ &= \Pr\{A\} \ . \end{aligned}$$

■ (lemma)

Simple example

Determine the expected number of heads when we flip a fair coin one time.

- Sample space is $\{H, T\}$.
- $\Pr\{H\} = \Pr\{T\} = 1/2$.
- Define indicator random variable $X_H = \mathrm{I}\{H\}$. X_H counts the number of heads in one flip.
- Since $\Pr\{H\} = 1/2$, lemma says that $\mathrm{E}[X_H] = 1/2$.

Slightly more complicated example

Determine the expected number of heads in n coin flips.

- Let X be a random variable for the number of heads in n flips.
- Could compute $\mathrm{E}[X] = \sum_{k=0}^{n} k \cdot \Pr\{X = k\}$. In fact, this is what the book does in equation (C.37).
- Instead, we'll use indicator random variables.
- For $i = 1, 2, \ldots, n$, define $X_i = \mathrm{I}\{\text{the } i\text{th flip results in event } H\}$.
- Then $X = \sum_{i=1}^{n} X_i$.
- Lemma says that $\mathrm{E}[X_i] = \Pr\{H\} = 1/2$ for $i = 1, 2, \ldots, n$.
- Expected number of heads is $\mathrm{E}[X] = \mathrm{E}[\sum_{i=1}^{n} X_i]$.
- ***Problem:*** We want $\mathrm{E}[\sum_{i=1}^{n} X_i]$. We have only the individual expectations $\mathrm{E}[X_1], \mathrm{E}[X_2], \ldots, \mathrm{E}[X_n]$.
- ***Solution:*** Linearity of expectation says that the expectation of the sum equals the sum of the expectations. Thus,

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} \mathrm{E}[X_i] \\ &= \sum_{i=1}^{n} 1/2 \\ &= n/2 \ . \end{aligned}$$

- Linearity of expectation applies even when there is dependence among the random variables. *[Not an issue in this example, but it can be a great help. The hat-check problem of Exercise 5.2-4 is a problem with lots of dependence. See the solution on page 5-11 of this manual.]*

Analysis of the hiring problem

Assume that the candidates arrive in a random order.

Let X be a random variable that equals the number of times we hire a new office assistant.

Define indicator random variables $X_1, X_2, \ldots, X_n$, where

$X_i = \mathrm{I}\{\text{candidate } i \text{ is hired}\}$.

Useful properties:

- $X = X_1 + X_2 + \cdots + X_n$.
- Lemma $\Rightarrow \mathrm{E}[X_i] = \Pr\{\text{candidate } i \text{ is hired}\}$.

We need to compute $\Pr\{\text{candidate } i \text{ is hired}\}$.

- Candidate i is hired if and only if candidate i is better than each of candidates $1, 2, \ldots, i-1$.
- Assumption that the candidates arrive in random order $\Rightarrow$ candidates $1, 2, \ldots, i$ arrive in random order $\Rightarrow$ any one of these first i candidates is equally likely to be the best one so far.
- Thus, $\Pr\{\text{candidate } i \text{ is the best so far}\} = 1/i$.
- Which implies $\mathrm{E}[X_i] = 1/i$.

Now compute $\mathrm{E}[X]$:

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \sum_{i=1}^{n} 1/i \\
&= \ln n + O(1) \quad \text{(equation (A.7): the sum is a harmonic series)} \ .
\end{aligned}$$

Thus, the expected hiring cost is $O(c_h \ln n)$, which is much better than the worst-case cost of $O(n c_h)$.

Randomized algorithms

Instead of assuming a distribution of the inputs, we impose a distribution.

The hiring problem

For the hiring problem, the algorithm is deterministic:

- For any given input, the number of times we hire a new office assistant will always be the same.

- The number of times we hire a new office assistant depends only on the input.
- In fact, it depends only on the ordering of the candidates' ranks that it is given.
- Some rank orderings will always produce a high hiring cost. Example: $\langle 1, 2, 3, 4, 5, 6\rangle$, where each candidate is hired.
- Some will always produce a low hiring cost. Example: any ordering in which the best candidate is the first one interviewed. Then only the best candidate is hired.
- Some may be in between.

Instead of always interviewing the candidates in the order presented, what if we first randomly permuted this order?

- The randomization is now in the algorithm, not in the input distribution.
- Given a particular input, we can no longer say what its hiring cost will be. Each time we run the algorithm, we can get a different hiring cost.
- In other words, each time we run the algorithm, the execution depends on the random choices made.
- No particular input always elicits worst-case behavior.
- Bad behavior occurs only if we get "unlucky" numbers from the random-number generator.

Pseudocode for randomized hiring problem

RANDOMIZED-HIRE-ASSISTANT(n)

  randomly permute the list of candidates
  HIRE-ASSISTANT(n)

Lemma

The expected hiring cost of RANDOMIZED-HIRE-ASSISTANT is $O(c_h \ln n)$.

Proof After permuting the input array, we have a situation identical to the probabilistic analysis of deterministic HIRE-ASSISTANT. ■

Randomly permuting an array

[The book considers two methods of randomly permuting an n-element array. The first method assigns a random priority in the range 1 to n^3 to each position and then reorders the array elements into increasing priority order. We omit this method from these notes. The second method is better: it works in place (unlike the priority-based method), it runs in linear time without requiring sorting, and it needs fewer random bits (n random numbers in the range 1 to n rather than the range 1 to n^3). We present and analyze the second method in these notes.]

Goal

Produce a uniform random permutation (each of the $n!$ permutations is equally likely to be produced).

Non-goal: Show that for each element $A[i]$, the probability that $A[i]$ moves to position j is $1/n$. (See Exercise 5.3-4, whose solution is on page 5-14 of this manual.)

The following procedure permutes the array $A[1..n]$ in place (i.e., no auxiliary array is required).

RANDOMIZE-IN-PLACE(A, n)

for $i = 1$ **to** n
    swap $A[i]$ with $A[\text{RANDOM}(i, n)]$

Idea

- In iteration i, choose $A[i]$ randomly from $A[i..n]$.
- Will never alter $A[i]$ after iteration i.

Time

$O(1)$ per iteration $\Rightarrow$ $O(n)$ total.

Correctness

Given a set of n elements, a ***k-permutation*** is a sequence containing k of the n elements. There are $n!/(n-k)!$ possible k-permutations.

Lemma

RANDOMIZE-IN-PLACE computes a uniform random permutation.

Proof Use a loop invariant:

> **Loop invariant:** Just prior to the ith iteration of the **for** loop, for each possible $(i-1)$-permutation, subarray $A[1..i-1]$ contains this $(i-1)$-permutation with probability $(n-i+1)!/n!$.

Initialization: Just before first iteration, $i = 1$. Loop invariant says that for each possible 0-permutation, subarray $A[1..0]$ contains this 0-permutation with probability $n!/n! = 1$. $A[1..0]$ is an empty subarray, and a 0-permutation has no elements. So, $A[1..0]$ contains any 0-permutation with probability 1.

Maintenance: Assume that just prior to the ith iteration, each possible $(i-1)$-permutation appears in $A[1..i-1]$ with probability $(n-i+1)!/n!$. Will show that after the ith iteration, each possible i-permutation appears in $A[1..i]$ with probability $(n-i)!/n!$. Incrementing i for the next iteration then maintains the invariant.

Consider a particular i-permutation $\pi = \langle x_1, x_2, \ldots, x_i \rangle$. It consists of an $(i-1)$-permutation $\pi' = \langle x_1, x_2, \ldots, x_{i-1} \rangle$, followed by x_i.

Let E_1 be the event that the algorithm actually puts π' into $A[1..i-1]$. By the loop invariant, $\Pr\{E_1\} = (n-i+1)!/n!$.

Let E_2 be the event that the ith iteration puts x_i into $A[i]$.

We get the i-permutation π in $A[1..i]$ if and only if both E_1 and E_2 occur $\Rightarrow$ the probability that the algorithm produces π in $A[1..i]$ is $\Pr\{E_2 \cap E_1\}$.

Equation (C.14) $\Rightarrow \Pr\{E_2 \cap E_1\} = \Pr\{E_2 \mid E_1\} \Pr\{E_1\}$.

The algorithm chooses x_i randomly from the $n - i + 1$ possibilities in $A[i..n]$ $\Rightarrow \Pr\{E_2 \mid E_1\} = 1/(n - i + 1)$. Thus,

$$\begin{aligned}
\Pr\{E_2 \cap E_1\} &= \Pr\{E_2 \mid E_1\} \Pr\{E_1\} \\
&= \frac{1}{n - i + 1} \cdot \frac{(n - i + 1)!}{n!} \\
&= \frac{(n - i)!}{n!} \, .
\end{aligned}$$

Termination: At termination, $i = n + 1$, so we conclude that $A[1..n]$ is a given n-permutation with probability $(n - n)!/n! = 1/n!$. ■ (lemma)

Solutions for Chapter 5: Probabilistic Analysis and Randomized Algorithms

Solution to Exercise 5.1-3

To get an unbiased random bit, given only calls to BIASED-RANDOM, call BIASED-RANDOM twice. Repeatedly do so until the two calls return different values, and when this occurs, return the first of the two bits:

```
UNBIASED-RANDOM
  while TRUE
      x = BIASED-RANDOM
      y = BIASED-RANDOM
      if x ≠ y
          return x
```

To see that UNBIASED-RANDOM returns 0 and 1 each with probability $1/2$, observe that the probability that a given iteration returns 0 is

$$\Pr\{x = 0 \text{ and } y = 1\} = (1 - p)p \ ,$$

and the probability that a given iteration returns 1 is

$$\Pr\{x = 1 \text{ and } y = 0\} = p(1 - p) \ .$$

(We rely on the bits returned by BIASED-RANDOM being independent.) Thus, the probability that a given iteration returns 0 equals the probability that it returns 1. Since there is no other way for UNBIASED-RANDOM to return a value, it returns 0 and 1 each with probability $1/2$.

Assuming that each iteration takes $O(1)$ time, the expected running time of UNBIASED-RANDOM is linear in the expected number of iterations. We can view each iteration as a Bernoulli trial, where "success" means that the iteration returns a value. The probability of success equals the probability that 0 is returned plus the probability that 1 is returned, or $2p(1 - p)$. The number of trials until a success occurs is given by the geometric distribution, and by equation (C.32), the expected number of trials for this scenario is $1/(2p(1 - p))$. Thus, the expected running time of UNBIASED-RANDOM is $\Theta(1/(2p(1 - p))$.

Solution to Exercise 5.2-1

This solution is also posted publicly

Since HIRE-ASSISTANT always hires candidate 1, it hires exactly once if and only if no candidates other than candidate 1 are hired. This event occurs when candidate 1 is the best candidate of the n, which occurs with probability $1/n$.

HIRE-ASSISTANT hires n times if each candidate is better than all those who were interviewed (and hired) before. This event occurs precisely when the list of ranks given to the algorithm is $\langle 1, 2, \ldots, n \rangle$, which occurs with probability $1/n!$.

Solution to Exercise 5.2-2

We make three observations:

1. Candidate 1 is always hired.
2. The best candidate, i.e., the one whose rank is n, is always hired.
3. If the best candidate is candidate 1, then that is the only candidate hired.

Therefore, in order for HIRE-ASSISTANT to hire exactly twice, candidate 1 must have rank $i \le n-1$ and all candidates whose ranks are $i+1, i+2, \ldots, n-1$ must be interviewed after the candidate whose rank is n. (When $i = n-1$, this second condition vacuously holds.)

Let E_i be the event in which candidate 1 has rank i; clearly, $\Pr\{E_i\} = 1/n$ for any given value of i.

Letting j denote the position in the interview order of the best candidate, let F be the event in which candidates $2, 3, \ldots, j-1$ have ranks strictly less than the rank of candidate 1. Given that event E_i has occurred, event F occurs when the best candidate is the first one interviewed out of the $n-i$ candidates whose ranks are $i+1, i+2, \ldots, n$. Thus, $\Pr\{F \mid E_i\} = 1/(n-i)$.

Our final event is A, which occurs when HIRE-ASSISTANT hires exactly twice. Noting that the events $E_1, E_2, \ldots, E_n$ are disjoint, we have

$$\begin{aligned} A &= F \cap (E_1 \cup E_2 \cup \cdots \cup E_{n-1}) \\ &= (F \cap E_1) \cup (F \cap E_2) \cup \cdots \cup (F \cap E_{n-1}) \ . \end{aligned}$$

and

$$\Pr\{A\} = \sum_{i=1}^{n-1} \Pr\{F \cap E_i\} \ .$$

By equation (C.14),

$$\begin{aligned} \Pr\{F \cap E_i\} &= \Pr\{F \mid E_i\} \Pr\{E_i\} \\ &= \frac{1}{n-i} \cdot \frac{1}{n} \ , \end{aligned}$$

and so

$$\begin{aligned}
\Pr\{A\} &= \sum_{i=1}^{n-1} \frac{1}{n-i} \cdot \frac{1}{n} \\
&= \frac{1}{n} \sum_{i=1}^{n-1} \frac{1}{n-i} \\
&= \frac{1}{n} \left(\frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{1} \right) \\
&= \frac{1}{n} \cdot H_{n-1} \ ,
\end{aligned}$$

where H_{n-1} is the nth harmonic number.

Solution to Exercise 5.2-4

This solution is also posted publicly

Another way to think of the hat-check problem is that we want to determine the expected number of fixed points in a random permutation. (A ***fixed point*** of a permutation π is a value i for which $\pi(i) = i$.) We could enumerate all $n!$ permutations, count the total number of fixed points, and divide by $n!$ to determine the average number of fixed points per permutation. This would be a painstaking process, and the answer would turn out to be 1. We can use indicator random variables, however, to arrive at the same answer much more easily.

Define a random variable X that equals the number of customers that get back their own hat, so that we want to compute $\mathrm{E}[X]$.

For $i = 1, 2, \ldots, n$, define the indicator random variable

$X_i = \mathrm{I}\{\text{customer } i \text{ gets back his own hat}\}$.

Then $X = X_1 + X_2 + \cdots + X_n$.

Since the ordering of hats is random, each customer has a probability of $1/n$ of getting back his or her own hat. In other words, $\Pr\{X_i = 1\} = 1/n$, which, by Lemma 5.1, implies that $\mathrm{E}[X_i] = 1/n$.

Thus,

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i \right] \\
&= \sum_{i=1}^{n} \mathrm{E}[X_i] \qquad \text{(linearity of expectation)} \\
&= \sum_{i=1}^{n} 1/n \\
&= 1 \ ,
\end{aligned}$$

and so we expect that exactly 1 customer gets back his own hat.

Note that this is a situation in which the indicator random variables are *not* independent. For example, if $n = 2$ and $X_1 = 1$, then X_2 must also equal 1. Conversely, if $n = 2$ and $X_1 = 0$, then X_2 must also equal 0. Despite the dependence, $\Pr\{X_i = 1\} = 1/n$ for all i, and linearity of expectation holds. Thus, we can use the technique of indicator random variables even in the presence of dependence.

Solution to Exercise 5.2-5

This solution is also posted publicly

Let X_{ij} be an indicator random variable for the event where the pair $A[i], A[j]$ for $i < j$ is inverted, i.e., $A[i] > A[j]$. More precisely, we define $X_{ij} = \mathrm{I}\{A[i] > A[j]\}$ for $1 \le i < j \le n$. We have $\Pr\{X_{ij} = 1\} = 1/2$, because given two distinct random numbers, the probability that the first is bigger than the second is $1/2$. By Lemma 5.1, $\mathrm{E}[X_{ij}] = 1/2$.

Let X be the the random variable denoting the total number of inverted pairs in the array, so that

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \ .$$

We want the expected number of inverted pairs, so we take the expectation of both sides of the above equation to obtain

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right] \ .$$

We use linearity of expectation to get

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right] \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{E}[X_{ij}] \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1/2 \\
&= \binom{n}{2}\frac{1}{2} \\
&= \frac{n(n-1)}{2} \cdot \frac{1}{2} \\
&= \frac{n(n-1)}{4} \ .
\end{aligned}$$

Thus the expected number of inverted pairs is $n(n-1)/4$.

Solution to Exercise 5.3-1

Here's the rewritten procedure:

```
RANDOMIZE-IN-PLACE(A)
  n = A.length
  swap A[1] with A[RANDOM(1, n)]
  for i = 2 to n
      swap A[i] with A[RANDOM(i, n)]
```

The loop invariant becomes

> **Loop invariant:** Just prior to the iteration of the **for** loop for each value of $i = 2, \ldots, n$, for each possible $(i-1)$-permutation, the subarray $A[1 \,..\, i-1]$ contains this $(i-1)$-permutation with probability $(n-i+1)!/n!$.

The maintenance and termination parts remain the same. The initialization part is for the subarray $A[1 \,..\, 1]$, which contains any 1-permutation with probability $(n-1)!/n! = 1/n$.

Solution to Exercise 5.3-2

This solution is also posted publicly

Although PERMUTE-WITHOUT-IDENTITY will not produce the identity permutation, there are other permutations that it fails to produce. For example, consider its operation when $n = 3$, when it should be able to produce the $n! - 1 = 5$ non-identity permutations. The **for** loop iterates for $i = 1$ and $i = 2$. When $i = 1$, the call to RANDOM returns one of two possible values (either 2 or 3), and when $i = 2$, the call to RANDOM returns just one value (3). Thus, PERMUTE-WITHOUT-IDENTITY can produce only $2 \cdot 1 = 2$ possible permutations, rather than the 5 that are required.

Solution to Exercise 5.3-3

The PERMUTE-WITH-ALL procedure does not produce a uniform random permutation. Consider the permutations it produces when $n = 3$. The procedure makes 3 calls to RANDOM, each of which returns one of 3 values, and so calling PERMUTE-WITH-ALL has 27 possible outcomes. Since there are $3! = 6$ permutations, if PERMUTE-WITH-ALL did produce a uniform random permutation, then each permutation would occur $1/6$ of the time. That would mean that each permutation would have to occur an integer number m times, where $m/27 = 1/6$. No integer m satisfies this condition.

In fact, if we were to work out the possible permutations of $\langle 1, 2, 3 \rangle$ and how often they occur with PERMUTE-WITH-ALL, we would get the following probabilities:

permutation	probability
$\langle 1,2,3\rangle$	$4/27$
$\langle 1,3,2\rangle$	$5/27$
$\langle 2,1,3\rangle$	$5/27$
$\langle 2,3,1\rangle$	$5/27$
$\langle 3,1,2\rangle$	$4/27$
$\langle 3,2,1\rangle$	$4/27$

Although these probabilities sum to 1, none are equal to $1/6$.

Solution to Exercise 5.3-4

This solution is also posted publicly

PERMUTE-BY-CYCLIC chooses *offset* as a random integer in the range $1 \le \textit{offset} \le n$, and then it performs a cyclic rotation of the array. That is, $B[((i + \textit{offset} - 1) \bmod n) + 1] = A[i]$ for $i = 1, 2, \ldots, n$. (The subtraction and addition of 1 in the index calculation is due to the 1-origin indexing. If we had used 0-origin indexing instead, the index calculation would have simplied to $B[(i + \textit{offset}) \bmod n] = A[i]$ for $i = 0, 1, \ldots, n-1$.)

Thus, once *offset* is determined, so is the entire permutation. Since each value of *offset* occurs with probability $1/n$, each element $A[i]$ has a probability of ending up in position $B[j]$ with probability $1/n$.

This procedure does not produce a uniform random permutation, however, since it can produce only n different permutations. Thus, n permutations occur with probability $1/n$, and the remaining $n! - n$ permutations occur with probability 0.

Solution to Exercise 5.3-7

Since each recursive call reduces m by 1 and makes only one call to RANDOM, it's easy to see that there are a total of m calls to RANDOM. Moreover, since each recursive call adds exactly one element to the set, it's easy to see that the resulting set S contains exactly m elements.

Because the elements of set S are chosen independently of each other, it suffices to show that each of the n values appears in S with probability m/n. We use an inductive proof. The inductive hypothesis is that a call to RANDOM-SUBSET(m, n) returns a set S of m elements, each appearing with probability m/n. The base cases are for $m = 0$ and $m = 1$. When $m = 0$, the returned set is empty, and so it contains each element with probability 0. When $m = 1$, the returned set has one element, and it is equally likely to be any number in $\{1, 2, 3, \ldots, n\}$.

For the inductive step, we assume that the call RANDOM-SUBSET$(m-1, n-1)$ returns a set S' of $m-1$ elements in which each value in $\{1, 2, 3, \ldots, n-1\}$ occurs with probability $(m-1)/(n-1)$. After the line $i =$ RANDOM$(1, n)$, i is equally likely to be any value in $\{1, 2, 3, \ldots, n\}$. We consider separately the probabilities

that S contains $j < n$ and that S contains n. Let R_j be the event that the call RANDOM$(1, n)$ returns j, so that $\Pr\{R_j\} = 1/n$.

For $j < n$, the event that $j \in S$ is the union of two disjoint events:

- $j \in S'$, and
- $j \notin S'$ and R_j (these events are independent),

Thus,

$$\begin{aligned}
&\Pr\{j \in S\} \\
&\quad = \Pr\{j \in S'\} + \Pr\{j \notin S' \text{ and } R_j\} && \text{(the events are disjoint)} \\
&\quad = \frac{m-1}{n-1} + \left(1 - \frac{m-1}{n-1}\right) \cdot \frac{1}{n} && \text{(by the inductive hypothesis)} \\
&\quad = \frac{m-1}{n-1} + \left(\frac{n-1}{n-1} - \frac{m-1}{n-1}\right) \cdot \frac{1}{n} \\
&\quad = \frac{m-1}{n-1} \cdot \frac{n}{n} + \frac{n-m}{n-1} \cdot \frac{1}{n} \\
&\quad = \frac{(m-1)n + (n-m)}{(n-1)n} \\
&\quad = \frac{mn - n + n - m}{(n-1)n} \\
&\quad = \frac{m(n-1)}{(n-1)n} \\
&\quad = \frac{m}{n} .
\end{aligned}$$

The event that $n \in S$ is also the union of two disjoint events:

- R_n, and
- R_j and $j \in S'$ for some $j < n$ (these events are independent).

Thus,

$$\begin{aligned}
&\Pr\{n \in S\} \\
&= \Pr\{R_n\} + \Pr\{R_j \text{ and } j \in S' \text{ for some } j < n\} && \text{(the events are disjoint)} \\
&= \frac{1}{n} + \frac{n-1}{n} \cdot \frac{m-1}{n-1} && \text{(by the inductive hypothesis)} \\
&= \frac{1}{n} \cdot \frac{n-1}{n-1} + \frac{n-1}{n} \cdot \frac{m-1}{n-1} \\
&= \frac{n - 1 + nm - n - m + 1}{n(n-1)} \\
&= \frac{nm - m}{n(n-1)} \\
&= \frac{m(n-1)}{n(n-1)} \\
&= \frac{m}{n} .
\end{aligned}$$

Solution to Exercise 5.4-6

First we determine the expected number of empty bins. We define a random variable X to be the number of empty bins, so that we want to compute $\mathrm{E}[X]$. Next, for $i = 1, 2, \ldots, n$, we define the indicator random variable $Y_i = \mathrm{I}\{\text{bin } i \text{ is empty}\}$. Thus,

$$X = \sum_{i=1}^{n} Y_i \ ,$$

and so

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} Y_i\right] \\
&= \sum_{i=1}^{n} \mathrm{E}[Y_i] && \text{(by linearity of expectation)} \\
&= \sum_{i=1}^{n} \Pr\{\text{bin } i \text{ is empty}\} && \text{(by Lemma 5.1)} \ .
\end{aligned}$$

Let us focus on a specific bin, say bin i. We view a toss as a success if it misses bin i and as a failure if it lands in bin i. We have n independent Bernoulli trials, each with probability of success $1 - 1/n$. In order for bin i to be empty, we need n successes in n trials. Using a binomial distribution, therefore, we have that

$$\begin{aligned}
\Pr\{\text{bin } i \text{ is empty}\} &= \binom{n}{n}\left(1 - \frac{1}{n}\right)^n \left(\frac{1}{n}\right)^0 \\
&= \left(1 - \frac{1}{n}\right)^n \ .
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^n \\
&= n\left(1 - \frac{1}{n}\right)^n \ .
\end{aligned}$$

By equation (3.14), as n approaches ∞, the quantity $(1 - 1/n)^n$ approaches $1/e$, and so $\mathrm{E}[X]$ approaches n/e.

Now we determine the expected number of bins with exactly one ball. We redefine X to be number of bins with exactly one ball, and we redefine Y_i to be $\mathrm{I}\{\text{bin } i \text{ gets exactly one ball}\}$. As before, we find that

$$\mathrm{E}[X] = \sum_{i=1}^{n} \Pr\{\text{bin } i \text{ gets exactly one ball}\} \ .$$

Again focusing on bin i, we need exactly $n-1$ successes in n independent Bernoulli trials, and so

$$\begin{aligned}\Pr\{\text{bin } i \text{ gets exactly one ball}\} &= \binom{n}{n-1}\left(1-\frac{1}{n}\right)^{n-1}\left(\frac{1}{n}\right)^{1} \\ &= n\cdot\left(1-\frac{1}{n}\right)^{n-1}\frac{1}{n} \\ &= \left(1-\frac{1}{n}\right)^{n-1} ,\end{aligned}$$

and so

$$\begin{aligned}\mathrm{E}[X] &= \sum_{i=1}^{n}\left(1-\frac{1}{n}\right)^{n-1} \\ &= n\left(1-\frac{1}{n}\right)^{n-1} .\end{aligned}$$

Because

$$n\left(1-\frac{1}{n}\right)^{n-1} = \frac{n\left(1-\frac{1}{n}\right)^{n}}{1-\frac{1}{n}} ,$$

as n approaches ∞, we find that $\mathrm{E}[X]$ approaches

$$\frac{n/e}{1-1/n} = \frac{n^2}{e(n-1)} .$$

Solution to Problem 5-1

a. To determine the expected value represented by the counter after n INCREMENT operations, we define some random variables:

- For $j = 1, 2, \ldots, n$, let X_j denote the increase in the value represented by the counter due to the jth INCREMENT operation.
- Let V_n be the value represented by the counter after n INCREMENT operations.

Then $V_n = X_1 + X_2 + \cdots + X_n$. We want to compute $\mathrm{E}[V_n]$. By linearity of expectation,

$$\mathrm{E}[V_n] = \mathrm{E}[X_1 + X_2 + \cdots + X_n] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n] .$$

We shall show that $\mathrm{E}[X_j] = 1$ for $j = 1, 2, \ldots, n$, which will prove that $\mathrm{E}[V_n] = n$.

We actually show that $\mathrm{E}[X_j] = 1$ in two ways, the second more rigorous than the first:

1. Suppose that at the start of the jth INCREMENT operation, the counter holds the value i, which represents n_i. If the counter increases due to this INCREMENT operation, then the value it represents increases by $n_{i+1} - n_i$. The counter increases with probability $1/(n_{i+1} - n_i)$, and so

$$\begin{aligned}
\mathrm{E}[X_j] &= (0 \cdot \Pr\{\text{counter does not increase}\}) \\
&\quad + ((n_{i+1} - n_i) \cdot \Pr\{\text{counter increases}\}) \\
&= \left(0 \cdot \left(1 - \frac{1}{n_{i+1} - n_i}\right)\right) + \left((n_{i+1} - n_i) \cdot \frac{1}{n_{i+1} - n_i}\right) \\
&= 1 \,,
\end{aligned}$$

and so $\mathrm{E}[X_j] = 1$ regardless of the value held by the counter.

2. Let C_j be the random variable denoting the value held in the counter at the start of the jth INCREMENT operation. Since we can ignore values of C_j greater than $2^b - 1$, we use a formula for conditional expectation:

$$\begin{aligned}
\mathrm{E}[X_j] &= \mathrm{E}[\mathrm{E}[X_j \mid C_j]] \\
&= \sum_{i=0}^{2^b-1} \mathrm{E}[X_j \mid C_j = i] \cdot \Pr\{C_j = i\} \,.
\end{aligned}$$

To compute $\mathrm{E}[X_j \mid C_j = i]$, we note that

- $\Pr\{X_j = 0 \mid C_j = i\} = 1 - 1/(n_{i+1} - n_i)$,
- $\Pr\{X_j = n_{i+1} - n_i \mid C_j = i\} = 1/(n_{i+1} - n_i)$, and
- $\Pr\{X_j = k \mid C_j = i\} = 0$ for all other k.

Thus,

$$\begin{aligned}
\mathrm{E}[X_j \mid C_j = i] &= \sum_k k \cdot \Pr\{X_j = k \mid C_j = i\} \\
&= \left(0 \cdot \left(1 - \frac{1}{n_{i+1} - n_i}\right)\right) + \left((n_{i+1} - n_i) \cdot \frac{1}{n_{i+1} - n_i}\right) \\
&= 1 \,.
\end{aligned}$$

Therefore, noting that

$$\sum_{i=0}^{2^b-1} \Pr\{C_j = i\} = 1 \,,$$

we have

$$\begin{aligned}
\mathrm{E}[X_j] &= \sum_{i=0}^{2^b-1} 1 \cdot \Pr\{C_j = i\} \\
&= 1 \,.
\end{aligned}$$

Why is the second way more rigorous than the first? Both ways condition on the value held in the counter, but only the second way incorporates the conditioning into the expression for $\mathrm{E}[X_j]$.

b. Defining V_n and X_j as in part (a), we want to compute $\mathrm{Var}[V_n]$, where $n_i = 100i$. The X_j are pairwise independent, and so by equation (C.29), $\mathrm{Var}[V_n] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots + \mathrm{Var}[X_n]$.

Since $n_i = 100i$, we see that $n_{i+1} - n_i = 100(i+1) - 100i = 100$. Therefore, with probability 99/100, the increase in the value represented by the counter due to the jth INCREMENT operation is 0, and with probability 1/100, the

value represented increases by 100. Thus, by equation (C.27),

$$\begin{aligned}
\mathrm{Var}\left[X_j\right] &= \mathrm{E}\left[X_j^2\right] - \mathrm{E}^2\left[X_j\right] \\
&= \left(\left(0^2 \cdot \frac{99}{100}\right) + \left(100^2 \cdot \frac{1}{100}\right)\right) - 1^2 \\
&= 100 - 1 \\
&= 99\,.
\end{aligned}$$

Summing up the variances of the X_j gives $\mathrm{Var}\left[V_n\right] = 99n$.

Lecture Notes for Chapter 6: Heapsort

Chapter 6 overview

Heapsort

- $O(n \lg n)$ worst case—like merge sort.
- Sorts in place—like insertion sort.
- Combines the best of both algorithms.

To understand heapsort, we'll cover heaps and heap operations, and then we'll take a look at priority queues.

Heaps

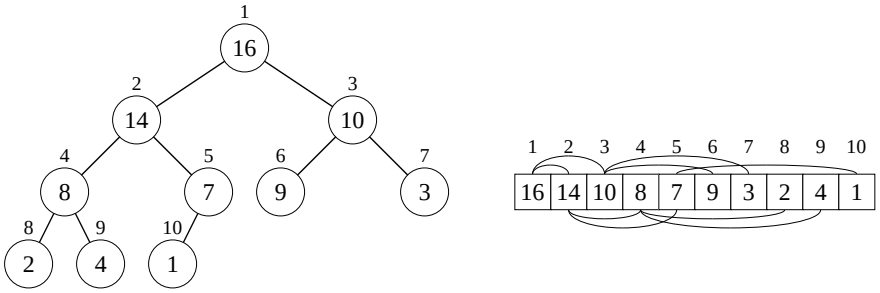
Heap data structure

- Heap A (*not* garbage-collected storage) is a nearly complete binary tree.
  - ***Height*** of node = # of edges on a longest simple path from the node down to a leaf.
  - ***Height*** of heap = height of root = $\Theta(\lg n)$.
- A heap can be stored as an array A.
  - Root of tree is $A[1]$.
  - Parent of $A[i] = A[\lfloor i/2 \rfloor]$.
  - Left child of $A[i] = A[2i]$.
  - Right child of $A[i] = A[2i + 1]$.
  - Computing is fast with binary representation implementation.

[In book, have length and heap-size attributes. Here, we bypass these attributes and use parameter values instead.]

Example

Of a max-heap. *[Arcs above and below the array on the right go between parents and children. There is no significance to whether an arc is drawn above or below the array.]*



Heap property

- For max-heaps (largest element at root), ***max-heap property:*** for all nodes i, excluding the root, $A[\text{PARENT}(i)] \geq A[i]$.
- For min-heaps (smallest element at root), ***min-heap property:*** for all nodes i, excluding the root, $A[\text{PARENT}(i)] \leq A[i]$.

By induction and transitivity of $\leq$, the max-heap property guarantees that the maximum element of a max-heap is at the root. Similar argument for min-heaps.

The heapsort algorithm we'll show uses max-heaps.

Note: In general, heaps can be k-ary tree instead of binary.

Maintaining the heap property

MAX-HEAPIFY is important for manipulating max-heaps. It is used to maintain the max-heap property.

- Before MAX-HEAPIFY, $A[i]$ may be smaller than its children.
- Assume left and right subtrees of i are max-heaps.
- After MAX-HEAPIFY, subtree rooted at i is a max-heap.

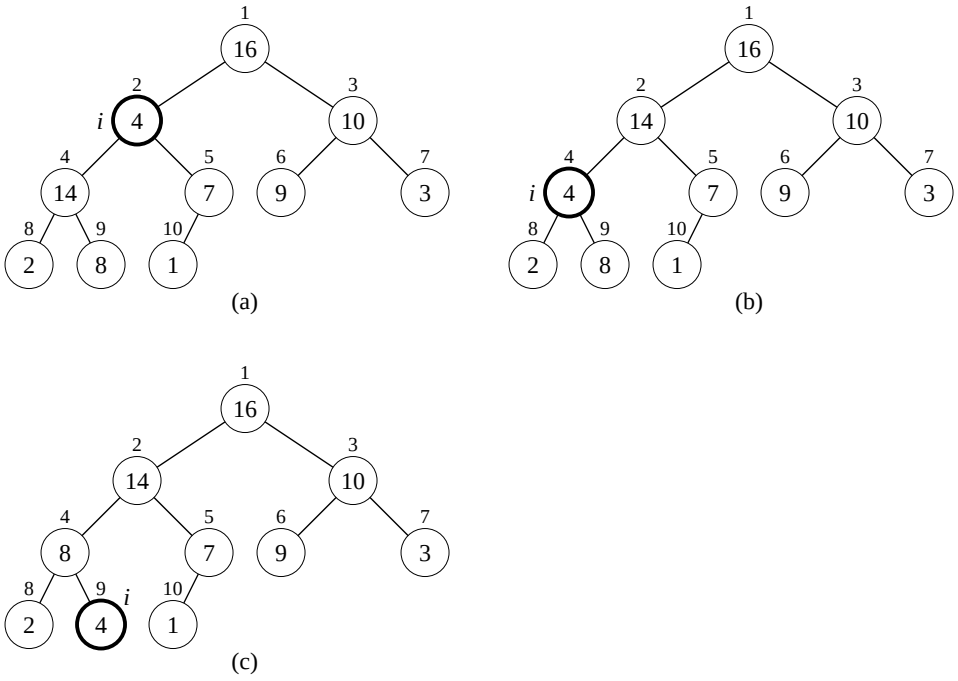
```
MAX-HEAPIFY(A, i, n)
  l = LEFT(i)
  r = RIGHT(i)
  if l ≤ n and A[l] > A[i]
      largest = l
  else largest = i
  if r ≤ n and A[r] > A[largest]
      largest = r
  if largest ≠ i
      exchange A[i] with A[largest]
      MAX-HEAPIFY(A, largest, n)
```

[Parameter n replaces attribute $A.heap\text{-}size$.]

The way MAX-HEAPIFY works:

- Compare $A[i]$, $A[\text{LEFT}(i)]$, and $A[\text{RIGHT}(i)]$.
- If necessary, swap $A[i]$ with the larger of the two children to preserve heap property.
- Continue this process of comparing and swapping down the heap, until subtree rooted at i is max-heap. If we hit a leaf, then the subtree rooted at the leaf is trivially a max-heap.

Run MAX-HEAPIFY on the following heap example.



- Node 2 violates the max-heap property.
- Compare node 2 with its children, and then swap it with the larger of the two children.
- Continue down the tree, swapping until the value is properly placed at the root of a subtree that is a max-heap. In this case, the max-heap is a leaf.

Time

$O(\lg n)$.

Analysis

[Instead of book's formal analysis with recurrence, just come up with $O(\lg n)$ intuitively.] Heap is almost-complete binary tree, hence must process $O(\lg n)$ levels, with constant work at each level (comparing 3 items and maybe swapping 2).

Building a heap

The following procedure, given an unordered array, will produce a max-heap.

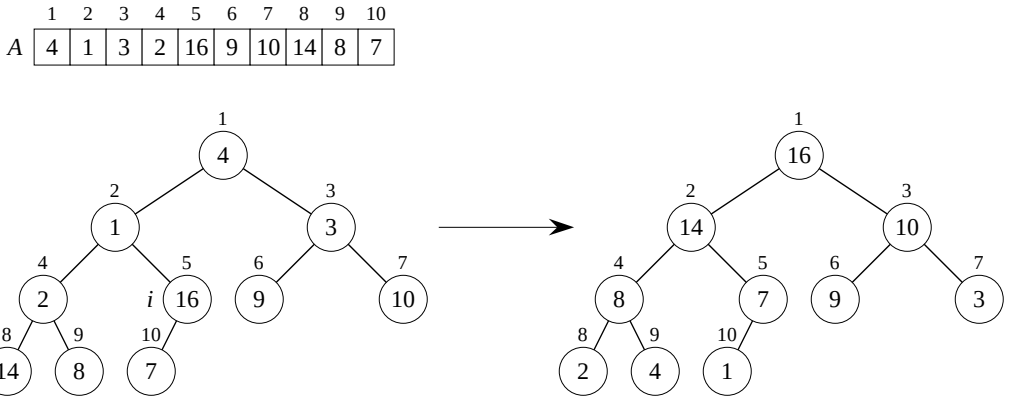
BUILD-MAX-HEAP(A, n)
  for $i = \lfloor n/2 \rfloor$ **downto** 1
    MAX-HEAPIFY(A, i, n)

[Parameter n replaces both attributes $A.length$ and $A.heap\text{-}size$.]

Example

Building a max-heap from the following unsorted array results in the first heap example.

- i starts off as 5.
- MAX-HEAPIFY is applied to subtrees rooted at nodes (in order): 16, 2, 3, 1, 4.



Correctness

Loop invariant: At start of every iteration of **for** loop, each node $i + 1$, $i + 2, \ldots, n$ is root of a max-heap.

Initialization: By Exercise 6.1-7, we know that each node $\lfloor n/2 \rfloor + 1$, $\lfloor n/2 \rfloor + 2$, $\ldots, n$ is a leaf, which is the root of a trivial max-heap. Since $i = \lfloor n/2 \rfloor$ before the first iteration of the **for** loop, the invariant is initially true.

Maintenance: Children of node i are indexed higher than i, so by the loop invariant, they are both roots of max-heaps. Correctly assuming that $i+1, i+2, \ldots, n$ are all roots of max-heaps, MAX-HEAPIFY makes node i a max-heap root. Decrementing i reestablishes the loop invariant at each iteration.

Termination: When $i = 0$, the loop terminates. By the loop invariant, each node, notably node 1, is the root of a max-heap.

Analysis

- ***Simple bound:*** $O(n)$ calls to MAX-HEAPIFY, each of which takes $O(\lg n)$ time $\Rightarrow O(n \lg n)$. (Note: A good approach to analysis in general is to start by proving easy bound, then try to tighten it.)
- ***Tighter analysis:*** Observation: Time to run MAX-HEAPIFY is linear in the height of the node it's run on, and most nodes have small heights. Have $\le \lceil n/2^{h+1} \rceil$ nodes of height h (see Exercise 6.3-3), and height of heap is $\lfloor \lg n \rfloor$ (Exercise 6.1-2).

  The time required by MAX-HEAPIFY when called on a node of height h is $O(h)$, so the total cost of BUILD-MAX-HEAP is

  $$\sum_{h=0}^{\lfloor \lg n \rfloor} \left\lceil \frac{n}{2^{h+1}} \right\rceil O(h) = O\left(n \sum_{h=0}^{\lfloor \lg n \rfloor} \frac{h}{2^h} \right) .$$

  Evaluate the last summation by substituting $x = 1/2$ in the formula (A.8) $\left(\sum_{k=0}^{\infty} k x^k\right)$, which yields

  $$\begin{aligned} \sum_{h=0}^{\infty} \frac{h}{2^h} &= \frac{1/2}{(1-1/2)^2} \\ &= 2 . \end{aligned}$$

  Thus, the running time of BUILD-MAX-HEAP is $O(n)$.

  Building a min-heap from an unordered array can be done by calling MIN-HEAPIFY instead of MAX-HEAPIFY, also taking linear time.

The heapsort algorithm

Given an input array, the heapsort algorithm acts as follows:

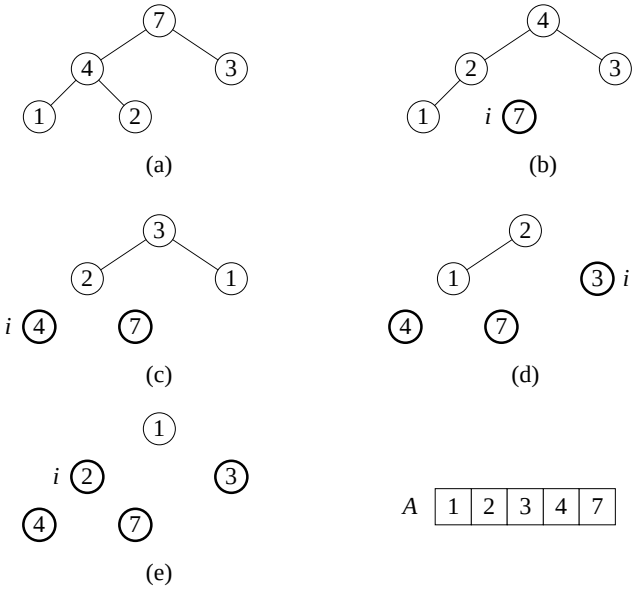
- Builds a max-heap from the array.
- Starting with the root (the maximum element), the algorithm places the maximum element into the correct place in the array by swapping it with the element in the last position in the array.
- "Discard" this last node (knowing that it is in its correct place) by decreasing the heap size, and calling MAX-HEAPIFY on the new (possibly incorrectly-placed) root.
- Repeat this "discarding" process until only one node (the smallest element) remains, and therefore is in the correct place in the array.

```
HEAPSORT(A, n)
  BUILD-MAX-HEAP(A, n)
  for i = n downto 2
      exchange A[1] with A[i]
      MAX-HEAPIFY(A, 1, i − 1)
```

[Parameter n replaces $A.length$, and parameter value $i - 1$ in MAX-HEAPIFY *call replaces decrementing of $A.heap\text{-}size$.]*

Example

Sort an example heap on the board. *[Nodes with heavy outline are no longer in the heap.]*



Analysis

- BUILD-MAX-HEAP: $O(n)$
- **for** loop: $n - 1$ times
- exchange elements: $O(1)$
- MAX-HEAPIFY: $O(\lg n)$

Total time: $O(n \lg n)$.

Though heapsort is a great algorithm, a well-implemented quicksort usually beats it in practice.

Heap implementation of priority queue

Heaps efficiently implement priority queues. These notes will deal with max-priority queues implemented with max-heaps. Min-priority queues are implemented with min-heaps similarly.

A heap gives a good compromise between fast insertion but slow extraction and vice versa. Both operations take $O(\lg n)$ time.

Priority queue

- Maintains a dynamic set S of elements.
- Each set element has a ***key***—an associated value.

- Max-priority queue supports dynamic-set operations:
  - INSERT(S, x): inserts element x into set S.
  - MAXIMUM(S): returns element of S with largest key.
  - EXTRACT-MAX(S): removes and returns element of S with largest key.
  - INCREASE-KEY(S, x, k): increases value of element x's key to k. Assume $k \geq x$'s current key value.
- Example max-priority queue application: schedule jobs on shared computer.
- Min-priority queue supports similar operations:
  - INSERT(S, x): inserts element x into set S.
  - MINIMUM(S): returns element of S with smallest key.
  - EXTRACT-MIN(S): removes and returns element of S with smallest key.
  - DECREASE-KEY(S, x, k): decreases value of element x's key to k. Assume $k \leq x$'s current key value.
- Example min-priority queue application: event-driven simulator.

Note: Actual implementations often have a ***handle*** in each heap element that allows access to an object in the application, and objects in the application often have a handle (likely an array index) to access the heap element.

Will examine how to implement max-priority queue operations.

Finding the maximum element

Getting the maximum element is easy: it's the root.

HEAP-MAXIMUM(A)
  return $A[1]$

Time

$\Theta(1)$.

Extracting max element

Given the array A:

- Make sure heap is not empty.
- Make a copy of the maximum element (the root).
- Make the last node in the tree the new root.
- Re-heapify the heap, with one fewer node.
- Return the copy of the maximum element.

Note: Because we need to decrement the heap size n in the following pseudocode, assume that it is passed by reference, not by value.

[This issue does not come up in the pseudocode in the book, because it uses the attribute $A.heap\text{-}size$ instead of passing in the heap size as a parameter.]

HEAP-EXTRACT-MAX(A, n)
if $n < 1$
    error "heap underflow"
$max = A[1]$
$A[1] = A[n]$
$n = n - 1$
MAX-HEAPIFY$(A, 1, n)$ // remakes heap
return *max*

Analysis

Constant-time assignments plus time for MAX-HEAPIFY.

Time

$O(\lg n)$.

Example

Run HEAP-EXTRACT-MAX on first heap example.

- Take 16 out of node 1.
- Move 1 from node 10 to node 1.
- Erase node 10.
- MAX-HEAPIFY from the root to preserve max-heap property.
- Note that successive extractions will remove items in reverse sorted order.

Increasing key value

Given set S, element x, and new key value k:

- Make sure $k \geq x$'s current key.
- Update x's key value to k.
- Traverse the tree upward comparing x to its parent and swapping keys if necessary, until x's key is smaller than its parent's key.

HEAP-INCREASE-KEY(A, i, key)
if $key < A[i]$
    error "new key is smaller than current key"
$A[i] = key$
while $i > 1$ and $A[\text{PARENT}(i)] < A[i]$
    exchange $A[i]$ with $A[\text{PARENT}(i)]$
    $i = \text{PARENT}(i)$

Analysis

Upward path from node i has length $O(\lg n)$ in an n-element heap.

Time

$O(\lg n)$.

Example

Increase key of node 9 in first heap example to have value 15. Exchange keys of nodes 4 and 9, then of nodes 2 and 4.

Inserting into the heap

Given a key k to insert into the heap:

- Increment the heap size.
- Insert a new node in the last position in the heap, with key $-\infty$.
- Increase the $-\infty$ key to k using the HEAP-INCREASE-KEY procedure defined above.

Note: Again, the parameter n is passed by reference, not by value.

MAX-HEAP-INSERT(A, key, n)
  $n = n + 1$
  $A[n] = -\infty$
  HEAP-INCREASE-KEY(A, n, key)

Analysis

Constant time assignments + time for HEAP-INCREASE-KEY.

Time

$O(\lg n)$.

Min-priority queue operations are implemented similarly with min-heaps.

Solutions for Chapter 6: Heapsort

Solution to Exercise 6.1-1

This solution is also posted publicly

Since a heap is an almost-complete binary tree (complete at all levels except possibly the lowest), it has at most $2^{h+1} - 1$ elements (if it is complete) and at least $2^h - 1 + 1 = 2^h$ elements (if the lowest level has just 1 element and the other levels are complete).

Solution to Exercise 6.1-2

This solution is also posted publicly

Given an n-element heap of height h, we know from Exercise 6.1-1 that

$$2^h \le n \le 2^{h+1} - 1 < 2^{h+1} .$$

Thus, $h \le \lg n < h + 1$. Since h is an integer, $h = \lfloor \lg n \rfloor$ (by definition of $\lfloor\ \rfloor$).

Solution to Exercise 6.1-3

Assume the claim is false—i.e., that there is a subtree whose root is not the largest element in the subtree. Then the maximum element is somewhere else in the subtree, possibly even at more than one location. Let m be the index at which the maximum appears (the lowest such index if the maximum appears more than once). Since the maximum is not at the root of the subtree, node m has a parent. Since the parent of a node has a lower index than the node, and m was chosen to be the smallest index of the maximum value, $A[\text{PARENT}(m)] < A[m]$. But by the max-heap property, we must have $A[\text{PARENT}(m)] \ge A[m]$. So our assumption is false, and the claim is true.

Solution to Exercise 6.2-6

This solution is also posted publicly

If you put a value at the root that is less than every value in the left and right subtrees, then MAX-HEAPIFY will be called recursively until a leaf is reached. To make the recursive calls traverse the longest path to a leaf, choose values that make MAX-HEAPIFY always recurse on the left child. It follows the left branch when the left child is greater than or equal to the right child, so putting 0 at the root and 1 at all the other nodes, for example, will accomplish that. With such values, MAX-HEAPIFY will be called h times (where h is the heap height, which is the number of edges in the longest path from the root to a leaf), so its running time will be $\Theta(h)$ (since each call does $\Theta(1)$ work), which is $\Theta(\lg n)$. Since we have a case in which MAX-HEAPIFY's running time is $\Theta(\lg n)$, its worst-case running time is $\Omega(\lg n)$.

Solution to Exercise 6.3-3

Let H be the height of the heap.

Two subtleties to beware of:

- Be careful not to confuse the height of a node (longest distance from a leaf) with its depth (distance from the root).
- If the heap is not a complete binary tree (bottom level is not full), then the nodes at a given level (depth) don't all have the same height. For example, although all nodes at depth H have height 0, nodes at depth $H - 1$ can have either height 0 or height 1.

For a complete binary tree, it's easy to show that there are $\lceil n/2^{h+1} \rceil$ nodes of height h. But the proof for an incomplete tree is tricky and is not derived from the proof for a complete tree.

Proof By induction on h.

Basis: Show that it's true for $h = 0$ (i.e., that # of leaves $\le \lceil n/2^{h+1} \rceil = \lceil n/2 \rceil$). In fact, we'll show that the # of leaves $= \lceil n/2 \rceil$.

The tree leaves (nodes at height 0) are at depths H and $H - 1$. They consist of

- all nodes at depth H, and
- the nodes at depth $H - 1$ that are not parents of depth-H nodes.

Let x be the number of nodes at depth H—that is, the number of nodes in the bottom (possibly incomplete) level.

Note that $n - x$ is odd, because the $n - x$ nodes above the bottom level form a complete binary tree, and a complete binary tree has an odd number of nodes (1 less than a power of 2). Thus if n is odd, x is even, and if n is even, x is odd.

To prove the base case, we must consider separately the case in which n is even (x is odd) and the case in which n is odd (x is even). Here are two ways to do this: The first requires more cleverness, and the second requires more algebraic manipulation.

1. First method of proving the base case:

   - If n is odd, then x is even, so all nodes have siblings—i.e., all internal nodes have 2 children. Thus (see Exercise B.5-3), # of internal nodes $=$ # of leaves $- 1$.

     So, $n =$ # of nodes $=$ # of leaves $+$ # of internal nodes $= 2 \cdot$ # of leaves $- 1$.

     Thus, # of leaves $= (n+1)/2 = \lceil n/2 \rceil$. (The latter equality holds because n is odd.)
   - If n is even, then x is odd, and some leaf doesn't have a sibling. If we gave it a sibling, we would have $n + 1$ nodes, where $n + 1$ is odd, so the case we analyzed above would apply. Observe that we would also increase the number of leaves by 1, since we added a node to a parent that already had a child. By the odd-node case above, # of leaves $+ 1 = \lceil (n + 1)/2 \rceil = \lceil n/2 \rceil + 1$. (The latter equality holds because n is even.)

   In either case, # of leaves $= \lceil n/2 \rceil$.
2. Second method of proving the base case:

   Note that at any depth $d < H$ there are 2^d nodes, because all such tree levels are complete.

   - If x is even, there are $x/2$ nodes at depth $H - 1$ that are parents of depth H nodes, hence $2^{H-1} - x/2$ nodes at depth $H - 1$ that are not parents of depth-H nodes. Thus,

     $$\begin{aligned} \text{total \# of height-0 nodes} &= x + 2^{H-1} - x/2 \\ &= 2^{H-1} + x/2 \\ &= (2^H + x)/2 \\ &= \lceil (2^H + x - 1)/2 \rceil \qquad \text{(because } x \text{ is even)} \\ &= \lceil n/2 \rceil \ . \end{aligned}$$

     ($n = 2^H + x - 1$ because the complete tree down to depth $H - 1$ has $2^H - 1$ nodes and depth H has x nodes.)
   - If x is odd, by an argument similar to the even case, we see that

     $$\begin{aligned} \text{\# of height-0 nodes} &= x + 2^{H-1} - (x + 1)/2 \\ &= 2^{H-1} + (x - 1)/2 \\ &= (2^H + x - 1)/2 \\ &= n/2 \\ &= \lceil n/2 \rceil \qquad \text{(because } x \text{ odd} \Rightarrow n \text{ even)} \ . \end{aligned}$$

Inductive step: Show that if it's true for height $h - 1$, it's true for h.

Let n_h be the number of nodes at height h in the n-node tree T.

Consider the tree T' formed by removing the leaves of T. It has $n' = n - n_0$ nodes. We know from the base case that $n_0 = \lceil n/2 \rceil$, so $n' = n - n_0 = n - \lceil n/2 \rceil = \lfloor n/2 \rfloor$.

Note that the nodes at height h in T would be at height $h - 1$ if the leaves of the tree were removed—that is, they are at height $h - 1$ in T'. Letting n'_{h-1} denote the number of nodes at height $h - 1$ in T', we have

$$n_h = n'_{h-1} .$$

By induction, we can bound n'_{h-1}:

$$n_h = n'_{h-1} \le \lceil n'/2^h \rceil = \lceil \lfloor n/2 \rfloor / 2^h \rceil \le \lceil (n/2)/2^h \rceil = \lceil n/2^{h+1} \rceil \ . \qquad \blacksquare$$

Alternative solution

An alternative solution relies on four facts:

1. Every node *not* on the unique simple path from the last leaf to the root is the root of a complete binary subtree.
2. A node that is the root of a complete binary subtree and has height h is the ancestor of 2^h leaves.
3. By Exercise 6.1-7, an n-element heap has $\lceil n/2 \rceil$ leaves.
4. For nonnegative reals a and b, we have $\lceil a \rceil \cdot b \ge \lceil ab \rceil$.

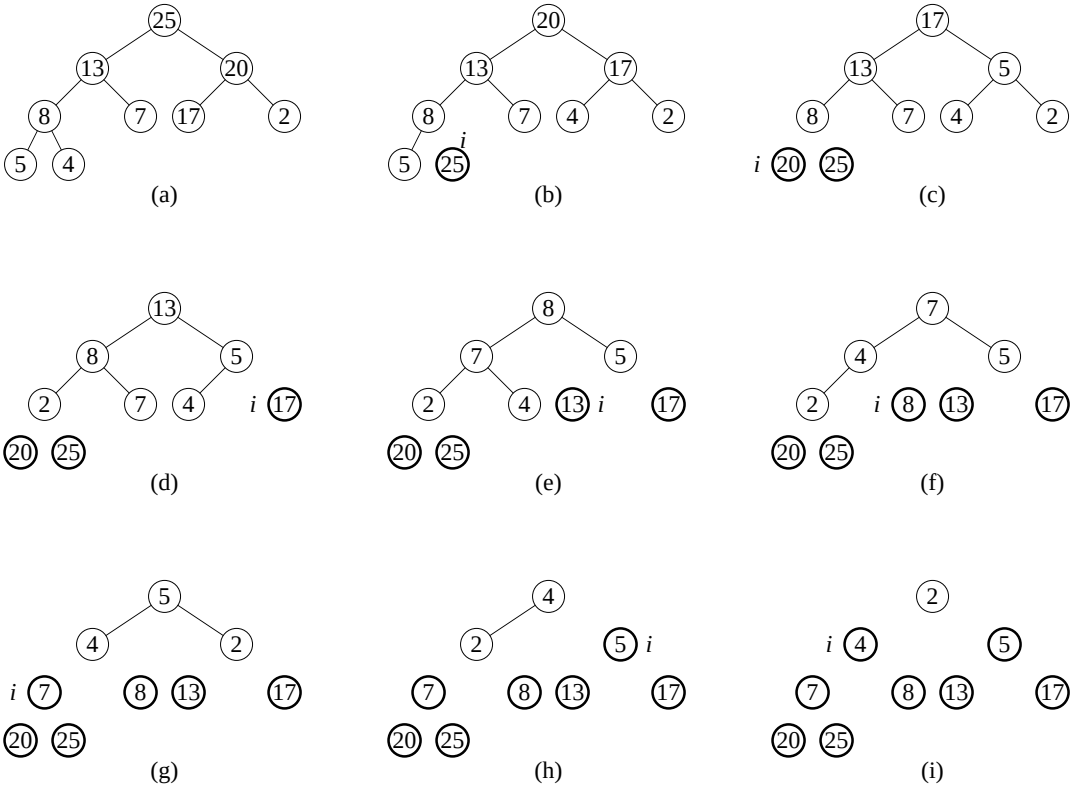
The proof is by contradiction. Assume that an n-element heap contains at least $\lceil n/2^{h+1} \rceil + 1$ nodes of height h. Exactly one node of height h is on the unique simple path from the last leaf to the root, and the subtree rooted at this node has at least one leaf (that being the last leaf). All other nodes of height h, of which the heap contains at least $\lceil n/2^{h+1} \rceil$, are the roots of complete binary subtrees, and each such node is the root of a subtree with 2^h leaves. Moreover, each subtree whose root is at height h is disjoint. Therefore, the number of leaves in the entire heap is at least

$$\begin{aligned} \left\lceil \frac{n}{2^{h+1}} \right\rceil \cdot 2^h + 1 &\ge \left\lceil \frac{n}{2^{h+1}} \cdot 2^h \right\rceil + 1 \\ &= \left\lceil \frac{n}{2} \right\rceil + 1 , \end{aligned}$$

which contradicts the property that an n-element heap has $\lceil n/2 \rceil$ leaves. $\blacksquare$

Solution to Exercise 6.4-1

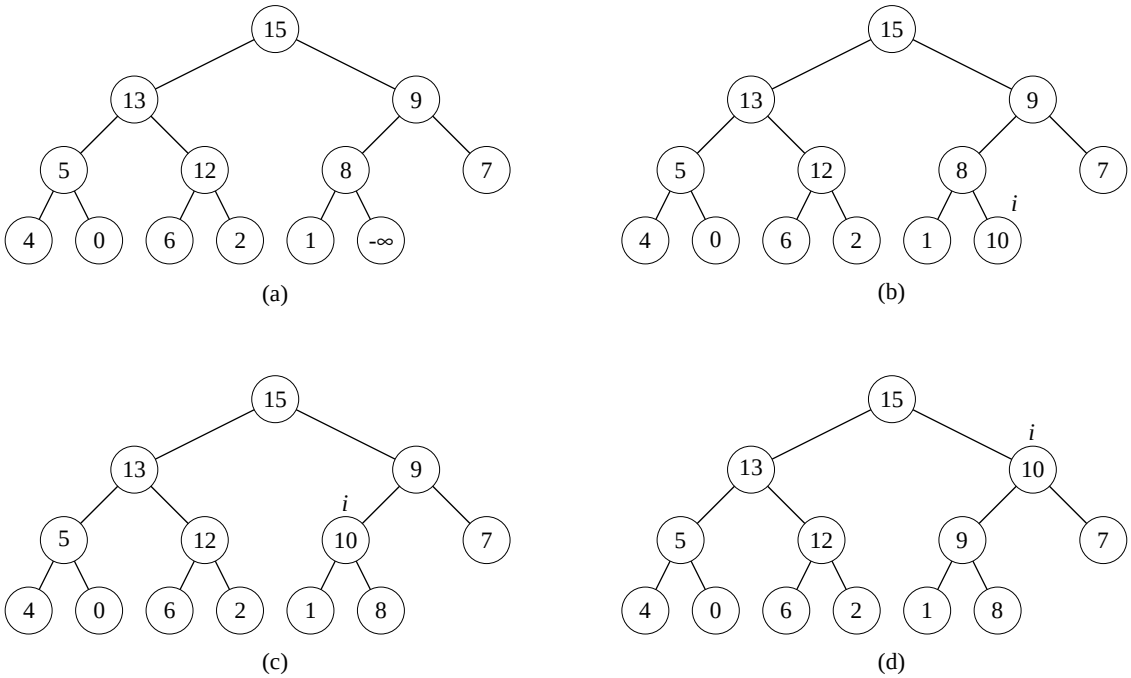
This solution is also posted publicly



A	2	4	5	7	8	13	17	20	25

Solution to Exercise 6.5-2

This solution is also posted publicly



Solution to Exercise 6.5-6

Change the procedure to the following:

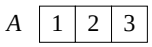
HEAP-INCREASE-KEY(A, i, key)
if $key < A[i]$
    error "new key is smaller than current key"
$A[i] = key$
while $i > 1$ and $A[\text{PARENT}(i)] < A[i]$
    $A[i] = A[\text{PARENT}(i)]$
    $i = \text{PARENT}(i)$
$A[i] = key$

Solution to Problem 6-1

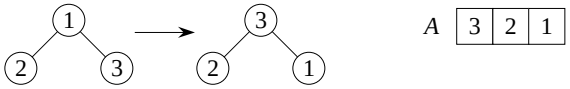
This solution is also posted publicly

a. The procedures BUILD-MAX-HEAP and BUILD-MAX-HEAP′ do not always create the same heap when run on the same input array. Consider the following counterexample.

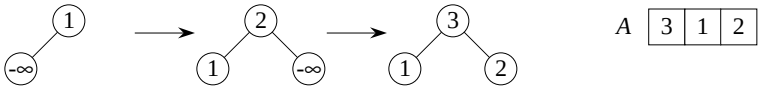
Input array A:



BUILD-MAX-HEAP(A):



BUILD-MAX-HEAP$'(A)$:



b. An upper bound of $O(n \lg n)$ time follows immediately from there being $n - 1$ calls to MAX-HEAP-INSERT, each taking $O(\lg n)$ time. For a lower bound of $\Omega(n \lg n)$, consider the case in which the input array is given in strictly increasing order. Each call to MAX-HEAP-INSERT causes HEAP-INCREASE-KEY to go all the way up to the root. Since the depth of node i is $\lfloor \lg i \rfloor$, the total time is

$$\begin{aligned}
\sum_{i=1}^{n} \Theta(\lfloor \lg i \rfloor) &\geq \sum_{i=\lceil n/2 \rceil}^{n} \Theta(\lfloor \lg \lceil n/2 \rceil \rfloor) \\
&\geq \sum_{i=\lceil n/2 \rceil}^{n} \Theta(\lfloor \lg(n/2) \rfloor) \\
&= \sum_{i=\lceil n/2 \rceil}^{n} \Theta(\lfloor \lg n - 1 \rfloor) \\
&\geq n/2 \cdot \Theta(\lg n) \\
&= \Omega(n \lg n) \; .
\end{aligned}$$

In the worst case, therefore, BUILD-MAX-HEAP$'$ requires $\Theta(n \lg n)$ time to build an n-element heap.

Solution to Problem 6-2

a. We can represent a d-ary heap in a 1-dimensional array as follows. The root resides in $A[1]$, its d children reside in order in $A[2]$ through $A[d + 1]$, their children reside in order in $A[d + 2]$ through $A[d^2 + d + 1]$, and so on. The following two procedures map a node with index i to its parent and to its jth child (for $1 \leq j \leq d$), respectively.

```
D-ARY-PARENT(i)
  return ⌊(i − 2)/d + 1⌋

D-ARY-CHILD(i, j)
  return d(i − 1) + j + 1
```

To convince yourself that these procedures really work, verify that

$$\text{D-ARY-PARENT}(\text{D-ARY-CHILD}(i, j)) = i \ ,$$

for any $1 \le j \le d$. Notice that the binary heap procedures are a special case of the above procedures when $d = 2$.

b. Since each node has d children, the height of a d-ary heap with n nodes is $\Theta(\log_d n) = \Theta(\lg n / \lg d)$.

c. The procedure HEAP-EXTRACT-MAX given in the text for binary heaps works fine for d-ary heaps too. The change needed to support d-ary heaps is in MAX-HEAPIFY, which must compare the argument node to all d children instead of just 2 children. The running time of HEAP-EXTRACT-MAX is still the running time for MAX-HEAPIFY, but that now takes worst-case time proportional to the product of the height of the heap by the number of children examined at each node (at most d), namely $\Theta(d \log_d n) = \Theta(d \lg n / \lg d)$.

d. The procedure MAX-HEAP-INSERT given in the text for binary heaps works fine for d-ary heaps too, assuming that HEAP-INCREASE-KEY works for d-ary heaps. The worst-case running time is still $\Theta(h)$, where h is the height of the heap. (Since only parent pointers are followed, the number of children a node has is irrelevant.) For a d-ary heap, this is $\Theta(\log_d n) = \Theta(\lg n / \lg d)$.

e. The HEAP-INCREASE-KEY procedure with two small changes works for d-ary heaps. First, because the problem specifies that the new key is given by the parameter k, change instances of the variable *key* to k. Second, change calls of PARENT to calls of D-ARY-PARENT from part (a).

In the worst case, the entire height of the tree must be traversed, so the worst-case running time is $\Theta(h) = \Theta(\log_d n) = \Theta(\lg n / \lg d)$.

Lecture Notes for Chapter 7: Quicksort

Chapter 7 overview

[The treatment in the second and third editions differs from that of the first edition. We use a different partitioning method—known as "Lomuto partitioning"—in the second and third editions, rather than the "Hoare partitioning" used in the first edition. Using Lomuto partitioning helps simplify the analysis, which uses indicator random variables in the second edition.]

Quicksort

- Worst-case running time: $\Theta(n^2)$.
- Expected running time: $\Theta(n \lg n)$.
- Constants hidden in $\Theta(n \lg n)$ are small.
- Sorts in place.

Description of quicksort

Quicksort is based on the three-step process of divide-and-conquer.

- To sort the subarray $A[p \,..\, r]$:

  Divide: Partition $A[p \,..\, r]$, into two (possibly empty) subarrays $A[p \,..\, q-1]$ and $A[q+1 \,..\, r]$, such that each element in the first subarray $A[p \,..\, q-1]$ is $\le A[q]$ and $A[q]$ is $\le$ each element in the second subarray $A[q+1 \,..\, r]$.

  Conquer: Sort the two subarrays by recursive calls to QUICKSORT.

  Combine: No work is needed to combine the subarrays, because they are sorted in place.

- Perform the divide step by a procedure PARTITION, which returns the index q that marks the position separating the subarrays.

```
QUICKSORT(A, p, r)
  if p < r
      q = PARTITION(A, p, r)
      QUICKSORT(A, p, q − 1)
      QUICKSORT(A, q + 1, r)
```

Initial call is QUICKSORT$(A, 1, n)$.

Partitioning

Partition subarray $A[p\,..\,r]$ by the following procedure:

```
PARTITION(A, p, r)
  x = A[r]
  i = p − 1
  for j = p to r − 1
      if A[j] ≤ x
          i = i + 1
          exchange A[i] with A[j]
  exchange A[i + 1] with A[r]
  return i + 1
```

- PARTITION always selects the last element $A[r]$ in the subarray $A[p\,..\,r]$ as the ***pivot***—the element around which to partition.
- As the procedure executes, the array is partitioned into four regions, some of which may be empty:

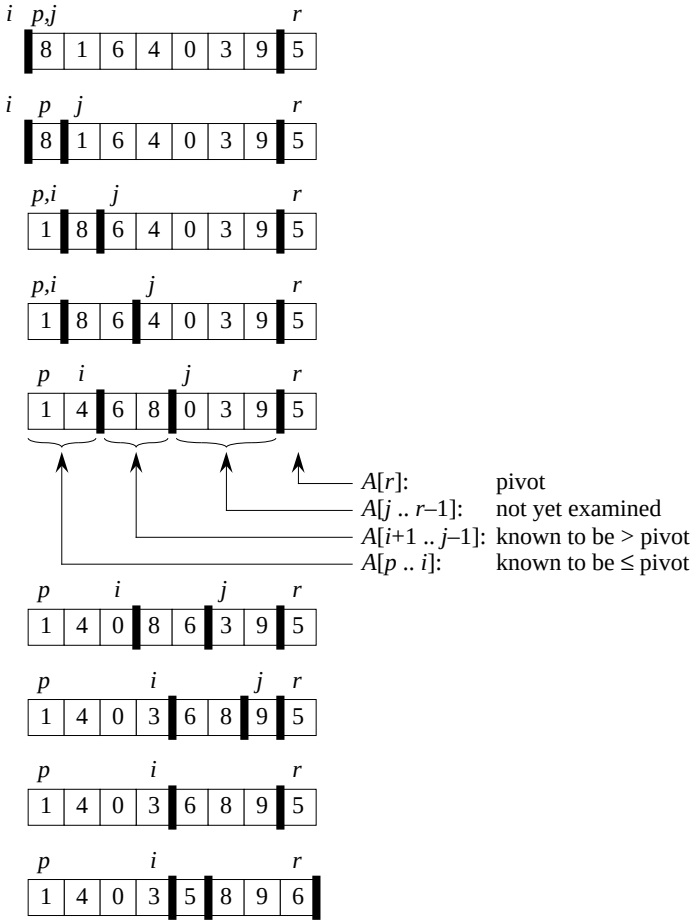
  Loop invariant:

  1. All entries in $A[p\,..\,i]$ are $\le$ pivot.
  2. All entries in $A[i + 1\,..\,j - 1]$ are $>$ pivot.
  3. $A[r] =$ pivot.

  It's not needed as part of the loop invariant, but the fourth region is $A[j\,..\,r-1]$, whose entries have not yet been examined, and so we don't know how they compare to the pivot.

Example

On an 8-element subarray.



[The index j disappears because it is no longer needed once the ***for*** *loop is exited.]*

Correctness

Use the loop invariant to prove correctness of PARTITION:

Initialization: Before the loop starts, all the conditions of the loop invariant are satisfied, because r is the pivot and the subarrays $A[p\,..\,i]$ and $A[i+1\,..\,j-1]$ are empty.

Maintenance: While the loop is running, if $A[j] \le$ pivot, then $A[j]$ and $A[i+1]$ are swapped and then i and j are incremented. If $A[j] >$ pivot, then increment only j.

Termination: When the loop terminates, $j = r$, so all elements in A are partitioned into one of the three cases: $A[p\,..\,i] \le$ pivot, $A[i+1\,..\,r-1] >$ pivot, and $A[r] =$ pivot.

The last two lines of PARTITION move the pivot element from the end of the array to between the two subarrays. This is done by swapping the pivot and the first element of the second subarray, i.e., by swapping $A[i+1]$ and $A[r]$.

Time for partitioning

$\Theta(n)$ to partition an n-element subarray.

Performance of quicksort

The running time of quicksort depends on the partitioning of the subarrays:

- If the subarrays are balanced, then quicksort can run as fast as mergesort.
- If they are unbalanced, then quicksort can run as slowly as insertion sort.

Worst case

- Occurs when the subarrays are completely unbalanced.
- Have 0 elements in one subarray and $n - 1$ elements in the other subarray.
- Get the recurrence
$$\begin{aligned} T(n) &= T(n-1) + T(0) + \Theta(n) \\ &= T(n-1) + \Theta(n) \\ &= \Theta(n^2) \ . \end{aligned}$$
- Same running time as insertion sort.
- In fact, the worst-case running time occurs when quicksort takes a sorted array as input, but insertion sort runs in $O(n)$ time in this case.

Best case

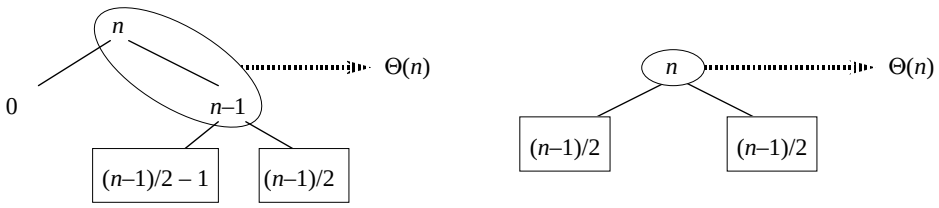
- Occurs when the subarrays are completely balanced every time.
- Each subarray has $\leq n/2$ elements.
- Get the recurrence
$$\begin{aligned} T(n) &= 2T(n/2) + \Theta(n) \\ &= \Theta(n \lg n) \ . \end{aligned}$$

Balanced partitioning

- Quicksort's average running time is much closer to the best case than to the worst case.
- Imagine that PARTITION always produces a 9-to-1 split.
- Get the recurrence
$$\begin{aligned} T(n) &\leq T(9n/10) + T(n/10) + \Theta(n) \\ &= O(n \lg n) \ . \end{aligned}$$
- Intuition: look at the recursion tree.
  - It's like the one for $T(n) = T(n/3) + T(2n/3) + O(n)$ in Section 4.4.
  - Except that here the constants are different; we get $\log_{10} n$ full levels and $\log_{10/9} n$ levels that are nonempty.
  - As long as it's a constant, the base of the log doesn't matter in asymptotic notation.
  - Any split of constant proportionality will yield a recursion tree of depth $\Theta(\lg n)$.

Intuition for the average case

- Splits in the recursion tree will not always be constant.
- There will usually be a mix of good and bad splits throughout the recursion tree.
- To see that this doesn't affect the asymptotic running time of quicksort, assume that levels alternate between best-case and worst-case splits.



- The extra level in the left-hand figure only adds to the constant hidden in the Θ-notation.
- There are still the same number of subarrays to sort, and only twice as much work was done to get to that point.
- Both figures result in $O(n \lg n)$ time, though the constant for the figure on the left is higher than that of the figure on the right.

Randomized version of quicksort

- We have assumed that all input permutations are equally likely.
- This is not always true.
- To correct this, we add randomization to quicksort.
- We could randomly permute the input array.
- Instead, we use ***random sampling***, or picking one element at random.
- Don't always use $A[r]$ as the pivot. Instead, randomly pick an element from the subarray that is being sorted.

RANDOMIZED-PARTITION(A, p, r)
  i = RANDOM(p, r)
  exchange $A[r]$ with $A[i]$
  return PARTITION(A, p, r)

Randomly selecting the pivot element will, on average, cause the split of the input array to be reasonably well balanced.

RANDOMIZED-QUICKSORT(A, p, r)
  if $p < r$
    q = RANDOMIZED-PARTITION(A, p, r)
    RANDOMIZED-QUICKSORT$(A, p, q-1)$
    RANDOMIZED-QUICKSORT$(A, q+1, r)$

Randomization of quicksort stops any specific type of array from causing worst-case behavior. For example, an already-sorted array causes worst-case behavior in non-randomized QUICKSORT, but not in RANDOMIZED-QUICKSORT.

Analysis of quicksort

We will analyze

- the worst-case running time of QUICKSORT and RANDOMIZED-QUICKSORT (the same), and
- the expected (average-case) running time of RANDOMIZED-QUICKSORT.

Worst-case analysis

We will prove that a worst-case split at every level produces a worst-case running time of $O(n^2)$.

- Recurrence for the worst-case running time of QUICKSORT:

  $$T(n) = \max_{0 \le q \le n-1} (T(q) + T(n-q-1)) + \Theta(n) .$$

- Because PARTITION produces two subproblems, totaling size $n-1$, q ranges from 0 to $n-1$.
- ***Guess:*** $T(n) \le cn^2$, for some c.
- Substituting our guess into the above recurrence:

  $$\begin{aligned} T(n) &\le \max_{0 \le q \le n-1} (cq^2 + c(n-q-1)^2) + \Theta(n) \\ &= c \cdot \max_{0 \le q \le n-1} (q^2 + (n-q-1)^2) + \Theta(n) . \end{aligned}$$

- The maximum value of $(q^2 + (n-q-1)^2)$ occurs when q is either 0 or $n-1$. (Second derivative with respect to q is positive.) Therefore,

  $$\begin{aligned} \max_{0 \le q \le n-1} (q^2 + (n-q-1)^2) &\le (n-1)^2 \\ &= n^2 - 2n + 1 . \end{aligned}$$

- And thus,

  $$\begin{aligned} T(n) &\le cn^2 - c(2n-1) + \Theta(n) \\ &\le cn^2 \qquad \text{if } c(2n-1) \ge \Theta(n) . \end{aligned}$$

- Pick c so that $c(2n-1)$ dominates $\Theta(n)$.
- Therefore, the worst-case running time of quicksort is $O(n^2)$.
- Can also show that the recurrence's solution is $\Omega(n^2)$. Thus, the worst-case running time is $\Theta(n^2)$.

Average-case analysis

- The dominant cost of the algorithm is partitioning.
- PARTITION removes the pivot element from future consideration each time.
- Thus, PARTITION is called at most n times.
- QUICKSORT recurses on the partitions.
- The amount of work that each call to PARTITION does is a constant plus the number of comparisons that are performed in its **for** loop.
- Let $X =$ the total number of comparisons performed in all calls to PARTITION.
- Therefore, the total work done over the entire execution is $O(n + X)$.

We will now compute a bound on the overall number of comparisons.

For ease of analysis:

- Rename the elements of A as $z_1, z_2, \ldots, z_n$, with z_i being the ith smallest element.
- Define the set $Z_{ij} = \{z_i, z_{i+1}, \ldots, z_j\}$ to be the set of elements between z_i and z_j, inclusive.

Each pair of elements is compared at most once, because elements are compared only to the pivot element, and then the pivot element is never in any later call to PARTITION.

Let $X_{ij} = \mathrm{I}\{z_i \text{ is compared to } z_j\}$.
(Considering whether z_i is compared to z_j at any time during the entire quicksort algorithm, not just during one call of PARTITION.)

Since each pair is compared at most once, the total number of comparisons performed by the algorithm is

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \ .$$

Take expectations of both sides, use Lemma 5.1 and linearity of expectation:

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right] \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{E}[X_{ij}] \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr\{z_i \text{ is compared to } z_j\} \ . \end{aligned}$$

Now all we have to do is find the probability that two elements are compared.

- Think about when two elements are *not* compared.
  - For example, numbers in separate partitions will not be compared.
  - In the previous example, $\langle 8, 1, 6, 4, 0, 3, 9, 5\rangle$ and the pivot is 5, so that none of the set $\{1, 4, 0, 3\}$ will ever be compared to any of the set $\{8, 6, 9\}$.

- Once a pivot x is chosen such that $z_i < x < z_j$, then z_i and z_j will never be compared at any later time.
- If either z_i or z_j is chosen before any other element of Z_{ij}, then it will be compared to all the elements of Z_{ij}, except itself.
- The probability that z_i is compared to z_j is the probability that either z_i or z_j is the first element chosen.
- There are $j-i+1$ elements, and pivots are chosen randomly and independently. Thus, the probability that any particular one of them is the first one chosen is $1/(j-i+1)$.

Therefore,

$$\begin{aligned}
\Pr\{z_i \text{ is compared to } z_j\} &= \Pr\{z_i \text{ or } z_j \text{ is the first pivot chosen from } Z_{ij}\} \\
&= \Pr\{z_i \text{ is the first pivot chosen from } Z_{ij}\} \\
&\quad + \Pr\{z_j \text{ is the first pivot chosen from } Z_{ij}\} \\
&= \frac{1}{j-i+1} + \frac{1}{j-i+1} \\
&= \frac{2}{j-i+1} .
\end{aligned}$$

[The second line follows because the two events are mutually exclusive.]

Substituting into the equation for $\mathrm{E}[X]$:

$$\mathrm{E}[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} .$$

Evaluate by using a change in variables ($k = j-i$) and the bound on the harmonic series in equation (A.7):

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \frac{2}{k+1} \\
&< \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \\
&= \sum_{i=1}^{n-1} O(\lg n) \\
&= O(n \lg n) .
\end{aligned}$$

So the expected running time of quicksort, using RANDOMIZED-PARTITION, is $O(n \lg n)$.

Solutions for Chapter 7: Quicksort

Solution to Exercise 7.2-3

This solution is also posted publicly

PARTITION does a "worst-case partitioning" when the elements are in decreasing order. It reduces the size of the subarray under consideration by only 1 at each step, which we've seen has running time $\Theta(n^2)$.

In particular, PARTITION, given a subarray $A[p\,..\,r]$ of distinct elements in decreasing order, produces an empty partition in $A[p\,..\,q-1]$, puts the pivot (originally in $A[r]$) into $A[p]$, and produces a partition $A[p+1\,..\,r]$ with only one fewer element than $A[p\,..\,r]$. The recurrence for QUICKSORT becomes $T(n) = T(n-1) + \Theta(n)$, which has the solution $T(n) = \Theta(n^2)$.

Solution to Exercise 7.2-5

This solution is also posted publicly

The minimum depth follows a path that always takes the smaller part of the partition—i.e., that multiplies the number of elements by α. One iteration reduces the number of elements from n to αn, and i iterations reduces the number of elements to $\alpha^i n$. At a leaf, there is just one remaining element, and so at a minimum-depth leaf of depth m, we have $\alpha^m n = 1$. Thus, $\alpha^m = 1/n$. Taking logs, we get $m \lg \alpha = -\lg n$, or $m = -\lg n / \lg \alpha$.

Similarly, maximum depth corresponds to always taking the larger part of the partition, i.e., keeping a fraction $1-\alpha$ of the elements each time. The maximum depth M is reached when there is one element left, that is, when $(1-\alpha)^M n = 1$. Thus, $M = -\lg n / \lg(1-\alpha)$.

All these equations are approximate because we are ignoring floors and ceilings.

Solution to Exercise 7.3-1

We may be interested in the worst-case performance, but in that case, the randomization is irrelevant: it won't improve the worst case. What randomization can do is make the chance of encountering a worst-case scenario small.

Solution to Exercise 7.4-2

To show that quicksort's best-case running time is $\Omega(n \lg n)$, we use a technique similar to the one used in Section 7.4.1 to show that its worst-case running time is $O(n^2)$.

Let $T(n)$ be the best-case time for the procedure QUICKSORT on an input of size n. We have the recurrence

$$T(n) = \min_{1 \le q \le n-1} (T(q) + T(n-q-1)) + \Theta(n) \; .$$

We guess that $T(n) \ge cn \lg n$ for some constant c. Substituting this guess into the recurrence, we obtain

$$\begin{aligned} T(n) &\ge \min_{1 \le q \le n-1} (cq \lg q + c(n-q-1) \lg(n-q-1)) + \Theta(n) \\ &= c \cdot \min_{1 \le q \le n-1} (q \lg q + (n-q-1) \lg(n-q-1)) + \Theta(n) \; . \end{aligned}$$

As we'll show below, the expression $q \lg q + (n-q-1) \lg(n-q-1)$ achieves a minimum over the range $1 \le q \le n-1$ when $q = n-q-1$, or $q = (n-1)/2$, since the first derivative of the expression with respect to q is 0 when $q = (n-1)/2$ and the second derivative of the expression is positive. (It doesn't matter that q is not an integer when n is even, since we're just trying to determine the minimum value of a function, knowing that when we constrain q to integer values, the function's value will be no lower.)

Choosing $q = (n-1)/2$ gives us the bound

$$\begin{aligned} &\min_{1 \le q \le n-1} (q \lg q + (n-q-1) \lg(n-q-1) \\ &\qquad \ge \frac{n-1}{2} \lg \frac{n-1}{2} + \left(n - \frac{n-1}{2} - 1 \right) \lg \left(n - \frac{n-1}{2} - 1 \right) \\ &\qquad = (n-1) \lg \frac{n-1}{2} \; . \end{aligned}$$

Continuing with our bounding of $T(n)$, we obtain, for $n \ge 2$,

$$\begin{aligned} T(n) &\ge c(n-1) \lg \frac{n-1}{2} + \Theta(n) \\ &= c(n-1) \lg(n-1) - c(n-1) + \Theta(n) \\ &= cn \lg(n-1) - c \lg(n-1) - c(n-1) + \Theta(n) \\ &\ge cn \lg(n/2) - c \lg(n-1) - c(n-1) + \Theta(n) \qquad \text{(since } n \ge 2\text{)} \\ &= cn \lg n - cn - c \lg(n-1) - cn + c + \Theta(n) \\ &= cn \lg n - (2cn + c \lg(n-1) - c) + \Theta(n) \\ &\ge cn \lg n \; , \end{aligned}$$

since we can pick the constant c small enough so that the $\Theta(n)$ term dominates the quantity $2cn + c\lg(n-1) - c$. Thus, the best-case running time of quicksort is $\Omega(n \lg n)$.

Letting $f(q) = q\lg q + (n-q-1)\lg(n-q-1)$, we now show how to find the minimum value of this function in the range $1 \le q \le n-1$. We need to find the value of q for which the derivative of f with respect to q is 0. We rewrite this function as

$$f(q) = \frac{q\ln q + (n-q-1)\ln(n-q-1)}{\ln 2} ,$$

and so

$$\begin{aligned} f'(q) &= \frac{d}{dq}\left(\frac{q\ln q + (n-q-1)\ln(n-q-1)}{\ln 2}\right) \\ &= \frac{\ln q + 1 - \ln(n-q-1) - 1}{\ln 2} \\ &= \frac{\ln q - \ln(n-q-1)}{\ln 2} . \end{aligned}$$

The derivative $f'(q)$ is 0 when $q = n-q-1$, or when $q = (n-1)/2$. To verify that $q = (n-1)/2$ is indeed a minimum (not a maximum or an inflection point), we need to check that the second derivative of f is positive at $q = (n-1)/2$:

$$\begin{aligned} f''(q) &= \frac{d}{dq}\left(\frac{\ln q - \ln(n-q-1)}{\ln 2}\right) \\ &= \frac{1}{\ln 2}\left(\frac{1}{q} + \frac{1}{n-q-1}\right) \\ f''\left(\frac{n-1}{2}\right) &= \frac{1}{\ln 2}\left(\frac{2}{n-1} + \frac{2}{n-1}\right) \\ &= \frac{1}{\ln 2}\cdot\frac{4}{n-1} \\ &> 0 \qquad (\text{since } n \ge 2) . \end{aligned}$$

Solution to Problem 7-2

a. If all elements are equal, then when PARTITION returns, $q = r$ and all elements in $A[p\,..\,q-1]$ are equal. We get the recurrence $T(n) = T(n-1)+T(0)+\Theta(n)$ for the running time, and so $T(n) = \Theta(n^2)$.

b. The PARTITION′ procedure:

```
PARTITION′(A, p, r)
  x = A[p]
  i = h = p
  for j = p + 1 to r
      // Invariant: A[p..i − 1] < x, A[i..h] = x,
            A[h + 1..j − 1] > x, A[j..r] unknown.
      if A[j] < x
            y = A[j]
            A[j] = A[h + 1]
            A[h + 1] = A[i]
            A[i] = y
            i = i + 1
            h = h + 1
      elseif A[j] == x
            exchange A[h + 1] with A[j]
            h = h + 1
  return (i, h)
```

c. RANDOMIZED-PARTITION′ is the same as RANDOMIZED-PARTITION, but with the call to PARTITION replaced by a call to PARTITION′.

```
QUICKSORT′(A, p, r)
  if p < r
      (q, t) = RANDOMIZED-PARTITION′(A, p, r)
      QUICKSORT′(A, p, q − 1)
      QUICKSORT′(A, t + 1, r)
```

d. Putting elements equal to the pivot in the same partition as the pivot can only help us, because we do not recurse on elements equal to the pivot. Thus, the subproblem sizes with QUICKSORT′, even with equal elements, are no larger than the subproblem sizes with QUICKSORT when all elements are distinct.

Solution to Problem 7-4

a. QUICKSORT′ does exactly what QUICKSORT does; hence it sorts correctly.

QUICKSORT and QUICKSORT′ do the same partitioning, and then each calls itself with arguments $A, p, q-1$. QUICKSORT then calls itself again, with arguments $A, q+1, r$. QUICKSORT′ instead sets $p = q+1$ and performs another iteration of its **while** loop. This executes the same operations as calling itself with $A, q+1, r$, because in both cases, the first and third arguments (A and r) have the same values as before, and p has the old value of $q+1$.

b. The stack depth of QUICKSORT′ will be $\Theta(n)$ on an n-element input array if there are $\Theta(n)$ recursive calls to QUICKSORT′. This happens if every call to PARTITION(A, p, r) returns $q = r$. The sequence of recursive calls in this scenario is

$\text{QUICKSORT}'(A, 1, n)$,
$\text{QUICKSORT}'(A, 1, n-1)$,
$\text{QUICKSORT}'(A, 1, n-2)$,
$\vdots$
$\text{QUICKSORT}'(A, 1, 1)$.

Any array that is already sorted in increasing order will cause QUICKSORT′ to behave this way.

c. The problem demonstrated by the scenario in part (b) is that each invocation of QUICKSORT′ calls QUICKSORT′ again with almost the same range. To avoid such behavior, we must change QUICKSORT′ so that the recursive call is on a smaller interval of the array. The following variation of QUICKSORT′ checks which of the two subarrays returned from PARTITION is smaller and recurses on the smaller subarray, which is at most half the size of the current array. Since the array size is reduced by at least half on each recursive call, the number of recursive calls, and hence the stack depth, is $\Theta(\lg n)$ in the worst case. Note that this method works no matter how partitioning is performed (as long as the PARTITION procedure has the same functionality as the procedure given in Section 7.1).

```
QUICKSORT″(A, p, r)
  while p < r
      // Partition and sort the small subarray first.
      q = PARTITION(A, p, r)
      if q − p < r − q
          QUICKSORT″(A, p, q − 1)
          p = q + 1
      else QUICKSORT″(A, q + 1, r)
          r = q − 1
```

The expected running time is not affected, because exactly the same work is done as before: the same partitions are produced, and the same subarrays are sorted.

Lecture Notes for Chapter 8: Sorting in Linear Time

Chapter 8 overview

How fast can we sort?

We will prove a lower bound, then beat it by playing a different game.

Comparison sorting

- The only operation that may be used to gain order information about a sequence is comparison of pairs of elements.
- All sorts seen so far are comparison sorts: insertion sort, selection sort, merge sort, quicksort, heapsort, treesort.

Lower bounds for sorting

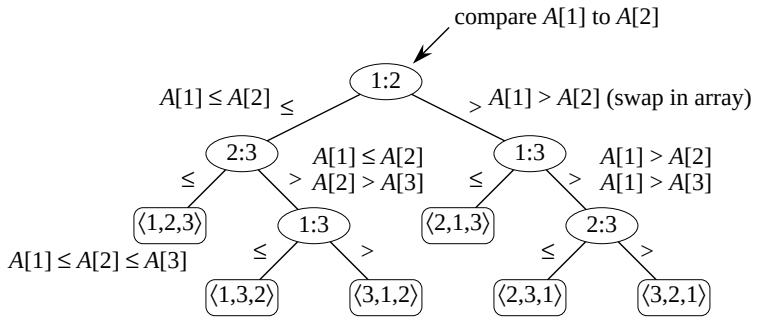
Lower bounds

- $\Omega(n)$ to examine all the input.
- All sorts seen so far are $\Omega(n \lg n)$.
- We'll show that $\Omega(n \lg n)$ is a lower bound for comparison sorts.

Decision tree

- Abstraction of any comparison sort.
- Represents comparisons made by
  - a specific sorting algorithm
  - on inputs of a given size.
- Abstracts away everything else: control and data movement.
- We're counting *only* comparisons.

For insertion sort on 3 elements:



[Each internal node is labeled by indices of array elements ***from their original positions****. Each leaf is labeled by the permutation of orders that the algorithm determines.]*

How many leaves on the decision tree? There are $\geq n!$ leaves, because every permutation appears at least once.

For any comparison sort,

- 1 tree for each n.
- View the tree as if the algorithm splits in two at each node, based on the information it has determined up to that point.
- The tree models all possible execution traces.

What is the length of the longest path from root to leaf?

- Depends on the algorithm
- Insertion sort: $\Theta(n^2)$
- Merge sort: $\Theta(n \lg n)$

Lemma

Any binary tree of height h has $\leq 2^h$ leaves.

In other words:

- $l =$ # of leaves,
- $h =$ height,
- Then $l \leq 2^h$.

(We'll prove this lemma later.)

Why is this useful?

Theorem

Any decision tree that sorts n elements has height $\Omega(n \lg n)$.

Proof

- $l \geq n!$
- By lemma, $n! \leq l \leq 2^h$ or $2^h \geq n!$
- Take logs: $h \geq \lg(n!)$
- Use Stirling's approximation: $n! > (n/e)^n$ (by equation (3.17))

$$\begin{aligned} h &\geq \lg(n/e)^n \\ &= n\lg(n/e) \\ &= n\lg n - n\lg e \\ &= \Omega(n\lg n)\ . \end{aligned}$$

■ (theorem)

Now to prove the lemma:

Proof By induction on h.

Basis: $h = 0$. Tree is just one node, which is a leaf. $2^h = 1$.

Inductive step: Assume true for height $= h - 1$. Extend tree of height $h - 1$ by making as many new leaves as possible. Each leaf becomes parent to two new leaves.

$$\begin{aligned} \#\text{ of leaves for height } h &= 2 \cdot (\#\text{ of leaves for height } h-1) \\ &= 2 \cdot 2^{h-1} && \text{(ind. hypothesis)} \\ &= 2^h\ . \end{aligned}$$

■ (lemma)

Corollary

Heapsort and merge sort are asymptotically optimal comparison sorts.

Sorting in linear time

Non-comparison sorts.

Counting sort

Depends on a *key assumption*: numbers to be sorted are integers in $\{0, 1, \ldots, k\}$.

Input: $A[1..n]$, where $A[j] \in \{0, 1, \ldots, k\}$ for $j = 1, 2, \ldots, n$. Array A and values n and k are given as parameters.

Output: $B[1..n]$, sorted. B is assumed to be already allocated and is given as a parameter.

Auxiliary storage: $C[0..k]$

```
COUNTING-SORT(A, B, n, k)
  let C[0..k] be a new array
  for i = 0 to k
      C[i] = 0
  for j = 1 to n
      C[A[j]] = C[A[j]] + 1
  for i = 1 to k
      C[i] = C[i] + C[i − 1]
  for j = n downto 1
      B[C[A[j]]] = A[j]
      C[A[j]] = C[A[j]] − 1
```

Do an example for $A = 2_1, 5_1, 3_1, 0_1, 2_2, 3_2, 0_2, 3_3$

Counting sort is ***stable*** (keys with same value appear in same order in output as they did in input) because of how the last loop works.

Analysis

$\Theta(n + k)$, which is $\Theta(n)$ if $k = O(n)$.

How big a k is practical?

- Good for sorting 32-bit values? No.
- 16-bit? Probably not.
- 8-bit? Maybe, depending on n.
- 4-bit? Probably (unless n is really small).

Counting sort will be used in radix sort.

Radix sort

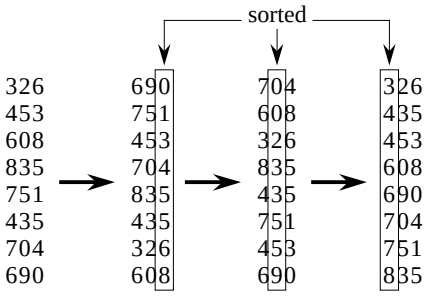
How IBM made its money. Punch card readers for census tabulation in early 1900's. Card sorters, worked on one column at a time. It's the algorithm for using the machine that extends the technique to multi-column sorting. The human operator was part of the algorithm!

Key idea: Sort *least* significant digits first.

To sort d digits:

```
RADIX-SORT(A, d)
  for i = 1 to d
      use a stable sort to sort array A on digit i
```

Example



Correctness

- Induction on number of passes (i in pseudocode).
- Assume digits $1, 2, \ldots, i-1$ are sorted.
- Show that a stable sort on digit i leaves digits $1, \ldots, i$ sorted:
  - If 2 digits in position i are different, ordering by position i is correct, and positions $1, \ldots, i-1$ are irrelevant.
  - If 2 digits in position i are equal, numbers are already in the right order (by inductive hypothesis). The stable sort on digit i leaves them in the right order.

This argument shows why it's so important to use a stable sort for intermediate sort.

Analysis

Assume that we use counting sort as the intermediate sort.

- $\Theta(n+k)$ per pass (digits in range $0, \ldots, k$)
- d passes
- $\Theta(d(n+k))$ total
- If $k = O(n)$, time $= \Theta(dn)$.

How to break each key into digits?

- n words.
- b bits/word.
- Break into r-bit digits. Have $d = \lceil b/r \rceil$.
- Use counting sort, $k = 2^r - 1$.

  Example: 32-bit words, 8-bit digits. $b = 32$, $r = 8$, $d = \lceil 32/8 \rceil = 4$, $k = 2^8 - 1 = 255$.
- Time $= \Theta\left(\frac{b}{r}(n + 2^r)\right)$.

How to choose r? Balance b/r and $n + 2^r$. Choosing $r \approx \lg n$ gives us $\Theta\left(\frac{b}{\lg n}(n+n)\right) = \Theta(bn/\lg n)$.

- If we choose $r < \lg n$, then $b/r > b/\lg n$, and $n + 2^r$ term doesn't improve.
- If we choose $r > \lg n$, then $n + 2^r$ term gets big. Example: $r = 2\lg n \Rightarrow 2^r = 2^{2\lg n} = (2^{\lg n})^2 = n^2$.

So, to sort 2^{16} 32-bit numbers, use $r = \lg 2^{16} = 16$ bits. $\lceil b/r \rceil = 2$ passes.

Compare radix sort to merge sort and quicksort:

- 1 million (2^{20}) 32-bit integers.
- Radix sort: $\lceil 32/20 \rceil = 2$ passes.
- Merge sort/quicksort: $\lg n = 20$ passes.
- Remember, though, that each radix sort "pass" is really 2 passes—one to take census, and one to move data.

How does radix sort violate the ground rules for a comparison sort?

- Using counting sort allows us to gain information about keys by means other than directly comparing 2 keys.
- Used keys as array indices.

Bucket sort

Assumes the input is generated by a random process that distributes elements uniformly over $[0, 1)$.

Idea

- Divide $[0, 1)$ into n equal-sized *buckets*.
- Distribute the n input values into the buckets.
- Sort each bucket.
- Then go through buckets in order, listing elements in each one.

Input: $A[1\,..\,n]$, where $0 \le A[i] < 1$ for all i.

Auxiliary array: $B[0\,..\,n-1]$ of linked lists, each list initially empty.

BUCKET-SORT(A, n)

```
let B[0..n − 1] be a new array
for i = 1 to n − 1
    make B[i] an empty list
for i = 1 to n
    insert A[i] into list B[⌊n · A[i]⌋]
for i = 0 to n − 1
    sort list B[i] with insertion sort
concatenate lists B[0], B[1], . . . , B[n − 1] together in order
return the concatenated lists
```

Correctness

Consider $A[i]$, $A[j]$. Assume without loss of generality that $A[i] \le A[j]$. Then $\lfloor n \cdot A[i] \rfloor \le \lfloor n \cdot A[j] \rfloor$. So $A[i]$ is placed into the same bucket as $A[j]$ or into a bucket with a lower index.

- If same bucket, insertion sort fixes up.
- If earlier bucket, concatenation of lists fixes up.

Analysis

- Relies on no bucket getting too many values.
- All lines of algorithm except insertion sorting take $\Theta(n)$ altogether.
- Intuitively, if each bucket gets a constant number of elements, it takes $O(1)$ time to sort each bucket $\Rightarrow$ $O(n)$ sort time for all buckets.
- We "expect" each bucket to have few elements, since the average is 1 element per bucket.
- But we need to do a careful analysis.

Define a random variable:

- n_i = the number of elements placed in bucket $B[i]$.

Because insertion sort runs in quadratic time, bucket sort time is

$$T(n) = \Theta(n) + \sum_{i=0}^{n-1} O(n_i^2) \ .$$

Take expectations of both sides:

$$\begin{aligned} \mathrm{E}[T(n)] &= \mathrm{E}\left[\Theta(n) + \sum_{i=0}^{n-1} O(n_i^2)\right] \\ &= \Theta(n) + \sum_{i=0}^{n-1} \mathrm{E}\left[O(n_i^2)\right] \qquad \text{(linearity of expectation)} \\ &= \Theta(n) + \sum_{i=0}^{n-1} O(\mathrm{E}\left[n_i^2\right]) \qquad (\mathrm{E}[aX] = a\mathrm{E}[X]) \end{aligned}$$

Claim

$\mathrm{E}[n_i^2] = 2 - (1/n)$ for $i = 0, \ldots, n-1$.

Proof of claim

Define indicator random variables:

- $X_{ij} = \mathrm{I}\{A[j] \text{ falls in bucket } i\}$
- $\Pr\{A[j] \text{ falls in bucket } i\} = 1/n$
- $n_i = \sum_{j=1}^{n} X_{ij}$

Then

$$\begin{aligned}
\mathrm{E}\left[n_i^2\right] &= \mathrm{E}\left[\left(\sum_{j=1}^{n} X_{ij}\right)^2\right] \\
&= \mathrm{E}\left[\sum_{j=1}^{n} X_{ij}^2 + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} X_{ij}X_{ik}\right] \\
&= \sum_{j=1}^{n} \mathrm{E}\left[X_{ij}^2\right] + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \mathrm{E}\left[X_{ij}X_{ik}\right] \quad \text{(linearity of expectation)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[X_{ij}^2\right] &= 0^2 \cdot \Pr\{A[j] \text{ doesn't fall in bucket } i\} + 1^2 \cdot \Pr\{A[j] \text{ falls in bucket } i\} \\
&= 0 \cdot \left(1 - \frac{1}{n}\right) + 1 \cdot \frac{1}{n} \\
&= \frac{1}{n}
\end{aligned}$$

$\mathrm{E}[X_{ij}X_{ik}]$ for $j \neq k$: Since $j \neq k$, X_{ij} and X_{ik} are independent random variables

$$\begin{aligned}
\Rightarrow \mathrm{E}[X_{ij}X_{ik}] &= \mathrm{E}[X_{ij}]\,\mathrm{E}[X_{ik}] \\
&= \frac{1}{n} \cdot \frac{1}{n} \\
&= \frac{1}{n^2}
\end{aligned}$$

Therefore:

$$\begin{aligned}
\mathrm{E}\left[n_i^2\right] &= \sum_{j=1}^{n} \frac{1}{n} + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \frac{1}{n^2} \\
&= n \cdot \frac{1}{n} + 2\binom{n}{2}\frac{1}{n^2} \\
&= 1 + 2 \cdot \frac{n(n-1)}{2} \cdot \frac{1}{n^2} \\
&= 1 + \frac{n-1}{n} \\
&= 1 + 1 - \frac{1}{n} \\
&= 2 - \frac{1}{n} \qquad \blacksquare \text{ (claim)}
\end{aligned}$$

Therefore:

$$\begin{aligned}
\mathrm{E}[T(n)] &= \Theta(n) + \sum_{i=0}^{n-1} O(2 - 1/n) \\
&= \Theta(n) + O(n) \\
&= \Theta(n)
\end{aligned}$$

- Again, not a comparison sort. Used a function of key values to index into an array.

- This is a ***probabilistic analysis***—we used probability to analyze an algorithm whose running time depends on the distribution of inputs.
- Different from a ***randomized algorithm***, where we use randomization to *impose* a distribution.
- With bucket sort, if the input isn't drawn from a uniform distribution on $[0, 1)$, all bets are off (performance-wise, but the algorithm is still correct).

Solutions for Chapter 8: Sorting in Linear Time

Solution to Exercise 8.1-3

This solution is also posted publicly

If the sort runs in linear time for m input permutations, then the height h of the portion of the decision tree consisting of the m corresponding leaves and their ancestors is linear.

Use the same argument as in the proof of Theorem 8.1 to show that this is impossible for $m = n!/2$, $n!/n$, or $n!/2^n$.

We have $2^h \geq m$, which gives us $h \geq \lg m$. For all the possible m's given here, $\lg m = \Omega(n \lg n)$, hence $h = \Omega(n \lg n)$.

In particular,

$$\begin{aligned}
\lg \frac{n!}{2} &= \lg n! - 1 \geq n \lg n - n \lg e - 1 \ , \\
\lg \frac{n!}{n} &= \lg n! - \lg n \geq n \lg n - n \lg e - \lg n \ , \\
\lg \frac{n!}{2^n} &= \lg n! - n \geq n \lg n - n \lg e - n \ .
\end{aligned}$$

Solution to Exercise 8.1-4

Let S be a sequence of n elements divided into n/k subsequences each of length k where all of the elements in any subsequence are larger than all of the elements of a preceding subsequence and smaller than all of the elements of a succeeding subsequence.

Claim

Any comparison-based sorting algorithm to sort s must take $\Omega(n \lg k)$ time in the worst case.

Proof First notice that, as pointed out in the hint, we cannot prove the lower bound by multiplying together the lower bounds for sorting each subsequence. That would only prove that there is no faster algorithm *that sorts the subsequences*

independently. This was not what we are asked to prove; we cannot introduce *any* extra assumptions.

Now, consider the decision tree of height h for any comparison sort for S. Since the elements of each subsequence can be in any order, any of the $k!$ permutations correspond to the final sorted order of a subsequence. And, since there are n/k such subsequences, each of which can be in any order, there are $(k!)^{n/k}$ permutations of S that could correspond to the sorting of some input order. Thus, any decision tree for sorting S must have at least $(k!)^{n/k}$ leaves. Since a binary tree of height h has no more than 2^h leaves, we must have $2^h \geq (k!)^{n/k}$ or $h \geq \lg((k!)^{n/k})$. We therefore obtain

$$\begin{aligned} h &\geq \lg((k!)^{n/k} \\ &= (n/k)\lg(k!) \\ &\geq (n/k)\lg((k/2)^{k/2}) \\ &= (n/2)\lg(k/2)\ . \end{aligned}$$

The third line comes from $k!$ having its $k/2$ largest terms being at least $k/2$ each. (We implicitly assume here that k is even. We could adjust with floors and ceilings if k were odd.)

Since there exists at least one path in any decision tree for sorting S that has length at least $(n/2)\lg(k/2)$, the worst-case running time of any comparison-based sorting algorithm for S is $\Omega(n \lg k)$. ■

Solution to Exercise 8.2-3

This solution is also posted publicly

[The following solution also answers Exercise 8.2-2.]

Notice that the correctness argument in the text does not depend on the order in which A is processed. The algorithm is correct no matter what order is used!

But the modified algorithm is not stable. As before, in the final **for** loop an element equal to one taken from A earlier is placed before the earlier one (i.e., at a lower index position) in the output arrray B. The original algorithm was stable because an element taken from A later started out with a lower index than one taken earlier. But in the modified algorithm, an element taken from A later started out with a higher index than one taken earlier.

In particular, the algorithm still places the elements with value k in positions $C[k-1]+1$ through $C[k]$, but in the reverse order of their appearance in A.

Solution to Exercise 8.2-4

Compute the C array as is done in counting sort. The number of integers in the range $[a\,..\,b]$ is $C[b] - C[a-1]$, where we interpret $C[-1]$ as 0.

Solution to Exercise 8.3-2

Insertion sort is stable. When inserting $A[j]$ into the sorted sequence $A[1 \ldots j-1]$, we do it the following way: compare $A[j]$ to $A[i]$, starting with $i = j - 1$ and going down to $i = 1$. Continue at long as $A[j] < A[i]$.

Merge sort as defined is stable, because when two elements compared are equal, the tie is broken by taking the element from array L which keeps them in the original order.

Heapsort and quicksort are not stable.

One scheme that makes a sorting algorithm stable is to store the index of each element (the element's place in the original ordering) with the element. When comparing two elements, compare them by their values and break ties by their indices.

Additional space requirements: For n elements, their indices are $1 \ldots n$. Each can be written in $\lg n$ bits, so together they take $O(n \lg n)$ additional space.

Additional time requirements: The worst case is when all elements are equal. The asymptotic time does not change because we add a constant amount of work to each comparison.

Solution to Exercise 8.3-3

This solution is also posted publicly

Basis: If $d = 1$, there's only one digit, so sorting on that digit sorts the array.

Inductive step: Assuming that radix sort works for $d - 1$ digits, we'll show that it works for d digits.

Radix sort sorts separately on each digit, starting from digit 1. Thus, radix sort of d digits, which sorts on digits $1, \ldots, d$ is equivalent to radix sort of the low-order $d - 1$ digits followed by a sort on digit d. By our induction hypothesis, the sort of the low-order $d - 1$ digits works, so just before the sort on digit d, the elements are in order according to their low-order $d - 1$ digits.

The sort on digit d will order the elements by their dth digit. Consider two elements, a and b, with dth digits a_d and b_d respectively.

- If $a_d < b_d$, the sort will put a before b, which is correct, since $a < b$ regardless of the low-order digits.
- If $a_d > b_d$, the sort will put a after b, which is correct, since $a > b$ regardless of the low-order digits.
- If $a_d = b_d$, the sort will leave a and b in the same order they were in, because it is stable. But that order is already correct, since the correct order of a and b is determined by the low-order $d - 1$ digits when their dth digits are equal, and the elements are already sorted by their low-order $d - 1$ digits.

If the intermediate sort were not stable, it might rearrange elements whose dth digits were equal—elements that *were* in the right order after the sort on their lower-order digits.

Solution to Exercise 8.3-4

This solution is also posted publicly

Treat the numbers as 3-digit numbers in radix n. Each digit ranges from 0 to $n-1$. Sort these 3-digit numbers with radix sort.

There are 3 calls to counting sort, each taking $\Theta(n+n) = \Theta(n)$ time, so that the total time is $\Theta(n)$.

Solution to Exercise 8.4-2

The worst-case running time for the bucket-sort algorithm occurs when the assumption of uniformly distributed input does not hold. If, for example, all the input ends up in the first bucket, then in the insertion sort phase it needs to sort all the input, which takes $O(n^2)$ time.

A simple change that will preserve the linear expected running time and make the worst-case running time $O(n \lg n)$ is to use a worst-case $O(n \lg n)$-time algorithm, such as merge sort, instead of insertion sort when sorting the buckets.

Solution to Problem 8-1

This solution is also posted publicly

a. For a comparison algorithm A to sort, no two input permutations can reach the same leaf of the decision tree, so there must be at least $n!$ leaves reached in T_A, one for each possible input permutation. Since A is a deterministic algorithm, it must always reach the same leaf when given a particular permutation as input, so at most $n!$ leaves are reached (one for each permutation). Therefore exactly $n!$ leaves are reached, one for each input permutation.

These $n!$ leaves will each have probability $1/n!$, since each of the $n!$ possible permutations is the input with the probability $1/n!$. Any remaining leaves will have probability 0, since they are not reached for any input.

Without loss of generality, we can assume for the rest of this problem that paths leading only to 0-probability leaves aren't in the tree, since they cannot affect the running time of the sort. That is, we can assume that T_A consists of only the $n!$ leaves labeled $1/n!$ and their ancestors.

b. If $k > 1$, then the root of T is not a leaf. This implies that all of T's leaves are leaves in LT and RT. Since every leaf at depth h in LT or RT has depth $h+1$ in T, $D(T)$ must be the sum of $D(LT)$, $D(RT)$, and k, the total number of leaves. To prove this last assertion, let $d_T(x) =$ depth of node x in tree T. Then,

$$\begin{aligned}
D(T) &= \sum_{x \in \text{leaves}(T)} d_T(x) \\
&= \sum_{x \in \text{leaves}(LT)} d_T(x) + \sum_{x \in \text{leaves}(RT)} d_T(x) \\
&= \sum_{x \in \text{leaves}(LT)} (d_{LT}(x) + 1) + \sum_{x \in \text{leaves}(RT)} (d_{RT}(x) + 1) \\
&= \sum_{x \in \text{leaves}(LT)} d_{LT}(x) + \sum_{x \in \text{leaves}(RT)} d_{RT}(x) + \sum_{x \in \text{leaves}(T)} 1 \\
&= D(LT) + D(RT) + k \; .
\end{aligned}$$

c. To show that $d(k) = \min_{1 \le i \le k-1} \{d(i) + d(k-i) + k\}$ we will show separately that

$$d(k) \le \min_{1 \le i \le k-1} \{d(i) + d(k-i) + k\}$$

and

$$d(k) \ge \min_{1 \le i \le k-1} \{d(i) + d(k-i) + k\} \; .$$

- To show that $d(k) \le \min_{1 \le i \le k-1} \{d(i) + d(k-i) + k\}$, we need only show that $d(k) \le d(i) + d(k-i) + k$, for $i = 1, 2, \ldots, k-1$. For any i from 1 to $k-1$ we can find trees RT with i leaves and LT with $k-i$ leaves such that $D(RT) = d(i)$ and $D(LT) = d(k-i)$. Construct T such that RT and LT are the right and left subtrees of T's root respectively. Then

  $$\begin{aligned}
  d(k) &\le D(T) && \text{(by definition of } d \text{ as min } D(T) \text{ value)} \\
  &= D(RT) + D(LT) + k && \text{(by part (b))} \\
  &= d(i) + d(k-i) + k && \text{(by choice of } RT \text{ and } LT) \; .
  \end{aligned}$$

- To show that $d(k) \ge \min_{1 \le i \le k-1} \{d(i) + d(k-i) + k\}$, we need only show that $d(k) \ge d(i) + d(k-i) + k$, for some i in $\{1, 2, \ldots, k-1\}$. Take the tree T with k leaves such that $D(T) = d(k)$, let RT and LT be T's right and left subtree, respecitvely, and let i be the number of leaves in RT. Then $k-i$ is the number of leaves in LT and

  $$\begin{aligned}
  d(k) &= D(T) && \text{(by choice of } T) \\
  &= D(RT) + D(LT) + k && \text{(by part (b))} \\
  &\ge d(i) + d(k-i) + k && \text{(by defintion of } d \text{ as min } D(T) \text{ value)} \; .
  \end{aligned}$$

  Neither i nor $k-i$ can be 0 (and hence $1 \le i \le k-1$), since if one of these were 0, either RT or LT would contain all k leaves of T, and that k-leaf subtree would have a D equal to $D(T) - k$ (by part (b)), contradicting the choice of T as the k-leaf tree with the minimum D.

d. Let $f_k(i) = i \lg i + (k-i) \lg(k-i)$. To find the value of i that minimizes f_k, find the i for which the derivative of f_k with respect to i is 0:

$$\begin{aligned}
f_k'(i) &= \frac{d}{di}\left(\frac{i \ln i + (k-i)\ln(k-i)}{\ln 2}\right) \\
&= \frac{\ln i + 1 - \ln(k-i) - 1}{\ln 2} \\
&= \frac{\ln i - \ln(k-i)}{\ln 2}
\end{aligned}$$

is 0 at $i = k/2$. To verify this is indeed a minimum (not a maximum), check that the second derivative of f_k is positive at $i = k/2$:

$$\begin{aligned} f_k''(i) &= \frac{d}{di}\left(\frac{\ln i - \ln(k-i)}{\ln 2}\right) \\ &= \frac{1}{\ln 2}\left(\frac{1}{i} + \frac{1}{k-i}\right) . \\ f_k''(k/2) &= \frac{1}{\ln 2}\left(\frac{2}{k} + \frac{2}{k}\right) \\ &= \frac{1}{\ln 2} \cdot \frac{4}{k} \\ &> 0 \qquad\qquad \text{since } k > 1 . \end{aligned}$$

Now we use substitution to prove $d(k) = \Omega(k \lg k)$. The base case of the induction is satisfied because $d(1) \geq 0 = c \cdot 1 \cdot \lg 1$ for any constant c. For the inductive step we assume that $d(i) \geq ci \lg i$ for $1 \leq i \leq k-1$, where c is some constant to be determined.

$$\begin{aligned} d(k) &= \min_{1 \leq i \leq k-1} \{d(i) + d(k-i) + k\} \\ &\geq \min_{1 \leq i \leq k-1} \{c(i \lg i + (k-i)\lg(k-i)) + k\} \\ &= \min_{1 \leq i \leq k-1} \{c f_k(i) + k\} \\ &= c\left(\frac{k}{2}\lg\frac{k}{2}\left(k - \frac{k}{2}\right)\lg\left(k - \frac{k}{2}\right)\right) + k \\ &= ck \lg\left(\frac{k}{2}\right) + k \\ &= c(k \lg k - k) + k \\ &= ck \lg k + (k - ck) \\ &\geq ck \lg k \qquad \text{if } c \leq 1 , \end{aligned}$$

and so $d(k) = \Omega(k \lg k)$.

e. Using the result of part (d) and the fact that T_A (as modified in our solution to part (a)) has $n!$ leaves, we can conclude that

$$D(T_A) \geq d(n!) = \Omega(n! \lg(n!)) .$$

$D(T_A)$ is the sum of the decision-tree path lengths for sorting all input permutations, and the path lengths are proportional to the run time. Since the $n!$ permutations have equal probability $1/n!$, the expected time to sort n random elements (1 input permutation) is the total time for all permutations divided by $n!$:

$$\frac{\Omega(n! \lg(n!))}{n!} = \Omega(\lg(n!)) = \Omega(n \lg n) .$$

f. We will show how to modify a randomized decision tree (algorithm) to define a deterministic decision tree (algorithm) that is at least as good as the randomized one in terms of the average number of comparisons.

At each randomized node, pick the child with the smallest subtree (the subtree with the smallest average number of comparisons on a path to a leaf). Delete all

the other children of the randomized node and splice out the randomized node itself.

The deterministic algorithm corresponding to this modified tree still works, because the randomized algorithm worked no matter which path was taken from each randomized node.

The average number of comparisons for the modified algorithm is no larger than the average number for the original randomized tree, since we discarded the higher-average subtrees in each case. In particular, each time we splice out a randomized node, we leave the overall average less than or equal to what it was, because

- the same set of input permutations reaches the modified subtree as before, but those inputs are handled in less than or equal to average time than before, and
- the rest of the tree is unmodified.

The randomized algorithm thus takes at least as much time on average as the corresponding deterministic one. (We've shown that the expected running time for a deterministic comparison sort is $\Omega(n \lg n)$, hence the expected time for a randomized comparison sort is also $\Omega(n \lg n)$.)

Solution to Problem 8-3

a. The usual, unadorned radix sort algorithm will not solve this problem in the required time bound. The number of passes, d, would have to be the number of digits in the largest integer. Suppose that there are m integers; we always have $m \le n$. In the worst case, we would have one integer with $n/2$ digits and $n/2$ integers with one digit each. We assume that the range of a single digit is constant. Therefore, we would have $d = n/2$ and $m = n/2 + 1$, and so the running time would be $\Theta(dm) = \Theta(n^2)$.

Let us assume without loss of generality that all the integers are positive and have no leading zeros. (If there are negative integers or 0, deal with the positive numbers, negative numbers, and 0 separately.) Under this assumption, we can observe that integers with more digits are always greater than integers with fewer digits. Thus, we can first sort the integers by number of digits (using counting sort), and then use radix sort to sort each group of integers with the same length. Noting that each integer has between 1 and n digits, let m_i be the number of integers with i digits, for $i = 1, 2, \ldots, n$. Since there are n digits altogether, we have $\sum_{i=1}^{n} i \cdot m_i = n$.

It takes $O(n)$ time to compute how many digits all the integers have and, once the numbers of digits have been computed, it takes $O(m + n) = O(n)$ time to group the integers by number of digits. To sort the group with m_i digits by radix sort takes $\Theta(i \cdot m_i)$ time. The time to sort all groups, therefore, is

$$\begin{aligned}\sum_{i=1}^{n} \Theta(i \cdot m_i) &= \Theta\left(\sum_{i=1}^{n} i \cdot m_i\right) \\ &= \Theta(n) \, .\end{aligned}$$

b. One way to solve this problem is by a radix sort from right to left. Since the strings have varying lengths, however, we have to pad out all strings that are shorter than the longest string. The padding is on the right end of the string, and it's with a special character that is lexicographically less than any other character (e.g., in C, the character `'\0'` with ASCII value 0). Of course, we don't have to actually change any string; if we want to know the jth character of a string whose length is k, then if $j > k$, the jth character is the pad character.

Unfortunately, this scheme does not always run in the required time bound. Suppose that there are m strings and that the longest string has d characters. In the worst case, one string has $n/2$ characters and, before padding, $n/2$ strings have one character each. As in part (a), we would have $d = n/2$ and $m = n/2 + 1$. We still have to examine the pad characters in each pass of radix sort, even if we don't actually create them in the strings. Assuming that the range of a single character is constant, the running time of radix sort would be $\Theta(dm) = \Theta(n^2)$.

To solve the problem in $O(n)$ time, we use the property that, if the first letter of string x is lexicographically less that the first letter of string y, then x is lexicographically less than y, regardless of the lengths of the two strings. We take advantage of this property by sorting the strings on the first letter, using counting sort. We take an empty string as a special case and put it first. We gather together all strings with the same first letter as a group. Then we recurse, *within each group*, based on each string with the first letter removed.

The correctness of this algorithm is straightforward. Analyzing the running time is a bit trickier. Let us count the number of times that each string is sorted by a call of counting sort. Suppose that the ith string, s_i, has length l_i. Then s_i is sorted by at most $l_i + 1$ counting sorts. (The "$+1$" is because it may have to be sorted as an empty string at some point; for example, `ab` and `a` end up in the same group in the first pass and are then ordered based on `b` and the empty string in the second pass. The string `a` is sorted its length, 1, time plus one more time.) A call of counting sort on t strings takes $\Theta(t)$ time (remembering that the number of different characters on which we are sorting is a constant.) Thus, the total time for all calls of counting sort is

$$\begin{aligned} O\left(\sum_{i=1}^{m}(l_i + 1)\right) &= O\left(\sum_{i=1}^{m} l_i + m\right) \\ &= O(n + m) \\ &= O(n)\ , \end{aligned}$$

where the second line follows from $\sum_{i=1}^{m} l_i = n$, and the last line is because $m \le n$.

Solution to Problem 8-4

a. Compare each red jug with each blue jug. Since there are n red jugs and n blue jugs, that will take $\Theta(n^2)$ comparisons in the worst case.

b. To solve the problem, an algorithm has to perform a series of comparisons until it has enough information to determine the matching. We can view the computation of the algorithm in terms of a decision tree. Every internal node is labeled with two jugs (one red, one blue) which we compare, and has three outgoing edges (red jug smaller, same size, or larger than the blue jug). The leaves are labeled with a unique matching of jugs.

The height of the decision tree is equal to the worst-case number of comparisons the algorithm has to make to determine the matching. To bound that size, let us first compute the number of possible matchings for n red and n blue jugs.

If we label the red jugs from 1 to n and we also label the blue jugs from 1 to n before starting the comparisons, every outcome of the algorithm can be represented as a set

$$\{(i, \pi(i)) : 1 \le i \le n \text{ and } \pi \text{ is a permutation on } \{1, \ldots, n\}\} \ ,$$

which contains the pairs of red jugs (first component) and blue jugs (second component) that are matched up. Since every permutation π corresponds to a different outcome, there must be exactly $n!$ different results.

Now we can bound the height h of our decision tree. Every tree with a branching factor of 3 (every inner node has at most three children) has at most 3^h leaves. Since the decison tree must have at least $n!$ children, it follows that

$$3^h \ge n! \ge (n/e)^n \Rightarrow h \ge n \log_3 n - n \log_3 e = \Omega(n \lg n) \ .$$

So any algorithm solving the problem must use $\Omega(n \lg n)$ comparisons.

c. Assume that the red jugs are labeled with numbers $1, 2, \ldots, n$ and so are the blue jugs. The numbers are arbitrary and do not correspond to the volumes of jugs, but are just used to refer to the jugs in the algorithm description. Moreover, the output of the algorithm will consist of n distinct pairs (i, j), where the red jug i and the blue jug j have the same volume.

The procedure MATCH-JUGS takes as input two sets representing jugs to be matched: $R \subseteq \{1, \ldots, n\}$, representing red jugs, and $B \subseteq \{1, \ldots, n\}$, representing blue jugs. We will call the procedure only with inputs that can be matched; one necessary condition is that $|R| = |B|$.

```
MATCH-JUGS(R, B)
  if |R| == 0                 // sets are empty
      return
  if |R| == 1                 // sets contain just one jug each
      let R = {r} and B = {b}
      output "(r, b)"
      return
  else r = a randomly chosen jug in R
      compare r to every jug of B
      B_< = the set of jugs in B that are smaller than r
      B_> = the set of jugs in B that are larger than r
      b = the one jug in B with the same size as r
      compare b to every jug of R − {r}
      R_< = the set of jugs in R that are smaller than b
      R_> = the set of jugs in R that are larger than b
      output "(r, b)"
      MATCH-JUGS(R_<, B_<)
      MATCH-JUGS(R_>, B_>)
```

Correctness can be seen as follows (remember that $|R| = |B|$ in each call). Once we pick r randomly from R, there will be a matching among the jugs in volume smaller than r (which are in the sets $R_<$ and $B_<$), and likewise between the jugs larger than r (which are in $R_>$ and $B_>$). Termination is also easy to see: since $|R_<| + |R_>| < |R|$ in every recursive step, the size of the first parameter reduces with every recursive call. It eventually must reach 0 or 1, in which case the recursion terminates.

What about the running time? The analysis of the expected number of comparisons is similar to that of the quicksort algorithm in Section 7.4.2. Let us order the jugs as $r_1, \ldots, r_n$ and $b_1, \ldots, b_n$ where $r_i < r_{i+1}$ and $b_i < b_{i+1}$ for $i = 1, \ldots, n$, and $r_i = b_i$. Our analysis uses indicator random variables

$$X_{ij} = \mathrm{I}\{\text{red jug } r_i \text{ is compared to blue jug } b_j\} \ .$$

As in quicksort, a given pair r_i and b_j is compared at most once. When we compare r_i to every jug in B, jug r_i will not be put in either $R_<$ or $R_>$. When we compare b_i to every jug in $R - \{r_i\}$, jug b_i is not put into either $B_<$ or $B_>$. The total number of comparisons is

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \ .$$

To calculate the expected value of X, we follow the quicksort analysis to arrive at

$$\mathrm{E}[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr\{r_i \text{ is compared to } b_j\} \ .$$

As in the quicksort analysis, once we choose a jug r_k such that $r_i < r_k < b_j$, we will put r_i in $R_<$ and b_j in $R_>$, and so r_i and b_j will never be compared

again. Let us denote $R_{ij} = \{r_i, \ldots, r_j\}$. Then jugs r_i and b_j will be compared if and only if the first jug in R_{ij} to be chosen is either r_i or r_j.

Still following the quicksort analysis, until a jug from R_{ij} is chosen, the entire set R_{ij} is together. Any jug in R_{ij} is equally likely to be first one chosen. Since $|R_{ij}| = j - i + 1$, the probability of any given jug being the first one chosen in R_{ij} is $1/(j-i+1)$. The remainder of the analysis is the same as the quicksort analysis, and we arrive at the solution of $O(n \lg n)$ comparisons.

Just like in quicksort, in the worst case we always choose the largest (or smallest) jug to partition the sets, which reduces the set sizes by only 1. The running time then obeys the recurrence $T(n) = T(n-1) + \Theta(n)$, and the number of comparisons we make in the worst case is $T(n) = \Theta(n^2)$.

Solution to Problem 8-7

a. $A[q]$ must go the wrong place, because it goes where $A[p]$ should go. Since $A[p]$ is the smallest value in array A that goes to the wrong array location, $A[p]$ must be smaller than $A[q]$.

b. From how we have defined the array B, we have that if $A[i] \le A[j]$ then $B[i] \le B[j]$. Therefore, algorithm X performs the same sequence of exchanges on array B as it does on array A. The output produced on array A is of the form $\ldots A[q] \ldots A[p] \ldots$, and so the output produced on array B is of the form $\ldots B[q] \ldots B[p] \ldots$, or $\ldots 1 \ldots 0 \ldots$. Hence algorithm X fails to sort array B correctly.

c. The even steps perform fixed permutations. The odd steps sort each column by some sorting algorithm, which might not be an oblivious compare-exchange algorithm. But the result of sorting each column would be the same as if we did use an oblivious compare-exchange algorithm.

d. After step 1, each column has 0s on top and 1s on the bottom, with at most one transition between 0s and 1s, and it is a $0 \to 1$ transition. (As we read the array in column-major order, all $1 \to 0$ transitions occur between adjacent columns.) After step 2, therefore, each consecutive group of r/s rows, read in row-major order, has at most one transition, and again it is a $0 \to 1$ transition. All $1 \to 0$ transitions occur at the end of a group of r/s rows. Since there are s groups of r/s rows, there are at most s dirty rows, and the rest of the rows are clean. Step 3 moves the 0s to the top rows and the 1s to the bottom rows. The s dirty rows are somewhere in the middle.

e. The dirty area after step 3 is at most s rows high and s columns wide, and so its area is at most s^2. Step 4 turns the clean 0s in the top rows into a clean area on the left, the clean 1s in the bottom rows into a clean area on the right, and the dirty area of size s^2 is between the two clean areas.

f. First, we argue that if the dirty area after step 4 has size at most $r/2$, then steps 5–8 complete the sorting. If the dirty area has size at most $r/2$ (half a column), then it either resides entirely in one column or it resides in the bottom

half of one column and the top half of the next column. In the former case, step 5 sorts the column containing the dirty area, and steps 6–8 maintain that the array is sorted. In the latter case, step 5 cannot increase the size of the dirty area, step 6 moves the entire dirty area into the same column, step 7 sorts it, and step 8 moves it back.

Second, we argue that the dirty area after step 4 has size at most $r/2$. But that follows immediately from the requirement that $r \geq 2s^2$ and the property that after step 4, the dirty area has size at most s^2.

g. If s does not divide r, then after step 2, we can see up to s $0 \rightarrow 1$ transitions and $s-1$ $1 \rightarrow 0$ transitions in the rows. After step 3, we would have up to $2s-1$ dirty rows, for a dirty area size of at most $2s^2 - s$. To push the correctness proof through, we need $2s^2 - s \leq r/2$, or $r \geq 4s^2 - 2s$.

h. We can reduce the number of transitions in the rows after step 2 back down to at most s by sorting every other column in reverse order in step 1. Now if we have a transition (either $1 \rightarrow 0$ or $0 \rightarrow 1$) between columns after step 1, then either one of the columns had all 1s or the other had all 0s, in which case we would not have a transition within one of the columns.

Lecture Notes for Chapter 9: Medians and Order Statistics

Chapter 9 overview

- ***i th order statistic*** is the ith smallest element of a set of n elements.
- The ***minimum*** is the first order statistic ($i = 1$).
- The ***maximum*** is the nth order statistic ($i = n$).
- A ***median*** is the "halfway point" of the set.
- When n is odd, the median is unique, at $i = (n + 1)/2$.
- When n is even, there are two medians:
  - The ***lower median***, at $i = n/2$, and
  - The ***upper median***, at $i = n/2 + 1$.
  - We mean lower median when we use the phrase "the median."

The ***selection problem***:

Input: A set A of n distinct numbers and a number i, with $1 \le i \le n$.

Output: The element $x \in A$ that is larger than exactly $i - 1$ other elements in A. In other words, the ith smallest element of A.

We can easily solve the selection problem in $O(n \lg n)$ time:

- Sort the numbers using an $O(n \lg n)$-time algorithm, such as heapsort or merge sort.
- Then return the ith element in the sorted array.

There are faster algorithms, however.

- First, we'll look at the problem of selecting the minimum and maximum of a set of elements.
- Then, we'll look at a simple general selection algorithm with a time bound of $O(n)$ in the average case.
- Finally, we'll look at a more complicated general selection algorithm with a time bound of $O(n)$ in the worst case.

Minimum and maximum

We can easily obtain an upper bound of $n-1$ comparisons for finding the minimum of a set of n elements.

- Examine each element in turn and keep track of the smallest one.
- This is the best we can do, because each element, except the minimum, must be compared to a smaller element at least once.

The following pseudocode finds the minimum element in array $A[1..n]$:

```
MINIMUM(A, n)
  min = A[1]
  for i = 2 to n
      if min > A[i]
          min = A[i]
  return min
```

The maximum can be found in exactly the same way by replacing the $>$ with $<$ in the above algorithm.

Simultaneous minimum and maximum

Some applications need both the minimum and maximum of a set of elements.

- For example, a graphics program may need to scale a set of (x, y) data to fit onto a rectangular display. To do so, the program must first find the minimum and maximum of each coordinate.

A simple algorithm to find the minimum and maximum is to find each one independently. There will be $n-1$ comparisons for the minimum and $n-1$ comparisons for the maximum, for a total of $2n-2$ comparisons. This will result in $\Theta(n)$ time.

In fact, at most $3\lfloor n/2 \rfloor$ comparisons suffice to find both the minimum and maximum:

- Maintain the minimum and maximum of elements seen so far.
- Don't compare each element to the minimum and maximum separately.
- Process elements in pairs.
- Compare the elements of a pair to each other.
- Then compare the larger element to the maximum so far, and compare the smaller element to the minimum so far.

This leads to only 3 comparisons for every 2 elements.

Setting up the initial values for the min and max depends on whether n is odd or even.

- If n is even, compare the first two elements and assign the larger to max and the smaller to min. Then process the rest of the elements in pairs.
- If n is odd, set both min and max to the first element. Then process the rest of the elements in pairs.

Analysis of the total number of comparisons

- If n is even, we do 1 initial comparison and then $3(n-2)/2$ more comparisons.

$$\begin{aligned} \text{\# of comparisons} &= \frac{3(n-2)}{2} + 1 \\ &= \frac{3n-6}{2} + 1 \\ &= \frac{3n}{2} - 3 + 1 \\ &= \frac{3n}{2} - 2 \,. \end{aligned}$$

- If n is odd, we do $3(n-1)/2 = 3\lfloor n/2 \rfloor$ comparisons.

In either case, the maximum number of comparisons is $\leq 3\lfloor n/2 \rfloor$.

Selection in expected linear time

Selection of the ith smallest element of the array A can be done in $\Theta(n)$ time.

The function RANDOMIZED-SELECT uses RANDOMIZED-PARTITION from the quicksort algorithm in Chapter 7. RANDOMIZED-SELECT differs from quicksort because it recurses on one side of the partition only.

```
RANDOMIZED-SELECT(A, p, r, i)
  if p == r
      return A[p]
  q = RANDOMIZED-PARTITION(A, p, r)
  k = q - p + 1
  if i == k            // pivot value is the answer
      return A[q]
  elseif i < k
      return RANDOMIZED-SELECT(A, p, q - 1, i)
  else return RANDOMIZED-SELECT(A, q + 1, r, i - k)
```

After the call to RANDOMIZED-PARTITION, the array is partitioned into two subarrays $A[p\,..\,q-1]$ and $A[q+1\,..\,r]$, along with a ***pivot*** element $A[q]$.

- The elements of subarray $A[p\,..\,q-1]$ are all $\leq A[q]$.
- The elements of subarray $A[q+1\,..\,r]$ are all $> A[q]$.
- The pivot element is the kth element of the subarray $A[p\,..\,r]$, where $k = q-p+1$.
- If the pivot element is the ith smallest element (i.e., $i = k$), return $A[q]$.
- Otherwise, recurse on the subarray containing the ith smallest element.
  - If $i < k$, this subarray is $A[p\,..\,q-1]$, and we want the ith smallest element.
  - If $i > k$, this subarray is $A[q+1\,..\,r]$ and, since there are k elements in $A[p\,..\,r]$ that precede $A[q+1\,..\,r]$, we want the $(i-k)$th smallest element of this subarray.

Analysis

Worst-case running time

$\Theta(n^2)$, because we could be extremely unlucky and always recurse on a subarray that is only 1 element smaller than the previous subarray.

Expected running time

RANDOMIZED-SELECT works well on average. Because it is randomized, no particular input brings out the worst-case behavior consistently.

The running time of RANDOMIZED-SELECT is a random variable that we denote by $T(n)$. We obtain an upper bound on $\mathrm{E}[T(n)]$ as follows:

- RANDOMIZED-PARTITION is equally likely to return any element of A as the pivot.
- For each k such that $1 \le k \le n$, the subarray $A[p..q]$ has k elements (all $\le$ pivot) with probability $1/n$. *[Note that we're now considering a subarray that includes the pivot, along with elements less than the pivot.]*
- For $k = 1, 2, \ldots, n$, define indicator random variable

  $X_k = \mathrm{I}\{\text{subarray } A[p..q] \text{ has exactly } k \text{ elements}\}$.
- Since $\Pr\{\text{subarray } A[p..q] \text{ has exactly } k \text{ elements}\} = 1/n$, Lemma 5.1 says that $\mathrm{E}[X_k] = 1/n$.
- When we call RANDOMIZED-SELECT, we don't know if it will terminate immediately with the correct answer, recurse on $A[p..q-1]$, or recurse on $A[q+1..r]$. It depends on whether the ith smallest element is less than, equal to, or greater than the pivot element $A[q]$.
- To obtain an upper bound, we assume that $T(n)$ is monotonically increasing and that the ith smallest element is always in the larger subarray.
- For a given call of RANDOMIZED-SELECT, $X_k = 1$ for exactly one value of k, and $X_k = 0$ for all other k.
- When $X_k = 1$, the two subarrays have sizes $k-1$ and $n-k$.
- For a subproblem of size n, RANDOMIZED-PARTITION takes $O(n)$ time. *[Actually, it takes $\Theta(n)$ time, but $O(n)$ suffices, since we're obtaining only an upper bound on the expected running time.]*
- Therefore, we have the recurrence

$$\begin{aligned} T(n) &\le \sum_{k=1}^{n} X_k \cdot (T(\max(k-1, n-k)) + O(n)) \\ &= \sum_{k=1}^{n} X_k \cdot T(\max(k-1, n-k)) + O(n) \; . \end{aligned}$$

- Taking expected values gives

$$\begin{aligned} &\mathrm{E}[T(n)] \\ &\le \mathrm{E}\left[\sum_{k=1}^{n} X_k \cdot T(\max(k-1, n-k)) + O(n)\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \mathrm{E}\left[X_k \cdot T(\max(k-1, n-k))\right] + O(n) \qquad \text{(linearity of expectation)} \\
&= \sum_{k=1}^{n} \mathrm{E}\left[X_k\right] \cdot \mathrm{E}\left[T(\max(k-1, n-k))\right] + O(n) \quad \text{(equation (C.24))} \\
&= \sum_{k=1}^{n} \frac{1}{n} \cdot \mathrm{E}\left[T(\max(k-1, n-k))\right] + O(n) \; .
\end{aligned}
$$

- We rely on X_k and $T(\max(k-1, n-k))$ being independent random variables in order to apply equation (C.24).
- Looking at the expression $\max(k-1, n-k)$, we have

$$
\max(k-1, n-k) = \begin{cases} k-1 & \text{if } k > \lceil n/2 \rceil \; , \\ n-k & \text{if } k \le \lceil n/2 \rceil \; . \end{cases}
$$

  - If n is even, each term from $T(\lceil n/2 \rceil)$ up to $T(n-1)$ appears exactly twice in the summation.
  - If n is odd, these terms appear twice and $T(\lfloor n/2 \rfloor)$ appears once.

- Either way,

$$
\mathrm{E}\left[T(n)\right] \le \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} \mathrm{E}\left[T(k)\right] + O(n) \; .
$$

- Solve this recurrence by substitution:
  - Guess that $T(n) \le cn$ for some constant c that satisfies the initial conditions of the recurrence.
  - Assume that $T(n) = O(1)$ for $n <$ some constant. We'll pick this constant later.
  - Also pick a constant a such that the function described by the $O(n)$ term is bounded from above by an for all $n > 0$.
  - Using this guess and constants c and a, we have

$$
\begin{aligned}
\mathrm{E}\left[T(n)\right] &\le \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} ck + an \\
&= \frac{2c}{n} \left(\sum_{k=1}^{n-1} k - \sum_{k=1}^{\lfloor n/2 \rfloor - 1} k \right) + an \\
&= \frac{2c}{n} \left(\frac{(n-1)n}{2} - \frac{(\lfloor n/2 \rfloor - 1)\,\lfloor n/2 \rfloor}{2} \right) + an \\
&\le \frac{2c}{n} \left(\frac{(n-1)n}{2} - \frac{(n/2-2)(n/2-1)}{2} \right) + an \\
&= \frac{2c}{n} \left(\frac{n^2-n}{2} - \frac{n^2/4 - 3n/2 + 2}{2} \right) + an \\
&= \frac{c}{n} \left(\frac{3n^2}{4} + \frac{n}{2} - 2 \right) + an
\end{aligned}
$$

$$\begin{aligned} &= c\left(\frac{3n}{4} + \frac{1}{2} - \frac{2}{n}\right) + an \\ &\le \frac{3cn}{4} + \frac{c}{2} + an \\ &= cn - \left(\frac{cn}{4} - \frac{c}{2} - an\right) . \end{aligned}$$

- To complete this proof, we choose c such that

$$\begin{aligned} cn/4 - c/2 - an &\ge 0 \\ cn/4 - an &\ge c/2 \\ n(c/4 - a) &\ge c/2 \\ n &\ge \frac{c/2}{c/4 - a} \\ n &\ge \frac{2c}{c - 4a} . \end{aligned}$$

- Thus, as long as we assume that $T(n) = O(1)$ for $n < 2c/(c-4a)$, we have $\mathrm{E}[T(n)] = O(n)$.

Therefore, we can determine any order statistic in linear time on average.

Selection in worst-case linear time

We can find the ith smallest element in $O(n)$ time *in the worst case*. We'll describe a procedure SELECT that does so.

SELECT recursively partitions the input array.

- ***Idea:*** Guarantee a good split when the array is partitioned.
- Will use the deterministic procedure PARTITION, but with a small modification. Instead of assuming that the last element of the subarray is the pivot, the modified PARTITION procedure is told which element to use as the pivot.

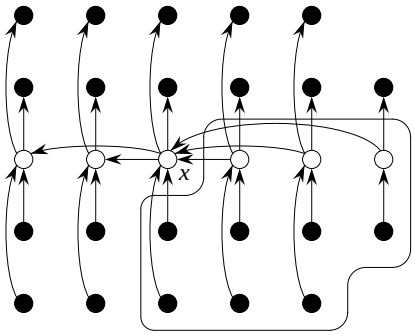
SELECT works on an array of $n > 1$ elements. It executes the following steps:

1. Divide the n elements into groups of 5. Get $\lceil n/5 \rceil$ groups: $\lfloor n/5 \rfloor$ groups with exactly 5 elements and, if 5 does not divide n, one group with the remaining n mod 5 elements.
2. Find the median of each of the $\lceil n/5 \rceil$ groups:
   - Run insertion sort on each group. Takes $O(1)$ time per group since each group has ≤ 5 elements.
   - Then just pick the median from each group, in $O(1)$ time.
3. Find the median x of the $\lceil n/5 \rceil$ medians by a recursive call to SELECT. (If $\lceil n/5 \rceil$ is even, then follow our convention and find the lower median.)
4. Using the modified version of PARTITION that takes the pivot element as input, partition the input array around x. Let x be the kth element of the array after partitioning, so that there are $k-1$ elements on the low side of the partition and $n-k$ elements on the high side.

5. Now there are three possibilities:
   - If $i = k$, just return x.
   - If $i < k$, return the ith smallest element on the low side of the partition by making a recursive call to SELECT.
   - If $i > k$, return the $(i-k)$th smallest element on the high side of the partition by making a recursive call to SELECT.

Analysis

Start by getting a lower bound on the number of elements that are greater than the partitioning element x:



[Each group is a column. Each white circle is the median of a group, as found in step 2. Arrows go from larger elements to smaller elements, based on what we know after step 4. Elements in the region on the lower right are known to be greater than x.]

- At least half of the medians found in step 2 are $\geq x$.
- Look at the groups containing these medians that are $\geq x$. All of them contribute 3 elements that are $> x$ (the median of the group and the 2 elements in the group greater than the group's median), except for 2 of the groups: the group containing x (which has only 2 elements $> x$) and the group with < 5 elements.
- Forget about these 2 groups. That leaves $\geq \left\lceil \frac{1}{2} \left\lceil \frac{n}{5} \right\rceil \right\rceil - 2$ groups with 3 elements known to be $> x$.
- Thus, we know that at least

  $$3\left(\left\lceil \frac{1}{2} \left\lceil \frac{n}{5} \right\rceil \right\rceil - 2\right) \geq \frac{3n}{10} - 6$$

  elements are $> x$.

Symmetrically, the number of elements that are $< x$ is at least $3n/10 - 6$.

Therefore, when we call SELECT recursively in step 5, it's on $\leq 7n/10 + 6$ elements.

Develop a recurrence for the worst-case running time of SELECT:

- Steps 1, 2, and 4 each take $O(n)$ time:

  - Step 1: making groups of 5 elements takes $O(n)$ time.
  - Step 2: sorting $\lceil n/5 \rceil$ groups in $O(1)$ time each.
  - Step 4: partitioning the n-element array around x takes $O(n)$ time.
- Step 3 takes time $T(\lceil n/5 \rceil)$.
- Step 5 takes time $\le T(7n/10 + 6)$, assuming that $T(n)$ is monotonically increasing.
- Assume that $T(n) = O(1)$ for small enough n. We'll use $n < 140$ as "small enough." Why 140? We'll see why later.
- Thus, we get the recurrence

$$T(n) \le \begin{cases} O(1) & \text{if } n < 140 , \\ T(\lceil n/5 \rceil) + T(7n/10 + 6) + O(n) & \text{if } n \ge 140 . \end{cases}$$

Solve this recurrence by substitution:

- ***Inductive hypothesis:*** $T(n) \le cn$ for some constant c and all $n > 0$.
- Assume that c is large enough that $T(n) \le cn$ for all $n < 140$. So we are concerned only with the case $n \ge 140$.
- Pick a constant a such that the function described by the $O(n)$ term in the recurrence is $\le an$ for all $n > 0$.
- Substitute the inductive hypothesis in the right-hand side of the recurrence:

$$\begin{aligned} T(n) &\le c\lceil n/5 \rceil + c(7n/10 + 6) + an \\ &\le cn/5 + c + 7cn/10 + 6c + an \\ &= 9cn/10 + 7c + an \\ &= cn + (-cn/10 + 7c + an) . \end{aligned}$$

- This last quantity is $\le cn$ if

$$\begin{aligned} -cn/10 + 7c + an &\le 0 \\ cn/10 - 7c &\ge an \\ cn - 70c &\ge 10an \\ c(n - 70) &\ge 10an \\ c &\ge 10a(n/(n - 70)) . \end{aligned}$$

- Because we assumed that $n \ge 140$, we have $n/(n - 70) \le 2$.
- Thus, $20a \ge 10a(n/(n-70))$, so choosing $c \ge 20a$ gives $c \ge 10a(n/(n-70))$, which in turn gives us the condition we need to show that $T(n) \le cn$.
- We conclude that $T(n) = O(n)$, so that SELECT runs in linear time in all cases.
- Why 140? We could have used any integer strictly greater than 70.
  - Observe that for $n > 70$, the fraction $n/(n - 70)$ decreases as n increases.
  - We picked $n \ge 140$ so that the fraction would be ≤ 2, which is an easy constant to work with.
  - We could have picked, say, $n \ge 71$, so that for all $n \ge 71$, the fraction would be $\le 71/(71 - 70) = 71$. Then we would have had $20a \ge 710a$, so we'd have needed to choose $c \ge 710a$.

Notice that SELECT and RANDOMIZED-SELECT determine information about the relative order of elements only by comparing elements.

- Sorting requires $\Omega(n \lg n)$ time in the comparison model.
- Sorting algorithms that run in linear time need to make assumptions about their input.
- Linear-time *selection* algorithms do not require any assumptions about their input.
- Linear-time selection algorithms solve the selection problem without sorting and therefore are not subject to the $\Omega(n \lg n)$ lower bound.

Solutions for Chapter 9: Medians and Order Statistics

Solution to Exercise 9.1-1

The smallest of n numbers can be found with $n-1$ comparisons by conducting a tournament as follows: Compare all the numbers in pairs. Only the smaller of each pair could possibly be the smallest of all n, so the problem has been reduced to that of finding the smallest of $\lceil n/2 \rceil$ numbers. Compare those numbers in pairs, and so on, until there's just one number left, which is the answer.

To see that this algorithm does exactly $n-1$ comparisons, notice that each number except the smallest loses exactly once. To show this more formally, draw a binary tree of the comparisons the algorithm does. The n numbers are the leaves, and each number that came out smaller in a comparison is the parent of the two numbers that were compared. Each non-leaf node of the tree represents a comparison, and there are $n-1$ internal nodes in an n-leaf full binary tree (see Exercise (B.5-3)), so exactly $n-1$ comparisons are made.

In the search for the smallest number, the second smallest number must have come out smallest in every comparison made with it until it was eventually compared with the smallest. So the second smallest is among the elements that were compared with the smallest during the tournament. To find it, conduct another tournament (as above) to find the smallest of these numbers. At most $\lceil \lg n \rceil$ (the height of the tree of comparisons) elements were compared with the smallest, so finding the smallest of these takes $\lceil \lg n \rceil - 1$ comparisons in the worst case.

The total number of comparisons made in the two tournaments was

$$n - 1 + \lceil \lg n \rceil - 1 = n + \lceil \lg n \rceil - 2$$

in the worst case.

Solution to Exercise 9.3-1

This solution is also posted publicly

For groups of 7, the algorithm still works in linear time. The number of elements greater than x (and similarly, the number less than x) is at least

$$4\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{7} \right\rceil \right\rceil - 2\right) \geq \frac{2n}{7} - 8\,,$$

and the recurrence becomes

$$T(n) \le T(\lceil n/7 \rceil) + T(5n/7 + 8) + O(n) ,$$

which can be shown to be $O(n)$ by substitution, as for the groups of 5 case in the text.

For groups of 3, however, the algorithm no longer works in linear time. The number of elements greater than x, and the number of elements less than x, is at least

$$2\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{3} \right\rceil \right\rceil - 2\right) \ge \frac{n}{3} - 4 ,$$

and the recurrence becomes

$$T(n) \le T(\lceil n/3 \rceil) + T(2n/3 + 4) + O(n) ,$$

which does not have a linear solution.

We can prove that the worst-case time for groups of 3 is $\Omega(n \lg n)$. We do so by deriving a recurrence for a particular case that takes $\Omega(n \lg n)$ time.

In counting up the number of elements greater than x (and similarly, the number less than x), consider the particular case in which there are exactly $\left\lceil \frac{1}{2}\left\lceil \frac{n}{3} \right\rceil \right\rceil$ groups with medians $\ge x$ and in which the "leftover" group does contribute 2 elements greater than x. Then the number of elements greater than x is exactly $2\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{3} \right\rceil \right\rceil - 1\right) + 1$ (the -1 discounts x's group, as usual, and the $+1$ is contributed by x's group) $= 2\lceil n/6 \rceil - 1$, and the recursive step for elements $\le x$ has $n - (2\lceil n/6 \rceil - 1) \ge n - (2(n/6 + 1) - 1) = 2n/3 - 1$ elements. Observe also that the $O(n)$ term in the recurrence is really $\Theta(n)$, since the partitioning in step 4 takes $\Theta(n)$ (not just $O(n)$) time. Thus, we get the recurrence

$$T(n) \ge T(\lceil n/3 \rceil) + T(2n/3 - 1) + \Theta(n) \ge T(n/3) + T(2n/3 - 1) + \Theta(n) ,$$

from which you can show that $T(n) \ge cn \lg n$ by substitution. You can also see that $T(n)$ is nonlinear by noticing that each level of the recursion tree sums to n.

[In fact, any odd group size ≥ 5 works in linear time.]

Solution to Exercise 9.3-3

This solution is also posted publicly

A modification to quicksort that allows it to run in $O(n \lg n)$ time in the worst case uses the deterministic PARTITION algorithm that was modified to take an element to partition around as an input parameter.

SELECT takes an array A, the bounds p and r of the subarray in A, and the rank i of an order statistic, and in time linear in the size of the subarray $A[p \,..\, r]$ it returns the ith smallest element in $A[p \,..\, r]$.

```
BEST-CASE-QUICKSORT(A, p, r)
  if p < r
      i = ⌊(r − p + 1)/2⌋
      x = SELECT(A, p, r, i)
      q = PARTITION(x)
      BEST-CASE-QUICKSORT(A, p, q − 1)
      BEST-CASE-QUICKSORT(A, q + 1, r)
```

For an n-element array, the largest subarray that BEST-CASE-QUICKSORT recurses on has $n/2$ elements. This situation occurs when $n = r - p + 1$ is even; then the subarray $A[q + 1 \,..\, r]$ has $n/2$ elements, and the subarray $A[p \,..\, q - 1]$ has $n/2 - 1$ elements.

Because BEST-CASE-QUICKSORT always recurses on subarrays that are at most half the size of the original array, the recurrence for the worst-case running time is $T(n) \le 2T(n/2) + \Theta(n) = O(n \lg n)$.

Solution to Exercise 9.3-5

This solution is also posted publicly

We assume that are given a procedure MEDIAN that takes as parameters an array A and subarray indices p and r, and returns the value of the median element of $A[p \,..\, r]$ in $O(n)$ time in the worst case.

Given MEDIAN, here is a linear-time algorithm SELECT$'$ for finding the ith smallest element in $A[p \,..\, r]$. This algorithm uses the deterministic PARTITION algorithm that was modified to take an element to partition around as an input parameter.

```
SELECT'(A, p, r, i)
  if p == r
      return A[p]
  x = MEDIAN(A, p, r)
  q = PARTITION(x)
  k = q − p + 1
  if i == k
      return A[q]
  elseif i < k
      return SELECT'(A, p, q − 1, i)
  else return SELECT'(A, q + 1, r, i − k)
```

Because x is the median of $A[p \,..\, r]$, each of the subarrays $A[p \,..\, q - 1]$ and $A[q + 1 \,..\, r]$ has at most half the number of elements of $A[p \,..\, r]$. The recurrence for the worst-case running time of SELECT$'$ is $T(n) \le T(n/2) + O(n) = O(n)$.

Solution to Exercise 9.3-8

Let's start out by supposing that the median (the lower median, since we know we have an even number of elements) is in X. Let's call the median value m, and let's suppose that it's in $X[k]$. Then k elements of X are less than or equal to m and $n-k$ elements of X are greater than or equal to m. We know that in the two arrays combined, there must be n elements less than or equal to m and n elements greater than or equal to m, and so there must be $n-k$ elements of Y that are less than or equal to m and $n-(n-k)=k$ elements of Y that are greater than or equal to m.

Thus, we can check that $X[k]$ is the lower median by checking whether $Y[n-k] \le X[k] \le Y[n-k+1]$. A boundary case occurs for $k=n$. Then $n-k=0$, and there is no array entry $Y[0]$; we only need to check that $X[n] \le Y[1]$.

Now, if the median is in X but is not in $X[k]$, then the above condition will not hold. If the median is in $X[k']$, where $k' < k$, then $X[k]$ is above the median, and $Y[n-k+1] < X[k]$. Conversely, if the median is in $X[k'']$, where $k'' > k$, then $X[k]$ is below the median, and $X[k] < Y[n-k]$.

Thus, we can use a binary search to determine whether there is an $X[k]$ such that either $k < n$ and $Y[n-k] \le X[k] \le Y[n-k+1]$ or $k=n$ and $X[k] \le Y[n-k+1]$; if we find such an $X[k]$, then it is the median. Otherwise, we know that the median is in Y, and we use a binary search to find a $Y[k]$ such that either $k < n$ and $X[n-k] \le Y[k] \le X[n-k+1]$ or $k=n$ and $Y[k] \le X[n-k+1]$; such a $Y[k]$ is the median. Since each binary search takes $O(\lg n)$ time, we spend a total of $O(\lg n)$ time.

Here's how we write the algorithm in pseudocode:

```
TWO-ARRAY-MEDIAN(X, Y)
  n = X.length                    // n also equals Y.length
  median = FIND-MEDIAN(X, Y, n, 1, n)
  if median == NOT-FOUND
      median = FIND-MEDIAN(Y, X, n, 1, n)
  return median

FIND-MEDIAN(A, B, n, low, high)
  if low > high
      return NOT-FOUND
  else k = ⌊(low + high)/2⌋
      if k == n and A[n] ≤ B[1]
          return A[n]
      elseif k < n and B[n − k] ≤ A[k] ≤ B[n − k + 1]
          return A[k]
      elseif A[k] > B[n − k + 1]
          return FIND-MEDIAN(A, B, n, low, k − 1)
      else return FIND-MEDIAN(A, B, n, k + 1, high)
```

Solution to Exercise 9.3-9

In order to find the optimal placement for Professor Olay's pipeline, we need only find the median(s) of the y-coordinates of his oil wells, as the following proof explains.

Claim

The optimal y-coordinate for Professor Olay's east-west oil pipeline is as follows:

- If n is even, then on either the oil well whose y-coordinate is the lower median or the one whose y-coordinate is the upper median, or anywhere between them.
- If n is odd, then on the oil well whose y-coordinate is the median.

Proof We examine various cases. In each case, we will start out with the pipeline at a particular y-coordinate and see what happens when we move it. We'll denote by s the sum of the north-south spurs with the pipeline at the starting location, and s' will denote the sum after moving the pipeline.

We start with the case in which n is even. Let us start with the pipeline somewhere on or between the two oil wells whose y-coordinates are the lower and upper medians. If we move the pipeline by a vertical distance d without crossing either of the median wells, then $n/2$ of the wells become d farther from the pipeline and $n/2$ become d closer, and so $s' = s + dn/2 - dn/2 = s$; thus, all locations on or between the two medians are equally good.

Now suppose that the pipeline goes through the oil well whose y-coordinate is the upper median. What happens when we increase the y-coordinate of the pipeline by $d > 0$ units, so that it moves above the oil well that achieves the upper median? All oil wells whose y-coordinates are at or below the upper median become d units farther from the pipeline, and there are at least $n/2 + 1$ such oil wells (the upper median, and every well at or below the lower median). There are at most $n/2 - 1$ oil wells whose y-coordinates are above the upper median, and each of these oil wells becomes at most d units closer to the pipeline when it moves up. Thus, we have a lower bound on s' of $s' \geq s + d(n/2 + 1) - d(n/2 - 1) = s + 2d > s$. We conclude that moving the pipeline up from the oil well at the upper median increases the total spur length. A symmetric argument shows that if we start with the pipeline going through the oil well whose y-coordinate is the lower median and move it down, then the total spur length increases.

We see, therefore, that when n is even, an optimal placement of the pipeline is anywhere on or between the two medians.

Now we consider the case when n is odd. We start with the pipeline going through the oil well whose y-coordinate is the median, and we consider what happens when we move it up by $d > 0$ units. All oil wells at or below the median become d units farther from the pipeline, and there are at least $(n + 1)/2$ such wells (the one at the median and the $(n - 1)/2$ at or below the median. There are at most $(n - 1)/2$ oil wells above the median, and each of these becomes at most d units closer to the pipeline. We get a lower bound on s' of $s' \geq s + d(n + 1)/2 - d(n - 1)/2 = s + d > s$, and we conclude that moving the pipeline up from the oil well at the

median increases the total spur length. A symmetric argument shows that moving the pipeline down from the median also increases the total spur length, and so the optimal placement of the pipeline is on the median. ■ (claim)

Since we know we are looking for the median, we can use the linear-time median-finding algorithm.

Solution to Problem 9-1

This solution is also posted publicly

We assume that the numbers start out in an array.

a. Sort the numbers using merge sort or heapsort, which take $\Theta(n \lg n)$ worst-case time. (Don't use quicksort or insertion sort, which can take $\Theta(n^2)$ time.) Put the i largest elements (directly accessible in the sorted array) into the output array, taking $\Theta(i)$ time.

Total worst-case running time: $\Theta(n \lg n + i) = \Theta(n \lg n)$ (because $i \le n$).

b. Implement the priority queue as a heap. Build the heap using BUILD-HEAP, which takes $\Theta(n)$ time, then call HEAP-EXTRACT-MAX i times to get the i largest elements, in $\Theta(i \lg n)$ worst-case time, and store them in reverse order of extraction in the output array. The worst-case extraction time is $\Theta(i \lg n)$ because

- i extractions from a heap with $O(n)$ elements takes $i \cdot O(\lg n) = O(i \lg n)$ time, and
- half of the i extractions are from a heap with $\ge n/2$ elements, so those $i/2$ extractions take $(i/2)\Omega(\lg(n/2)) = \Omega(i \lg n)$ time in the worst case.

Total worst-case running time: $\Theta(n + i \lg n)$.

c. Use the SELECT algorithm of Section 9.3 to find the ith largest number in $\Theta(n)$ time. Partition around that number in $\Theta(n)$ time. Sort the i largest numbers in $\Theta(i \lg i)$ worst-case time (with merge sort or heapsort).

Total worst-case running time: $\Theta(n + i \lg i)$.

Note that method (c) is always asymptotically at least as good as the other two methods, and that method (b) is asymptotically at least as good as (a). (Comparing (c) to (b) is easy, but it is less obvious how to compare (c) and (b) to (a). (c) and (b) are asymptotically at least as good as (a) because n, $i \lg i$, and $i \lg n$ are all $O(n \lg n)$. The sum of two things that are $O(n \lg n)$ is also $O(n \lg n)$.)

Solution to Problem 9-2

a. The median x of the elements $x_1, x_2, \ldots, x_n$, is an element $x = x_k$ satisfying $|\{x_i : 1 \le i \le n \text{ and } x_i < x\}| \le n/2$ and $|\{x_i : 1 \le i \le n \text{ and } x_i > x\}| \le n/2$. If each element x_i is assigned a weight $w_i = 1/n$, then we get

$$\begin{aligned}
\sum_{x_i<x} w_i &= \sum_{x_i<x} \frac{1}{n} \\
&= \frac{1}{n} \cdot \sum_{x_i<x} 1 \\
&= \frac{1}{n} \cdot |\{x_i : 1 \le i \le n \text{ and } x_i < x\}| \\
&\le \frac{1}{n} \cdot \frac{n}{2} \\
&= \frac{1}{2} ,
\end{aligned}$$

and

$$\begin{aligned}
\sum_{x_i>x} w_i &= \sum_{x_i>x} \frac{1}{n} \\
&= \frac{1}{n} \cdot \sum_{x_i>x} 1 \\
&= \frac{1}{n} \cdot |\{x_i : 1 \le i \le n \text{ and } x_i > x\}| \\
&\le \frac{1}{n} \cdot \frac{n}{2} \\
&= \frac{1}{2} ,
\end{aligned}$$

which proves that x is also the weighted median of $x_1, x_2, \ldots, x_n$ with weights $w_i = 1/n$, for $i = 1, 2, \ldots, n$.

b. We first sort the n elements into increasing order by x_i values. Then we scan the array of sorted x_i's, starting with the smallest element and accumulating weights as we scan, until the total exceeds $1/2$. The last element, say x_k, whose weight caused the total to exceed $1/2$, is the weighted median. Notice that the total weight of all elements smaller than x_k is less than $1/2$, because x_k was the first element that caused the total weight to exceed $1/2$. Similarly, the total weight of all elements larger than x_k is also less than $1/2$, because the total weight of all the other elements exceeds $1/2$.

The sorting phase can be done in $O(n \lg n)$ worst-case time (using merge sort or heapsort), and the scanning phase takes $O(n)$ time. The total running time in the worst case, therefore, is $O(n \lg n)$.

c. We find the weighted median in $\Theta(n)$ worst-case time using the $\Theta(n)$ worst-case median algorithm in Section 9.3. (Although the first paragraph of the section only claims an $O(n)$ upper bound, it is easy to see that the more precise

running time of $\Theta(n)$ applies as well, since steps 1, 2, and 4 of SELECT actually take $\Theta(n)$ time.)

The weighted-median algorithm works as follows. If $n \le 2$, we just return the brute-force solution. Otherwise, we proceed as follows. We find the actual median x_k of the n elements and then partition around it. We then compute the total weights of the two halves. If the weights of the two halves are each strictly less than $1/2$, then the weighted median is x_k. Otherwise, the weighted median should be in the half with total weight exceeding $1/2$. The total weight of the "light" half is lumped into the weight of x_k, and the search continues within the half that weighs more than $1/2$. Here's pseudocode, which takes as input a set $X = \{x_1, x_2, \ldots, x_n\}$:

```
WEIGHTED-MEDIAN(X)
if n == 1
    return x_1
elseif n == 2
    if w_1 ≥ w_2
        return x_1
    else return x_2
else find the median x_k of X = {x_1, x_2, ..., x_n}
    partition the set X around x_k
    compute W_L = Σ_{x_i < x_k} w_i and W_G = Σ_{x_i > x_k} w_i
    if W_L < 1/2 and W_G < 1/2
        return x_k
    elseif W_L > 1/2
        w_k = w_k + W_G
        X' = {x_i ∈ X : x_i ≤ x_k}
        return WEIGHTED-MEDIAN(X')
    else w_k = w_k + W_L
        X' = {x_i ∈ X : x_i ≥ x_k}
        return WEIGHTED-MEDIAN(X')
```

The recurrence for the worst-case running time of WEIGHTED-MEDIAN is $T(n) = T(n/2+1) + \Theta(n)$, since there is at most one recursive call on half the number of elements, plus the median element x_k, and all the work preceding the recursive call takes $\Theta(n)$ time. The solution of the recurrence is $T(n) = \Theta(n)$.

d. Let the n points be denoted by their coordinates $x_1, x_2, \ldots, x_n$, let the corresponding weights be $w_1, w_2, \ldots, w_n$, and let $x = x_k$ be the weighted median. For any point p, let $f(p) = \sum_{i=1}^{n} w_i \, |p - x_i|$; we want to find a point p such that $f(p)$ is minimum. Let y be any point (real number) other than x. We show the optimality of the weighted median x by showing that $f(y) - f(x) \ge 0$. We examine separately the cases in which $y > x$ and $x > y$. For any x and y, we have

$$\begin{aligned} f(y) - f(x) &= \sum_{i=1}^{n} w_i \, |y - x_i| - \sum_{i=1}^{n} w_i \, |x - x_i| \\ &= \sum_{i=1}^{n} w_i (|y - x_i| - |x - x_i|) \; . \end{aligned}$$

When $y > x$, we bound the quantity $|y - x_i| - |x - x_i|$ from below by examining three cases:

1. $x < y \le x_i$: Here, $|x - y| + |y - x_i| = |x - x_i|$ and $|x - y| = y - x$, which imply that $|y - x_i| - |x - x_i| = -|x - y| = x - y$.
2. $x < x_i \le y$: Here, $|y - x_i| \ge 0$ and $|x_i - x| \le y - x$, which imply that $|y - x_i| - |x - x_i| \ge -(y - x) = x - y$.
3. $x_i \le x < y$: Here, $|x - x_i| + |y - x| = |y - x_i|$ and $|y - x| = y - x$, which imply that $|y - x_i| - |x - x_i| = |y - x| = y - x$.

Separating out the first two cases, in which $x < x_i$, from the third case, in which $x \ge x_i$, we get

$$\begin{aligned} f(y) - f(x) &= \sum_{i=1}^{n} w_i (|y - x_i| - |x - x_i|) \\ &\ge \sum_{x<x_i} w_i (x - y) + \sum_{x \ge x_i} w_i (y - x) \\ &= (y - x) \left(\sum_{x \ge x_i} w_i - \sum_{x<x_i} w_i \right) \; . \end{aligned}$$

The property that $\sum_{x_i<x} w_i < 1/2$ implies that $\sum_{x \ge x_i} w_i \ge 1/2$. This fact, combined with $y - x > 0$ and $\sum_{x<x_i} w_i \le 1/2$, yields that $f(y) - f(x) \ge 0$.

When $x > y$, we again bound the quantity $|y - x_i| - |x - x_i|$ from below by examining three cases:

1. $x_i \le y < x$: Here, $|y - x_i| + |x - y| = |x - x_i|$ and $|x - y| = x - y$, which imply that $|y - x_i| - |x - x_i| = -|x - y| = y - x$.
2. $y \le x_i < x$: Here, $|y - x_i| \ge 0$ and $|x - x_i| \le x - y$, which imply that $|y - x_i| - |x - x_i| \ge -(x - y) = y - x$.
3. $y < x \le x_i$. Here, $|x - y| + |x - x_i| = |y - x_i|$ and $|x - y| = x - y$, which imply that $|y - x_i| - |x - x_i| = |x - y| = x - y$.

Separating out the first two cases, in which $x > x_i$, from the third case, in which $x \le x_i$, we get

$$\begin{aligned} f(y) - f(x) &= \sum_{i=1}^{n} w_i (|y - x_i| - |x - x_i|) \\ &\ge \sum_{x>x_i} w_i (y - x) + \sum_{x \le x_i} w_i (x - y) \\ &= (x - y) \left(\sum_{x \le x_i} w_i - \sum_{x>x_i} w_i \right) \; . \end{aligned}$$

The property that $\sum_{x_i>x} w_i \le 1/2$ implies that $\sum_{x \le x_i} w_i > 1/2$. This fact, combined with $x - y > 0$ and $\sum_{x>x_i} w_i < 1/2$, yields that $f(y) - f(x) > 0$.

e. We are given n 2-dimensional points $p_1, p_2, \ldots, p_n$, where each p_i is a pair of real numbers $p_i = (x_i, y_i)$, and positive weights $w_1, w_2, \ldots, w_n$. The goal is to find a point $p = (x, y)$ that minimizes the sum

$$f(x, y) = \sum_{i=1}^{n} w_i \left(|x - x_i| + |y - y_i|\right) .$$

We can express the cost function of the two variables, $f(x, y)$, as the sum of two functions of one variable each: $f(x, y) = g(x) + h(y)$, where $g(x) = \sum_{i=1}^{n} w_i |x - x_i|$, and $h(y) = \sum_{i=1}^{n} w_i |y - y_i|$. The goal of finding a point $p = (x, y)$ that minimizes the value of $f(x, y)$ can be achieved by treating each dimension independently, because g does not depend on y and h does not depend on x. Thus,

$$\begin{aligned}
\min_{x,y} f(x, y) &= \min_{x,y} \left(g(x) + h(y)\right) \\
&= \min_{x} \left(\min_{y}(g(x) + h(y))\right) \\
&= \min_{x} \left(g(x) + \min_{y} h(y)\right) \\
&= \min_{x} g(x) + \min_{y} h(y) .
\end{aligned}$$

Consequently, finding the best location in 2 dimensions can be done by finding the weighted median x_k of the x-coordinates and then finding the weighted median y_j of the y-coordinates. The point (x_k, y_j) is an optimal solution for the 2-dimensional post-office location problem.

Solution to Problem 9-3

a. Our algorithm relies on a particular property of SELECT: that not only does it return the ith smallest element, but that it also partitions the input array so that the first i positions contain the i smallest elements (though not necessarily in sorted order). To see that SELECT has this property, observe that there are only two ways in which returns a value: when $n = 1$, and when immediately after partitioning in step 4, it finds that there are exactly i elements on the low side of the partition.

Taking the hint from the book, here is our modified algorithm to select the ith smallest element of n elements. Whenever it is called with $i \geq n/2$, it just calls SELECT and returns its result; in this case, $U_i(n) = T(n)$.

When $i < n/2$, our modified algorithm works as follows. Assume that the input is in a subarray $A[p + 1 \,..\, p + n]$, and let $m = \lfloor n/2 \rfloor$. In the initial call, $p = 1$.

1. Divide the input as follows. If n is even, divide the input into two parts: $A[p + 1 \,..\, p + m]$ and $A[p + m + 1 \,..\, p + n]$. If n is odd, divide the input into three parts: $A[p + 1 \,..\, p + m]$, $A[p + m + 1 \,..\, p + n - 1]$, and $A[p + n]$ as a leftover piece.
2. Compare $A[p + i]$ and $A[p + i + m]$ for $i = 1, 2, \ldots, m$, putting the smaller of the the two elements into $A[p + i + m]$ and the larger into $A[p + i]$.

3. Recursively find the ith smallest element in $A[p+m+1\,..\,p+n]$, but with an additional action performed by the partitioning procedure: whenever it exchanges $A[j]$ and $A[k]$ (where $p+m+1 \le j,k \le p+2m$), it also exchanges $A[j-m]$ and $A[k-m]$. The idea is that after recursively finding the ith smallest element in $A[p+m+1\,..\,p+n]$, the subarray $A[p+m+1\,..\,p+m+i]$ contains the i smallest elements that had been in $A[p+m+1\,..\,p+n]$ and the subarray $A[p+1\,..\,p+i]$ contains their larger counterparts, as found in step 1. The ith smallest element of $A[p+1\,..\,p+n]$ must be either one of the i smallest, as placed into $A[p+m+1\,..\,p+m+i]$, or it must be one of the larger counterparts, as placed into $A[p+1\,..\,p+i]$.
4. Collect the subarrays $A[p+1\,..\,p+i]$ and $A[p+m+1\,..\,p+m+i]$ into a single array $B[1\,..\,2i]$, call SELECT to find the ith smallest element of B, and return the result of this call to SELECT.

The number of comparisons in each step is as follows:

1. No comparisons.
2. $m = \lfloor n/2 \rfloor$ comparisons.
3. Since we recurse on $A[p+m+1\,..\,p+n]$, which has $\lceil n/2 \rceil$ elements, the number of comparisons is $U_i(\lceil n/2 \rceil)$.
4. Since we call SELECT on an array with $2i$ elements, the number of comparisons is $T(2i)$.

Thus, when $i < n/2$, the total number of comparisons is $\lfloor n/2 \rfloor + U_i(\lceil n/2 \rceil) + T(2i)$.

b. We show by substitution that if $i < n/2$, then $U_i(n) = n + O(T(2i)\lg(n/i))$. In particular, we show that $U_i(n) \le n + cT(2i)\lg(n/i) - d(\lg\lg n)T(2i) = n + cT(2i)\lg n - cT(2i)\lg i - d(\lg\lg n)T(2i)$ for some positive constant c, some positive constant d to be chosen later, and $n \ge 4$. We have

$$\begin{aligned}
U_i(n) &= \lfloor n/2 \rfloor + U_i(\lceil n/2 \rceil) + T(2i) \\
&\le \lfloor n/2 \rfloor + \lceil n/2 \rceil + cT(2i)\lg\lceil n/2 \rceil - cT(2i)\lg i \\
&\qquad - d(\lg\lg\lceil n/2 \rceil)T(2i) \\
&= n + cT(2i)\lg\lceil n/2 \rceil - cT(2i)\lg i - d(\lg\lg\lceil n/2 \rceil)T(2i) \\
&\le n + cT(2i)\lg(n/2+1) - cT(2i)\lg i - d(\lg\lg(n/2))T(2i) \\
&= n + cT(2i)\lg(n/2+1) - cT(2i)\lg i - d(\lg(\lg n - 1))T(2i) \\
&\le n + cT(2i)\lg n - cT(2i)\lg i - d(\lg\lg n)T(2i)
\end{aligned}$$

if $cT(2i)\lg(n/2+1) - d(\lg(\lg n-1))T(2i) \le cT(2i)\lg n - d(\lg\lg n)T(2i)$. Simple algebraic manipulations gives the following sequence of equivalent conditions:

$$cT(2i)\lg(n/2+1) - d(\lg(\lg n-1))T(2i) \le cT(2i)\lg n - d(\lg\lg n)T(2i)$$

$$c\lg(n/2+1) - d(\lg(\lg n-1)) \le c\lg n - d(\lg\lg n)$$

$$c(\lg(n/2+1) - \lg n) \le d(\lg(\lg n-1) - \lg\lg n)$$

$$c\left(\lg\frac{n/2+1}{n}\right) \le d\lg\frac{\lg n - 1}{\lg n}$$

$$c\left(\lg\left(\frac{1}{2}+\frac{1}{n}\right)\right) \le d\lg\frac{\lg n - 1}{\lg n}$$

Observe that $1/2+1/n$ decreases as n increases, but $(\lg n-1)/\lg n$ increases as n increases. When $n = 4$, we have $1/2+1/n = 3/4$ and $(\lg n-1)/\lg n = 1/2$. Thus, we just need to choose d such that $c\lg(3/4) \le d\lg(1/2)$ or, equivalently, $c\lg(3/4) \le -d$. Multiplying both sides by -1, we get $d \le -c\lg(3/4) = c\lg(4/3)$. Thus, any value of d that is at most $c\lg(4/3)$ suffices.

c. When i is a constant, $T(2i) = O(1)$ and $\lg(n/i) = \lg n - \lg i = O(\lg n)$. Thus, when i is a constant less than $n/2$, we have that

$$\begin{aligned} U_i(n) &= n + O(T(2i)\lg(n/i)) \\ &= n + O(O(1)\cdot O(\lg n)) \\ &= n + O(\lg n)\ . \end{aligned}$$

d. Suppose that $i = n/k$ for $k \ge 2$. Then $i \le n/2$. If $k > 2$, then $i < n/2$, and we have

$$\begin{aligned} U_i(n) &= n + O(T(2i)\lg(n/i)) \\ &= n + O(T(2n/k)\lg(n/(n/k)) \\ &= n + O(T(2n/k)\lg k)\ . \end{aligned}$$

If $k = 2$, then $n = 2i$ and $\lg k = 1$. We have

$$\begin{aligned} U_i(n) &= T(n) \\ &= n + (T(n) - n) \\ &\le n + (T(2i) - n) \\ &= n + (T(2n/k) - n) \\ &= n + (T(2n/k)\lg k - n) \\ &= n + O(T(2n/k)\lg k)\ . \end{aligned}$$

Solution to Problem 9-4

a. As in the quicksort analysis, elements z_i and z_j will not be compared with each other if any element in $\{z_{i+1}, z_{i+2}, \ldots, z_{j-1}\}$ is chosen as a pivot element before either z_i or z_j, because z_i and z_j would then lie in separate partitions. There can be another reason that z_i and z_j might not be compared, however. Suppose that $k < i$, so that $z_k < z_i$, and suppose further that the element chosen as the pivot is z_l, where $k \le l < i$. In this case, because $k \le l$, the recursion won't consider elements indexed higher than l. Therefore, the recursion will never look at z_i or z_j, and they will never be compared with each other. Similarly, if $j < k$ and the pivot element z_l is such that $j < l \le k$, then the recursion won't consider elements indexed less than l, and again z_i and z_j will never be compared with each other. The final case is when $i \le k \le j$ (but disallowing $i = j$), so that $z_i \le z_k \le z_j$; in this case, we have the same analysis as for quicksort: z_i and z_j are compared with each other only if one of them is chosen as the pivot element.

Getting back to the case in which $k < i$, it is again true that z_i and z_j are compared with each other only if one of them is chosen as the pivot element. As we know, they won't be compared with each other if the pivot element is

between them, and we argued above that they won't be compared with each other if the pivot element is z_l for $l < i$. Similarly, when $j < k$, elements z_i and z_j are compared with each other only if one of them is chosen as the pivot element.

Now we need to compute the probability that z_i and z_j are compared with each other. Let Z_{ijk} be the set of elements that includes $z_i, \ldots, z_j$, along with $z_k, \ldots, z_{i-1}$ if $k < i$ or $z_{j+1}, \ldots, z_k$ if $j < k$. In other words,

$$Z_{ijk} = \begin{cases} \{z_i, z_{i+1}, \ldots, z_j\} & \text{if } i \le k \le j \ , \\ \{z_k, z_{k+1}, \ldots, z_j\} & \text{if } k < i \ , \\ \{z_i, z_{i+1}, \ldots, z_k\} & \text{if } j < k \ . \end{cases}$$

With this definition of Z_{ijk}, we have that

$$|Z_{ijk}| = \max(j - i + 1, j - k + 1, k - i + 1) \ .$$

As in the quicksort analysis, we observe that until an element from Z_{ijk} is chosen as the pivot, the whole set Z_{ijk} is together in the same partition, and so each element of Z_{ijk} is equally likely to be the first one chosen as the pivot.

Letting C be the event that z_i is compared with z_j when finding z_k sometime during the execution of the algorithm, we have that

$$\begin{aligned} \mathrm{E}[X_{ijk}] &= \Pr\{C\} \\ &= \Pr\{z_i \text{ or } z_j \text{ is the first pivot chosen from } Z_{ijk}\} \\ &= \Pr\{z_i \text{ is the first pivot chosen from } Z_{ijk}\} \\ &\qquad + \Pr\{z_j \text{ is the first pivot chosen from } Z_{ijk}\} \\ &= \frac{1}{|Z_{ijk}|} + \frac{1}{|Z_{ijk}|} \\ &= \frac{2}{\max(j - i + 1, j - k + 1, k - i + 1)} \ . \end{aligned}$$

b. Adding up all the possible pairs that might be compared gives

$$X_k = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ijk} \ ,$$

and so, by linearity of expectation, we have

$$\begin{aligned} \mathrm{E}[X_k] &= \mathrm{E}\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ijk}\right] \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{E}[X_{ijk}] \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{\max(j - i + 1, j - k + 1, k - i + 1)} \ . \end{aligned}$$

We break this sum into the same three cases as before: $i \le k \le j$, $k < i$, and $j < k$. With k fixed, we vary i and j. We get an inequality because we cannot

have $i = k = j$, but our summation will allow it:

$$\begin{aligned}
\mathrm{E}[X_k] &\le 2\left(\sum_{i=1}^{k}\sum_{j=k}^{n}\frac{1}{j-i+1} + \sum_{j=k+1}^{n}\sum_{i=k+1}^{j-1}\frac{1}{j-k+1}\right. \\
&\qquad \left. + \sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\frac{1}{k-i+1}\right) \\
&= 2\left(\sum_{i=1}^{k}\sum_{j=k}^{n}\frac{1}{j-i+1} + \sum_{j=k+1}^{n}\frac{j-k-1}{j-k+1} + \sum_{i=1}^{k-2}\frac{k-i-1}{k-i+1}\right) .
\end{aligned}$$

c. First, let's focus on the latter two summations. Each one sums fractions that are strictly less than 1. The middle summation has $n-k$ terms, and the right-hand summation has $k-2$ terms, and so the latter two summations sum to less than n.

Now we look at the first summation. Let $m = j - i$. There is only one way for m to equal 0: if $i = k = j$. There are only two ways for m to equal 1: if $i = k-1$ and $j = k$, or if $i = k$ and $j = k+1$. There are only three ways for m to equal 2: if $i = k-2$ and $j = k$, if $i = k-1$ and $j = k+1$, or if $i = k$ and $j = k+2$. Continuing on, we see that there are at most $m+1$ ways for $j - i$ to equal m. Since $j - i \le n-1$, we can rewrite the first summation as

$$\sum_{m=0}^{n-1}\frac{m+1}{m+1} = n .$$

Thus, we have

$$\begin{aligned}
\mathrm{E}[X_k] &< 2(n+n) \\
&= 4n .
\end{aligned}$$

d. To show that RANDOMIZED-SELECT runs in expected time $O(n)$, we adapt Lemma 7.1 for RANDOMIZED-SELECT. The adaptation is trivial: just replace the variable X in the lemma statement by the random variable X_k that we just analyzed. Thus, the expected running time of RANDOMIZED-SELECT is $O(n + X_k) = O(n)$.

Lecture Notes for Chapter 11: Hash Tables

Chapter 11 overview

Many applications require a dynamic set that supports only the ***dictionary operations*** INSERT, SEARCH, and DELETE. Example: a symbol table in a compiler.

A hash table is effective for implementing a dictionary.

- The expected time to search for an element in a hash table is $O(1)$, under some reasonable assumptions.
- Worst-case search time is $\Theta(n)$, however.

A hash table is a generalization of an ordinary array.

- With an ordinary array, we store the element whose key is k in position k of the array.
- Given a key k, we find the element whose key is k by just looking in the kth position of the array. This is called ***direct addressing***.
- Direct addressing is applicable when we can afford to allocate an array with one position for every possible key.

We use a hash table when we do not want to (or cannot) allocate an array with one position per possible key.

- Use a hash table when the number of keys actually stored is small relative to the number of possible keys.
- A hash table is an array, but it typically uses a size proportional to the number of keys to be stored (rather than the number of possible keys).
- Given a key k, don't just use k as the index into the array.
- Instead, compute a function of k, and use that value to index into the array. We call this function a ***hash function***.

Issues that we'll explore in hash tables:

- How to compute hash functions. We'll look at the multiplication and division methods.
- What to do when the hash function maps multiple keys to the same table entry. We'll look at chaining and open addressing.

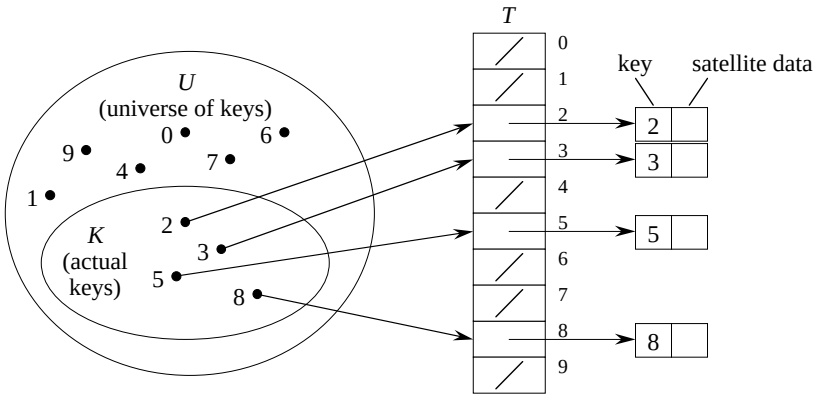
Direct-address tables

Scenario

- Maintain a dynamic set.
- Each element has a key drawn from a universe $U = \{0, 1, \ldots, m-1\}$ where m isn't too large.
- No two elements have the same key.

Represent by a ***direct-address table***, or array, $T[0 \ldots m-1]$:

- Each ***slot***, or position, corresponds to a key in U.
- If there's an element x with key k, then $T[k]$ contains a pointer to x.
- Otherwise, $T[k]$ is empty, represented by NIL.



Dictionary operations are trivial and take $O(1)$ time each:

DIRECT-ADDRESS-SEARCH(T, k)
  return $T[k]$

DIRECT-ADDRESS-INSERT(T, x)
  $T[key[x]] = x$

DIRECT-ADDRESS-DELETE(T, x)
  $T[key[x]] =$ NIL

Hash tables

The problem with direct addressing is if the universe U is large, storing a table of size $|U|$ may be impractical or impossible.

Often, the set K of keys actually stored is small, compared to U, so that most of the space allocated for T is wasted.

- When K is much smaller than U, a hash table requires much less space than a direct-address table.
- Can reduce storage requirements to $\Theta(|K|)$.
- Can still get $O(1)$ search time, but in the *average case*, not the *worst case*.

Idea

Instead of storing an element with key k in slot k, use a function h and store the element in slot $h(k)$.

- We call h a ***hash function***.
- $h : U \to \{0, 1, \ldots, m-1\}$, so that $h(k)$ is a legal slot number in T.
- We say that k ***hashes*** to slot $h(k)$.

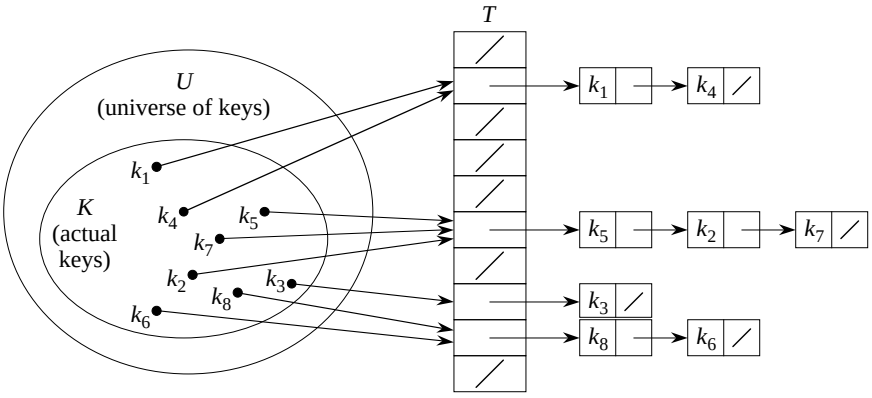
Collisions

When two or more keys hash to the same slot.

- Can happen when there are more possible keys than slots ($|U| > m$).
- For a given set K of keys with $|K| \leq m$, may or may not happen. Definitely happens if $|K| > m$.
- Therefore, must be prepared to handle collisions in all cases.
- Use two methods: chaining and open addressing.
- Chaining is usually better than open addressing. We'll examine both.

Collision resolution by chaining

Put all elements that hash to the same slot into a linked list.



[This figure shows singly linked lists. If we want to delete elements, it's better to use doubly linked lists.]

- Slot j contains a pointer to the head of the list of all stored elements that hash to j *[or to the sentinel if using a circular, doubly linked list with a sentinel]* ,
- If there are no such elements, slot j contains NIL.

How to implement dictionary operations with chaining:

- ***Insertion:***

  CHAINED-HASH-INSERT(T, x)
  insert x at the head of list $T[h(key[x])]$

  - Worst-case running time is $O(1)$.
  - Assumes that the element being inserted isn't already in the list.
  - It would take an additional search to check if it was already inserted.

- ***Search:***

  CHAINED-HASH-SEARCH(T, k)
  search for an element with key k in list $T[h(k)]$

  Running time is proportional to the length of the list of elements in slot $h(k)$.

- ***Deletion:***

  CHAINED-HASH-DELETE(T, x)
  delete x from the list $T[h(key[x])]$

  - Given pointer x to the element to delete, so no search is needed to find this element.
  - Worst-case running time is $O(1)$ time if the lists are doubly linked.
  - If the lists are singly linked, then deletion takes as long as searching, because we must find x's predecessor in its list in order to correctly update *next* pointers.

Analysis of hashing with chaining

Given a key, how long does it take to find an element with that key, or to determine that there is no element with that key?

- Analysis is in terms of the ***load factor*** $\alpha = n/m$:
  - $n =$ # of elements in the table.
  - $m =$ # of slots in the table $=$ # of (possibly empty) linked lists.
  - Load factor is average number of elements per linked list.
  - Can have $\alpha < 1$, $\alpha = 1$, or $\alpha > 1$.
- Worst case is when all n keys hash to the same slot $\Rightarrow$ get a single list of length n $\Rightarrow$ worst-case time to search is $\Theta(n)$, plus time to compute hash function.
- Average case depends on how well the hash function distributes the keys among the slots.

We focus on average-case performance of hashing with chaining.

- Assume ***simple uniform hashing***: any given element is equally likely to hash into any of the m slots.

- For $j = 0, 1, \ldots, m-1$, denote the length of list $T[j]$ by n_j. Then $n = n_0 + n_1 + \cdots + n_{m-1}$.
- Average value of n_j is $\mathrm{E}[n_j] = \alpha = n/m$.
- Assume that we can compute the hash function in $O(1)$ time, so that the time required to search for the element with key k depends on the length $n_{h(k)}$ of the list $T[h(k)]$.

We consider two cases:

- If the hash table contains no element with key k, then the search is unsuccessful.
- If the hash table does contain an element with key k, then the search is successful.

[In the theorem statements that follow, we omit the assumptions that we're resolving collisions by chaining and that simple uniform hashing applies.]

Unsuccessful search

Theorem

An unsuccessful search takes expected time $\Theta(1 + \alpha)$.

Proof Simple uniform hashing $\Rightarrow$ any key not already in the table is equally likely to hash to any of the m slots.

To search unsuccessfully for any key k, need to search to the end of the list $T[h(k)]$. This list has expected length $\mathrm{E}[n_{h(k)}] = \alpha$. Therefore, the expected number of elements examined in an unsuccessful search is α.

Adding in the time to compute the hash function, the total time required is $\Theta(1 + \alpha)$. ■

Successful search

- The expected time for a successful search is also $\Theta(1 + \alpha)$.
- The circumstances are slightly different from an unsuccessful search.
- The probability that each list is searched is proportional to the number of elements it contains.

Theorem

A successful search takes expected time $\Theta(1 + \alpha)$.

Proof Assume that the element x being searched for is equally likely to be any of the n elements stored in the table.

The number of elements examined during a successful search for x is 1 more than the number of elements that appear before x in x's list. These are the elements inserted *after* x was inserted (because we insert at the head of the list).

So we need to find the average, over the n elements x in the table, of how many elements were inserted into x's list after x was inserted.

For $i = 1, 2, \ldots, n$, let x_i be the ith element inserted into the table, and let $k_i = key[x_i]$.

For all i and j, define indicator random variable $X_{ij} = \mathrm{I}\{h(k_i) = h(k_j)\}$.

Simple uniform hashing $\Rightarrow \Pr\{h(k_i) = h(k_j)\} = 1/m \Rightarrow \mathrm{E}[X_{ij}] = 1/m$ (by Lemma 5.1).

Expected number of elements examined in a successful search is

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(1+\sum_{j=i+1}^{n}X_{ij}\right)\right] \\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\left(1+\sum_{j=i+1}^{n}\mathrm{E}[X_{ij}]\right) \quad \text{(linearity of expectation)} \\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\left(1+\sum_{j=i+1}^{n}\frac{1}{m}\right) \\
&\quad= 1+\frac{1}{nm}\sum_{i=1}^{n}(n-i) \\
&\quad= 1+\frac{1}{nm}\left(\sum_{i=1}^{n}n-\sum_{i=1}^{n}i\right) \\
&\quad= 1+\frac{1}{nm}\left(n^2-\frac{n(n+1)}{2}\right) \quad \text{(equation (A.1))} \\
&\quad= 1+\frac{n-1}{2m} \\
&\quad= 1+\frac{\alpha}{2}-\frac{\alpha}{2n}\,.
\end{aligned}$$

Adding in the time for computing the hash function, we get that the expected total time for a successful search is $\Theta(2+\alpha/2-\alpha/2n) = \Theta(1+\alpha)$.

Alternative analysis, using indicator random variables even more

For each slot l and for each pair of keys k_i and k_j, define the indicator random variable $X_{ijl} = \mathrm{I}\{\text{the search is for } x_i, h(k_i) = l, \text{ and } h(k_j) = l\}$. $X_{ijl} = 1$ when keys k_i and k_j collide at slot l and when we are searching for x_i.

Simple uniform hashing $\Rightarrow \Pr\{h(k_i) = l\} = 1/m$ and $\Pr\{h(k_j) = l\} = 1/m$. Also have $\Pr\{\text{the search is for } x_i\} = 1/n$. These events are all independent $\Rightarrow \Pr\{X_{ijl} = 1\} = 1/nm^2 \Rightarrow \mathrm{E}[X_{ijl}] = 1/nm^2$ (by Lemma 5.1).

Define, for each element x_j, the indicator random variable

$Y_j = \mathrm{I}\{x_j \text{ appears in a list prior to the element being searched for}\}$.

$Y_j = 1$ if and only if there is some slot l that has both elements x_i and x_j in its list, and also $i < j$ (so that x_i appears after x_j in the list). Therefore,

$$Y_j = \sum_{i=1}^{j-1}\sum_{l=0}^{m-1}X_{ijl}\,.$$

One final random variable: Z, which counts how many elements appear in the list prior to the element being searched for: $Z = \sum_{j=1}^{n} Y_j$. We must count the element being searched for as well as all those preceding it in its list $\Rightarrow$ compute $\mathrm{E}[Z+1]$:

$$\begin{aligned}
\mathrm{E}[Z+1] &= \mathrm{E}\left[1 + \sum_{j=1}^{n} Y_j\right] \\
&= 1 + \mathrm{E}\left[\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{l=0}^{m-1} X_{ijl}\right] \quad \text{(linearity of expectation)} \\
&= 1 + \sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{l=0}^{m-1} \mathrm{E}[X_{ijl}] \quad \text{(linearity of expectation)} \\
&= 1 + \sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{l=0}^{m-1} \frac{1}{nm^2} \\
&= 1 + \binom{n}{2} \cdot m \cdot \frac{1}{nm^2} \\
&= 1 + \frac{n(n-1)}{2} \cdot \frac{1}{nm} \\
&= 1 + \frac{n-1}{2m} \\
&= 1 + \frac{n}{2m} - \frac{1}{2m} \\
&= 1 + \frac{\alpha}{2} - \frac{\alpha}{2n} \,.
\end{aligned}$$

Adding in the time for computing the hash function, we get that the expected total time for a successful search is $\Theta(2 + \alpha/2 - \alpha/2n) = \Theta(1+\alpha)$. ■

Interpretation

If $n = O(m)$, then $\alpha = n/m = O(m)/m = O(1)$, which means that searching takes constant time on average.

Since insertion takes $O(1)$ worst-case time and deletion takes $O(1)$ worst-case time when the lists are doubly linked, all dictionary operations take $O(1)$ time on average.

Hash functions

We discuss some issues regarding hash-function design and present schemes for hash function creation.

What makes a good hash function?

- Ideally, the hash function satisfies the assumption of simple uniform hashing.

- In practice, it's not possible to satisfy this assumption, since we don't know in advance the probability distribution that keys are drawn from, and the keys may not be drawn independently.
- Often use heuristics, based on the domain of the keys, to create a hash function that performs well.

Keys as natural numbers

- Hash functions assume that the keys are natural numbers.
- When they're not, have to interpret them as natural numbers.
- ***Example:*** Interpret a character string as an integer expressed in some radix notation. Suppose the string is CLRS:
  - ASCII values: C = 67, L = 76, R = 82, S = 83.
  - There are 128 basic ASCII values.
  - So interpret CLRS as $(67 \cdot 128^3) + (76 \cdot 128^2) + (82 \cdot 128^1) + (83 \cdot 128^0) =$ 141,764,947.

Division method

$h(k) = k \bmod m$.

Example: $m = 20$ and $k = 91 \Rightarrow h(k) = 11$.

Advantage: Fast, since requires just one division operation.

Disadvantage: Have to avoid certain values of m:

- Powers of 2 are bad. If $m = 2^p$ for integer p, then $h(k)$ is just the least significant p bits of k.
- If k is a character string interpreted in radix 2^p (as in CLRS example), then $m = 2^p - 1$ is bad: permuting characters in a string does not change its hash value (Exercise 11.3-3).

Good choice for m: A prime not too close to an exact power of 2.

Multiplication method

1. Choose constant A in the range $0 < A < 1$.
2. Multiply key k by A.
3. Extract the fractional part of kA.
4. Multiply the fractional part by m.
5. Take the floor of the result.

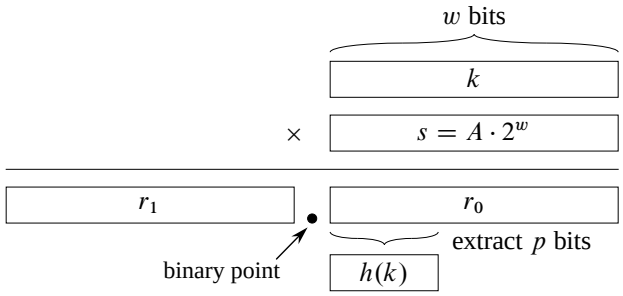
Put another way, $h(k) = \lfloor m\,(k\,A \bmod 1) \rfloor$, where $k\,A \bmod 1 = kA - \lfloor kA \rfloor =$ fractional part of kA.

Disadvantage: Slower than division method.

Advantage: Value of m is not critical.

(Relatively) easy implementation:

- Choose $m = 2^p$ for some integer p.
- Let the word size of the machine be w bits.
- Assume that k fits into a single word. (k takes w bits.)
- Let s be an integer in the range $0 < s < 2^w$. (s takes w bits.)
- Restrict A to be of the form $s/2^w$.



- Multiply k by s.
- Since we're multiplying two w-bit words, the result is $2w$ bits, $r_1 2^w + r_0$, where r_1 is the high-order word of the product and r_0 is the low-order word.
- r_1 holds the integer part of kA ($\lfloor kA \rfloor$) and r_0 holds the fractional part of kA ($k\ A \bmod 1 = kA - \lfloor kA \rfloor$). Think of the "binary point" (analog of decimal point, but for binary representation) as being between r_1 and r_0. Since we don't care about the integer part of kA, we can forget about r_1 and just use r_0.
- Since we want $\lfloor m\ (k\ A \bmod 1) \rfloor$, we could get that value by shifting r_0 to the left by $p = \lg m$ bits and then taking the p bits that were shifted to the left of the binary point.
- We don't need to shift. The p bits that would have been shifted to the left of the binary point are the p most significant bits of r_0. So we can just take these bits after having formed r_0 by multiplying k by s.
- ***Example:*** $m = 8$ (implies $p = 3$), $w = 5$, $k = 21$. Must have $0 < s < 2^5$; choose $s = 13 \Rightarrow A = 13/32$.
  - Using just the formula to compute $h(k)$: $kA = 21 \cdot 13/32 = 273/32 = 8\frac{17}{32}$ $\Rightarrow k\ A \bmod 1 = 17/32 \Rightarrow m\ (k\ A \bmod 1) = 8 \cdot 17/32 = 17/4 = 4\frac{1}{4} \Rightarrow$ $\lfloor m\ (k\ A \bmod 1) \rfloor = 4$, so that $h(k) = 4$.
  - Using the implementation: $ks = 21 \cdot 13 = 273 = 8 \cdot 2^5 + 17 \Rightarrow r_1 = 8$, $r_0 = 17$. Written in $w = 5$ bits, $r_0 = 10001$. Take the $p = 3$ most significant bits of r_0, get 100 in binary, or 4 in decimal, so that $h(k) = 4$.

How to choose A:

- The multiplication method works with any legal value of A.
- But it works better with some values than with others, depending on the keys being hashed.
- Knuth suggests using $A \approx (\sqrt{5} - 1)/2$.

Universal hashing

[We just touch on universal hashing in these notes. See the book for a full treatment.]

Suppose that a malicious adversary, who gets to choose the keys to be hashed, has seen your hashing program and knows the hash function in advance. Then he could choose keys that all hash to the same slot, giving worst-case behavior.

One way to defeat the adversary is to use a different hash function each time. You choose one at random at the beginning of your program. Unless the adversary knows how you'll be randomly choosing which hash function to use, he cannot intentionally defeat you.

Just because we choose a hash function randomly, that doesn't mean it's a good hash function. What we want is to randomly choose a single hash function from a set of good candidates.

Consider a finite collection $\mathcal{H}$ of hash functions that map a universe U of keys into the range $\{0, 1, \ldots, m-1\}$. $\mathcal{H}$ is ***universal*** if for each pair of keys $k, l \in U$, where $k \neq l$, the number of hash functions $h \in \mathcal{H}$ for which $h(k) = h(l)$ is $\leq |\mathcal{H}|/m$.

Put another way, $\mathcal{H}$ is universal if, with a hash function h chosen randomly from $\mathcal{H}$, the probability of a collision between two different keys is no more than than $1/m$ chance of just choosing two slots randomly and independently.

Why are universal hash functions good?

- They give good hashing behavior:

  Theorem

  Using chaining and universal hashing on key k:

  - If k is not in the table, the expected length $\mathrm{E}[n_{h(k)}]$ of the list that k hashes to is $\leq \alpha$.
  - If k is in the table, the expected length $\mathrm{E}[n_{h(k)}]$ of the list that holds k is $\leq 1 + \alpha$.

  Corollary

  Using chaining and universal hashing, the expected time for each SEARCH operation is $O(1)$.

- They are easy to design.

[See book for details of behavior and design of a universal class of hash functions.]

Open addressing

An alternative to chaining for handling collisions.

Idea

- Store all keys in the hash table itself.
- Each slot contains either a key or NIL.
- To search for key k:
  - Compute $h(k)$ and examine slot $h(k)$. Examining a slot is known as a ***probe***.
  - If slot $h(k)$ contains key k, the search is successful. If this slot contains NIL, the search is unsuccessful.
  - There's a third possibility: slot $h(k)$ contains a key that is not k. We compute the index of some other slot, based on k and on which probe (count from 0: 0th, 1st, 2nd, etc.) we're on.
  - Keep probing until we either find key k (successful search) or we find a slot holding NIL (unsuccessful search).
- We need the sequence of slots probed to be a permutation of the slot numbers $\langle 0, 1, \ldots, m-1 \rangle$ (so that we examine all slots if we have to, and so that we don't examine any slot more than once).
- Thus, the hash function is $h : U \times \underbrace{\{0, 1, \ldots, m-1\}}_{\text{probe number}} \to \underbrace{\{0, 1, \ldots, m-1\}}_{\text{slot number}}$.
- The requirement that the sequence of slots be a permutation of $\langle 0, 1, \ldots, m-1 \rangle$ is equivalent to requiring that the ***probe sequence*** $\langle h(k, 0), h(k, 1), \ldots, h(k, m-1) \rangle$ be a permutation of $\langle 0, 1, \ldots, m-1 \rangle$.
- To insert, act as though we're searching, and insert at the first NIL slot we find.

Pseudocode for searching

HASH-SEARCH(T, k)

```
i = 0
repeat
    j = h(k, i)
    if T[j] == k
        return j
    i = i + 1
until T[j] == NIL or i = m
return NIL
```

HASH-SEARCH returns the index of a slot containing key k, or NIL if the search is unsuccessful.

Pseudocode for insertion

```
HASH-INSERT(T, k)
 i = 0
 repeat
     j = h(k, i)
     if T[j] == NIL
         T[j] = k
         return j
     else i = i + 1
 until i == m
 error "hash table overflow"
```

HASH-INSERT returns the number of the slot that gets key k, or it flags a "hash table overflow" error if there is no empty slot in which to put key k.

Deletion

Cannot just put NIL into the slot containing the key we want to delete.

- Suppose we want to delete key k in slot j.
- And suppose that sometime after inserting key k, we were inserting key k', and during this insertion we had probed slot j (which contained key k).
- And suppose we then deleted key k by storing NIL into slot j.
- And then we search for key k'.
- During the search, we would probe slot j *before* probing the slot into which key k' was eventually stored.
- Thus, the search would be unsuccessful, even though key k' is in the table.

Solution: Use a special value DELETED instead of NIL when marking a slot as empty during deletion.

- Search should treat DELETED as though the slot holds a key that does not match the one being searched for.
- Insertion should treat DELETED as though the slot were empty, so that it can be reused.

The disadvantage of using DELETED is that now search time is no longer dependent on the load factor α.

How to compute probe sequences

The ideal situation is ***uniform hashing***: each key is equally likely to have any of the $m!$ permutations of $\langle 0, 1, \ldots, m-1 \rangle$ as its probe sequence. (This generalizes simple uniform hashing for a hash function that produces a whole probe sequence rather than just a single number.)

It's hard to implement true uniform hashing, so we approximate it with techniques that at least guarantee that the probe sequence is a permutation of $\langle 0, 1, \ldots, m-1 \rangle$.

None of these techniques can produce all $m!$ probe sequences. They will make use of ***auxiliary hash functions***, which map $U \to \{0, 1, \ldots, m-1\}$.

Linear probing

Given auxiliary hash function h', the probe sequence starts at slot $h'(k)$ and continues sequentially through the table, wrapping after slot $m-1$ to slot 0.

Given key k and probe number i ($0 \le i < m$), $h(k,i) = (h'(k) + i) \bmod m$.

The initial probe determines the entire sequence $\Rightarrow$ only m possible sequences.

Linear probing suffers from ***primary clustering***: long runs of occupied sequences build up. And long runs tend to get longer, since an empty slot preceded by i full slots gets filled next with probability $(i+1)/m$. Result is that the average search and insertion times increase.

Quadratic probing

As in linear probing, the probe sequence starts at $h'(k)$. Unlike linear probing, it jumps around in the table according to a quadratic function of the probe number: $h(k,i) = (h'(k) + c_1 i + c_2 i^2) \bmod m$, where $c_1, c_2 \ne 0$ are constants.

Must constrain c_1, c_2, and m in order to ensure that we get a full permutation of $\langle 0, 1, \ldots, m-1 \rangle$. (Problem 11-3 explores one way to implement quadratic probing.)

Can get ***secondary clustering***: if two distinct keys have the same h' value, then they have the same probe sequence.

Double hashing

Use two auxiliary hash functions, h_1 and h_2. h_1 gives the initial probe, and h_2 gives the remaining probes: $h(k,i) = (h_1(k) + i h_2(k)) \bmod m$.

Must have $h_2(k)$ be relatively prime to m (no factors in common other than 1) in order to guarantee that the probe sequence is a full permutation of $\langle 0, 1, \ldots, m-1 \rangle$.

- Could choose m to be a power of 2 and h_2 to always produce an odd number > 1.
- Could let m be prime and have $1 < h_2(k) < m$.

$\Theta(m^2)$ different probe sequences, since each possible combination of $h_1(k)$ and $h_2(k)$ gives a different probe sequence.

Analysis of open-address hashing

Assumptions

- Analysis is in terms of load factor α. We will assume that the table never completely fills, so we always have $0 \le n < m \Rightarrow 0 \le \alpha < 1$.
- Assume uniform hashing.
- No deletion.
- In a successful search, each key is equally likely to be searched for.

Theorem

The expected number of probes in an unsuccessful search is at most $1/(1-\alpha)$.

Proof Since the search is unsuccessful, every probe is to an occupied slot, except for the last probe, which is to an empty slot.

Define random variable $X = \#$ of probes made in an unsuccessful search.

Define events A_i, for $i = 1, 2, \ldots$, to be the event that there is an ith probe and that it's to an occupied slot.

$X \geq i$ if and only if probes $1, 2, \ldots, i-1$ are made and are to occupied slots $\Rightarrow$ $\Pr\{X \geq i\} = \Pr\{A_1 \cap A_2 \cap \cdots \cap A_{i-1}\}$.

By Exercise C.2-5,

$$\begin{aligned}\Pr\{A_1 \cap A_2 \cap \cdots \cap A_{i-1}\} &= \Pr\{A_1\} \cdot \Pr\{A_2 \mid A_1\} \cdot \Pr\{A_3 \mid A_1 \cap A_2\} \cdots \\ &\quad \Pr\{A_{i-1} \mid A_1 \cap A_2 \cap \cdots \cap A_{i-2}\} \ .\end{aligned}$$

Claim

$\Pr\{A_j \mid A_1 \cap A_2 \cap \cdots \cap A_{j-1}\} = (n-j+1)/(m-j+1)$. Boundary case: $j = 1$ $\Rightarrow \Pr\{A_1\} = n/m$.

Proof For the boundary case $j = 1$, there are n stored keys and m slots, so the probability that the first probe is to an occupied slot is n/m.

Given that $j-1$ probes were made, all to occupied slots, the assumption of uniform hashing says that the probe sequence is a permutation of $\langle 0, 1, \ldots, m-1 \rangle$, which in turn implies that the next probe is to a slot that we have not yet probed. There are $m - j + 1$ slots remaining, $n - j + 1$ of which are occupied. Thus, the probability that the jth probe is to an occupied slot is $(n - j + 1)/(m - j + 1)$. ■ (claim)

Using this claim,

$$\Pr\{X \geq i\} = \underbrace{\frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \frac{n-2}{m-2} \cdots \frac{n-i+2}{m-i+2}}_{i-1 \text{ factors}} \ .$$

$n < m \Rightarrow (n-j)/(m-j) \leq n/m$ for $j \geq 0$, which implies

$$\begin{aligned}\Pr\{X \geq i\} &\leq \left(\frac{n}{m}\right)^{i-1} \\ &= \alpha^{i-1} \ .\end{aligned}$$

By equation (C.25),

$$\begin{aligned}\mathrm{E}[X] &= \sum_{i=1}^{\infty} \Pr\{X \geq i\} \\ &\leq \sum_{i=1}^{\infty} \alpha^{i-1} \\ &= \sum_{i=0}^{\infty} \alpha^{i} \\ &= \frac{1}{1-\alpha} \qquad \text{(equation (A.6))} \ .\end{aligned}$$

■ (theorem)

Interpretation

If α is constant, an unsuccessful search takes $O(1)$ time.

- If $\alpha = 0.5$, then an unsuccessful search takes an average of $1/(1-0.5) = 2$ probes.
- If $\alpha = 0.9$, takes an average of $1/(1-0.9) = 10$ probes.

Corollary

The expected number of probes to insert is at most $1/(1-\alpha)$.

Proof Since there is no deletion, insertion uses the same probe sequence as an unsuccessful search. ■

Theorem

The expected number of probes in a successful search is at most $\frac{1}{\alpha} \ln \frac{1}{1-\alpha}$.

Proof A successful search for key k follows the same probe sequence as when key k was inserted.

By the previous corollary, if k was the $(i+1)$st key inserted, then α equaled i/m at the time. Thus, the expected number of probes made in a search for k is at most $1/(1-i/m) = m/(m-i)$.

That was assuming that k was the $(i+1)$st key inserted. We need to average over all n keys:

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1}\frac{m}{m-i} &= \frac{m}{n}\sum_{i=0}^{n-1}\frac{1}{m-i} \\
&= \frac{1}{\alpha}\sum_{k=m-n+1}^{m}\frac{1}{k} \\
&\le \frac{1}{\alpha}\int_{m-n}^{m}(1/x)\,dx \qquad \text{(by inequality (A.12))} \\
&= \frac{1}{\alpha}\ln\frac{m}{m-n} \\
&= \frac{1}{\alpha}\ln\frac{1}{1-\alpha}
\end{aligned}$$

■ (theorem)

Solutions for Chapter 11: Hash Tables

Solution to Exercise 11.1-4

We denote the huge array by T and, taking the hint from the book, we also have a stack implemented by an array S. The size of S equals the number of keys actually stored, so that S should be allocated at the dictionary's maximum size. The stack has an attribute $S.top$, so that only entries $S[1\,..\,S.top]$ are valid.

The idea of this scheme is that entries of T and S validate each other. If key k is actually stored in T, then $T[k]$ contains the index, say j, of a valid entry in S, and $S[j]$ contains the value k. Let us call this situation, in which $1 \le T[k] \le S.top$, $S[T[k]] = k$, and $T[S[j]] = j$, a ***validating cycle***.

Assuming that we also need to store pointers to objects in our direct-address table, we can store them in an array that is parallel to either T or S. Since S is smaller than T, we'll use an array S', allocated to be the same size as S, for these pointers. Thus, if the dictionary contains an object x with key k, then there is a validating cycle and $S'[T[k]]$ points to x.

The operations on the dictionary work as follows:

- Initialization: Simply set $S.top = 0$, so that there are no valid entries in the stack.
- SEARCH: Given key k, we check whether we have a validating cycle, i.e., whether $1 \le T[k] \le S.top$ and $S[T[k]] = k$. If so, we return $S'[T[k]]$, and otherwise we return NIL.
- INSERT: To insert object x with key k, assuming that this object is not already in the dictionary, we increment $S.top$, set $S[S.top] = k$, set $S'[S.top] = x$, and set $T[k] = S.top$.
- DELETE: To delete object x with key k, assuming that this object is in the dictionary, we need to break the validating cycle. The trick is to also ensure that we don't leave a "hole" in the stack, and we solve this problem by moving the top entry of the stack into the position that we are vacating—and then fixing up *that* entry's validating cycle. That is, we execute the following sequence of assignments:

$S[T[k]] = S[S.top]$
$S'[T[k]] = S'[S.top]$
$T[S[T[k]]] = T[k]$
$T[k] = 0$
$S.top = S.top - 1$

Each of these operations—initialization, SEARCH, INSERT, and DELETE—takes $O(1)$ time.

Solution to Exercise 11.2-1

This solution is also posted publicly

For each pair of keys k, l, where $k \neq l$, define the indicator random variable $X_{kl} = \mathrm{I}\{h(k) = h(l)\}$. Since we assume simple uniform hashing, $\Pr\{X_{kl} = 1\} = \Pr\{h(k) = h(l)\} = 1/m$, and so $\mathrm{E}[X_{kl}] = 1/m$.

Now define the random variable Y to be the total number of collisions, so that $Y = \sum_{k \neq l} X_{kl}$. The expected number of collisions is

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}\left[\sum_{k \neq l} X_{kl}\right] \\
&= \sum_{k \neq l} \mathrm{E}[X_{kl}] \qquad \text{(linearity of expectation)} \\
&= \binom{n}{2}\frac{1}{m} \\
&= \frac{n(n-1)}{2} \cdot \frac{1}{m} \\
&= \frac{n(n-1)}{2m} .
\end{aligned}$$

Solution to Exercise 11.2-4

This solution is also posted publicly

The flag in each slot will indicate whether the slot is free.

- A free slot is in the free list, a doubly linked list of all free slots in the table. The slot thus contains two pointers.
- A used slot contains an element and a pointer (possibly NIL) to the next element that hashes to this slot. (Of course, that pointer points to another slot in the table.)

Operations

- ***Insertion:***
  - If the element hashes to a free slot, just remove the slot from the free list and store the element there (with a NIL pointer). The free list must be doubly linked in order for this deletion to run in $O(1)$ time.
  - If the element hashes to a used slot j, check whether the element x already there "belongs" there (its key also hashes to slot j).
    - If so, add the new element to the chain of elements in this slot. To do so, allocate a free slot (e.g., take the head of the free list) for the new element and put this new slot at the head of the list pointed to by the hashed-to slot (j).
    - If not, E is part of another slot's chain. Move it to a new slot by allocating one from the free list, copying the old slot's (j's) contents (element x and pointer) to the new slot, and updating the pointer in the slot that pointed to j to point to the new slot. Then insert the new element in the now-empty slot as usual.
      To update the pointer to j, it is necessary to find it by searching the chain of elements starting in the slot x hashes to.
- ***Deletion:*** Let j be the slot the element x to be deleted hashes to.
  - If x is the only element in j (j doesn't point to any other entries), just free the slot, returning it to the head of the free list.
  - If x is in j but there's a pointer to a chain of other elements, move the first pointed-to entry to slot j and free the slot it was in.
  - If x is found by following a pointer from j, just free x's slot and splice it out of the chain (i.e., update the slot that pointed to x to point to x's successor).
- ***Searching:*** Check the slot the key hashes to, and if that is not the desired element, follow the chain of pointers from the slot.

  All the operations take expected $O(1)$ times for the same reason they do with the version in the book: The expected time to search the chains is $O(1 + \alpha)$ regardless of where the chains are stored, and the fact that all the elements are stored in the table means that $\alpha \leq 1$. If the free list were singly linked, then operations that involved removing an arbitrary slot from the free list would not run in $O(1)$ time.

Solution to Exercise 11.2-6

We can view the hash table as if it had m rows and L columns; each row stores one chain. The array has mL entries storing n keys, and $mL - n$ empty values. The procedure picks array positions at random until it finds a key, which it returns. The probability of success on one draw is n/mL, so $mL/n = L/\alpha$ trials are needed. Each trial takes time $O(1)$, since the individual chain sizes are known. The chain for the last draw needs to be scanned to find the desired element, however, costing $O(L)$.

Solution to Exercise 11.3-3

First, we observe that we can generate any permutation by a sequence of interchanges of pairs of characters. One can prove this property formally, but informally, consider that both heapsort and quicksort work by interchanging pairs of elements and that they have to be able to produce any permutation of their input array. Thus, it suffices to show that if string x can be derived from string y by interchanging a single pair of characters, then x and y hash to the same value.

Let us denote the ith character in x by x_i, and similarly for y. The interpretation of x in radix 2^p is $\sum_{i=0}^{n-1} x_i 2^{ip}$, and so $h(x) = \left(\sum_{i=0}^{n-1} x_i 2^{ip}\right) \bmod (2^p - 1)$. Similarly, $h(y) = \left(\sum_{i=0}^{n-1} y_i 2^{ip}\right) \bmod (2^p - 1)$.

Suppose that x and y are identical strings of n characters except that the characters in positions a and b are interchanged: $x_a = y_b$ and $y_a = x_b$. Without loss of generality, let $a > b$. We have

$$h(x) - h(y) = \left(\sum_{i=0}^{n-1} x_i 2^{ip}\right) \bmod (2^p - 1) - \left(\sum_{i=0}^{n-1} y_i 2^{ip}\right) \bmod (2^p - 1) \; .$$

Since $0 \le h(x), h(y) < 2^p - 1$, we have that $-(2^p - 1) < h(x) - h(y) < 2^p - 1$. If we show that $(h(x) - h(y)) \bmod (2^p - 1) = 0$, then $h(x) = h(y)$.

Since the sums in the hash functions are the same except for indices a and b, we have

$$\begin{aligned}
&(h(x) - h(y)) \bmod (2^p - 1) \\
&\quad = ((x_a 2^{ap} + x_b 2^{bp}) - (y_a 2^{ap} + y_b 2^{bp})) \bmod (2^p - 1) \\
&\quad = ((x_a 2^{ap} + x_b 2^{bp}) - (x_b 2^{ap} + x_a 2^{bp})) \bmod (2^p - 1) \\
&\quad = ((x_a - x_b) 2^{ap} - (x_a - x_b) 2^{bp}) \bmod (2^p - 1) \\
&\quad = ((x_a - x_b)(2^{ap} - 2^{bp})) \bmod (2^p - 1) \\
&\quad = ((x_a - x_b) 2^{bp} (2^{(a-b)p} - 1)) \bmod (2^p - 1) \; .
\end{aligned}$$

By equation (A.5),

$$\sum_{i=0}^{a-b-1} 2^{pi} = \frac{2^{(a-b)p} - 1}{2^p - 1} \; ,$$

and multiplying both sides by $2^p - 1$, we get $2^{(a-b)p} - 1 = \left(\sum_{i=0}^{a-b-1} 2^{pi}\right)(2^p - 1)$. Thus,

$$\begin{aligned}
&(h(x) - h(y)) \bmod (2^p - 1) \\
&\quad = \left((x_a - x_b) 2^{bp} \left(\sum_{i=0}^{a-b-1} 2^{pi}\right)(2^p - 1)\right) \bmod (2^p - 1) \\
&\quad = 0 \; ,
\end{aligned}$$

since one of the factors is $2^p - 1$.

We have shown that $(h(x) - h(y)) \bmod (2^p - 1) = 0$, and so $h(x) = h(y)$.

Solution to Exercise 11.3-5

Let $b = |B|$ and $u = |U|$. We start by showing that the total number of collisions is minimized by a hash function that maps u/b elements of U to each of the b values in B. For a given hash function, let u_j be the number of elements that map to $j \in B$. We have $u = \sum_{j \in B} u_j$. We also have that the number of collisions for a given value of $j \in B$ is $\binom{u_j}{2} = u_j(u_j - 1)/2$.

Lemma

The total number of collisions is minimized when $u_j = u/b$ for each $j \in B$.

Proof If $u_j \le u/b$, let us call j ***underloaded***, and if $u_j \ge u/b$, let us call j ***overloaded***. Consider an unbalanced situation in which $u_j \ne u/b$ for at least one value $j \in B$. We can think of converting a balanced situation in which all u_j equal u/b into the unbalanced situation by repeatedly moving an element that maps to an underloaded value to map instead to an overloaded value. (If you think of the values of B as representing buckets, we are repeatedly moving elements from buckets containing at most u/b elements to buckets containing at least u/b elements.)

We now show that each such move increases the number of collisions, so that all the moves together must increase the number of collisions. Suppose that we move an element from an underloaded value j to an overloaded value k, and we leave all other elements alone. Because j is underloaded and k is overloaded, $u_j \le u/b \le u_k$. Considering just the collisions for values j and k, we have $u_j(u_j - 1)/2 + u_k(u_k - 1)/2$ collisions before the move and $(u_j - 1)(u_j - 2)/2 + (u_k + 1)u_k/2$ collisions afterward. We wish to show that $u_j(u_j - 1)/2 + u_k(u_k - 1)/2 < (u_j - 1)(u_j - 2)/2 + (u_k + 1)u_k/2$. We have the following sequence of equivalent inequalities:

$$\begin{aligned}
u_j &< u_k + 1 \\
2u_j &< 2u_k + 2 \\
-u_k &< u_k - 2u_j + 2 \\
u_j^2 - u_j + u_k^2 - u_k &< u_j^2 - 3u_j + 2 + u_k^2 + u_k \\
u_j(u_j - 1) + u_k(u_k - 1) &< (u_j - 1)(u_j - 2) + (u_k + 1)u_k \\
u_j(u_j - 1)/2 + u_k(u_k - 1)/2 &< (u_j - 1)(u_j - 2)/2 + (u_k + 1)u_k/2 \,.
\end{aligned}$$

Thus, each move increases the number of collisions. We conclude that the number of collisions is minimized when $u_j = u/b$ for each $j \in B$. ■

By the above lemma, for any hash function, the total number of collisions must be at least $b(u/b)(u/b - 1)/2$. The number of pairs of distinct elements is $\binom{u}{2} = u(u-1)/2$. Thus, the number of collisions per pair of distinct elements must be at least

$$\begin{aligned}\frac{b(u/b)(u/b-1)/2}{u(u-1)/2} &= \frac{u/b-1}{u-1}\\ &> \frac{u/b-1}{u}\\ &= \frac{1}{b}-\frac{1}{u}\,.\end{aligned}$$

Thus, the bound ϵ on the probability of a collision for any pair of distinct elements can be no less than $1/b - 1/u = 1/|B| - 1/|U|$.

Solution to Problem 11-1

a. Since we assume uniform hashing, we can use the same observation as is used in Corollary 11.7: that inserting a key entails an unsuccessful search followed by placing the key into the first empty slot found. As in the proof of Theorem 11.6, if we let X be the random variable denoting the number of probes in an unsuccessful search, then $\Pr\{X \geq i\} \leq \alpha^{i-1}$. Since $n \leq m/2$, we have $\alpha \leq 1/2$. Letting $i = k+1$, we have $\Pr\{X > k\} = \Pr\{X \geq k+1\} \leq (1/2)^{(k+1)-1} = 2^{-k}$.

b. Substituting $k = 2\lg n$ into the statement of part (a) yields that the probability that the ith insertion requires more than $k = 2\lg n$ probes is at most $2^{-2\lg n} = (2^{\lg n})^{-2} = n^{-2} = 1/n^2$.

We must deal with the possibility that $2\lg n$ is not an integer, however. Then the event that the ith insertion requires more than $2\lg n$ probes is the same as the event that the ith insertion requires more than $\lfloor 2\lg n \rfloor$ probes. Since $\lfloor 2\lg n \rfloor > 2\lg n - 1$, we have that the probability of this event is at most $2^{-\lfloor 2\lg n \rfloor} < 2^{-(2\lg n - 1)} = 2/n^2 = O(1/n^2)$.

c. Let the event A be $X > 2\lg n$, and for $i = 1, 2, \ldots, n$, let the event A_i be $X_i > 2\lg n$. In part (b), we showed that $\Pr\{A_i\} = O(1/n^2)$ for $i = 1, 2, \ldots, n$. From how we defined these events, $A = A_1 \cup A_2 \cup \cdots \cup A_n$. Using Boole's inequality, (C.19), we have

$$\begin{aligned}\Pr\{A\} &\leq \Pr\{A_1\} + \Pr\{A_2\} + \cdots + \Pr\{A_n\}\\ &\leq n \cdot O(1/n^2)\\ &= O(1/n)\,.\end{aligned}$$

d. We use the definition of expectation and break the sum into two parts:

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=1}^{n} k \cdot \Pr\{X = k\} \\
&= \sum_{k=1}^{\lceil 2\lg n \rceil} k \cdot \Pr\{X = k\} + \sum_{k=\lceil 2\lg n \rceil + 1}^{n} k \cdot \Pr\{X = k\} \\
&\le \sum_{k=1}^{\lceil 2\lg n \rceil} \lceil 2\lg n \rceil \cdot \Pr\{X = k\} + \sum_{k=\lceil 2\lg n \rceil + 1}^{n} n \cdot \Pr\{X = k\} \\
&= \lceil 2\lg n \rceil \sum_{k=1}^{\lceil 2\lg n \rceil} \Pr\{X = k\} + n \sum_{k=\lceil 2\lg n \rceil + 1}^{n} \Pr\{X = k\} \ .
\end{aligned}$$

Since X takes on exactly one value, we have that $\sum_{k=1}^{\lceil 2\lg n \rceil} \Pr\{X = k\} = \Pr\{X \le \lceil 2\lg n \rceil\} \le 1$ and $\sum_{k=\lceil 2\lg n \rceil + 1}^{n} \Pr\{X = k\} \le \Pr\{X > 2\lg n\} = O(1/n)$, by part (c). Therefore,

$$\begin{aligned}
\mathrm{E}[X] &\le \lceil 2\lg n \rceil \cdot 1 + n \cdot O(1/n) \\
&= \lceil 2\lg n \rceil + O(1) \\
&= O(\lg n) \ .
\end{aligned}$$

Solution to Problem 11-2

This solution is also posted publicly

a. A particular key is hashed to a particular slot with probability $1/n$. Suppose we select a specific set of k keys. The probability that these k keys are inserted into the slot in question and that all other keys are inserted elsewhere is

$$\left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \ .$$

Since there are $\binom{n}{k}$ ways to choose our k keys, we get

$$Q_k = \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \binom{n}{k} \ .$$

b. For $i = 1, 2, \ldots, n$, let X_i be a random variable denoting the number of keys that hash to slot i, and let A_i be the event that $X_i = k$, i.e., that exactly k keys hash to slot i. From part (a), we have $\Pr\{A\} = Q_k$. Then,

$$\begin{aligned}
P_k &= \Pr\{M = k\} \\
&= \Pr\left\{\left(\max_{1 \le i \le n} X_i\right) = k\right\} \\
&= \Pr\{\text{there exists } i \text{ such that } X_i = k \text{ and that } X_i \le k \text{ for } i = 1, 2, \ldots, n\} \\
&\le \Pr\{\text{there exists } i \text{ such that } X_i = k\} \\
&= \Pr\{A_1 \cup A_2 \cup \cdots \cup A_n\} \\
&\le \Pr\{A_1\} + \Pr\{A_2\} + \cdots + \Pr\{A_n\} \qquad \text{(by inequality (C.19))} \\
&= nQ_k \ .
\end{aligned}$$

c. We start by showing two facts. First, $1 - 1/n < 1$, which implies $(1 - 1/n)^{n-k} < 1$. Second, $n!/(n-k)! = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) < n^k$. Using these facts, along with the simplification $k! > (k/e)^k$ of equation (3.18), we have

$$\begin{aligned} Q_k &= \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \frac{n!}{k!(n-k)!} \\ &< \frac{n!}{n^k k!(n-k)!} && ((1 - 1/n)^{n-k} < 1) \\ &< \frac{1}{k!} && (n!/(n-k)! < n^k) \\ &< \frac{e^k}{k^k} && (k! > (k/e)^k) \ . \end{aligned}$$

d. Notice that when $n = 2$, $\lg \lg n = 0$, so to be precise, we need to assume that $n \geq 3$.

In part (c), we showed that $Q_k < e^k / k^k$ for any k; in particular, this inequality holds for k_0. Thus, it suffices to show that $e^{k_0}/{k_0}^{k_0} < 1/n^3$ or, equivalently, that $n^3 < {k_0}^{k_0}/e^{k_0}$.

Taking logarithms of both sides gives an equivalent condition:

$$\begin{aligned} 3 \lg n &< k_0 (\lg k_0 - \lg e) \\ &= \frac{c \lg n}{\lg \lg n} (\lg c + \lg \lg n - \lg \lg \lg n - \lg e) \ . \end{aligned}$$

Dividing both sides by $\lg n$ gives the condition

$$\begin{aligned} 3 &< \frac{c}{\lg \lg n} (\lg c + \lg \lg n - \lg \lg \lg n - \lg e) \\ &= c \left(1 + \frac{\lg c - \lg e}{\lg \lg n} - \frac{\lg \lg \lg n}{\lg \lg n}\right) \ . \end{aligned}$$

Let x be the last expression in parentheses:

$$x = \left(1 + \frac{\lg c - \lg e}{\lg \lg n} - \frac{\lg \lg \lg n}{\lg \lg n}\right) \ .$$

We need to show that there exists a constant $c > 1$ such that $3 < cx$.

Noting that $\lim_{n \to \infty} x = 1$, we see that there exists n_0 such that $x \geq 1/2$ for all $n \geq n_0$. Thus, any constant $c > 6$ works for $n \geq n_0$.

We handle smaller values of n—in particular, $3 \leq n < n_0$—as follows. Since n is constrained to be an integer, there are a finite number of n in the range $3 \leq n < n_0$. We can evaluate the expression x for each such value of n and determine a value of c for which $3 < cx$ for all values of n. The final value of c that we use is the larger of

- 6, which works for all $n \geq n_0$, and
- $\max_{3 \leq n < n_0} \{c : 3 < cx\}$, i.e., the largest value of c that we chose for the range $3 \leq n < n_0$.

Thus, we have shown that $Q_{k_0} < 1/n^3$, as desired.

To see that $P_k < 1/n^2$ for $k \geq k_0$, we observe that by part (b), $P_k \leq n Q_k$ for all k. Choosing $k = k_0$ gives $P_{k_0} \leq n Q_{k_0} < n \cdot (1/n^3) = 1/n^2$. For

$k > k_0$, we will show that we can pick the constant c such that $Q_k < 1/n^3$ for all $k \geq k_0$, and thus conclude that $P_k < 1/n^2$ for all $k \geq k_0$.

To pick c as required, we let c be large enough that $k_0 > 3 > e$. Then $e/k < 1$ for all $k \geq k_0$, and so e^k/k^k decreases as k increases. Thus,

$$\begin{aligned} Q_k &< e^k/k^k \\ &\leq e^{k_0}/k^{k_0} \\ &< 1/n^3 \end{aligned}$$

for $k \geq k_0$.

e. The expectation of M is

$$\begin{aligned} \mathrm{E}[M] &= \sum_{k=0}^{n} k \cdot \Pr\{M = k\} \\ &= \sum_{k=0}^{k_0} k \cdot \Pr\{M = k\} + \sum_{k=k_0+1}^{n} k \cdot \Pr\{M = k\} \\ &\leq \sum_{k=0}^{k_0} k_0 \cdot \Pr\{M = k\} + \sum_{k=k_0+1}^{n} n \cdot \Pr\{M = k\} \\ &\leq k_0 \sum_{k=0}^{k_0} \Pr\{M = k\} + n \sum_{k=k_0+1}^{n} \Pr\{M = k\} \\ &= k_0 \cdot \Pr\{M \leq k_0\} + n \cdot \Pr\{M > k_0\} \ , \end{aligned}$$

which is what we needed to show, since $k_0 = c \lg n / \lg \lg n$.

To show that $\mathrm{E}[M] = O(\lg n / \lg \lg n)$, note that $\Pr\{M \leq k_0\} \leq 1$ and

$$\begin{aligned} \Pr\{M > k_0\} &= \sum_{k=k_0+1}^{n} \Pr\{M = k\} \\ &= \sum_{k=k_0+1}^{n} P_k \\ &< \sum_{k=k_0+1}^{n} 1/n^2 \qquad \text{(by part (d))} \\ &< n \cdot (1/n^2) \\ &= 1/n \ . \end{aligned}$$

We conclude that

$$\begin{aligned} \mathrm{E}[M] &\leq k_0 \cdot 1 + n \cdot (1/n) \\ &= k_0 + 1 \\ &= O(\lg n / \lg \lg n) \ . \end{aligned}$$

Solution to Problem 11-3

a. From how the probe-sequence computation is specified, it is easy to see that the probe sequence is $\langle h(k), h(k) + 1, h(k) + 1 + 2, h(k) + 1 + 2 + 3,$

$\ldots, h(k) + 1 + 2 + 3 + \cdots + i, \ldots\rangle$, where all the arithmetic is modulo m. Starting the probe numbers from 0, the ith probe is offset (modulo m) from $h(k)$ by

$$\sum_{j=0}^{i} j = \frac{i(i+1)}{2} = \frac{1}{2}\,i^2 + \frac{1}{2}\,i\ .$$

Thus, we can write the probe sequence as

$$h'(k,i) = \left(h(k) + \frac{1}{2}\,i + \frac{1}{2}\,i^2\right) \bmod m\ ,$$

which demonstrates that this scheme is a special case of quadratic probing.

b. Let $h'(k,i)$ denote the ith probe of our scheme. We saw in part (a) that $h'(k,i) = (h(k) + i(i+1)/2) \bmod m$. To show that our algorithm examines every table position in the worst case, we show that for a given key, each of the first m probes hashes to a distinct value. That is, for any key k and for any probe numbers i and j such that $0 \le i < j < m$, we have $h'(k,i) \ne h'(k,j)$. We do so by showing that $h'(k,i) = h'(k,j)$ yields a contradiction.

Let us assume that there exists a key k and probe numbers i and j satsifying $0 \le i < j < m$ for which $h'(k,i) = h'(k,j)$. Then

$$h(k) + i(i+1)/2 \equiv h(k) + j(j+1)/2 \pmod m\ ,$$

which in turn implies that

$$i(i+1)/2 \equiv j(j+1)/2 \pmod m\ ,$$

or

$$j(j+1)/2 - i(i+1)/2 \equiv 0 \pmod m\ .$$

Since $j(j+1)/2 - i(i+1)/2 = (j-i)(j+i+1)/2$, we have

$$(j-i)(j+i+1)/2 \equiv 0 \pmod m\ .$$

The factors $j-i$ and $j+i+1$ must have different parities, i.e., $j-i$ is even if and only if $j+i+1$ is odd. (Work out the various cases in which i and j are even and odd.) Since $(j-i)(j+i+1)/2 \equiv 0 \pmod m$, we have $(j-i)(j+i+1)/2 = rm$ for some integer r or, equivalently, $(j-i)(j+i+1) = r \cdot 2m$. Using the assumption that m is a power of 2, let $m = 2^p$ for some nonnegative integer p, so that now we have $(j-i)(j+i+1) = r \cdot 2^{p+1}$. Because exactly one of the factors $j-i$ and $j+i+1$ is even, 2^{p+1} must divide one of the factors. It cannot be $j-i$, since $j-i < m < 2^{p+1}$. But it also cannot be $j+i+1$, since $j+i+1 \le (m-1) + (m-2) + 1 = 2m-2 < 2^{p+1}$. Thus we have derived the contradiction that 2^{p+1} divides neither of the factors $j-i$ and $j+i+1$. We conclude that $h'(k,i) \ne h'(k,j)$.

Lecture Notes for Chapter 12: Binary Search Trees

Chapter 12 overview

Search trees

- Data structures that support many dynamic-set operations.
- Can be used as both a dictionary and as a priority queue.
- Basic operations take time proportional to the height of the tree.
  - For complete binary tree with n nodes: worst case $\Theta(\lg n)$.
  - For linear chain of n nodes: worst case $\Theta(n)$.
- Different types of search trees include binary search trees, red-black trees (covered in Chapter 13), and B-trees (covered in Chapter 18).

We will cover binary search trees, tree walks, and operations on binary search trees.

Binary search trees

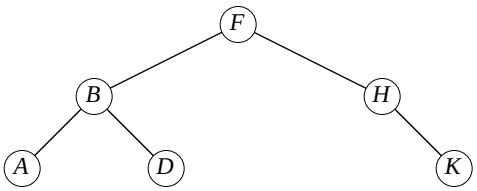
Binary search trees are an important data structure for dynamic sets.

- Accomplish many dynamic-set operations in $O(h)$ time, where h = height of tree.
- As in Section 10.4, we represent a binary tree by a linked data structure in which each node is an object.
- $T.root$ points to the root of tree T.
- Each node contains the attributes
  - *key* (and possibly other satellite data).
  - *left*: points to left child.
  - *right*: points to right child.
  - p: points to parent. $T.root.p = \text{NIL}$.

- Stored keys must satisfy the ***binary-search-tree property***.
  - If y is in left subtree of x, then $y.key \leq x.key$.
  - If y is in right subtree of x, then $y.key \geq x.key$.

Draw sample tree.

[This is Figure 12.1(a) from the text, using A, B, D, F, H, K in place of 2, 3, 5, 5, 7, 8, with alphabetic comparisons. It's OK to have duplicate keys, though there are none in this example. Show that the binary-search-tree property holds.]



The binary-search-tree property allows us to print keys in a binary search tree in order, recursively, using an algorithm called an ***inorder tree walk***. Elements are printed in monotonically increasing order.

How INORDER-TREE-WALK works:

- Check to make sure that x is not NIL.
- Recursively, print the keys of the nodes in x's left subtree.
- Print x's key.
- Recursively, print the keys of the nodes in x's right subtree.

```
INORDER-TREE-WALK(x)
  if x ≠ NIL
      INORDER-TREE-WALK(x.left)
      print key[x]
      INORDER-TREE-WALK(x.right)
```

Example

Do the inorder tree walk on the example above, getting the output $ABDFHK$.

Correctness

Follows by induction directly from the binary-search-tree property.

Time

Intuitively, the walk takes $\Theta(n)$ time for a tree with n nodes, because we visit and print each node once. *[Book has formal proof.]*

Querying a binary search tree

Searching

TREE-SEARCH(x, k)
  if x == NIL or k == *key*$[x]$
    return x
  if $k < x.key$
    return TREE-SEARCH$(x.left, k)$
  else return TREE-SEARCH$(x.right, k)$

Initial call is TREE-SEARCH$(T.root, k)$.

Example

Search for values D and C in the example tree from above.

Time

The algorithm recurses, visiting nodes on a downward path from the root. Thus, running time is $O(h)$, where h is the height of the tree.

[The text also gives an iterative version of TREE-SEARCH*, which is more efficient on most computers. The above recursive procedure is more straightforward, however.]*

Minimum and maximum

The binary-search-tree property guarantees that

- the minimum key of a binary search tree is located at the leftmost node, and
- the maximum key of a binary search tree is located at the rightmost node.

Traverse the appropriate pointers (*left* or *right*) until NIL is reached.

TREE-MINIMUM(x)
  while $x.left \neq$ NIL
    $x = x.left$
  return x

TREE-MAXIMUM(x)
  while $x.right \neq$ NIL
    $x = x.right$
  return x

Time

Both procedures visit nodes that form a downward path from the root to a leaf. Both procedures run in $O(h)$ time, where h is the height of the tree.

Successor and predecessor

Assuming that all keys are distinct, the successor of a node x is the node y such that $y.key$ is the smallest key $> x.key$. (We can find x's successor based entirely on the tree structure. No key comparisons are necessary.) If x has the largest key in the binary search tree, then we say that x's successor is NIL.

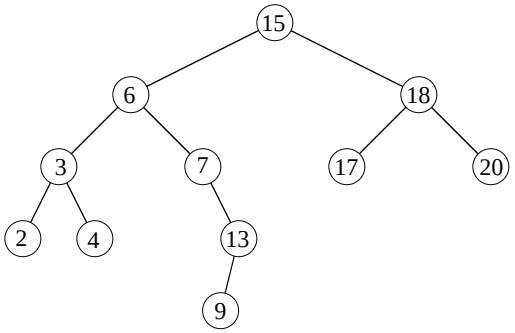
There are two cases:

1. If node x has a non-empty right subtree, then x's successor is the minimum in x's right subtree.
2. If node x has an empty right subtree, notice that:
   - As long as we move to the left up the tree (move up through right children), we're visiting smaller keys.
   - x's successor y is the node that x is the predecessor of (x is the maximum in y's left subtree).

TREE-SUCCESSOR(x)
```
if x.right ≠ NIL
    return TREE-MINIMUM(x.right)
y = x.p
while y ≠ NIL and x == y.right
    x = y
    y = y.p
return y
```

TREE-PREDECESSOR is symmetric to TREE-SUCCESSOR.

Example



- Find the successor of the node with key value 15. (Answer: Key value 17)
- Find the successor of the node with key value 6. (Answer: Key value 7)
- Find the successor of the node with key value 4. (Answer: Key value 6)
- Find the predecessor of the node with key value 6. (Answer: Key value 4)

Time

For both the TREE-SUCCESSOR and TREE-PREDECESSOR procedures, in both cases, we visit nodes on a path down the tree or up the tree. Thus, running time is $O(h)$, where h is the height of the tree.

Insertion and deletion

Insertion and deletion allows the dynamic set represented by a binary search tree to change. The binary-search-tree property must hold after the change. Insertion is more straightforward than deletion.

Insertion

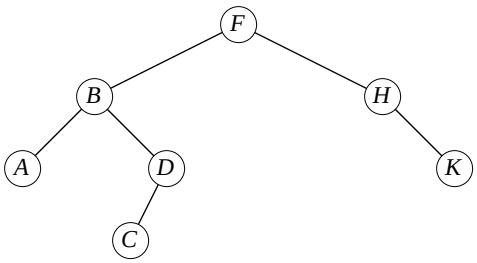
TREE-INSERT(T, z)

```
y = NIL
x = T.root
while x ≠ NIL
      y = x
      if z.key < x.key
            x = x.left
      else x = x.right
z.p = y
if y == NIL
      T.root = z          // tree T was empty
elseif z.key < y.key
      y.left = z
else y.right = z
```

- To insert value v into the binary search tree, the procedure is given node z, with $z.key = v$, $z.left =$ NIL, and $z.right =$ NIL.
- Beginning at root of the tree, trace a downward path, maintaining two pointers.
  - Pointer x: traces the downward path.
  - Pointer y: "trailing pointer" to keep track of parent of x.
- Traverse the tree downward by comparing the value of node at x with v, and move to the left or right child accordingly.
- When x is NIL, it is at the correct position for node z.
- Compare z's value with y's value, and insert z at either y's *left* or *right*, appropriately.

Example

Run TREE-INSERT(T, C) on the first sample binary search tree. Result:



Time

Same as TREE-SEARCH. On a tree of height h, procedure takes $O(h)$ time.

TREE-INSERT can be used with INORDER-TREE-WALK to sort a given set of numbers. (See Exercise 12.3-3.)

Deletion

[Deletion from a binary search tree changed in the third edition. In the first two editions, when the node z passed to TREE-DELETE *had two children, z's successor y was the node actually removed, with y's contents copied into z. The problem with that approach is that if there are external pointers into the binary search tree, then a pointer to y from outside the binary search tree becomes stale. In the third edition, the node z passed to* TREE-DELETE *is always the node actually removed, so that all external pointers to nodes other than z remain valid.]*

Conceptually, deleting node z from binary search tree T has three cases:

1. If z has no children, just remove it.
2. If z has just one child, then make that child take z's position in the tree, dragging the child's subtree along.
3. If z has two children, then find z's successor y and replace z by y in the tree. y must be in z's right subtree and have no left child. The rest of z's original right subtree becomes y's new right subtree, and z's left subtree becomes y's new left subtree.

   This case is a little tricky because the exact sequence of steps taken depends on whether y is z's right child.

The code organizes the cases a bit differently. Since it will move subtrees around within the binary search tree, it uses a subroutine, TRANSPLANT, to replace one subtree as the child of its parent by another subtree.

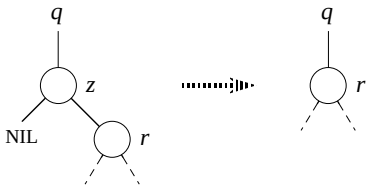
```
TRANSPLANT(T, u, v)
 if u.p == NIL
     T.root = v
 elseif u == u.p.left
     u.p.left = v
 else u.p.right = v
 if v ≠ NIL
     v.p = u.p
```

TRANSPLANT(T, u, v) replaces the subtree rooted at u by the subtree rooted at v:

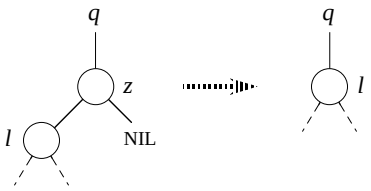
- Makes u's parent become v's parent (unless u is the root, in which case it makes v the root).
- u's parent gets v as either its left or right child, depending on whether u was a left or right child.
- Doesn't update $v.left$ or $v.right$, leaving that up to TRANSPLANT's caller.

TREE-DELETE(T, z) has four cases when deleting node z from binary search tree T:

- If z has no left child, replace z by its right child. The right child may or may not be NIL. (If z's right child is NIL, then this case handles the situation in which z has no children.)



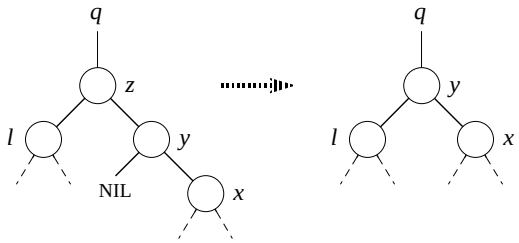
- If z has just one child, and that child is its left child, then replace z by its left child.



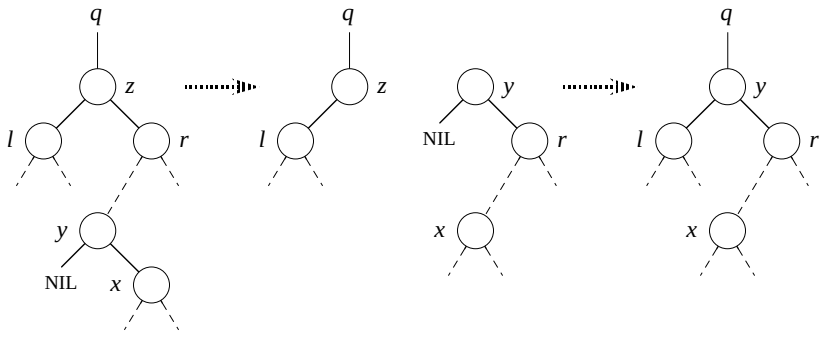
- Otherwise, z has two children. Find z's successor y. y must lie in z's right subtree and have no left child (the solution to Exercise 12.2-5 on page 12-15 of this manual shows why).

  Goal is to replace z by y, splicing y out of its current location.

  - If y is z's right child, replace z by y and leave y's right child alone.



  - Otherwise, y lies within z's right subtree but is not the root of this subtree. Replace y by its own right child. Then replace z by y.

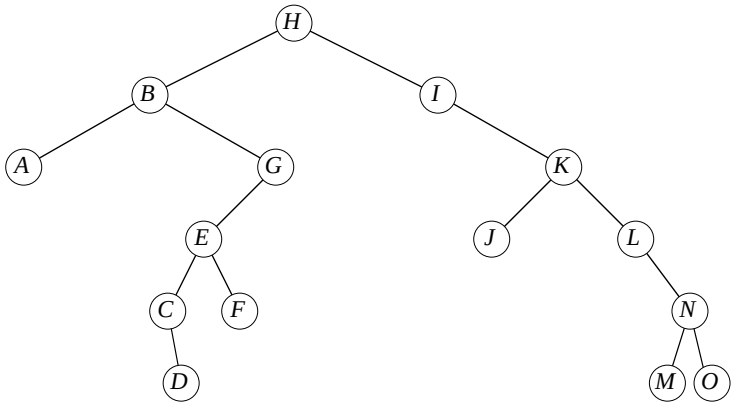


```
TREE-DELETE(T, z)
  if z.left == NIL
      TRANSPLANT(T, z, z.right)          // z has no left child
  elseif z.right == NIL
      TRANSPLANT(T, z, z.left)           // z has just a left child
  else // z has two children.
      y = TREE-MINIMUM(z.right)          // y is z's successor
      if y.p ≠ z
          // y lies within z's right subtree but is not the root of this subtree.
          TRANSPLANT(T, y, y.right)
          y.right = z.right
          y.right.p = y
      // Replace z by y.
      TRANSPLANT(T, z, y)
      y.left = z.left
      y.left.p = y
```

Note that the last three lines execute when z has two children, regardless of whether y is z's right child.

Example

On this binary search tree T,



run the following. *[You can either start with the original tree each time or start with the result of the previous call. The tree is designed so that either way will elicit all four cases.]*

- TREE-DELETE(T, I) shows the case in which the node deleted has no left child.
- TREE-DELETE(T, G) shows the case in which the node deleted has a left child but no right child.
- TREE-DELETE(T, K) shows the case in which the node deleted has both children and its successor is its right child.
- TREE-DELETE(T, B) shows the case in which the node deleted has both children and its successor is not its right child.

Time

$O(h)$, on a tree of height h. Everything is $O(1)$ except for the call to TREE-MINIMUM.

Minimizing running time

We've been analyzing running time in terms of h (the height of the binary search tree), instead of n (the number of nodes in the tree).

- Problem: Worst case for binary search tree is $\Theta(n)$—no better than linked list.
- Solution: Guarantee small height (balanced tree)—$h = O(\lg n)$.

In later chapters, by varying the properties of binary search trees, we will be able to analyze running time in terms of n.

- Method: Restructure the tree if necessary. Nothing special is required for querying, but there may be extra work when changing the structure of the tree (inserting or deleting).

Red-black trees are a special class of binary trees that avoids the worst-case behavior of $O(n)$ that we can see in "plain" binary search trees. Red-black trees are covered in detail in Chapter 13.

Expected height of a randomly built binary search tree

[These are notes on a starred section in the book. I covered this material in an optional lecture.]

Given a set of n distinct keys. Insert them in random order into an initially empty binary search tree.

- Each of the $n!$ permutations is equally likely.
- Different from assuming that every binary search tree on n keys is equally likely.

  Try it for $n = 3$. Will get 5 different binary search trees. When we look at the binary search trees resulting from each of the 3! input permutations, 4 trees will appear once and 1 tree will appear twice. *[This gives the idea for the solution to Exercise 12.4-3.]*
- Forget about deleting keys.

We will show that the expected height of a randomly built binary search tree is $O(\lg n)$.

Random variables

Define the following random variables:

- X_n = height of a randomly built binary search tree on n keys.

- $Y_n = 2^{X_n}$ = ***exponential height***.
- R_n = rank of the root within the set of n keys used to build the binary search tree.
  - Equally likely to be any element of $\{1, 2, \ldots, n\}$.
  - If $R_n = i$, then
    - Left subtree is a randomly-built binary search tree on $i - 1$ keys.
    - Right subtree is a randomly-built binary search tree on $n - i$ keys.

Foreshadowing

We will need to relate $\mathrm{E}[Y_n]$ to $\mathrm{E}[X_n]$.

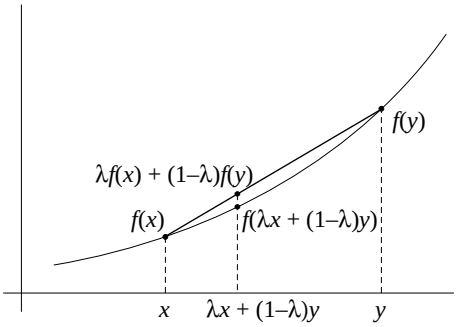
We'll use ***Jensen's inequality***:

$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X])$, *[leave on board]*

provided

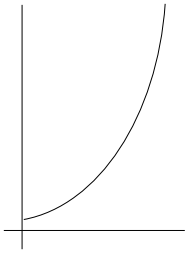
- the expectations exist and are finite, and
- $f(x)$ is ***convex***: for all x, y and all $0 \leq \lambda \leq 1$

  $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$.



Convex ≡ "curves upward"

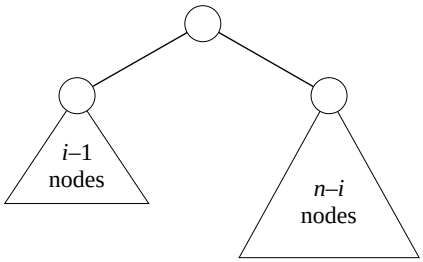
We'll use Jensen's inequality for $f(x) = 2^x$.



Since 2^x curves upward, it's convex.

Formula for Y_n

Think about Y_n, if we know that $R_n = i$:



Height of root is 1 more than the maximum height of its children:

$Y_n = 2 \cdot \max(Y_{i-1}, Y_{n-i})$.

Base cases:

- $Y_1 = 1$ (expected height of a 1-node tree is $2^0 = 1$).
- Define $Y_0 = 0$.

Define indicator random variables $Z_{n,1}, Z_{n,2}, \ldots, Z_{n,n}$:

$Z_{n,i} = \mathrm{I}\{R_n = i\}$.

R_n is equally likely to be any element of $\{1, 2, \ldots, n\}$
$\Rightarrow \Pr\{R_n = i\} = 1/n$
$\Rightarrow \mathrm{E}[Z_{n,i}] = 1/n$ *[leave on board]*
(since $\mathrm{E}[\mathrm{I}\{A\}] = \Pr\{A\}$)

Consider a given n-node binary search tree (which could be a subtree). Exactly one $Z_{n,i}$ is 1, and all others are 0. Hence,

$$Y_n = \sum_{i=1}^{n} Z_{n,i} \cdot (2 \cdot \max(Y_{i-1}, Y_{n-i})) \; . \qquad \textit{[leave on board]}$$

[Recall: $Y_n = 2 \cdot \max(Y_{i-1}, Y_{n-i})$ was assuming that $R_n = i$.]

Bounding $\mathrm{E}[Y_n]$

We will show that $\mathrm{E}[Y_n]$ is polynomial in n, which will imply that $\mathrm{E}[X_n] = O(\lg n)$.

Claim
$Z_{n,i}$ is independent of Y_{i-1} and Y_{n-i}.

Justification If we choose the root such that $R_n = i$, the left subtree contains $i - 1$ nodes, and it's like any other randomly built binary search tree with $i - 1$ nodes. Other than the number of nodes, the left subtree's structure has nothing to do with it being the left subtree of the root. Hence, Y_{i-1} and $Z_{n,i}$ are independent.

Similarly, Y_{n-i} and $Z_{n,i}$ are independent. ■ (claim)

Fact
If X and Y are nonnegative random variables, then $\mathrm{E}[\max(X,Y)] \le \mathrm{E}[X]+\mathrm{E}[Y]$.
[Leave on board. This is Exercise C.3-4 from the text.]

Thus,

$$\begin{aligned}
\mathrm{E}[Y_n] &= \mathrm{E}\left[\sum_{i=1}^{n} Z_{n,i}\,(2\cdot\max(Y_{i-1},Y_{n-i}))\right] \\
&= \sum_{i=1}^{n}\mathrm{E}[Z_{n,i}\cdot(2\cdot\max(Y_{i-1},Y_{n-i}))] && \text{(linearity of expectation)} \\
&= \sum_{i=1}^{n}\mathrm{E}[Z_{n,i}]\cdot\mathrm{E}[2\cdot\max(Y_{i-1},Y_{n-i})] && \text{(independence)} \\
&= \sum_{i=1}^{n}\frac{1}{n}\cdot\mathrm{E}[2\cdot\max(Y_{i-1},Y_{n-i})] && (\mathrm{E}[Z_{n,i}] = 1/n) \\
&= \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}[\max(Y_{i-1},Y_{n-i})] && (\mathrm{E}[aX] = a\,\mathrm{E}[X]) \\
&\le \frac{2}{n}\sum_{i=1}^{n}(\mathrm{E}[Y_{i-1}]+\mathrm{E}[Y_{n-i}]) && \text{(earlier fact)}\ .
\end{aligned}$$

Observe that the last summation is

$$(\mathrm{E}[Y_0]+\mathrm{E}[Y_{n-1}])+(\mathrm{E}[Y_1]+\mathrm{E}[Y_{n-2}])+(\mathrm{E}[Y_2]+\mathrm{E}[Y_{n-3}]) \\ +\cdots+(\mathrm{E}[Y_{n-1}]+\mathrm{E}[Y_0]) = 2\sum_{i=0}^{n-1}\mathrm{E}[Y_i]\ ,$$

and so we get the recurrence

$$\mathrm{E}[Y_n] \le \frac{4}{n}\sum_{i=0}^{n-1}\mathrm{E}[Y_i]\ . \qquad \textit{[leave on board]}$$

Solving the recurrence

We will show that for all integers $n > 0$, this recurrence has the solution

$$\mathrm{E}[Y_n] \le \frac{1}{4}\binom{n+3}{3}\ .$$

Lemma

$$\sum_{i=0}^{n-1}\binom{i+3}{3} = \binom{n+3}{4}\ .$$

[This lemma solves Exercise 12.4-1.]

Proof Use Pascal's identity (Exercise C.1-7): $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

Also using the simple identity $\binom{4}{4} = 1 = \binom{3}{3}$, we have

$$\begin{aligned}
\binom{n+3}{4} &= \binom{n+2}{3} + \binom{n+2}{4} \\
&= \binom{n+2}{3} + \binom{n+1}{3} + \binom{n+1}{4} \\
&= \binom{n+2}{3} + \binom{n+1}{3} + \binom{n}{3} + \binom{n}{4} \\
&\vdots \\
&= \binom{n+2}{3} + \binom{n+1}{3} + \binom{n}{3} + \cdots + \binom{4}{3} + \binom{4}{4} \\
&= \binom{n+2}{3} + \binom{n+1}{3} + \binom{n}{3} + \cdots + \binom{4}{3} + \binom{3}{3} \\
&= \sum_{i=0}^{n-1} \binom{i+3}{3} .
\end{aligned}$$

■ (lemma)

We solve the recurrence by induction on n.

Basis: $n = 1$.

$$1 = Y_1 = \mathrm{E}[Y_1] \le \frac{1}{4}\binom{1+3}{3} = \frac{1}{4} \cdot 4 = 1 .$$

Inductive step: Assume that $\mathrm{E}[Y_i] \le \frac{1}{4}\binom{i+3}{3}$ for all $i < n$. Then

$$\begin{aligned}
\mathrm{E}[Y_n] &\le \frac{4}{n}\sum_{i=0}^{n-1} \mathrm{E}[Y_i] && \text{(from before)} \\
&\le \frac{4}{n}\sum_{i=0}^{n-1} \frac{1}{4}\binom{i+3}{3} && \text{(inductive hypothesis)} \\
&= \frac{1}{n}\sum_{i=0}^{n-1} \binom{i+3}{3} \\
&= \frac{1}{n}\binom{n+3}{4} && \text{(lemma)} \\
&= \frac{1}{n} \cdot \frac{(n+3)!}{4!\,(n-1)!} \\
&= \frac{1}{4} \cdot \frac{(n+3)!}{3!\,n!}
\end{aligned}$$

$$= \frac{1}{4}\binom{n+3}{3}.$$

Thus, we've proven that $\mathrm{E}[Y_n] \le \frac{1}{4}\binom{n+3}{3}$.

Bounding $\mathrm{E}[X_n]$

With our bound on $\mathrm{E}[Y_n]$, we use Jensen's inequality to bound $\mathrm{E}[X_n]$:

$2^{\mathrm{E}[X_n]} \le \mathrm{E}\left[2^{X_n}\right] = \mathrm{E}[Y_n]$.

Thus,

$$\begin{aligned} 2^{\mathrm{E}[X_n]} &\le \frac{1}{4}\binom{n+3}{3} \\ &= \frac{1}{4} \cdot \frac{(n+3)(n+2)(n+1)}{6} \\ &= O(n^3) . \end{aligned}$$

Taking logs of both sides gives $\mathrm{E}[X_n] = O(\lg n)$.

Done!

Solutions for Chapter 12: Binary Search Trees

Solution to Exercise 12.1-2

This solution is also posted publicly

In a heap, a node's key is $\geq$ both of its children's keys. In a binary search tree, a node's key is $\geq$ its left child's key, but $\leq$ its right child's key.

The heap property, unlike the binary-searth-tree property, doesn't help print the nodes in sorted order because it doesn't tell which subtree of a node contains the element to print before that node. In a heap, the largest element smaller than the node could be in either subtree.

Note that if the heap property could be used to print the keys in sorted order in $O(n)$ time, we would have an $O(n)$-time algorithm for sorting, because building the heap takes only $O(n)$ time. But we know (Chapter 8) that a comparison sort must take $\Omega(n \lg n)$ time.

Solution to Exercise 12.2-5

Let x be a node with two children. In an inorder tree walk, the nodes in x's left subtree immediately precede x and the nodes in x's right subtree immediately follow x. Thus, x's predecessor is in its left subtree, and its successor is in its right subtree.

Let s be x's successor. Then s cannot have a left child, for a left child of s would come between x and s in the inorder walk. (It's after x because it's in x's right subtree, and it's before s because it's in s's left subtree.) If any node were to come between x and s in an inorder walk, then s would not be x's successor, as we had supposed.

Symmetrically, x's predecessor has no right child.

Solution to Exercise 12.2-7

This solution is also posted publicly

Note that a call to TREE-MINIMUM followed by $n-1$ calls to TREE-SUCCESSOR performs exactly the same inorder walk of the tree as does the procedure INORDER-TREE-WALK. INORDER-TREE-WALK prints the TREE-MINIMUM first, and by definition, the TREE-SUCCESSOR of a node is the next node in the sorted order determined by an inorder tree walk.

This algorithm runs in $\Theta(n)$ time because:

- It requires $\Omega(n)$ time to do the n procedure calls.
- It traverses each of the $n-1$ tree edges at most twice, which takes $O(n)$ time.

To see that each edge is traversed at most twice (once going down the tree and once going up), consider the edge between any node u and either of its children, node v. By starting at the root, we must traverse (u, v) downward from u to v, before traversing it upward from v to u. The only time the tree is traversed downward is in code of TREE-MINIMUM, and the only time the tree is traversed upward is in code of TREE-SUCCESSOR when we look for the successor of a node that has no right subtree.

Suppose that v is u's left child.

- Before printing u, we must print all the nodes in its left subtree, which is rooted at v, guaranteeing the downward traversal of edge (u, v).
- After all nodes in u's left subtree are printed, u must be printed next. Procedure TREE-SUCCESSOR traverses an upward path to u from the maximum element (which has no right subtree) in the subtree rooted at v. This path clearly includes edge (u, v), and since all nodes in u's left subtree are printed, edge (u, v) is never traversed again.

Now suppose that v is u's right child.

- After u is printed, TREE-SUCCESSOR(u) is called. To get to the minimum element in u's right subtree (whose root is v), the edge (u, v) must be traversed downward.
- After all values in u's right subtree are printed, TREE-SUCCESSOR is called on the maximum element (again, which has no right subtree) in the subtree rooted at v. TREE-SUCCESSOR traverses a path up the tree to an element after u, since u was already printed. Edge (u, v) must be traversed upward on this path, and since all nodes in u's right subtree have been printed, edge (u, v) is never traversed again.

Hence, no edge is traversed twice in the same direction.

Therefore, this algorithm runs in $\Theta(n)$ time.

Solution to Exercise 12.3-3

This solution is also posted publicly

Here's the algorithm:

```
TREE-SORT(A)
  let T be an empty binary search tree
  for i = 1 to n
      TREE-INSERT(T, A[i])
  INORDER-TREE-WALK(T.root)
```

Worst case: $\Theta(n^2)$—occurs when a linear chain of nodes results from the repeated TREE-INSERT operations.

Best case: $\Theta(n \lg n)$—occurs when a binary tree of height $\Theta(\lg n)$ results from the repeated TREE-INSERT operations.

Solution to Exercise 12.4-2

We will answer the second part first. We shall show that if the average depth of a node is $\Theta(\lg n)$, then the height of the tree is $O(\sqrt{n \lg n})$. Then we will answer the first part by exhibiting that this bound is tight: there is a binary search tree with average node depth $\Theta(\lg n)$ and height $\Theta(\sqrt{n \lg n}) = \omega(\lg n)$.

Lemma

If the average depth of a node in an n-node binary search tree is $\Theta(\lg n)$, then the height of the tree is $O(\sqrt{n \lg n})$.

Proof Suppose that an n-node binary search tree has average depth $\Theta(\lg n)$ and height h. Then there exists a path from the root to a node at depth h, and the depths of the nodes on this path are $0, 1, \ldots, h$. Let P be the set of nodes on this path and Q be all other nodes. Then the average depth of a node is

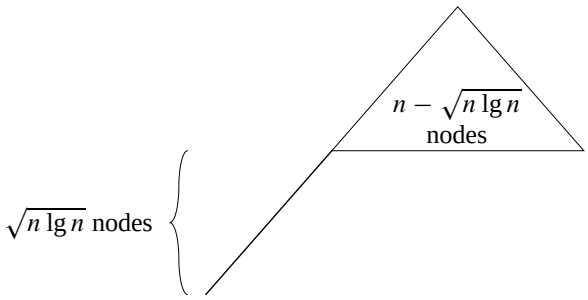
$$\begin{aligned}
\frac{1}{n}\left(\sum_{x \in P} \text{depth}(x) + \sum_{y \in Q} \text{depth}(y)\right) &\geq \frac{1}{n}\sum_{x \in P} \text{depth}(x) \\
&= \frac{1}{n}\sum_{d=0}^{h} d \\
&= \frac{1}{n} \cdot \Theta(h^2) \; .
\end{aligned}$$

For the purpose of contradiction, suppose that h is not $O(\sqrt{n \lg n})$, so that $h = \omega(\sqrt{n \lg n})$. Then we have

$$\begin{aligned}
\frac{1}{n} \cdot \Theta(h^2) &= \frac{1}{n} \cdot \omega(n \lg n) \\
&= \omega(\lg n) \; ,
\end{aligned}$$

which contradicts the assumption that the average depth is $\Theta(\lg n)$. Thus, the height is $O(\sqrt{n \lg n})$. ■

Here is an example of an n-node binary search tree with average node depth $\Theta(\lg n)$ but height $\omega(\lg n)$:



In this tree, $n - \sqrt{n \lg n}$ nodes are a complete binary tree, and the other $\sqrt{n \lg n}$ nodes protrude from below as a single chain. This tree has height

$$\begin{aligned}\Theta(\lg(n - \sqrt{n \lg n})) + \sqrt{n \lg n} &= \Theta(\sqrt{n \lg n}) \\ &= \omega(\lg n) .\end{aligned}$$

To compute an upper bound on the average depth of a node, we use $O(\lg n)$ as an upper bound on the depth of each of the $n - \sqrt{n \lg n}$ nodes in the complete binary tree part and $O(\lg n + \sqrt{n \lg n})$ as an upper bound on the depth of each of the $\sqrt{n \lg n}$ nodes in the protruding chain. Thus, the average depth of a node is bounded from above by

$$\begin{aligned}\frac{1}{n} \cdot O(\sqrt{n \lg n}\ (\lg n + \sqrt{n \lg n}) + (n - \sqrt{n \lg n}) \lg n) &= \frac{1}{n} \cdot O(n \lg n) \\ &= O(\lg n) .\end{aligned}$$

To bound the average depth of a node from below, observe that the bottommost level of the complete binary tree part has $\Theta(n - \sqrt{n \lg n})$ nodes, and each of these nodes has depth $\Theta(\lg n)$. Thus, the average node depth is at least

$$\begin{aligned}\frac{1}{n} \cdot \Theta((n - \sqrt{n \lg n}) \lg n) &= \frac{1}{n} \cdot \Omega(n \lg n) \\ &= \Omega(\lg n) .\end{aligned}$$

Because the average node depth is both $O(\lg n)$ and $\Omega(\lg n)$, it is $\Theta(\lg n)$.

Solution to Exercise 12.4-4

We'll go one better than showing that the function 2^x is convex. Instead, we'll show that the function c^x is convex, for any positive constant c. According to the definition of convexity on page 1199 of the text, a function $f(x)$ is convex if for all x and y and for all $0 \le \lambda \le 1$, we have $f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda) f(y)$. Thus, we need to show that for all $0 \le \lambda \le 1$, we have $c^{\lambda x + (1-\lambda)y} \le \lambda c^x + (1 - \lambda)c^y$.

We start by proving the following lemma.

Lemma

For any real numbers a and b and any positive real number c,

$$c^a \geq c^b + (a-b)c^b \ln c \ .$$

Proof We first show that for all real r, we have $c^r \geq 1 + r \ln c$. By equation (3.12) from the text, we have $e^x \geq 1 + x$ for all real x. Let $x = r \ln c$, so that $e^x = e^{r \ln c} = (e^{\ln c})^r = c^r$. Then we have $c^r = e^{r \ln c} \geq 1 + r \ln c$.

Substituting $a - b$ for r in the above inequality, we have $c^{a-b} \geq 1 + (a-b) \ln c$. Multiplying both sides by c^b gives $c^a \geq c^b + (a-b)c^b \ln c$. ■ (lemma)

Now we can show that $c^{\lambda x + (1-\lambda)y} \leq \lambda c^x + (1-\lambda)c^y$ for all $0 \leq \lambda \leq 1$. For convenience, let $z = \lambda x + (1-\lambda)y$.

In the inequality given by the lemma, substitute x for a and z for b, giving

$$c^x \geq c^z + (x-z)c^z \ln c \ .$$

Also substitute y for a and z for b, giving

$$c^y \geq c^z + (y-z)c^z \ln c \ .$$

If we multiply the first inequality by λ and the second by $1 - \lambda$ and then add the resulting inequalities, we get

$$\begin{aligned}
&\lambda c^x + (1-\lambda)c^y \\
&\quad \geq \lambda(c^z + (x-z)c^z \ln c) + (1-\lambda)(c^z + (y-z)c^z \ln c) \\
&\quad = \lambda c^z + \lambda x c^z \ln c - \lambda z c^z \ln c + (1-\lambda)c^z + (1-\lambda)y c^z \ln c \\
&\qquad\qquad - (1-\lambda)z c^z \ln c \\
&\quad = (\lambda + (1-\lambda))c^z + (\lambda x + (1-\lambda)y)c^z \ln c - (\lambda + (1-\lambda))z c^z \ln c \\
&\quad = c^z + z c^z \ln c - z c^z \ln c \\
&\quad = c^z \\
&\quad = c^{\lambda x + (1-\lambda)y} \ ,
\end{aligned}$$

as we wished to show.

Solution to Problem 12-2

This solution is also posted publicly

To sort the strings of S, we first insert them into a radix tree, and then use a preorder tree walk to extract them in lexicographically sorted order. The tree walk outputs strings only for nodes that indicate the existence of a string (i.e., those that are lightly shaded in Figure 12.5 of the text).

Correctness

The preorder ordering is the correct order because:

- Any node's string is a prefix of all its descendants' strings and hence belongs before them in the sorted order (rule 2).

- A node's left descendants belong before its right descendants because the corresponding strings are identical up to that parent node, and in the next position the left subtree's strings have 0 whereas the right subtree's strings have 1 (rule 1).

Time

$\Theta(n)$.

- Insertion takes $\Theta(n)$ time, since the insertion of each string takes time proportional to its length (traversing a path through the tree whose length is the length of the string), and the sum of all the string lengths is n.
- The preorder tree walk takes $O(n)$ time. It is just like INORDER-TREE-WALK (it prints the current node and calls itself recursively on the left and right subtrees), so it takes time proportional to the number of nodes in the tree. The number of nodes is at most 1 plus the sum (n) of the lengths of the binary strings in the tree, because a length-i string corresponds to a path through the root and i other nodes, but a single node may be shared among many string paths.

Solution to Problem 12-3

a. The total path length $P(T)$ is defined as $\sum_{x \in T} d(x, T)$. Dividing both quantities by n gives the desired equation.

b. For any node x in T_L, we have $d(x, T_L) = d(x, T) - 1$, since the distance to the root of T_L is one less than the distance to the root of T. Similarly, for any node x in T_R, we have $d(x, T_R) = d(x, T) - 1$. Thus, if T has n nodes, we have

$$P(T) = P(T_L) + P(T_R) + n - 1 \ ,$$

since each of the n nodes of T (except the root) is in either T_L or T_R.

c. If T is a randomly built binary search tree, then the root is equally likely to be any of the n elements in the tree, since the root is the first element inserted. It follows that the number of nodes in subtree T_L is equally likely to be any integer in the set $\{0, 1, \ldots, n-1\}$. The definition of $P(n)$ as the average total path length of a randomly built binary search tree, along with part (b), gives us the recurrence

$$P(n) = \frac{1}{n} \sum_{i=0}^{n-1} (P(i) + P(n-i-1) + n - 1) \ .$$

d. Since $P(0) = 0$, and since for $k = 1, 2, \ldots, n-1$, each term $P(k)$ in the summation appears once as $P(i)$ and once as $P(n-i-1)$, we can rewrite the equation from part (c) as

$$P(n) = \frac{2}{n} \sum_{k=1}^{n-1} P(k) + \Theta(n) \ .$$

e. Observe that if, in the recurrence (7.6) in part (c) of Problem 7-3, we replace $\mathrm{E}[T(\cdot)]$ by $P(\cdot)$ and we replace q by k, we get almost the same recurrence as in part (d) of Problem 12-3. The remaining difference is that in Problem 12-3(d), the summation starts at 1 rather than 2. Observe, however, that a binary tree with just one node has a total path length of 0, so that $P(1) = 0$. Thus, we can rewrite the recurrence in Problem 12-3(d) as

$$P(n) = \frac{2}{n}\sum_{k=2}^{n-1} P(k) + \Theta(n)$$

and use the same technique as was used in Problem 7-3 to solve it.

We start by solving part (d) of Problem 7-3: showing that

$$\sum_{k=2}^{n-1} k \lg k \le \frac{1}{2}n^2 \lg n - \frac{1}{8}n^2 .$$

Following the hint in Problem 7-3(d), we split the summation into two parts:

$$\sum_{k=2}^{n-1} k \lg k = \sum_{k=2}^{\lceil n/2 \rceil - 1} k \lg k + \sum_{k=\lceil n/2 \rceil}^{n-1} k \lg k .$$

The $\lg k$ in the first summation on the right is less than $\lg(n/2) = \lg n - 1$, and the $\lg k$ in the second summation is less than $\lg n$. Thus,

$$\begin{aligned}
\sum_{k=2}^{n-1} k \lg k &< (\lg n - 1) \sum_{k=2}^{\lceil n/2 \rceil - 1} k + \lg n \sum_{k=\lceil n/2 \rceil}^{n-1} k \\
&= \lg n \sum_{k=2}^{n-1} k - \sum_{k=2}^{\lceil n/2 \rceil - 1} k \\
&\le \frac{1}{2}n(n-1)\lg n - \frac{1}{2}\left(\frac{n}{2} - 1\right)\frac{n}{2} \\
&\le \frac{1}{2}n^2 \lg n - \frac{1}{8}n^2
\end{aligned}$$

if $n \ge 2$.

Now we show that the recurrence

$$P(n) = \frac{2}{n}\sum_{k=2}^{n-1} P(k) + \Theta(n)$$

has the solution $P(n) = O(n \lg n)$. We use the substitution method. Assume inductively that $P(n) \le an \lg n + b$ for some positive constants a and b to be determined. We can pick a and b sufficiently large so that $an \lg n + b \ge P(1)$. Then, for $n > 1$, we have by substitution

$$\begin{aligned}
P(n) &= \frac{2}{n}\sum_{k=2}^{n-1} P(k) + \Theta(n) \\
&\le \frac{2}{n}\sum_{k=2}^{n-1}(ak \lg k + b) + \Theta(n)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2a}{n}\sum_{k=2}^{n-1} k\lg k + \frac{2b}{n}(n-2) + \Theta(n) \\
&\le \frac{2a}{n}\left(\frac{1}{2}n^2\lg n - \frac{1}{8}n^2\right) + \frac{2b}{n}(n-2) + \Theta(n) \\
&\le an\lg n - \frac{a}{4}n + 2b + \Theta(n) \\
&= an\lg n + b + \left(\Theta(n) + b - \frac{a}{4}n\right) \\
&\le an\lg n + b\,,
\end{aligned}
$$

since we can choose a large enough so that $\frac{a}{4}n$ dominates $\Theta(n) + b$. Thus, $P(n) = O(n\lg n)$.

f. We draw an analogy between inserting an element into a subtree of a binary search tree and sorting a subarray in quicksort. Observe that once an element x is chosen as the root of a subtree T, all elements that will be inserted after x into T will be compared to x. Similarly, observe that once an element y is chosen as the pivot in a subarray S, all other elements in S will be compared to y. Therefore, the quicksort implementation in which the comparisons are the same as those made when inserting into a binary search tree is simply to consider the pivots in the same order as the order in which the elements are inserted into the tree.

Lecture Notes for Chapter 13: Red-Black Trees

Chapter 13 overview

Red-black trees

- A variation of binary search trees.
- ***Balanced***: height is $O(\lg n)$, where n is the number of nodes.
- Operations will take $O(\lg n)$ time in the worst case.

[These notes are a bit simpler than the treatment in the book, to make them more amenable to a lecture situation. Our students first see red-black trees in a course that precedes our algorithms course. This set of lecture notes is intended as a refresher for the students, bearing in mind that some time may have passed since they last saw red-black trees.

The procedures in this chapter are rather long sequences of pseudocode. You might want to make arrangements to project them rather than spending time writing them on a board.]

Red-black trees

A ***red-black tree*** is a binary search tree + 1 bit per node: an attribute *color*, which is either red or black.

All leaves are empty (nil) and colored black.

- We use a single sentinel, $T.nil$, for all the leaves of red-black tree T.
- $T.nil.color$ is black.
- The root's parent is also $T.nil$.

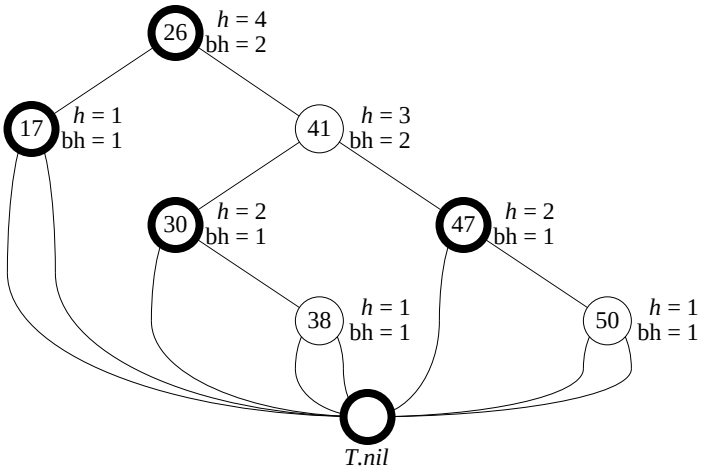
All other attributes of binary search trees are inherited by red-black trees (*key*, *left*, *right*, and p). We don't care about the key in $T.nil$.

Red-black properties

[Leave these up on the board.]

1. Every node is either red or black.
2. The root is black.
3. Every leaf ($T.nil$) is black.
4. If a node is red, then both its children are black. (Hence no two reds in a row on a simple path from the root to a leaf.)
5. For each node, all paths from the node to descendant leaves contain the same number of black nodes.

Example:



[Nodes with bold outline indicate black nodes. Don't add heights and black-heights yet. We won't bother with drawing T.nil any more.]

Height of a red-black tree

- ***Height of a node*** is the number of edges in a longest path to a leaf.
- ***Black-height*** of a node x: $\text{bh}(x)$ is the number of black nodes (including $T.nil$) on the path from x to leaf, not counting x. By property 5, black-height is well defined.

[Now label the example tree with height h and bh *values.]*

Claim

Any node with height h has black-height $\geq h/2$.

Proof By property 4, $\leq h/2$ nodes on the path from the node to a leaf are red. Hence $\geq h/2$ are black. ■ (claim)

Claim

The subtree rooted at any node x contains $\geq 2^{\text{bh}(x)} - 1$ internal nodes.

Proof By induction on height of x.

Basis: Height of $x = 0 \Rightarrow x$ is a leaf $\Rightarrow \text{bh}(x) = 0$. The subtree rooted at x has 0 internal nodes. $2^0 - 1 = 0$.

Inductive step: Let the height of x be h and $\text{bh}(x) = b$. Any child of x has height $h - 1$ and black-height either b (if the child is red) or $b - 1$ (if the child is black). By the inductive hypothesis, each child has $\geq 2^{\text{bh}(x)-1} - 1$ internal nodes. Thus, the subtree rooted at x contains $\geq 2 \cdot (2^{\text{bh}(x)-1} - 1) + 1 = 2^{\text{bh}(x)} - 1$ internal nodes. (The $+1$ is for x itself.) ■ (claim)

Lemma
A red-black tree with n internal nodes has height $\leq 2\lg(n + 1)$.

Proof Let h and b be the height and black-height of the root, respectively. By the above two claims,

$$n \geq 2^b - 1 \geq 2^{h/2} - 1 .$$

Adding 1 to both sides and then taking logs gives $\lg(n + 1) \geq h/2$, which implies that $h \leq 2\lg(n + 1)$. ■ (theorem)

Operations on red-black trees

The non-modifying binary-search-tree operations MINIMUM, MAXIMUM, SUCCESSOR, PREDECESSOR, and SEARCH run in $O(\text{height})$ time. Thus, they take $O(\lg n)$ time on red-black trees.

Insertion and deletion are not so easy.

If we insert, what color to make the new node?

- Red? Might violate property 4.
- Black? Might violate property 5.

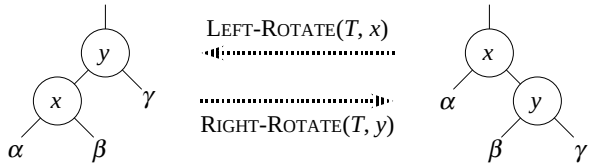
If we delete, thus removing a node, what color was the node that was removed?

- Red? OK, since we won't have changed any black-heights, nor will we have created two red nodes in a row. Also, cannot cause a violation of property 2, since if the removed node was red, it could not have been the root.
- Black? Could cause there to be two reds in a row (violating property 4), and can also cause a violation of property 5. Could also cause a violation of property 2, if the removed node was the root and its child—which becomes the new root—was red.

Rotations

- The basic tree-restructuring operation.
- Needed to maintain red-black trees as balanced binary search trees.
- Changes the local pointer structure. (Only pointers are changed.)

- Won't upset the binary-search-tree property.
- Have both left rotation and right rotation. They are inverses of each other.
- A rotation takes a red-black-tree and a node within the tree.



LEFT-ROTATE(T, x)

```
y = x.right                 // set y
x.right = y.left            // turn y's left subtree into x's right subtree
if y.left ≠ T.nil
    y.left.p = x
y.p = x.p                   // link x's parent to y
if x.p == T.nil
    T.root = y
elseif x == x.p.left
    x.p.left = y
else x.p.right = y
y.left = x                  // put x on y's left
x.p = y
```

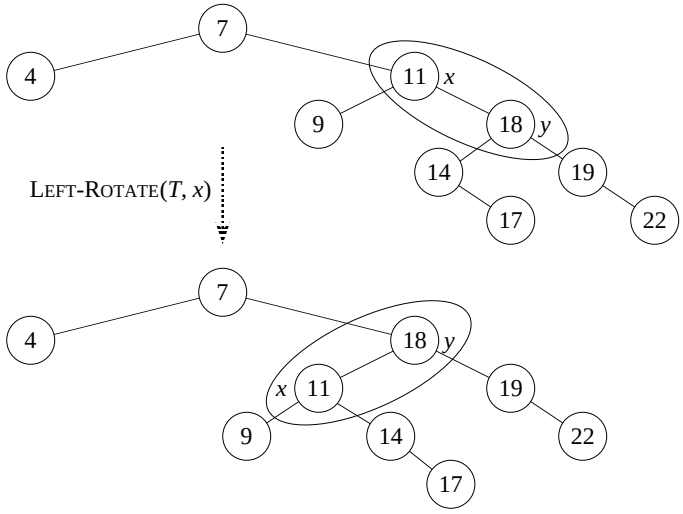
The pseudocode for LEFT-ROTATE assumes that

- $x.right \neq T.nil$, and
- root's parent is $T.nil$.

Pseudocode for RIGHT-ROTATE is symmetric: exchange *left* and *right* everywhere.

Example

[Use to demonstrate that rotation maintains inorder ordering of keys. Node colors omitted.]



- Before rotation: keys of x's left subtree $\leq 11 \leq$ keys of y's left subtree $\leq 18 \leq$ keys of y's right subtree.
- Rotation makes y's left subtree into x's right subtree.
- After rotation: keys of x's left subtree $\leq 11 \leq$ keys of x's right subtree $\leq 18 \leq$ keys of y's right subtree.

Time

$O(1)$ for both LEFT-ROTATE and RIGHT-ROTATE, since a constant number of pointers are modified.

Notes

- Rotation is a very basic operation, also used in AVL trees and splay trees.
- Some books talk of rotating on an edge rather than on a node.

Insertion

Start by doing regular binary-search-tree insertion:

RB-INSERT(T, z)

```
y = T.nil
x = T.root
while x ≠ T.nil
      y = x
      if z.key < x.key
            x = x.left
      else x = x.right
z.p = y
if y == T.nil
      T.root = z
elseif z.key < y.key
      y.left = z
else y.right = z
z.left = T.nil
z.right = T.nil
z.color = RED
```

RB-INSERT-FIXUP(T, z)

- RB-INSERT ends by coloring the new node z red.
- Then it calls RB-INSERT-FIXUP because we could have violated a red-black property.

Which property might be violated?

1. OK.

2. If z is the root, then there's a violation. Otherwise, OK.
3. OK.
4. If $z.p$ is red, there's a violation: both z and $z.p$ are red.
5. OK.

Remove the violation by calling RB-INSERT-FIXUP:

```
RB-INSERT-FIXUP(T, z)
  while z.p.color == RED
      if z.p == z.p.p.left
          y = z.p.p.right
          if y.color == RED
              z.p.color = BLACK                          // case 1
              y.color = BLACK                            // case 1
              z.p.p.color = RED                          // case 1
              z = z.p.p                                  // case 1
          else if z == z.p.right
                  z = z.p                                // case 2
                  LEFT-ROTATE(T, z)                      // case 2
              z.p.color = BLACK                          // case 3
              z.p.p.color = RED                          // case 3
              RIGHT-ROTATE(T, z.p.p)                     // case 3
      else (same as then clause with "right" and "left" exchanged)
  T.root.color = BLACK
```

Loop invariant:

At the start of each iteration of the **while** loop,

a. z is red.

b. There is at most one red-black violation:

- Property 2: z is a red root, or
- Property 4: z and $z.p$ are both red.

[The book has a third part of the loop invariant, but we omit it for lecture.]

Initialization: We've already seen why the loop invariant holds initially.

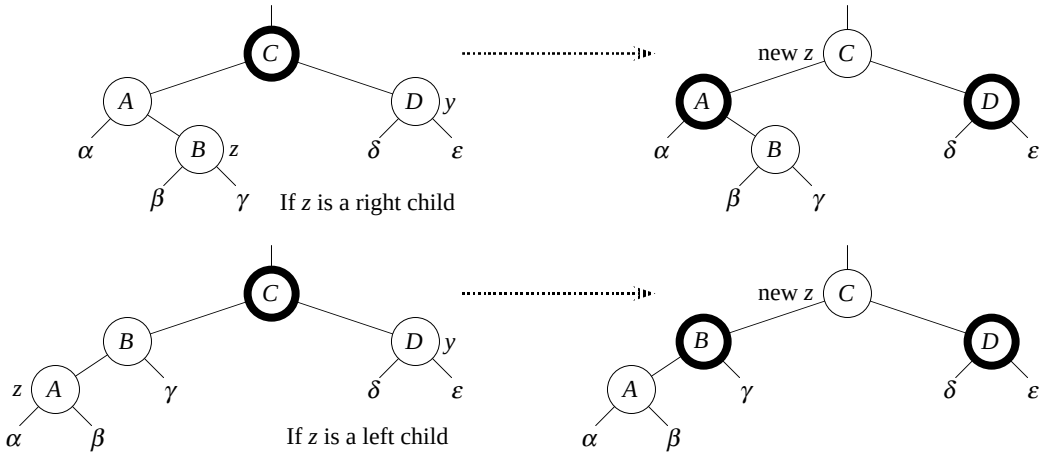
Termination: The loop terminates because $z.p$ is black. Hence, property 4 is OK. Only property 2 might be violated, and the last line fixes it.

Maintenance: We drop out when z is the root (since then $z.p$ is the sentinel $T.nil$, which is black). When we start the loop body, the only violation is of property 4.

There are 6 cases, 3 of which are symmetric to the other 3. The cases are not mutually exclusive. We'll consider cases in which $z.p$ is a left child.

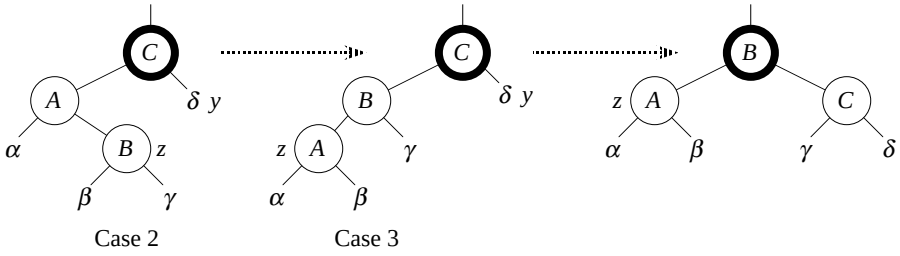
Let y be z's uncle ($z.p$'s sibling).

Case 1: y is red



- $z.p.p$ (z's grandparent) must be black, since z and $z.p$ are both red and there are no other violations of property 4.
- Make $z.p$ and y black $\Rightarrow$ now z and $z.p$ are not both red. But property 5 might now be violated.
- Make $z.p.p$ red $\Rightarrow$ restores property 5.
- The next iteration has $z.p.p$ as the new z (i.e., z moves up 2 levels).

Case 2: y is black, z is a right child



- Left rotate around $z.p$ $\Rightarrow$ now z is a left child, and both z and $z.p$ are red.
- Takes us immediately to case 3.

Case 3: y is black, z is a left child

- Make $z.p$ black and $z.p.p$ red.
- Then right rotate on $z.p.p$.
- No longer have 2 reds in a row.
- $z.p$ is now black $\Rightarrow$ no more iterations.

Analysis

$O(\lg n)$ time to get through RB-INSERT up to the call of RB-INSERT-FIXUP.

Within RB-INSERT-FIXUP:

- Each iteration takes $O(1)$ time.
- Each iteration is either the last one or it moves z up 2 levels.
- $O(\lg n)$ levels $\Rightarrow$ $O(\lg n)$ time.
- Also note that there are at most 2 rotations overall.

Thus, insertion into a red-black tree takes $O(\lg n)$ time.

Deletion

[Because deletion from a binary search tree changed in the third edition, so did deletion from a red-black tree. As with deletion from a binary search tree, the node z deleted from a red-black tree is always the node z passed to the deletion procedure.]

Based on the TREE-DELETE procedure for binary search trees:

```
RB-DELETE(T, z)
 y = z
 y-original-color = y.color
 if z.left == T.nil
      x = z.right
      RB-TRANSPLANT(T, z, z.right)
 elseif z.right == T.nil
      x = z.left
      RB-TRANSPLANT(T, z, z.left)
 else y = TREE-MINIMUM(z.right)
      y-original-color = y.color
      x = y.right
      if y.p == z
           x.p = y
      else RB-TRANSPLANT(T, y, y.right)
           y.right = z.right
           y.right.p = y
      RB-TRANSPLANT(T, z, y)
      y.left = z.left
      y.left.p = y
      y.color = z.color
 if y-original-color == BLACK
      RB-DELETE-FIXUP(T, x)
```

RB-DELETE calls a special version of TRANSPLANT (used in deletion from binary search trees), customized for red-black trees:

RB-TRANSPLANT(T, u, v)

```
if u.p == T.nil
    T.root = v
elseif u == u.p.left
    u.p.left = v
else u.p.right = v
v.p = u.p
```

Differences between RB-TRANSPLANT and TRANSPLANT:

- RB-TRANSPLANT references the sentinel $T.nil$ instead of NIL.
- Assignment to $v.p$ occurs even if v points to the sentinel. In fact, we exploit the ability to assign to $v.p$ when v points to the sentinel.

RB-DELETE has almost twice as many lines as TREE-DELETE, but you can find each line of TREE-DELETE within RB-DELETE (with NIL replaced by $T.nil$ and calls to TRANSPLANT replaced by calls to RB-TRANSPLANT).

Differences between RB-DELETE and TREE-DELETE:

- y is the node either removed from the tree (when z has fewer than 2 children) or moved within the tree (when z has 2 children).
- Need to save y's original color (in *y-original-color*) to test it at the end, because if it's black, then removing or moving y could cause red-black properties to be violated.
- x is the node that moves into y's original position. It's either y's only child, or $T.nil$ if y has no children.
- Sets $x.p$ to point to the original position of y's parent, even if $x = T.nil$. $x.p$ is set in one of two ways:
  - If z is not y's original parent, $x.p$ is set in the last line of RB-TRANSPLANT.
  - If z is y's original parent, then y will move up to take z's position in the tree. The assignment $x.p = y$ makes $x.p$ point to the original position of y's parent, even if x is $T.nil$.
- If y's original color was black, the changes to the tree structure might cause red-black properties to be violated, and we call RB-DELETE-FIXUP at the end to resolve the violations.

If y was originally black, what violations of red-black properties could arise?

1. No violation.
2. If y is the root and x is red, then the root has become red.
3. No violation.
4. Violation if $x.p$ and x are both red.
5. Any simple path containing y now has 1 fewer black node.
   - Correct by giving x an "extra black."
   - Add 1 to count of black nodes on paths containing x.
   - Now property 5 is OK, but property 1 is not.

- x is either ***doubly black*** (if $x.color = \text{BLACK}$) or ***red & black*** (if $x.color = \text{RED}$).
- The attribute $x.color$ is still either RED or BLACK. No new values for *color* attribute.
- In other words, the extra blackness on a node is by virtue of x pointing to the node.

Remove the violations by calling RB-DELETE-FIXUP:

```
RB-DELETE-FIXUP(T, x)
  while x ≠ T.root and x.color == BLACK
      if x == x.p.left
          w = x.p.right
          if w.color == RED
              w.color = BLACK                          // case 1
              x.p.color = RED                          // case 1
              LEFT-ROTATE(T, x.p)                      // case 1
              w = x.p.right                            // case 1
          if w.left.color == BLACK and w.right.color == BLACK
              w.color = RED                            // case 2
              x = x.p                                  // case 2
          else if w.right.color == BLACK
                  w.left.color = BLACK                 // case 3
                  w.color = RED                        // case 3
                  RIGHT-ROTATE(T, w)                   // case 3
                  w = x.p.right                        // case 3
              w.color = x.p.color                      // case 4
              x.p.color = BLACK                        // case 4
              w.right.color = BLACK                    // case 4
              LEFT-ROTATE(T, x.p)                      // case 4
              x = T.root                               // case 4
      else (same as then clause with "right" and "left" exchanged)
  x.color = BLACK
```

Idea

Move the extra black up the tree until

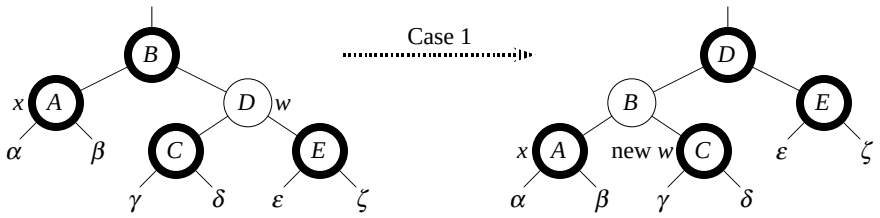
- x points to a red & black node $\Rightarrow$ turn it into a black node,
- x points to the root $\Rightarrow$ just remove the extra black, or
- we can do certain rotations and recolorings and finish.

Within the **while** loop:

- x always points to a nonroot doubly black node.
- w is x's sibling.
- w cannot be $T.nil$, since that would violate property 5 at $x.p$.

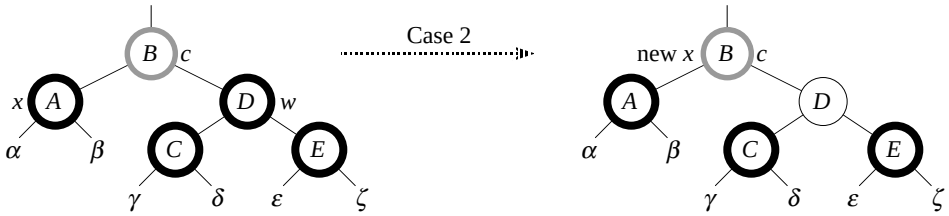
There are 8 cases, 4 of which are symmetric to the other 4. As with insertion, the cases are not mutually exclusive. We'll look at cases in which x is a left child.

Case 1: w is red



- w must have black children.
- Make w black and $x.p$ red.
- Then left rotate on $x.p$.
- New sibling of x was a child of w before rotation $\Rightarrow$ must be black.
- Go immediately to case 2, 3, or 4.

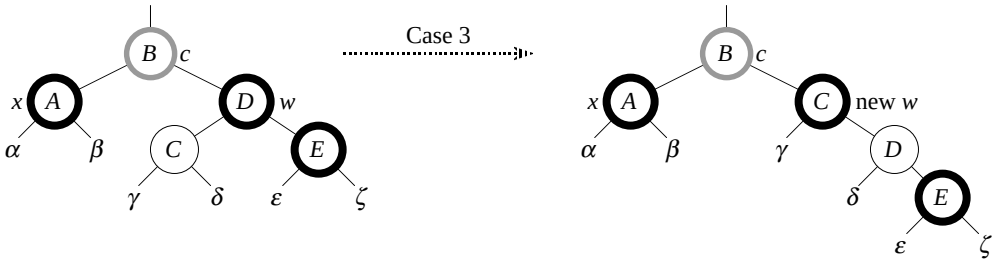
Case 2: w is black and both of w's children are black



[Node with gray outline is of unknown color, denoted by c.]

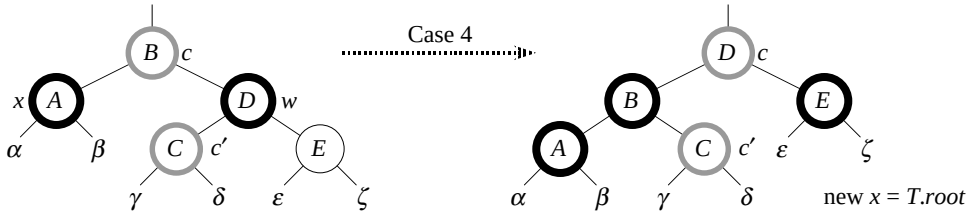
- Take 1 black off x ($\Rightarrow$ singly black) and off w ($\Rightarrow$ red).
- Move that black to $x.p$.
- Do the next iteration with $x.p$ as the new x.
- If entered this case from case 1, then $x.p$ was red $\Rightarrow$ new x is red & black $\Rightarrow$ color attribute of new x is RED $\Rightarrow$ loop terminates. Then new x is made black in the last line.

Case 3: w is black, w's left child is red, and w's right child is black



- Make w red and w's left child black.
- Then right rotate on w.
- New sibling w of x is black with a red right child $\Rightarrow$ case 4.

Case 4: w is black, w's left child is black, and w's right child is red



[Now there are two nodes of unknown colors, denoted by c and c'.]

- Make w be $x.p$'s color (c).
- Make $x.p$ black and w's right child black.
- Then left rotate on $x.p$.
- Remove the extra black on x ($\Rightarrow$ x is now singly black) without violating any red-black properties.
- All done. Setting x to root causes the loop to terminate.

Analysis

$O(\lg n)$ time to get through RB-DELETE up to the call of RB-DELETE-FIXUP.

Within RB-DELETE-FIXUP:

- Case 2 is the only case in which more iterations occur.
  - x moves up 1 level.
  - Hence, $O(\lg n)$ iterations.
- Each of cases 1, 3, and 4 has 1 rotation $\Rightarrow \leq 3$ rotations in all.
- Hence, $O(\lg n)$ time.

[In Chapter 14, we'll see a theorem that relies on red-black tree operations causing at most a constant number of rotations. This is where red-black trees enjoy an advantage over AVL trees: in the worst case, an operation on an n-node AVL tree causes $\Omega(\lg n)$ rotations.]

Solutions for Chapter 13: Red-Black Trees

Solution to Exercise 13.1-3

If we color the root of a relaxed red-black tree black but make no other changes, the resulting tree is a red-black tree. Not even any black-heights change.

Solution to Exercise 13.1-4

This solution is also posted publicly

After absorbing each red node into its black parent, the degree of each node black node is

- 2, if both children were already black,
- 3, if one child was black and one was red, or
- 4, if both children were red.

All leaves of the resulting tree have the same depth.

Solution to Exercise 13.1-5

This solution is also posted publicly

In the longest path, at least every other node is black. In the shortest path, at most every node is black. Since the two paths contain equal numbers of black nodes, the length of the longest path is at most twice the length of the shortest path.

We can say this more precisely, as follows:

Since every path contains $\text{bh}(x)$ black nodes, even the shortest path from x to a descendant leaf has length at least $\text{bh}(x)$. By definition, the longest path from x to a descendant leaf has length $\text{height}(x)$. Since the longest path has $\text{bh}(x)$ black nodes and at least half the nodes on the longest path are black (by property 4), $\text{bh}(x) \geq \text{height}(x)/2$, so

$$\text{length of longest path} = \text{height}(x) \leq 2 \cdot \text{bh}(x) \leq \text{ twice length of shortest path .}$$

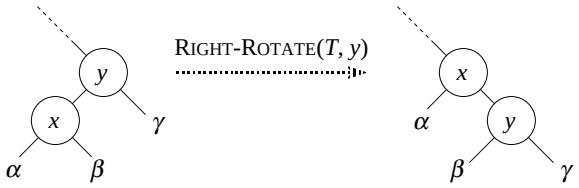
Solution to Exercise 13.2-4

Since the exercise asks about binary search trees rather than the more specific red-black trees, we assume here that leaves are full-fledged nodes, and we ignore the sentinels.

Taking the book's hint, we start by showing that with at most $n - 1$ right rotations, we can convert any binary search tree into one that is just a right-going chain.

The idea is simple. Let us define the ***right spine*** as the root and all descendants of the root that are reachable by following only *right* pointers from the root. A binary search tree that is just a right-going chain has all n nodes in the right spine.

As long as the tree is not just a right spine, repeatedly find some node y on the right spine that has a non-leaf left child x and then perform a right rotation on y:



(In the above figure, note that any of α, β, and γ can be an empty subtree.)

Observe that this right rotation adds x to the right spine, and no other nodes leave the right spine. Thus, this right rotation increases the number of nodes in the right spine by 1. Any binary search tree starts out with at least one node—the root—in the right spine. Moreover, if there are any nodes not on the right spine, then at least one such node has a parent on the right spine. Thus, at most $n - 1$ right rotations are needed to put all nodes in the right spine, so that the tree consists of a single right-going chain.

If we knew the sequence of right rotations that transforms an arbitrary binary search tree T to a single right-going chain T', then we could perform this sequence in reverse—turning each right rotation into its inverse left rotation—to transform T' back into T.

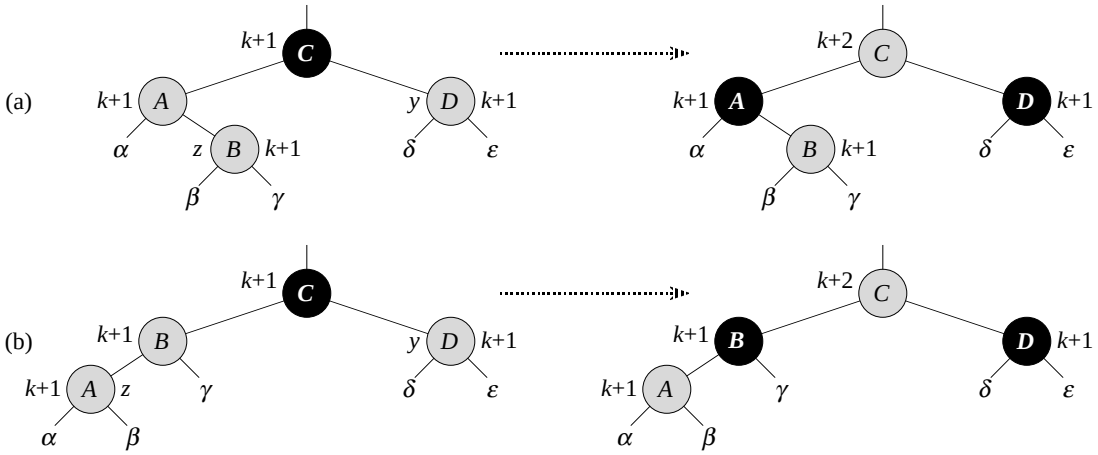
Therefore, here is how we can transform any binary search tree T_1 into any other binary search tree T_2. Let T' be the unique right-going chain consisting of the nodes of T_1 (which is the same as the nodes of T_2). Let $r = \langle r_1, r_2, \ldots, r_k \rangle$ be a sequence of right rotations that transforms T_1 to T', and let $r' = \langle r'_1, r'_2, \ldots, r'_{k'} \rangle$ be a sequence of right rotations that transforms T_2 to T'. We know that there exist sequences r and r' with $k, k' \le n - 1$. For each right rotation r'_i, let l'_i be the corresponding inverse left rotation. Then the sequence $\langle r_1, r_2, \ldots, r_k, l'_{k'}, l'_{k'-1}, \ldots, l'_2, l'_1 \rangle$ transforms T_1 to T_2 in at most $2n - 2$ rotations.

Solution to Exercise 13.3-3

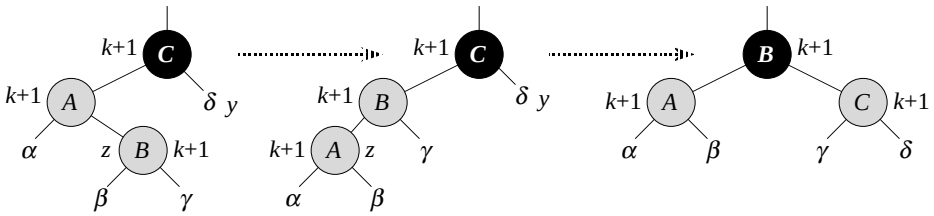
This solution is also posted publicly

In Figure 13.5, nodes A, B, and D have black-height $k + 1$ in all cases, because each of their subtrees has black-height k and a black root. Node C has black-

height $k+1$ on the left (because its red children have black-height $k+1$) and black-height $k+2$ on the right (because its black children have black-height $k+1$).



In Figure 13.6, nodes A, B, and C have black-height $k+1$ in all cases. At left and in the middle, each of A's and B's subtrees has black-height k and a black root, while C has one such subtree and a red child with black-height $k+1$. At the right, each of A's and C's subtrees has black-height k and a black root, while B's red children each have black-height $k+1$.



Case 2 Case 3

Property 5 is preserved by the transformations. We have shown above that the black-height is well-defined within the subtrees pictured, so property 5 is preserved within those subtrees. Property 5 is preserved for the tree containing the subtrees pictured, because every path through these subtrees to a leaf contributes $k+2$ black nodes.

Solution to Exercise 13.3-4

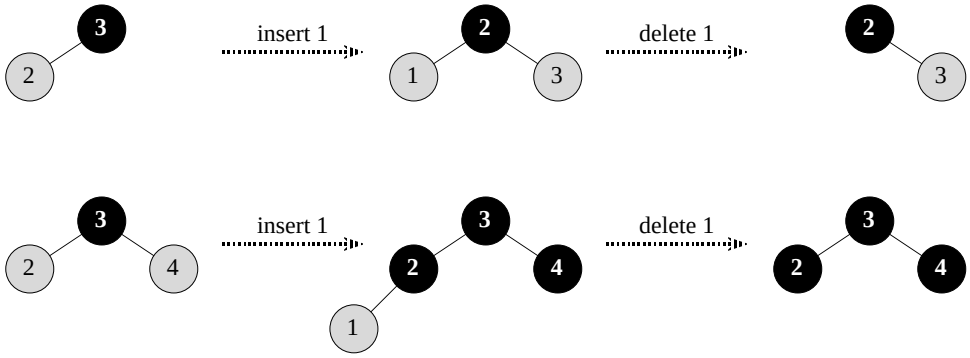
Colors are set to red only in cases 1 and 3, and in both situations, it is $z.p.p$ that is reddened. If $z.p.p$ is the sentinel, then $z.p$ is the root. By part (b) of the loop invariant and line 1 of RB-INSERT-FIXUP, if $z.p$ is the root, then we have dropped out of the loop. The only subtlety is in case 2, where we set $z = z.p$ before coloring $z.p.p$ red. Because we rotate before the recoloring, the identity of $z.p.p$ is the same before and after case 2, so there's no problem.

Solution to Exercise 13.4-6

Case 1 occurs only if x's sibling w is red. If $x.p$ were red, then there would be two reds in a row, namely $x.p$ (which is also $w.p$) and w, and we would have had these two reds in a row even before calling RB-DELETE.

Solution to Exercise 13.4-7

No, the red-black tree will not necessarily be the same. Here are two examples: one in which the tree's shape changes, and one in which the shape remains the same but the node colors change.



Solution to Problem 13-1

This solution is also posted publicly

a. When inserting key k, all nodes on the path from the root to the added node (a new leaf) must change, since the need for a new child pointer propagates up from the new node to all of its ancestors.

When deleting a node, let y be the node actually removed and z be the node given to the delete procedure.

- If z has at most one child, it will be spliced out, so that all ancestors of z will be changed. (As with insertion, the need for a new child pointer propagates up from the removed node.)
- If z has two children, then its successor y will be spliced out and moved to z's position. Therefore all ancestors of both z and y must be changed. Because z is an ancestor of y, we can just say that all ancestors of y must be changed.

In either case, y's children (if any) are unchanged, because we have assumed that there is no parent attribute.

b. We assume that we can call two procedures:

- MAKE-NEW-NODE(k) creates a new node whose *key* attribute has value k and with *left* and *right* attributes NIL, and it returns a pointer to the new node.
- COPY-NODE(x) creates a new node whose *key*, *left*, and *right* attributes have the same values as those of node x, and it returns a pointer to the new node.

Here are two ways to write PERSISTENT-TREE-INSERT. The first is a version of TREE-INSERT, modified to create new nodes along the path to where the new node will go, and to not use parent attributes. It returns the root of the new tree.

```
PERSISTENT-TREE-INSERT(T, k)
z = MAKE-NEW-NODE(k)
new-root = COPY-NODE(T.root)
y = NIL
x = new-root
while x ≠ NIL
    y = x
    if z.key < x.key
        x = COPY-NODE(x.left)
        y.left = x
    else x = COPY-NODE(x.right)
        y.right = x
if y == NIL
    new-root = z
elseif z.key < y.key
    y.left = z
else y.right = z
return new-root
```

The second is a rather elegant recursive procedure. The initial call should have $T.root$ as its first argument. It returns the root of the new tree.

```
PERSISTENT-TREE-INSERT(r, k)
if r == NIL
    x = MAKE-NEW-NODE(k)
else x = COPY-NODE(r)
    if k < r.key
        x.left = PERSISTENT-TREE-INSERT(r.left, k)
    else x.right = PERSISTENT-TREE-INSERT(r.right, k)
return x
```

c. Like TREE-INSERT, PERSISTENT-TREE-INSERT does a constant amount of work at each node along the path from the root to the new node. Since the length of the path is at most h, it takes $O(h)$ time.

Since it allocates a new node (a constant amount of space) for each ancestor of the inserted node, it also needs $O(h)$ space.

d. If there were parent attributes, then because of the new root, every node of the tree would have to be copied when a new node is inserted. To see why, observe that the children of the root would change to point to the new root, then their children would change to point to them, and so on. Since there are n nodes, this change would cause insertion to create $\Omega(n)$ new nodes and to take $\Omega(n)$ time.

e. From parts (a) and (c), we know that insertion into a persistent binary search tree of height h, like insertion into an ordinary binary search tree, takes worst-case time $O(h)$. A red-black tree has $h = O(\lg n)$, so insertion into an ordinary red-black tree takes $O(\lg n)$ time. We need to show that if the red-black tree is persistent, insertion can still be done in $O(\lg n)$ time. To do this, we will need to show two things:

- How to still find the parent pointers we need in $O(1)$ time without using a parent attribute. We cannot use a parent attribute because a persistent tree with parent attributes uses $\Omega(n)$ time for insertion (by part (d)).
- That the additional node changes made during red-black tree operations (by rotation and recoloring) don't cause more than $O(\lg n)$ additional nodes to change.

Each parent pointer needed during insertion can be found in $O(1)$ time without having a parent attribute as follows:

To insert into a red-black tree, we call RB-INSERT, which in turn calls RB-INSERT-FIXUP. Make the same changes to RB-INSERT as we made to TREE-INSERT for persistence. Additionally, as RB-INSERT walks down the tree to find the place to insert the new node, have it build a stack of the nodes it traverses and pass this stack to RB-INSERT-FIXUP. RB-INSERT-FIXUP needs parent pointers to walk back up the same path, and at any given time it needs parent pointers only to find the parent and grandparent of the node it is working on. As RB-INSERT-FIXUP moves up the stack of parents, it needs only parent pointers that are at known locations a constant distance away in the stack. Thus, the parent information can be found in $O(1)$ time, just as if it were stored in a parent attribute.

Rotation and recoloring change nodes as follows:

- RB-INSERT-FIXUP performs at most 2 rotations, and each rotation changes the child pointers in 3 nodes (the node around which we rotate, that node's parent, and one of the children of the node around which we rotate). Thus, at most 6 nodes are directly modified by rotation during RB-INSERT-FIXUP. In a persistent tree, all ancestors of a changed node are copied, so RB-INSERT-FIXUP's rotations take $O(\lg n)$ time to change nodes due to rotation. (Actually, the changed nodes in this case share a single $O(\lg n)$-length path of ancestors.)
- RB-INSERT-FIXUP recolors some of the inserted node's ancestors, which are being changed anyway in persistent insertion, and some children of ancestors (the "uncles" referred to in the algorithm description). There are at most $O(\lg n)$ ancestors, hence at most $O(\lg n)$ color changes of uncles. Recoloring uncles doesn't cause any additional node changes due to persistence, because the ancestors of the uncles are the same nodes (ancestors of

the inserted node) that are being changed anyway due to persistence. Thus, recoloring does not affect the $O(\lg n)$ running time, even with persistence.

We could show similarly that deletion in a persistent tree also takes worst-case time $O(h)$.

- We already saw in part (a) that $O(h)$ nodes change.
- We could write a persistent RB-DELETE procedure that runs in $O(h)$ time, analogous to the changes we made for persistence in insertion. But to do so without using parent pointers we need to walk down the tree to the node to be deleted, to build up a stack of parents as discussed above for insertion. This is a little tricky if the set's keys are not distinct, because in order to find the path to the node to delete—a particular node with a given key—we have to make some changes to how we store things in the tree, so that duplicate keys can be distinguished. The easiest way is to have each key take a second part that is unique, and to use this second part as a tiebreaker when comparing keys.

Then the problem of showing that deletion needs only $O(\lg n)$ time in a persistent red-black tree is the same as for insertion.

- As for insertion, we can show that the parents needed by RB-DELETE-FIXUP can be found in $O(1)$ time (using the same technique as for insertion).
- Also, RB-DELETE-FIXUP performs at most 3 rotations, which as discussed above for insertion requires $O(\lg n)$ time to change nodes due to persistence. It also does $O(\lg n)$ color changes, which (as for insertion) take only $O(\lg n)$ time to change ancestors due to persistence, because the number of copied nodes is $O(\lg n)$.

Lecture Notes for Chapter 14: Augmenting Data Structures

Chapter 14 overview

We'll be looking at methods for *designing* algorithms. In some cases, the design will be intermixed with analysis. In other cases, the analysis is easy, and it's the design that's harder.

Augmenting data structures

- It's unusual to have to design an all-new data structure from scratch.
- It's more common to take a data structure that you know and store additional information in it.
- With the new information, the data structure can support new operations.
- But you have to figure out how to *correctly maintain* the new information *without loss of efficiency*.

We'll look at a couple of situations in which we augment red-black trees.

Dynamic order statistics

We want to support the usual dynamic-set operations from R-B trees, plus:

- OS-SELECT(x, i): return pointer to node containing the ith smallest key of the subtree rooted at x.
- OS-RANK(T, x): return the rank of x in the linear order determined by an inorder walk of T.

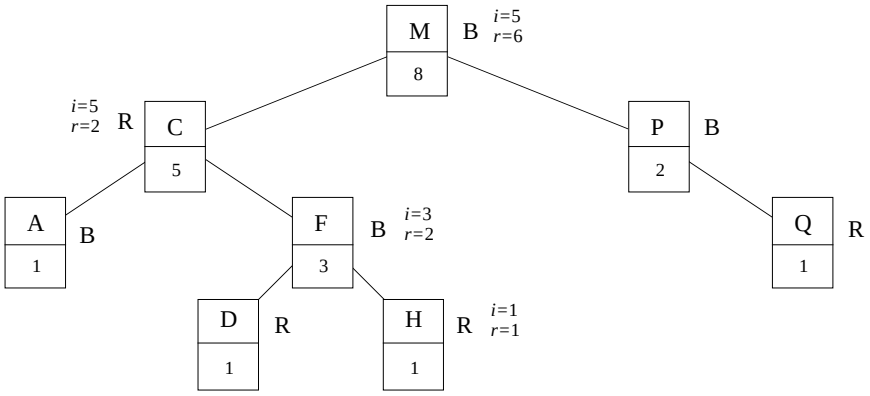
Augment by storing in each node x:

$x.size$ = # of nodes in subtree rooted at x .

- Includes x itself.
- Does not include leaves (sentinels).

Define for sentinel $T.nil.size = 0$.

Then $x.size = x.left.size + x.right.size + 1$.



*[**Example above:** Ignore colors, but legal coloring shown with "R" and "B" notations. Values of i and r are for the example below.]*

Note: OK for keys to not be distinct. Rank is defined with respect to position in inorder walk. So if we changed D to C, rank of original C is 2, rank of D changed to C is 3.

OS-SELECT(x, i)

$r = x.left.size + 1$
if $i == r$
  return x
elseif $i < r$
  return OS-SELECT$(x.left, i)$
else return OS-SELECT$(x.right, i - r)$

Initial call: OS-SELECT$(T.root, i)$

Try OS-SELECT$(T.root, 5)$. *[Values shown in figure above. Returns node whose key is H.]*

Correctness

$r =$ rank of x within subtree rooted at x.

- If $i = r$, then we want x.
- If $i < r$, then ith smallest element is in x's left subtree, and we want the ith smallest element in the subtree.
- If $i > r$, then ith smallest element is in x's right subtree, but subtract off the r elements in x's subtree that precede those in x's right subtree.
- Like the randomized SELECT algorithm.

Analysis

Each recursive call goes down one level. Since R-B tree has $O(\lg n)$ levels, have $O(\lg n)$ calls $\Rightarrow O(\lg n)$ time.

OS-RANK(T, x)

```
r = x.left.size + 1
y = x
while y ≠ T.root
    if y == y.p.right
        r = r + y.p.left.size + 1
    y = y.p
return r
```

Demo: Node D.

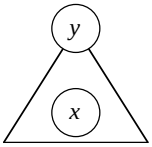
Why does this work?

> **Loop invariant:** At start of each iteration of **while** loop, r = rank of $x.key$ in subtree rooted at y.

Initialization: Initially, r = rank of $x.key$ in subtree rooted at x, and $y = x$.

Termination: Loop terminates when $y = T.root \Rightarrow$ subtree rooted at y is entire tree. Therefore, r = rank of $x.key$ in entire tree.

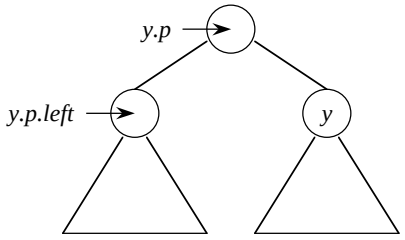
Maintenance: At end of each iteration, set $y = y.p$. So, show that if r = rank of $x.key$ in subtree rooted at y at start of loop body, then r = rank of $x.key$ in subtree rooted at $y.p$ at end of loop body.



[r = # of nodes in subtree rooted at y preceding x in inorder walk]

Must add nodes in y's sibling's subtree.

- If y is a left child, its sibling's subtree follows all nodes in y's subtree $\Rightarrow$ don't change r.
- If y is a right child, all nodes in y's sibling's subtree precede all nodes in y's subtree $\Rightarrow$ add size of y's sibling's subtree, plus 1 for $y.p$, into r.



Analysis

y goes up one level in each iteration $\Rightarrow O(\lg n)$ time.

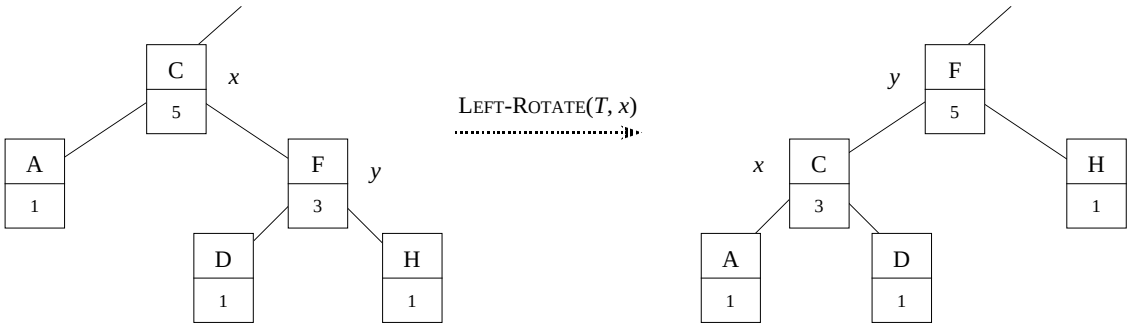
Maintaining subtree sizes

- Need to maintain *size* attributes during insert and delete operations.
- Need to maintain them efficiently. Otherwise, might have to recompute them all, at a cost of $\Omega(n)$.

Will see how to maintain without increasing $O(\lg n)$ time for insert and delete.

Insert

- During pass downward, we know that the new node will be a descendant of each node we visit, and only of these nodes. Therefore, increment *size* attribute of each node visited.
- Then there's the fixup pass:
  - Goes up the tree.
  - Changes colors $O(\lg n)$ times.
  - Performs ≤ 2 rotations.
- Color changes don't affect subtree sizes.
- Rotations do!
- But we can determine new sizes based on old sizes and sizes of children.



$$y.size = x.size$$
$$x.size = x.left.size + x.right.size + 1$$

- Similar for right rotation.
- Therefore, can update in $O(1)$ time per rotation $\Rightarrow$ $O(1)$ time spent updating *size* attributes during fixup.
- Therefore, $O(\lg n)$ to insert.

Delete

Also 2 phases:

1. Splice out some node y.
2. Fixup.

After splicing out y, traverse a path $y \to root$, decrementing *size* in each node on path. $O(\lg n)$ time.

During fixup, like insertion, only color changes and rotations.

- ≤ 3 rotations $\Rightarrow O(1)$ time spent updating *size* attributes during fixup.
- Therefore, $O(\lg n)$ to delete.

Done!

Methodology for augmenting a data structure

1. Choose an underlying data structure.
2. Determine additional information to maintain.
3. Verify that we can maintain additional information for existing data structure operations.
4. Develop new operations.

Don't need to do these steps in strict order! Usually do a little of each, in parallel.

How did we do them for OS trees?

1. R-B tree.
2. $x.size$.
3. Showed how to maintain *size* during insert and delete.
4. Developed OS-SELECT and OS-RANK.

Red-black trees are particularly amenable to augmentation.

Theorem

Augment a R-B tree with attribute f, where $x.f$ depends only on information in x, $x.left$, and $x.right$ (including $x.left.f$ and $x.right.f$). Then can maintain values of f in all nodes during insert and delete without affecting $O(\lg n)$ performance.

Proof Since $x.f$ depends only on x and its children, when we alter information in x, changes propagate only upward (to $x.p$, $x.p.p$, $x.p.p.p$, $\ldots$, *root*).

Height = $O(\lg n) \Rightarrow O(\lg n)$ updates, at $O(1)$ each.

Insertion

Insert a node as child of existing node. Even if can't update f on way down, can go up from inserted node to update f. During fixup, only changes come from color changes (no effect on f) and rotations. Each rotation affects f of ≤ 3 nodes (x, y, and parent), and can recompute each in $O(1)$ time. Then, if necessary, propagate changes up the tree. Therefore, $O(\lg n)$ time per rotation. Since ≤ 2 rotations, $O(\lg n)$ time to update f during fixup.

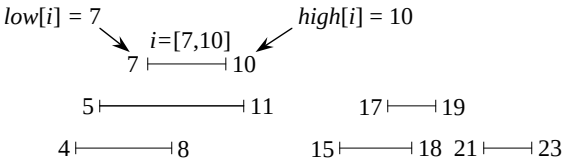
Delete

Same idea. After splicing out a node, go up from there to update f. Fixup has ≤ 3 rotations. $O(\lg n)$ per rotation $\Rightarrow$ $O(\lg n)$ to update f during fixup. ■ (theorem)

For some attributes, can get away with $O(1)$ per rotation. Example: *size* attribute.

Interval trees

Maintain a set of intervals. For instance, time intervals.



[leave on board]

Operations

- INTERVAL-INSERT(T, x): $x.int$ already filled in.
- INTERVAL-DELETE(T, x)
- INTERVAL-SEARCH(T, i): return pointer to a node x in T such that $x.int$ overlaps interval i. Any overlapping node in T is OK. Return pointer to sentinel $T.nil$ if no overlapping node in T.

Interval i has $i.low$, $i.high$.

i and j overlap if and only if
$i.low \leq j.high$ and $j.low \leq i.high$.

(Go through examples of proper inclusion, overlap without proper inclusion, no overlap.)

Another way: i and j *don't* overlap if and only if
$i.low > j.high$ or $j.low > i.high$.
[leave this on board]

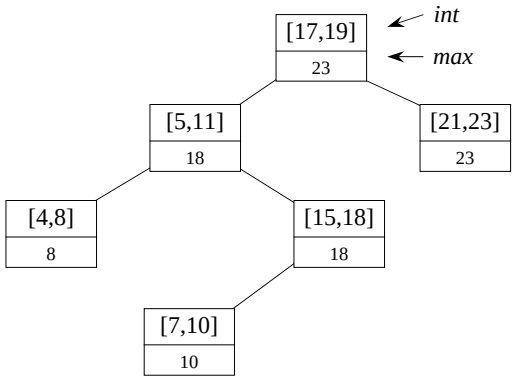
Recall the 4-part methodology.

For interval trees

1. Use R-B trees.
   - Each node x contains interval $x.int$.
   - Key is low endpoint ($x.int.low$).
   - Inorder walk would list intervals sorted by low endpoint.

2. Each node x contains

   $x.max$ = max endpoint value in subtree rooted at x .



   [leave on board]

$$x.max = max \begin{cases} x.int.high\ , \\ x.left.max\ , \\ x.right.max \end{cases}$$

   Could $x.left.max > x.right.max$? Sure. Position in tree is determined only by low endpoints, not high endpoints.

3. Maintaining the information.

   - This is easy—$x.max$ depends only on:
     - information in x: $x.int.high$
     - information in $x.left$: $x.left.max$
     - information in $x.right$: $x.right.max$
   - Apply the theorem.
   - In fact, can update *max* on way down during insertion, and in $O(1)$ time per rotation.

4. Developing new operations.

   INTERVAL-SEARCH(T, i)

```
x = T.root
while x ≠ T.nil and i does not overlap x.int
    if x.left ≠ T.nil and x.left.max ≥ i.low
        x = x.left
    else x = x.right
return x
```

Examples

Search for $[14, 16]$ and $[12, 14]$.

Time

$O(\lg n)$.

Correctness

Key idea: need check only 1 of node's 2 children.

Theorem

If search goes right, then either:

- There is an overlap in right subtree, or
- There is no overlap in either subtree.

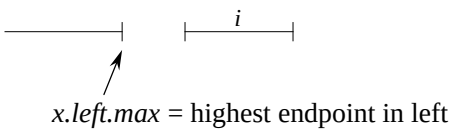
If search goes left, then either:

- There is an overlap in left subtree, or
- There is no overlap in either subtree.

Proof If search goes right:

- If there is an overlap in right subtree, done.
- If there is no overlap in right, show there is no overlap in left. Went right because
  - $x.left = T.nil \Rightarrow$ no overlap in left.

    OR
  - $x.left.max < i.low \Rightarrow$ no overlap in left.



x.left.max = highest endpoint in left

If search goes left:

- If there is an overlap in left subtree, done.
- If there is no overlap in left, show there is no overlap in right.
  - Went left because:

    $i.low \leq x.left.max$

    $= j.high$ for some j in left subtree .
  - Since there is no overlap in left, i and j don't overlap.
  - Refer back to: no overlap if

    $i.low > j.high$ or $j.low > i.high$.
  - Since $i.low \leq j.high$, must have $j.low > i.high$.
  - Now consider *any* interval k in *right* subtree.
  - Because keys are low endpoint,

    $\underbrace{j.low}_{\text{in left}} \leq \underbrace{k.low}_{\text{in right}}$.
  - Therefore, $i.high < j.low \leq k.low$.
  - Therefore, $i.high < k.low$.
  - Therefore, i and k do not overlap. ■ (theorem)

Solutions for Chapter 14: Augmenting Data Structures

Solution to Exercise 14.1-5

Given an element x in an n-node order-statistic tree T and a natural number i, the following procedure retrieves the ith successor of x in the linear order of T:

OS-SUCCESSOR(T, x, i)
  $r =$ OS-RANK(T, x)
  $s = r + i$
  return OS-SELECT$(T.root, s)$

Since OS-RANK and OS-SELECT each take $O(\lg n)$ time, so does the procedure OS-SUCCESSOR.

Solution to Exercise 14.1-6

When inserting node z, we search down the tree for the proper place for z. For each node x on this path, add 1 to $x.rank$ if z is inserted within x's left subtree, and leave $x.rank$ unchanged if z is inserted within x's right subtree. Similarly when deleting, subtract 1 from $x.rank$ whenever the spliced-out node had been in x's left subtree.

We also need to handle the rotations that occur during the fixup procedures for insertion and deletion. Consider a left rotation on node x, where the pre-rotation right child of x is y (so that x becomes y's left child after the left rotation). We leave $x.rank$ unchanged, and letting $r = y.rank$ before the rotation, we set $y.rank = r + x.rank$. Right rotations are handled in an analogous manner.

Solution to Exercise 14.1-7

This solution is also posted publicly

Let $A[1..n]$ be the array of n distinct numbers.

One way to count the inversions is to add up, for each element, the number of larger elements that precede it in the array:

$$\#\text{ of inversions} = \sum_{j=1}^{n} |Inv(j)| \ ,$$

where $Inv(j) = \{i : i < j \text{ and } A[i] > A[j]\}$.

Note that $|Inv(j)|$ is related to $A[j]$'s rank in the subarray $A[1..j]$ because the elements in $Inv(j)$ are the reason that $A[j]$ is not positioned according to its rank. Let $r(j)$ be the rank of $A[j]$ in $A[1..j]$. Then $j = r(j) + |Inv(j)|$, so we can compute

$$|Inv(j)| = j - r(j)$$

by inserting $A[1], \ldots, A[n]$ into an order-statistic tree and using OS-RANK to find the rank of each $A[j]$ in the tree immediately after it is inserted into the tree. (This OS-RANK value is $r(j)$.)

Insertion and OS-RANK each take $O(\lg n)$ time, and so the total time for n elements is $O(n \lg n)$.

Solution to Exercise 14.2-2

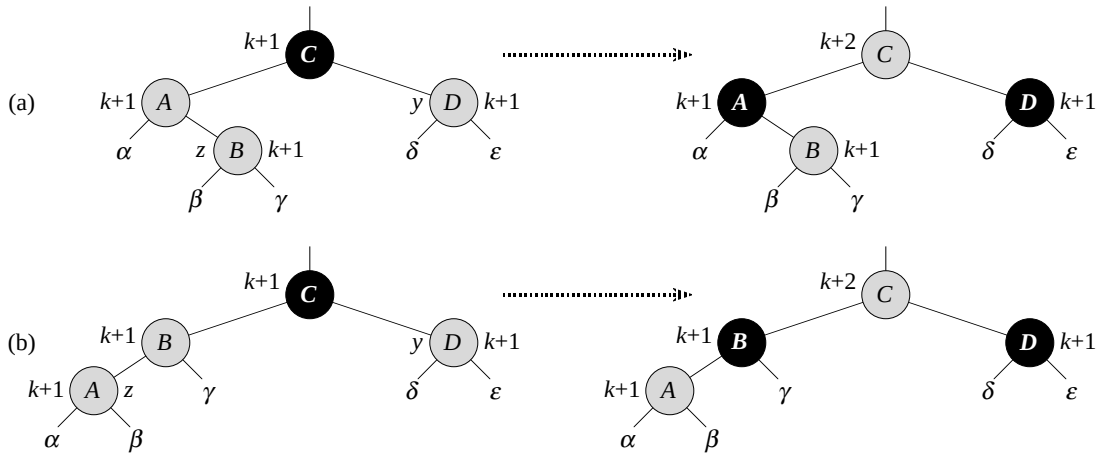
This solution is also posted publicly

Yes, we can maintain black-heights as attributes in the nodes of a red-black tree without affecting the asymptotic performance of the red-black tree operations. We appeal to Theorem 14.1, because the black-height of a node can be computed from the information at the node and its two children. Actually, the black-height can be computed from just one child's information: the black-height of a node is the black-height of a red child, or the black height of a black child plus one. The second child does not need to be checked because of property 5 of red-black trees.

Within the RB-INSERT-FIXUP and RB-DELETE-FIXUP procedures are color changes, each of which potentially cause $O(\lg n)$ black-height changes. Let us show that the color changes of the fixup procedures cause only local black-height changes and thus are constant-time operations. Assume that the black-height of each node x is kept in the attribute $x.bh$.

For RB-INSERT-FIXUP, there are 3 cases to examine.

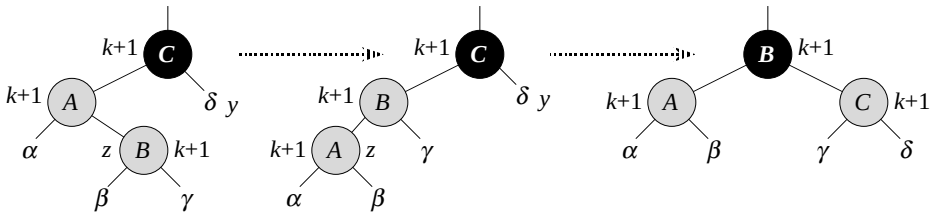
Case 1: z's uncle is red.



- Before color changes, suppose that all subtrees $\alpha, \beta, \gamma, \delta, \epsilon$ have the same black-height k with a black root, so that nodes A, B, C, and D have black-heights of $k + 1$.
- After color changes, the only node whose black-height changed is node C. To fix that, add $z.p.p.bh = z.p.p.bh + 1$ after line 7 in RB-INSERT-FIXUP.
- Since the number of black nodes between $z.p.p$ and z remains the same, nodes above $z.p.p$ are not affected by the color change.

Case 2: z's uncle y is black, and z is a right child.

Case 3: z''s uncle y is black, and z is a left child.



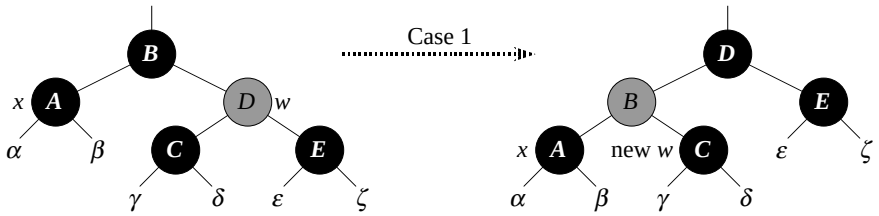
Case 2 Case 3

- With subtrees $\alpha, \beta, \gamma, \delta, \epsilon$ of black-height k, we see that even with color changes and rotations, the black-heights of nodes A, B, and C remain the same ($k + 1$).

Thus, RB-INSERT-FIXUP maintains its original $O(\lg n)$ time.

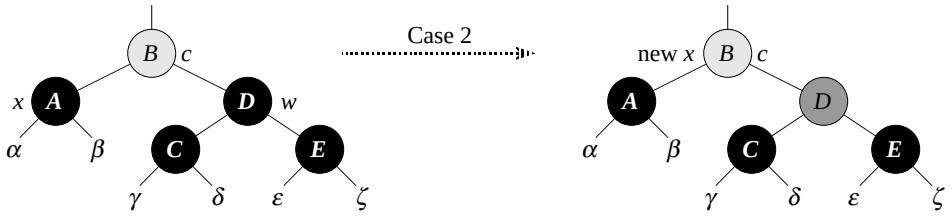
For RB-DELETE-FIXUP, there are 4 cases to examine.

Case 1: x's sibling w is red.



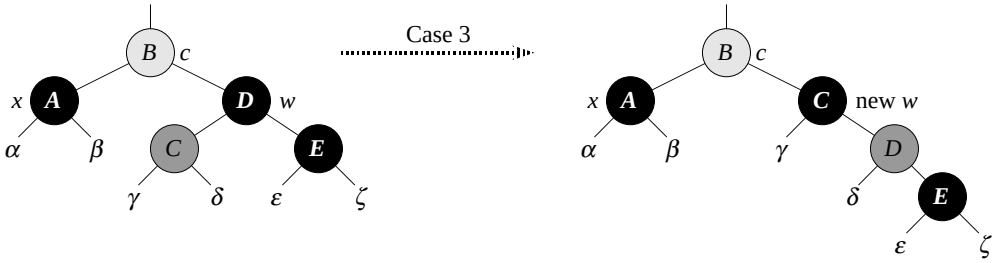
- Even though case 1 changes colors of nodes and does a rotation, black-heights are not changed.
- Case 1 changes the structure of the tree, but waits for cases 2, 3, and 4 to deal with the "extra black" on x.

Case 2: x's sibling w is black, and both of w's children are black.



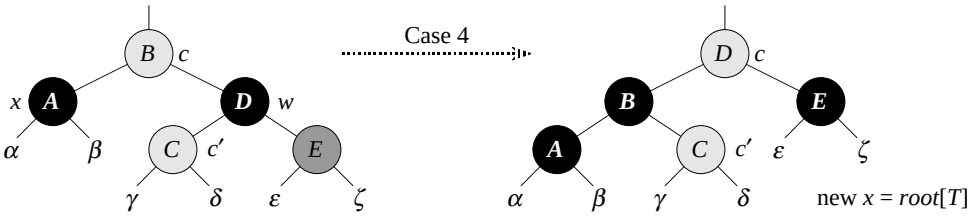
- w is colored red, and x's "extra" black is moved up to $x.p$.
- Now we can add $x.p.bh = x.bh$ after line 10 in RB-DELETE-FIXUP.
- This is a constant-time update. Then, keep looping to deal with the extra black on $x.p$.

Case 3: x's sibling w is black, w's left child is red, and w's right child is black.



- Regardless of the color changes and rotation of this case, the black-heights don't change.
- Case 3 just sets up the structure of the tree, so it can fall correctly into case 4.

Case 4: x's sibling w is black, and w's right child is red.



- Nodes A, C, and E keep the same subtrees, so their black-heights don't change.
- Add these two constant-time assignments in RB-DELETE-FIXUP after line 20:

  $x.p.bh = x.bh + 1$
  $x.p.p.bh = x.p.bh + 1$

- The extra black is taken care of. Loop terminates.

Thus, RB-DELETE-FIXUP maintains its original $O(\lg n)$ time.

Therefore, we conclude that black-heights of nodes can be maintained as attributes in red-black trees without affecting the asymptotic performance of red-black tree operations.

For the second part of the question, no, we cannot maintain node depths without affecting the asymptotic performance of red-black tree operations. The depth of a node depends on the depth of its parent. When the depth of a node changes, the depths of all nodes below it in the tree must be updated. Updating the root node causes $n - 1$ other nodes to be updated, which would mean that operations on the tree that change node depths might not run in $O(n \lg n)$ time.

Solution to Exercise 14.3-3

As it travels down the tree, INTERVAL-SEARCH first checks whether current node x overlaps the query interval i and, if it does not, goes down to either the left or right child. If node x overlaps i, and some node in the right subtree overlaps i, but no node in the left subtree overlaps i, then because the keys are low endpoints, this order of checking (first x, then one child) will return the overlapping interval with the minimum low endpoint. On the other hand, if there is an interval that overlaps i in the left subtree of x, then checking x before the left subtree might cause the procedure to return an interval whose low endpoint is not the minimum of those that overlap i. Therefore, if there is a possibility that the left subtree might contain an interval that overlaps i, we need to check the left subtree first. If there is no overlap in the left subtree but node x overlaps i, then we return x. We check the right subtree under the same conditions as in INTERVAL-SEARCH: the left subtree cannot contain an interval that overlaps i, and node x does not overlap i, either.

Because we might search the left subtree first, it is easier to write the pseudocode to use a recursive procedure MIN-INTERVAL-SEARCH-FROM(T, x, i), which returns the node overlapping i with the minimum low endpoint in the subtree rooted at x, or $T.nil$ if there is no such node.

```
MIN-INTERVAL-SEARCH(T, i)
  return MIN-INTERVAL-SEARCH-FROM(T, T.root, i)
```

```
MIN-INTERVAL-SEARCH-FROM(T, x, i)
  if x.left ≠ T.nil and x.left.max ≥ i.low
      y = MIN-INTERVAL-SEARCH-FROM(T, x.left, i)
      if y ≠ T.nil
          return y
      elseif i overlaps x.int
          return x
      else return T.nil
  elseif i overlaps x.int
      return x
  else return MIN-INTERVAL-SEARCH-FROM(T, x.right, i)
```

The call MIN-INTERVAL-SEARCH(T, i) takes $O(\lg n)$ time, since each recursive call of MIN-INTERVAL-SEARCH-FROM goes one node lower in the tree, and the height of the tree is $O(\lg n)$.

Solution to Exercise 14.3-6

1. Underlying data structure:
   A red-black tree in which the numbers in the set are stored simply as the keys of the nodes.

   SEARCH is then just the ordinary TREE-SEARCH for binary search trees, which runs in $O(\lg n)$ time on red-black trees.
2. Additional information:
   The red-black tree is augmented by the following attributes in each node x:
   - $x.\mathit{min\text{-}gap}$ contains the minimum gap in the subtree rooted at x. It has the magnitude of the difference of the two closest numbers in the subtree rooted at x. If x is a leaf (its children are all $T.\mathit{nil}$), let $x.\mathit{min\text{-}gap} = \infty$.
   - $x.\mathit{min\text{-}val}$ contains the minimum value (key) in the subtree rooted at x.
   - $x.\mathit{max\text{-}val}$ contains the maximum value (key) in the subtree rooted at x.
3. Maintaining the information:
   The three attributes added to the tree can each be computed from information in the node and its children. Hence by Theorem 14.1, they can be maintained during insertion and deletion without affecting the $O(\lg n)$ running time:

$$x.\mathit{min\text{-}val} = \begin{cases} x.\mathit{left}.\mathit{min\text{-}val} & \text{if there is a left subtree ,} \\ x.\mathit{key} & \text{otherwise ,} \end{cases}$$

$$x.\mathit{max\text{-}val} = \begin{cases} x.\mathit{right}.\mathit{max\text{-}val} & \text{if there is a right subtree ,} \\ x.\mathit{key} & \text{otherwise ,} \end{cases}$$

$$x.\mathit{min\text{-}gap} = \min \begin{cases} x.\mathit{left}.\mathit{min\text{-}gap} & (\infty \text{ if no left subtree}) , \\ x.\mathit{right}.\mathit{min\text{-}gap} & (\infty \text{ if no right subtree}) , \\ x.\mathit{key} - x.\mathit{left}.\mathit{max\text{-}val} & (\infty \text{ if no left subtree}) , \\ x.\mathit{right}.\mathit{min\text{-}val} - x.\mathit{key} & (\infty \text{ if no right subtree}) . \end{cases}$$

In fact, the reason for defining the *min-val* and *max-val* attributes is to make it possible to compute *min-gap* from information at the node and its children.

4. New operation:
   MIN-GAP simply returns the *min-gap* stored at the tree root. Thus, its running time is $O(1)$.

   Note that in addition (not asked for in the exercise), it is possible to find the two closest numbers in $O(\lg n)$ time. Starting from the root, look for where the minimum gap (the one stored at the root) came from. At each node x, simulate the computation of $x.$*min-gap* to figure out where $x.$*min-gap* came from. If it came from a subtree's *min-gap* attribute, continue the search in that subtree. If it came from a computation with x's key, then x and that other number are the closest numbers.

Solution to Exercise 14.3-7

This solution is also posted publicly

General idea: Move a sweep line from left to right, while maintaining the set of rectangles currently intersected by the line in an interval tree. The interval tree will organize all rectangles whose x interval includes the current position of the sweep line, and it will be based on the y intervals of the rectangles, so that any overlapping y intervals in the interval tree correspond to overlapping rectangles.

Details:

1. Sort the rectangles by their x-coordinates. (Actually, each rectangle must appear twice in the sorted list—once for its left x-coordinate and once for its right x-coordinate.)
2. Scan the sorted list (from lowest to highest x-coordinate).
   - When an x-coordinate of a left edge is found, check whether the rectangle's y-coordinate interval overlaps an interval in the tree, and insert the rectangle (keyed on its y-coordinate interval) into the tree.
   - When an x-coordinate of a right edge is found, delete the rectangle from the interval tree.

   The interval tree always contains the set of "open" rectangles intersected by the sweep line. If an overlap is ever found in the interval tree, there are overlapping rectangles.

Time: $O(n \lg n)$

- $O(n \lg n)$ to sort the rectangles (we can use merge sort or heap sort).
- $O(n \lg n)$ for interval-tree operations (insert, delete, and check for overlap).

Solution to Problem 14-1

a. Assume for the purpose of contradiction that there is no point of maximum overlap in an endpoint of a segment. The maximum overlap point p is in the

interior of m segments. Actually, p is in the interior of the intersection of those m segments. Now look at one of the endpoints p' of the intersection of the m segments. Point p' has the same overlap as p because it is in the same intersection of m segments, and so p' is also a point of maximum overlap. Moreover, p' is in the endpoint of a segment (otherwise the intersection would not end there), which contradicts our assumption that there is no point of maximum overlap in an endpoint of a segment. Thus, there is always a point of maximum overlap which is an endpoint of one of the segments.

b. Keep a balanced binary search tree of the endpoints. That is, to insert an interval, we insert its endpoints separately. With each left endpoint e, associate a value $p(e) = +1$ (increasing the overlap by 1). With each right endpoint e associate a value $p(e) = -1$ (decreasing the overlap by 1). When multiple endpoints have the same value, insert all the left endpoints with that value before inserting any of the right endpoints with that value.

Here's some intuition. Let $e_1, e_2, \ldots, e_n$ be the sorted sequence of endpoints corresponding to our intervals. Let $s(i, j)$ denote the sum $p(e_i) + p(e_{i+1}) + \cdots + p(e_j)$ for $1 \le i \le j \le n$. We wish to find an i maximizing $s(1, i)$.

For each node x in the tree, let $l(x)$ and $r(x)$ be the indices in the sorted order of the leftmost and rightmost endpoints, respectively, in the subtree rooted at x. Then the subtree rooted at x contains the endpoints $e_{l(x)}, e_{l(x)+1}, \ldots, e_{r(x)}$.

Each node x stores three new attributes. We store $x.v = s(l(x), r(x))$, the sum of the values of all nodes in the subtree rooted at x. We also store $x.m$, the maximum value obtained by the expression $s(l(x), i)$ for any i in $\{l(x), l(x)+1, \ldots, r(x)\}$. Finally, we store $x.o$ as the value of i for which $x.m$ achieves its maximum. For the sentinel, we define $T.nil.v = T.nil.m = 0$.

We can compute these attributes in a bottom-up fashion to satisfy the requirements of Theorem 14.1:

$$x.v = x.left.v + p(x) + x.right.v ,$$

$$x.m = \max \begin{cases} x.left.m & \text{(max is in } x\text{'s left subtree)} , \\ x.left.v + p(x) & \text{(max is at } x\text{)} , \\ x.left.v + p(x) + x.right.m & \text{(max is in } x\text{'s right subtree)} . \end{cases}$$

Computing $x.v$ is straightforward. Computing $x.m$ bears further explanation. Recall that it is the maximum value of the sum of the p values for the nodes in the subtree rooted at x, starting at the node for $e_{l(x)}$, which is the leftmost endpoint in x's subtree, and ending at any node for e_i in x's subtree. The endpoint e_i that maximizes this sum—let's call it e_{i*}—corresponds to either a node in x's left subtree, x itself, or a node in x's right subtree. If e_{i*} corresponds to a node in x's left subtree, then $x.left.m$ represents a sum starting at the node for $e_{l(x)}$ and ending at a node in x's left subtree, and hence $x.m = x.left.m$. If e_{i*} corresponds to x itself, then $x.m$ represents the sum of all p values in x's left subtree, plus $p(x)$, so that $x.m = x.left.v + p(x)$. Finally, if e_{i*} corresponds to a node in x's right subtree, then $x.m$ represents the sum of all p values in x's left subtree, plus $p(x)$, plus the sum of some subset of p values in x's right subtree. Moreover, the values taken from x's right subtree must start from the leftmost endpoint stored in the right subtree. To maximize this sum,

we need to maximize the sum from the right subtree, and that value is precisely $x.right.m$. Hence, in this case, $x.m = x.left.v + p(x) + x.right.m$.

Once we understand how to compute $x.m$, it is straightforward to compute $x.o$ from the information in x and its two children. Thus, we can implement the operations as follows:

- INTERVAL-INSERT: insert two nodes, one for each endpoint of the interval.
- FIND-POM: return the interval whose endpoint is represented by $T.root.o$.

(Note that because we are building a binary search tree of all the endpoints and then determining $T.root.o$, we have no need to delete any nodes from the tree.)

Because of how we have defined the new attributes, Theorem 14.1 says that each operation runs in $O(\lg n)$ time. In fact, FIND-POM takes only $O(1)$ time.

Solution to Problem 14-2

a. We use a circular list in which each element has two attributes, *key* and *next*. At the beginning, we initialize the list to contain the keys $1, 2, \ldots, n$ in that order. This initialization takes $O(n)$ time, since there is only a constant amount of work per element (i.e., setting its *key* and its *next* attributes). We make the list circular by letting the *next* attribute of the last element point to the first element.

We then start scanning the list from the beginning. We output and then delete every mth element, until the list becomes empty. The output sequence is the (n, m)-Josephus permutation. This process takes $O(m)$ time per element, for a total time of $O(mn)$. Since m is a constant, we get $O(mn) = O(n)$ time, as required.

b. We can use an order-statistic tree, straight out of Section 14.1. Why? Suppose that we are at a particular spot in the permutation, and let's say that it's the jth largest remaining person. Suppose that there are $k \le n$ people remaining. Then we will remove person j, decrement k to reflect having removed this person, and then go on to the $(j+m-1)$th largest remaining person (subtract 1 because we have just removed the jth largest). But that assumes that $j+m \le k$. If not, then we use a little modular arithmetic, as shown below.

In detail, we use an order-statistic tree T, and we call the procedures OS-INSERT, OS-DELETE, OS-RANK, and OS-SELECT:

```
JOSEPHUS(n, m)
  initialize T to be empty
  for j = 1 to n
      create a node x with x.key == j
      OS-INSERT(T, x)
  k = n
  j = m
  while k > 2
      x = OS-SELECT(T.root, j)
      print x.key
      OS-DELETE(T, x)
      k = k − 1
      j = ((j + m − 2) mod k) + 1
  print OS-SELECT(T.root, 1).key
```

The above procedure is easier to understand. Here's a streamlined version:

```
JOSEPHUS(n, m)
  initialize T to be empty
  for j = 1 to n
      create a node x with x.key == j
      OS-INSERT(T, x)
  j = 1
  for k = n downto 1
      j = ((j + m − 2) mod k) + 1
      x = OS-SELECT(T.root, j)
      print x.key
      OS-DELETE(T, x)
```

Either way, it takes $O(n \lg n)$ time to build up the order-statistic tree T, and then we make $O(n)$ calls to the order-statistic-tree procedures, each of which takes $O(\lg n)$ time. Thus, the total time is $O(n \lg n)$.

Lecture Notes for Chapter 15: Dynamic Programming

Dynamic Programming

- Not a specific algorithm, but a technique (like divide-and-conquer).
- Developed back in the day when "programming" meant "tabular method" (like linear programming). Doesn't really refer to computer programming.
- Used for optimization problems:
  - Find *a* solution with *the* optimal value.
  - Minimization or maximization. (We'll see both.)

Four-step method

1. Characterize the structure of an optimal solution.
2. Recursively define the value of an optimal solution.
3. Compute the value of an optimal solution, typically in a bottom-up fashion.
4. Construct an optimal solution from computed information.

Rod cutting

[New in the third edition of the book.]

How to cut steel rods into pieces in order to maximize the revenue you can get? Each cut is free. Rod lengths are always an integral number of inches.

Input: A length n and table of prices p_i, for $i = 1, 2, \ldots, n$.

Output: The maximum revenue obtainable for rods whose lengths sum to n, computed as the sum of the prices for the individual rods.

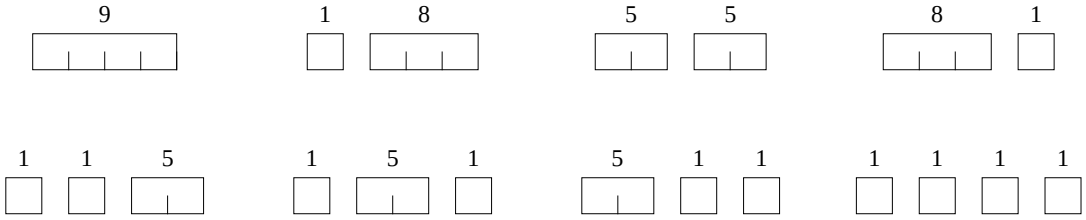
If p_n is large enough, an optimal solution might require no cuts, i.e., just leave the rod as n inches long.

Example: *[Using the first 8 values from the example in the book.]*

length i	1	2	3	4	5	6	7	8
price p_i	1	5	8	9	10	17	17	20

Can cut up a rod in 2^{n-1} different ways, because can choose to cut or not cut after each of the first $n-1$ inches.

Here are all 8 ways to cut a rod of length 4, with the costs from the example:



The best way is to cut it into two 2-inch pieces, getting a revenue of $p_2 + p_2 = 5 + 5 = 10$.

Let r_i be the maximum revenue for a rod of length i. Can express a solution as a sum of individual rod lengths.

Can determine optimal revenues r_i for the example, by inspection:

i	r_i	optimal solution
1	1	1 (no cuts)
2	5	2 (no cuts)
3	8	3 (no cuts)
4	10	$2 + 2$
5	13	$2 + 3$
6	17	6 (no cuts)
7	18	$1 + 6$ or $2 + 2 + 3$
8	22	$2 + 6$

Can determine optimal revenue r_n by taking the maximum of

- p_n: the price we get by not making a cut,
- $r_1 + r_{n-1}$: the maximum revenue from a rod of 1 inch and a rod of $n-1$ inches,
- $r_2 + r_{n-2}$: the maximum revenue from a rod of 2 inches and a rod of $n-2$ inches, ...
- $r_{n-1} + r_1$.

That is,

$$r_n = \max(p_n, r_1 + r_{n-1}, r_2 + r_{n-2}, \ldots, r_{n-1} + r_1) .$$

Optimal substructure: To solve the original problem of size n, solve subproblems on smaller sizes. After making a cut, we have two subproblems. The optimal solution to the original problem incorporates optimal solutions to the subproblems. We may solve the subproblems independently.

Example: For $n = 7$, one of the optimal solutions makes a cut at 3 inches, giving two subproblems, of lengths 3 and 4. We need to solve both of them optimally. The optimal solution for the problem of length 4, cutting into 2 pieces, each of length 2, is used in the optimal solution to the original problem with length 7.

A simpler way to decompose the problem: Every optimal solution has a leftmost cut. In other words, there's some cut that gives a first piece of length i cut off the left end, and a remaining piece of length $n - i$ on the right.

- Need to divide only the remainder, not the first piece.
- Leaves only one subproblem to solve, rather than two subproblems.
- Say that the solution with no cuts has first piece size $i = n$ with revenue p_n, and remainder size 0 with revenue $r_0 = 0$.
- Gives a simpler version of the equation for r_n:

$$r_n = \max_{1 \le i \le n} (p_i + r_{n-i}) .$$

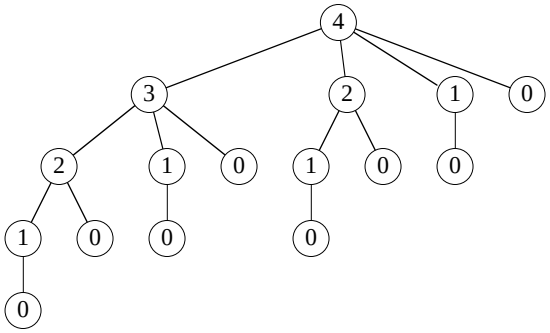
Recursive top-down solution

Direct implementation of the simpler equation for r_n.
The call CUT-ROD(p, n) returns the optimal revenue r_n:

```
CUT-ROD(p, n)
  if n == 0
      return 0
  q = -∞
  for i = 1 to n
      q = max(q, p[i] + CUT-ROD(p, n - i))
  return q
```

This procedure works, but it is terribly *inefficient*. If you code it up and run it, it could take more than an hour for $n = 40$. Running time almost doubles each time n increases by 1.

Why so inefficient?: CUT-ROD calls itself repeatedly, even on subproblems it has already solved. Here's a tree of recursive calls for $n = 4$. Inside each node is the value of n for the call represented by the node:



Lots of repeated subproblems. Solve the subproblem for size 2 twice, for size 1 four times, and for size 0 eight times.

Exponential growth: Let $T(n)$ equal the number of calls to CUT-ROD with second parameter equal to n. Then

$$T(n) = \begin{cases} 1 & \text{if } n = 0 , \\ 1 + \sum_{j=0}^{n-1} T(j) & \text{if } n \geq 1 . \end{cases}$$

Summation counts calls where second parameter is $j = n - i$.
Solution to recurrence is $T(n) = 2^n$.

Dynamic-programming solution

Instead of solving the same subproblems repeatedly, arrange to solve each subproblem just once.
Save the solution to a subproblem in a table, and refer back to the table whenever we revisit the subproblem.
"Store, don't recompute" $\Rightarrow$ time-memory trade-off.
Can turn an exponential-time solution into a polynomial-time solution.

Two basic approaches: top-down with memoization, and bottom-up.

Top-down with memoization

Solve recursively, but store each result in a table.
To find the solution to a subproblem, first look in the table. If the answer is there, use it. Otherwise, compute the solution to the subproblem and then store the solution in the table for future use.

Memoizing is remembering what we have computed previously.

Memoized version of the recursive solution, storing the solution to the subproblem of length i in array entry $r[i]$:

```
MEMOIZED-CUT-ROD(p, n)
  let r[0..n] be a new array
  for i = 0 to n
      r[i] = -∞
  return MEMOIZED-CUT-ROD-AUX(p, n, r)


MEMOIZED-CUT-ROD-AUX(p, n, r)
  if r[n] ≥ 0
      return r[n]
  if n == 0
      q = 0
  else q = -∞
      for i = 1 to n
          q = max(q, p[i] + MEMOIZED-CUT-ROD-AUX(p, n - i, r))
  r[n] = q
  return q
```

Bottom-up

Sort the subproblems by size and solve the smaller ones first. That way, when solving a subproblem, have already solved the smaller subproblems we need.

```
BOTTOM-UP-CUT-ROD(p, n)
  let r[0..n] be a new array
  r[0] = 0
  for j = 1 to n
      q = -∞
      for i = 1 to j
          q = max(q, p[i] + r[j - i])
      r[j] = q
  return r[n]
```

Running time

Both the top-down and bottom-up versions run in $\Theta(n^2)$ time.

- Bottom-up: Doubly nested loops. Number of iterations of inner **for** loop forms an arithmetic series.
- Top-down: MEMOIZED-CUT-ROD solves each subproblem just once, and it solves subproblems for sizes $0, 1, \ldots, n$. To solve a subproblem of size n, the **for** loop iterates n times $\Rightarrow$ over all recursive calls, total number of iterations forms an arithmetic series. *[Actually using aggregate analysis, which Chapter 17 covers.]*

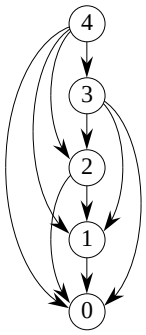
Subproblem graphs

How to understand the subproblems involved and how they depend on each other.

Directed graph:

- One vertex for each distinct subproblem.
- Has a directed edge (x, y) if computing an optimal solution to subproblem x *directly* requires knowing an optimal solution to subproblem y.

Example: For rod-cutting problem with $n = 4$:



Can think of the subproblem graph as a collapsed version of the tree of recursive calls, where all nodes for the same subproblem are collapsed into a single vertex, and all edges go from parent to child.

Subproblem graph can help determine running time. Because we solve each subproblem just once, running time is sum of times needed to solve each subproblem.

- Time to compute solution to a subproblem is typically linear in the out-degree (number of outgoing edges) of its vertex.
- Number of subproblems equals number of vertices.

When these conditions hold, running time is linear in number of vertices and edges.

Reconstructing a solution

So far, have focused on computing the *value* of an optimal solution, rather than the *choices* that produced an optimal solution.

Extend the bottom-up approach to record not just optimal values, but optimal choices. Save the optimal choices in a separate table. Then use a separate procedure to print the optimal choices.

```
EXTENDED-BOTTOM-UP-CUT-ROD(p, n)
  let r[0..n] and s[0..n] be new arrays
  r[0] = 0
  for j = 1 to n
      q = -∞
      for i = 1 to j
          if q < p[i] + r[j - i]
              q = p[i] + r[j - i]
              s[j] = i
      r[j] = q
  return r and s
```

Saves the first cut made in an optimal solution for a problem of size i in $s[i]$.

To print out the cuts made in an optimal solution:

```
PRINT-CUT-ROD-SOLUTION(p, n)
  (r, s) = EXTENDED-BOTTOM-UP-CUT-ROD(p, n)
  while n > 0
      print s[n]
      n = n - s[n]
```

Example: For the example, EXTENDED-BOTTOM-UP-CUT-ROD returns

i	0	1	2	3	4	5	6	7	8
$r[i]$	0	1	5	8	10	13	17	18	22
$s[i]$	0	1	2	3	2	2	6	1	2

A call to PRINT-CUT-ROD-SOLUTION$(p, 8)$ calls EXTENDED-BOTTOM-UP-CUT-ROD to compute the above r and s tables. Then it prints 2, sets n to 6, prints 6, and finishes (because n becomes 0).

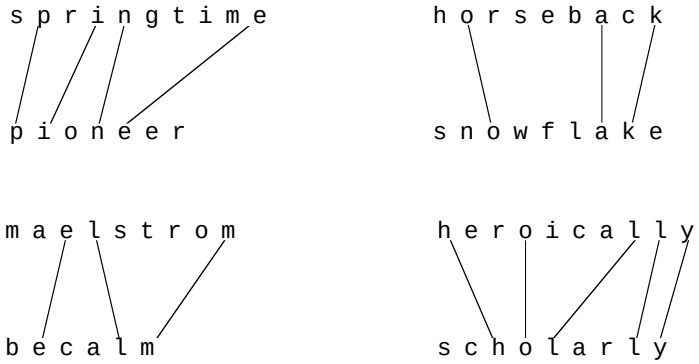
Longest common subsequence

Problem: Given 2 sequences, $X = \langle x_1, \ldots, x_m \rangle$ and $Y = \langle y_1, \ldots, y_n \rangle$. Find a subsequence common to both whose length is longest. A subsequence doesn't have to be consecutive, but it has to be in order.

[To come up with examples of longest common subsequences, search the dictionary for all words that contain the word you are looking for as a subsequence. On a UNIX system, for example, to find all the words with `pine` *as a subsequence, use the command* `grep '.*p.*i.*n.*e.*' dict`*, where* `dict` *is your local dictionary. Then check if that word is actually a longest common subsequence. Working C code for finding a longest common subsequence of two strings appears at http://www.cs.dartmouth.edu/~thc/code/lcs.c]*

Examples

[The examples are of different types of trees.]



Brute-force algorithm:

For every subsequence of X, check whether it's a subsequence of Y.

Time: $\Theta(n2^m)$.

- 2^m subsequences of X to check.
- Each subsequence takes $\Theta(n)$ time to check: scan Y for first letter, from there scan for second, and so on.

Optimal substructure

Notation:

X_i = prefix $\langle x_1, \ldots, x_i \rangle$

Y_i = prefix $\langle y_1, \ldots, y_i \rangle$

Theorem

Let $Z = \langle z_1, \ldots, z_k \rangle$ be any LCS of X and Y.

1. If $x_m = y_n$, then $z_k = x_m = y_n$ and Z_{k-1} is an LCS of X_{m-1} and Y_{n-1}.
2. If $x_m \neq y_n$, then $z_k \neq x_m \Rightarrow Z$ is an LCS of X_{m-1} and Y.
3. If $x_m \neq y_n$, then $z_k \neq y_n \Rightarrow Z$ is an LCS of X and Y_{n-1}.

Proof

1. First show that $z_k = x_m = y_n$. Suppose not. Then make a subsequence $Z' = \langle z_1, \ldots, z_k, x_m \rangle$. It's a common subsequence of X and Y and has length $k + 1 \Rightarrow Z'$ is a longer common subsequence than $Z \Rightarrow$ contradicts Z being an LCS.

   Now show Z_{k-1} is an LCS of X_{m-1} and Y_{n-1}. Clearly, it's a common subsequence. Now suppose there exists a common subsequence W of X_{m-1} and Y_{n-1} that's longer than $Z_{k-1} \Rightarrow$ length of $W \geq k$. Make subsequence W' by appending x_m to W. W' is common subsequence of X and Y, has length $\geq k + 1$ $\Rightarrow$ contradicts Z being an LCS.
2. If $z_k \neq x_m$, then Z is a common subsequence of X_{m-1} and Y. Suppose there exists a subsequence W of X_{m-1} and Y with length $> k$. Then W is a common subsequence of X and $Y \Rightarrow$ contradicts Z being an LCS.
3. Symmetric to 2. ■ (theorem)

Therefore, an LCS of two sequences contains as a prefix an LCS of prefixes of the sequences.

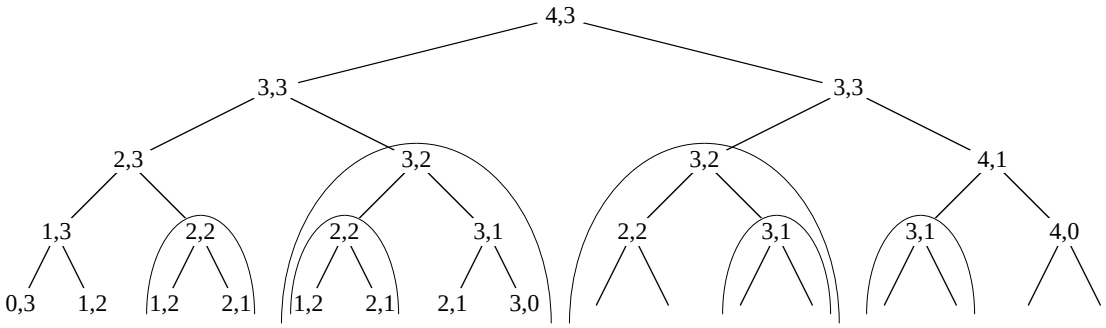
Recursive formulation

Define $c[i, j]$ = length of LCS of X_i and Y_j. We want $c[m, n]$.

$$c[i, j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 , \\ c[i-1, j-1] + 1 & \text{if } i, j > 0 \text{ and } x_i = y_j , \\ \max(c[i-1, j], c[i, j-1]) & \text{if } i, j > 0 \text{ and } x_i \neq y_j . \end{cases}$$

Again, we could write a recursive algorithm based on this formulation.

Try with bozo, bat.



- Lots of repeated subproblems.
- Instead of recomputing, store in a table.

Compute length of optimal solution

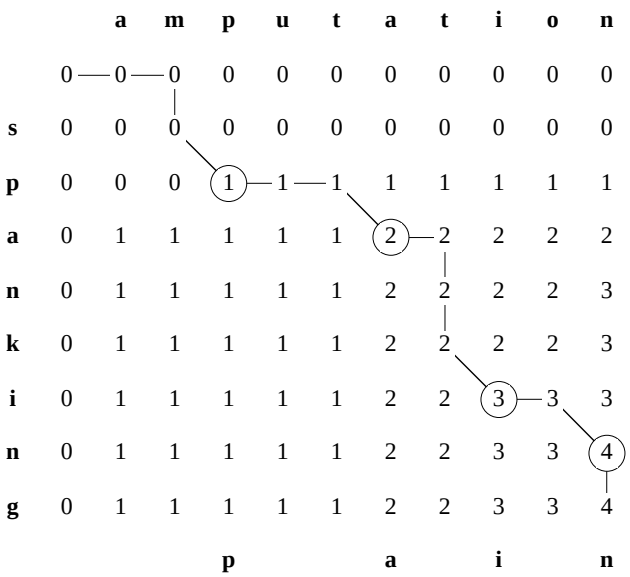
```
LCS-LENGTH(X, Y, m, n)
  let b[1..m, 1..n] and c[0..m, o..n] be new tables
  for i = 1 to m
      c[i, 0] = 0
  for j = 0 to n
      c[0, j] = 0
  for i = 1 to m
      for j = 1 to n
          if x_i == y_j
              c[i, j] = c[i − 1, j − 1] + 1
              b[i, j] = "↖"
          else if c[i − 1, j] ≥ c[i, j − 1]
                  c[i, j] = c[i − 1, j]
                  b[i, j] = "↑"
              else c[i, j] = c[i, j − 1]
                  b[i, j] = "←"
  return c and b
```

```
PRINT-LCS(b, X, i, j)
  if i == 0 or j = 0
      return
  if b[i, j] == "↖"
      PRINT-LCS(b, X, i − 1, j − 1)
      print x_i
  elseif b[i, j] == "↑"
      PRINT-LCS(b, X, i − 1, j)
  else PRINT-LCS(b, X, i, j − 1)
```

- Initial call is PRINT-LCS(b, X, m, n).
- $b[i, j]$ points to table entry whose subproblem we used in solving LCS of X_i and Y_j.
- When $b[i, j] = \nwarrow$, we have extended LCS by one character. So longest common subsequence = entries with $\nwarrow$ in them.

Demonstration

What do `spanking` and `amputation` have in common? *[Show only $c[i, j]$.]*



Answer: `pain`.

Time

$\Theta(mn)$

Optimal binary search trees

[Added in the second edition.]

- Given sequence $K = \langle k_1, k_2, \ldots, k_n \rangle$ of n distinct keys, sorted ($k_1 < k_2 < \cdots < k_n$).
- Want to build a binary search tree from the keys.
- For k_i, have probability p_i that a search is for k_i.
- Want BST with minimum expected search cost.
- Actual cost = # of items examined.

  For key k_i, cost = $\text{depth}_T(k_i) + 1$, where $\text{depth}_T(k_i)$ = depth of k_i in BST T.

E [search cost in T]

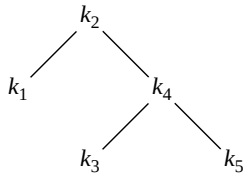
$$\begin{aligned} &= \sum_{i=1}^{n} (\text{depth}_T(k_i) + 1) \cdot p_i \\ &= \sum_{i=1}^{n} \text{depth}_T(k_i) \cdot p_i + \sum_{i=1}^{n} p_i \\ &= 1 + \sum_{i=1}^{n} \text{depth}_T(k_i) \cdot p_i \qquad \text{(since probabilities sum to 1)} \qquad (*) \end{aligned}$$

[Keep equation (*) on board.]

[Similar to optimal BST problem in the book, but simplified here: we assume that all searches are successful. Book has probabilities of searches between keys in tree.]

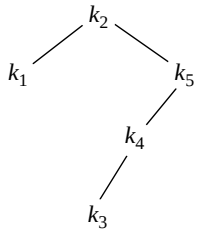
Example

i	1	2	3	4	5
p_i	.25	.2	.05	.2	.3



i	$\text{depth}_T(k_i)$	$\text{depth}_T(k_i) \cdot p_i$
1	1	.25
2	0	0
3	2	.1
4	1	.2
5	2	.6
		1.15

Therefore, E [search cost] $= 2.15$.



i	$\text{depth}_T(k_i)$	$\text{depth}_T(k_i) \cdot p_i$
1	1	.25
2	0	0
3	3	.15
4	2	.4
5	1	.3
		1.10

Therefore, E [search cost] $= 2.10$, which turns out to be optimal.

Observations

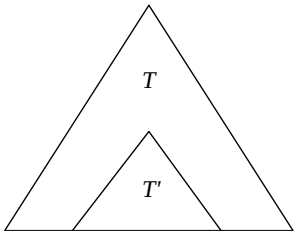
- Optimal BST might not have smallest height.
- Optimal BST might not have highest-probability key at root.

Build by exhaustive checking?

- Construct each n-node BST.
- For each, put in keys.
- Then compute expected search cost.
- But there are $\Omega(4^n / n^{3/2})$ different BSTs with n nodes.

Optimal substructure

Consider any subtree of a BST. It contains keys in a contiguous range $k_i, \ldots, k_j$ for some $1 \le i \le j \le n$.

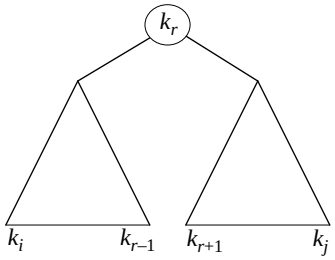


If T is an optimal BST and T contains subtree T' with keys $k_i, \ldots, k_j$, then T' must be an optimal BST for keys $k_i, \ldots, k_j$.

Proof Cut and paste. ■

Use optimal substructure to construct an optimal solution to the problem from optimal solutions to subproblems:

- Given keys $k_i, \ldots, k_j$ (the problem).
- One of them, k_r, where $i \le r \le j$, must be the root.
- Left subtree of k_r contains $k_i, \ldots, k_{r-1}$.
- Right subtree of k_r contains $k_{r+1}, \ldots, k_j$.



- If
  - we examine all candidate roots k_r, for $i \le r \le j$, and
  - we determine all optimal BSTs containing $k_i, \ldots, k_{r-1}$ and containing $k_{r+1}, \ldots, k_j$,

  then we're guaranteed to find an optimal BST for $k_i, \ldots, k_j$.

Recursive solution

Subproblem domain:

- Find optimal BST for $k_i, \ldots, k_j$, where $i \geq 1$, $j \leq n$, $j \geq i - 1$.
- When $j = i - 1$, the tree is empty.

Define $e[i, j]$ = expected search cost of optimal BST for $k_i, \ldots, k_j$.

If $j = i - 1$, then $e[i, j] = 0$.

If $j \geq i$,

- Select a root k_r, for some $i \leq r \leq j$.
- Make an optimal BST with $k_i, \ldots, k_{r-1}$ as the left subtree.
- Make an optimal BST with $k_{r+1}, \ldots, k_j$ as the right subtree.
- Note: when $r = i$, left subtree is $k_i, \ldots, k_{i-1}$; when $r = j$, right subtree is $k_{j+1}, \ldots, k_j$.

When a subtree becomes a subtree of a node:

- Depth of every node in subtree goes up by 1.
- Expected search cost increases by

$$w(i, j) = \sum_{l=i}^{j} p_l \qquad \text{(refer to equation } (*)\text{)} .$$

If k_r is the root of an optimal BST for $k_i, \ldots, k_j$:

$$e[i, j] = p_r + (e[i, r-1] + w(i, r-1)) + (e[r+1, j] + w(r+1, j)) .$$

But $w(i, j) = w(i, r-1) + p_r + w(r+1, j)$.

Therefore, $e[i, j] = e[i, r-1] + e[r+1, j] + w(i, j)$.

This equation assumes that we already know which key is k_r.

We don't.

Try all candidates, and pick the best one:

$$e[i, j] = \begin{cases} 0 & \text{if } j = i - 1 , \\ \min\limits_{i \leq r \leq j} \{e[i, r-1] + e[r+1, j] + w(i, j)\} & \text{if } i \leq j . \end{cases}$$

Could write a recursive algorithm...

Computing an optimal solution

As "usual," we'll store the values in a table:

$$e[\underbrace{1\,..\,n+1}_{\substack{\text{can store} \\ e[n+1, n]}} , \underbrace{0\,..\,n}_{\substack{\text{can store} \\ e[1, 0]}}]$$

- Will use only entries $e[i, j]$, where $j \geq i - 1$.

- Will also compute

  $root[i, j]$ = root of subtree with keys $k_i, \ldots, k_j$, for $1 \le i \le j \le n$.

One other table: don't recompute $w(i, j)$ from scratch every time we need it. (Would take $\Theta(j - i)$ additions.)

Instead:

- Table $w[1..n+1, 0..n]$
- $w[i, i-1] = 0$ for $1 \le i \le n$
- $w[i, j] = w[i, j-1] + p_j$ for $1 \le i \le j \le n$

Can compute all $\Theta(n^2)$ values in $O(1)$ time each.

OPTIMAL-BST(p, q, n)

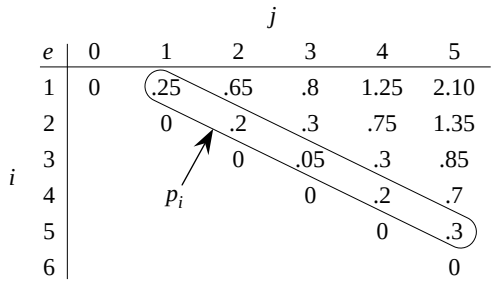
```
let e[1..n+1, 0..n], w[1..n+1, 0..n], and root[1..n, 1..n] be new tables
for i = 1 to n+1
    e[i, i-1] = 0
    w[i, i-1] = 0
for l = 1 to n
    for i = 1 to n-l+1
        j = i+l-1
        e[i, j] = ∞
        w[i, j] = w[i, j-1] + p_j
        for r = i to j
            t = e[i, r-1] + e[r+1, j] + w[i, j]
            if t < e[i, j]
                e[i, j] = t
                root[i, j] = r
return e and root
```

First **for** loop initializes e, w entries for subtrees with 0 keys.

Main **for** loop:

- Iteration for l works on subtrees with l keys.
- Idea: compute in order of subtree sizes, smaller (1 key) to larger (n keys).

For example at beginning:



e	0	1	2	3	4	5
1	0	.25	.65	.8	1.25	2.10
2		0	.2	.3	.75	1.35
3			0	.05	.3	.85
4				0	.2	.7
5					0	.3
6						0

		j					
	w	0	1	2	3	4	5
	1	0	.25	.45	.5	.7	1.0
	2		0	.2	.25	.45	.75
i	3			0	.05	.25	.55
	4				0	.2	.5
	5					0	.3
	6						0

		j				
	root	1	2	3	4	5
	1	1	1	1	2	2
	2		2	2	2	4
i	3			3	4	5
	4				4	5
	5					5

Time

$O(n^3)$: for loops nested 3 deep, each loop index takes on $\leq n$ values. Can also show $\Omega(n^3)$. Therefore, $\Theta(n^3)$.

Construct an optimal solution

CONSTRUCT-OPTIMAL-BST(*root*)

$r = root[1, n]$
print "k"$_r$ "is the root"
CONSTRUCT-OPT-SUBTREE$(1, r-1, r,$ "left", $root)$
CONSTRUCT-OPT-SUBTREE$(r+1, n, r,$ "right", $root)$

CONSTRUCT-OPT-SUBTREE$(i, j, r, dir, root)$

if $i \leq j$
  $t = root[i, j]$
  print "k"$_t$ "is" *dir* "child of k"$_r$
  CONSTRUCT-OPT-SUBTREE$(i, t-1, t,$ "left", $root)$
  CONSTRUCT-OPT-SUBTREE$(t+1, j, t,$ "right", $root)$

Elements of dynamic programming

Mentioned already:

- optimal substructure
- overlapping subproblems

Optimal substructure

- Show that a solution to a problem consists of making a choice, which leaves one or subproblems to solve.

- Suppose that you are given this last choice that leads to an optimal solution. *[We find that students often have trouble understanding the relationship between optimal substructure and determining which choice is made in an optimal solution. One way that helps them understand optimal substructure is to imagine that the dynamic-programming gods tell you what was the last choice made in an optimal solution.]*
- Given this choice, determine which subproblems arise and how to characterize the resulting space of subproblems.
- Show that the solutions to the subproblems used within the optimal solution must themselves be optimal. Usually use cut-and-paste:
  - Suppose that one of the subproblem solutions is not optimal.
  - *Cut* it out.
  - *Paste* in an optimal solution.
  - Get a better solution to the original problem. Contradicts optimality of problem solution.

That was optimal substructure.

Need to ensure that you consider a wide enough range of choices and subproblems that you get them all. *[The dynamic-programming gods are too busy to tell you what that last choice really was.]* Try all the choices, solve all the subproblems resulting from each choice, and pick the choice whose solution, along with subproblem solutions, is best.

How to characterize the space of subproblems?

- Keep the space as simple as possible.
- Expand it as necessary.

Examples

Rod cutting

- Space of subproblems was rods of length $n - i$, for $1 \le i \le n$.
- No need to try a more general space of subproblems.

Optimal binary search trees

- Suppose we had tried to constrain space of subproblems to subtrees with keys $k_1, k_2, \ldots, k_j$.
- An optimal BST would have root k_r, for some $1 \le r \le j$.
- Get subproblems $k_1, \ldots, k_{r-1}$ and $k_{r+1}, \ldots, k_j$.
- Unless we could guarantee that $r = j$, so that subproblem with $k_{r+1}, \ldots, k_j$ is empty, then this subproblem is *not* of the form $k_1, k_2, \ldots, k_j$.
- Thus, needed to allow the subproblems to vary at "both ends," i.e., allow both i and j to vary.

Optimal substructure varies across problem domains:

1. *How many subproblems* are used in an optimal solution.
2. *How many choices* in determining which subproblem(s) to use.

- Rod cutting:
  - 1 subproblem (of size $n - i$)
  - n choices
- Longest common subsequence:
  - 1 subproblem
  - Either
    - 1 choice (if $x_i = y_j$, LCS of X_{i-1} and Y_{j-1}), or
    - 2 choices (if $x_i \neq y_j$, LCS of X_{i-1} and Y, and LCS of X and Y_{j-1})
- Optimal binary search tree:
  - 2 subproblems ($k_i, \ldots, k_{r-1}$ and $k_{r+1}, \ldots, k_j$)
  - $j - i + 1$ choices for k_r in $k_i, \ldots, k_j$. Once we determine optimal solutions to subproblems, we choose from among the $j - i + 1$ candidates for k_r.

Informally, running time depends on (# of subproblems overall) $\times$ (# of choices).

- Rod cutting: $\Theta(n)$ subproblems, $\leq n$ choices for each
  $\Rightarrow O(n^2)$ running time.
- Longest common subsequence: $\Theta(mn)$ subproblems, ≤ 2 choices for each
  $\Rightarrow \Theta(mn)$ running time.
- Optimal binary search tree: $\Theta(n^2)$ subproblems, $O(n)$ choices for each
  $\Rightarrow O(n^3)$ running time.

Can use the subproblem graph to get the same analysis: count the number of edges.

- Each vertex corresponds to a subproblem.
- Choices for a subproblem are vertices that the subproblem has edges going to.
- For rod cutting, subproblem graph has n vertices and $\leq n$ edges per vertex
  $\Rightarrow O(n^2)$ running time.
  In fact, can get an exact count of the edges: for $i = 0, 1, \ldots, n$, vertex for subproblem size i has out-degree $i \Rightarrow$ # of edges $= \sum_{i=0}^{n} i = n(n+1)/2$.
- Subproblem graph for matrix-chain multiplication would have $\Theta(n^2)$ vertices, each with degree $\leq n - 1$
  $\Rightarrow O(n^3)$ running time.

Dynamic programming uses optimal substructure *bottom up*.

- *First* find optimal solutions to subproblems.
- *Then* choose which to use in optimal solution to the problem.

When we look at greedy algorithms, we'll see that they work *top down*: *first* make a choice that looks best, *then* solve the resulting subproblem.

Don't be fooled into thinking optimal substructure applies to all optimization problems. It doesn't.

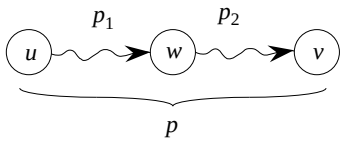
Here are two problems that look similar. In both, we're given an *unweighted, directed graph* $G = (V, E)$.

- V is a set of *vertices*.
- E is a set of *edges*.

And we ask about finding a ***path*** (sequence of connected edges) from vertex u to vertex v.

- ***Shortest path***: find path $u \rightsquigarrow v$ with fewest edges. Must be ***simple*** (no *cycles*), since removing a cycle from a path gives a path with fewer edges.
- ***Longest simple path***: find *simple* path $u \rightsquigarrow v$ with most edges. If didn't require simple, could repeatedly traverse a cycle to make an arbitrarily long path.

Shortest path has optimal substructure.



- Suppose p is shortest path $u \rightsquigarrow v$.
- Let w be any vertex on p.
- Let p_1 be the portion of p going $u \rightsquigarrow w$.
- Then p_1 is a shortest path $u \rightsquigarrow w$.

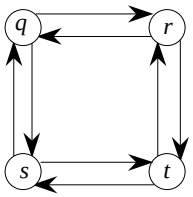
Proof Suppose there exists a shorter path p_1' going $u \rightsquigarrow w$. Cut out p_1, replace it with p_1', get path $u \overset{p_1'}{\rightsquigarrow} w \overset{p_2}{\rightsquigarrow} v$ with fewer edges than p. ■

Therefore, can find shortest path $u \rightsquigarrow v$ by considering all intermediate vertices w, then finding shortest paths $u \rightsquigarrow w$ and $w \rightsquigarrow v$.

Same argument applies to p_2.

Does longest path have optimal substructure?

- It seems like it should.
- It does *not*.



Consider $q \to r \to t$ = longest path $q \rightsquigarrow t$. Are its subpaths longest paths?

No!

- Subpath $q \rightsquigarrow r$ is $q \to r$.
- Longest simple path $q \rightsquigarrow r$ is $q \to s \to t \to r$.
- Subpath $r \rightsquigarrow t$ is $r \to t$.
- Longest simple path $r \rightsquigarrow t$ is $r \to q \to s \to t$.

Not only isn't there optimal substructure, but we can't even assemble a legal solution from solutions to subproblems.

Combine longest simple paths:

$q \to s \to t \to r \to q \to s \to t$

Not simple!

In fact, this problem is NP-complete (so it probably has no optimal substructure to find.)

What's the big difference between shortest path and longest path?

- Shortest path has ***independent*** subproblems.
- Solution to one subproblem does not affect solution to another subproblem of the same problem.
- Longest simple path: subproblems are *not* independent.
- Consider subproblems of longest simple paths $q \leadsto r$ and $r \leadsto t$.
- Longest simple path $q \leadsto r$ uses s and t.
- Cannot use s and t to solve longest simple path $r \leadsto t$, since if we do, the path isn't simple.
- But we *have* to use t to find longest simple path $r \leadsto t$!
- Using resources (vertices) to solve one subproblem renders them unavailable to solve the other subproblem.

  [For shortest paths, if we look at a shortest path $u \overset{p_1}{\leadsto} w \overset{p_2}{\leadsto} v$, no vertex other than w can appear in p_1 and p_2. Otherwise, we have a cycle.]

Independent subproblems in our examples:

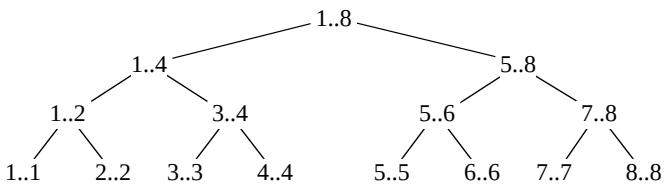
- Rod cutting and longest common subsequence
  - 1 subproblem $\Rightarrow$ automatically independent.
- Optimal binary search tree
  - $k_i, \ldots, k_{r-1}$ and $k_{r+1}, \ldots, k_j \Rightarrow$ independent.

Overlapping subproblems

These occur when a recursive algorithm revisits the same problem over and over.

Good divide-and-conquer algorithms usually generate a brand new problem at each stage of recursion.

Example: merge sort



Won't go through exercise of showing repeated subproblems.

Book has a good example for matrix-chain multiplication.

Alternative approach to dynamic programming: ***memoization***

- "Store, don't recompute."
- Make a table indexed by subproblem.
- When solving a subproblem:
  - Lookup in table.
  - If answer is there, use it.
  - Else, compute answer, then store it.
- In bottom-up dynamic programming, we go one step further. We determine in what order we'd want to access the table, and fill it in that way.

Solutions for Chapter 15: Dynamic Programming

Solution to Exercise 15.1-1

We can verify that $T(n) = 2^n$ is a solution to the given recurrence by the substitution method. We note that for $n = 0$, the formula is true since $2^0 = 1$. For $n > 0$, substituting into the recurrence and using the formula for summing a geometric series yields

$$\begin{aligned} T(n) &= 1 + \sum_{j=0}^{n-1} 2^j \\ &= 1 + (2^n - 1) \\ &= 2^n \ . \end{aligned}$$

Solution to Exercise 15.1-2

Here is a counterexample for the "greedy" strategy:

length i	1	2	3	4
price p_i	1	20	33	36
p_i / i	1	10	11	1

Let the given rod length be 4. According to a greedy strategy, we first cut out a rod of length 3 for a price of 33, which leaves us with a rod of length 1 of price 1. The total price for the rod is 34. The optimal way is to cut it into two rods of length 2 each fetching us 40 dollars.

Solution to Exercise 15.1-3

```
MODIFIED-CUT-ROD(p, n, c)
  let r[0..n] be a new array
  r[0] = 0
  for j = 1 to n
      q = p[j]
      for i = 1 to j − 1
          q = max(q, p[i] + r[j − i] − c)
      r[j] = q
  return r[n]
```

The major modification required is in the body of the inner **for** loop, which now reads $q = \max(q, p[i] + r[j - i] - c)$. This change reflects the fixed cost of making the cut, which is deducted from the revenue. We also have to handle the case in which we make no cuts (when i equals j); the total revenue in this case is simply $p[j]$. Thus, we modify the inner **for** loop to run from i to $j - 1$ instead of to j. The assignment $q = p[j]$ takes care of the case of no cuts. If we did not make these modifications, then even in the case of no cuts, we would be deducting c from the total revenue.

Solution to Exercise 15.1-4

```
MEMOIZED-CUT-ROD(p, n)
  let r[0..n] and s[0..n] be new arrays
  for i = 0 to n
      r[i] = −∞
  (val, s) = MEMOIZED-CUT-ROD-AUX(p, n, r, s)
  print "The optimal value is " val " and the cuts are at "
  j = n
  while j > 0
      print s[j]
      j = j − s[j]
```

```
MEMOIZED-CUT-ROD-AUX(p, n, r, s)
 if r[n] ≥ 0
     return r[n]
 if n == 0
     q = 0
 else q = −∞
     for i = 1 to n
         (val, s) = MEMOIZED-CUT-ROD-AUX(p, n − i, r, s)
         if q < p[i] + val
             q = p[i] + val
             s[n] = i
 r[n] = q
 return (q, s)
```

PRINT-CUT-ROD-SOLUTION constructs the actual lengths where a cut should happen. Array entry $s[i]$ contains the value j indicating that an optimal cut for a rod of length i is j inches. The next cut is given by $s[i-j]$, and so on.

Solution to Exercise 15.1-5

```
FIBONACCI(n)
 let fib[0..n] be a new array
 fib[0] = fib[1] = 1
 for i = 2 to n
     fib[i] = fib[i − 1] + fib[i − 2]
 return fib[n]
```

FIBONACCI directly implements the recurrence relation of the Fibonacci sequence. Each number in the sequence is the sum of the two previous numbers in the sequence. The running time is clearly $O(n)$.

The subproblem graph consists of $n+1$ vertices, $v_0, v_1, \ldots, v_n$. For $i = 2, 3, \ldots, n$, vertex v_i has two leaving edges: to vertex v_{i-1} and to vertex v_{i-2}. No edges leave vertices v_0 or v_1. Thus, the subproblem graph has $2n-2$ edges.

Solution to Exercise 15.2-4

The vertices of the subproblem graph are the ordered pairs v_{ij}, where $i \le j$. If $i = j$, then there are no edges out of v_{ij}. If $i < j$, then for every k such that $i \le k < j$, the subproblem graph contains edges (v_{ij}, v_{ik}) and $(v_{ij}, v_{k+1,j})$. These edges indicate that to solve the subproblem of optimally parenthesizing the product $A_i \cdots A_j$, we need to solve subproblems of optimally parenthesizing the products $A_i \cdots A_k$ and $A_{k+1} \cdots A_j$. The number of vertices is

$$\sum_{i=1}^{n} \sum_{j=i}^{n} 1 = \frac{n(n+1)}{2} ,$$

and the number of edges is

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=i}^{n}(j-i) &= \sum_{i=1}^{n}\sum_{t=0}^{n-i}t \qquad \text{(substituting } t = j-i\text{)}\\ &= \sum_{i=1}^{n}\frac{(n-i)(n-i+1)}{2}\,.\end{aligned}$$

Substituting $r = n - i$ and reversing the order of summation, we obtain

$$\begin{aligned}&\sum_{i=1}^{n}\frac{(n-i)(n-i+1)}{2}\\ &= \frac{1}{2}\sum_{r=0}^{n-1}(r^2+r)\\ &= \frac{1}{2}\left(\frac{(n-1)n(2n-1)}{6}+\frac{(n-1)n}{2}\right) \qquad \text{(by equations (A.3) and (A.1))}\\ &= \frac{(n-1)n(n+1)}{6}\,.\end{aligned}$$

Thus, the subproblem graph has $\Theta(n^2)$ vertices and $\Theta(n^3)$ edges.

Solution to Exercise 15.2-5

This solution is also posted publicly

Each time the l-loop executes, the i-loop executes $n - l + 1$ times. Each time the i-loop executes, the k-loop executes $j - i = l - 1$ times, each time referencing m twice. Thus the total number of times that an entry of m is referenced while computing other entries is $\sum_{l=2}^{n}(n-l+1)(l-1)2$. Thus,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=i}^{n}R(i,j) &= \sum_{l=2}^{n}(n-l+1)(l-1)2\\ &= 2\sum_{l=1}^{n-1}(n-l)l\\ &= 2\sum_{l=1}^{n-1}nl - 2\sum_{l=1}^{n-1}l^2\\ &= 2\frac{n(n-1)n}{2} - 2\frac{(n-1)n(2n-1)}{6}\\ &= n^3 - n^2 - \frac{2n^3-3n^2+n}{3}\\ &= \frac{n^3-n}{3}\,.\end{aligned}$$

Solution to Exercise 15.3-1

This solution is also posted publicly

Running RECURSIVE-MATRIX-CHAIN is asymptotically more efficient than enumerating all the ways of parenthesizing the product and computing the number of multiplications for each.

Consider the treatment of subproblems by the two approaches.

- For each possible place to split the matrix chain, the enumeration approach finds all ways to parenthesize the left half, finds all ways to parenthesize the right half, and looks at all possible combinations of the left half with the right half. The amount of work to look at each combination of left- and right-half subproblem results is thus the product of the number of ways to do the left half and the number of ways to do the right half.
- For each possible place to split the matrix chain, RECURSIVE-MATRIX-CHAIN finds the best way to parenthesize the left half, finds the best way to parenthesize the right half, and combines just those two results. Thus the amount of work to combine the left- and right-half subproblem results is $O(1)$.

Section 15.2 argued that the running time for enumeration is $\Omega(4^n/n^{3/2})$. We will show that the running time for RECURSIVE-MATRIX-CHAIN is $O(n3^{n-1})$.

To get an upper bound on the running time of RECURSIVE-MATRIX-CHAIN, we'll use the same approach used in Section 15.2 to get a lower bound: Derive a recurrence of the form $T(n) \le \ldots$ and solve it by substitution. For the lower-bound recurrence, the book assumed that the execution of lines 1–2 and 6–7 each take at least unit time. For the upper-bound recurrence, we'll assume those pairs of lines each take at most constant time c. Thus, we have the recurrence

$$T(n) \le \begin{cases} c & \text{if } n = 1 , \\ c + \displaystyle\sum_{k=1}^{n-1} (T(k) + T(n-k) + c) & \text{if } n \ge 2 . \end{cases}$$

This is just like the book's $\ge$ recurrence except that it has c instead of 1, and so we can be rewrite it as

$$T(n) \le 2 \sum_{i=1}^{n-1} T(i) + cn .$$

We shall prove that $T(n) = O(n3^{n-1})$ using the substitution method. (Note: Any upper bound on $T(n)$ that is $o(4^n/n^{3/2})$ will suffice. You might prefer to prove one that is easier to think up, such as $T(n) = O(3.5^n)$.) Specifically, we shall show that $T(n) \le cn3^{n-1}$ for all $n \ge 1$. The basis is easy, since $T(1) \le c = c \cdot 1 \cdot 3^{1-1}$. Inductively, for $n \ge 2$ we have

$$
\begin{aligned}
T(n) &\le 2\sum_{i=1}^{n-1} T(i) + cn \\
&\le 2\sum_{i=1}^{n-1} ci\,3^{i-1} + cn \\
&\le c\cdot\left(2\sum_{i=1}^{n-1} i\,3^{i-1} + n\right) \\
&= c\cdot\left(2\cdot\left(\frac{n3^{n-1}}{3-1} + \frac{1-3^n}{(3-1)^2}\right) + n\right) \qquad \text{(see below)} \\
&= cn3^{n-1} + c\cdot\left(\frac{1-3^n}{2} + n\right) \\
&= cn3^{n-1} + \frac{c}{2}(2n+1-3^n) \\
&\le cn3^{n-1} \text{ for all } c > 0,\, n \ge 1\ .
\end{aligned}
$$

Running RECURSIVE-MATRIX-CHAIN takes $O(n3^{n-1})$ time, and enumerating all parenthesizations takes $\Omega(4^n/n^{3/2})$ time, and so RECURSIVE-MATRIX-CHAIN is more efficient than enumeration.

Note: The above substitution uses the following fact:

$$\sum_{i=1}^{n-1} i x^{i-1} = \frac{nx^{n-1}}{x-1} + \frac{1-x^n}{(x-1)^2}\ .$$

This equation can be derived from equation (A.5) by taking the derivative. Let

$$f(x) = \sum_{i=1}^{n-1} x^i = \frac{x^n-1}{x-1} - 1\ .$$

Then

$$\sum_{i=1}^{n-1} i x^{i-1} = f'(x) = \frac{nx^{n-1}}{x-1} + \frac{1-x^n}{(x-1)^2}\ .$$

Solution to Exercise 15.3-5

We say that a problem exhibits the optimal substructure property when optimal solutions to a problem incorporate optimal solutions to related subproblems, *which we may solve independently* (i.e., they do not share resources). When we impose a limit l_i on the number of pieces of size i that we are permitted to produce, the subproblems can no longer be solved *independently*. For example, consider a rod of length 4 with the following prices and limits:

length i	1	2	3	4
price p_i	15	20	33	36
limit l_i	2	1	1	1

This instance has only three solutions that do not violate the limits: length 4 with price 36; lengths 1 and 3 with price 48; and lengths 1, 1, and 2 with price 50. The

optimal solution, therefore is to cut into lengths 1, 1, and 2. When we look at the subproblem for length 2, it has two solutions that do not violate the limits: length 2 with price 20, and lengths 1 and 1 with price 30. The optimal solution for length 2, therefore, is to cut into lengths 1 and 1. But we cannot use this optimal solution for the subproblem in the optimal solution for the original problem, because it would result in using four rods of length 1 to solve the original problem, violating the limit of two length-1 rods.

Solution to Exercise 15.3-6

Any solution must add the additional assumption that no currency can be repeated in a sequence of trades. Without this assumption, if $r_{ij} > 1/r_{ji}$ for some currencies i and j, we could repeatedly exchange $i \to j \to i \to j \to \cdots$ and make an unbounded profit.

To see that this problem has optimal substructure when $c_k = 0$ for all k, observe that the problem of exchanging currency a for currency b is equivalent to finding a sequence of currencies $k_1, k_2, \ldots, k_m$ such that $k_1 = a$, $k_m = b$, and the product $r_{k_1k_2}r_{k_2k_3}\cdots r_{k_{m-1}k_m}$ is maximized.

We use the usual cut-and-paste argument. Suppose that an optimal solution contains a sequence $\langle k_i, k_{i+1}, \ldots, k_j\rangle$ of currencies, and suppose that there exists a sequence $\langle k'_i, k'_{i+1}, \ldots, k'_j\rangle$, such that $k'_i = k_i$, $k'_j = k_j$, and $r_{k'_ik'_{i+1}}\cdots r_{k'_{j-1}k'_j} > r_{k_ik_{i+1}}\cdots r_{k_{j-1}k_j}$. Then we could substitute the sequence $\langle k'_i, k'_{i+1}, \ldots, k'_j\rangle$ for the sequence $\langle k_i, k_{i+1}, \ldots, k_j\rangle$ in the optimal solution to create an even better solution.

We show that optimal substructure does not hold when the c_k are arbitrary values by means of an example. Suppose we have four currencies, with the following exchange rates:

		j			
	r_{ij}	1	2	3	4
	1	1	2	5/2	6
	2	1/2	1	3/2	3
i	3	2/5	2/3	1	3
	4	1/6	1/3	1/3	1

Let $c_1 = 2$ and $c_2 = c_3 = 3$. Note that this example is not too badly contrived, in that $r_{ji} = 1/r_{ij}$ for all i and j.

To see how this example does not exhibit optimal substructure, let's examine an optimal solution for exchanging currency 1 for currency 4. There are five possible exchange sequences, with the following costs:

$$\begin{aligned}
\langle 1,4\rangle &: 6-2 &&= 4\,,\\
\langle 1,2,4\rangle &: 2\cdot 3-3 &&= 3\,,\\
\langle 1,3,4\rangle &: 5/2\cdot 3-3 &&= 9/2\,,\\
\langle \mathbf{1},\mathbf{2},\mathbf{3},\mathbf{4}\rangle &: \mathbf{2\cdot 3/2\cdot 3-3} &&= \mathbf{6}\\
\langle 1,3,2,4\rangle &: 5/2\cdot 2/3\cdot 3-3 &&= 2
\end{aligned}$$

The optimal exchange sequence, $\langle 1,2,3,4\rangle$, appears in boldface.

Let's examine the subproblem of exchanging currency 1 for currency 3. Allowing currency 4 to be part of the exchange sequence, there are again five possible exchange sequences with the following costs and the optimal one in boldface:

$$\begin{array}{lcll}
\langle \mathbf{1}, \mathbf{3} \rangle & : & \mathbf{5/2 - 2} & = \mathbf{1/2} \\
\langle 1, 2, 3 \rangle & : & 2 \cdot 3/2 - 3 & = 0 \\
\langle 1, 4, 3 \rangle & : & 6 \cdot 1/3 - 3 & = -1 \\
\langle 1, 2, 4, 3 \rangle & : & 2 \cdot 3 \cdot 1/3 - 3 & = -1 \\
\langle 1, 4, 2, 3 \rangle & : & 6 \cdot 1/3 \cdot 3/2 = 3 & = 0
\end{array}$$

We see that the solution to the original problem includes the subproblem of exchanging currency 1 for currency 3, yet the solution $\langle 1, 2, 3 \rangle$ to the subproblem used in the optimal solution to the original problem is not the optimal solution $\langle 1, 3 \rangle$ to the subproblem on its own.

Solution to Exercise 15.4-4

This solution is also posted publicly

When computing a particular row of the c table, no rows before the previous row are needed. Thus only two rows—$2 \cdot Y.length$ entries—need to be kept in memory at a time. (Note: Each row of c actually has $Y.length + 1$ entries, but we don't need to store the column of 0's—instead we can make the program "know" that those entries are 0.) With this idea, we need only $2 \cdot \min(m, n)$ entries if we always call LCS-LENGTH with the shorter sequence as the Y argument.

We can thus do away with the c table as follows:

- Use two arrays of length $\min(m, n)$, *previous-row* and *current-row*, to hold the appropriate rows of c.
- Initialize *previous-row* to all 0 and compute *current-row* from left to right.
- When *current-row* is filled, if there are still more rows to compute, copy *current-row* into *previous-row* and compute the new *current-row*.

Actually only a little more than one row's worth of c entries—$\min(m, n) + 1$ entries—are needed during the computation. The only entries needed in the table when it is time to compute $c[i, j]$ are $c[i, k]$ for $k \leq j - 1$ (i.e., earlier entries in the current row, which will be needed to compute the next row); and $c[i - 1, k]$ for $k \geq j - 1$ (i.e., entries in the previous row that are still needed to compute the rest of the current row). This is one entry for each k from 1 to $\min(m, n)$ except that there are two entries with $k = j - 1$, hence the additional entry needed besides the one row's worth of entries.

We can thus do away with the c table as follows:

- Use an array a of length $\min(m, n) + 1$ to hold the appropriate entries of c. At the time $c[i, j]$ is to be computed, a will hold the following entries:
  - $a[k] = c[i, k]$ for $1 \leq k < j - 1$ (i.e., earlier entries in the current "row"),
  - $a[k] = c[i - 1, k]$ for $k \geq j - 1$ (i.e., entries in the previous "row"),

  - $a[0] = c[i, j-1]$ (i.e., the previous entry computed, which couldn't be put into the "right" place in a without erasing the still-needed $c[i-1, j-1]$).
- Initialize a to all 0 and compute the entries from left to right.
  - Note that the 3 values needed to compute $c[i, j]$ for $j > 1$ are in $a[0] = c[i, j-1]$, $a[j-1] = c[i-1, j-1]$, and $a[j] = c[i-1, j]$.
  - When $c[i, j]$ has been computed, move $a[0]$ ($c[i, j-1]$) to its "correct" place, $a[j-1]$, and put $c[i, j]$ in $a[0]$.

Solution to Problem 15-1

We will make use of the optimal substructure property of longest paths in *acyclic* graphs. Let u be some vertex of the graph. If $u = t$, then the longest path from u to t has zero weight. If $u \neq t$, let p be a longest path from u to t. Path p has at least two vertices. Let v be the second vertex on the path. Let p' be the subpath of p from v to t (p' might be a zero-length path). That is, the path p looks like $u \to v \overset{p'}{\rightsquigarrow} t$.

We claim that p' is a longest path from v to t.

To prove the claim, we use a cut-and-paste argument. If p' were not a longest path, then there exists a longer path p'' from v to t. We could cut out p' and paste in p'' to produce a path $u \to v \overset{p''}{\rightsquigarrow} t$ which is longer than p, thus contradicting the assumption that p is a longest path from u to t.

It is important to note that the graph is *acyclic*. Because the graph is acyclic, path p'' cannot include the vertex u, for otherwise there would be a cycle of the form $u \to v \rightsquigarrow u$ in the graph. Thus, we can indeed use p'' to construct a longer path. The acyclicity requirement ensures that by pasting in path p'', the overall path is still a *simple* path (there is no cycle in the path). This difference between the cyclic and the acyclic case allows us to use dynamic programming to solve the acyclic case.

Let $dist[u]$ denote the weight of a longest path from u to t. The optimal substructure property allows us to write a recurrence for $dist[u]$ as

$$dist[u] = \begin{cases} 0 & \text{if } u = t\ , \\ \max\limits_{(u,v) \in E} \{w(u, v) + dist[v]\} & \text{otherwise}\ . \end{cases}$$

This recurrence allows us to construct the following procedure:

```
LONGEST-PATH-AUX(G, u, t, dist, next)
  if u == t
      dist[u] = 0
      return (dist, next)
  elseif next[u] ≥ 0
      return (dist, next)
  else next[u] = 0
      for each vertex v ∈ G.Adj[u]
          (dist, next) = LONGEST-PATH-AUX(G, v, t, dist, next)
          if w(u, v) + dist[v] > dist[u]
              dist[u] = w(u, v) + dist[v]
              next[u] = v
  return (dist, next)
```

(See Section 22.1 for an explanation of the notation $G.Adj[u]$.)

LONGEST-PATH-AUX is a memoized, recursive procedure, which returns the tuple $(dist, next)$. The array *dist* is the memoized array that holds the solution to subproblems. That is, after the procedure returns, $dist[u]$ will hold the weight of a longest path from u to t. The array *next* serves two purposes:

- It holds information necessary for printing out an actual path. Specifically, if u is a vertex on the longest path that the procedure found, then $next[u]$ is the next vertex on the path.
- The value in $next[u]$ is used to check whether the current subproblem has been solved earlier. A value of at least zero indicates that this subproblem has been solved earlier.

The first **if** condition checks for the base case $u = t$. The second **if** condition checks whether the current subproblem has already been solved. The **for** loop iterates over each adjacent edge (u, v) and updates the longest distance in $dist[u]$.

What is the running time of LONGEST-PATH-AUX? Each subproblem represented by a vertex u is solved at most once due to the memoization. For each vertex, we examine its adjacent edges. Thus, each edge is examined at most once, and the overall running time is $O(E)$. (Section 22.1 discusses how we achieve $O(E)$ time by representing the graph with adjacency lists.)

The PRINT-PATH procedure prints out the path using information stored in the *next* array:

```
PRINT-PATH(s, t, next)
  u = s
  print u
  while u ≠ t
      print "→" next[u]
      u = next[u]
```

The LONGEST-PATH-MAIN procedure is the main driver. It creates and initializes the *dist* and the *next* arrays. It then calls LONGEST-PATH-AUX to find a path and PRINT-PATH to print out the actual path.

```
LONGEST-PATH-MAIN(G, s, t)
  n = |G.V|
  let dist[1..n] and next[1..n] be new arrays
  for i = 1 to n
      dist[i] = −∞
      next[i] = −1
  (dist, next) = LONGEST-PATH-AUX(G, s, t, dist, next)
  if dist[s] == −∞
      print "No path exists"
  else print "The weight of the longest path is " dist[s]
      PRINT-PATH(s, t, next)
```

Initializating the *dist* and *next* arrays takes $O(V)$ time. Thus the overall running time of LONGEST-PATH-MAIN is $O(V + E)$.

Alternative solution

We can also solve the problem using a bottom-up aproach. To do so, we need to ensure that we solve "smaller" subproblems before we solve "larger" ones. In our case, we can use a *topological sort* (see Section 22.4) to obtain a bottom-up procedure, imposing the required ordering on the vertices in $\Theta(V + E)$ time.

```
LONGEST-PATH2(G, s, t)
  let dist[1..n] and next[1..n] be new arrays
  topologically sort the vertices of G
  for i = 1 to |G.V|
      dist[i] = −∞
  dist[s] = 0
  for each u in topological order, starting from s
      for each edge (u, v) ∈ G.Adj[u]
          if dist[u] + w(u, v) > dist[v]
              dist[v] = dist[u] + w(u, v)
              next[u] = v
  print "The longest distance is " dist[t]
  PRINT-PATH(s, t, next)
```

The running time of LONGEST-PATH2 is $\Theta(V + E)$.

Solution to Problem 15-2

We solve the longest palindrome subsequence (LPS) problem in a manner similar to how we compute the longest common subsequence in Section 15.4.

Step 1: Characterizing a longest palindrome subsequence

The LPS problem has an optimal-substructure property, where the subproblems correspond to pairs of indices, starting and ending, of the input sequence.

For a sequence $X = \langle x_1, x_2, \ldots, x_n \rangle$, we denote the subsequence starting at x_i and ending at x_j by $X_{ij} = \langle x_i, x_{i+1}, \ldots, x_j \rangle$.

Theorem (Optimal substructure of an LPS)
Let $X = \langle x_1, x_2, \ldots, x_n \rangle$ be the input sequence, and let $Z = \langle z_1, z_2, \ldots, z_m \rangle$ be any LPS of X.

1. If $n = 1$, then $m = 1$ and $z_1 = x_1$.
2. If $n = 2$ and $x_1 = x_2$, then $m = 2$ and $z_1 = z_2 = x_1 = x_2$.
3. If $n = 2$ and $x_1 \neq x_2$, then $m = 1$ and z_1 is equal to either x_1 or x_n.
4. If $n > 2$ and $x_1 = x_n$, then $m > 2$, $z_1 = z_m = x_1 = x_n$, and $Z_{2,m-1}$ is an LPS of $X_{2,n-1}$.
5. If $n > 2$ and $x_1 \neq x_n$, then $z_1 \neq x_1$ implies that $Z_{1,m}$ is an LPS of $X_{2,n}$.
6. If $n > 2$ and $x_1 \neq x_n$, then $z_m \neq x_n$ implies that $Z_{1,m}$ is an LPS of $X_{1,n-1}$.

Proof Properties (1), (2), and (3) follow trivially from the definition of LPS.

(4) If $n > 2$ and $x_1 = x_n$, then we can choose x_1 and x_n as the ends of Z and at least one more element of X as part of Z. Thus, it follows that $m > 2$. If $z_1 \neq x_1$, then we could append $x_1 = x_n$ to the ends of Z to obtain a palindrome subsequence of X with length $m + 2$, contradicting the supposition that Z is a *longest* palindrome subsequence of X. Thus, we must have $z_1 = x_1$ $(= x_n = z_m)$. Now, $Z_{2,m-1}$ is a length-$(m-2)$ palindrome subsequence of $X_{2,n-1}$. We wish to show that it is an LPS. Suppose for the purpose of contradiction that there exists a palindrome subsequence W of $X_{2,n-1}$ with length greater than $m - 2$. Then, appending $x_1 = x_n$ to the ends of W produces a palindrome subsequence of X whose length is greater than m, which is a contradiction.

(5) If $z_1 \neq x_1$, then Z is a palindrome subsequence of $X_{2,n}$. If there were a palindrome subsequence W of $X_{2,n}$ with length greater than m, then W would also be a palindrome subsequence of X, contradicting the assumption that Z is an LPS of X.

(6) The proof is symmetric to (2). ■

The way that the theorem characterizes longest palindrome subsequences tells us that an LPS of a sequence contains within it an LPS of a subsequence of the sequence. Thus, the LPS problem has an optimal-substructure property.

Step 2: A recursive solution

The theorem implies that we should examine either one or two subproblems when finding an LPS of $X = \langle x_1, x_2, \ldots, x_n \rangle$, depending on whether $x_1 = x_n$.

Let us define $p[i, j]$ to be the length of an LPS of the subsequence X_{ij}. If $i = j$, the LPS has length 1. If $j = i + 1$, then the LPS has length either 1 or 2, depending on whether $x_i = x_j$. The optimal substructure of the LPS problem gives the following recursive formula:

$$p[i,j] = \begin{cases} 1 & \text{if } i = j\ , \\ 2 & \text{if } j = i+1 \text{ and } x_i = x_j\ , \\ 1 & \text{if } j = i+1 \text{ and } x_i \neq x_j\ , \\ p[i+1, j-1] + 2 & \text{if } j > i+1 \text{ and } x_i = x_j\ , \\ \max(p[i, j-1], p[i+1, j]) & \text{if } j > i+1 \text{ and } x_i \neq x_j\ . \end{cases}$$

Step 3: Computing the length of an LPS

Procedure LONGEST-PALINDROME takes a sequence $X = \langle x_1, x_2, \ldots, x_n \rangle$ as input. The procedure fills cells $p[i,i]$, where $1 \le i \le n$, and $p[i,i+1]$, where $1 \le i \le n-1$, as the base cases. It then starts filling cells $p[i,j]$, where $j > i+1$. The procedure fills the p table row by row, starting with row $n-2$ and moving toward row 1. (Rows $n-1$ and n are already filled as part of the base cases.) Within each row, the procedure fills the entries from left to right. The procedure also maintains the table $b[1..n, 1..n]$ to help us construct an optimal solution. Intuitively, $b[i,j]$ points to the table entry corresponding to the optimal subproblem solution chosen when computing $p[i,j]$. The procedure returns the b and p tables; $p[1,n]$ contains the length of an LPS of X. The running time of LONGEST-PALINDROME is clearly $\Theta(n^2)$.

```
LONGEST-PALINDROME(X)
  n = X.length
  let b[1..n, 1..n] and p[0..n, 0..n] be new tables
  for i = 1 to n − 1
      p[i, i] = 1
      j = i + 1
      if x_i == x_j
          p[i, j] = 2
          b[i, j] = "↙"
      else p[i, j] = 1
          b[i, j] = "↓"
  p[n, n] = 1
  for i = n − 2 downto 1
      for j = i + 2 to n
          if x_i == x_j
              p[i, j] = p[i + 1, j − 1] + 2
              b[i, j] = "↙"
          elseif p[i + 1, j] ≥ p[i, j − 1]
              p[i, j] = p[i + 1, j]
              b[i, j] = "↓"
          else p[i, j] = p[i, j − 1]
              b[i, j] = "←"
  return p and b
```

Step 4: Constructing an LPS

The b table returned by LONGEST-PALINDROME enables us to quickly construct an LPS of $X = \langle x_1, x_2, \ldots, x_m \rangle$. We simply begin at $b[1, n]$ and trace through the table by following the arrows. Whenever we encounter a "$\swarrow$" in entry $b[i, j]$, it implies that $x_i = y_j$ are the first and last elements of the LPS that LONGEST-PALINDROME found. The following recursive procedure returns a sequence S that contains an LPS of X. The initial call is GENERATE-LPS$(b, X, 1, X.length, \langle\rangle)$, where $\langle\rangle$ denotes an empty sequence. Within the procedure, the symbol $||$ denotes concatenation of a symbol and a sequence.

```
GENERATE-LPS(b, X, i, j, S)
  if i > j
      return S
  elseif i == j
      return S || x_i
  elseif b[i, j] == "↙"
      return x_i || GENERATE-LPS(b, X, i + 1, j − 1, S) || x_i
  elseif b[i, j] == "↓"
      return GENERATE-LPS(b, X, i + 1, j, S)
  else return GENERATE-LPS(b, X, i, j − 1, S)
```

Solution to Problem 15-3

Taking the book's hint, we sort the points by x-coordinate, left to right, in $O(n \lg n)$ time. Let the sorted points be, left to right, $\langle p_1, p_2, p_3, \ldots, p_n \rangle$. Therefore, p_1 is the leftmost point, and p_n is the rightmost.

We define as our subproblems paths of the following form, which we call bitonic paths. A ***bitonic path*** $P_{i,j}$, where $i \le j$, includes all points $p_1, p_2, \ldots, p_j$; it starts at some point p_i, goes strictly left to point p_1, and then goes strictly right to point p_j. By "going strictly left," we mean that each point in the path has a lower x-coordinate than the previous point. Looked at another way, the indices of the sorted points form a strictly decreasing sequence. Likewise, "going strictly right" means that the indices of the sorted points form a strictly increasing sequence. Moreover, $P_{i,j}$ contains all the points $p_1, p_2, p_3, \ldots, p_j$. Note that p_j is the rightmost point in $P_{i,j}$ and is on the rightgoing subpath. The leftgoing subpath may be degenerate, consisting of just p_1.

Let us denote the euclidean distance between any two points p_i and p_j by $|p_i p_j|$. And let us denote by $b[i, j]$, for $1 \le i \le j \le n$, the length of the shortest bitonic path $P_{i,j}$. Since the leftgoing subpath may be degenerate, we can easily compute all values $b[1, j]$. The only value of $b[i, i]$ that we will need is $b[n, n]$, which is the length of the shortest bitonic tour. We have the following formulation of $b[i, j]$ for $1 \le i \le j \le n$:

$$
\begin{aligned}
b[1,2] &= |p_1 p_2| \ , \\
b[i,j] &= b[i,j-1] + |p_{j-1} p_j| \quad \text{for } i < j-1 \ , \\
b[j-1,j] &= \min_{1 \le k < j-1} \{ b[k,j-1] + |p_k p_j| \} \ .
\end{aligned}
$$

Why are these formulas correct? Any bitonic path ending at p_2 has p_2 as its rightmost point, so it consists only of p_1 and p_2. Its length, therefore, is $|p_1 p_2|$.

Now consider a shortest bitonic path $P_{i,j}$. The point p_{j-1} is somewhere on this path. If it is on the rightgoing subpath, then it immediately preceeds p_j on this subpath. Otherwise, it is on the leftgoing subpath, and it must be the rightmost point on this subpath, so $i = j - 1$. In the first case, the subpath from p_i to p_{j-1} must be a shortest bitonic path $P_{i,j-1}$, for otherwise we could use a cut-and-paste argument to come up with a shorter bitonic path than $P_{i,j}$. (This is part of our optimal substructure.) The length of $P_{i,j}$, therefore, is given by $b[i, j-1] + |p_{j-1} p_j|$. In the second case, p_j has an immediate predecessor p_k, where $k < j - 1$, on the rightgoing subpath. Optimal substructure again applies: the subpath from p_k to p_{j-1} must be a shortest bitonic path $P_{k,j-1}$, for otherwise we could use cut-and-paste to come up with a shorter bitonic path than $P_{i,j}$. (We have implicitly relied on paths having the same length regardless of which direction we traverse them.) The length of $P_{i,j}$, therefore, is given by $\min_{1 \le k \le j-1} \{b[k, j-1] + |p_k p_j|\}$.

We need to compute $b[n, n]$. In an optimal bitonic tour, one of the points adjacent to p_n must be p_{n-1}, and so we have

$$b[n, n] = b[n-1, n] + |p_{n-1} p_n| \ .$$

To reconstruct the points on the shortest bitonic tour, we define $r[i, j]$ to be the index of the immediate predecessor of p_j on the shortest bitonic path $P_{i,j}$. Because the immediate predecessor of p_2 on $P_{1,2}$ is p_1, we know that $r[1, 2]$ must be 1. The pseudocode below shows how we compute $b[i, j]$ and $r[i, j]$. It fills in only entries $b[i, j]$ where $1 \le i \le n-1$ and $i + 1 \le j \le n$, or where $i = j = n$, and only entries $r[i, j]$ where $1 \le i \le n-2$ and $i + 2 \le j \le n$.

```
EUCLIDEAN-TSP(p)
  sort the points so that ⟨p_1, p_2, p_3, ..., p_n⟩ are in order of increasing x-coordinate
  let b[1..n, 2..n] and r[1..n-2, 3..n] be new arrays
  b[1,2] = |p_1 p_2|
  for j = 3 to n
      for i = 1 to j - 2
          b[i, j] = b[i, j-1] + |p_{j-1} p_j|
          r[i, j] = j - 1
      b[j-1, j] = ∞
      for k = 1 to j - 2
          q = b[k, j-1] + |p_k p_j|
          if q < b[j-1, j]
              b[j-1, j] = q
              r[j-1, j] = k
  b[n, n] = b[n-1, n] + |p_{n-1} p_n|
  return b and r
```

We print out the tour we found by starting at p_n, then a leftgoing subpath that includes p_{n-1}, from right to left, until we hit p_1. Then we print right-to-left the remaining subpath, which does not include p_{n-1}. For the example in Figure 15.11(b) on page 405, we wish to print the sequence $p_7, p_6, p_4, p_3, p_1, p_2, p_5$. Our code is recursive. The right-to-left subpath is printed as we go deeper into the recursion, and the left-to-right subpath is printed as we back out.

```
PRINT-TOUR(r, n)
  print p_n
  print p_{n-1}
  k = r[n − 1, n]
  PRINT-PATH(r, k, n − 1)
  print p_k

PRINT-PATH(r, i, j)
  if i < j
      k = r[i, j]
      if k ≠ i
          print p_k
      if k > 1
          PRINT-PATH(r, i, k)
  else k = r[j, i]
      if k > 1
          PRINT-PATH(r, k, j)
          print p_k
```

The relative values of the parameters i and j in each call of PRINT-PATH indicate which subpath we're working on. If $i < j$, we're on the right-to-left subpath, and if $i > j$, we're on the left-to-right subpath. The test for $k \neq i$ prevents us from printing p_1 an extra time, which could occur when we call PRINT-PATH$(r, 1, 2)$.

The time to run EUCLIDEAN-TSP is $O(n^2)$ since the outer loop on j iterates $n-2$ times and the inner loops on i and k each run at most $n-2$ times. The sorting step at the beginning takes $O(n \lg n)$ time, which the loop times dominate. The time to run PRINT-TOUR is $O(n)$, since each point is printed just once.

Solution to Problem 15-4

This solution is also posted publicly

Note: We assume that no word is longer than will fit into a line, i.e., $l_i \leq M$ for all i.

First, we'll make some definitions so that we can state the problem more uniformly. Special cases about the last line and worries about whether a sequence of words fits in a line will be handled in these definitions, so that we can forget about them when framing our overall strategy.

- Define $extras[i, j] = M - j + i - \sum_{k=i}^{j} l_k$ to be the number of extra spaces at the end of a line containing words i through j. Note that *extras* may be negative.
- Now define the cost of including a line containing words i through j in the sum we want to minimize:

$$lc[i, j] = \begin{cases} \infty & \text{if } extras[i, j] < 0 \text{ (i.e., words } i, \ldots, j \text{ don't fit)}, \\ 0 & \text{if } j = n \text{ and } extras[i, j] \geq 0 \text{ (last line costs 0)}, \\ (extras[i, j])^3 & \text{otherwise}. \end{cases}$$

By making the line cost infinite when the words don't fit on it, we prevent such an arrangement from being part of a minimal sum, and by making the cost 0 for the last line (if the words fit), we prevent the arrangement of the last line from influencing the sum being minimized.

We want to minimize the sum of lc over all lines of the paragraph.

Our subproblems are how to optimally arrange words $1, \ldots, j$, where $j = 1, \ldots, n$.

Consider an optimal arrangement of words $1, \ldots, j$. Suppose we know that the last line, which ends in word j, begins with word i. The preceding lines, therefore, contain words $1, \ldots, i-1$. In fact, they must contain an optimal arrangement of words $1, \ldots, i-1$. (The usual type of cut-and-paste argument applies.)

Let $c[j]$ be the cost of an optimal arrangement of words $1, \ldots, j$. If we know that the last line contains words $i, \ldots, j$, then $c[j] = c[i-1] + lc[i, j]$. As a base case, when we're computing $c[1]$, we need $c[0]$. If we set $c[0] = 0$, then $c[1] = lc[1, 1]$, which is what we want.

But of course we have to figure out which word begins the last line for the subproblem of words $1, \ldots, j$. So we try all possibilities for word i, and we pick the one that gives the lowest cost. Here, i ranges from 1 to j. Thus, we can define $c[j]$ recursively by

$$c[j] = \begin{cases} 0 & \text{if } j = 0\,, \\ \min\limits_{1 \le i \le j} (c[i-1] + lc[i, j]) & \text{if } j > 0\,. \end{cases}$$

Note that the way we defined lc ensures that

- all choices made will fit on the line (since an arrangement with $lc = \infty$ cannot be chosen as the minimum), and
- the cost of putting words $i, \ldots, j$ on the last line will not be 0 unless this really is the last line of the paragraph ($j = n$) or words $i \ldots j$ fill the entire line.

We can compute a table of c values from left to right, since each value depends only on earlier values.

To keep track of what words go on what lines, we can keep a parallel p table that points to where each c value came from. When $c[j]$ is computed, if $c[j]$ is based on the value of $c[k-1]$, set $p[j] = k$. Then after $c[n]$ is computed, we can trace the pointers to see where to break the lines. The last line starts at word $p[n]$ and goes through word n. The previous line starts at word $p[p[n]]$ and goes through word $p[n] - 1$, etc.

In pseudocode, here's how we construct the tables:

```
PRINT-NEATLY(l, n, M)
  let extras[1..n, 1..n], lc[1..n, 1..n], and c[0..n] be new arrays
  // Compute extras[i, j] for 1 ≤ i ≤ j ≤ n.
  for i = 1 to n
      extras[i, i] = M − l_i
      for j = i + 1 to n
          extras[i, j] = extras[i, j − 1] − l_j − 1
  // Compute lc[i, j] for 1 ≤ i ≤ j ≤ n.
  for i = 1 to n
      for j = i to n
          if extras[i, j] < 0
              lc[i, j] = ∞
          elseif j == n and extras[i, j] ≥ 0
              lc[i, j] = 0
          else lc[i, j] = (extras[i, j])^3
  // Compute c[j] and p[j] for 1 ≤ j ≤ n.
  c[0] = 0
  for j = 1 to n
      c[j] = ∞
      for i = 1 to j
          if c[i − 1] + lc[i, j] < c[j]
              c[j] = c[i − 1] + lc[i, j]
              p[j] = i
  return c and p
```

Quite clearly, both the time and space are $\Theta(n^2)$.

In fact, we can do a bit better: we can get both the time and space down to $\Theta(nM)$. The key observation is that at most $\lceil M/2 \rceil$ words can fit on a line. (Each word is at least one character long, and there's a space between words.) Since a line with words $i, \ldots, j$ contains $j - i + 1$ words, if $j - i + 1 > \lceil M/2 \rceil$ then we know that $lc[i, j] = \infty$. We need only compute and store $extras[i, j]$ and $lc[i, j]$ for $j - i + 1 \le \lceil M/2 \rceil$. And the inner **for** loop header in the computation of $c[j]$ and $p[j]$ can run from $\max(1, j - \lceil M/2 \rceil + 1)$ to j.

We can reduce the space even further to $\Theta(n)$. We do so by not storing the *lc* and *extras* tables, and instead computing the value of $lc[i, j]$ as needed in the last loop. The idea is that we could compute $lc[i, j]$ in $O(1)$ time if we knew the value of $extras[i, j]$. And if we scan for the minimum value in *descending* order of i, we can compute that as $extras[i, j] = extras[i + 1, j] - l_i - 1$. (Initially, $extras[j, j] = M - l_j$.) This improvement reduces the space to $\Theta(n)$, since now the only tables we store are c and p.

Here's how we print which words are on which line. The printed output of GIVE-LINES(p, j) is a sequence of triples (k, i, j), indicating that words $i, \ldots, j$ are printed on line k. The return value is the line number k.

```
GIVE-LINES(p, j)
  i = p[j]
  if i == 1
      k = 1
  else k = GIVE-LINES(p, i − 1) + 1
  print (k, i, j)
  return k
```

The initial call is GIVE-LINES(p, n). Since the value of j decreases in each recursive call, GIVE-LINES takes a total of $O(n)$ time.

Solution to Problem 15-5

a. Dynamic programming is the ticket. This problem is slightly similar to the longest-common-subsequence problem. In fact, we'll define the notational conveniences X_i and Y_j in the similar manner as we did for the LCS problem: $X_i = x[1..i]$ and $Y_j = y[1..j]$.

Our subproblems will be determining an optimal sequence of operations that converts X_i to Y_j, for $0 \le i \le m$ and $0 \le j \le n$. We'll call this the "$X_i \to Y_j$ problem." The original problem is the $X_m \to Y_n$ problem.

Let's suppose for the moment that we know what was the last operation used to convert X_i to Y_j. There are six possibilities. We denote by $c[i, j]$ the cost of an optimal solution to the $X_i \to Y_j$ problem.

- If the last operation was a copy, then we must have had $x[i] = y[j]$. The subproblem that remains is converting X_{i-1} to Y_{j-1}. And an optimal solution to the $X_i \to Y_j$ problem must include an optimal solution to the $X_{i-1} \to Y_{j-1}$ problem. The cut-and-paste argument applies. Thus, assuming that the last operation was a copy, we have $c[i, j] = c[i-1, j-1] + \text{cost(copy)}$.
- If it was a replace, then we must have had $x[i] \ne y[j]$. (Here, we assume that we cannot replace a character with itself. It is a straightforward modification if we allow replacement of a character with itself.) We have the same optimal substructure argument as for copy, and assuming that the last operation was a replace, we have $c[i, j] = c[i-1, j-1] + \text{cost(replace)}$.
- If it was a twiddle, then we must have had both $x[i] = y[j-1]$ and $x[i-1] = y[j]$, along with the implicit assumption that $i, j \ge 2$. Now our subproblem is $X_{i-2} \to Y_{j-2}$ and, assuming that the last operation was a twiddle, we have $c[i, j] = c[i-2, j-2] + \text{cost(twiddle)}$.
- If it was a delete, then we have no restrictions on x or y. Since we can view delete as removing a character from X_i and leaving Y_j alone, our subproblem is $X_{i-1} \to Y_j$. Assuming that the last operation was a delete, we have $c[i, j] = c[i-1, j] + \text{cost(delete)}$.
- If it was an insert, then we have no restrictions on x or y. Our subproblem is $X_i \to Y_{j-1}$. Assuming that the last operation was an insert, we have $c[i, j] = c[i, j-1] + \text{cost(insert)}$.

- If it was a kill, then we had to have completed converting X_m to Y_n, so that the current problem must be the $X_m \to Y_n$ problem. In other words, we must have $i = m$ and $j = n$. If we think of a kill as a multiple delete, we can get any $X_i \to Y_n$, where $0 \le i < m$, as a subproblem. We pick the best one, and so assuming that the last operation was a kill, we have

  $$c[m,n] = \min_{0 \le i < m} \{c[i,n]\} + \text{cost(kill)} \; .$$

We have not handled the base cases, in which $i = 0$ or $j = 0$. These are easy. X_0 and Y_0 are the empty strings. We convert an empty string into Y_j by a sequence of j inserts, so that $c[0, j] = j \cdot \text{cost(insert)}$. Similarly, we convert X_i into Y_0 by a sequence of i deletes, so that $c[i, 0] = i \cdot \text{cost(delete)}$. When $i = j = 0$, either formula gives us $c[0, 0] = 0$, which makes sense, since there's no cost to convert the empty string to the empty string.

For $i, j > 0$, our recursive formulation for $c[i, j]$ applies the above formulas in the situations in which they hold:

$$c[i,j] = \min \begin{cases} c[i-1, j-1] + \text{cost(copy)} & \text{if } x[i] = y[j] \; , \\ c[i-1, j-1] + \text{cost(replace)} & \text{if } x[i] \ne y[j] \; , \\ c[i-2, j-2] + \text{cost(twiddle)} & \text{if } i, j \ge 2, \\ & \quad x[i] = y[j-1], \\ & \quad \text{and } x[i-1] = y[j] \; , \\ c[i-1, j] + \text{cost(delete)} & \text{always} \; , \\ c[i, j] = c[i, j-1] + \text{cost(insert)} & \text{always} \; , \\ \min\limits_{0 \le i < m} \{c[i,n]\} + \text{cost(kill)} & \text{if } i = m \text{ and } j = n \; . \end{cases}$$

Like we did for LCS, our pseudocode fills in the table in row-major order, i.e., row-by-row from top to bottom, and left to right within each row. Column-major order (column-by-column from left to right, and top to bottom within each column) would also work. Along with the $c[i, j]$ table, we fill in the table $op[i, j]$, holding which operation was used.

```
EDIT-DISTANCE(x, y, m, n)
  let c[0..m, 0..n] and op[0..m, 0..n] be new arrays
  for i = 0 to m
      c[i, 0] = i · cost(delete)
      op[i, 0] = DELETE
  for j = 0 to n
      c[0, j] = j · cost(insert)
      op[0, j] = INSERT
  for i = 1 to m
      for j = 1 to n
          c[i, j] = ∞
          if x[i] == y[j]
              c[i, j] = c[i − 1, j − 1] + cost(copy)
              op[i, j] = COPY
          if x[i] ≠ y[j] and c[i − 1, j − 1] + cost(replace) < c[i, j]
              c[i, j] = c[i − 1, j − 1] + cost(replace)
              op[i, j] = REPLACE(by y[j])
          if i ≥ 2 and j ≥ 2 and x[i] == y[j − 1] and
              x[i − 1] == y[j] and
              c[i − 2, j − 2] + cost(twiddle) < c[i, j]
              c[i, j] = c[i − 2, j − 2] + cost(twiddle)
              op[i, j] = TWIDDLE
          if c[i − 1, j] + cost(delete) < c[i, j]
              c[i, j] = c[i − 1, j] + cost(delete)
              op[i, j] = DELETE
          if c[i, j − 1] + cost(insert) < c[i, j]
              c[i, j] = c[i, j − 1] + cost(insert)
              op[i, j] = INSERT(y[j])
  for i = 0 to m − 1
      if c[i, n] + cost(kill) < c[m, n]
          c[m, n] = c[i, n] + cost(kill)
          op[m, n] = KILL i
  return c and op
```

The time and space are both $\Theta(mn)$. If we store a KILL operation in $op[m, n]$, we also include the index i after which we killed, to help us reconstruct the optimal sequence of operations. (We don't need to store $y[i]$ in the *op* table for replace or insert operations.)

To reconstruct this sequence, we use the *op* table returned by EDIT-DISTANCE. The procedure OP-SEQUENCE(op, i, j) reconstructs the optimal operation sequence that we found to transform X_i into Y_j. The base case is when $i = j = 0$. The first call is OP-SEQUENCE(op, m, n).

```
OP-SEQUENCE(op, i, j)
  if i == 0 and j = 0
      return
  if op[i, j] == COPY or op[i, j] = REPLACE
      i' = i − 1
      j' = j − 1
  elseif op[i, j] == TWIDDLE
      i' = i − 2
      j' = j − 2
  elseif op[i, j] == DELETE
      i' = i − 1
      j' = j
  elseif op[i, j] == INSERT        // don't care yet what character is inserted
      i' = i
      j' = j − 1
  else          // must be KILL, and must have i = m and j = n
      let op[i, j] == KILL k
      i' = k
      j' = j
  OP-SEQUENCE(op, i', j')
  print op[i, j]
```

This procedure determines which subproblem we used, recurses on it, and then prints its own last operation.

b. The DNA-alignment problem is just the edit-distance problem, with

$$\begin{aligned}\text{cost(copy)} &= -1\,,\\ \text{cost(replace)} &= +1\,,\\ \text{cost(delete)} &= +2\,,\\ \text{cost(insert)} &= +2\,,\end{aligned}$$

and the twiddle and kill operations are not permitted.

The score that we are trying to maximize in the DNA-alignment problem is precisely the negative of the cost we are trying to minimize in the edit-distance problem. The negative cost of copy is not an impediment, since we can only apply the copy operation when the characters are equal.

Solution to Problem 15-8

a. Let us set up a recurrence for the number of valid seams as a function of m. Suppose we are in the process of carving out a seam row by row, starting from the first row. Let the last pixel carved out be $A[i, j]$. How many choices do we have for the pixel in row $i + 1$ such that the pixel continues the seam? If the last pixel $A[i, j]$ were on the column boundary ($i = 1$ or $i = n$), then there would be two choices for the next pixel. For example, when $i = 1$, the two choices for the next pixel are $A[i + 1, j]$ and $A[i + 1, j + 1]$. Otherwise, there would

be three choices for the next pixel: $A[i+1, j-1]$, $A[i+1, j]$, $A[i+1, j+1]$. Thus, for a general pixel $A[i, j]$, there are at least two possible choices for a pixel p in the next row such that p continues a seam ending in $A[i, j]$. Let $T(i)$ denote the number of possible seams from row 1 to row i. Then, for $i = 1$, we have $T(i) = n$, and for $i > 1$,

$$T(i) \geq 2T(i-1) \ .$$

It is easy to guess that $T(i) \geq n2^{i-1}$, which we verify by direct substitution. For $i = 1$, we have $T(1) = n \geq n \cdot 2^0$. For $i > 1$, we have

$$\begin{aligned} T(i) &\geq 2T(i-1) \\ &\geq 2 \cdot n2^{i-2} \\ &= n2^{i-1} \ . \end{aligned}$$

Thus, the total number $T(m)$ of seams is at least $n2^{m-1}$. We conclude that the number of seams grows at least exponentially in m.

b. As proved in the previous part, it is infeasible to systematically check every seam, since the number of possible seams grows exponentially.

The structure of the problem allows us to build the solution row by row. Consider a pixel $A[i, j]$. We ask the question: "If i were the first row of the picture, what is the minimum disruptive measure of seams that start with the pixel $A[i, j]$?"

Let S^* be a seam of minimum disruptive measure among all seams that start with pixel $A[i, j]$. Let $A[i+1, p]$, where $p \in \{j-1, j, j+1\}$, be the pixel of S^* in the next row. Let S' be the sub-seam of S^* that starts with $A[i+1, p]$. We claim that S' has the minimum disruptive measure among seams that start with $A[i+1, p]$. Why? Suppose there exists another seam S'' that starts with $A[i+1, p]$ and has disruptive measure less than that of S'. By using S'' as the sub-seam instead of S', we can obtain another seam that starts with $A[i, j]$ and has a disruptive measure which is less than that of S^*. Thus, we obtain a contradiction to our assumption that S^* is a seam of minimum disruptive measure.

Let $disr[i, j]$ be the value of the minimum disruptive measure among all seams that start with pixel $A[i, j]$. For row m, the seam with the minimum disruptive measure consists of just one point. We can now state a recurrence for $disr[i, j]$ as follows. In the base case, $disr[m, j] = d[m, j]$ for $j = 1, 2, \ldots, n$. In the recursive case, for $j = 1, 2, \ldots, n$,

$$disr[i, j] = d[i, j] + \min_{k \in K} \{disr[i+i, j+k]\} \ ,$$

where the set K of index offsets is

$$K = \begin{cases} \{0, 1\} & \text{if } j = 1 \ , \\ \{-1, 0, 1\} & \text{if } 1 < j < m \ , \\ \{-1, 0\} & \text{if } j = n \ . \end{cases}$$

Since every seam has to start with a pixel of the first row, we simply find the minimum $disr[1, j]$ for pixels in the first row to obtain the minimum disruptive measure.

```
COMPRESS-IMAGE(d)
  m = d.rows
  n = d.columns
  let disr[1..m, 1..n] and next[1..m, 1..n] be new tables
  for j = 1 to n
      disr[m, j] = d[m, j]
  for i = m − 1 downto 1
      for j = 1 to n
          low = max(−1, 1 − j)
          high = min(1, n − j)
          disr[i, j] = ∞
          for k = low to high
              if disr[i + 1, j + k] < disr[i, j]
                  disr[i, j] = disr[i + 1, j + k]
                  next[i, j] = j + k
          disr[i, j] = disr[i, j] + d[i, j]
  val = ∞
  start = 1
  for j = 1 to n
      if disr[1, j] < val
          val = disr[1, j]
          start = j
  print "The minimum value of the disruptive measure is " val
  for i = 1 to m
      print "cut point at " (i, start)
      start = next[i, start]
```

The procedure COMPRESS-IMAGE is simply an implementation of this recurrence in a bottom-up fashion.

We first carry out the initialization of the base cases, which are the cases when row $i = m$. The minimum disruptive measure for the base cases is simply $d[m, j]$.

The next **for** loop runs down from $m - 1$ to 1. Thus, $disr[i + 1, j]$ is already available before computing $disr[i, j]$ for pixels of row i.

The assignments to *low* and *high* allow the index offset k to range over the correct set K from above. We set *low* to 0 when $j = 1$ and to -1 when $j > 1$, and we set *high* to 0 when $j = n$ and to 1 when $j < n$. The innermost **for** loop sets $disr[i, j]$ to the minimum value of $disr[i + 1, j + k]$ for all $k \in K$, and the line that follows this loop adds in $d[i, j]$.

We use the *next* table to reconstruct the actual seam. For a given pixel, it records which pixel was used as the next pixel. Specifically, for a pixel $A[i, j]$, if $next[i, j] = p$, where $p \in \{j - 1, j, j + 1\}$, then the next pixel of the seam is $A[i + 1, p]$.

The last line of the **for** loop adds the disruptive measure of the current pixel to the disruptive measure of the seam.

The next **for** loop finds the minimum disruptive measure of pixels in the first row. We print the minimum disruptive measure as the answer.

The rest of the code reconstructs the actual seam, using the information stored in the *next* array.

Noting that the innermost **for** loop runs over at most three values of k, we see that the running time of COMPRESS-IMAGE is $O(mn)$. The space requirement is also $O(mn)$. We can improve upon the space requirement by observing that row i of the *disr* table depends on only row $i + 1$. Therefore, we can store just two rows at any time. Thus, we can improve the space requirement of COMPRESS-IMAGE to $O(n)$.

Solution to Problem 15-9

Our first step will be to identify the subproblems that satisfy the optimal-substructure property. Before we frame the subproblem, we make two simplifying modifications to the input:

- We sort L so that the indices in L are in ascending order.
- We prepend the index 0 to the beginning of L and append n to the end of L.

Let $L[i\,..\,j]$ denote a subarray of L that starts from index i and ends at index j. Define the subproblem denoted by (i, j) as "What is the cheapest sequence of breaks to break the substring $S[L[i] + 1\,..\,L[j]]$?" Note that the first and last elements of the subarray $L[i\,..\,j]$ define the ends of the substring, and we have to worry about only the indices of the subarray $L[i + 1\,..\,j - 1]$.

For example, let $L = \langle 20, 17, 14, 11, 25\rangle$ and $n = 30$. First, we sort L. Then, we prepend 0 and append n as explained to get $L = \langle 0, 11, 14, 17, 20, 25, 30\rangle$. Now, what is the subproblem $(2, 6)$? We obtain a substring by breaking S after character $L[2] = 11$ and character $L[6] = 25$. We ask "What is the cheapest sequence of breaks to break the substring $S[12\,..\,25]$?" We have to worry about only indices in the subarray $L[3\,..\,5] = \langle 14, 17, 20\rangle$, since the other indices are not present in the substring.

At this point, the problem looks similar to matrix-chain multiplication (see Section 15.2). We can make the first break at any element of $L[i + 1\,..\,j - 1]$.

Suppose that an optimal sequence of breaks σ for subproblem (i, j) makes the first break at $L[k]$, where $i < k < j$. This break gives rise to two subproblems:

- The "prefix" subproblem (i, k), covering the subarray $L[i + 1\,..\,k - 1]$,
- The "suffix" subproblem (k, j), covering the subarray $L[k + 1\,..\,j - 1]$.

The overall cost can be expressed as the sum of the length of the substring, the prefix cost, and the suffix cost.

We show optimal substructure by claiming that the sequence of breaks in σ for the prefix subproblem (i, k) must be an optimal one. Why? If there were a less costly way to break the substring $S[L[i]+1\,..\,L[k]]$ represented by the subproblem (i, k), then substituting that sequence of breaks in σ would produce another sequence of breaks whose cost is lower than that of σ, which would be a contradiction. A similar observation holds for the sequence of breaks for the suffix subproblem (k, j): it must be an optimal sequence of breaks.

Let $cost[i, j]$ denote the cost of the cheapest solution to subproblem (i, j). We write the recurrence relation for *cost* as

$$cost[i, j] = \begin{cases} 0 & \text{if } j - i \le 1 , \\ \min\limits_{i<k<j} \bigl\{cost[i, k] + cost[k, j] + (L[j] - L[i])\bigr\} & \text{if } j - i > 1 . \end{cases}$$

Thus, our approach to solving the subproblem (i, j) will be to try to split the respective substring at all possible values of k and then choosing a break that results in the minimum cost. We need to be careful to solve smaller subproblems before we solve larger subproblems. In particular, we solve subproblems in increasing order of the length $j - i$.

```
BREAK-STRING(n, L)
  prepend 0 to the start of L and append n to the end of L
  m = L.length
  sort L into increasing order
  let cost[1..m, 1..m] and break[1..m, 1..m] be new tables
  for i = 1 to m − 1
      cost[i, i] = cost[i, i + 1] = 0
  cost[m, m] = 0
  for len = 3 to m
      for i = 1 to m − len + 1
          j = i + len − 1
          cost[i, j] = ∞
          for k = i + 1 to j − 1
              if cost[i, k] + cost[k, j] < cost[i, j]
                  cost[i, j] = cost[i, k] + cost[k, j]
                  break[i, j] = k
          cost[i, j] = cost[i, j] + L[j] − L[i]
  print "The minimum cost of breaking the string is " cost[1, m]
  PRINT-BREAKS(L, break, 1, m)
```

After sorting L, we initialize the base cases, in which $i = j$ or $j = i + 1$.

The nested **for** loops represent the main computation. The outermost **for** loop runs for $len = 3$ to m, which means that we need to consider subarrays of L with length at least 3, since the first and the last element define the substring, and we need at least one more element to specify a break. The increasing values of *len* also ensures that we solve subproblems with smaller length before we solve subproblems with greater length.

The inner **for** loop on i runs from 1 to $m - len + 1$. The upper bound of $m - len + 1$ is the largest value that the start index i can take such that $i + len - 1 \le m$.

In the innermost **for** loop, we try each possible location k as the place to make the first break for subproblem (i, j). The first such place is $L[i+1]$, and not $L[i]$, since $L[i]$ represents the start of the substring (and thus not a valid place for a break). Similarly, the last valid place is $L[j - 1]$, because $L[j]$ represents the end of the substring.

The **if** condition tests whether k is the best place for a break found so far, and it updates the best value in $cost[i, j]$ if so. We use $break[i, j]$ to record that the

best place for the first break is k. Specifically, if $break[i, j] = k$, then an optimal sequence of breaks for (i, j) makes the first break at $L[k]$.

Finally, we add the length of the substring $L[j] - L[i]$ to $cost[i, j]$ because, irrespective of what we choose as the first break, it costs us a price equal to the length of the substring to make a break.

The lowest cost for the original problem ends up in $cost[1, m]$. By our initialization, $L[1] = 0$ and $L[m] = n$. Thus, $cost[1, m]$ will hold the optimum price of cutting the substring from $L[1] + 1 = 1$ to $L[m] = n$, which is the entire string.

The running time is $\Theta(m^3)$, and it is dictated by the three nested **for** loops. They fill in the entries above the main diagonal of the two tables, except for entries in which $j = i + 1$. That is, they fill in rows $i = 1, 2, \ldots, m - 2$, entries $j = i + 2, i + 3, \ldots, m$. When filling in entry $[i, j]$, we check values of k running from $i + 1$ to $j - 1$, or $j - i - 1$ entries. Thus, the total number of iterations of the innermost **for** loop is

$$\begin{aligned}
\sum_{i=1}^{m-2} \sum_{j=i+2}^{m} (j - i - 1) &= \sum_{i=1}^{m-2} \sum_{d=1}^{m-i-1} d && (d = j - i - 1) \\
&= \sum_{i=1}^{m-2} \Theta((m - i)^2) && \text{(equation (A.2))} \\
&= \sum_{h=2}^{m-1} \Theta(h^2) && (h = m - i) \\
&= \Theta(m^3) && \text{(equation (A.3))} .
\end{aligned}$$

Since each iteration of the innermost **for** loop takes constant time, the total running time is $\Theta(m^3)$. Note in particular that the running time is independent of the length of the string n.

```
PRINT-BREAKS(L, break, i, j)
  if j − i ≥ 2
      k = break[i, j]
      print "Break at " L[k]
      PRINT-BREAKS(L, break, i, k)
      PRINT-BREAKS(L, break, k, j)
```

PRINT-BREAKS uses the information stored in *break* to print out the actual sequence of breaks.

Solution to Problem 15-11

We state the subproblem (k, s) as "What is the cheapest way to satisfy all the demands of months $k, \ldots, n$ when we start with a surplus of s before the kth month?" A *plan* for the subproblem (k, s) would specify the number of machines to manufacture for each month $k, \ldots, n$ such that demands are satisfied.

In some optimal plan P to (k, s), let f^* machines be maufactured in month k. Thus, the surplus s' in month $k + 1$ is $s + f^* - d_k$. Let P' be the part of the

plan P for months $k+1, \ldots, n$. We claim that P' is an optimal plan for the subproblem $(k+1, s')$. Why? Suppose P' were not an optimal plan and let P'' be an optimal plan for $(k+1, s')$. If we modify plan P by cutting out P' and pasting in P'' (i.e., by using plan P'' for months $k+1, \ldots, n$), we obtain another plan for (k, s) which is cheaper than plan P. Thus, we obtain a contradiction to the assumption that plan P was optimal.

Let $cost[k, s]$ denote the cost of an optimal plan for (k, s), and let f denote the number of machines that can be manufactured in month k. The bounds for f are as follows:

- At least the number of machines so that (along with surplus s) there are enough machines to satisfy the current month's demand. Let us denote this lower bound by $L(k, s)$. We have

  $$L(k, s) = \max(d_k - s, 0) \ .$$

- At most the number of machines such that there are enough machines to satisfy the demands of all the following months. Let us denote this upper bound by $U(k, s)$. We have

  $$U(k, s) = \left(\sum_{i=k}^{n} d_i \right) - s \ .$$

For the last month, we need only manufacture the minimum required number of machines, given by $L(n, s)$. For other months, we examine the costs of manufacturing all feasible numbers of machines and see which choice gives us the cheapest plan. We can now write the recurrence for *cost* as the following:

$$cost[k, s] = \begin{cases} c \cdot \max(L(n, s) - m, 0) & \\ \qquad + h(s + L(n, s) - d_n) & \text{if } k = n \ , \\ \min\limits_{L(k,s) \le f \le U(k,s)} \Big\{ cost[k+1, s + f - d_k] & \\ \qquad + c \cdot \max(f - m, 0) & \\ \qquad + h(s + f - d_k) \Big\} & \text{if } 0 < k < n \ . \end{cases}$$

The recurrence suggests how to build an optimal plan in a bottom-up fashion. We now present the algorithm for constructing an optimal plan.

```
INVENTORY-PLANNING(n, m, c, D, d, h)
  let cost[1..n, 0..D] and make[1..n, 0..D] be new tables
  // Compute cost[n, 0..D] and make[n, 0..D].
  for s = 0 to D
      f = max(d_n − s, 0)
      cost[n, s] = c · max(f − m, 0) + h(s + f − d_n)
      make[n, s] = f
  // Compute cost[1..n − 1, 0..D] and make[1..n − 1, 0..D].
  U = d_n
  for k = n − 1 downto 1
      U = U + d_k
      for s = 0 to D
          cost[k, s] = ∞
          for f = max(d_k − s, 0) to U − s
              val = cost[k + 1, s + f − d_k]
                        + c · max(f − m, 0) + h(s + f − d_k)
              if val < cost[k, s]
                  cost[k, s] = val
                  make[k, s] = f
  print cost[1,0]
  PRINT-PLAN(make, n, d)

PRINT-PLAN(make, n, d)
  s = 0
  for k = 1 to n
      print "For month " k " manufacture " make[k, s] " machines"
      s = s + make[k, s] − d_k
```

In INVENTORY-PLANNING, we build the solution month by month, starting from month n, moving backward toward month 1. First, we solve the subproblem for the last month, for all surpluses. Then, for each month and for each surplus entering that month, we calculate the cheapest way to satisfy demand for that month based on the solved subproblems of the next month.

- f is the number of machines that we try to manufacture in month k.
- $cost[k, s]$ holds the cheapest way to satisfy demands of months $k, \ldots, n$, with a net surplus of s left over at the beginning of month k.
- $make[k, s]$ holds the number of machines to manufacture in month k and the surplus s of an optimal plan. We will use this table to reconstruct the optimal plan.

We first initialize the base cases, which are the cases for month n starting with surplus s, for $s = 0, \ldots, D$. If $d_n > s$, it suffices to manufacture $d_n - s$ machines, since we need not keep any surplus after month n. If $d_n \le s$, we need not manufacture any machines at all.

We then calculate the total cost for month n as the sum of hiring extra labor $c \cdot \max(f - m, 0)$ and the inventory costs for leftover surplus $h(s + f - d_n)$, which can be nonzero if we had started out with a large surplus.

The outer **for** loop of the next block of code runs down from month $n-1$ to 1, thus ensuring that when we consider month k, we have already solved the subproblems of month $k+1$.

The next inner **for** loop iterates through all possible values of f as described.

For every choice of f for a given month k, the total cost of (k, s) is given by the cost of extra labor (if any) plus the cost of inventory (if there is a surplus) plus the cost of the subproblem $(k+1, s+f-d_k)$. This value is checked and updated.

Finally, the required answer is the answer to the subproblem $(1, 0)$, which appears in $cost[1, 0]$. That is, it is the cheapest way to satisfy all the demands of months $1, \ldots, n$ when we start with a surplus of 0.

The running time of INVENTORY-PLANNING is clearly $O(nD^2)$. The space requirement is $O(nD)$. We can improve upon the space requirement by noting that we need only store the solution to subproblems of the next month. With this observation, we can construct an algorithm that uses $O(n+D)$ space.

Solution to Problem 15-12

Let $p.cost$ denote the cost and $p.vorp$ denote the VORP of player p. We shall assume that all dollar amounts are expressed in units of \$100,000.

Since the order of choosing players for the positions does not matter, we may assume that we make our decisions starting from position 1, moving toward position N. For each position, we decide to either sign one player or sign no players. Suppose we decide to sign player p, who plays position 1. Then, we are left with an amount of $X - p.cost$ dollars to sign players at positions $2, \ldots, N$. This observation guides us in how to frame the subproblems.

We define the cost and VORP of a *set* of players as the sum of costs and the sum of VORPs of all players in that set. Let (i, x) denote the following subproblem: "Suppose we consider only positions $i, i+1, \ldots, N$ and we can spend at most x dollars. What set of players (with at most one player for each position under consideration) has the maximum VORP?" A *valid* set of players for (i, x) is one in which each player in the set plays one of the positions $i, i+1, \ldots, n$, each position has at most one player, and the cost of the players in the set is at most x dollars. An *optimal* set of players for (i, x) is a valid set with the maximum VORP. We now show that the problem exhibits optimal substructure.

Theorem (Optimal substructure of the VORP maximization problem)
Let $L = \{p_1, p_2, \ldots, p_k\}$ be a set of players, possibly empty, with maximum VORP for the subproblem (i, x).

1. If $i = N$, then L has at most one player. If all players in position N have cost more than x, then L has no players. Otherwise, $L = \{p_1\}$, where p_1 has the maximum VORP among players for position N with cost at most x.
2. If $i < N$ and L includes player p for position i, then $L' = L - \{p\}$ is an optimal set for the subproblem $(i+1, x - p.cost)$.
3. If $i < N$ and L does not include a player for position i, then L is an optimal set for the subproblem $(i+1, x)$.

Proof Property (1) follows trivially from the problem statement.

(2) Suppose that L' is not an optimal set for the subproblem $(i+1, x - p.cost)$. Then, there exists another valid set L'' for $(i+1, x - p.cost)$ that has VORP more than L'. Let $L''' = L'' \cup \{p\}$. The cost of L''' is at most x, since L'' has a cost at most $x - p.cost$. Moreover, L''' has at most one player for each position $i, i+1, \ldots, N$. Thus, L''' is a valid set for (i, x). But L''' has VORP more than L, thus contradicting the assumption that L had the maximum VORP for (i, x).

(3) Clearly, any valid set for $(i+1, x)$ is also a valid set for (i, x). If L were not an optimal set for $(i+1, x)$, then there exists another valid set L' for $(i+1, x)$ with VORP more than L. The set L' would also be a valid set for (i, x), which contradicts the assumption that L had the maximum VORP for (i, x). ■

The theorem suggests that when $i < N$, we examine two subproblems and choose the better of the two. Let $v[i, x]$ denote the maximum VORP for (i, x). Let $S(i, x)$ be the set of players who play position i and cost at most x. In the following recurrence for $v[i, x]$, we assume that the max function returns $-\infty$ when invoked over an empty set:

$$v[i, x] = \begin{cases} \max\limits_{p \in S(N,x)} \{p.vorp\} & \text{if } i = N\ , \\ \max\Big\{ v[i+1, x], & \\ \qquad \max\limits_{p \in S(i,x)} \{p.vorp + v[i+1, x - p.cost]\} \Big\} & \text{if } i < N\ . \end{cases}$$

This recurrence lends itself to implementation in a straightforward way. Let p_{ij} denote the jth player who plays position i.

```
FREE-AGENT-VORP(p, N, P, X)
  let v[1..N][0..X] and who[1..N][0..X] be new tables
  for x = 0 to X
      v[N, x] = −∞
      who[N, x] = 0
      for k = 1 to P
          if p_Nk.cost ≤ x and p_Nk.vorp > v[N, x]
              v[N, x] = p_Nk.vorp
              who[N, x] = k
  for i = N − 1 downto 1
      for x = 0 to X
          v[i, x] = v[i + 1, x]
          who[i, x] = 0
          for k = 1 to P
              if p_ik.cost ≤ x and v[i + 1, x − p_ik.cost] + p_ik.vorp > v[i, x]
                  v[i, x] = v[i + 1, x − p_ik.cost] + p_ik.vorp
                  who[i, x] = k
  print "The maximum value of VORP is " v[1, X]
  amt = X
  for i = 1 to N
      k = who[i, amt]
      if k ≠ 0
          print "sign player " p_ik
          amt = amt − p_ik.cost
  print "The total money spent is " X − amt
```

The input to FREE-AGENT-VORP is the list of players p and N, P, and X, as given in the problem. The table $v[i, x]$ holds the maximum VORP for the subproblem (i, x). The table $who[i, x]$ holds information necessary to reconstruct the actual solution. Specifically, $who[i, x]$ holds the index of player to sign for position i, or 0 if no player should be signed for position i. The first set of nested **for** loops initializes the base cases, in which $i = N$. For every amount x, the inner loop simply picks the player with the highest VORP who plays position N and whose cost is at most x.

The next set of three nested **for** loops represents the main computation. The outermost **for** loop runs down from position $N - 1$ to 1. This order ensures that smaller subproblems are solved before larger ones. We initialize $v[i, x]$ as $v[i + 1, x]$. This way, we already take care of the case in which we decide not to sign any player who plays position i. The innermost **for** loop tries to sign each player (if we have enough money) in turn, and it keeps track of the maximum VORP possible.

The maximum VORP for the entire problem ends up in $v[1, X]$. The final **for** loop uses the information in *who* table to print out which players to sign. The running time of FREE-AGENT-VORP is clearly $\Theta(NPX)$, and it uses $\Theta(NX)$ space.

Lecture Notes for Chapter 16: Greedy Algorithms

Chapter 16 Introduction

Similar to dynamic programming.

Used for optimization problems.

Idea

When we have a choice to make, make the one that looks best *right now*. Make a *locally optimal choice* in hope of getting a *globally optimal solution.*

Greedy algorithms don't always yield an optimal solution. But sometimes they do. We'll see a problem for which they do. Then we'll look at some general characteristics of when greedy algorithms give optimal solutions.

[We do not cover Huffman codes or matroids in these notes.]

Activity selection

n ***activities*** require *exclusive* use of a common resource. For example, scheduling the use of a classroom.

Set of activities $S = \{a_1, \ldots, a_n\}$.

a_i needs resource during period $[s_i, f_i)$, which is a half-open interval, where $s_i =$ start time and $f_i =$ finish time.

Goal

Select the largest possible set of nonoverlapping (***mutually compatible***) activities.

Note

Could have many other objectives:

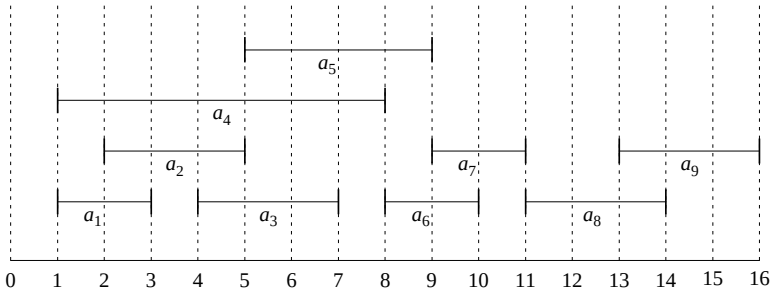
- Schedule room for longest time.
- Maximize income rental fees.

Assume that activities are sorted by finish time: $f_1 \le f_2 \le f_3 \le \cdots \le f_{n-1} \le f_n$.

Example

S sorted by finish time: *[Leave on board]*

i	1	2	3	4	5	6	7	8	9
s_i	1	2	4	1	5	8	9	11	13
f_i	3	5	7	8	9	10	11	14	16



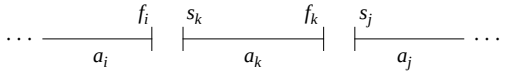
Maximum-size mutually compatible set: $\{a_1, a_3, a_6, a_8\}$.

Not unique: also $\{a_2, a_5, a_7, a_9\}$.

Optimal substructure of activity selection

$S_{ij} = \{a_k \in S : f_i \le s_k < f_k \le s_j\}$ *[Leave on board]*

$=$ activities that start after a_i finishes and finish before a_j starts .



Activities in S_{ij} are compatible with

- all activities that finish by f_i, and
- all activities that start no earlier than s_j.

Let A_{ij} be a maximum-size set of mutually compatible activities in S_{ij}.

Let $a_k \in A_{ij}$ be some activity in A_{ij}. Then we have two subproblems:

- Find mutually compatible activities in S_{ik} (activities that start after a_i finishes and that finish before a_k starts).
- Find mutually compatible activities in S_{kj} (activities that start after a_k finishes and that finish before a_j starts).

Let

$A_{ik} = A_{ij} \cap S_{ik}$ = activities in A_{ij} that finish before a_k starts ,

$A_{kj} = A_{ij} \cap S_{kj}$ = activities in A_{ij} that start afer a_k finishes .

Then $A_{ij} = A_{ik} \cup \{a_k\} \cup A_{kj}$

$\Rightarrow |A_{ij}| = |A_{ik}| + |A_{kj}| + 1$.

Claim

Optimal solution A_{ij} must include optimal solutions for the two subproblems for S_{ik} and S_{kj}.

Proof Use the usual cut-and-paste argument. Will show the claim for S_{kj}; proof for S_{ik} is symmetric.

Suppose we could find a set A'_{kj} of mutually compatible activities in S_{kj}, where $|A'_{kj}| > |A_{kj}|$. Then use A'_{kj} instead of A_{kj} when solving the subproblem for S_{ij}. Size of resulting set of mutually compatible activities would be $|A_{ik}| + |A'_{kj}| + 1 > |A_{ik}| + |A_{kj}| + 1 = |A|$. Contradicts assumption that A_{ij} is optimal. ■ (claim)

One recursive solution

Since optimal solution A_{ij} must include optimal solutions to the subproblems for S_{ik} and S_{kj}, could solve by dynamic programming.

Let $c[i, j] =$ size of optimal solution for S_{ij}. Then

$c[i, j] = c[i, k] + c[k, j] + 1$.

But we don't know which activity a_k to choose, so we have to try them all:

$$c[i, j] = \begin{cases} 0 & \text{if } S_{ij} = \emptyset \text{ ,} \\ \max\limits_{a_k \in S_{ij}} \{c[i, k] + c[k, j] + 1\} & \text{if } S_{ij} \neq \emptyset \text{ .} \end{cases}$$

Could then develop a recursive algorithm and memoize it. Or could develop a bottom-up algorithm and fill in table entries.

Instead, we will look at a greedy approach.

Making the greedy choice

Choose an activity to add to optimal solution *before* solving subproblems. For activity-selection problem, we can get away with considering only the greedy choice: the activity that leaves the resource available for as many other activities as possible.

Question: Which activity leaves the resource available for the most other activities? Answer: The first activity to finish. (If more than one activity has earliest finish time, can choose any such activity.)

Since activities are sorted by finish time, just choose activity a_1.

That leaves only one subproblem to solve: finding a maximum size set of mutually compatible activities that start after a_1 finishes. (Don't have to worry about activities that finish before a_1 starts, because $s_1 < f_1$ and no activity a_i has finish time $f_i < f_1 \Rightarrow$ no activity a_i has $f_i \leq s_1$.)

Since have only subproblem to solve, simplify notation:

$S_k = \{a_i \in S : s_i \geq f_k\} =$ activities that start after a_k finishes .

Making greedy choice of $a_1 \Rightarrow S_1$ remains as only subproblem to solve. *[Slight abuse of notation: referring to S_k not only as a set of activities but as a subproblem consisting of these activities.]*

By optimal substructure, if a_1 is in an optimal solution, then an optimal solution to the original problem consists of a_1 plus all activities in an optimal solution to S_1.

But need to prove that a_1 is always part of some optimal solution.

Theorem

If S_k is nonempty and a_m has the earliest finish time in S_k, then a_m is included in some optimal solution.

Proof Let A_k be an optimal solution to S_k, and let a_j have the earliest finish time of any activity in A_k. If $a_j = a_m$, done. Otherwise, let $A'_k = A_k - \{a_j\} \cup \{a_m\}$ be A_k but with a_m substituted for a_j.

Claim

Activities in A'_k are disjoint.

Proof Activities in A_k are disjoint, a_j is first activity in A_k to finish, and $f_m \leq f_j$. ■ (claim)

Since $|A'_k| = |A_k|$, conclude that A'_k is an optimal solution to S_k, and it includes a_m. ■ (theorem)

So, don't need full power of dynamic programming. Don't need to work bottom-up.

Instead, can just repeatedly choose the activity that finishes first, keep only the activities that are compatible with that one, and repeat until no activities remain.

Can work top-down: make a choice, then solve a subproblem. Don't have to solve subproblems before making a choice.

Recursive greedy algorithm

Start and finish times are represented by arrays s and f, where f is assumed to be already sorted in monotonically increasing order.

To start, add fictitious activity a_0 with $f_0 = 0$, so that $S_0 = S$, the entire set of activities.

Procedure REC-ACTIVITY-SELECTOR takes as parameters the arrays s and f, index k of current subproblem, and number n of activities in the original problem.

```
REC-ACTIVITY-SELECTOR(s, f, k, n)
  m = k + 1
  while m ≤ n and s[m] < f[k]          // find the first activity in S_k to finish
      m = m + 1
  if m ≤ n
      return {a_m} ∪ REC-ACTIVITY-SELECTOR(s, f, m, n)
  else return ∅
```

Initial call

REC-ACTIVITY-SELECTOR$(s, f, 0, n)$.

Idea

The **while** loop checks $a_{k+1}, a_{k+2}, \ldots, a_n$ until it finds an activity a_m that is compatible with a_k (need $s_m \geq f_k$).

- If the loop terminates because a_m is found ($m \leq n$), then recursively solve S_m, and return this solution, along with a_m.
- If the loop never finds a compatible a_m ($m > n$), then just return empty set.

Go through example given earlier. Should get $\{a_1, a_3, a_6, a_8\}$.

Time

$\Theta(n)$—each activity examined exactly once, assuming that activities are already sorted by finish times.

Iterative greedy algorithm

Can convert the recursive algorithm to an iterative one. It's already almost tail recursive.

GREEDY-ACTIVITY-SELECTOR(s, f)

```
n = s.length
A = {a_1}
k = 1
for m = 2 to n
    if s[m] ≥ f[k]
        A = A ∪ {a_m}
        k = m
return A
```

Go through example given earlier. Should again get $\{a_1, a_3, a_6, a_8\}$.

Time

$\Theta(n)$, if activities are already sorted by finish times.

For both the recursive and iterative algorithms, add $O(n \lg n)$ time if activities need to be sorted.

Greedy strategy

The choice that seems best at the moment is the one we go with.

What did we do for activity selection?

1. Determine the optimal substructure.
2. Develop a recursive solution.
3. Show that if we make the greedy choice, only one subproblem remains.

4. Prove that it's always safe to make the greedy choice.
5. Develop a recursive greedy algorithm.
6. Convert it to an iterative algorithm.

At first, it looked like dynamic programming. In the activity-selection problem, we started out by defining subproblems S_{ij}, where both i and j varied. But then found that making the greedy choice allowed us to restrict the subproblems to be of the form S_k.

Could instead have gone straight for the greedy approach: in our first crack at defining subproblems, use the S_k form. Could then have proven that the greedy choice a_m (the first activity to finish), combined with optimal solution to the remaining compatible activities S_m, gives an optimal solution to S_k.

Typically, we streamline these steps:

1. Cast the optimization problem as one in which we make a choice and are left with one subproblem to solve.
2. Prove that there's always an optimal solution that makes the greedy choice, so that the greedy choice is always safe.
3. Demonstrate optimal substructure by showing that, having made the greedy choice, combining an optimal solution to the remaining subproblem with the greedy choice gives an optimal solution to the original problem.

No general way to tell whether a greedy algorithm is optimal, but two key ingredients are

1. greedy-choice property and
2. optimal substructure.

Greedy-choice property

Can assemble a globally optimal solution by making locally optimal (greedy) choices.

Dynamic programming

- Make a choice at each step.
- Choice depends on knowing optimal solutions to subproblems. Solve subproblems *first*.
- Solve *bottom-up*.

Greedy

- Make a choice at each step.
- Make the choice *before* solving the subproblems.
- Solve *top-down*.

Typically show the greedy-choice property by what we did for activity selection:

- Look at an optimal solution.

- If it includes the greedy choice, done.
- Otherwise, modify the optimal solution to include the greedy choice, yielding another solution that's just as good.

Can get efficiency gains from greedy-choice property.

- Preprocess input to put it into greedy order.
- Or, if dynamic data, use a priority queue.

Optimal substructure

Just show that optimal solution to subproblem and greedy choice $\Rightarrow$ optimal solution to problem.

Greedy vs. dynamic programming

The knapsack problem is a good example of the difference.

0-1 knapsack problem

- n items.
- Item i is worth $\$v_i$, weighs w_i pounds.
- Find a most valuable subset of items with total weight $\leq W$.
- Have to either take an item or not take it—can't take part of it.

Fractional knapsack problem

Like the 0-1 knapsack problem, but can take fraction of an item.

Both have optimal substructure.

But the fractional knapsack problem has the greedy-choice property, and the 0-1 knapsack problem does not.

To solve the fractional problem, rank items by value/weight: v_i/w_i. Let $v_i/w_i \geq v_{i+1}/w_{i+1}$ for all i. Take items in decreasing order of value/weight. Will take all of the items with the greatest value/weight, and possibly a fraction of the next item.

```
FRACTIONAL-KNAPSACK(v, w, W)
  load = 0
  i = 1
  while load < W and i ≤ n
      if w_i ≤ W − load
          take all of item i
      else take (W − load)/w_i of item i
      add what was taken to load
      i = i + 1
```

Time: $O(n \lg n)$ to sort, $O(n)$ thereafter.

Greedy doesn't work for the 0-1 knapsack problem. Might get empty space, which lowers the average value per pound of the items taken.

i	1	2	3
v_i	60	100	120
w_i	10	20	30
v_i/w_i	6	5	4

$W = 50$.

Greedy solution:

- Take items 1 and 2.
- value $= 160$, weight $= 30$.

Have 20 pounds of capacity left over.

Optimal solution:

- Take items 2 and 3.
- value $= 220$, weight $= 50$.

No leftover capacity.

Solutions for Chapter 16: Greedy Algorithms

Solution to Exercise 16.1-1

The tricky part is determining which activities are in the set S_{ij}. If activity k is in S_{ij}, then we must have $i < k < j$, which means that $j - i \geq 2$, but we must also have that $f_i \leq s_k$ and $f_k \leq s_j$. If we start k at $j - 1$ and decrement k, we can stop once k reaches i, but we can also stop once we find that $f_k \leq f_i$, since then activities $i + 1$ through k cannot be compatible with activity i.

We create two fictitious activities, a_0 with $f_0 = 0$ and a_{n+1} with $s_{n+1} = \infty$. We are interested in a maximum-size set $A_{0,n+1}$ of mutually compatible activities in $S_{0,n+1}$. We'll use tables $c[0\,..\,n+1, 0\,..\,n+1]$, as in recurrence (16.2) (so that $c[i, j] = |A_{ij}|$), and $act[0\,..\,n+1, 0\,..\,n+1]$, where $act[i, j]$ is the activity k that we choose to put into A_{ij}.

We fill the tables in according to increasing difference $j - i$, which we denote by l in the pseudocode. Since $S_{ij} = \emptyset$ if $j - i < 2$, we initialize $c[i, i] = 0$ for all i and $c[i, i+1] = 0$ for $0 \leq i \leq n$. As in RECURSIVE-ACTIVITY-SELECTOR and GREEDY-ACTIVITY-SELECTOR, the start and finish times are given as arrays s and f, where we assume that the arrays already include the two fictitious activities and that the activities are sorted by monotonically increasing finish time.

```
DYNAMIC-ACTIVITY-SELECTOR(s, f, n)
  let c[0..n+1, 0..n+1] and act[0..n+1, 0..n+1] be new tables
  for i = 0 to n
      c[i, i] = 0
      c[i, i+1] = 0
  c[n+1, n+1] = 0
  for l = 2 to n+1
      for i = 0 to n-l+1
          j = i + l
          c[i, j] = 0
          k = j - 1
          while f[i] < f[k]
              if f[i] ≤ s[k] and f[k] ≤ s[j] and c[i, k] + c[k, j] + 1 > c[i, j]
                  c[i, j] = c[i, k] + c[k, j] + 1
                  act[i, j] = k
              k = k - 1
  print "A maximum size set of mutually compatible activities has size " c[0, n+1]
  print "The set contains "
  PRINT-ACTIVITIES(c, act, 0, n+1)

PRINT-ACTIVITIES(c, act, i, j)
  if c[i, j] > 0
      k = act[i, j]
      print k
      PRINT-ACTIVITIES(c, act, i, k)
      PRINT-ACTIVITIES(c, act, k, j)
```

The PRINT-ACTIVITIES procedure recursively prints the set of activities placed into the optimal solution A_{ij}. It first prints the activity k that achieved the maximum value of $c[i, j]$, and then it recurses to print the activities in A_{ik} and A_{kj}. The recursion bottoms out when $c[i, j] = 0$, so that $A_{ij} = \emptyset$.

Whereas GREEDY-ACTIVITY-SELECTOR runs in $\Theta(n)$ time, the DYNAMIC-ACTIVITY-SELECTOR procedure runs in $O(n^3)$ time.

Solution to Exercise 16.1-2

The proposed approach—selecting the last activity to start that is compatible with all previously selected activities—is really the greedy algorithm but starting from the end rather than the beginning.

Another way to look at it is as follows. We are given a set $S = \{a_1, a_2, \ldots, a_n\}$ of activities, where $a_i = [s_i, f_i)$, and we propose to find an optimal solution by selecting the last activity to start that is compatible with all previously selected activities. Instead, let us create a set $S' = \{a'_1, a'_2, \ldots, a'_n\}$, where $a'_i = [f_i, s_i)$. That is, a'_i is a_i in reverse. Clearly, a subset of $\{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\} \subseteq S$ is mutually compatible if and only if the corresponding subset $\{a'_{i_1}, a'_{i_2}, \ldots, a'_{i_k}\} \subseteq S'$ is also

mutually compatible. Thus, an optimal solution for S maps directly to an optimal solution for S' and vice versa.

The proposed approach of selecting the last activity to start that is compatible with all previously selected activities, when run on S, gives the same answer as the greedy algorithm from the text—selecting the first activity to finish that is compatible with all previously selected activities—when run on S'. The solution that the proposed approach finds for S corresponds to the solution that the text's greedy algorithm finds for S', and so it is optimal.

Solution to Exercise 16.1-3

- For the approach of selecting the activity of least duration from those that are compatible with previously selected activities:

i	1	2	3
s_i	0	2	3
f_i	3	4	6
duration	3	2	3

  This approach selects just $\{a_2\}$, but the optimal solution selects $\{a_1, a_3\}$.

- For the approach of always selecting the compatible activity that overlaps the fewest other remaining activities:

i	1	2	3	4	5	6	7	8	9	10	11
s_i	0	1	1	1	2	3	4	5	5	5	6
f_i	2	3	3	3	4	5	6	7	7	7	8
# of overlapping activities	3	4	4	4	4	2	4	4	4	4	3

  This approach first selects a_6, and after that choice it can select only two other activities (one of a_1, a_2, a_3, a_4 and one of a_8, a_9, a_{10}, a_{11}). An optimal solution is $\{a_1, a_5, a_7, a_{11}\}$.

- For the approach of always selecting the compatible remaining activity with the earliest start time, just add one more activity with the interval $[0, 14)$ to the example in Section 16.1. It will be the first activity selected, and no other activities are compatible with it.

Solution to Exercise 16.1-4

This solution is also posted publicly

Let S be the set of n activities.

The "obvious" solution of using GREEDY-ACTIVITY-SELECTOR to find a maximum-size set S_1 of compatible activities from S for the first lecture hall, then using it again to find a maximum-size set S_2 of compatible activities from $S - S_1$ for the second hall, (and so on until all the activities are assigned), requires $\Theta(n^2)$ time in the worst case. Moreover, it can produce a result that uses more lecture halls

than necessary. Consider activities with the intervals $\{[1,4),[2,5),[6,7),[4,8)\}$. GREEDY-ACTIVITY-SELECTOR would choose the activities with intervals $[1,4)$ and $[6,7)$ for the first lecture hall, and then each of the activities with intervals $[2,5)$ and $[4,8)$ would have to go into its own hall, for a total of three halls used. An optimal solution would put the activities with intervals $[1,4)$ and $[4,8)$ into one hall and the activities with intervals $[2,5)$ and $[6,7)$ into another hall, for only two halls used.

There is a correct algorithm, however, whose asymptotic time is just the time needed to sort the activities by time—$O(n \lg n)$ time for arbitrary times, or possibly as fast as $O(n)$ if the times are small integers.

The general idea is to go through the activities in order of start time, assigning each to any hall that is available at that time. To do this, move through the set of events consisting of activities starting and activities finishing, in order of event time. Maintain two lists of lecture halls: Halls that are busy at the current event-time t (because they have been assigned an activity i that started at $s_i \le t$ but won't finish until $f_i > t$) and halls that are free at time t. (As in the activity-selection problem in Section 16.1, we are assuming that activity time intervals are half open—i.e., that if $s_i \ge f_j$, then activities i and j are compatible.) When t is the start time of some activity, assign that activity to a free hall and move the hall from the free list to the busy list. When t is the finish time of some activity, move the activity's hall from the busy list to the free list. (The activity is certainly in some hall, because the event times are processed in order and the activity must have started before its finish time t, hence must have been assigned to a hall.)

To avoid using more halls than necessary, always pick a hall that has already had an activity assigned to it, if possible, before picking a never-used hall. (This can be done by always working at the front of the free-halls list—putting freed halls onto the front of the list and taking halls from the front of the list—so that a new hall doesn't come to the front and get chosen if there are previously-used halls.)

This guarantees that the algorithm uses as few lecture halls as possible: The algorithm will terminate with a schedule requiring $m \le n$ lecture halls. Let activity i be the first activity scheduled in lecture hall m. The reason that i was put in the mth lecture hall is that the first $m-1$ lecture halls were busy at time s_i. So at this time there are m activities occurring simultaneously. Therefore any schedule must use at least m lecture halls, so the schedule returned by the algorithm is optimal.

Run time:

- Sort the $2n$ activity-starts/activity-ends events. (In the sorted order, an activity-ending event should precede an activity-starting event that is at the same time.) $O(n \lg n)$ time for arbitrary times, possibly $O(n)$ if the times are restricted (e.g., to small integers).
- Process the events in $O(n)$ time: Scan the $2n$ events, doing $O(1)$ work for each (moving a hall from one list to the other and possibly associating an activity with it).

Total: $O(n + \text{time to sort})$

[The idea of this algorithm is related to the rectangle-overlap algorithm in Exercise 14.3-7.]

Solution to Exercise 16.1-5

We can no longer use the greedy algorithm to solve this problem. However, as we show, the problem still has an optimal substructure which allows us to formulate a dynamic programming solution. The analysis here follows closely the analysis of Section 16.1 in the book. We define the value of a set of compatible events as the sum of values of events in that set. Let S_{ij} be defined as in Section 16.1. An *optimal solution* to S_{ij} is a subset of mutually compatible events of S_{ij} that has maximum value. Let A_{ij} be an optimal solution to S_{ij}. Suppose A_{ij} includes an event a_k. Let A_{ik} and A_{kj} be defined as in Section 16.1. Thus, we have $A_{ij} = A_{ik} \cup \{a_k\} \cup A_{kj}$, and so the value of maximum-value set A_{ij} is equal to the value of A_{ik} plus the value of A_{kj} plus v_k.

The usual cut-and-paste argument shows that the optimal solution A_{ij} must also include optimal solutions to the two subproblems for S_{ik} and S_{kj}. If we could find a set A'_{kj} of mutually compatible activities in S_{kj} where the value of A'_{kj} is greater than the value of A_{kj}, then we could use A'_{kj}, rather than A_{kj}, in a solution to the subproblem for S_{ij}. We would have constructed a set of mutually compatible activities with greater value than that of A_{ij}, which contradicts the assumption that A_{ij} is an optimal solution. A symmetric argument applies to the activities in S_{ik}.

Let us denote the value of an optimal solution for the set S_{ij} by $val[i, j]$. Then, we would have the recurrence

$$val[i, j] = val[i, k] + val[k, j] + v_k \ .$$

Of course, since we do not know that an optimal solution for the set S_{ij} includes activity a_k, we would have to examine all activities in S_{ij} to find which one to choose, so that

$$val[i, j] = \begin{cases} 0 & \text{if } S_{ij} = \emptyset \ , \\ \max\limits_{a_k \in S_{ij}} \{val[i, k] + val[k, j] + v_k\} & \text{if } S_{ij} \neq \emptyset \ . \end{cases}$$

While implementing the recurrence, the tricky part is determining which activities are in the set S_{ij}. If activity k is in S_{ij}, then we must have $i < k < j$, which means that $j - i \geq 2$, but we must also have that $f_i \leq s_k$ and $f_k \leq s_j$. If we start k at $j - 1$ and decrement k, we can stop once k reaches i, but we can also stop once we find that $f_k \leq f_i$, since then activities $i + 1$ through k cannot be compatible with activity i.

We create two fictitious activities, a_0 with $f_0 = 0$ and a_{n+1} with $s_{n+1} = \infty$. We are interested in a maximum-size set $A_{0,n+1}$ of mutually compatible activities in $S_{0,n+1}$. We'll use tables $val[0..n+1, 0..n+1]$, as in the recurrence, and $act[0..n+1, 0..n+1]$, where $act[i, j]$ is the activity k that we choose to put into A_{ij}.

We fill the tables in according to increasing difference $j - i$, which we denote by l in the pseudocode. Since $S_{ij} = \emptyset$ if $j - i < 2$, we initialize $val[i, i] = 0$ for all i and $val[i, i + 1] = 0$ for $0 \leq i \leq n$. As in RECURSIVE-ACTIVITY-SELECTOR and GREEDY-ACTIVITY-SELECTOR, the start and finish times are given as arrays s and f, where we assume that the arrays already include the two fictitious activities

and that the activities are sorted by monotonically increasing finish time. The array v specifies the value of each activity.

```
MAX-VALUE-ACTIVITY-SELECTOR(s, f, v, n)
  let val[0..n+1, 0..n+1] and act[0..n+1, 0..n+1] be new tables
  for i = 0 to n
      val[i, i] = 0
      val[i, i+1] = 0
  val[n+1, n+1] = 0
  for l = 2 to n+1
      for i = 0 to n-l+1
          j = i + l
          val[i, j] = 0
          k = j - 1
          while f[i] < f[k]
              if f[i] ≤ s[k] and f[k] ≤ s[j] and
                        val[i, k] + val[k, j] + v_k > val[i, j]
                  val[i, j] = val[i, k] + val[k, j] + v_k
                  act[i, j] = k
              k = k - 1
  print "A maximum-value set of mutually compatible activities has value "
          val[0, n+1]
  print "The set contains "
  PRINT-ACTIVITIES(val, act, 0, n+1)

PRINT-ACTIVITIES(val, act, i, j)
  if val[i, j] > 0
      k = act[i, j]
      print k
      PRINT-ACTIVITIES(val, act, i, k)
      PRINT-ACTIVITIES(val, act, k, j)
```

The PRINT-ACTIVITIES procedure recursively prints the set of activities placed into the optimal solution A_{ij}. It first prints the activity k that achieved the maximum value of $val[i, j]$, and then it recurses to print the activities in A_{ik} and A_{kj}. The recursion bottoms out when $val[i, j] = 0$, so that $A_{ij} = \emptyset$.

Whereas GREEDY-ACTIVITY-SELECTOR runs in $\Theta(n)$ time, the MAX-VALUE-ACTIVITY-SELECTOR procedure runs in $O(n^3)$ time.

Solution to Exercise 16.2-2

This solution is also posted publicly

The solution is based on the optimal-substructure observation in the text: Let i be the highest-numbered item in an optimal solution S for W pounds and items $1, \ldots, n$. Then $S' = S - \{i\}$ must be an optimal solution for $W - w_i$ pounds and items $1, \ldots, i-1$, and the value of the solution S is v_i plus the value of the subproblem solution S'.

We can express this relationship in the following formula: Define $c[i, w]$ to be the value of the solution for items $1, \ldots, i$ and maximum weight w. Then

$$c[i,w] = \begin{cases} 0 & \text{if } i = 0 \text{ or } w = 0 \ , \\ c[i-1,w] & \text{if } w_i > w \ , \\ \max(v_i + c[i-1, w-w_i], c[i-1,w]) & \text{if } i > 0 \text{ and } w \geq w_i \ . \end{cases}$$

The last case says that the value of a solution for i items either includes item i, in which case it is v_i plus a subproblem solution for $i-1$ items and the weight excluding w_i, or doesn't include item i, in which case it is a subproblem solution for $i-1$ items and the same weight. That is, if the thief picks item i, he takes v_i value, and he can choose from items $1, \ldots, i-1$ up to the weight limit $w - w_i$, and get $c[i-1, w-w_i]$ additional value. On the other hand, if he decides not to take item i, he can choose from items $1, \ldots, i-1$ up to the weight limit w, and get $c[i-1, w]$ value. The better of these two choices should be made.

The algorithm takes as inputs the maximum weight W, the number of items n, and the two sequences $v = \langle v_1, v_2, \ldots, v_n \rangle$ and $w = \langle w_1, w_2, \ldots, w_n \rangle$. It stores the $c[i, j]$ values in a table $c[0..n, 0..W]$ whose entries are computed in row-major order. (That is, the first row of c is filled in from left to right, then the second row, and so on.) At the end of the computation, $c[n, W]$ contains the maximum value the thief can take.

```
DYNAMIC-0-1-KNAPSACK(v, w, n, W)
  let c[0..n, 0..W] be a new array
  for w = 0 to W
      c[0, w] = 0
  for i = 1 to n
      c[i, 0] = 0
      for w = 1 to W
          if w_i ≤ w
              if v_i + c[i − 1, w − w_i] > c[i − 1, w]
                  c[i, w] = v_i + c[i − 1, w − w_i]
              else c[i, w] = c[i − 1, w]
          else c[i, w] = c[i − 1, w]
```

We can use the c table to deduce the set of items to take by starting at $c[n, W]$ and tracing where the optimal values came from. If $c[i, w] = c[i-1, w]$, then item i is not part of the solution, and we continue tracing with $c[i-1, w]$. Otherwise item i is part of the solution, and we continue tracing with $c[i-1, w-w_i]$.

The above algorithm takes $\Theta(nW)$ time total:

- $\Theta(nW)$ to fill in the c table: $(n+1) \cdot (W+1)$ entries, each requiring $\Theta(1)$ time to compute.
- $O(n)$ time to trace the solution (since it starts in row n of the table and moves up one row at each step).

Solution to Exercise 16.2-4

The optimal strategy is the obvious greedy one. Starting with both bottles full, Professor Gekko should go to the westernmost place that he can refill his bottles within m miles of Grand Forks. Fill up there. Then go to the westernmost refilling location he can get to within m miles of where he filled up, fill up there, and so on.

Looked at another way, at each refilling location, Professor Gekko should check whether he can make it to the next refilling location without stopping at this one. If he can, skip this one. If he cannot, then fill up. Professor Gekko doesn't need to know how much water he has or how far the next refilling location is to implement this approach, since at each fillup, he can determine which is the next location at which he'll need to stop.

This problem has optimal substructure. Suppose there are n possible refilling locations. Consider an optimal solution with s refilling locations and whose first stop is at the kth location. Then the rest of the optimal solution must be an optimal solution to the subproblem of the remaining $n - k$ stations. Otherwise, if there were a better solution to the subproblem, i.e., one with fewer than $s - 1$ stops, we could use it to come up with a solution with fewer than s stops for the full problem, contradicting our supposition of optimality.

This problem also has the greedy-choice property. Suppose there are k refilling locations beyond the start that are within m miles of the start. The greedy solution chooses the kth location as its first stop. No station beyond the kth works as a first stop, since Professor Gekko would run out of water first. If a solution chooses a location $j < k$ as its first stop, then Professor Gekko could choose the kth location instead, having at least as much water when he leaves the kth location as if he'd chosen the jth location. Therefore, he would get at least as far without filling up again if he had chosen the kth location.

If there are n refilling locations on the map, Professor Gekko needs to inspect each one just once. The running time is $O(n)$.

Solution to Exercise 16.2-6

Use a linear-time median algorithm to calculate the median m of the v_i/w_i ratios. Next, partition the items into three sets: $G = \{i : v_i/w_i > m\}$, $E = \{i : v_i/w_i = m\}$, and $L = \{i : v_i/w_i < m\}$; this step takes linear time. Compute $W_G = \sum_{i \in G} w_i$ and $W_E = \sum_{i \in E} w_i$, the total weight of the items in sets G and E, respectively.

- If $W_G > W$, then do not yet take any items in set G, and instead recurse on the set of items G and knapsack capacity W.
- Otherwise ($W_G \le W$), take all items in set G, and take as much of the items in set E as will fit in the remaining capacity $W - W_G$.
- If $W_G + W_E \ge W$ (i.e., there is no capacity left after taking all the items in set G and all the items in set E that fit in the remaining capacity $W - W_G$), then we are done.

- Otherwise ($W_G + W_E < W$), then after taking all the items in sets G and E, recurse on the set of items L and knapsack capacity $W - W_G - W_E$.

To analyze this algorithm, note that each recursive call takes linear time, exclusive of the time for a recursive call that it may make. When there is a recursive call, there is just one, and it's for a problem of at most half the size. Thus, the running time is given by the recurrence $T(n) \le T(n/2) + \Theta(n)$, whose solution is $T(n) = O(n)$.

Solution to Exercise 16.2-7

This solution is also posted publicly

Sort A and B into monotonically decreasing order.

Here's a proof that this method yields an optimal solution. Consider any indices i and j such that $i < j$, and consider the terms ${a_i}^{b_i}$ and ${a_j}^{b_j}$. We want to show that it is no worse to include these terms in the payoff than to include ${a_i}^{b_j}$ and ${a_j}^{b_i}$, i.e., that ${a_i}^{b_i}{a_j}^{b_j} \ge {a_i}^{b_j}{a_j}^{b_i}$. Since A and B are sorted into monotonically decreasing order and $i < j$, we have $a_i \ge a_j$ and $b_i \ge b_j$. Since a_i and a_j are positive and $b_i - b_j$ is nonnegative, we have ${a_i}^{b_i - b_j} \ge {a_j}^{b_i - b_j}$. Multiplying both sides by ${a_i}^{b_j}{a_j}^{b_j}$ yields ${a_i}^{b_i}{a_j}^{b_j} \ge {a_i}^{b_j}{a_j}^{b_i}$.

Since the order of multiplication doesn't matter, sorting A and B into monotonically increasing order works as well.

Solution to Exercise 16.3-1

We are given that $x.freq \le y.freq$ are the two lowest frequencies in order, and that $a.freq \le b.freq$. Now,

$$\begin{aligned} & b.freq = x.freq \\ \Rightarrow\ & a.freq \le x.freq \\ \Rightarrow\ & a.freq = x.freq \quad \text{(since } x.freq \text{ is the lowest frequency)} , \end{aligned}$$

and since $y.freq \le b.freq$,

$$\begin{aligned} & b.freq = x.freq \\ \Rightarrow\ & y.freq \le x.freq \\ \Rightarrow\ & y.freq = x.freq \quad \text{(since } x.freq \text{ is the lowest frequency)} . \end{aligned}$$

Thus, if we assume that $x.freq = b.freq$, then we have that each of $a.freq$, $b.freq$, and $y.freq$ equals $x.freq$, and so $a.freq = b.freq = x.freq = y.freq$.

Solution to Exercise 16.4-2

We need to show three things to prove that $(S, \mathcal{I})$ is a matroid:

1. S is finite. That's because S is the set of of m columns of matrix T.

2. $\mathcal{I}$ is hereditary. That's because if $B \in \mathcal{I}$, then the columns in B are linearly independent. If $A \subseteq B$, then the columns of A must also be linearly independent, and so $A \in \mathcal{I}$.
3. $(S, \mathcal{I})$ satisfies the exchange property. To see why, let us suppose that $A, B \in \mathcal{I}$ and $|A| < |B|$.

   We will use the following properties of matrices:

   - The rank of a matrix is the number of columns in a maximal set of linearly independent columns (see page 1223 of the text). The rank is also equal to the dimension of the column space of the matrix.
   - If the column space of matrix B is a subspace of the column space of matrix A, then $\text{rank}(B) \leq \text{rank}(A)$.

   Because the columns in A are linearly independent, if we take just these columns as a matrix A, we have that $\text{rank}(A) = |A|$. Similarly, if we take the columns of B as a matrix B, we have $\text{rank}(B) = |B|$. Since $|A| < |B|$, we have $\text{rank}(A) < \text{rank}(B)$.

   We shall show that there is some column $b \in B$ that is not a linear combination of the columns in A, and so $A \cup \{b\}$ is linearly independent. The proof proceeds by contradiction. Assume that each column in B is a linear combination of the columns of A. That means that any vector that is a linear combination of the columns of B is also a linear combination of the columns of A, and so, treating the columns of A and B as matrices, the column space of B is a subspace of the column space of A. By the second property above, we have $\text{rank}(B) \leq \text{rank}(A)$. But we have already shown that $\text{rank}(A) < \text{rank}(B)$, a contradiction. Therefore, some column in B is not a linear combination of the columns of A, and $(S, \mathcal{I})$ satisfies the exchange property.

Solution to Exercise 16.4-3

[This exercise defines what is commonly known as the dual of a matroid, and it asks to prove that the dual of a matroid is itself a matroid. The literature contains simpler proofs of this fact, but they depend on other (equivalent) definitions of a matroid. The proof given here is more complicated, but it relies only on the definition given in the text.]

We need to show three things to prove that $(S, \mathcal{I}')$ is a matroid:

1. S is finite. We are given that.
2. $\mathcal{I}'$ is hereditary. Suppose that $B' \in \mathcal{I}'$ and $A' \subseteq B'$. Since $B' \in \mathcal{I}'$, there is some maximal set $B \in \mathcal{I}$ such that $B \subseteq S - B'$. But $A' \subseteq B'$ implies that $S - B' \subseteq S - A'$, and so $B \subseteq S - B' \subseteq S - A'$. Thus, there exists a maximal set $B \in \mathcal{I}$ such that $B \subseteq S - A'$, proving that $A' \in \mathcal{I}'$.
3. $(S, \mathcal{I}')$ satisfies the exchange property. We start with two preliminary facts about sets. The proofs of these facts are omitted.

   Fact 1: $|X - Y| = |X| - |X \cap Y|$.

Fact 2: Let S be the universe of elements. If $X - Y \subseteq Z$ and $Z \subseteq S - Y$, then $|X \cap Z| = |X| - |X \cap Y|$.

To show that $(S, \mathcal{I}')$ satisfies the exchange property, let us assume that $A' \in \mathcal{I}'$, $B' \in \mathcal{I}'$, and that $|A'| < |B'|$. We need to show that there exists some $x \in B' - A'$ such that $A' \cup \{x\} \in \mathcal{I}'$. Because $A' \in \mathcal{I}'$ and $B' \in \mathcal{I}'$, there are maximal sets $A \subseteq S - A'$ and $B \subseteq S - B'$ such that $A \in \mathcal{I}$ and $B \in \mathcal{I}$.

Define the set $X = B' - A' - A$, so that X consists of elements in B' but not in A' or A.

If X is nonempty, then let x be any element of X. By how we defined set X, we know that $x \in B'$ and $x \notin A'$, so that $x \in B' - A'$. Since $x \notin A$, we also have that $A \subseteq S - A' - \{x\} = S - (A' \cup \{x\})$, and so $A' \cup \{x\} \in \mathcal{I}'$.

If X is empty, the situation is more complicated. Because $|A'| < |B'|$, we have that $B' - A' \neq \emptyset$, and so X being empty means that $B' - A' \subseteq A$.

Claim

There is an element $y \in B - A'$ such that $(A - B') \cup \{y\} \in \mathcal{I}$.

Proof First, observe that because $A - B' \subseteq A$ and $A \in \mathcal{I}$, we have that $A - B' \in \mathcal{I}$. Similarly, $B - A' \subseteq B$ and $B \in \mathcal{I}$, and so $B - A' \in \mathcal{I}$. If we show that $|A - B'| < |B - A'|$, the assumption that $(S, \mathcal{I})$ is a matroid proves the existence of y.

Because $B' - A' \subseteq A$ and $A \subseteq S - A'$, we can apply Fact 2 to conclude that $|B' \cap A| = |B'| - |B' \cap A'|$. We claim that $|B \cap A'| \leq |A' - B'|$. To see why, observe that $A' - B' = A' \cap (S - B')$ and $B \subseteq S - B'$, and so $B \cap A' \subseteq (S - B') \cap A' = A' \cap (S - B') = A' - B'$. Applying Fact 1, we see that $|A' - B'| = |A'| - |A' \cap B'| = |A'| - |B' \cap A'|$, and hence $|B \cap A'| \leq |A'| - |B' \cap A'|$.

Now, we have

$$
\begin{array}{rcll}
|A'| & < & |B'| & \text{(by assumption)} \\
|A'| - |B' \cap A'| & < & |B'| - |B' \cap A'| & \text{(subtracting same quantity)} \\
|B \cap A'| & < & |B'| - |B' \cap A'| & (|B \cap A'| \leq |A'| - |B' \cap A'|) \\
|B \cap A'| & < & |B' \cap A| & (|B' \cap A| = |B'| - |B' \cap A'|) \\
|B| - |B \cap A'| & > & |A| - |B' \cap A| & (|A| = |B|) \\
|B - A'| & > & |A - B'| & \text{(Fact 1)}
\end{array}
$$

■ (claim)

Now we know there is an element $y \in B - A'$ such that $(A - B') \cup \{y\} \in \mathcal{I}$. Moreover, we claim that $y \notin A$. To see why, we know that by the exchange property, we can, without loss of generality, choose y so that $y \notin A - B'$. In order for y to be in A, it would have to be in $A \cap B'$. But $y \in B$, which means that $y \notin B'$, and hence $y \notin A \cap B'$. Therefore $y \notin A$.

Applying the exchange property, we add such an element y in $B - A'$ to $A - B'$, maintaining that the set we get, say C, is in $\mathcal{I}$. Then we keep applying the exchange property, adding a new element in $A - C$ to C, maintaining that C is in $\mathcal{I}$, until $|C| = |A|$. Once $|C| = |A|$, there must exist some element $x \in A$

that we have not added into C. We know that such an element exists because the element y that we first added into C was not in A, and so some element x in A must be left over. Also, we must have $x \in B'$ because all the elements in $A - B'$ are initially in C. Therefore, we have $x \in B' - A'$.

The set C so constructed is maximal, because it has the same cardinality as A, which is maximal, and $C \in \mathcal{I}$. All the elements but one in C are also in A; the one exception is in $B - A'$, and so C contains no elements in A'. Because we never added x to C, we have that $C \subseteq S - A' - \{x\} = S - (A' \cup \{x\})$. Therefore, $A' \cup \{x\} \in \mathcal{I}'$, as we needed to show.

Solution to Problem 16-1

Before we go into the various parts of this problem, let us first prove once and for all that the coin-changing problem has optimal substructure.

Suppose we have an optimal solution for a problem of making change for n cents, and we know that this optimal solution uses a coin whose value is c cents; let this optimal solution use k coins. We claim that this optimal solution for the problem of n cents must contain within it an optimal solution for the problem of $n - c$ cents. We use the usual cut-and-paste argument. Clearly, there are $k - 1$ coins in the solution to the $n - c$ cents problem used within our optimal solution to the n cents problem. If we had a solution to the $n - c$ cents problem that used fewer than $k - 1$ coins, then we could use this solution to produce a solution to the n cents problem that uses fewer than k coins, which contradicts the optimality of our solution.

a. A greedy algorithm to make change using quarters, dimes, nickels, and pennies works as follows:

- Give $q = \lfloor n/25 \rfloor$ quarters. That leaves $n_q = n \bmod 25$ cents to make change.
- Then give $d = \lfloor n_q/10 \rfloor$ dimes. That leaves $n_d = n_q \bmod 10$ cents to make change.
- Then give $k = \lfloor n_d/5 \rfloor$ nickels. That leaves $n_k = n_d \bmod 5$ cents to make change.
- Finally, give $p = n_k$ pennies.

An equivalent formulation is the following. The problem we wish to solve is making change for n cents. If $n = 0$, the optimal solution is to give no coins. If $n > 0$, determine the largest coin whose value is less than or equal to n. Let this coin have value c. Give one such coin, and then recursively solve the subproblem of making change for $n - c$ cents.

To prove that this algorithm yields an optimal solution, we first need to show that the greedy-choice property holds, that is, that some optimal solution to making change for n cents includes one coin of value c, where c is the largest coin value such that $c \leq n$. Consider some optimal solution. If this optimal solution includes a coin of value c, then we are done. Otherwise, this optimal solution does not include a coin of value c. We have four cases to consider:

- If $1 \le n < 5$, then $c = 1$. A solution may consist only of pennies, and so it must contain the greedy choice.
- If $5 \le n < 10$, then $c = 5$. By supposition, this optimal solution does not contain a nickel, and so it consists of only pennies. Replace five pennies by one nickel to give a solution with four fewer coins.
- If $10 \le n < 25$, then $c = 10$. By supposition, this optimal solution does not contain a dime, and so it contains only nickels and pennies. Some subset of the nickels and pennies in this solution adds up to 10 cents, and so we can replace these nickels and pennies by a dime to give a solution with (between 1 and 9) fewer coins.
- If $25 \le n$, then $c = 25$. By supposition, this optimal solution does not contain a quarter, and so it contains only dimes, nickels, and pennies. If it contains three dimes, we can replace these three dimes by a quarter and a nickel, giving a solution with one fewer coin. If it contains at most two dimes, then some subset of the dimes, nickels, and pennies adds up to 25 cents, and so we can replace these coins by one quarter to give a solution with fewer coins.

Thus, we have shown that there is always an optimal solution that includes the greedy choice, and that we can combine the greedy choice with an optimal solution to the remaining subproblem to produce an optimal solution to our original problem. Therefore, the greedy algorithm produces an optimal solution.

For the algorithm that chooses one coin at a time and then recurses on subproblems, the running time is $\Theta(k)$, where k is the number of coins used in an optimal solution. Since $k \le n$, the running time is $O(n)$. For our first description of the algorithm, we perform a constant number of calculations (since there are only 4 coin types), and the running time is $O(1)$.

b. When the coin denominations are $c^0, c^1, \ldots, c^k$, the greedy algorithm to make change for n cents works by finding the denomination c^j such that $j = \max\{0 \le i \le k : c^i \le n\}$, giving one coin of denomination c^j, and recursing on the subproblem of making change for $n - c^j$ cents. (An equivalent, but more efficient, algorithm is to give $\lfloor n/c^k \rfloor$ coins of denomination c^k and $\lfloor (n \bmod c^{i+1})/c^i \rfloor$ coins of denomination c^i for $i = 0, 1, \ldots, k-1$.)

To show that the greedy algorithm produces an optimal solution, we start by proving the following lemma:

Lemma

For $i = 0, 1, \ldots, k$, let a_i be the number of coins of denomination c^i used in an optimal solution to the problem of making change for n cents. Then for $i = 0, 1, \ldots, k-1$, we have $a_i < c$.

Proof If $a_i \ge c$ for some $0 \le i < k$, then we can improve the solution by using one more coin of denomination c^{i+1} and c fewer coins of denomination c^i. The amount for which we make change remains the same, but we use $c - 1 > 0$ fewer coins. ■ (lemma)

To show that the greedy solution is optimal, we show that any non-greedy solution is not optimal. As above, let $j = \max\{0 \le i \le k : c^i \le n\}$, so that the

greedy solution uses at least one coin of denomination c^j. Consider a non-greedy solution, which must use no coins of denomination c^j or higher. Let the non-greedy solution use a_i coins of denomination c^i, for $i = 0, 1, \ldots, j-1$; thus we have $\sum_{i=0}^{j-1} a_i c^i = n$. Since $n \geq c^j$, we have that $\sum_{i=0}^{j-1} a_i c^i \geq c^j$. Now suppose that the non-greedy solution is optimal. By the above lemma, $a_i \leq c-1$ for $i = 0, 1, \ldots, j-1$. Thus,

$$\begin{aligned}
\sum_{i=0}^{j-1} a_i c^i &\leq \sum_{i=0}^{j-1} (c-1)c^i \\
&= (c-1) \sum_{i=0}^{j-1} c^i \\
&= (c-1)\,\frac{c^j - 1}{c-1} \\
&= c^j - 1 \\
&< c^j \,,
\end{aligned}$$

which contradicts our earlier assertion that $\sum_{i=0}^{j-1} a_i c^i \geq c^j$. We conclude that the non-greedy solution is not optimal.

Since any algorithm that does not produce the greedy solution fails to be optimal, only the greedy algorithm produces the optimal solution.

The problem did not ask for the running time, but for the more efficient greedy-algorithm formulation, it is easy to see that the running time is $O(k)$, since we have to perform at most k each of the division, floor, and mod operations.

c. With actual U.S. coins, we can use coins of denomination 1, 10, and 25. When $n = 30$ cents, the greedy solution gives one quarter and five pennies, for a total of six coins. The non-greedy solution of three dimes is better.

The smallest integer numbers we can use are 1, 3, and 4. When $n = 6$ cents, the greedy solution gives one 4-cent coin and two 1-cent coins, for a total of three coins. The non-greedy solution of two 3-cent coins is better.

d. Since we have optimal substructure, dynamic programming might apply. And indeed it does.

Let us define $c[j]$ to be the minimum number of coins we need to make change for j cents. Let the coin denominations be $d_1, d_2, \ldots, d_k$. Since one of the coins is a penny, there is a way to make change for any amount $j \geq 1$.

Because of the optimal substructure, if we knew that an optimal solution for the problem of making change for j cents used a coin of denomination d_i, we would have $c[j] = 1 + c[j - d_i]$. As base cases, we have that $c[j] = 0$ for all $j \leq 0$.

To develop a recursive formulation, we have to check all denominations, giving

$$c[j] = \begin{cases} 0 & \text{if } j \leq 0\,, \\ 1 + \min\limits_{1 \leq i \leq k} \{c[j - d_i]\} & \text{if } j > 1\,. \end{cases}$$

We can compute the $c[j]$ values in order of increasing j by using a table. The following procedure does so, producing a table $c[1..n]$. It avoids even examining $c[j]$ for $j \leq 0$ by ensuring that $j \geq d_i$ before looking up $c[j - d_i]$. The

procedure also produces a table $denom[1..n]$, where $denom[j]$ is the denomination of a coin used in an optimal solution to the problem of making change for j cents.

```
COMPUTE-CHANGE(n, d, k)
  let c[1..n] and denom[1..n] be new arrays
  for j = 1 to n
      c[j] = ∞
      for i = 1 to k
          if j ≥ d_i and 1 + c[j − d_i] < c[j]
              c[j] = 1 + c[j − d_i]
              denom[j] = d_i
  return c and denom
```

This procedure obviously runs in $O(nk)$ time.

We use the following procedure to output the coins used in the optimal solution computed by COMPUTE-CHANGE:

```
GIVE-CHANGE(j, denom)
  if j > 0
      give one coin of denomination denom[j]
      GIVE-CHANGE(j − denom[j], denom)
```

The initial call is GIVE-CHANGE$(n, denom)$. Since the value of the first parameter decreases in each recursive call, this procedure runs in $O(n)$ time.

Solution to Problem 16-5

a. The procedure CACHE-MANAGER is a generic procedure, which initializes a cache by calling INITIALIZE-CACHE and then calls ACCESS with each data element in turn. The inputs are a sequence $R = \langle r_1, r_2, \ldots, r_n \rangle$ of memory requests and a cache size k.

```
CACHE-MANAGER(R, k)
  INITIALIZE-CACHE(R, k)
  for i = 1 to n
      ACCESS(r_i)
```

The running time of CACHE-MANAGER of course depends heavily on how ACCESS is implemented. We have several choices for how to implement the greedy strategy outlined in the problem. A straightforward way of implementing the greedy strategy is that when processing request r_i, for each of the at most k elements currently in the cache, scan through requests $r_{i+1}, \ldots, r_n$ to find which of the elements in the cache and r_i has its next access furthest in the future, and evict this element. Because each scan takes $O(n)$ time, each request entails $O(k)$ scans, and there are n requests, the running time of this straightforward approach is $O(kn^2)$.

Instead, we describe an asymptotically faster algorithm, which uses a red-black tree to check whether a given element is currently in the cache, a max-priority queue to retrieve the data element with the furthest access time, and a hash table (resolving collisions by chaining) to map data elements to integer indices. We assume that the data elements can be linearly ordered, so that it makes sense to put them into a red-black tree and a max-priority queue. The following procedure INITIALIZE-CACHE creates and initializes some global data structures that are used by ACCESS.

```
INITIALIZE-CACHE(R, k)
  let T be a new red-black tree
  let P be a new max-priority queue
  let H be a new hash table
  ind = 1
  for i = 1 to n
      j = HASH-SEARCH(r_i)
      if j == NIL
          HASH-INSERT(r_i, ind)
          let S_ind be a new linked list
          j = ind
          ind = ind + 1
      append i to S_j
```

In the above procedure, here is the meaning of various variables:

- The red-black tree T has at most k nodes and holds the distinct data elements that are currently in the cache. We assume that the red-black tree procedures are modified to keep track of the number of nodes currently in the tree, and that the procedure TREE-SIZE returns this value. Because red-black tree T has at most k nodes, we can insert into, delete from, or search in it in $O(\lg k)$ worst-case time.
- The max-priority queue P contains elements with two attributes: *key* is the next access time of a data element, and *value* is the actual data element for each data element in the cache. *key* gives the key and *value* is satellite data in the priority queue. Like the red-black tree T, the max-priority queue contains only elements currently in the cache. We need to maintain T and P separately, however, because T is keyed on the data elements and P is keyed on access times. Using a max-heap to implement P, we can extract the maximum element or insert a new element in $O(\lg k)$ time, and we can find the maximum element in $\Theta(1)$ time.
- The hash table H is a dictionary or a map, which maps each data element to a unique integer. This integer is used to index linked lists, which are described next. We assume that the HASH-INSERT procedure uses the table-expansion technique of Section 17.4.1 to keep the hash table's load factor to be at most some constant α. In this way, the amortized cost per insertion is $\Theta(1)$ and, under the assumption of simple uniform hashing, then by Theorems 11.1 and 11.2, the average-case search time is also $\Theta(1)$.
- For every distinct data element r_i, we create a linked list S_{ind} (where *ind* is obtained through the hash table) holding the indices in the in-

put array where r_i occurs. For example, if the input sequence is $\langle d, b, d, b, d, a, c, d, b, a, c, b\rangle$, then we create four linked lists: S_1 for a, S_2 for b, S_3 for c, and S_4 for d. S_1 holds the indices where a is accessed, and so $S_1 = \langle 6, 10\rangle$. Similarly, $S_2 = \langle 2, 4, 9, 12\rangle$, $S_3 = \langle 7, 11\rangle$ and $S_4 = \langle 1, 3, 5, 8\rangle$.

For each data element r_i, we first check whether there is already a linked list associated with r_i and create a new linked list if not. We retrieve the linked list associated with r_i and append i to it, indicating that an access to r_i occurs at access i.

```
ACCESS(r_i)
  // Compute the next access time for r_i.
  ind = HASH-SEARCH(r_i)
  time = ∞
  delete the head of S_ind
  if S_ind is not empty
      time = head of S_ind
  // Check to see whether r_i is currently in the cache.
  if TREE-SEARCH(T.root, r_i) ≠ NIL
      print "cache hit"
  elseif TREE-SIZE(T) < k
      // Insert in an empty slot in the cache.
      let z be a new node for T
      z.key = r_i
      RB-INSERT(T, z)
      let event be a new object for P
      event.key = time
      event.value = r_i
      INSERT(P, event)
      print "cache miss, inserted " r_i " in empty slot"
  else event = MAXIMUM(P)
      if event.key ≤ time          // r_i has the furthest access time
          print "cache miss, no data element evicted"
      else // evict the element with furthest access time
          print "cache miss, evict data element " event.value
          event = EXTRACT-MAX(P)
          RB-DELETE(T, TREE-SEARCH(T.root, event.value))
          event.key = time
          event.value = r_i
          INSERT(P, event)
          let z be a new node for T
          z.key = r_i
          RB-INSERT(T, z)
```

The procedure ACCESS takes an input r_i and decides which element to evict, if any, from the cache. The first **if** condition properly sets *time* to the next access time of r_i. The head of the linked list associated with r_i contains i; we remove this element from the list, and the new head contains the next access

time for r_i. Then, we check to see whether r_i is already present in the cache. If r_i is not present in the cache, we check to see whether we can store r_i in an empty slot. If there are no empty slots, we have to evict the element with the furthest access time. We retrieve the element with the furthest access time from the max-priority queue and compare it with that of r_i. If r_i's next access is sooner, we evict the element with the furthest access time from the cache (deleting the element from the tree and from the priority queue) and insert r_i into the tree and priority queue.

Under the assumption of simple uniform hashing, the average-case running time of ACCESS is $O(\lg k)$, since it performs a constant number of operations on the red-black tree, priority queue, and hash table. Thus, the average-case running time of CACHE-MANAGER is $O(n \lg k)$.

b. To show that the problem exhibits optimal substructure, we define the subproblem (C, i) as the contents of the cache just before the ith request, where C is a subset of the set of input data elements containing at most k of them. A *solution* to (C, i) is a sequence of decisions that specifies which element to evict (if any) for each request $i, i+1, \ldots, n$. An *optimal solution* to (C, i) is a solution that minimizes the number of cache misses.

Let S be an optimal solution to (C, i). Let S' be the subsolution of S for requests $i+1, i+2, \ldots, n$. If a cache hit occurs on the ith request, then the cache remains unchanged. If a cache miss occurs, then the ith request results in the contents of the cache changing to C' (possibly with $C' = C$ if no element was evicted). We claim that S' is an optimal solution to $(C', i+1)$. Why? If S' were not an optimal solution to $(C', i+1)$, then there exists another solution S'' to $(C', i+1)$ that makes fewer cache misses than S'. By combining S'' with the decision of S at the ith request, we obtain another solution that makes fewer cache misses than S, which contradicts our assumption that S is an optimal solution to (C, i).

Suppose the ith request results in a cache miss. Let P_C be the set of all cache states that can be reached from C through a single decision of the cache manager. The set P_C contains up to $k+1$ states: k of them arising from different elements of the cache being evicted and one arising from the decision of evicting no element. For example, if $C = \{r_1, r_2, r_3\}$ and the requested data element is r_4, then $P_C = \{\{r_1, r_2, r_3\}, \{r_1, r_2, r_4\}, \{r_1, r_3, r_4\}, \{r_2, r_3, r_4\}\}$.

Let $miss(C, i)$ denote the minimum number of cache misses for (C, i). We can state a recurrence for $miss(C, i)$ as

$$miss(C, i) = \begin{cases} 0 & \text{if } i = n \text{ and } r_n \in C \ , \\ 1 & \text{if } i = n \text{ and } r_n \notin C \ , \\ miss(C, i+1) & \text{if } i < n \text{ and } r_i \in C \ , \\ 1 + \min\limits_{C' \in P_C} \{miss(C', i+1)\} & \text{if } i < n \text{ and } r_i \notin C \ . \end{cases}$$

Thus, we conclude that the problem exhibits optimal substructure.

c. To prove that the furthest-in-future strategy yields an optimal solution, we show that the problem exhibits the greedy-choice property. Combined with the optimal-substructure property from part (b), the greedy-choice property will

prove that furthest-in-future produces the minimum possible number of cache misses.

We use the definitions of subproblem, solution, and optimal solution from part (b). Since we will be comparing different solutions, let us define C_{Ai} as the state of the cache for solution A just before the ith request. The following theorem is the key.

Theorem (Greedy-choice property)
Let A be some optimal solution to (C, i). Let b be the element in $C_{Ai} \cup \{r_i\}$ whose next access at the time of the ith request is furthest in the future, at time m. Then, we can construct another solution A' to (C, i) that has the following properties:

1. On the ith request, A' evicts b.
2. For $i + 1 \leq j \leq m$, the caches C_{Aj} and $C_{A'j}$ differ by at most one element. If they differ, then $b \in C_{Aj}$ is always the element in C_{Aj} that is not in $C_{A'j}$. Equivalently, if C_{Aj} and $C_{A'j}$ differ, we can write $C_{Aj} = D_j \cup \{b\}$ and $C_{A'j} = D_j \cup \{x\}$, where D_j is a size-$(k-1)$ set and $x \neq b$ is some data element.
3. For requests $i, \ldots, m-1$, if A has a cache hit, then A' has a cache hit.
4. $C_{Aj} = C_{A'j}$ for $j > m$.
5. For requests $i, \ldots, m$, the number of cache misses produced by A' is at most the number of cache misses produced by A.

Proof If A evicts b at request i, then the proof of the theorem is trivial. Therefore, suppose A evicts data element a on request i, where $a \neq b$. We will prove the theorem by constructing A' inductively for each request.

(1) At request i, A' evicts b instead of a.

(2) We proceed with induction on j, where $i + 1 \leq j \leq m$. The construction for property 1 establishes the base case because $C_{A,i+1}$ and $C_{A',i+1}$ differ by just one element and b is the element in $C_{A,i+1}$ that is not in $C_{A',i+1}$.

For the induction step, suppose property 2 is true for some request j, where $i + 1 \leq j < m$. If A does not evict any element or evicts an element in D_j, then construct A' to make the same decision on request j as A makes. If A evicts b on request j, then construct A' to evict x and keep the same element as A keeps, namely r_j. This construction conserves property 2 for $j + 1$. Note that this construction might sometimes insert duplicate elements in the cache. This situation can easily be dealt with by introducing a dummy element for x.

(3) Suppose A has a cache hit for request j, where $i \leq j \leq m - 1$. Then, $r_j \in D_j$ since $r_j \neq b$. Thus, $r_j \in C_{A'j}$ and A' has a cache hit, too.

(4) By property 2, the cache C_{Am} differs from $C_{A'm}$ by at most one element, with b being the element in C_{Am} that might not be in $C_{A'm}$. If $C_{Am} = C_{A'm}$, then construct A' to make the same decision on request m as A. Otherwise, $C_{Am} \neq C_{A'm}$ and $b \in C_{Am}$. Construct A' to evict x and keep b on request m. Since the mth request is for element b and $b \in C_{Am}$, A has a cache hit so that it does not evict any element. Thus, we can ensure that $C_{A,m+1} = C_{A',m+1}$. From the $(m + 1)$st request on, A' simply makes the same decisions as A.

(5) By property 3, for requests $i, \ldots, m-1$, whenever we have a cache hit for A, we also have a cache hit for A'. Thus, we have to concern ourselves with only the mth request. If A has a cache miss on the mth request, we are done. Otherwise, A has a cache hit on the mth request, and we will prove that there exists at least one request j, where $i+1 \le j \le m-1$, such that the jth request results in a cache miss for A and a cache hit for A'. Because A evicts data element a in request i, then, by our construction of A', $C_{A',i+1} = D_{i+1} \cup \{a\}$. The mth request is for data element b. If A has a cache hit, then because none of the requests $i+1, \ldots, m-1$ were for b, A could not have evicted b and brought it back. Moreover, because A has a cache hit on the mth request, $b \in C_{Am}$. Therefore, A did not evict b in any of requests $i, \ldots, m-1$. By our construction, A' did not evict a. But a request for a occurs at least once before the mth request. Consider the first such instance. At this instance, A has a cache miss and A' has a cache hit. ■

The above theorem and the optimal-substructure property proved in part (b) imply that furthest-in-future produces the minimum number of cache misses.

Lecture Notes for Chapter 17: Amortized Analysis

Chapter 17 overview

Amortized analysis

- Analyze a *sequence* of operations on a data structure.
- ***Goal:*** Show that although some individual operations may be expensive, *on average* the cost per operation is small.

Average in this context does not mean that we're averaging over a distribution of inputs.

- No probability is involved.
- We're talking about *average cost in the worst case*.

Organization

We'll look at 3 methods:

- aggregate analysis
- accounting method
- potential method

Using 3 examples:

- stack with multipop operation
- binary counter
- dynamic tables (later on)

Aggregate analysis

Stack operations

- PUSH(S, x): $O(1)$ each $\Rightarrow$ $O(n)$ for any sequence of n operations.
- POP(S): $O(1)$ each $\Rightarrow$ $O(n)$ for any sequence of n operations.

- MULTIPOP(S, k)
  while S is not empty and $k > 0$
  POP(S)
  $k = k - 1$

Running time of MULTIPOP:

- Linear in # of POP operations.
- Let each PUSH/POP cost 1.
- # of iterations of **while** loop is $\min(s, k)$, where $s =$ # of objects on stack.
- Therefore, total cost $= \min(s, k)$.

Sequence of n PUSH, POP, MULTIPOP operations:

- Worst-case cost of MULTIPOP is $O(n)$.
- Have n operations.
- Therefore, worst-case cost of sequence is $O(n^2)$.

Observation

- Each object can be popped only once per time that it's pushed.
- Have $\leq n$ PUSHes $\Rightarrow \leq n$ POPs, including those in MULTIPOP.
- Therefore, total cost $= O(n)$.
- Average over the n operations $\Rightarrow O(1)$ per operation on average.

Again, notice no probability.

- Showed *worst-case* $O(n)$ cost for sequence.
- Therefore, $O(1)$ per operation on average.

This technique is called ***aggregate analysis***.

Binary counter

- k-bit binary counter $A[0\,..\,k-1]$ of bits, where $A[0]$ is the least significant bit and $A[k-1]$ is the most significant bit.
- Counts upward from 0.
- Value of counter is $\sum_{i=0}^{k-1} A[i] \cdot 2^i$.
- Initially, counter value is 0, so $A[0\,..\,k-1] = 0$.
- To increment, add 1 $\pmod{2^k}$:

  INCREMENT(A, k)
  $i = 0$
  while $i < k$ and $A[i] == 1$
  $A[i] = 0$
  $i = i + 1$
  if $i < k$
  $A[i] = 1$

Example: $k = 3$

[Underlined bits flip. Show costs later.]

counter value	A 2 1 0	cost
0	0 0 $\underline{0}$	0
1	0 $\underline{0\ 1}$	1
2	0 1 $\underline{0}$	3
3	$\underline{0\ 1\ 1}$	4
4	1 0 $\underline{0}$	7
5	1 $\underline{0\ 1}$	8
6	1 1 $\underline{0}$	10
7	$\underline{1\ 1\ 1}$	11
0	0 0 $\underline{0}$	14
$\vdots$	$\vdots$	15

Cost of INCREMENT $= \Theta($# of bits flipped$)$.

Analysis

Each call could flip k bits, so n INCREMENTs takes $O(nk)$ time.

Observation

Not every bit flips every time.

[Show costs from above.]

bit	flips how often	times in n INCREMENTs
0	every time	n
1	$1/2$ the time	$\lfloor n/2 \rfloor$
2	$1/4$ the time	$\lfloor n/4 \rfloor$
	$\vdots$	
i	$1/2^i$ the time	$\lfloor n/2^i \rfloor$
	$\vdots$	
$i \geq k$	never	0

$$\begin{aligned}
\text{Therefore, total \# of flips} &= \sum_{i=0}^{k-1} \lfloor n/2^i \rfloor \\
&< n \sum_{i=0}^{\infty} 1/2^i \\
&= n \left(\frac{1}{1-1/2} \right) \\
&= 2n \ .
\end{aligned}$$

Therefore, n INCREMENTs costs $O(n)$.

Average cost per operation $= O(1)$.

Accounting method

Assign different charges to different operations.

- Some are charged more than actual cost.
- Some are charged less.

Amortized cost = amount we charge.

When amortized cost > actual cost, store the difference *on specific objects* in the data structure as ***credit***.

Use credit later to pay for operations whose actual cost > amortized cost.

Differs from aggregate analysis:

- In the accounting method, different operations can have different costs.
- In aggregate analysis, all operations have same cost.

Need credit to never go negative.

- Otherwise, have a sequence of operations for which the amortized cost is not an upper bound on actual cost.
- Amortized cost would tell us *nothing*.

Let c_i = actual cost of ith operation ,
$\hat{c}_i$ = amortized cost of ith operation .

Then require $\sum_{i=1}^{n} \hat{c}_i \geq \sum_{i=1}^{n} c_i$ for *all* sequences of n operations.

Total credit stored $= \sum_{i=1}^{n} \hat{c}_i - \sum_{i=1}^{n} c_i \underbrace{\geq}_{\text{had better be}} 0$.

Stack

operation	actual cost	amortized cost
PUSH	1	2
POP	1	0
MULTIPOP	$\min(k, s)$	0

Intuition

When pushing an object, pay \$2.

- \$1 pays for the PUSH.
- \$1 is prepayment for it being popped by either POP or MULTIPOP.
- Since each object has \$1, which is credit, the credit can never go negative.
- Therefore, total amortized cost, $= O(n)$, is an upper bound on total actual cost.

Binary counter

Charge \$2 to set a bit to 1.

- \$1 pays for setting a bit to 1.
- \$1 is prepayment for flipping it back to 0.
- Have \$1 of credit for every 1 in the counter.
- Therefore, credit ≥ 0.

Amortized cost of INCREMENT:

- Cost of resetting bits to 0 is paid by credit.
- At most 1 bit is set to 1.
- Therefore, amortized cost $\leq$ \$2.
- For n operations, amortized cost $= O(n)$.

Potential method

Like the accounting method, but think of the credit as *potential* stored with the entire data structure.

- Accounting method stores credit with specific objects.
- Potential method stores potential in the data structure as a whole.
- Can release potential to pay for future operations.
- Most flexible of the amortized analysis methods.

Let D_i = data structure after ith operation ,
D_0 = initial data structure ,
c_i = actual cost of ith operation ,
$\hat{c}_i$ = amortized cost of ith operation .

Potential function $\Phi : D_i \to \mathbb{R}$

$\Phi(D_i)$ is the *potential* associated with data structure D_i.

$$\begin{aligned}\hat{c}_i &= c_i + \Phi(D_i) - \Phi(D_{i-1}) \\ &= c_i + \underbrace{\Delta\Phi(D_i)}_{\text{increase in potential due to } i\text{th operation}} .\end{aligned}$$

$$\begin{aligned}\text{Total amortized cost} &= \sum_{i=1}^{n} \hat{c}_i \\ &= \sum_{i=1}^{n} (c_i + \Phi(D_i) - \Phi(D_{i-1})) \\ &\quad \text{(telescoping sum: every term other than } D_0 \text{ and } D_n \\ &\qquad \text{is added once and subtracted once)} \\ &= \sum_{i=1}^{n} c_i + \Phi(D_n) - \Phi(D_0) .\end{aligned}$$

If we require that $\Phi(D_i) \geq \Phi(D_0)$ for all i, then the amortized cost is always an upper bound on actual cost.

In practice: $\Phi(D_0) = 0$, $\Phi(D_i) \geq 0$ for all i.

Stack

Φ = # of objects in stack
(= # of \$1 bills in accounting method)

D_0 = empty stack $\Rightarrow \Phi(D_0) = 0$.

Since # of objects in stack is always ≥ 0, $\Phi(D_i) \geq 0 = \Phi(D_0)$ for all i.

operation	actual cost	$\Delta\Phi$	amortized cost
PUSH	1	$(s+1) - s = 1$ where s = # of objects initially	$1 + 1 = 2$
POP	1	$(s-1) - s = -1$	$1 - 1 = 0$
MULTIPOP	$k' = \min(k, s)$	$(s - k') - s = -k'$	$k' - k' = 0$

Therefore, amortized cost of a sequence of n operations $= O(n)$.

Binary counter

$\Phi = b_i$ = # of 1's after ith INCREMENT

Suppose ith operation resets t_i bits to 0.

$c_i \leq t_i + 1$ (resets t_i bits, sets ≤ 1 bit to 1)

- If $b_i = 0$, the ith operation reset all k bits and didn't set one, so $b_{i-1} = t_i = k \Rightarrow b_i = b_{i-1} - t_i$.
- If $b_i > 0$, the ith operation reset t_i bits, set one, so $b_i = b_{i-1} - t_i + 1$.
- Either way, $b_i \leq b_{i-1} - t_i + 1$.
- Therefore,

$$\begin{aligned} \Delta\Phi(D_i) &\leq (b_{i-1} - t_i + 1) - b_{i-1} \\ &= 1 - t_i \ . \end{aligned}$$

$$\begin{aligned} \widehat{c_i} &= c_i + \Delta\Phi(D_i) \\ &\leq (t_i + 1) + (1 - t_i) \\ &= 2 \ . \end{aligned}$$

If counter starts at 0, $\Phi(D_0) = 0$.

Therefore, amortized cost of n operations $= O(n)$.

Dynamic tables

A nice use of amortized analysis.

Scenario

- Have a table—maybe a hash table.
- Don't know in advance how many objects will be stored in it.
- When it fills, must reallocate with a larger size, copying all objects into the new, larger table.
- When it gets sufficiently small, *might* want to reallocate with a smaller size.

Details of table organization not important.

Goals

1. $O(1)$ amortized time per operation.
2. Unused space always $\leq$ constant fraction of allocated space.

Load factor $\alpha = num/size$, where num = # items stored, $size$ = allocated size.

If $size = 0$, then $num = 0$. Call $\alpha = 1$.

Never allow $\alpha > 1$.

Keep $\alpha >$ a constant fraction $\Rightarrow$ goal (2).

Table expansion

Consider only insertion.

- When the table becomes full, double its size and reinsert all existing items.
- Guarantees that $\alpha \geq 1/2$.
- Each time we actually insert an item into the table, it's an ***elementary insertion***.

```
TABLE-INSERT(T, x)
  if T.size == 0
      allocate T.table with 1 slot
      T.size = 1
  if T.num == T.size                          // expand?
      allocate new-table with 2 · T.size slots
      insert all items in T.table into new-table   // T.num elem insertions
      free T.table
      T.table = new-table
      T.size = 2 · T.size
  insert x into T.table                       // 1 elem insertion
  T.num = T.num + 1
```

Initially, $T.num = T.size = 0$.

Running time

Charge 1 per elementary insertion. Count only elementary insertions, since all other costs together are constant per call.

c_i = actual cost of ith operation

- If not full, $c_i = 1$.
- If full, have $i - 1$ items in the table at the start of the ith operation. Have to copy all $i - 1$ existing items, then insert ith item $\Rightarrow c_i = i$.

n operations $\Rightarrow c_i = O(n) \Rightarrow O(n^2)$ time for n operations.

Of course, we don't always expand:

$$c_i = \begin{cases} i & \text{if } i-1 \text{ is exact power of 2 ,} \\ 1 & \text{otherwise .} \end{cases}$$

$$\begin{aligned} \text{Total cost} &= \sum_{i=1}^{n} c_i \\ &\le n + \sum_{j=0}^{\lfloor \lg n \rfloor} 2^j \\ &= n + \frac{2^{\lfloor \lg n \rfloor + 1} - 1}{2 - 1} \\ &< n + 2n \\ &= 3n \end{aligned}$$

Therefore, **aggregate analysis** says amortized cost per operation = 3.

Accounting method

Charge \$3 per insertion of x.

- \$1 pays for x's insertion.
- \$1 pays for x to be moved in the future.
- \$1 pays for some other item to be moved.

Suppose we've just expanded, *size* $= m$ before next expansion, *size* $= 2m$ after next expansion.

- Assume that the expansion used up all the credit, so that there's no credit stored after the expansion.
- Will expand again after another m insertions.
- Each insertion will put \$1 on one of the m items that were in the table just after expansion and will put \$1 on the item inserted.
- Have \$$2m$ of credit by next expansion, when there are $2m$ items to move. Just enough to pay for the expansion, with no credit left over!

Potential method

$\Phi(T) = 2 \cdot T.num - T.size$

- Initially, $num = size = 0 \Rightarrow \Phi = 0$.
- Just after expansion, $size = 2 \cdot num \Rightarrow \Phi = 0$.
- Just before expansion, $size = num \Rightarrow \Phi = num \Rightarrow$ have enough potential to pay for moving all items.
- Need $\Phi \geq 0$, always.

  Always have

$$\begin{array}{rcll} size & \geq & num & \geq & size/2 & \Rightarrow \\ & & 2 \cdot num & \geq & size & \Rightarrow \\ & & \Phi & \geq & 0 \, . \end{array}$$

Amortized cost of i th operation

$$\begin{aligned} num_i &= num \text{ after } i \text{th operation} , \\ size_i &= size \text{ after } i \text{th operation} , \\ \Phi_i &= \Phi \text{ after } i \text{th operation} . \end{aligned}$$

- If no expansion:

$$\begin{aligned} size_i &= size_{i-1} , \\ num_i &= num_{i-1} + 1 , \\ c_i &= 1 . \end{aligned}$$

  Then we have

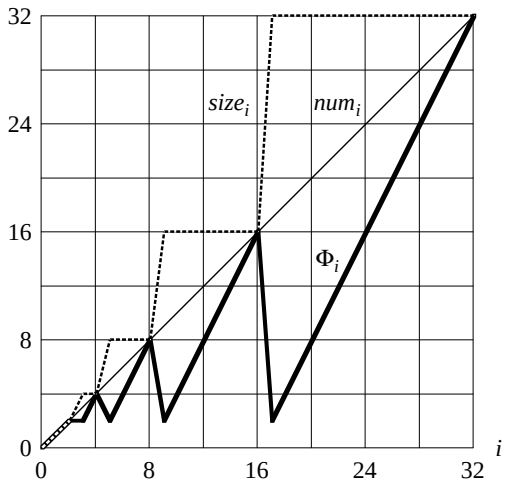
$$\begin{aligned} \widehat{c}_i &= c_i + \Phi_i - \Phi_{i-1} \\ &= 1 + (2 \cdot num_i - size_i) - (2 \cdot num_{i-1} - size_{i-1}) \\ &= 1 + (2 \cdot num_i - size_i) - (2(num_i - 1) - size_i) \\ &= 1 + 2 \\ &= 3 . \end{aligned}$$

- If expansion:

$$\begin{aligned} size_i &= 2 \cdot size_{i-1} , \\ size_{i-1} &= num_{i-1} = num_i - 1 , \\ c_i &= num_{i-1} + 1 = num_i . \end{aligned}$$

  Then we have

$$\begin{aligned} \widehat{c}_i &= c_i + \Phi_i + \Phi_{i-1} \\ &= num_i + (2 \cdot num_i - size_i) - (2 \cdot num_{i-1} - size_{i-1}) \\ &= num_i + (2 \cdot num_i - 2(num_i - 1)) - (2(num_i - 1) - (num_i - 1)) \\ &= num_i + 2 - (num_i - 1) \\ &= 3 . \end{aligned}$$



Expansion and contraction

When α drops too low, contract the table.

- Allocate a new, smaller one.
- Copy all items.

Still want

- α bounded from below by a constant,
- amortized cost per operation $= O(1)$.

Measure cost in terms of elementary insertions and deletions.

"Obvious strategy"

- Double size when inserting into a full table (when $\alpha = 1$, so that after insertion α would become > 1).
- Halve size when deletion would make table less than half full (when $\alpha = 1/2$, so that after deletion α would become $< 1/2$).
- Then always have $1/2 \le \alpha \le 1$.
- Suppose we fill table.

  Then insert $\Rightarrow$ double
  2 deletes $\Rightarrow$ halve
  2 inserts $\Rightarrow$ double
  2 deletes $\Rightarrow$ halve
  $\ldots$

Not performing enough operations after expansion or contraction to pay for the next one.

Simple solution

- Double as before: when inserting with $\alpha = 1 \Rightarrow$ after doubling, $\alpha = 1/2$.
- Halve size when deleting with $\alpha = 1/4 \Rightarrow$ after halving, $\alpha = 1/2$.
- Thus, immediately after either expansion or contraction, have $\alpha = 1/2$.
- Always have $1/4 \le \alpha \le 1$.

Intuition

- Want to make sure that we perform enough operations between consecutive expansions/contractions to pay for the change in table size.
- Need to delete half the items before contraction.
- Need to double number of items before expansion.
- Either way, number of operations between expansions/contractions is at least a constant fraction of number of items copied.

$$\Phi(T) = \begin{cases} 2 \cdot T.num - T.size & \text{if } \alpha \ge 1/2 \text{ ,} \\ T.size/2 - T.num & \text{if } \alpha < 1/2 \text{ .} \end{cases}$$

T empty $\Rightarrow \Phi = 0$.

$\alpha \ge 1/2 \Rightarrow num \ge size/2 \Rightarrow 2 \cdot num \ge size \Rightarrow \Phi \ge 0$.

$\alpha < 1/2 \Rightarrow num < size/2 \Rightarrow \Phi \ge 0$.

Further intuition

Φ measures how far from $\alpha = 1/2$ we are.

- $\alpha = 1/2 \Rightarrow \Phi = 2 \cdot num - 2 \cdot num = 0$.
- $\alpha = 1 \Rightarrow \Phi = 2 \cdot num - num = num$.
- $\alpha = 1/4 \Rightarrow \Phi = size/2 - num = 4 \cdot num/2 - num = num$.
- Therefore, when we double or halve, have enough potential to pay for moving all *num* items.
- Potential increases linearly between $\alpha = 1/2$ and $\alpha = 1$, and it also increases linearly between $\alpha = 1/2$ and $\alpha = 1/4$.
- Since α has different distances to go to get to 1 or $1/4$, starting from $1/2$, rate of increase of Φ differs.
  - For α to go from $1/2$ to 1, *num* increases from $size/2$ to *size*, for a total increase of $size/2$. Φ increases from 0 to *size*. Thus, Φ needs to increase by 2 for each item inserted. That's why there's a coefficient of 2 on the $T.num$ term in the formula for Φ when $\alpha \ge 1/2$.
  - For α to go from $1/2$ to $1/4$, *num* decreases from $size/2$ to $size/4$, for a total decrease of $size/4$. Φ increases from 0 to $size/4$. Thus, Φ needs to increase by 1 for each item deleted. That's why there's a coefficient of -1 on the $T.num$ term in the formula for Φ when $\alpha < 1/2$.

Amortized costs: more cases

- insert, delete
- $\alpha \ge 1/2$, $\alpha < 1/2$ (use α_i, since α can vary a lot)
- *size* does/doesn't change

Insert

- $\alpha_{i-1} \geq 1/2$, same analysis as before $\Rightarrow \hat{c}_i = 3$.
- $\alpha_{i-1} < 1/2 \Rightarrow$ *no expansion* (only occurs when $\alpha_{i-1} = 1$).
  - If $\alpha_{i-1} < 1/2$ and $\alpha_i < 1/2$:
    $$\begin{aligned}\hat{c}_i &= c_i + \Phi_i + \Phi_{i-1} \\ &= 1 + (size_i/2 - num_i) - (size_{i-1}/2 - num_{i-1}) \\ &= 1 + (size_i/2 - num_i) - (size_i/2 - (num_i - 1)) \\ &= 0 .\end{aligned}$$
  - If $\alpha_{i-1} < 1/2$ and $\alpha_i \geq 1/2$:
    $$\begin{aligned}\hat{c}_i &= 1 + (2 \cdot num_i - size_i) - (size_{i-1}/2 - num_{i-1}) \\ &= 1 + (2(num_{i-1} + 1) - size_{i-1}) - (size_{i-1}/2 - num_{i-1}) \\ &= 3 \cdot num_{i-1} - \frac{3}{2} \cdot size_{i-1} + 3 \\ &= 3 \cdot \alpha_{i-1} size_{i-1} - \frac{3}{2} \cdot size_{i-1} + 3 \\ &< \frac{3}{2} \cdot size_{i-1} - \frac{3}{2} \cdot size_{i-1} + 3 \\ &= 3 .\end{aligned}$$
    Therefore, amortized cost of insert is < 3.

Delete

- If $\alpha_{i-1} < 1/2$, then $\alpha_i < 1/2$.
  - If no contraction:
    $$\begin{aligned}\hat{c}_i &= 1 + (size_i/2 - num_i) - (size_{i-1}/2 - num_{i-1}) \\ &= 1 + (size_i/2 - num_i) - (size_i/2 - (num_i + 1)) \\ &= 2 .\end{aligned}$$
  - If contraction:
    $$\begin{aligned}\hat{c}_i &= \underbrace{(num_i + 1)}_{\text{move + delete}} + (size_i/2 - num_i) - (size_{i-1}/2 - num_{i-1}) \\ &\quad [size_i/2 = size_{i-1}/4 = num_{i-1} = num_i + 1] \\ &= (num_i + 1) + ((num_i + 1) - num_i) - ((2 \cdot num_i + 2) - (num_i + 1)) \\ &= 1 .\end{aligned}$$
- If $\alpha_{i-1} \geq 1/2$, then no contraction.
  - If $\alpha_i \geq 1/2$:
    $$\begin{aligned}\hat{c}_i &= 1 + (2 \cdot num_i - size_i) - (2 \cdot num_{i-1} - size_{i-1}) \\ &= 1 + (2 \cdot num_i - size_i) - (2 \cdot num_i + 2 - size_i) \\ &= -1 .\end{aligned}$$

- If $\alpha_i < 1/2$, since $\alpha_{i-1} \geq 1/2$, have

  $$num_i = num_{i-1} - 1 \geq \frac{1}{2} \cdot size_{i-1} - 1 = \frac{1}{2} \cdot size_i - 1 \ .$$

  Thus,

  $$\begin{aligned}
  \widehat{c}_i &= 1 + (size_i/2 - num_i) - (2 \cdot num_{i-1} - size_{i-1}) \\
  &= 1 + (size_i/2 - num_i) - (2 \cdot num_i + 2 - size_i) \\
  &= -1 + \frac{3}{2} \cdot size_i - 3 \cdot num_i \\
  &\leq -1 + \frac{3}{2} \cdot size_i - 3\left(\frac{1}{2} \cdot size_i - 1\right) \\
  &= 2 \ .
  \end{aligned}$$

Therefore, amortized cost of delete is ≤ 2.

Solutions for Chapter 17: Amortized Analysis

Solution to Exercise 17.1-3
This solution is also posted publicly

Let c_i = cost of ith operation.

$$c_i = \begin{cases} i & \text{if } i \text{ is an exact power of 2 ,} \\ 1 & \text{otherwise .} \end{cases}$$

Operation	Cost
1	1
2	2
3	1
4	4
5	1
6	1
7	1
8	8
9	1
10	1
⋮	⋮

n operations cost

$$\sum_{i=1}^{n} c_i \le n + \sum_{j=0}^{\lg n} 2^j = n + (2n - 1) < 3n \ .$$

(Note: Ignoring floor in upper bound of $\sum 2^j$.)

Average cost of operation $= \dfrac{\text{Total cost}}{\text{\# operations}} < 3$.

By aggregate analysis, the amortized cost per operation $= O(1)$.

Solution to Exercise 17.2-1

[We assume that the only way in which COPY *is invoked is automatically, after every sequence of* k PUSH *and* POP *operations.]*

Charge \$2 for each PUSH and POP operation and \$0 for each COPY. When we call PUSH, we use \$1 to pay for the operation, and we store the other \$1 on the item pushed. When we call POP, we again use \$1 to pay for the operation, and we store the other \$1 in the stack itself. Because the stack size never exceeds k, the actual cost of a COPY operation is at most \$$k$, which is paid by the \$$k$ found in the items in the stack and the stack itself. Since k PUSH and POP operations occur between two consecutive COPY operations, \$$k$ of credit are stored, either on individual items (from PUSH operations) or in the stack itself (from POP operations) by the time a COPY occurs. Since the amortized cost of each operation is $O(1)$ and the amount of credit never goes negative, the total cost of n operations is $O(n)$.

Solution to Exercise 17.2-2

This solution is also posted publicly

Let c_i = cost of ith operation.

$$c_i = \begin{cases} i & \text{if } i \text{ is an exact power of 2 ,} \\ 1 & \text{otherwise .} \end{cases}$$

Charge each operation \$3 (amortized cost $\hat{c}_i$).

- If i is not an exact power of 2, pay \$1, and store \$2 as credit.
- If i is an exact power of 2, pay \$$i$, using stored credit.

Operation	Cost	Actual cost	Credit remaining
1	3	1	2
2	3	2	3
3	3	1	5
4	3	4	4
5	3	1	6
6	3	1	8
7	3	1	10
8	3	8	5
9	3	1	7
10	3	1	9
⋮	⋮	⋮	⋮

Since the amortized cost is \$3 per operation, $\sum_{i=1}^{n} \hat{c}_i = 3n$.

We know from Exercise 17.1-3 that $\sum_{i=1}^{n} c_i < 3n$.

Then we have $\sum_{i=1}^{n} \hat{c}_i \geq \sum_{i=1}^{n} c_i \Rightarrow$ credit = amortized cost − actual cost ≥ 0.

Since the amortized cost of each operation is $O(1)$, and the amount of credit never goes negative, the total cost of n operations is $O(n)$.

Solution to Exercise 17.2-3

This solution is also posted publicly

We introduce a new field $A.max$ to hold the index of the high-order 1 in A. Initially, $A.max$ is set to -1, since the low-order bit of A is at index 0, and there are initially no 1's in A. The value of $A.max$ is updated as appropriate when the counter is incremented or reset, and we use this value to limit how much of A must be looked at to reset it. By controlling the cost of RESET in this way, we can limit it to an amount that can be covered by credit from earlier INCREMENTs.

```
INCREMENT(A)
 i = 0
 while i < A.length and A[i] == 1
      A[i] = 0
      i = i + 1
 if i < A.length
      A[i] = 1
      // Additions to book's INCREMENT start here.
      if i > A.max
           A.max = i
 else A.max = −1
```

```
RESET(A)
 for i = 0 to A.max
      A[i] = 0
 A.max = −1
```

As for the counter in the book, we assume that it costs \$1 to flip a bit. In addition, we assume it costs \$1 to update $A.max$.

Setting and resetting of bits by INCREMENT will work exactly as for the original counter in the book: \$1 will pay to set one bit to 1; \$1 will be placed on the bit that is set to 1 as credit; the credit on each 1 bit will pay to reset the bit during incrementing.

In addition, we'll use \$1 to pay to update *max*, and if *max* increases, we'll place an additional \$1 of credit on the new high-order 1. (If *max* doesn't increase, we can just waste that \$1—it won't be needed.) Since RESET manipulates bits at positions only up to $A.max$, and since each bit up to there must have become the high-order 1

at some time before the high-order 1 got up to $A.max$, every bit seen by RESET has \$1 of credit on it. So the zeroing of bits of A by RESET can be completely paid for by the credit stored on the bits. We just need \$1 to pay for resetting *max*.

Thus charging \$4 for each INCREMENT and \$1 for each RESET is sufficient, so the sequence of n INCREMENT and RESET operations takes $O(n)$ time.

Solution to Exercise 17.3-3

Let D_i be the heap after the ith operation, and let D_i consist of n_i elements. Also, let k be a constant such that each INSERT or EXTRACT-MIN operation takes at most $k \ln n$ time, where $n = \max(n_{i-1}, n_i)$. (We don't want to worry about taking the log of 0, and at least one of n_{i-1} and n_i is at least 1. We'll see later why we use the natural log.)

Define

$$\Phi(D_i) = \begin{cases} 0 & \text{if } n_i = 0\,, \\ k\, n_i \ln n_i & \text{if } n_i > 0\,. \end{cases}$$

This function exhibits the characteristics we like in a potential function: if we start with an empty heap, then $\Phi(D_0) = 0$, and we always maintain that $\Phi(D_i) \ge 0$.

Before proving that we achieve the desired amortized times, we show that if $n \ge 2$, then $n \ln \frac{n}{n-1} \le 2$. We have

$$\begin{aligned} n \ln \frac{n}{n-1} &= n \ln \left(1 + \frac{1}{n-1}\right) \\ &= \ln \left(1 + \frac{1}{n-1}\right)^n \\ &\le \ln \left(e^{\frac{1}{n-1}}\right)^n && \text{(since } 1 + x \le e^x \text{ for all real } x\text{)} \\ &= \ln e^{\frac{n}{n-1}} \\ &= \frac{n}{n-1} \\ &\le 2\,, \end{aligned}$$

assuming that $n \ge 2$. (The equation $\ln e^{\frac{n}{n-1}} = \frac{n}{n-1}$ is why we use the natural log.)

If the ith operation is an INSERT, then $n_i = n_{i-1} + 1$. If the ith operation inserts into an empty heap, then $n_i = 1$, $n_{i-1} = 0$, and the amortized cost is

$$\begin{aligned} \hat{c}_i &= c_i + \Phi(D_i) - \Phi(D_{i-1}) \\ &\le k \ln 1 + k \cdot 1 \ln 1 - 0 \\ &= 0\,. \end{aligned}$$

If the ith operation inserts into a nonempty heap, then $n_i = n_{i-1} + 1$, and the amortized cost is

$$\begin{aligned} \hat{c}_i &= c_i + \Phi(D_i) - \Phi(D_{i-1}) \\ &\le k \ln n_i + k\, n_i \ln n_i - k\, n_{i-1} \ln n_{i-1} \\ &= k \ln n_i + k\, n_i \ln n_i - k(n_i - 1) \ln (n_i - 1) \end{aligned}$$

$$
\begin{aligned}
&= k \ln n_i + k n_i \ln n_i - k n_i \ln(n_i - 1) + k \ln(n_i - 1) \\
&< 2k \ln n_i + k n_i \ln \frac{n_i}{n_i - 1} \\
&\le 2k \ln n_i + 2k \\
&= O(\lg n_i) \, .
\end{aligned}
$$

If the ith operation is an EXTRACT-MIN, then $n_i = n_{i-1} - 1$. If the ith operation extracts the one and only heap item, then $n_i = 0$, $n_{i-1} = 1$, and the amortized cost is

$$
\begin{aligned}
\widehat{c}_i &= c_i + \Phi(D_i) - \Phi(D_{i-1}) \\
&\le k \ln 1 + 0 - k \cdot 1 \ln 1 \\
&= 0 \, .
\end{aligned}
$$

If the ith operation extracts from a heap with more than 1 item, then $n_i = n_{i-1} - 1$ and $n_{i-1} \ge 2$, and the amortized cost is

$$
\begin{aligned}
\widehat{c}_i &= c_i + \Phi(D_i) - \Phi(D_{i-1}) \\
&\le k \ln n_{i-1} + k n_i \ln n_i - k n_{i-1} \ln n_{i-1} \\
&= k \ln n_{i-1} + k(n_{i-1} - 1) \ln(n_{i-1} - 1) - k n_{i-1} \ln n_{i-1} \\
&= k \ln n_{i-1} + k n_{i-1} \ln(n_{i-1} - 1) - k \ln(n_{i-1} - 1) - k n_{i-1} \ln n_{i-1} \\
&= k \ln \frac{n_{i-1}}{n_{i-1} - 1} + k n_{i-1} \ln \frac{n_{i-1} - 1}{n_{i-1}} \\
&< k \ln \frac{n_{i-1}}{n_{i-1} - 1} + k n_{i-1} \ln 1 \\
&= k \ln \frac{n_{i-1}}{n_{i-1} - 1} \\
&\le k \ln 2 \qquad \text{(since } n_{i-1} \ge 2\text{)} \\
&= O(1) \, .
\end{aligned}
$$

A slightly different potential function—which may be easier to work with—is as follows. For each node x in the heap, let $d_i(x)$ be the depth of x in D_i. Define

$$
\begin{aligned}
\Phi(D_i) &= \sum_{x \in D_i} k(d_i(x) + 1) \\
&= k \left(n_i + \sum_{x \in D_i} d_i(x) \right) \, ,
\end{aligned}
$$

where k is defined as before.

Initially, the heap has no items, which means that the sum is over an empty set, and so $\Phi(D_0) = 0$. We always have $\Phi(D_i) \ge 0$, as required.

Observe that after an INSERT, the sum changes only by an amount equal to the depth of the new last node of the heap, which is $\lfloor \lg n_i \rfloor$. Thus, the change in potential due to an INSERT is $k(1 + \lfloor \lg n_i \rfloor)$, and so the amortized cost is $O(\lg n_i) + O(\lg n_i) = O(\lg n_i) = O(\lg n)$.

After an EXTRACT-MIN, the sum changes by the negative of the depth of the old last node in the heap, and so the potential *decreases* by $k(1 + \lfloor \lg n_{i-1} \rfloor)$. The amortized cost is at most $k \lg n_{i-1} - k(1 + \lfloor \lg n_{i-1} \rfloor) = O(1)$.

Solution to Problem 17-2

a. The SEARCH operation can be performed by searching each of the individually sorted arrays. Since all the individual arrays are sorted, searching one of them using a binary search algorithm takes $O(\lg m)$ time, where m is the size of the array. In an unsuccessful search, the time is $\Theta(\lg m)$. In the worst case, we may assume that all the arrays $A_0, A_1, \ldots, A_{k-1}$ are full, $k = \lceil \lg(n+1) \rceil$, and we perform an unsuccessful search. The total time taken is

$$\begin{aligned} T(n) &= \Theta(\lg 2^{k-1} + \lg 2^{k-2} + \cdots + \lg 2^1 + \lg 2^0) \\ &= \Theta((k-1) + (k-2) + \cdots + 1 + 0) \\ &= \Theta(k(k-1)/2) \\ &= \Theta(\lceil \lg(n+1) \rceil (\lceil \lg(n+1) \rceil - 1)/2) \\ &= \Theta\left(\lg^2 n\right) . \end{aligned}$$

Thus, the worst-case running time is $\Theta(\lg^2 n)$.

b. We create a new sorted array of size 1 containing the new element to be inserted. If array A_0 (which has size 1) is empty, then we replace A_0 with the new sorted array. Otherwise, we merge sort the two arrays into another sorted array of size 2. If A_1 is empty, then we replace A_1 with the new array; otherwise we merge sort the arrays as before and continue. Since array A_i is of size 2^i, if we merge sort two arrays of size 2^i each, we obtain one of size 2^{i+1}, which is the size of A_{i+1}. Thus, this method will result in another list of arrays in the same structure that we had before.

Let us analyze its worst-case running time. We will assume that merge sort takes $2m$ time to merge two sorted lists of size m each. If all the arrays $A_0, A_1, \ldots, A_{k-2}$ are full, then the running time to fill array A_{k-1} would be

$$\begin{aligned} T(n) &= 2\left(2^0 + 2^1 + \cdots + 2^{k-2}\right) \\ &= 2(2^{k-1} - 1) \\ &= 2^k - 2 \\ &= \Theta(n) . \end{aligned}$$

Therefore, the worst-case time to insert an element into this data structure is $\Theta(n)$.

However, let us now analyze the amortized running time. Using the aggregate method, we compute the total cost of a sequence of n inserts, starting with the empty data structure. Let r be the position of the rightmost 0 in the binary representation $\langle n_{k-1}, n_{k-2}, \ldots, n_0 \rangle$ of n, so that $n_j = 1$ for $j = 0, 1, \ldots, r-1$. The cost of an insertion when n items have already been inserted is

$$\sum_{j=0}^{r-1} 2 \cdot 2^j = O(2^r) .$$

Furthermore, $r = 0$ half the time, $r = 1$ a quarter of the time, and so on. There are at most $\lceil n/2^r \rceil$ insertions for each value of r. The total cost of the n operations is therefore bounded by

$$O\left(\sum_{r=0}^{\lceil \lg(n+1) \rceil} \left(\left\lceil \frac{n}{2^r} \right\rceil\right) 2^r\right) = O(n \lg n) .$$

The amortized cost per INSERT operation, therefore is $O(\lg n)$.

We can also use the accounting method to analyze the running time. We can charge $\$k$ to insert an element. \$1 pays for the insertion, and we put $\$(k-1)$ on the inserted item to pay for it being involved in merges later on. Each time it is merged, it moves to a higher-indexed array, i.e., from A_i to A_{i+1}. It can move to a higher-indexed array at most $k-1$ times, and so the $\$(k-1)$ on the item suffices to pay for all the times it will ever be involved in merges. Since $k = \Theta(\lg n)$, we have an amortized cost of $\Theta(\lg n)$ per insertion.

c. DELETE(x) will be implemented as follows:

1. Find the smallest j for which the array A_j with 2^j elements is full. Let y be the last element of A_j.
2. Let x be in the array A_i. If necessary, find which array this is by using the search procedure.
3. Remove x from A_i and put y into A_i. Then move y to its correct place in A_i.
4. Divide A_j (which now has $2^j - 1$ elements left): The first element goes into array A_0, the next 2 elements go into array A_1, the next 4 elements go into array A_2, and so forth. Mark array A_j as empty. The new arrays are created already sorted.

The cost of DELETE is $\Theta(n)$ in the worst case, where $i = k-1$ and $j = k-2$: $\Theta(\lg n)$ to find A_j, $\Theta(\lg^2 n)$ to find A_i, $\Theta(2^i) = \Theta(n)$ to put y in its correct place in array A_i, and $\Theta(2^j) = \Theta(n)$ to divide array A_j. The following sequence of n operations, where $n/3$ is a power of 2, yields an amortized cost that is no better: perform $n/3$ INSERT operations, followed by $n/3$ pairs of DELETE and INSERT. It costs $O(n \lg n)$ to do the first $n/3$ INSERT operations. This creates a single full array. Each subsequent DELETE/INSERT pair costs $\Theta(n)$ for the DELETE to divide the full array and another $\Theta(n)$ for the INSERT to recombine it. The total is then $\Theta(n^2)$, or $\Theta(n)$ per operation.

Solution to Problem 17-4

a. For RB-INSERT, consider a complete red-black tree in which the colors alternate between levels. That is, the root is black, the children of the root are red, the grandchildren of the root are black, the great-grandchildren of the root are red, and so on. When a node is inserted as a red child of one of the red leaves, then case 1 of RB-INSERT-FIXUP occurs $(\lg(n+1))/2$ times, so that there are $\Omega(\lg n)$ color changes to fix the colors of nodes on the path from the inserted node to the root.

For RB-DELETE, consider a complete red-black tree in which all nodes are black. If a leaf is deleted, then the double blackness will be pushed all the way up to the root, with a color change at each level (case 2 of RB-DELETE-FIXUP), for a total of $\Omega(\lg n)$ color changes.

b. All cases except for case 1 of RB-INSERT-FIXUP and case 2 of RB-DELETE-FIXUP are terminating.

c. Case 1 of RB-INSERT-FIXUP reduces the number of red nodes by 1. As Figure 13.5 shows, node z's parent and uncle change from red to black, and z's grandparent changes from black to red. Hence, $\Phi(T') = \Phi(T) - 1$.

d. Lines 1–16 of RB-INSERT cause one node insertion and a unit increase in potential. The nonterminating case of RB-INSERT-FIXUP (Case 1) makes three color changes and decreases the potential by 1. The terminating cases of RB-INSERT-FIXUP (cases 2 and 3) cause one rotation each and do not affect the potential. (Although case 3 makes color changes, the potential does not change. As Figure 13.6 shows, node z's parent changes from red to black, and z's grandparent changes from black to red.)

e. The number of structural modifications and amount of potential change resulting from lines 1–16 of RB-INSERT and from the terminating cases of RB-INSERT-FIXUP are $O(1)$, and so the amortized number of structural modifications of these parts is $O(1)$. The nonterminating case of RB-INSERT-FIXUP may repeat $O(\lg n)$ times, but its amortized number of structural modifications is 0, since by our assumption the unit decrease in the potential pays for the structural modifications needed. Therefore, the amortized number of structural modifications performed by RB-INSERT is $O(1)$.

f. From Figure 13.5, we see that case 1 of RB-INSERT-FIXUP makes the following changes to the tree:

- Changes a black node with two red children (node C) to a red node, resulting in a potential change of -2.
- Changes a red node (node A in part (a) and node B in part (b)) to a black node with one red child, resulting in no potential change.
- Changes a red node (node D) to a black node with no red children, resulting in a potential change of 1.

The total change in potential is -1, which pays for the structural modifications performed, and thus the amortized number of structural modifications in case 1 (the nonterminating case) is 0. The terminating cases of RB-INSERT-FIXUP cause $O(1)$ structural changes. Because $w(v)$ is based solely on node colors and the number of color changes caused by terminating cases is $O(1)$, the change in potential in terminating cases is $O(1)$. Hence, the amortized number of structural modifications in the terminating cases is $O(1)$. The overall amortized number of structural modifications in RB-INSERT, therefore, is $O(1)$.

g. Figure 13.7 shows that case 2 of RB-DELETE-FIXUP makes the following changes to the tree:

- Changes a black node with no red children (node D) to a red node, resulting in a potential change of -1.
- If B is red, then it loses a black child, with no effect on potential.
- If B is black, then it goes from having no red children to having one red child, resulting in a potential change of -1.

The total change in potential is either -1 or -2, depending on the color of B. In either case, one unit of potential pays for the structural modifications performed, and thus the amortized number of structural modifications in case 2 (the nonterminating case) is at most 0. The terminating cases of RB-DELETE cause $O(1)$ structural changes. Because $w(v)$ is based solely on node colors and the number of color changes caused by terminating cases is $O(1)$, the change in potential in terminating cases is $O(1)$. Hence, the amortized number of structural changes in the terminating cases is $O(1)$. The overall amortized number of structural modifications in RB-DELETE-FIXUP, therefore, is $O(1)$.

h. Since the amortized number structural modification in each operation is $O(1)$, the actual number of structural modifications for any sequence of m RB-INSERT and RB-DELETE operations on an initially empty red-black tree is $O(m)$ in the worst case.

Lecture Notes for Chapter 21: Data Structures for Disjoint Sets

Chapter 21 overview

Disjoint-set data structures

- Also known as "union find."
- Maintain collection $\mathcal{S} = \{S_1, \ldots, S_k\}$ of disjoint dynamic (changing over time) sets.
- Each set is identified by a *representative*, which is some member of the set.

  Doesn't matter which member is the representative, as long as if we ask for the representative twice without modifying the set, we get the same answer both times.

[We do not include notes for the proof of running time of the disjoint-set forest implementation, which is covered in Section 21.4.]

Operations

- MAKE-SET(x): make a new set $S_i = \{x\}$, and add S_i to $\mathcal{S}$.
- UNION(x, y): if $x \in S_x$, $y \in S_y$, then $\mathcal{S} = \mathcal{S} - S_x - S_y \cup \{S_x \cup S_y\}$.
  - Representative of new set is any member of $S_x \cup S_y$, often the representative of one of S_x and S_y.
  - Destroys S_x and S_y (since sets must be disjoint).
- FIND-SET(x): return representative of set containing x.

Analysis in terms of:

- n = # of elements = # of MAKE-SET operations,
- m = total # of operations.

Analysis

- Since MAKE-SET counts toward total # of operations, $m \geq n$.
- Can have at most $n - 1$ UNION operations, since after $n - 1$ UNIONs, only 1 set remains.
- Assume that the first n operations are MAKE-SET (helpful for analysis, usually not really necessary).

Application

Dynamic connected components.

For a graph $G = (V, E)$, vertices u, v are in same connected component if and only if there's a path between them.

- Connected components partition vertices into equivalence classes.

```
CONNECTED-COMPONENTS(G)
  for each vertex v ∈ G.V
      MAKE-SET(v)
  for each edge (u, v) ∈ G.E
      if FIND-SET(u) ≠ FIND-SET(v)
          UNION(u, v)
```

```
SAME-COMPONENT(u, v)
  if FIND-SET(u) == FIND-SET(v)
      return TRUE
  else return FALSE
```

Note

If actually implementing connected components,

- each vertex needs a handle to its object in the disjoint-set data structure,
- each object in the disjoint-set data structure needs a handle to its vertex.

Linked list representation

- Each set is a singly linked list, represented by an object with attributes
  - *head*: the first element in the list, assumed to be the set's representative, and
  - *tail*: the last element in the list.

  Objects may appear within the list in any order.
- Each object in the list has attributes for
  - the set member,
  - pointer to the set object, and
  - next.

MAKE-SET: create a singleton list.

FIND-SET: follow the pointer back to the list object, and then follow the *head* pointer to the representative.

UNION: a couple of ways to do it.

1. UNION(x, y): append y's list onto end of x's list. Use x's tail pointer to find the end.
   - Need to update the pointer back to the set object for every node on y's list.
   - If appending a large list onto a small list, it can take a while.

Operation	# objects updated
UNION(x_2, x_1)	1
UNION(x_3, x_2)	2
UNION(x_4, x_3)	3
UNION(x_5, x_4)	4
$\vdots$	$\vdots$
UNION(x_n, x_{n-1})	$n-1$
	$\Theta(n^2)$ total

   Amortized time per operation $= \Theta(n)$.

2. ***Weighted-union heuristic:*** Always append the smaller list to the larger list. (Break ties arbitrarily.)

   A single union can still take $\Omega(n)$ time, e.g., if both sets have $n/2$ members.

   Theorem
   With weighted union, a sequence of m operations on n elements takes $O(m + n \lg n)$ time.

   Sketch of proof Each MAKE-SET and FIND-SET still takes $O(1)$. How many times can each object's representative pointer be updated? It must be in the smaller set each time.

times updated	size of resulting set
1	≥ 2
2	≥ 4
3	≥ 8
$\vdots$	$\vdots$
k	$\geq 2^k$
$\vdots$	$\vdots$
$\lg n$	$\geq n$

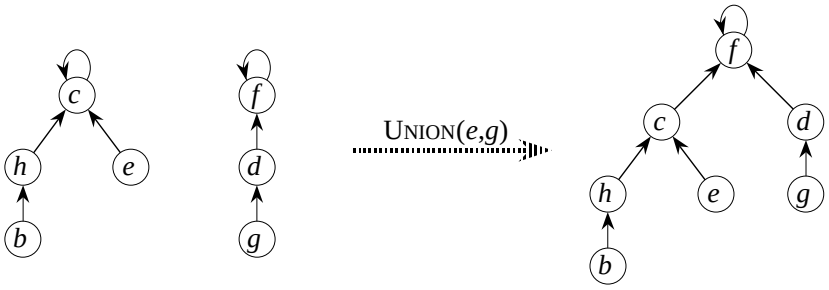
   Therefore, each representative is updated $\leq \lg n$ times. ■ (theorem)

   Seems pretty good, but we can do much better.

Disjoint-set forest

Forest of trees.

- 1 tree per set. Root is representative.
- Each node points only to its parent.

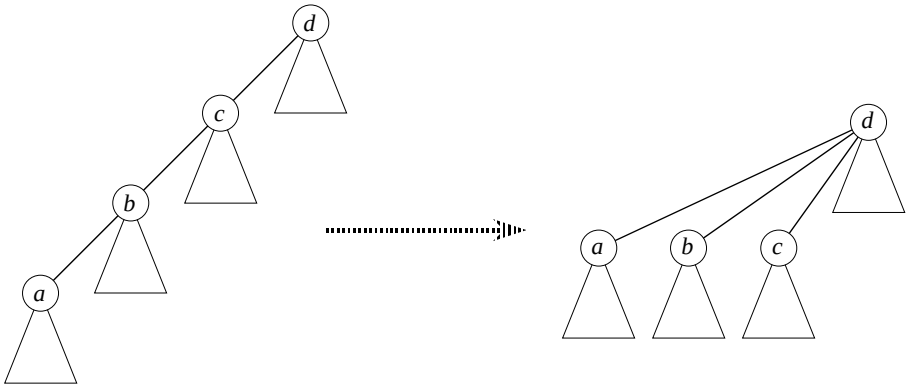


- MAKE-SET: make a single-node tree.
- UNION: make one root a child of the other.
- FIND-SET: follow pointers to the root.

Not so good—could get a linear chain of nodes.

Great heuristics

- ***Union by rank:*** make the root of the smaller tree (fewer nodes) a child of the root of the larger tree.
  - Don't actually use *size*.
  - Use *rank*, which is an upper bound on height of node.
  - Make the root with the smaller rank into a child of the root with the larger rank.
- ***Path compression: Find path*** = nodes visited during FIND-SET on the trip to the root. Make all nodes on the find path direct children of root.



Each node has two attributes, p (parent) and *rank*.

```
MAKE-SET(x)
  x.p = x
  x.rank = 0

UNION(x, y)
  LINK(FIND-SET(x), FIND-SET(y))

LINK(x, y)
  if x.rank > y.rank
      y.p = x
  else x.p = y
      // If equal ranks, choose y as parent and increment its rank.
      if x.rank == y.rank
          y.rank = y.rank + 1

FIND-SET(x)
  if x ≠ x.p
      x.p = FIND-SET(x.p)
  return x.p
```

FIND-SET makes a pass up to find the root, and a pass down as recursion unwinds to update each node on find path to point directly to root.

Running time

If use both union by rank and path compression, $O(m\,\alpha(n))$.

n	$\alpha(n)$
0–2	0
3	1
4–7	2
8–2047	3
2048–$A_4(1)$	4

What's $A_4(1)$? See Section 21.4, if you dare. It's $\gg 10^{80} \approx$ # of atoms in observable universe.

This bound is tight—there exists a sequence of operations that takes $\Omega(m\,\alpha(n))$ time.

Solutions for Chapter 21: Data Structures for Disjoint Sets

Solution to Exercise 21.2-3

This solution is also posted publicly

We want to show that we can assign $O(1)$ charges to MAKE-SET and FIND-SET and an $O(\lg n)$ charge to UNION such that the charges for a sequence of these operations are enough to cover the cost of the sequence—$O(m + n \lg n)$, according to the theorem. When talking about the charge for each kind of operation, it is helpful to also be able to talk about the number of each kind of operation.

Consider the usual sequence of m MAKE-SET, UNION, and FIND-SET operations, n of which are MAKE-SET operations, and let $l < n$ be the number of UNION operations. (Recall the discussion in Section 21.1 about there being at most $n - 1$ UNION operations.) Then there are n MAKE-SET operations, l UNION operations, and $m - n - l$ FIND-SET operations.

The theorem didn't separately name the number l of UNIONs; rather, it bounded the number by n. If you go through the proof of the theorem with l UNIONs, you get the time bound $O(m - l + l \lg l) = O(m + l \lg l)$ for the sequence of operations. That is, the actual time taken by the sequence of operations is at most $c(m + l \lg l)$, for some constant c.

Thus, we want to assign operation charges such that

$$\begin{array}{lcl} (\text{MAKE-SET charge}) & \cdot & n \\ +(\text{FIND-SET charge}) & \cdot & (m - n - l) \\ +(\text{UNION charge}) & \cdot & l \\ \hline \geq c(m + l \lg l)\ , & & \end{array}$$

so that the amortized costs give an upper bound on the actual costs.

The following assignments work, where c' is some constant $\geq c$:

- MAKE-SET: c'
- FIND-SET: c'
- UNION: $c'(\lg n + 1)$

Substituting into the above sum, we get

$$\begin{aligned} c'n + c'(m - n - l) + c'(\lg n + 1)l &= c'm + c'l \lg n \\ &= c'(m + l \lg n) \\ &> c(m + l \lg l)\ . \end{aligned}$$

Solution to Exercise 21.2-5

As the hint suggests, make the representative of each set be the tail of its linked list. Except for the tail element, each element's representative pointer points to the tail. The tail's representative pointer points to the head. An element is the tail if its next pointer is NIL. Now we can get to the tail in $O(1)$ time: if $x.next$ == NIL, then $tail = x$, else $tail = x.rep$. We can get to the head in $O(1)$ time as well: if $x.next$ == NIL, then $head = x.rep$, else $head = x.rep.rep$. The set object needs only to store a pointer to the tail, though a pointer to any list element would suffice.

Solution to Exercise 21.2-6

This solution is also posted publicly

Let's call the two lists A and B, and suppose that the representative of the new list will be the representative of A. Rather than appending B to the end of A, instead splice B into A right after the first element of A. We have to traverse B to update pointers to the set object anyway, so we can just make the last element of B point to the second element of A.

Solution to Exercise 21.3-3

You need to find a sequence of m operations on n elements that takes $\Omega(m \lg n)$ time. Start with n MAKE-SETs to create singleton sets $\{x_1\}, \{x_2\}, \ldots, \{x_n\}$. Next perform the $n-1$ UNION operations shown below to create a single set whose tree has depth $\lg n$.

$\text{UNION}(x_1, x_2)$	$n/2$ of these
$\text{UNION}(x_3, x_4)$	
$\text{UNION}(x_5, x_6)$	
$\vdots$	
$\text{UNION}(x_{n-1}, x_n)$	
$\text{UNION}(x_2, x_4)$	$n/4$ of these
$\text{UNION}(x_6, x_8)$	
$\text{UNION}(x_{10}, x_{12})$	
$\vdots$	
$\text{UNION}(x_{n-2}, x_n)$	
$\text{UNION}(x_4, x_8)$	$n/8$ of these
$\text{UNION}(x_{12}, x_{16})$	
$\text{UNION}(x_{20}, x_{24})$	
$\vdots$	
$\text{UNION}(x_{n-4}, x_n)$	
$\vdots$	
$\text{UNION}(x_{n/2}, x_n)$	1 of these

Finally, perform $m - 2n + 1$ FIND-SET operations on the deepest element in the tree. Each of these FIND-SET operations takes $\Omega(\lg n)$ time. Letting $m \geq 3n$, we have more than $m/3$ FIND-SET operations, so that the total cost is $\Omega(m \lg n)$.

Solution to Exercise 21.3-4

Maintain a circular, singly linked list of the nodes of each set. To print, just follow the list until you get back to node x, printing each member of the list. The only other operations that change are FIND-SET, which sets $x.next = x$, and LINK, which exchanges the pointers $x.next$ and $y.next$.

Solution to Exercise 21.3-5

With the path-compression heuristic, the sequence of m MAKE-SET, FIND-SET, and LINK operations, where all the LINK operations take place before any of the FIND-SET operations, runs in $O(m)$ time. The key observation is that once a node x appears on a find path, x will be either a root or a child of a root at all times thereafter.

We use the accounting method to obtain the $O(m)$ time bound. We charge a MAKE-SET operation two dollars. One dollar pays for the MAKE-SET, and one dollar remains on the node x that is created. The latter pays for the first time that x appears on a find path and is turned into a child of a root.

We charge one dollar for a LINK operation. This dollar pays for the actual linking of one node to another.

We charge one dollar for a FIND-SET. This dollar pays for visiting the root and its child, and for the path compression of these two nodes, during the FIND-SET. All other nodes on the find path use their stored dollar to pay for their visitation and path compression. As mentioned, after the FIND-SET, all nodes on the find path become children of a root (except for the root itself), and so whenever they are visited during a subsequent FIND-SET, the FIND-SET operation itself will pay for them.

Since we charge each operation either one or two dollars, a sequence of m operations is charged at most $2m$ dollars, and so the total time is $O(m)$.

Observe that nothing in the above argument requires union by rank. Therefore, we get an $O(m)$ time bound regardless of whether we use union by rank.

Solution to Exercise 21.4-4

Clearly, each MAKE-SET and LINK operation takes $O(1)$ time. Because the rank of a node is an upper bound on its height, each find path has length $O(\lg n)$, which in turn implies that each FIND-SET takes $O(\lg n)$ time. Thus, any sequence of m MAKE-SET, LINK, and FIND-SET operations on n elements takes $O(m \lg n)$ time. It is easy to prove an analogue of Lemma 21.7 to show that if we convert a sequence of m' MAKE-SET, UNION, and FIND-SET operations into a sequence of m MAKE-SET, LINK, and FIND-SET operations that take $O(m \lg n)$ time, then the sequence of m' MAKE-SET, UNION, and FIND-SET operations takes $O(m' \lg n)$ time.

Solution to Exercise 21.4-5

Professor Dante is mistaken. Take the following scenario. Let $n = 16$, and make 16 separate singleton sets using MAKE-SET. Then do 8 UNION operations to link the sets into 8 pairs, where each pair has a root with rank 0 and a child with rank 1. Now do 4 UNIONs to link pairs of these trees, so that there are 4 trees, each with a root of rank 2, children of the root of ranks 1 and 0, and a node of rank 0 that is the child of the rank-1 node. Now link pairs of these trees together, so that there are two resulting trees, each with a root of rank 3 and each containing a path from a leaf to the root with ranks 0, 1, and 3. Finally, link these two trees together, so that there is a path from a leaf to the root with ranks 0, 1, 3, and 4. Let x and y be the nodes on this path with ranks 1 and 3, respectively. Since $A_1(1) = 3$, $\text{level}(x) = 1$, and since $A_0(3) = 4$, $\text{level}(y) = 0$. Yet y follows x on the find path.

Solution to Exercise 21.4-6

First, $\alpha'(2^{2047} - 1) = \min\{k : A_k(1) \geq 2047\} = 3$, and $2^{2047} - 1 \gg 10^{80}$.

Second, we need that $0 \le \text{level}(x) \le \alpha'(n)$ for all nonroots x with $x.rank \ge 1$. With this definition of $\alpha'(n)$, we have $A_{\alpha'(n)}(x.rank) \ge A_{\alpha'(n)}(1) \ge \lg(n+1) > \lg n \ge x.p.rank$. The rest of the proof goes through with $\alpha'(n)$ replacing $\alpha(n)$.

Solution to Problem 21-1

a. For the input sequence

$$4, 8, \text{E}, 3, \text{E}, 9, 2, 6, \text{E}, \text{E}, \text{E}, 1, 7, \text{E}, 5\ ,$$

the values in the *extracted* array would be $4, 3, 2, 6, 8, 1$.

The following table shows the situation after the ith iteration of the **for** loop when we use OFF-LINE-MINIMUM on the same input. (For this input, $n = 9$ and m—the number of extractions—is 6).

i	K_1	K_2	K_3	K_4	K_5	K_6	K_7	*extracted* 1	2	3	4	5	6
0	{4, 8}	{3}	{9, 2, 6}	{}	{}	{1, 7}	{5}						
1	{4, 8}	{3}	{9, 2, 6}	{}	{}		{5, 1, 7}						1
2	{4, 8}	{3}		{9, 2, 6}	{}		{5, 1, 7}			2			1
3	{4, 8}			{9, 2, 6, 3}	{}		{5, 1, 7}		3	2			1
4				{9, 2, 6, 3, 4, 8}	{}		{5, 1, 7}	4	3	2			1
5				{9, 2, 6, 3, 4, 8}	{}		{5, 1, 7}	4	3	2			1
6					{9, 2, 6, 3, 4, 8}		{5, 1, 7}	4	3	2	6		1
7					{9, 2, 6, 3, 4, 8}		{5, 1, 7}	4	3	2	6		1
8							{5, 1, 7, 9, 2, 6, 3, 4, 8}	4	3	2	6	8	1

Because $j = m + 1$ in the iterations for $i = 5$ and $i = 7$, no changes occur in these iterations.

b. We want to show that the array *extracted* returned by OFF-LINE-MINIMUM is correct, meaning that for $i = 1, 2, \ldots, m$, $extracted[j]$ is the key returned by the jth EXTRACT-MIN call.

We start with n INSERT operations and m EXTRACT-MIN operations. The smallest of all the elements will be extracted in the first EXTRACT-MIN after its insertion. So we find j such that the minimum element is in K_j, and put the minimum element in $extracted[j]$, which corresponds to the EXTRACT-MIN after the minimum element insertion.

Now we reduce to a similar problem with $n - 1$ INSERT operations and $m - 1$ EXTRACT-MIN operations in the following way: the INSERT operations are the same but without the insertion of the smallest that was extracted, and the EXTRACT-MIN operations are the same but without the extraction that extracted the smallest element.

Conceptually, we unite I_j and I_{j+1}, removing the extraction between them and also removing the insertion of the minimum element from $\text{I}_j \cup \text{I}_{j+1}$. Uniting I_j and I_{j+1} is accomplished by line 6. We need to determine which set is K_l, rather than just using K_{j+1} unconditionally, because K_{j+1} may have been destroyed when it was united into a higher-indexed set by a previous execution of line 6.

Because we process extractions in increasing order of the minimum value found, the remaining iterations of the **for** loop correspond to solving the reduced problem.

There are two other points worth making. First, if the smallest remaining element had been inserted after the last EXTRACT-MIN (i.e., $j = m + 1$), then no changes occur, because this element is not extracted. Second, there may be smaller elements within the K_j sets than the the one we are currently looking for. These elements do not affect the result, because they correspond to elements that were already extracted, and their effect on the algorithm's execution is over.

c. To implement this algorithm, we place each element in a disjoint-set forest. Each root has a pointer to its K_i set, and each K_i set has a pointer to the root of the tree representing it. All the valid sets K_i are in a linked list.

Before OFF-LINE-MINIMUM, there is initialization that builds the initial sets K_i according to the I_i sequences.

- Line 2 ("determine j such that $i \in K_j$") turns into $j = \text{FIND-SET}(i)$.
- Line 5 ("let l be the smallest value greater than j for which set K_l exists") turns into $K_l = K_j.next$.
- Line 6 ("$K_l = K_j \cup K_l$, destroying K_j") turns into $l = \text{LINK}(j, l)$ and remove K_j from the linked list.

To analyze the running time, we note that there are n elements and that we have the following disjoint-set operations:

- n MAKE-SET operations
- at most $n - 1$ UNION operations before starting
- n FIND-SET operations
- at most n LINK operations

Thus the number m of overall operations is $O(n)$. The total running time is $O(m\,\alpha(n)) = O(n\,\alpha(n))$.

[The "tight bound" wording that this question uses does not refer to an "asymptotically tight" bound. Instead, the question is merely asking for a bound that is not too "loose."]

Solution to Problem 21-2

a. Denote the number of nodes by n, and let $n = (m + 1)/3$, so that $m = 3n - 1$. First, perform the n operations MAKE-TREE(v_1), MAKE-TREE(v_2), ..., MAKE-TREE(v_n). Then perform the sequence of $n - 1$ GRAFT operations GRAFT(v_1, v_2), GRAFT(v_2, v_3), ..., GRAFT(v_{n-1}, v_n); this sequence produces a single disjoint-set tree that is a linear chain of n nodes with v_n at the root and v_1 as the only leaf. Then perform FIND-DEPTH(v_1) repeatedly, n times. The total number of operations is $n + (n - 1) + n = 3n - 1 = m$.

Each MAKE-TREE and GRAFT operation takes $O(1)$ time. Each FIND-DEPTH operation has to follow an n-node find path, and so each of the n FIND-DEPTH operations takes $\Theta(n)$ time. The total time is $n \cdot \Theta(n) + (2n - 1) \cdot O(1) = \Theta(n^2) = \Theta(m^2)$.

b. MAKE-TREE is like MAKE-SET, except that it also sets the d value to 0:

```
MAKE-TREE(v)
  v.p = v
  v.rank = 0
  v.d = 0
```

It is correct to set $v.d$ to 0, because the depth of the node in the single-node disjoint-set tree is 0, and the sum of the depths on the find path for v consists only of $v.d$.

c. FIND-DEPTH will call a procedure FIND-ROOT:

```
FIND-ROOT(v)
  if v.p ≠ v.p.p
      y = v.p
      v.p = FIND-ROOT(y)
      v.d = v.d + y.d
  return v.p

FIND-DEPTH(v)
  FIND-ROOT(v)              // no need to save the return value
  if v == v.p
      return v.d
  else return v.d + v.p.d
```

FIND-ROOT performs path compression and updates pseudodistances along the find path from v. It is similar to FIND-SET on page 571, but with three changes. First, when v is either the root or a child of a root (one of these conditions holds if and only if $v.p = v.p.p$) in the disjoint-set forest, we don't have to recurse; instead, we just return $v.p$. Second, when we do recurse, we save the pointer $v.p$ into a new variable y. Third, when we recurse, we update $v.d$ by adding into it the d values of all nodes on the find path that are no longer proper ancestors of v after path compression; these nodes are precisely the proper ancestors of v other than the root. Thus, as long as v does not start out the FIND-ROOT call as either the root or a child of the root, we add $y.d$ into $v.d$. Note that $y.d$ has been updated prior to updating $v.d$, if y is also neither the root nor a child of the root.

FIND-DEPTH first calls FIND-ROOT to perform path compression and update pseudodistances. Afterward, the find path from v consists of either just v (if v is a root) or just v and $v.p$ (if v is not a root, in which case it is a child of the root after path compression). In the former case, the depth of v is just $v.d$, and in the latter case, the depth is $v.d + v.p.d$.

d. Our procedure for GRAFT is a combination of UNION and LINK:

```
GRAFT(r, ν)
  r' = FIND-ROOT(r)
  ν' = FIND-ROOT(ν)
  z = FIND-DEPTH(ν)
  if r'.rank > ν'.rank
      ν'.p = r'
      r'.d = r'.d + z + 1
      ν'.d = ν'.d − r'.d
  else r'.p = ν'
      r'.d = r'.d + z + 1 − ν'.d
      if r'.rank == ν'.rank
          ν'.rank = ν'.rank + 1
```

This procedure works as follows. First, we call FIND-ROOT on r and ν in order to find the roots r' and ν', respectively, of their trees in the disjoint-set forest. As we saw in part (c), these FIND-ROOT calls also perform path compression and update pseudodistances on the find paths from r and ν. We then call FIND-DEPTH(ν), saving the depth of ν in the variable z. (Since we have just compressed ν's find path, this call of FIND-DEPTH takes $O(1)$ time.) Next, we emulate the action of LINK, by making the root (r' or ν') of smaller rank a child of the root of larger rank; in case of a tie, we make r' a child of ν'.

If ν' has the smaller rank, then all nodes in r's tree will have their depths increased by the depth of ν plus 1 (because r is to become a child of ν). Altering the psuedodistance of the root of a disjoint-set tree changes the computed depth of all nodes in that tree, and so adding $z + 1$ to $r'.d$ accomplishes this update for all nodes in r's disjoint-set tree. Since ν' will become a child of r' in the disjoint-set forest, we have just increased the computed depth of all nodes in the disjoint-set tree rooted at ν' by $r'.d$. These computed depths should not have changed, however. Thus, we subtract off $r'.d$ from $\nu'.d$, so that the sum $\nu'.d + r'.d$ after making ν' a child of r' equals $\nu'.d$ before making ν' a child of r'.

On the other hand, if r' has the smaller rank, or if the ranks are equal, then r' becomes a child of ν' in the disjoint-set forest. In this case, ν' remains a root in the disjoint-set forest afterward, and we can leave $\nu'.d$ alone. We have to update $r'.d$, however, so that after making r' a child of ν', the depth of each node in r's disjoint-set tree is increased by $z + 1$. We add $z + 1$ to $r'.d$, but we also subtract out $\nu'.d$, since we have just made r' a child of ν'. Finally, if the ranks of r' and ν' are equal, we increment the rank of ν', as is done in the LINK procedure.

e. The asymptotic running times of MAKE-TREE, FIND-DEPTH, and GRAFT are equivalent to those of MAKE-SET, FIND-SET, and UNION, respectively. Thus, a sequence of m operations, n of which are MAKE-TREE operations, takes $\Theta(m\,\alpha(n))$ time in the worst case.

Lecture Notes for Chapter 22: Elementary Graph Algorithms

Graph representation

Given graph $G = (V, E)$. In pseudocode, represent vertex set by $G.V$ and edge set by $G.E$.

- G may be either directed or undirected.
- Two common ways to represent graphs for algorithms:
  1. Adjacency lists.
  2. Adjacency matrix.

When expressing the running time of an algorithm, it's often in terms of both $|V|$ and $|E|$. In asymptotic notation—and *only* in asymptotic notation—we'll drop the cardinality. Example: $O(V + E)$.

[The introduction to Part VI talks more about this.]

Adjacency lists

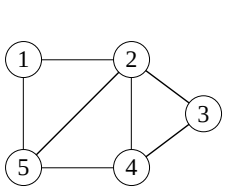
Array *Adj* of $|V|$ lists, one per vertex.

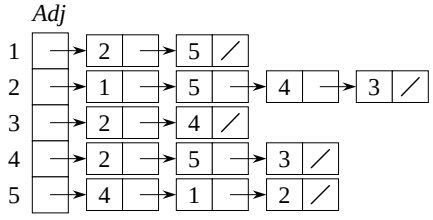
Vertex u's list has all vertices v such that $(u, v) \in E$. (Works for both directed and undirected graphs.)

In pseudocode, denote the array as attribute $G.Adj$, so will see notation such as $G.Adj[u]$.

Example

For an undirected graph:





If edges have *weights*, can put the weights in the lists.

Weight: $w : E \to \mathbb{R}$

We'll use weights later on for spanning trees and shortest paths.

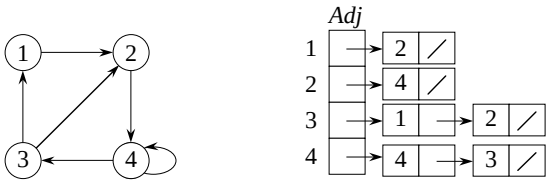
Space: $\Theta(V + E)$.

Time: to list all vertices adjacent to u: $\Theta(\text{degree}(u))$.

Time: to determine whether $(u, v) \in E$: $O(\text{degree}(u))$.

Example

For a directed graph:



Same asymptotic space and time.

Adjacency matrix

$|V| \times |V|$ matrix $A = (a_{ij})$

$$a_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E \ , \\ 0 & \text{otherwise .} \end{cases}$$

	1	2	3	4	5
1	0	1	0	0	1
2	1	0	1	1	1
3	0	1	0	1	0
4	0	1	1	0	1
5	1	1	0	1	0

	1	2	3	4
1	0	1	0	0
2	0	0	0	1
3	1	1	0	0
4	0	0	1	1

Space: $\Theta(V^2)$.

Time: to list all vertices adjacent to u: $\Theta(V)$.

Time: to determine whether $(u, v) \in E$: $\Theta(1)$.

Can store weights instead of bits for weighted graph.

We'll use both representations in these lecture notes.

Representing graph attributes

Graph algorithms usually need to maintain attributes for vertices and/or edges. Use the usual dot-notation: denote attribute d of vertex v by $v.d$.

Use the dot-notation for edges, too: denote attribute f of edge (u, v) by $(u, v).f$.

Implementing graph attributes

No one best way to implement. Depends on the programming language, the algorithm, and how the rest of the program interacts with the graph.

If representing the graph with adjacency lists, can represent vertex attributes in additional arrays that parallel the Adj array, e.g., $d[1..|V|]$, so that if vertices adjacent to u are in $Adj[u]$, store $u.d$ in array entry $d[u]$.

But can represent attributes in other ways. Example: represent vertex attributes as instance variables within a subclass of a `Vertex` class.

Breadth-first search

Input: Graph $G = (V, E)$, either directed or undirected, and ***source vertex*** $s \in V$.

Output: $v.d$ = distance (smallest # of edges) from s to v, for all $v \in V$.
In book, also $v.\pi$ such that (u, v) is last edge on shortest path $s \rightsquigarrow v$.

- u is v's ***predecessor***.
- set of edges $\{(v.\pi, v) : v \neq s\}$ forms a tree.

Later, we'll see a generalization of breadth-first search, with edge weights. For now, we'll keep it simple.

- Compute only $v.d$, not $v.\pi$. *[See book for $v.\pi$.]*
- Omitting colors of vertices. *[Used in book to reason about the algorithm. We'll skip them here.]*

Idea

Send a wave out from s.

- First hits all vertices 1 edge from s.
- From there, hits all vertices 2 edges from s.
- Etc.

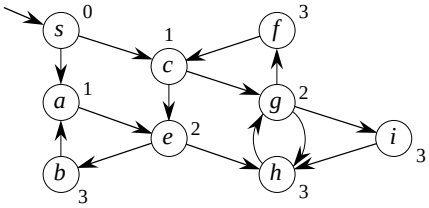
Use FIFO queue Q to maintain wavefront.

- $v \in Q$ if and only if wave has hit v but has not come out of v yet.

```
BFS(V, E, s)
  for each u ∈ V − {s}
      u.d = ∞
  s.d = 0
  Q = ∅
  ENQUEUE(Q, s)
  while Q ≠ ∅
      u = DEQUEUE(Q)
      for each v ∈ G.Adj[u]
          if v.d == ∞
              v.d = u.d + 1
              ENQUEUE(Q, v)
```

Example

directed graph *[undirected example in book]* .



Can show that Q consists of vertices with d values.

$i \quad i \quad i \quad \ldots \quad i \quad i+1 \quad i+1 \quad \ldots \quad i+1$

- Only 1 or 2 values.
- If 2, differ by 1 and all smallest are first.

Since each vertex gets a finite d value at most once, values assigned to vertices are monotonically increasing over time.

Actual proof of correctness is a bit trickier. See book.

BFS may not reach all vertices.

Time $= O(V + E)$.

- $O(V)$ because every vertex enqueued at most once.
- $O(E)$ because every vertex dequeued at most once and we examine (u, v) only when u is dequeued. Therefore, every edge examined at most once if directed, at most twice if undirected.

Depth-first search

Input: $G = (V, E)$, directed or undirected. No source vertex given!

Output: 2 ***timestamps*** on each vertex:

- $v.d =$ ***discovery time***
- $v.f =$ ***finishing time***

These will be useful for other algorithms later on.

Can also compute $v.\pi$. *[See book.]*

Will methodically explore *every* edge.

- Start over from different vertices as necessary.

As soon as we discover a vertex, explore from it.

- Unlike BFS, which puts a vertex on a queue so that we explore from it later.

As DFS progresses, every vertex has a ***color***:

- WHITE = undiscovered
- GRAY = discovered, but not finished (not done exploring from it)
- BLACK = finished (have found everything reachable from it)

Discovery and finishing times:

- Unique integers from 1 to $2|V|$.
- For all v, $v.d < v.f$.

In other words, $1 \le v.d < v.f \le 2|V|$.

Pseudocode

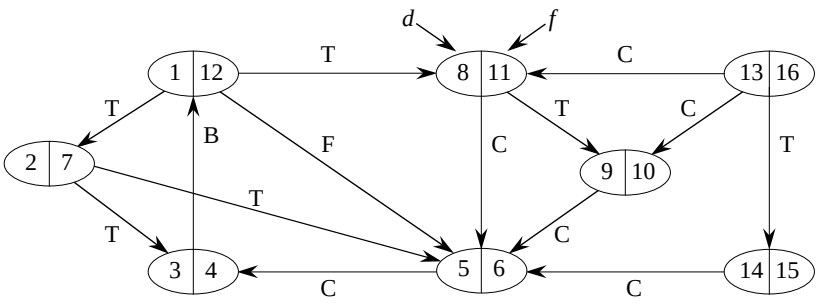
Uses a global timestamp *time*.

```
DFS(G)
  for each u ∈ G.V
      u.color = WHITE
  time = 0
  for each u ∈ G.V
      if u.color == WHITE
          DFS-VISIT(G, u)

DFS-VISIT(G, u)
  time = time + 1
  u.d = time
  u.color = GRAY              // discover u
  for each v ∈ G.Adj[u]       // explore (u, v)
      if v.color == WHITE
          DFS-VISIT(v)
  u.color = BLACK
  time = time + 1
  u.f = time                  // finish u
```

Example

[Go through this example, adding in the d and f values as they're computed. Show colors as they change. Don't put in the edge types yet.]



Time $= \Theta(V + E)$.

- Similar to BFS analysis.
- Θ, not just O, since guaranteed to examine every vertex and edge.

DFS forms a ***depth-first forest*** comprised of > 1 ***depth-first trees***. Each tree is made of edges (u, v) such that u is gray and v is white when (u, v) is explored.

Theorem (Parenthesis theorem)
[Proof omitted.]

For all u, v, exactly one of the following holds:

1. $u.d < u.f < v.d < v.f$ or $v.d < v.f < u.d < u.f$ (i.e., the intervals $[u.d, u.f]$ and $[v.d, v.f]$ are disjoint) and neither of u and v is a descendant of the other.
2. $u.d < v.d < v.f < u.f$ and v is a descendant of u.
3. $v.d < u.d < u.f < v.f$ and u is a descendant of v.

So $u.d < v.d < u.f < v.f$ *cannot* happen.

Like parentheses:

- OK: ( ) [ ] ([ ]) [( )]
- Not OK: ([ )] [( ])

Corollary
v is a proper descendant of u if and only if $u.d < v.d < v.f < u.f$.

Theorem (White-path theorem)
[Proof omitted.]

v is a descendant of u if and only if at time $u.d$, there is a path $u \rightsquigarrow v$ consisting of only white vertices. (Except for u, which was *just* colored gray.)

Classification of edges

- ***Tree edge:*** in the depth-first forest. Found by exploring (u, v).
- ***Back edge:*** (u, v), where u is a descendant of v.
- ***Forward edge:*** (u, v), where v is a descendant of u, but not a tree edge.
- ***Cross edge:*** any other edge. Can go between vertices in same depth-first tree or in different depth-first trees.

[Now label the example from above with edge types.]

In an undirected graph, there may be some ambiguity since (u, v) and (v, u) are the same edge. Classify by the first type above that matches.

Theorem
[Proof omitted.]

In DFS of an *un*directed graph, we get only tree and back edges. No forward or cross edges.

Topological sort

Directed acyclic graph (dag)

A directed graph with no cycles.

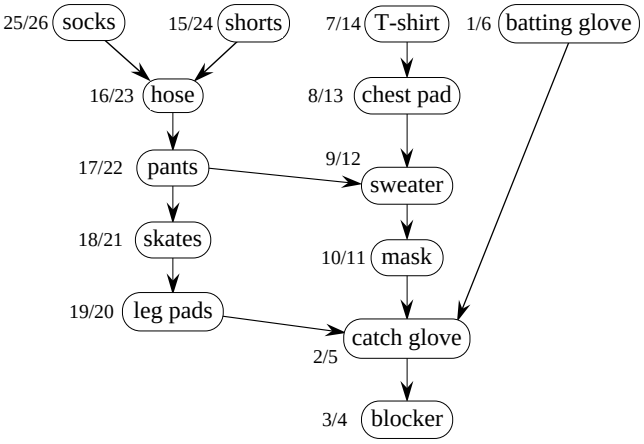
Good for modeling processes and structures that have a ***partial order:***

- $a > b$ and $b > c \Rightarrow a > c$.
- But may have a and b such that neither $a > b$ nor $b > c$.

Can always make a ***total order*** (either $a > b$ or $b > a$ for all $a \neq b$) from a partial order. In fact, that's what a topological sort will do.

Example

Dag of dependencies for putting on goalie equipment: *[Leave on board, but show without discovery and finishing times. Will put them in later.]*

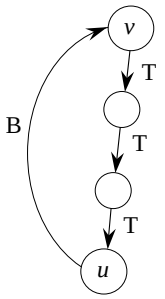


Lemma

A directed graph G is acyclic if and only if a DFS of G yields no back edges.

Proof $\Rightarrow$: Show that back edge $\Rightarrow$ cycle.

Suppose there is a back edge (u, v). Then v is ancestor of u in depth-first forest.



Therefore, there is a path $v \rightsquigarrow u$, so $v \rightsquigarrow u \rightarrow v$ is a cycle.

$\Leftarrow$: Show that cycle $\Rightarrow$ back edge.

Suppose G contains cycle c. Let v be the first vertex discovered in c, and let (u, v) be the preceding edge in c. At time $v.d$, vertices of c form a white path $v \rightsquigarrow u$ (since v is the first vertex discovered in c). By white-path theorem, u is descendant of v in depth-first forest. Therefore, (u, v) is a back edge. ■ (lemma)

Topological sort of a dag: a linear ordering of vertices such that if $(u, v) \in E$, then u appears somewhere before v. (Not like sorting numbers.)

TOPOLOGICAL-SORT(G)

call DFS(G) to compute finishing times $v.f$ for all $v \in G.V$
output vertices in order of *decreasing* finishing times

Don't need to sort by finishing times.

- Can just output vertices as they're finished and understand that we want the *reverse* of this list.
- Or put them onto the *front* of a linked list as they're finished. When done, the list contains vertices in topologically sorted order.

Time

$\Theta(V + E)$.

Do example. *[Now write discovery and finishing times in goalie equipment example.]*

Order:

26 socks
24 shorts
23 hose
22 pants
21 skates
20 leg pads
14 t-shirt
13 chest pad
12 sweater
11 mask
6 batting glove
5 catch glove
4 blocker

Correctness

Just need to show if $(u, v) \in E$, then $v.f < u.f$.
When we explore (u, v), what are the colors of u and v?

- u is gray.

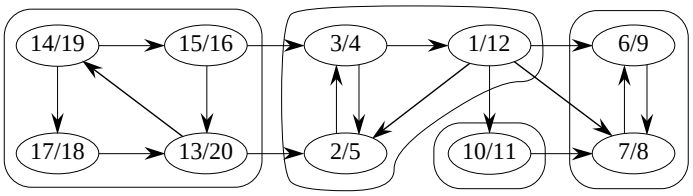
- Is v gray, too?
  - *No*, because then v would be ancestor of u.
    $\Rightarrow (u, v)$ is a back edge.
    $\Rightarrow$ contradiction of previous lemma (dag has no back edges).
- Is v white?
  - Then becomes descendant of u.
    By parenthesis theorem, $u.d < v.d < \underline{v.f < u.f}$.
- Is v black?
  - Then v is already finished.
    Since we're exploring (u, v), we have not yet finished u.
    Therefore, $v.f < u.f$. ■

Strongly connected components

Given directed graph $G = (V, E)$.
A ***strongly connected component*** (***SCC***) of G is a maximal set of vertices $C \subseteq V$ such that for all $u, v \in C$, both $u \rightsquigarrow v$ and $v \rightsquigarrow u$.

Example

[Just show SCC's at first. Do DFS a little later.]



Algorithm uses $G^{\mathrm{T}} =$ ***transpose*** of G.

- $G^{\mathrm{T}} = (V, E^{\mathrm{T}})$, $E^{\mathrm{T}} = \{(u, v) : (v, u) \in E\}$.
- G^{T} is G with all edges reversed.

Can create G^{T} in $\Theta(V + E)$ time if using adjacency lists.

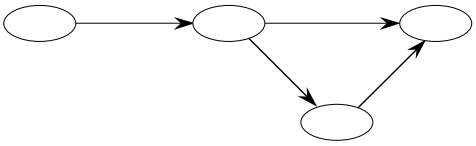
Observation

G and G^{T} have the *same* SCC's. (u and v are reachable from each other in G if and only if reachable from each other in G^{T}.)

Component graph

- $G^{\mathrm{SCC}} = (V^{\mathrm{SCC}}, E^{\mathrm{SCC}})$.
- V^{SCC} has one vertex for each SCC in G.
- E^{SCC} has an edge if there's an edge between the corresponding SCC's in G.

For our example:



Lemma

G^{SCC} is a dag. More formally, let C and C' be distinct SCC's in G, let $u, v \in C$, $u', v' \in C'$, and suppose there is a path $u \rightsquigarrow u'$ in G. Then there cannot also be a path $v' \rightsquigarrow v$ in G.

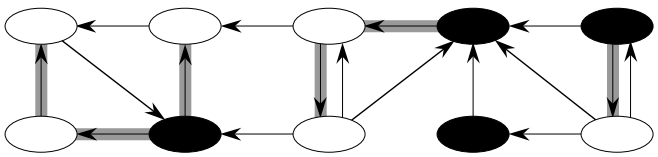
Proof Suppose there is a path $v' \rightsquigarrow v$ in G. Then there are paths $u \rightsquigarrow u' \rightsquigarrow v'$ and $v' \rightsquigarrow v \rightsquigarrow u$ in G. Therefore, u and v' are reachable from each other, so they are not in separate SCC's. ■ (lemma)

SCC(G)

call DFS(G) to compute finishing times $u.f$ for all u
compute G^T
call DFS(G^T), but in the main loop, consider vertices in order of decreasing $u.f$
(as computed in first DFS)
output the vertices in each tree of the depth-first forest formed in second DFS
as a separate SCC

Example:

1. Do DFS
2. G^T
3. DFS (roots blackened)



Time: $\Theta(V + E)$.

How can this possibly work?

Idea

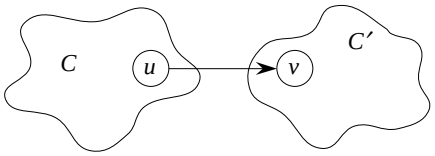
By considering vertices in second DFS in decreasing order of finishing times from first DFS, we are visiting vertices of the component graph in topological sort order.

To prove that it works, first deal with 2 notational issues:

- Will be discussing $u.d$ and $u.f$. These always refer to *first* DFS.
- Extend notation for d and f to sets of vertices $U \subseteq V$:
  - $d(U) = \min_{u \in U} \{u.d\}$ (earliest discovery time)
  - $f(U) = \max_{u \in U} \{u.f\}$ (latest finishing time)

Lemma

Let C and C' be distinct SCC's in $G = (V, E)$. Suppose there is an edge $(u, v) \in E$ such that $u \in C$ and $v \in C'$.



Then $f(C) > f(C')$.

Proof Two cases, depending on which SCC had the first discovered vertex during the first DFS.

- If $d(C) < d(C')$, let x be the first vertex discovered in C. At time $x.d$, all vertices in C and C' are white. Thus, there exist paths of white vertices from x to all vertices in C and C'.

  By the white-path theorem, all vertices in C and C' are descendants of x in depth-first tree.

  By the parenthesis theorem, $x.f = f(C) > f(C')$.
- If $d(C) > d(C')$, let y be the first vertex discovered in C'. At time $y.d$, all vertices in C' are white and there is a white path from y to each vertex in C' $\Rightarrow$ all vertices in C' become descendants of y. Again, $y.f = f(C')$.

  At time $y.d$, all vertices in C are white.

  By earlier lemma, since there is an edge (u, v), we cannot have a path from C' to C.

  So no vertex in C is reachable from y.

  Therefore, at time $y.f$, all vertices in C are still white.

  Therefore, for all $w \in C$, $w.f > y.f$, which implies that $f(C) > f(C')$.

  ■ (lemma)

Corollary

Let C and C' be distinct SCC's in $G = (V, E)$. Suppose there is an edge $(u, v) \in E^{\mathrm{T}}$, where $u \in C$ and $v \in C'$. Then $f(C) < f(C')$.

Proof $(u, v) \in E^{\mathrm{T}} \Rightarrow (v, u) \in E$. Since SCC's of G and G^{T} are the same, $f(C') > f(C)$. ■ (corollary)

Corollary

Let C and C' be distinct SCC's in $G = (V, E)$, and suppose that $f(C) > f(C')$. Then there cannot be an edge from C to C' in G^{T}.

Proof It's the contrapositive of the previous corollary. ■

Now we have the intuition to understand why the SCC procedure works.

When we do the second DFS, on G^{T}, start with SCC C such that $f(C)$ is maximum. The second DFS starts from some $x \in C$, and it visits all vertices in C.

Corollary says that since $f(C) > f(C')$ for all $C' \neq C$, there are no edges from C to C' in G^{T}.

Therefore, DFS will visit *only* vertices in C.

Which means that the depth-first tree rooted at x contains *exactly* the vertices of C.

The next root chosen in the second DFS is in SCC C' such that $f(C')$ is maximum over all SCC's other than C. DFS visits all vertices in C', but the only edges out of C' go to C, *which we've already visited*.

Therefore, the only tree edges will be to vertices in C'.

We can continue the process.

Each time we choose a root for the second DFS, it can reach only

- vertices in its SCC—get tree edges to these,
- vertices in SCC's *already visited* in second DFS—get *no* tree edges to these.

We are visiting vertices of $(G^{\mathrm{T}})^{\mathrm{SCC}}$ in reverse of topologically sorted order.

[The book has a formal proof.]

Solutions for Chapter 22: Elementary Graph Algorithms

Solution to Exercise 22.1-6

We start by observing that if $a_{ij} = 1$, so that $(i, j) \in E$, then vertex i cannot be a universal sink, for it has an outgoing edge. Thus, if row i contains a 1, then vertex i cannot be a universal sink. This observation also means that if there is a self-loop (i, i), then vertex i is not a universal sink. Now suppose that $a_{ij} = 0$, so that $(i, j) \notin E$, and also that $i \neq j$. Then vertex j cannot be a universal sink, for either its in-degree must be strictly less than $|V| - 1$ or it has a self-loop. Thus if column j contains a 0 in any position other than the diagonal entry (j, j), then vertex j cannot be a universal sink.

Using the above observations, the following procedure returns TRUE if vertex k is a universal sink, and FALSE otherwise. It takes as input a $|V| \times |V|$ adjacency matrix $A = (a_{ij})$.

```
IS-SINK(A, k)
  let A be |V| × |V|
  for j = 1 to |V|              // check for a 1 in row k
      if a_kj == 1
          return FALSE
  for i = 1 to |V|              // check for an off-diagonal 0 in column k
      if a_ik == 0 and i ≠ k
          return FALSE
  return TRUE
```

Because this procedure runs in $O(V)$ time, we may call it only $O(1)$ times in order to achieve our $O(V)$-time bound for determining whether directed graph G contains a universal sink.

Observe also that a directed graph can have at most one universal sink. This property holds because if vertex j is a universal sink, then we would have $(i, j) \in E$ for all $i \neq j$ and so no other vertex i could be a universal sink.

The following procedure takes an adjacency matrix A as input and returns either a message that there is no universal sink or a message containing the identity of the universal sink. It works by eliminating all but one vertex as a potential universal sink and then checking the remaining candidate vertex by a single call to IS-SINK.

```
UNIVERSAL-SINK(A)
  let A be |V| × |V|
  i = j = 1
  while i ≤ |V| and j ≤ |V|
      if a_ij == 1
          i = i + 1
      else j = j + 1
  if i > |V|
      return "there is no universal sink"
  elseif IS-SINK(A, i) == FALSE
      return "there is no universal sink"
  else return i "is a universal sink"
```

UNIVERSAL-SINK walks through the adjacency matrix, starting at the upper left corner and always moving either right or down by one position, depending on whether the current entry a_{ij} it is examining is 0 or 1. It stops once either i or j exceeds $|V|$.

To understand why UNIVERSAL-SINK works, we need to show that after the **while** loop terminates, the only vertex that might be a universal sink is vertex i. The call to IS-SINK then determines whether vertex i is indeed a universal sink.

Let us fix i and j to be values of these variables at the termination of the **while** loop. We claim that every vertex k such that $1 \le k < i$ cannot be a universal sink. That is because the way that i achieved its final value at loop termination was by finding a 1 in each row k for which $1 \le k < i$. As we observed above, any vertex k whose row contains a 1 cannot be a universal sink.

If $i > |V|$ at loop termination, then we have eliminated all vertices from consideration, and so there is no universal sink. If, on the other hand, $i \le |V|$ at loop termination, we need to show that every vertex k such that $i < k \le |V|$ cannot be a universal sink. If $i \le |V|$ at loop termination, then the **while** loop terminated because $j > |V|$. That means that we found a 0 in every column. Recall our earlier observation that if column k contains a 0 in an off-diagonal position, then vertex k cannot be a universal sink. Since we found a 0 in every column, we found a 0 in every column k such that $i < k \le |V|$. Moreover, we never examined any matrix entries in rows greater than i, and so we never examined the diagonal entry in any column k such that $i < k \le |V|$. Therefore, all the 0s that we found in columns k such that $i < k \le |V|$ were off-diagonal. We conclude that every vertex k such that $i < k \le |V|$ cannot be a universal sink.

Thus, we have shown that every vertex less than i and every vertex greater than i cannot be a universal sink. The only remaining possibility is that vertex i might be a universal sink, and the call to IS-SINK checks whether it is.

To see that UNIVERSAL-SINK runs in $O(V)$ time, observe that either i or j is incremented in each iteration of the **while** loop. Thus, the **while** loop makes at most $2|V| - 1$ iterations. Each iteration takes $O(1)$ time, for a total **while** loop time of $O(V)$ and, combined with the $O(V)$-time call to IS-SINK, we get a total running time of $O(V)$.

Solution to Exercise 22.1-7

This solution is also posted publicly

$$BB^T(i,j) = \sum_{e \in E} b_{ie} b_{ej}^T = \sum_{e \in E} b_{ie} b_{je}$$

- If $i = j$, then $b_{ie} b_{je} = 1$ (it is $1 \cdot 1$ or $(-1) \cdot (-1)$) whenever e enters or leaves vertex i, and 0 otherwise.
- If $i \neq j$, then $b_{ie} b_{je} = -1$ when $e = (i, j)$ or $e = (j, i)$, and 0 otherwise.

Thus,

$$BB^T(i,j) = \begin{cases} \text{degree of } i = \text{in-degree} + \text{out-degree} & \text{if } i = j \ , \\ -(\text{\# of edges connecting } i \text{ and } j) & \text{if } i \neq j \ . \end{cases}$$

Solution to Exercise 22.2-3

Note: This exercise changed in the third printing. This solution reflects the change.

The BFS procedure cares only whether a vertex is white or not. A vertex v must become non-white at the same time that $v.d$ is assigned a finite value so that we do not attempt to assign to $v.d$ again, and so we need to change vertex colors in lines 5 and 14. Once we have changed a vertex's color to non-white, we do not need to change it again.

Solution to Exercise 22.2-5

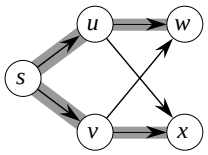
This solution is also posted publicly

The correctness proof for the BFS algorithm shows that $u.d = \delta(s, u)$, and the algorithm doesn't assume that the adjacency lists are in any particular order.

In Figure 22.3, if t precedes x in $Adj[w]$, we can get the breadth-first tree shown in the figure. But if x precedes t in $Adj[w]$ and u precedes y in $Adj[x]$, we can get edge (x, u) in the breadth-first tree.

Solution to Exercise 22.2-6

The edges in E_π are shaded in the following graph:



To see that E_π cannot be a breadth-first tree, let's suppose that $Adj[s]$ contains u before v. BFS adds edges (s,u) and (s,v) to the breadth-first tree. Since u is enqueued before v, BFS then adds edges (u,w) and (u,x). (The order of w and x in $Adj[u]$ doesn't matter.) Symmetrically, if $Adj[s]$ contains v before u, then BFS adds edges (s,v) and (s,u) to the breadth-first tree, v is enqueued before u, and BFS adds edges (v,w) and (v,x). (Again, the order of w and x in $Adj[v]$ doesn't matter.) BFS will never put both edges (u,w) and (v,x) into the breadth-first tree. In fact, it will also never put both edges (u,x) and (v,w) into the breadth-first tree.

Solution to Exercise 22.2-7

Create a graph G where each vertex represents a wrestler and each edge represents a rivalry. The graph will contain n vertices and r edges.

Perform as many BFS's as needed to visit all vertices. Assign all wrestlers whose distance is even to be babyfaces and all wrestlers whose distance is odd to be heels. Then check each edge to verify that it goes between a babyface and a heel. This solution would take $O(n+r)$ time for the BFS, $O(n)$ time to designate each wrestler as a babyface or heel, and $O(r)$ time to check edges, which is $O(n+r)$ time overall.

Solution to Exercise 22.3-4

Note: This exercise changed in the third printing. This solution reflects the change.

The DFS and DFS-VISIT procedures care only whether a vertex is white or not. By coloring vertex u gray when it is first visited, in line 3 of DFS-VISIT, we ensure that u will not be visited again. Once we have changed a vertex's color to non-white, we do not need to change it again.

Solution to Exercise 22.3-5

a. Edge (u,v) is a tree edge or forward edge if and only if v is a descendant of u in the depth-first forest. (If (u,v) is a back edge, then u is a descendant of v, and if (u,v) is a cross edge, then neither of u or v is a descendant of the other.) By Corollary 22.8, therefore, (u,v) is a tree edge or forward edge if and only if $u.d < v.d < v.f < u.f$.

b. First, suppose that (u,v) is a back edge. A self-loop is by definition a back edge. If (u,v) is a self-loop, then clearly $v.d = u.d < u.f = v.f$. If (u,v) is not a self-loop, then u is a descendant of v in the depth-first forest, and by Corollary 22.8, $v.d < u.d < u.f < v.f$.

Now, suppose that $v.d \le u.d < u.f \le v.f$. If u and v are the same vertex, then $v.d = u.d < u.f = v.f$, and (u,v) is a self-loop and hence a back edge. If u

and v are distinct, then $v.d < u.d < u.f < v.f$. By the parenthesis theorem, interval $[u.d, u.f]$ is contained entirely within the interval $[v.d, v.f]$, and u is a descendant of v in a depth-first tree. Thus, (u, v) is a back edge.

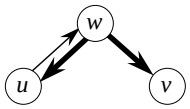
c. First, suppose that (u, v) is a cross edge. Since neither u nor v is an ancestor of the other, the parenthesis theorem says that the intervals $[u.d, u.f]$ and $[v.d, v.f]$ are entirely disjoint. Thus, we must have either $u.d < u.f < v.d < v.f$ or $v.d < v.f < u.d < u.f$. We claim that we cannot have $u.d < v.d$ if (u, v) is a cross edge. Why? If $u.d < v.d$, then v is white at time $u.d$. By the white-path theorem, v is a descendant of u, which contradicts (u, v) being a cross edge. Thus, we must have $v.d < v.f < u.d < u.f$.

Now suppose that $v.d < v.f < u.d < u.f$. By the parenthesis theorem, neither u nor v is a descendant of the other, which means that (u, v) must be a cross edge.

Solution to Exercise 22.3-8

Let us consider the example graph and depth-first search below.

	d	f
w	1	6
u	2	3
v	4	5

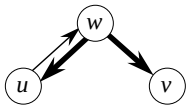


Clearly, there is a path from u to v in G. The bold edges are in the depth-first forest produced. We can see that $u.d < v.d$ in the depth-first search but v is not a descendant of u in the forest.

Solution to Exercise 22.3-9

Let us consider the example graph and depth-first search below.

	d	f
w	1	6
u	2	3
v	4	5

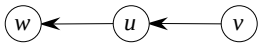


Clearly, there is a path from u to v in G. The bold edges of G are in the depth-first forest produced by the search. However, $v.d > u.f$ and the conjecture is false.

Solution to Exercise 22.3-11

Let us consider the example graph and depth-first search below.

	d	f
w	1	2
u	3	4
v	5	6



Clearly u has both incoming and outgoing edges in G but a depth-first search of G produced a depth-first forest where u is in a tree by itself.

Solution to Exercise 22.3-12

This solution is also posted publicly

The following pseudocode modifies the DFS and DFS-VISIT procedures to assign values to the *cc* attributes of vertices.

```
DFS(G)
  for each vertex u ∈ G.V
      u.color = WHITE
      u.π = NIL
  time = 0
  counter = 0
  for each vertex u ∈ G.V
      if u.color == WHITE
          counter = counter + 1
          DFS-VISIT(G, u, counter)

DFS-VISIT(G, u, counter)
  u.cc = counter              // label the vertex
  time = time + 1
  u.d = time
  u.color = GRAY
  for each v ∈ G.Adj[u]
      if v.color == WHITE
          v.π = u
          DFS-VISIT(G, v, counter)
  u.color = BLACK
  time = time + 1
  u.f = time
```

This DFS increments a counter each time DFS-VISIT is called to grow a new tree in the DFS forest. Every vertex visited (and added to the tree) by DFS-VISIT is labeled with that same counter value. Thus $u.cc = v.cc$ if and only if u and v are visited in the same call to DFS-VISIT from DFS, and the final value of the counter is the number of calls that were made to DFS-VISIT by DFS. Also, since every vertex is visited eventually, every vertex is labeled.

Thus all we need to show is that the vertices visited by each call to DFS-VISIT from DFS are exactly the vertices in one connected component of G.

- All vertices in a connected component are visited by one call to DFS-VISIT from DFS:

  Let u be the first vertex in component C visited by DFS-VISIT. Since a vertex becomes non-white only when it is visited, all vertices in C are white when DFS-VISIT is called for u. Thus, by the white-path theorem, all vertices in C become descendants of u in the forest, which means that all vertices in C are visited (by recursive calls to DFS-VISIT) before DFS-VISIT returns to DFS.
- All vertices visited by one call to DFS-VISIT from DFS are in the same connected component:

  If two vertices are visited in the same call to DFS-VISIT from DFS, they are in the same connected component, because vertices are visited only by following paths in G (by following edges found in adjacency lists, starting from some vertex).

Solution to Exercise 22.4-3

This solution is also posted publicly

An undirected graph is acyclic (i.e., a forest) if and only if a DFS yields no back edges.

- If there's a back edge, there's a cycle.
- If there's no back edge, then by Theorem 22.10, there are only tree edges. Hence, the graph is acyclic.

Thus, we can run DFS: if we find a back edge, there's a cycle.

- Time: $O(V)$. (Not $O(V + E)$!)
  If we ever see $|V|$ distinct edges, we must have seen a back edge because (by Theorem B.2 on p. 1174) in an acyclic (undirected) forest, $|E| \leq |V| - 1$.

Solution to Exercise 22.4-5

```
TOPOLOGICAL-SORT(G)
  // Initialize in-degree, Θ(V) time.
  for each vertex u ∈ G.V
      u.in-degree = 0
  // Compute in-degree, Θ(V + E) time.
  for each vertex u ∈ G.V
      for each v ∈ G.Adj[u]
          v.in-degree = v.in-degree + 1
  // Initialize Queue, Θ(V) time.
  Q = ∅
  for each vertex u ∈ G.V
      if u.in-degree == 0
          ENQUEUE(Q, u)
  // while loop takes O(V + E) time.
  while Q ≠ ∅
      u = DEQUEUE(Q)
      output u
      // for loop executes O(E) times total.
      for each v ∈ G.Adj[u]
          v.in-degree = v.in-degree − 1
          if v.in-degree == 0
              ENQUEUE(Q, v)
  // Check for cycles, O(V) time.
  for each vertex u ∈ G.V
      if u.in-degree ≠ 0
          report that there's a cycle
  // Another way to check for cycles would be to count the vertices
  // that are output and report a cycle if that number is < |V|.
```

To find and output vertices of in-degree 0, we first compute all vertices' in-degrees by making a pass through all the edges (by scanning the adjacency lists of all the vertices) and incrementing the in-degree of each vertex an edge enters.

- Computing all in-degrees takes $\Theta(V + E)$ time ($|V|$ adjacency lists accessed, $|E|$ edges total found in those lists, $\Theta(1)$ work for each edge).

We keep the vertices with in-degree 0 in a FIFO queue, so that they can be enqueued and dequeued in $O(1)$ time. (The order in which vertices in the queue are processed doesn't matter, so any kind of FIFO queue works.)

- Initializing the queue takes one pass over the vertices doing $\Theta(1)$ work, for total time $\Theta(V)$.

As we process each vertex from the queue, we effectively remove its outgoing edges from the graph by decrementing the in-degree of each vertex one of those edges enters, and we enqueue any vertex whose in-degree goes to 0. We do not need to actually remove the edges from the adjacency list, because that adjacency list

will never be processed again by the algorithm: Each vertex is enqueued/dequeued at most once because it is enqueued only if it starts out with in-degree 0 or if its in-degree becomes 0 after being decremented (and never incremented) some number of times.

- The processing of a vertex from the queue happens $O(V)$ times because no vertex can be enqueued more than once. The per-vertex work (dequeue and output) takes $O(1)$ time, for a total of $O(V)$ time.
- Because the adjacency list of each vertex is scanned only when the vertex is dequeued, the adjacency list of each vertex is scanned at most once. Since the sum of the lengths of all the adjacency lists is $\Theta(E)$, at most $O(E)$ time is spent in total scanning adjacency lists. For each edge in an adjacency list, $\Theta(1)$ work is done, for a total of $O(E)$ time.

Thus the total time taken by the algorithm is $O(V + E)$.

The algorithm outputs vertices in the right order (u before v for every edge (u, v)) because v will not be output until its in-degree becomes 0, which happens only when every edge (u, v) leading into v has been "removed" due to the processing (including output) of u.

If there are no cycles, all vertices are output.

- Proof: Assume that some vertex v_0 is not output. Vertex v_0 cannot start out with in-degree 0 (or it would be output), so there are edges into v_0. Since v_0's in-degree never becomes 0, at least one edge (v_1, v_0) is never removed, which means that at least one other vertex v_1 was not output. Similarly, v_1 not output means that some vertex v_2 such that $(v_2, v_1) \in E$ was not output, and so on. Since the number of vertices is finite, this path $(\cdots \to v_2 \to v_1 \to v_0)$ is finite, so we must have $v_i = v_j$ for some i and j in this sequence, which means there is a cycle.

If there are cycles, not all vertices will be output, because some in-degrees never become 0.

- Proof: Assume that a vertex in a cycle is output (its in-degree becomes 0). Let v be the first vertex in its cycle to be output, and let u be v's predecessor in the cycle. In order for v's in-degree to become 0, the edge (u, v) must have been "removed," which happens only when u is processed. But this cannot have happened, because v is the first vertex in its cycle to be processed. Thus no vertices in cycles are output.

Solution to Exercise 22.5-5

We have at our disposal an $O(V + E)$-time algorithm that computes strongly connected components. Let us assume that the output of this algorithm is a mapping $u.scc$, giving the number of the strongly connected component containing vertex u, for each vertex u. Without loss of generality, assume that $u.scc$ is an integer in the set $\{1, 2, \ldots, |V|\}$.

Construct the multiset (a set that can contain the same object more than once) $T = \{u.scc : u \in V\}$, and sort it by using counting sort. Since the values we are sorting are integers in the range 1 to $|V|$, the time to sort is $O(V)$. Go through the sorted multiset T and every time we find an element x that is distinct from the one before it, add x to V^{SCC}. (Consider the first element of the sorted set as "distinct from the one before it.") It takes $O(V)$ time to construct V^{SCC}.

Construct the set of ordered pairs

$$S = \{(x, y) : \text{there is an edge } (u, v) \in E, x = u.scc, \text{ and } y = v.scc\} .$$

We can easily construct this set in $\Theta(E)$ time by going through all edges in E and looking up $u.scc$ and $v.scc$ for each edge $(u, v) \in E$.

Having constructed S, remove all elements of the form (x, x). Alternatively, when we construct S, do not put an element in S when we find an edge (u, v) for which $u.scc = v.scc$. S now has at most $|E|$ elements.

Now sort the elements of S using radix sort. Sort on one component at a time. The order does not matter. In other words, we are performing two passes of counting sort. The time to do so is $O(V + E)$, since the values we are sorting on are integers in the range 1 to $|V|$.

Finally, go through the sorted set S, and every time we find an element (x, y) that is distinct from the element before it (again considering the first element of the sorted set as distinct from the one before it), add (x, y) to E^{SCC}. Sorting and then adding (x, y) only if it is distinct from the element before it ensures that we add (x, y) at most once. It takes $O(E)$ time to go through S in this way, once S has been sorted.

The total time is $O(V + E)$.

Solution to Exercise 22.5-6

The basic idea is to replace the edges within each SCC by one simple, directed cycle and then remove redundant edges between SCC's. Since there must be at least k edges within an SCC that has k vertices, a single directed cycle of k edges gives the k-vertex SCC with the fewest possible edges.

The algorithm works as follows:

1. Identify all SCC's of G. Time: $\Theta(V + E)$, using the SCC algorithm in Section 22.5.
2. Form the component graph G^{SCC}. Time: $O(V + E)$, by Exercise 22.5-5.
3. Start with $E' = \emptyset$. Time: $O(1)$.
4. For each SCC of G, let the vertices in the SCC be $v_1, v_2, \ldots, v_k$, and add to E' the directed edges $(v_1, v_2), (v_2, v_3), \ldots, (v_{k-1}, v_k), (v_k, v_1)$. These edges form a simple, directed cycle that includes all vertices of the SCC. Time for all SCC's: $O(V)$.
5. For each edge (u, v) in the component graph G^{SCC}, select any vertex x in u's SCC and any vertex y in v's SCC, and add the directed edge (x, y) to E'. Time: $O(E)$.

Thus, the total time is $\Theta(V + E)$.

Solution to Exercise 22.5-7

To determine whether $G = (V, E)$ is semiconnected, do the following:

1. Call STRONGLY-CONNECTED-COMPONENTS.
2. Form the component graph. (By Exercise 22.5-5, you may assume that this takes $O(V + E)$ time.)
3. Topologically sort the component graph. (Recall that it's a dag.) Assuming that G contains k SCC's, the topological sort gives a linear ordering $\langle v_1, v_2, \ldots, v_k \rangle$ of the vertices.
4. Verify that the sequence of vertices $\langle v_1, v_2, \ldots, v_k \rangle$ given by topological sort forms a linear chain in the component graph. That is, verify that the edges $(v_1, v_2), (v_2, v_3), \ldots, (v_{k-1}, v_k)$ exist in the component graph. If the vertices form a linear chain, then the original graph is semiconnected; otherwise it is not.

Because we know that all vertices in each SCC are mutually reachable from each other, it suffices to show that the component graph is semiconnected if and only if it contains a linear chain. We must also show that if there's a linear chain in the component graph, it's the one returned by topological sort.

We'll first show that if there's a linear chain in the component graph, then it's the one returned by topological sort. In fact, this is trivial. A topological sort has to respect every edge in the graph. So if there's a linear chain, a topological sort *must* give us the vertices in order.

Now we'll show that the component graph is semiconnected if and only if it contains a linear chain.

First, suppose that the component graph contains a linear chain. Then for every pair of vertices u, v in the component graph, there is a path between them. If u precedes v in the linear chain, then there's a path $u \rightsquigarrow v$. Otherwise, v precedes u, and there's a path $v \rightsquigarrow u$.

Conversely, suppose that the component graph does not contain a linear chain. Then in the list returned by topological sort, there are two consecutive vertices v_i and v_{i+1}, but the edge (v_i, v_{i+1}) is not in the component graph. Any edges out of v_i are to vertices v_j, where $j > i + 1$, and so there is no path from v_i to v_{i+1} in the component graph. And since v_{i+1} follows v_i in the topological sort, there cannot be any paths at all from v_{i+1} to v_i. Thus, the component graph is not semiconnected.

Running time of each step:

1. $\Theta(V + E)$.
2. $O(V + E)$.
3. Since the component graph has at most $|V|$ vertices and at most $|E|$ edges, $O(V + E)$.
4. Also $O(V + E)$. We just check the adjacency list of each vertex v_i in the component graph to verify that there's an edge (v_i, v_{i+1}). We'll go through each adjacency list once.

Thus, the total running time is $\Theta(V + E)$.

Solution to Problem 22-1

This solution is also posted publicly

a. 1. Suppose (u, v) is a back edge or a forward edge in a BFS of an undirected graph. Then one of u and v, say u, is a proper ancestor of the other (v) in the breadth-first tree. Since we explore all edges of u before exploring any edges of any of u's descendants, we must explore the edge (u, v) at the time we explore u. But then (u, v) must be a tree edge.

2. In BFS, an edge (u, v) is a tree edge when we set $v.\pi = u$. But we only do so when we set $v.d = u.d + 1$. Since neither $u.d$ nor $v.d$ ever changes thereafter, we have $v.d = u.d + 1$ when BFS completes.

3. Consider a cross edge (u, v) where, without loss of generality, u is visited before v. At the time we visit u, vertex v must already be on the queue, for otherwise (u, v) would be a tree edge. Because v is on the queue, we have $v.d \leq u.d + 1$ by Lemma 22.3. By Corollary 22.4, we have $v.d \geq u.d$. Thus, either $v.d = u.d$ or $v.d = u.d + 1$.

b. 1. Suppose (u, v) is a forward edge. Then we would have explored it while visiting u, and it would have been a tree edge.

2. Same as for undirected graphs.

3. For any edge (u, v), whether or not it's a cross edge, we cannot have $v.d > u.d + 1$, since we visit v at the latest when we explore edge (u, v). Thus, $v.d \leq u.d + 1$.

4. Clearly, $v.d \geq 0$ for all vertices v. For a back edge (u, v), v is an ancestor of u in the breadth-first tree, which means that $v.d \leq u.d$. (Note that since self-loops are considered to be back edges, we could have $u = v$.)

Solution to Problem 22-3

a. An Euler tour is a single cycle that traverses each edge of G exactly once, but it might not be a simple cycle. An Euler tour can be decomposed into a set of edge-disjoint simple cycles, however.

If G has an Euler tour, therefore, we can look at the simple cycles that, together, form the tour. In each simple cycle, each vertex in the cycle has one entering edge and one leaving edge. In each simple cycle, therefore, each vertex v has in-degree(v) = out-degree(v), where the degrees are either 1 (if v is on the simple cycle) or 0 (if v is not on the simple cycle). Adding the in- and out-degrees over all edges proves that if G has an Euler tour, then in-degree(v) = out-degree(v) for all vertices v.

We prove the converse—that if in-degree(v) = out-degree(v) for all vertices v, then G has an Euler tour—in two different ways. One proof is nonconstructive, and the other proof will help us design the algorithm for part (b).

First, we claim that if in-degree(v) = out-degree(v) for all vertices v, then we can pick any vertex u for which in-degree(u) = out-degree$(u) \geq 1$ and create

a cycle (not necessarily simple) that contains u. To prove this claim, let us start by placing vertex u on the cycle, and choose any leaving edge of u, say (u, v). Now we put v on the cycle. Since in-degree(v) = out-degree$(v) \geq 1$, we can pick some leaving edge of v and continue visiting edges and vertices. Each time we pick an edge, we can remove it from further consideration. At each vertex other than u, at the time we visit an entering edge, there must be an unvisited leaving edge, since in-degree(v) = out-degree(v) for all vertices v. The only vertex for which there might not be an unvisited leaving edge is u, since we started the cycle by visiting one of u's leaving edges. Since there's always a leaving edge we can visit from all vertices other than u, eventually the cycle must return to u, thus proving the claim.

The nonconstructive proof proves the contrapositive—that if G does not have an Euler tour, then in-degree$(v) \neq$ out-degree(v) for some vertex v—by contradiction. Choose a graph $G = (V, E)$ that does not have an Euler tour but has at least one edge and for which in-degree(v) = out-degree(v) for all vertices v, and let G have the fewest edges of any such graph. By the above claim, G contains a cycle. Let C be a cycle of G with the greatest number of edges, and let V_C be the set of vertices visited by cycle C. By our assumption, C is not an Euler tour, and so the set of edges $E' = E - C$ is nonempty. If we use the set V of vertices and the set E' of edges, we get the graph $G' = (V, E')$; this graph has in-degree(v) = out-degree(v) for all vertices v, since we have removed one entering edge and one leaving edge for each vertex on cycle C. Consider any component $G'' = (V'', E'')$ of G', and observe that G'' also has in-degree(v) = out-degree(v) for all vertices v. Since $E'' \subseteq E' \subsetneq E$, it follows from how we chose G that G'' must have an Euler tour, say C'. Because the original graph G is connected, there must be some vertex $x \in V'' \cup V_C$ and, without loss of generality, consider x to be the first and last vertex on both C and C'. But then the cycle C'' formed by first traversing C and then traversing C' is a cycle of G with more edges than C, contradicting our choice of C. We conclude that C must have been an Euler tour.

The constructive proof uses the same ideas. Let us start at a vertex u and, via random traversal of edges, create a cycle. We know that once we take any edge entering a vertex $v \neq u$, we can find an edge leaving v that we have not yet taken. Eventually, we get back to vertex u, and if there are still edges leaving u that we have not taken, we can continue the cycle. Eventually, we get back to vertex u and there are no untaken edges leaving u. If we have visited every edge in the graph G, we are done. Otherwise, since G is connected, there must be some unvisited edge leaving a vertex, say v, on the cycle. We can traverse a new cycle starting at v, visiting only previously unvisited edges, and we can splice this cycle into the cycle we already know. That is, if the original cycle is $\langle u, \ldots, v, w, \ldots, u\rangle$, and the new cycle is $\langle v, x, \ldots, v\rangle$, then we can create the cycle $\langle u, \ldots, v, x, \ldots, v, w, \ldots, u\rangle$. We continue this process of finding a vertex with an unvisited leaving edge on a visited cycle, visiting a cycle starting and ending at this vertex, and splicing in the newly visited cycle, until we have visited every edge.

b. The algorithm is based on the idea in the constructive proof above.

We assume that G is represented by adjacency lists, and we work with a copy of the adjacency lists, so that as we visit each edge, we can remove it from its adjacency list. The singly linked form of adjacency list will suffice. The output of this algorithm is a doubly linked list T of vertices which, read in list order, will give an Euler tour. The algorithm constructs T by finding cycles (also represented by doubly linked lists) and splicing them into T. By using doubly linked lists for cycles and the Euler tour, splicing a cycle into the Euler tour takes constant time.

We also maintain a singly linked list L, in which each list element consists of two parts:

1. a vertex v, and
2. a pointer to some appearance of v in T.

Initially, L contains one vertex, which may be any vertex of G.

Here is the algorithm:

```
EULER-TOUR(G)
  T = empty list
  L = (any vertex v ∈ G.V, NIL)
  while L is not empty
      remove (v, location-in-T) from L
      C = VISIT(G, L, v)
      if location-in-T == NIL
          T = C
      else splice C into T just before location-in-T
  return T

VISIT(G, L, v)
  C = empty sequence of vertices
  u = v
  while out-degree(u) > 0
      let w be the first vertex in G.Adj[u]
      remove w from G.Adj[u], decrementing out-degree(u)
      add u onto the end of C
      if out-degree(u) > 0
          add (u, u's location in C) to L
      u = w
  return C
```

The use of NIL in the initial assignment to L ensures that the first cycle C returned by VISIT becomes the current version of the Euler tour T. All cycles returned by VISIT thereafter are spliced into T. We assume that whenever an empty cycle is returned by VISIT, splicing it into T leaves T unchanged.

Each time that EULER-TOUR removes a vertex v from the list L, it calls VISIT(G, L, v) to find a cycle C, possibly empty and possibly not simple, that starts and ends at v; the cycle C is represented by a list that starts with v and ends with the last vertex on the cycle before the cycle ends at v. EULER-TOUR

then splices this cycle C into the Euler tour T just before some appearance of v in T.

When VISIT is at a vertex u, it looks for some vertex w such that the edge (u, w) has not yet been visited. Removing w from $Adj[u]$ ensures that we will never visit (u, w) again. VISIT adds u onto the cycle C that it constructs. If, after removing edge (u, w), vertex u still has any leaving edges, then u, along with its location in C, is added to L. The cycle construction continues from w, and it ceases once a vertex with no unvisited leaving edges is found. Using the argument from part (a), at that point, this vertex must close up a cycle. At that point, therefore, the cycle C is returned.

It is possible that a vertex u has unvisited leaving edges at the time it is added to list L in VISIT, but that by the time that u is removed from L in EULER-TOUR, all of its leaving edges have been visited. In this case, the **while** loop of VISIT executes 0 iterations, and VISIT returns an empty cycle.

Once the list L is empty, every edge has been visited. The resulting cycle T is then an Euler tour.

To see that EULER-TOUR takes $O(E)$ time, observe that because we remove each edge from its adjacency list as it is visited, no edge is visited more than once. Since each edge is visited at some time, the number of times that a vertex is added to L, and thus removed from L, is at most $|E|$. Thus, the **while** loop in EULER-TOUR executes at most E iterations. The **while** loop in VISIT executes one iteration per edge in the graph, and so it executes at most E iterations as well. Since adding vertex u to the doubly linked list C takes constant time and splicing C into T takes constant time, the entire algorithm takes $O(E)$ time.

Solution to Problem 22-4

Compute G^{T} in the usual way, so that G^{T} is G with its edges reversed. Then do a depth-first search on G^{T}, but in the main loop of DFS, consider the vertices in order of increasing values of $L(v)$. If vertex u is in the depth-first tree with root v, then $\min(u) = v$. Clearly, this algorithm takes $O(V + E)$ time.

To show correctness, first note that if u is in the depth-first tree rooted at v in G^{T}, then there is a path $v \rightsquigarrow u$ in G^{T}, and so there is a path $u \rightsquigarrow v$ in G. Thus, the minimum vertex label of all vertices reachable from u is at most $L(v)$, or in other words, $L(v) \geq \min\{L(w) : w \in R(u)\}$.

Now suppose that $L(v) > \min\{L(w) : w \in R(u)\}$, so that there is a vertex $w \in R(u)$ such that $L(w) < L(v)$. At the time $v.d$ that we started the depth-first search from v, we would have already discovered w, so that $w.d < v.d$. By the parenthesis theorem, either the intervals $[v.d, v.f]$, and $[w.d, w.f]$ are disjoint and neither v nor w is a descendant of the other, or we have the ordering $w.d < v.d < v.f < w.f$ and v is a descendant of w. The latter case cannot occur, since v is a root in the depth-first forest (which means that v cannot be a descendant of any other vertex). In the former case, since $w.d < v.d$, we must have $w.d < w.f < v.d < v.f$. In this case, since u is reachable from w in G^{T}, we would

have discovered u by the time $w.f$, so that $u.d < w.f$. Since we discovered u during a search that started at v, we have $v.d \le u.d$. Thus, $v.d \le u.d < w.f < v.d$, which is a contradiction. We conclude that no such vertex w can exist.

Lecture Notes for Chapter 23: Minimum Spanning Trees

Chapter 23 overview

Problem

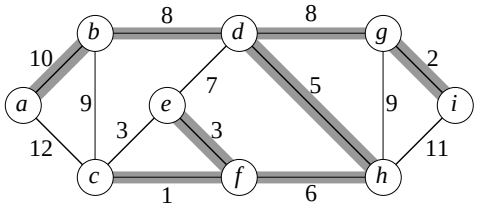
- A town has a set of houses and a set of roads.
- A road connects 2 and only 2 houses.
- A road connecting houses u and v has a repair cost $w(u, v)$.
- ***Goal:*** Repair enough (and no more) roads such that
  1. everyone stays connected: can reach every house from all other houses, and
  2. total repair cost is minimum.

Model as a graph:

- Undirected graph $G = (V, E)$.
- ***Weight*** $w(u, v)$ on each edge $(u, v) \in E$.
- Find $T \subseteq E$ such that
  1. T connects all vertices (T is a ***spanning tree***), and
  2. $w(T) = \sum_{(u,v)\in T} w(u, v)$ is minimized.

A spanning tree whose weight is minimum over all spanning trees is called a ***minimum spanning tree***, or ***MST***.

Example of such a graph *[edges in MST are shaded]* :



In this example, there is more than one MST. Replace edge (e, f) in the MST by (c, e). Get a different spanning tree with the same weight.

Growing a minimum spanning tree

Some properties of an MST:

- It has $|V| - 1$ edges.
- It has no cycles.
- It might not be unique.

Building up the solution

- We will build a set A of edges.
- Initially, A has no edges.
- As we add edges to A, maintain a loop invariant:

  Loop invariant: A is a subset of some MST.

- Add only edges that maintain the invariant. If A is a subset of some MST, an edge (u, v) is ***safe*** for A if and only if $A \cup \{(u, v)\}$ is also a subset of some MST. So we will add only safe edges.

Generic MST algorithm

GENERIC-MST(G, w)
  $A = \emptyset$
  while A is not a spanning tree
    find an edge (u, v) that is safe for A
    $A = A \cup \{(u, v)\}$
  return A

Use the loop invariant to show that this generic algorithm works.

Initialization: The empty set trivially satisfies the loop invariant.

Maintenance: Since we add only safe edges, A remains a subset of some MST.

Termination: All edges added to A are in an MST, so when we stop, A is a spanning tree that is also an MST.

Finding a safe edge

How do we find safe edges?

Let's look at the example. Edge (c, f) has the lowest weight of any edge in the graph. Is it safe for $A = \emptyset$?

Intuitively: Let $S \subset V$ be any set of vertices that includes c but not f (so that f is in $V - S$). In any MST, there has to be one edge (at least) that connects S with $V - S$. Why not choose the edge with minimum weight? (Which would be (c, f) in this case.)

Some definitions: Let $S \subset V$ and $A \subseteq E$.

- A ***cut*** $(S, V - S)$ is a partition of vertices into disjoint sets V and $S - V$.
- Edge $(u, v) \in E$ ***crosses*** cut $(S, V - S)$ if one endpoint is in S and the other is in $V - S$.
- A cut ***respects*** A if and only if no edge in A crosses the cut.
- An edge is a ***light edge*** crossing a cut if and only if its weight is minimum over all edges crossing the cut. For a given cut, there can be > 1 light edge crossing it.

Theorem

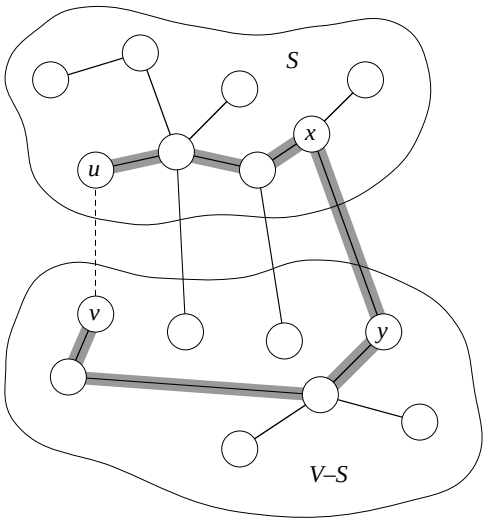
Let A be a subset of some MST, $(S, V - S)$ be a cut that respects A, and (u, v) be a light edge crossing $(S, V - S)$. Then (u, v) is safe for A.

Proof Let T be an MST that includes A.

If T contains (u, v), done.

So now assume that T does not contain (u, v). We'll construct a different MST T' that includes $A \cup \{(u, v)\}$.

Recall: a tree has unique path between each pair of vertices. Since T is an MST, it contains a unique path p between u and v. Path p must cross the cut $(S, V - S)$ at least once. Let (x, y) be an edge of p that crosses the cut. From how we chose (u, v), must have $w(u, v) \leq w(x, y)$.



[Except for the dashed edge (u, v), all edges shown are in T. A is some subset of the edges of T, but A cannot contain any edges that cross the cut $(S, V - S)$, since this cut respects A. Shaded edges are the path p.]

Since the cut respects A, edge (x, y) is not in A.

To form T' from T:

- Remove (x, y). Breaks T into two components.
- Add (u, v). Reconnects.

So $T' = T - \{(x, y)\} \cup \{(u, v)\}$.

T' is a spanning tree.

$$\begin{aligned} w(T') &= w(T) - w(x, y) + w(u, v) \\ &\leq w(T) , \end{aligned}$$

since $w(u, v) \leq w(x, y)$. Since T' is a spanning tree, $w(T') \leq w(T)$, and T is an MST, then T' must be an MST.

Need to show that (u, v) is safe for A:

- $A \subseteq T$ and $(x, y) \notin A \Rightarrow A \subseteq T'$.
- $A \cup \{(u, v)\} \subseteq T'$.
- Since T' is an MST, (u, v) is safe for A. ■ (theorem)

So, in GENERIC-MST:

- A is a forest containing connected components. Initially, each component is a single vertex.
- Any safe edge merges two of these components into one. Each component is a tree.
- Since an MST has exactly $|V| - 1$ edges, the **for** loop iterates $|V| - 1$ times. Equivalently, after adding $|V|-1$ safe edges, we're down to just one component.

Corollary

If $C = (V_C, E_C)$ is a connected component in the forest $G_A = (V, A)$ and (u, v) is a light edge connecting C to some other component in G_A (i.e., (u, v) is a light edge crossing the cut $(V_C, V - V_C)$), then (u, v) is safe for A.

Proof Set $S = V_C$ in the theorem. ■ (corollary)

This idea naturally leads to the algorithm known as Kruskal's algorithm to solve the minimum-spanning-tree problem.

Kruskal's algorithm

$G = (V, E)$ is a connected, undirected, weighted graph. $w : E \to \mathbb{R}$.

- Starts with each vertex being its own component.
- Repeatedly merges two components into one by choosing the light edge that connects them (i.e., the light edge crossing the cut between them).
- Scans the set of edges in monotonically increasing order by weight.
- Uses a disjoint-set data structure to determine whether an edge connects vertices in different components.

```
KRUSKAL(G, w)
  A = ∅
  for each vertex v ∈ G.V
      MAKE-SET(v)
  sort the edges of G.E into nondecreasing order by weight w
  for each (u, v) taken from the sorted list
      if FIND-SET(u) ≠ FIND-SET(v)
          A = A ∪ {(u, v)}
          UNION(u, v)
  return A
```

Run through the above example to see how Kruskal's algorithm works on it:

(c, f) : safe
(g, i) : safe
(e, f) : safe
(c, e) : reject
(d, h) : safe
(f, h) : safe
(e, d) : reject
(b, d) : safe
(d, g) : safe
(b, c) : reject
(g, h) : reject
(a, b) : safe

At this point, we have only one component, so all other edges will be rejected. *[We could add a test to the main loop of* KRUSKAL *to stop once* $|V| - 1$ *edges have been added to* A*.]*

Get the shaded edges shown in the figure.

Suppose we had examined (c, e) *before* (e, f). Then would have found (c, e) safe and would have rejected (e, f).

Analysis

Initialize A:	$O(1)$
First **for** loop:	$\|V\|$ MAKE-SETs
Sort E:	$O(E \lg E)$
Second **for** loop:	$O(E)$ FIND-SETs and UNIONs

- Assuming the implementation of disjoint-set data structure, already seen in Chapter 21, that uses union by rank and path compression:

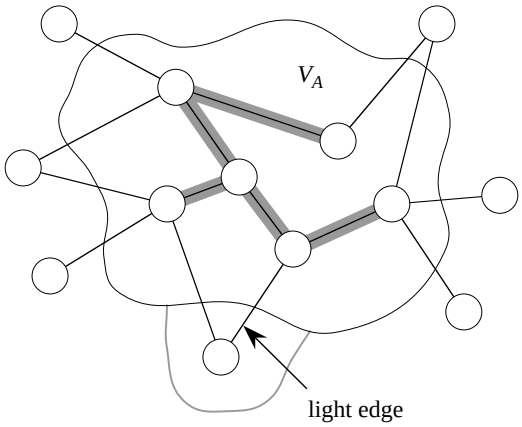
  $$O((V + E)\,\alpha(V)) + O(E \lg E)\ .$$

- Since G is connected, $|E| \geq |V| - 1 \Rightarrow O(E\,\alpha(V)) + O(E \lg E)$.
- $\alpha(|V|) = O(\lg V) = O(\lg E)$.
- Therefore, total time is $O(E \lg E)$.
- $|E| \leq |V|^2 \Rightarrow \lg |E| = O(2 \lg V) = O(\lg V)$.

- Therefore, $O(E \lg V)$ time. (If edges are already sorted, $O(E\,\alpha(V))$, which is almost linear.)

Prim's algorithm

- Builds one tree, so A is always a tree.
- Starts from an arbitrary "root" r.
- At each step, find a light edge crossing cut $(V_A, V - V_A)$, where V_A = vertices that A is incident on. Add this edge to A.



[Edges of A are shaded.]

How to find the light edge quickly?

Use a priority queue Q:

- Each object is a vertex in $V - V_A$.
- Key of v is minimum weight of any edge (u, v), where $u \in V_A$.
- Then the vertex returned by EXTRACT-MIN is v such that there exists $u \in V_A$ and (u, v) is light edge crossing $(V_A, V - V_A)$.
- Key of v is ∞ if v is not adjacent to any vertices in V_A.

The edges of A will form a rooted tree with root r:

- r is given as an input to the algorithm, but it can be any vertex.
- Each vertex knows its parent in the tree by the attribute $v.\pi$ = parent of v. $v.\pi$ = NIL if $v = r$ or v has no parent.
- As algorithm progresses, $A = \{(v, v.\pi) : v \in V - \{r\} - Q\}$.
- At termination, $V_A = V \Rightarrow Q = \emptyset$, so MST is $A = \{(v, v.\pi) : v \in V - \{r\}\}$.

[The pseudocode that follows differs from the book in that it explicitly calls INSERT *and* DECREASE-KEY *to operate on Q.]*

```
PRIM(G, w, r)
  Q = ∅
  for each u ∈ G.V
      u.key = ∞
      u.π = NIL
      INSERT(Q, u)
  DECREASE-KEY(Q, r, 0)        // r.key = 0
  while Q ≠ ∅
      u = EXTRACT-MIN(Q)
      for each v ∈ G.Adj[u]
          if v ∈ Q and w(u, v) < v.key
              v.π = u
              DECREASE-KEY(Q, v, w(u, v))
```

Do example from previous graph. *[Let a student pick the root.]*

Analysis

Depends on how the priority queue is implemented:

- Suppose Q is a binary heap.

Initialize Q and first **for** loop:	$O(V \lg V)$
Decrease key of r:	$O(\lg V)$
while loop:	$\lvert V\rvert$ EXTRACT-MIN calls $\Rightarrow O(V \lg V)$
	$\leq \lvert E\rvert$ DECREASE-KEY calls $\Rightarrow O(E \lg V)$
Total:	$O(E \lg V)$

- Suppose we could do DECREASE-KEY in $O(1)$ *amortized* time.

  Then $\leq \lvert E\rvert$ DECREASE-KEY calls take $O(E)$ time altogether $\Rightarrow$ total time becomes $O(V \lg V + E)$.

  In fact, there is a way to do DECREASE-KEY in $O(1)$ amortized time: Fibonacci heaps, in Chapter 19.

Solutions for Chapter 23: Minimum Spanning Trees

Solution to Exercise 23.1-1

This solution is also posted publicly

Theorem 23.1 shows this.

Let A be the empty set and S be any set containing u but not v.

Solution to Exercise 23.1-4

This solution is also posted publicly

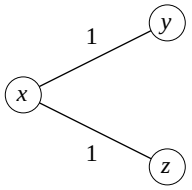
A triangle whose edge weights are all equal is a graph in which every edge is a light edge crossing some cut. But the triangle is cyclic, so it is not a minimum spanning tree.

Solution to Exercise 23.1-6

This solution is also posted publicly

Suppose that for every cut of G, there is a unique light edge crossing the cut. Let us consider two distinct minimum spanning trees, T and T', of G. Because T and T' are distinct, T contains some edge (u, v) that is not in T'. If we remove (u, v) from T, then T becomes disconnected, resulting in a cut $(S, V - S)$. The edge (u, v) is a light edge crossing the cut $(S, V - S)$ (by Exercise 23.1-3) and, by our assumption, it's the only light edge crossing this cut. Because (u, v) is the only light edge crossing $(S, V - S)$ and (u, v) is not in T', each edge in T' that crosses $(S, V - S)$ must have weight strictly greater than $w(u, v)$. As in the proof of Theorem 23.1, we can identify the unique edge (x, y) in T' that crosses $(S, V - S)$ and lies on the cycle that results if we add (u, v) to T'. By our assumption, we know that $w(u, v) < w(x, y)$. Then, we can then remove (x, y) from T' and replace it by (u, v), giving a spanning tree with weight strictly less than $w(T')$. Thus, T' was not a minimum spanning tree, contradicting the assumption that the graph had two unique minimum spanning trees.

Here's a counterexample for the converse:



Here, the graph is its own minimum spanning tree, and so the minimum spanning tree is unique. Consider the cut $(\{x\}, \{y, z\})$. Both of the edges (x, y) and (x, z) are light edges crossing the cut, and they are both light edges.

Solution to Exercise 23.1-10

Let $w(T) = \sum_{(x,y)\in T} w(x, y)$. We have $w'(T) = w(T) - k$. Consider any other spanning tree T', so that $w(T) \le w(T')$.

If $(x, y) \notin T'$, then $w'(T') = w(T') \ge w(T) > w'(T)$.

If $(x, y) \in T'$, then $w'(T') = w(T') - k \ge w(T) - k = w'(T)$.

Either way, $w'(T) \le w'(T')$, and so T is a minimum spanning tree for weight function w'.

Solution to Exercise 23.2-4

We know that Kruskal's algorithm takes $O(V)$ time for initialization, $O(E \lg E)$ time to sort the edges, and $O(E\,\alpha(V))$ time for the disjoint-set operations, for a total running time of $O(V + E \lg E + E\,\alpha(V)) = O(E \lg E)$.

If we knew that all of the edge weights in the graph were integers in the range from 1 to $|V|$, then we could sort the edges in $O(V + E)$ time using counting sort. Since the graph is connected, $V = O(E)$, and so the sorting time is reduced to $O(E)$. This would yield a total running time of $O(V + E + E\,\alpha(V)) = O(E\,\alpha(V))$, again since $V = O(E)$, and since $E = O(E\,\alpha(V))$. The time to process the edges, not the time to sort them, is now the dominant term. Knowledge about the weights won't help speed up any other part of the algorithm, since nothing besides the sort uses the weight values.

If the edge weights were integers in the range from 1 to W for some constant W, then we could again use counting sort to sort the edges more quickly. This time, sorting would take $O(E + W) = O(E)$ time, since W is a constant. As in the first part, we get a total running time of $O(E\,\alpha(V))$.

Solution to Exercise 23.2-5

The time taken by Prim's algorithm is determined by the speed of the queue operations. With the queue implemented as a Fibonacci heap, it takes $O(E + V \lg V)$ time.

Since the keys in the priority queue are edge weights, it might be possible to implement the queue even more efficiently when there are restrictions on the possible edge weights.

We can improve the running time of Prim's algorithm if W is a constant by implementing the queue as an array $Q[0..W+1]$ (using the $W+1$ slot for key$=\infty$), where each slot holds a doubly linked list of vertices with that weight as their key. Then EXTRACT-MIN takes only $O(W) = O(1)$ time (just scan for the first nonempty slot), and DECREASE-KEY takes only $O(1)$ time (just remove the vertex from the list it's in and insert it at the front of the list indexed by the new key). This gives a total running time of $O(E)$, which is the best possible asymptotic time (since $\Omega(E)$ edges must be processed).

However, if the range of edge weights is 1 to $|V|$, then EXTRACT-MIN takes $\Theta(V)$ time with this data structure. So the total time spent doing EXTRACT-MIN is $\Theta(V^2)$, slowing the algorithm to $\Theta(E + V^2) = \Theta(V^2)$. In this case, it is better to keep the Fibonacci-heap priority queue, which gave the $\Theta(E + V \lg V)$ time.

Other data structures yield better running times:

- van Emde Boas trees (see Chapter 20) give an upper bound of $O(E + V \lg\lg V)$ time for Prim's algorithm.
- A redistributive heap (used in the single-source shortest-paths algorithm of Ahuja, Mehlhorn, Orlin, and Tarjan, and mentioned in the chapter notes for Chapter 24) gives an upper bound of $O\left(E + V\sqrt{\lg V}\right)$ for Prim's algorithm.

Solution to Exercise 23.2-7

We start with the following lemma.

Lemma

Let T be a minimum spanning tree of $G = (V, E)$, and consider a graph $G' = (V', E')$ for which G is a subgraph, i.e., $V \subseteq V'$ and $E \subseteq E'$. Let $\overline{T} = E - T$ be the edges of G that are not in T. Then there is a minimum spanning tree of G' that includes no edges in $\overline{T}$.

Proof By Exercise 23.2-1, there is a way to order the edges of E so that Kruskal's algorithm, when run on G, produces the minimum spanning tree T. We will show that Kruskal's algorithm, run on G', produces a minimum spanning tree T' that includes no edges in $\overline{T}$. We assume that the edges in E are considered in the same relative order when Kruskal's algorithm is run on G and on G'. We first state and prove the following claim.

Claim

For any pair of vertices $u, v \in V$, if these vertices are in the same set after Kruskal's algorithm run on G considers any edge $(x, y) \in E$, then they are in the same set after Kruskal's algorithm run on G' considers (x, y).

Proof of claim Let us order the edges of E by nondecreasing weight as $\langle (x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k) \rangle$, where $k = |E|$. This sequence gives the order in which the edges of E are considered by Kruskal's algorithm, whether it is run on G or on G'. We will use induction, with the inductive hypothesis that if u and v are in the same set after Kruskal's algorithm run on G considers an edge (x_i, y_i), then they are in the same set after Kruskal's algorithm run on G' considers the same edge. We use induction on i.

Basis: For the basis, $i = 0$. Kruskal's algorithm run on G has not considered any edges, and so all vertices are in different sets. The inductive hypothesis holds trivially.

Inductive step: We assume that any vertices that are in the same set after Kruskal's algorithm run on G has considered edges $\langle (x_1, y_1), (x_2, y_2), \ldots, (x_{i-1}, y_{i-1}) \rangle$ are in the same set after Kruskal's algorithm run on G' has considered the same edges. When Kruskal's algorithm runs on G', after it considers (x_{i-1}, y_{i-1}), it may consider some edges in $E' - E$ before considering (x_i, y_i). The edges in $E' - E$ may cause UNION operations to occur, but sets are never divided. Hence, any vertices that are in the same set after Kruskal's algorithm run on G' considers (x_{i-1}, y_{i-1}) are still in the same set when (x_i, y_i) is considered.

When Kruskal's algorithm run on G considers (x_i, y_i), either x_i and y_i are found to be in the same set or they are not.

- If Kruskal's algorithm run on G finds x_i and y_i to be in the same set, then no UNION operation occurs. The sets of vertices remain the same, and so the inductive hypothesis continues to hold after considering (x_i, y_i).
- If Kruskal's algorithm run on G finds x_i and y_i to be in different sets, then the operation UNION(x_i, y_i) will occur. Kruskal's algorithm run on G' will find that either x_i and y_i are in the same set or they are not. By the inductive hypothesis, when edge (x_i, y_i) is considered, all vertices in x_i's set when Kruskal's algorithm runs on G are in x_i's set when Kruskal's algorithm runs on G', and the same holds for y_i. Regardless of whether Kruskal's algorithm run on G' finds x_i and y_i to already be in the same set, their sets are united after considering (x_i, y_i), and so the inductive hypothesis continues to hold after considering (x_i, y_i). ■ (claim)

With the claim in hand, we suppose that some edge $(u, v) \in \overline{T}$ is placed into T'. That means that Kruskal's algorithm run on G found u and v to be in the same set (since $(u, v) \in \overline{T}$) but Kruskal's algorithm run on G' found u and v to be in different sets (since (u, v) is placed into T'). This fact contradicts the claim, and we conclude that no edge in $\overline{T}$ is placed into T'. Thus, by running Kruskal's algorithm on G and G', we demonstrate that there exists a minimum spanning tree of G' that includes no edges in $\overline{T}$. ■ (lemma)

We use this lemma as follows. Let $G' = (V', E')$ be the graph $G = (V, E)$ with the one new vertex and its incident edges added. Suppose that we have a minimum

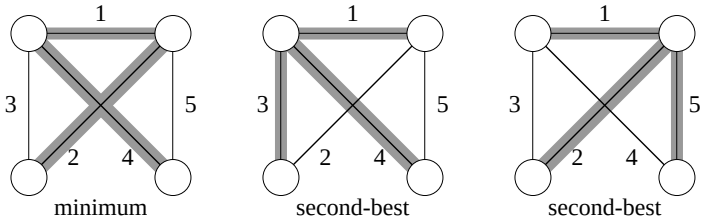
spanning tree T for G. We compute a minimum spanning tree for G' by creating the graph $G'' = (V', E'')$, where E'' consists of the edges of T and the edges in $E' - E$ (i.e., the edges added to G that made G'), and then finding a minimum spanning tree T' for G''. By the lemma, there is a minimum spanning tree for G' that includes no edges of $E - T$. In other words, G' has a minimum spanning tree that includes only edges in T and $E' - E$; these edges comprise exactly the set E''. Thus, the the minimum spanning tree T' of G'' is also a minimum spanning tree of G'.

Even though the proof of the lemma uses Kruskal's algorithm, we are not required to use this algorithm to find T'. We can find a minimum spanning tree by any means we choose. Let us use Prim's algorithm with a Fibonacci-heap priority queue. Since $|V'| = |V| + 1$ and $|E''| \le 2|V| - 1$ (E'' contains the $|V| - 1$ edges of T and at most $|V|$ edges in $E' - E$), it takes $O(V)$ time to construct G'', and the run of Prim's algorithm with a Fibonacci-heap priority queue takes time $O(E'' + V' \lg V') = O(V \lg V)$. Thus, if we are given a minimum spanning tree of G, we can compute a minimum spanning tree of G' in $O(V \lg V)$ time.

Solution to Problem 23-1

a. To see that the minimum spanning tree is unique, observe that since the graph is connected and all edge weights are distinct, then there is a unique light edge crossing every cut. By Exercise 23.1-6, the minimum spanning tree is unique.

To see that the second-best minimum spanning tree need not be unique, here is a weighted, undirected graph with a unique minimum spanning tree of weight 7 and two second-best minimum spanning trees of weight 8:



b. Since any spanning tree has exactly $|V| - 1$ edges, any second-best minimum spanning tree must have at least one edge that is not in the (best) minimum spanning tree. If a second-best minimum spanning tree has exactly one edge, say (x, y), that is not in the minimum spanning tree, then it has the same set of edges as the minimum spanning tree, except that (x, y) replaces some edge, say (u, v), of the minimum spanning tree. In this case, $T' = T - \{(u, v)\} \cup \{(x, y)\}$, as we wished to show.

Thus, all we need to show is that by replacing two or more edges of the minimum spanning tree, we cannot obtain a second-best minimum spanning tree. Let T be the minimum spanning tree of G, and suppose that there exists a second-best minimum spanning tree T' that differs from T by two or more

edges. There are at least two edges in $T - T'$, and let (u, v) be the edge in $T - T'$ with minimum weight. If we were to add (u, v) to T', we would get a cycle c. This cycle contains some edge (x, y) in $T' - T$ (since otherwise, T would contain a cycle).

We claim that $w(x, y) > w(u, v)$. We prove this claim by contradiction, so let us assume that $w(x, y) < w(u, v)$. (Recall the assumption that edge weights are distinct, so that we do not have to concern ourselves with $w(x, y) = w(u, v)$.) If we add (x, y) to T, we get a cycle c', which contains some edge (u', v') in $T - T'$ (since otherwise, T' would contain a cycle). Therefore, the set of edges $T'' = T - \{(u', v')\} \cup \{(x, y)\}$ forms a spanning tree, and we must also have $w(u', v') < w(x, y)$, since otherwise T'' would be a spanning tree with weight less than $w(T)$. Thus, $w(u', v') < w(x, y) < w(u, v)$, which contradicts our choice of (u, v) as the edge in $T - T'$ of minimum weight.

Since the edges (u, v) and (x, y) would be on a common cycle c if we were to add (u, v) to T', the set of edges $T' - \{(x, y)\} \cup \{(u, v)\}$ is a spanning tree, and its weight is less than $w(T')$. Moreover, it differs from T (because it differs from T' by only one edge). Thus, we have formed a spanning tree whose weight is less than $w(T')$ but is not T. Hence, T' was not a second-best minimum spanning tree.

c. We can fill in $max[u, v]$ for all $u, v \in V$ in $O(V^2)$ time by simply doing a search from each vertex u, having restricted the edges visited to those of the spanning tree T. It doesn't matter what kind of search we do: breadth-first, depth-first, or any other kind.

We'll give pseudocode for both breadth-first and depth-first approaches. Each approach differs from the pseudocode given in Chapter 22 in that we don't need to compute d or f values, and we'll use the *max* table itself to record whether a vertex has been visited in a given search. In particular, $max[u, v] = \text{NIL}$ if and only if $u = v$ or we have not yet visited vertex v in a search from vertex u. Note also that since we're visiting via edges in a spanning tree of an undirected graph, we are guaranteed that the search from each vertex u—whether breadth-first or depth-first—will visit all vertices. There will be no need to "restart" the search as is done in the DFS procedure of Section 22.3. Our pseudocode assumes that the adjacency list of each vertex consists only of edges in the spanning tree T.

Here's the breadth-first search approach:

```
BFS-FILL-MAX(G, T, w)
  let max be a new table with an entry max[u, v] for each u, v ∈ G.V
  for each vertex u ∈ G.V
      for each vertex v ∈ G.V
          max[u, v] = NIL
      Q = ∅
      ENQUEUE(Q, u)
      while Q ≠ ∅
          x = DEQUEUE(Q)
          for each v ∈ G.Adj[x]
              if max[u, v] == NIL and v ≠ u
                  if x == u or w(x, v) > max[u, x]
                      max[u, v] = (x, v)
                  else max[u, v] = max[u, x]
                  ENQUEUE(Q, v)
  return max
```

Here's the depth-first search approach:

```
DFS-FILL-MAX(G, T, w)
  let max be a new table with an entry max[u, v] for each u, v ∈ G.V
  for each vertex u ∈ G.V
      for each vertex v ∈ G.V
          max[u, v] = NIL
      DFS-FILL-MAX-VISIT(G, u, u, max)
  return max

DFS-FILL-MAX-VISIT(G, u, x, max)
  for each vertex v ∈ G.Adj[x]
      if max[u, v] == NIL and v ≠ u
          if x == u or w(x, v) > max[u, x]
              max[u, v] = (x, v)
          else max[u, v] = max[u, x]
          DFS-FILL-MAX-VISIT(G, u, v, max)
```

For either approach, we are filling in $|V|$ rows of the *max* table. Since the number of edges in the spanning tree is $|V| - 1$, each row takes $O(V)$ time to fill in. Thus, the total time to fill in the *max* table is $O(V^2)$.

d. In part (b), we established that we can find a second-best minimum spanning tree by replacing just one edge of the minimum spanning tree T by some edge (u, v) not in T. As we know, if we create spanning tree T' by replacing edge $(x, y) \in T$ by edge $(u, v) \notin T$, then $w(T') = w(T) - w(x, y) + w(u, v)$. For a given edge (u, v), the edge $(x, y) \in T$ that minimizes $w(T')$ is the edge of maximum weight on the unique path between u and v in T. If we have already computed the *max* table from part (c) based on T, then the identity of this edge is precisely what is stored in $max[u, v]$. All we have to do is determine an edge $(u, v) \notin T$ for which $w(max[u, v]) - w(u, v)$ is minimum.

Thus, our algorithm to find a second-best minimum spanning tree goes as follows:

1. Compute the minimum spanning tree T. Time: $O(E + V \lg V)$, using Prim's algorithm with a Fibonacci-heap implementation of the priority queue. Since $|E| < |V|^2$, this running time is $O(V^2)$.
2. Given the minimum spanning tree T, compute the *max* table, as in part (c). Time: $O(V^2)$.
3. Find an edge $(u, v) \notin T$ that minimizes $w(max[u, v]) - w(u, v)$. Time: $O(E)$, which is $O(V^2)$.
4. Having found an edge (u, v) in step 3, return $T' = T - \{max[u, v]\} \cup \{(u, v)\}$ as a second-best minimum spanning tree.

The total time is $O(V^2)$.

Lecture Notes for Chapter 24: Single-Source Shortest Paths

Shortest paths

How to find the shortest route between two points on a map.

Input:

- Directed graph $G = (V, E)$
- Weight function $w : E \to \mathbb{R}$

Weight of path $p = \langle v_0, v_1, \ldots, v_k \rangle$

$$= \sum_{i=1}^{k} w(v_{i-1}, v_i)$$

= sum of edge weights on path p .

Shortest-path weight u to v:

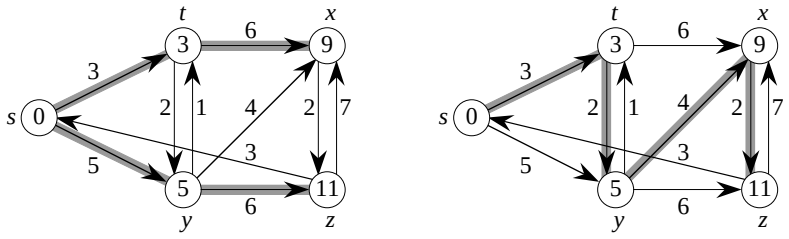
$$\delta(u, v) = \begin{cases} \min \left\{ w(p) : u \overset{p}{\rightsquigarrow} v \right\} & \text{if there exists a path } u \rightsquigarrow v \ , \\ \infty & \text{otherwise .} \end{cases}$$

Shortest path u to v is any path p such that $w(p) = \delta(u, v)$.

Example

shortest paths from s

[δ values appear inside vertices. Shaded edges show shortest paths.]



This example shows that the shortest path might not be unique.

It also shows that when we look at shortest paths from one vertex to all other vertices, the shortest paths are organized as a tree.

Can think of weights as representing any measure that

- accumulates linearly along a path, and
- we want to minimize.

Examples: time, cost, penalties, loss.

Generalization of breadth-first search to weighted graphs.

Variants

- ***Single-source:*** Find shortest paths from a given ***source*** vertex $s \in V$ to every vertex $v \in V$.
- ***Single-destination:*** Find shortest paths to a given destination vertex.
- ***Single-pair:*** Find shortest path from u to v. No way known that's better in worst case than solving single-source.
- ***All-pairs:*** Find shortest path from u to v for all $u, v \in V$. We'll see algorithms for all-pairs in the next chapter.

Negative-weight edges

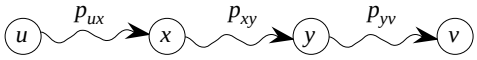
OK, as long as no negative-weight cycles are reachable from the source.

- If we have a negative-weight cycle, we can just keep going around it, and get $w(s, v) = -\infty$ for all v on the cycle.
- But OK if the negative-weight cycle is not reachable from the source.
- Some algorithms work only if there are no negative-weight edges in the graph. We'll be clear when they're allowed and not allowed.

Optimal substructure

Lemma
Any subpath of a shortest path is a shortest path.

Proof Cut-and-paste.



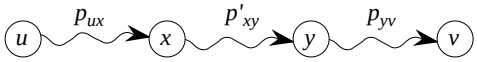
Suppose this path p is a shortest path from u to v.

Then $\delta(u, v) = w(p) = w(p_{ux}) + w(p_{xy}) + w(p_{yv})$.

Now suppose there exists a shorter path $x \overset{p'_{xy}}{\rightsquigarrow} y$.

Then $w(p'_{xy}) < w(p_{xy})$.

Construct p':



Then

$$\begin{aligned} w(p') &= w(p_{ux}) + w(p'_{xy}) + w(p_{yv}) \\ &< w(p_{ux}) + w(p_{xy}) + w(p_{yv}) \\ &= w(p) \, . \end{aligned}$$

Contradicts the assumption that p is a shortest path. ■ (lemma)

Cycles

Shortest paths can't contain cycles:

- Already ruled out negative-weight cycles.
- Positive-weight $\Rightarrow$ we can get a shorter path by omitting the cycle.
- Zero-weight: no reason to use them $\Rightarrow$ assume that our solutions won't use them.

Output of single-source shortest-path algorithm

For each vertex $v \in V$:

- $v.d = \delta(s, v)$.
  - Initially, $v.d = \infty$.
  - Reduces as algorithms progress. But always maintain $v.d \geq \delta(s, v)$.
  - Call $v.d$ a ***shortest-path estimate***.
- $v.\pi$ = predecessor of v on a shortest path from s.
  - If no predecessor, $v.\pi = \text{NIL}$.
  - π induces a tree—***shortest-path tree***.
  - We won't prove properties of π in lecture—see text.

Initialization

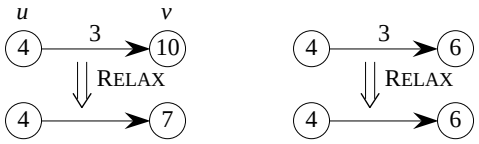
All the shortest-paths algorithms start with INIT-SINGLE-SOURCE.

```
INIT-SINGLE-SOURCE(G, s)
  for each v ∈ G.V
      v.d = ∞
      v.π = NIL
  s.d = 0
```

Relaxing an edge (u, v)

Can we improve the shortest-path estimate for v by going through u and taking (u, v)?

RELAX(u, v, w)
  if $v.d > u.d + w(u, v)$
    $v.d = u.d + w(u, v)$
    $v.\pi = u$



For all the single-source shortest-paths algorithms we'll look at,

- start by calling INIT-SINGLE-SOURCE,
- then relax edges.

The algorithms differ in the order and how many times they relax each edge.

Shortest-paths properties

Based on calling INIT-SINGLE-SOURCE once and then calling RELAX zero or more times.

Triangle inequality

For all $(u, v) \in E$, we have $\delta(s, v) \le \delta(s, u) + w(u, v)$.

Proof Weight of shortest path $s \rightsquigarrow v$ is $\le$ weight of any path $s \rightsquigarrow v$. Path $s \rightsquigarrow u \rightarrow v$ is a path $s \rightsquigarrow v$, and if we use a shortest path $s \rightsquigarrow u$, its weight is $\delta(s, u) + w(u, v)$. ■

Upper-bound property

Always have $v.d \ge \delta(s, v)$ for all v. Once $v.d = \delta(s, v)$, it never changes.

Proof Initially true.

Suppose there exists a vertex such that $v.d < \delta(s, v)$.

Without loss of generality, v is first vertex for which this happens.

Let u be the vertex that causes $v.d$ to change.

Then $v.d = u.d + w(u, v)$.

So,

$$\begin{aligned} v.d &< \delta(s, v) \\ &\le \delta(s, u) + w(u, v) && \text{(triangle inequality)} \\ &\le u.d + w(u, v) && (v \text{ is first violation}) \\ \Rightarrow v.d &< u.d + w(u, v)\ . \end{aligned}$$

Contradicts $v.d = u.d + w(u, v)$.

Once $v.d$ reaches $\delta(s, v)$, it never goes lower. It never goes up, since relaxations only lower shortest-path estimates. ■

No-path property

If $\delta(s, v) = \infty$, then $v.d = \infty$ always.

Proof $v.d \geq \delta(s, v) = \infty \Rightarrow v.d = \infty$. ■

Convergence property

If $s \rightsquigarrow u \rightarrow v$ is a shortest path, $u.d = \delta(s, u)$, and we call RELAX(u, v, w), then $v.d = \delta(s, v)$ afterward.

Proof After relaxation:

$$\begin{aligned} v.d &\leq u.d + w(u, v) && \text{(RELAX code)} \\ &= \delta(s, u) + w(u, v) \\ &= \delta(s, v) && \text{(lemma—optimal substructure)} \end{aligned}$$

Since $v.d \geq \delta(s, v)$, must have $v.d = \delta(s, v)$. ■

Path relaxation property

Let $p = \langle v_0, v_1, \ldots, v_k \rangle$ be a shortest path from $s = v_0$ to v_k. If we relax, *in order,* $(v_0, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_k)$, even intermixed with other relaxations, then $v_k.d = \delta(s, v_k)$.

Proof Induction to show that $v_i.d = \delta(s, v_i)$ after (v_{i-1}, v_i) is relaxed.

Basis: $i = 0$. Initially, $v_0.d = 0 = \delta(s, v_0) = \delta(s, s)$.

Inductive step: Assume $v_{i-1}.d = \delta(s, v_{i-1})$. Relax (v_{i-1}, v_i). By convergence property, $v_i.d = \delta(s, v_i)$ afterward and $v_i.d$ never changes. ■

The Bellman-Ford algorithm

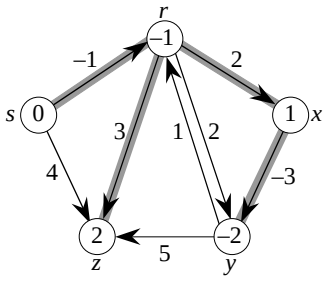
- Allows negative-weight edges.
- Computes $v.d$ and $v.\pi$ for all $v \in V$.
- Returns TRUE if no negative-weight cycles reachable from s, FALSE otherwise.

```
BELLMAN-FORD(G, w, s)
  INIT-SINGLE-SOURCE(G, s)
  for i = 1 to |G.V| − 1
      for each edge (u, v) ∈ G.E
          RELAX(u, v, w)
  for each edge (u, v) ∈ G.E
      if v.d > u.d + w(u, v)
          return FALSE
  return TRUE
```

Core: The nested **for** loops relax all edges $|V| - 1$ times.

Time: $\Theta(VE)$.

Example



Values you get on each pass and how quickly it converges depends on order of relaxation.

But guaranteed to converge after $|V| - 1$ passes, assuming no negative-weight cycles.

Proof Use path-relaxation property.

Let v be reachable from s, and let $p = \langle v_0, v_1, \ldots, v_k \rangle$ be a shortest path from s to v, where $v_0 = s$ and $v_k = v$. Since p is acyclic, it has $\leq |V| - 1$ edges, so $k \leq |V| - 1$.

Each iteration of the **for** loop relaxes all edges:

- First iteration relaxes (v_0, v_1).
- Second iteration relaxes (v_1, v_2).
- kth iteration relaxes (v_{k-1}, v_k).

By the path-relaxation property, $v.d = v_k.d = \delta(s, v_k) = \delta(s, v)$. ■

How about the TRUE/FALSE return value?

- Suppose there is no negative-weight cycle reachable from s.

  At termination, for all $(u, v) \in E$,

  $$\begin{aligned} v.d &= \delta(s, v) \\ &\leq \delta(s, u) + w(u, v) \quad \text{(triangle inequality)} \\ &= u.d + w(u, v) \ . \end{aligned}$$

  So BELLMAN-FORD returns TRUE.

- Now suppose there exists negative-weight cycle $c = \langle v_0, v_1, \ldots, v_k \rangle$, where $v_0 = v_k$, reachable from s.

  Then $\sum_{i=1}^{k} (v_{i-1}, v_i) < 0$.

  Suppose (for contradiction) that BELLMAN-FORD returns TRUE.

  Then $v_i.d \le v_{i-1}.d + w(v_{i-1}, v_i)$ for $i = 1, 2, \ldots, k$.

  Sum around c:

$$\begin{aligned} \sum_{i=1}^{k} v_i.d &\le \sum_{i=1}^{k} (v_{i-1}.d + w(v_{i-1}, v_i)) \\ &= \sum_{i=1}^{k} v_{i-1}.d + \sum_{i=1}^{k} w(v_{i-1}, v_i) \end{aligned}$$

  Each vertex appears once in each summation $\sum_{i=1}^{k} v_i.d$ and $\sum_{i=1}^{k} v_{i-1}.d \Rightarrow$

  $0 \le \sum_{i=1}^{k} w(v_{i-1}, v_i)$.

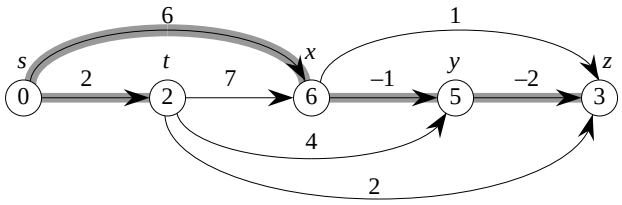
  Contradicts c being a negative-weight cycle. ■

Single-source shortest paths in a directed acyclic graph

Since a dag, we're guaranteed no negative-weight cycles.

DAG-SHORTEST-PATHS(G, w, s)
  topologically sort the vertices
  INIT-SINGLE-SOURCE(G, s)
  for each vertex u, taken in topologically sorted order
    for each vertex $v \in G.Adj[u]$
      RELAX(u, v, w)

Example



Time

$\Theta(V + E)$.

Correctness

Because we process vertices in topologically sorted order, edges of *any* path must be relaxed in order of appearance in the path.
$\Rightarrow$ Edges on any shortest path are relaxed in order.
$\Rightarrow$ By path-relaxation property, correct. ■

Dijkstra's algorithm

No negative-weight *edges*.

Essentially a weighted version of breadth-first search.

- Instead of a FIFO queue, uses a priority queue.
- Keys are shortest-path weights ($v.d$).

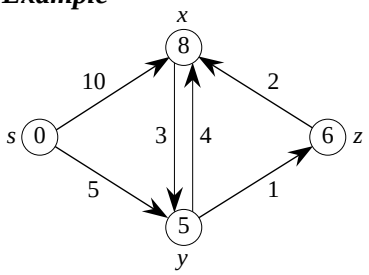
Have two sets of vertices:

- S = vertices whose final shortest-path weights are determined,
- Q = priority queue = $V - S$.

```
DIJKSTRA(G, w, s)
  INIT-SINGLE-SOURCE(G, s)
  S = ∅
  Q = G.V          // i.e., insert all vertices into Q
  while Q ≠ ∅
      u = EXTRACT-MIN(Q)
      S = S ∪ {u}
      for each vertex v ∈ G.Adj[u]
          RELAX(u, v, w)
```

- Looks a lot like Prim's algorithm, but computing $v.d$, and using shortest-path weights as keys.
- Dijkstra's algorithm can be viewed as greedy, since it always chooses the "lightest" ("closest"?) vertex in $V - S$ to add to S.

Example



Order of adding to S: s, y, z, x.

Correctness

Loop invariant: At the start of each iteration of the **while** loop, $v.d = \delta(s, v)$ for all $v \in S$.

Initialization: Initially, $S = \emptyset$, so trivially true.

Termination: At end, $Q = \emptyset \Rightarrow S = V \Rightarrow v.d = \delta(s, v)$ for all $v \in V$.

Maintenance: Need to show that $u.d = \delta(s, u)$ when u is added to S in each iteration.

Suppose there exists u such that $u.d \neq \delta(s, u)$. Without loss of generality, let u be the first vertex for which $u.d \neq \delta(s, u)$ when u is added to S.

Observations:

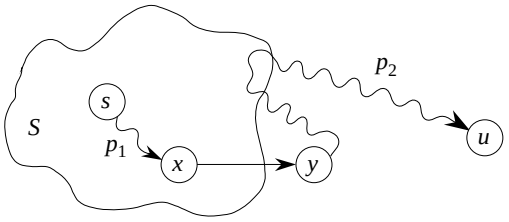
- $u \neq s$, since $s.d = \delta(s, s) = 0$.
- Therefore, $s \in S$, so $S \neq \emptyset$.
- There must be some path $s \rightsquigarrow u$, since otherwise $u.d = \delta(s, u) = \infty$ by no-path property.

So, there's a path $s \rightsquigarrow u$.

Then there's a shortest path $s \overset{p}{\rightsquigarrow} u$.

Just before u is added to S, path p connects a vertex in S (i.e., s) to a vertex in $V - S$ (i.e., u).

Let y be first vertex along p that's in $V - S$, and let $x \in S$ be y's predecessor.



Decompose p into $s \overset{p_1}{\rightsquigarrow} x \rightarrow y \overset{p_2}{\rightsquigarrow} u$. (Could have $x = s$ or $y = u$, so that p_1 or p_2 may have no edges.)

Claim

$y.d = \delta(s, y)$ when u is added to S.

Proof $x \in S$ and u is the first vertex such that $u.d \neq \delta(s, u)$ when u is added to $S \Rightarrow x.d = \delta(s, x)$ when x is added to S. Relaxed (x, y) at that time, so by the convergence property, $y.d = \delta(s, y)$. ■ (claim)

Now can get a contradiction to $u.d \neq \delta(s, u)$:

y is on shortest path $s \rightsquigarrow u$, and all edge weights are nonnegative
$\Rightarrow \delta(s, y) \leq \delta(s, u) \Rightarrow$

$$\begin{aligned} y.d &= \delta(s, y) \\ &\leq \delta(s, u) \\ &\leq u.d \qquad \text{(upper-bound property)} . \end{aligned}$$

Also, both y and u were in Q when we chose u, so

$u.d \le y.d \Rightarrow u.d = y.d$.

Therefore, $y.d = \delta(s, y) = \delta(s, u) = u.d$.

Contradicts assumption that $u.d \ne \delta(s, u)$. Hence, Dijkstra's algorithm is correct. ■

Analysis

Like Prim's algorithm, depends on implementation of priority queue.

- If binary heap, each operation takes $O(\lg V)$ time $\Rightarrow O(E \lg V)$.
- If a Fibonacci heap:
  - Each EXTRACT-MIN takes $O(1)$ amortized time.
  - There are $O(V)$ other operations, taking $O(\lg V)$ amortized time each.
  - Therefore, time is $O(V \lg V + E)$.

Difference constraints

Given a set of inequalities of the form $x_j - x_i \le b_k$.

- x's are variables, $1 \le i, j \le n$,
- b's are constants, $1 \le k \le m$.

Want to find a set of values for the x's that satisfy all m inequalities, or determine that no such values exist. Call such a set of values a ***feasible solution***.

Example

$$\begin{aligned} x_1 - x_2 &\le 5 \\ x_1 - x_3 &\le 6 \\ x_2 - x_4 &\le -1 \\ x_3 - x_4 &\le -2 \\ x_4 - x_1 &\le -3 \end{aligned}$$

Solution: $x = (0, -4, -5, -3)$

Also: $x = (5, 1, 0, 2) =$ [above solution] $+ 5$

Lemma

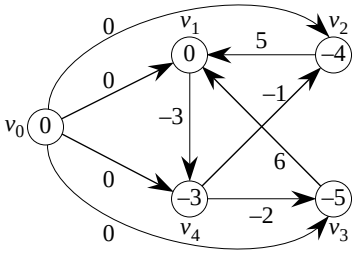
If x is a feasible solution, then so is $x + d$ for any constant d.

Proof x is a feasible solution $\Rightarrow x_j - x_i \le b_k$ for all i, j, k
$\Rightarrow (x_j + d) - (x_i + d) \le b_k$. ■ (lemma)

Constraint graph

$G = (V, E)$, weighted, directed.

- $V = \{v_0, v_1, v_2, \ldots, v_n\}$: one vertex per variable + v_0
- $E = \{(v_i, v_j) : x_j - x_i \le b_k \text{ is a constraint}\} \cup \{(v_0, v_1), (v_0, v_2), \ldots, (v_0, v_n)\}$
- $w(v_0, v_j) = 0$ for all j
- $w(v_i, v_j) = b_k$ if $x_j - x_i \le b_k$



Theorem

Given a system of difference constraints, let $G = (V, E)$ be the corresponding constraint graph.

1. If G has no negative-weight cycles, then
   $x = (\delta(v_0, v_1), \delta(v_0, v_2), \ldots, \delta(v_0, v_n))$
   is a feasible solution.
2. If G has a negative-weight cycle, then there is no feasible solution.

Proof

1. Show no negative-weight cycles $\Rightarrow$ feasible solution.
   Need to show that $x_j - x_i \le b_k$ for all constraints. Use
   $x_j = \delta(v_0, v_j)$
   $x_i = \delta(v_0, v_i)$
   $b_k = w(v_i, v_j)$.
   By the triangle inequality,
   $\delta(v_0, v_j) \le \delta(v_0, v_i) + w(v_i, v_j)$
   $x_j \le x_i + b_k$
   $x_j - x_i \le b_k$.
   Therefore, feasible.
2. Show negative-weight cycles $\Rightarrow$ no feasible solution.
   Without loss of generality, let a negative-weight cycle be $c = \langle v_1, v_2, \ldots, v_k \rangle$, where $v_1 = v_k$. (v_0 can't be on c, since v_0 has no entering edges.) c corresponds to the constraints

$$\begin{aligned} x_2 - x_1 &\le w(v_1, v_2) , \\ x_3 - x_2 &\le w(v_2, v_3) , \\ &\vdots \\ x_{k-1} - x_{k-2} &\le w(v_{k-2}, v_{k-1}) , \\ x_k - x_{k-1} &\le w(v_{k-1}, v_k) . \end{aligned}$$

If x is a solution satisfying these inequalities, it must satisfy their sum.
So add them up.
Each x_i is added once and subtracted once. ($v_1 = v_k \Rightarrow x_1 = x_k$.)
We get $0 \le w(c)$.
But $w(c) < 0$, since c is a negative-weight cycle.
Contradiction $\Rightarrow$ no such feasible solution x exists. ■ (theorem)

How to find a feasible solution

1. Form constraint graph.
   - $n + 1$ vertices.
   - $m + n$ edges.
   - $\Theta(m + n)$ time.
2. Run BELLMAN-FORD from v_0.
   - $O((n + 1)(m + n)) = O(n^2 + nm)$ time.
3. If BELLMAN-FORD returns FALSE $\Rightarrow$ no feasible solution.
   If BELLMAN-FORD returns TRUE $\Rightarrow$ set $x_i = \delta(v_0, v_i)$ for all i.

Solutions for Chapter 24: Single-Source Shortest Paths

Solution to Exercise 24.1-3

This solution is also posted publicly

If the greatest number of edges on any shortest path from the source is m, then the path-relaxation property tells us that after m iterations of BELLMAN-FORD, every vertex v has achieved its shortest-path weight in $v.d$. By the upper-bound property, after m iterations, no d values will ever change. Therefore, no d values will change in the $(m+1)$st iteration. Because we do not know m in advance, we cannot make the algorithm iterate exactly m times and then terminate. But if we just make the algorithm stop when nothing changes any more, it will stop after $m+1$ iterations.

```
BELLMAN-FORD-(M+1)(G, w, s)
  INITIALIZE-SINGLE-SOURCE(G, s)
  changes = TRUE
  while changes == TRUE
      changes = FALSE
      for each edge (u, v) ∈ G.E
          RELAX-M(u, v, w)

RELAX-M(u, v, w)
  if v.d > u.d + w(u, v)
      v.d = u.d + w(u, v)
      v.π = u
      changes = TRUE
```

The test for a negative-weight cycle (based on there being a d value that would change if another relaxation step was done) has been removed above, because this version of the algorithm will never get out of the **while** loop unless all d values stop changing.

Solution to Exercise 24.2-3

Instead of modifying the DAG-SHORTEST-PATHS procedure, we'll modify the structure of the graph so that we can run DAG-SHORTEST-PATHS on it. In fact,

we'll give two ways to transform a PERT chart $G = (V, E)$ with weights on vertices to a PERT chart $G' = (V', E')$ with weights on edges. In each way, we'll have that $|V'| \leq 2|V|$ and $|E'| \leq |V| + |E|$. We can then run on G' the same algorithm to find a longest path through a dag as is given in Section 24.2 of the text.

In the first way, we transform each vertex $v \in V$ into two vertices v' and v'' in V'. All edges in E that enter v will enter v' in E', and all edges in E that leave v will leave v'' in E'. In other words, if $(u, v) \in E$, then $(u'', v') \in E'$. All such edges have weight 0. We also put edges (v', v'') into E' for all vertices $v \in V$, and these edges are given the weight of the corresponding vertex v in G. Thus, $|V'| = 2|V|$, $|E'| = |V| + |E|$, and the edge weight of each path in G' equals the vertex weight of the corresponding path in G.

In the second way, we leave vertices in V alone, but we add one new source vertex s to V', so that $V' = V \cup \{s\}$. All edges of E are in E', and E' also includes an edge (s, v) for every vertex $v \in V$ that has in-degree 0 in G. Thus, the only vertex with in-degree 0 in G' is the new source s. The weight of edge $(u, v) \in E'$ is the weight of vertex v in G. In other words, the weight of each entering edge in G' is the weight of the vertex it enters in G. In effect, we have "pushed back" the weight of each vertex onto the edges that enter it. Here, $|V'| = |V| + 1$, $|E'| \leq |V| + |E|$ (since no more than $|V|$ vertices have in-degree 0 in G), and again the edge weight of each path in G' equals the vertex weight of the corresponding path in G.

Solution to Exercise 24.3-3

This solution is also posted publicly

Yes, the algorithm still works. Let u be the leftover vertex that does not get extracted from the priority queue Q. If u is not reachable from s, then $u.d = \delta(s, u) = \infty$. If u is reachable from s, then there is a shortest path $p = s \rightsquigarrow x \rightarrow u$. When the vertex x was extracted, $x.d = \delta(s, x)$ and then the edge (x, u) was relaxed; thus, $u.d = \delta(s, u)$.

Solution to Exercise 24.3-4

1. Verify that $s.d = 0$ and $s.\pi = \text{NIL}$.
2. Verify that $v.d = v.\pi.+w(v.\pi, v)$ for all $v \neq s$.
3. Verify that $v.d = \infty$ if and only if $v.ß = \text{NIL}$ for all $v \neq s$.
4. If any of the above verification tests fail, declare the output to be incorrect. Otherwise, run one pass of Bellman-Ford, i.e., relax each edge $(u, v) \in E$ one time. If any values of $v.d$ change, then declare the output to be incorrect; otherwise, declare the output to be correct.

Solution to Exercise 24.3-5

Let the graph have vertices s, x, y, z and edges $(s,x), (x,y), (y,z), (s,y)$, and let every edge have weight 0. Dijkstra's algorithm could relax edges in the order $(s,y), (s,x), (y,z), (x,y)$. The graph has two shortest paths from s to z: $\langle s, x, y, z\rangle$ and $\langle s, y, z\rangle$, both with weight 0. The edges on the shortest path $\langle s, x, y, z\rangle$ are relaxed out of order, because (x,y) is relaxed after (y,z).

Solution to Exercise 24.3-6

This solution is also posted publicly

To find the most reliable path between s and t, run Dijkstra's algorithm with edge weights $w(u,v) = -\lg r(u,v)$ to find shortest paths from s in $O(E+V\lg V)$ time. The most reliable path is the shortest path from s to t, and that path's reliability is the product of the reliabilities of its edges.

Here's why this method works. Because the probabilities are independent, the probability that a path will not fail is the product of the probabilities that its edges will not fail. We want to find a path $s \overset{p}{\rightsquigarrow} t$ such that $\prod_{(u,v)\in p} r(u,v)$ is maximized. This is equivalent to maximizing $\lg(\prod_{(u,v)\in p} r(u,v)) = \sum_{(u,v)\in p} \lg r(u,v)$, which is in turn equivalent to minimizing $\sum_{(u,v)\in p} -\lg r(u,v)$. (Note: $r(u,v)$ can be 0, and $\lg 0$ is undefined. So in this algorithm, define $\lg 0 = -\infty$.) Thus if we assign weights $w(u,v) = -\lg r(u,v)$, we have a shortest-path problem.

Since $\lg 1 = 0$, $\lg x < 0$ for $0 < x < 1$, and we have defined $\lg 0 = -\infty$, all the weights w are nonnegative, and we can use Dijkstra's algorithm to find the shortest paths from s in $O(E + V\lg V)$ time.

Alternative solution

You can also work with the original probabilities by running a modified version of Dijkstra's algorithm that maximizes the product of reliabilities along a path instead of minimizing the sum of weights along a path.

In Dijkstra's algorithm, use the reliabilities as edge weights and substitute

- max (and EXTRACT-MAX) for min (and EXTRACT-MIN) in relaxation and the queue,
- $\cdot$ for $+$ in relaxation,
- 1 (identity for $\cdot$) for 0 (identity for $+$) and $-\infty$ (identity for min) for ∞ (identity for max).

For example, we would use the following instead of the usual RELAX procedure:

```
RELAX-RELIABILITY(u, v, r)
  if v.d < u.d · r(u, v)
      v.d = u.d · r(u, v)
      v.π = u
```

This algorithm is isomorphic to the one above: it performs the same operations except that it is working with the original probabilities instead of the transformed ones.

Solution to Exercise 24.3-8

Observe that if a shortest-path estimate is not ∞, then it's at most $(|V| - 1)W$. Why? In order to have $v.d < \infty$, we must have relaxed an edge (u, v) with $u.d < \infty$. By induction, we can show that if we relax (u, v), then $v.d$ is at most the number of edges on a path from s to v times the maximum edge weight. Since any acyclic path has at most $|V| - 1$ edges and the maximum edge weight is W, we see that $v.d \leq (|V| - 1)W$. Note also that $v.d$ must also be an integer, unless it is ∞.

We also observe that in Dijkstra's algorithm, the values returned by the EXTRACT-MIN calls are monotonically increasing over time. Why? After we do our initial $|V|$ INSERT operations, we never do another. The only other way that a key value can change is by a DECREASE-KEY operation. Since edge weights are nonnegative, when we relax an edge (u, v), we have that $u.d \leq v.d$. Since u is the minimum vertex that we just extracted, we know that any other vertex we extract later has a key value that is at least $u.d$.

When keys are known to be integers in the range 0 to k and the key values extracted are monotonically increasing over time, we can implement a min-priority queue so that any sequence of m INSERT, EXTRACT-MIN, and DECREASE-KEY operations takes $O(m + k)$ time. Here's how. We use an array, say $A[0\,..\,k]$, where $A[j]$ is a linked list of each element whose key is j. Think of $A[j]$ as a bucket for all elements with key j. We implement each bucket by a circular, doubly linked list with a sentinel, so that we can insert into or delete from each bucket in $O(1)$ time. We perform the min-priority queue operations as follows:

- INSERT: To insert an element with key j, just insert it into the linked list in $A[j]$. Time: $O(1)$ per INSERT.
- EXTRACT-MIN: We maintain an index *min* of the value of the smallest key extracted. Initially, *min* is 0. To find the smallest key, look in $A[min]$ and, if this list is nonempty, use any element in it, removing the element from the list and returning it to the caller. Otherwise, we rely on the monotonicity property and increment *min* until we either find a list $A[min]$ that is nonempty (using any element in $A[min]$ as before) or we run off the end of the array A (in which case the min-priority queue is empty).

  Since there are at most m INSERT operations, there are at most m elements in the min-priority queue. We increment *min* at most k times, and we remove and return some element at most m times. Thus, the total time over all EXTRACT-MIN operations is $O(m + k)$.
- DECREASE-KEY: To decrease the key of an element from j to i, first check whether $i \leq j$, flagging an error if not. Otherwise, we remove the element from its list $A[j]$ in $O(1)$ time and insert it into the list $A[i]$ in $O(1)$ time. Time: $O(1)$ per DECREASE-KEY.

To apply this kind of min-priority queue to Dijkstra's algorithm, we need to let $k = (|V| - 1)W$, and we also need a separate list for keys with value ∞. The number of operations m is $O(V + E)$ (since there are $|V|$ INSERT and $|V|$ EXTRACT-MIN operations and at most $|E|$ DECREASE-KEY operations), and so the total time is $O(V + E + VW) = O(VW + E)$.

Solution to Exercise 24.3-9

First, observe that at any time, there are at most $W + 2$ distinct key values in the priority queue. Why? A key value is either ∞ or it is not. Consider what happens whenever a key value $v.d$ becomes finite. It must have occurred due to the relaxation of an edge (u, v). At that time, u was being placed into S, and $u.d \le y.d$ for all vertices $y \in V - S$. After relaxing edge (u, v), we have $v.d \le u.d + W$. Since any other vertex $y \in V - S$ with $y.d < \infty$ also had its estimate changed by a relaxation of some edge x with $x.d \le u.d$, we must have $y.d \le x.d + W \le u.d + W$. Thus, at the time that we are relaxing edges from a vertex u, we must have, for all vertices $v \in V - S$, that $u.d \le v.d \le u.d + W$ or $v.d = \infty$. Since shortest-path estimates are integer values (except for ∞), at any given moment we have at most $W + 2$ different ones: $u.d, u.d + 1, u.d + 2, \ldots, u.d + W$ and ∞.

Therefore, we can maintain the min-priorty queue as a binary min-heap in which each node points to a doubly linked list of all vertices with a given key value. There are at most $W + 2$ nodes in the heap, and so EXTRACT-MIN runs in $O(\lg W)$ time. To perform DECREASE-KEY, we need to be able to find the heap node corresponding to a given key in $O(\lg W)$ time. We can do so in $O(1)$ time as follows. First, keep a pointer *inf* to the node containing all the ∞ keys. Second, maintain an array $loc[0\,..\,W]$, where $loc[i]$ points to the unique heap entry whose key value is congruent to $i \pmod{(W + 1)}$. As keys move around in the heap, we can update this array in $O(1)$ time per movement.

Alternatively, instead of using a binary min-heap, we could use a red-black tree. Now INSERT, DELETE, MINIMUM, and SEARCH—from which we can construct the priority-queue operations—each run in $O(\lg W)$ time.

Solution to Exercise 24.4-4

Let $\delta(u)$ be the shortest-path weight from s to u. Then we want to find $\delta(t)$.

δ must satisfy

$$\begin{aligned} \delta(s) &= 0 \\ \delta(v) - \delta(u) &\le w(u, v) \text{ for all } (u, v) \in E \qquad \text{(Lemma 24.10)} , \end{aligned}$$

where $w(u, v)$ is the weight of edge (u, v).

Thus $x_v = \delta(v)$ is a solution to

$$\begin{aligned} x_s &= 0 \\ x_v - x_u &\le w(u, v) . \end{aligned}$$

To turn this into a set of inequalities of the required form, replace $x_s = 0$ by $x_s \le 0$ and $-x_s \le 0$ (i.e., $x_s \ge 0$). The constraints are now

$$\begin{aligned} x_s &\le 0 , \\ -x_s &\le 0 , \\ x_v - x_u &\le w(u,v) , \end{aligned}$$

which still has $x_v = \delta(v)$ as a solution.

However, δ isn't the only solution to this set of inequalities. (For example, if all edge weights are nonnegative, all $x_i = 0$ is a solution.) To force $x_t = \delta(t)$ as required by the shortest-path problem, add the requirement to maximize (the objective function) x_t. This is correct because

- $\max(x_t) \ge \delta(t)$ because $x_t = \delta(t)$ is part of one solution to the set of inequalities,
- $\max(x_t) \le \delta(t)$ can be demonstrated by a technique similar to the proof of Theorem 24.9:

  Let p be a shortest path from s to t. Then by definition,

  $$\delta(t) = \sum_{(u,v)\in p} w(u,v) .$$

  But for each edge (u,v) we have the inequality $x_v - x_u \le w(u,v)$, so

  $$\delta(t) = \sum_{(u,v)\in p} w(u,v) \ge \sum_{(u,v)\in p} (x_v - x_u) = x_t - x_s .$$

  But $x_s = 0$, so $x_t \le \delta(t)$.

Note: Maximizing x_t subject to the above inequalities solves the single-pair shortest-path problem when t is reachable from s and there are no negative-weight cycles. But if there's a negative-weight cycle, the inequalities have no feasible solution (as demonstrated in the proof of Theorem 24.9); and if t is not reachable from s, then x_t is unbounded.

Solution to Exercise 24.4-7

This solution is also posted publicly

Observe that after the first pass, all d values are at most 0, and that relaxing edges (v_0, v_i) will never again change a d value. Therefore, we can eliminate v_0 by running the Bellman-Ford algorithm on the constraint graph without the v_0 vertex but initializing all shortest path estimates to 0 instead of ∞.

Solution to Exercise 24.4-10

To allow for single-variable constraints, we add the variable x_0 and let it correspond to the source vertex v_0 of the constraint graph. The idea is that, if there are no

negative-weight cycles containing v_0, we will find that $\delta(v_0, v_0) = 0$. In this case, we set $x_0 = 0$, and so we can treat any single-variable constraint using x_i as if it were a 2-variable constraint with x_0 as the other variable.

Specifically, we treat the constraint $x_i \le b_k$ as if it were $x_i - x_0 \le b_k$, and we add the edge (v_0, v_i) with weight b_k to the constraint graph. We treat the constraint $-x_i \le b_k$ as if it were $x_0 - x_i \le b_k$, and we add the edge (v_i, v_0) with weight b_k to the constraint graph.

Once we find shortest-path weights from v_0, we set $x_i = \delta(v_0, v_i)$ for all $i = 0, 1, \ldots, n$; that is, we do as before but also include x_0 as one of the variables that we set to a shortest-path weight. Since v_0 is the source vertex, either $x_0 = 0$ or $x_0 < 0$.

If $\delta(v_0, v_0) = 0$, so that $x_0 = 0$, then setting $x_i = \delta(v_0, v_i)$ for all $i = 0, 1, \ldots, n$ gives a feasible solution for the system. The only new constraints beyond those in the text are those involving x_0. For constraints $x_i \le b_k$, we use $x_i - x_0 \le b_k$. By the triangle inequality, $\delta(v_0, v_i) \le \delta(v_0, v_0) + w(v_0, v_i) = b_k$, and so $x_i \le b_k$. For constraints $-x_i \le b_k$, we use $x_0 - x_i \le b_k$. By the triangle inequality, $0 = \delta(v_0, v_0) \le \delta(v_0, v_i) + w(v_i, v_0)$; thus, $0 \le x_i + b_k$ or, equivalently, $-x_i \le b_k$.

If $\delta(v_0, v_0) < 0$, so that $x_0 < 0$, then there is a negative-weight cycle containing v_0. The portion of the proof of Theorem 24.9 that deals with negative-weight cycles carries through but with v_0 on the negative-weight cycle, and we see that there is no feasible solution.

Solution to Exercise 24.5-4

This solution is also posted publicly

Whenever RELAX sets π for some vertex, it also reduces the vertex's d value. Thus if $s.\pi$ gets set to a non-NIL value, $s.d$ is reduced from its initial value of 0 to a negative number. But $s.d$ is the weight of some path from s to s, which is a cycle including s. Thus, there is a negative-weight cycle.

Solution to Exercise 24.5-7

Suppose we have a shortest-paths tree G_π. Relax edges in G_π according to the order in which a BFS would visit them. Then we are guaranteed that the edges along each shortest path are relaxed in order. By the path-relaxation property, we would then have $v.d = \delta(s, v)$ for all $v \in V$. Since G_π contains at most $|V| - 1$ edges, we need to relax only $|V| - 1$ edges to get $v.d = \delta(s, v)$ for all $v \in V$.

Solution to Exercise 24.5-8

Suppose that there is a negative-weight cycle $c = \langle v_0, v_1, \ldots, v_k \rangle$, where $v_0 = v_k$, that is reachable from the source vertex s; thus, $w(c) < 0$. Without loss of general-

ity, c is simple. There must be an acyclic path from s to some vertex of c that uses no other vertices in c. Without loss of generality let this vertex of c be v_0, and let this path from s to v_0 be $p = \langle u_0, u_1, \ldots, u_l \rangle$, where $u_0 = s$ and $u_l = v_0 = v_k$. (It may be the case that $u_l = s$, in which case path p has no edges.) After the call to INITIALIZE-SINGLE-SOURCE sets $v.d = \infty$ for all $v \in V - \{s\}$, perform the following sequence of relaxations. First, relax every edge in path p, in order. Then relax every edge in cycle c, in order, and repeatedly relax the cycle. That is, we relax the edges $(u_0, u_1), (u_1, u_2), \ldots, (u_{l-1}, v_0), (v_0, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_0)$, $(v_0, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_0), (v_0, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_0), \ldots$.

We claim that every edge relaxation in this sequence reduces a shortest-path estimate. Clearly, the first time we relax an edge (u_{i-1}, u_i) or (v_{j-1}, v_j), for $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, k-1$ (note that we have not yet relaxed the last edge of cycle c), we reduce $u_i.d$ or $v_j.d$ from ∞ to a finite value. Now consider the relaxation of any edge (v_{j-1}, v_j) after this opening sequence of relaxations. We use induction on the number of edge relaxations to show that this relaxation reduces $v_j.d$.

Basis: The next edge relaxed after the opening sequence is (v_{k-1}, v_k). Before relaxation, $v_k.d = w(p)$, and after relaxation, $v_k.d = w(p) + w(c) < w(p)$, since $w(c) < 0$.

Inductive step: Consider the relaxation of edge (v_{j-1}, v_j). Since c is a simple cycle, the last time $v_j.d$ was updated was by a relaxation of this same edge. By the inductive hypothesis, $v_{j-1}.d$ has just been reduced. Thus, $v_{j-1}.d + w(v_{j-1}, v_j) < v_j.d$, and so the relaxation will reduce the value of $v_j.d$.

Solution to Problem 24-1

a. Assume for the purpose contradiction that G_f is not acyclic; thus G_f has a cycle. A cycle must have at least one edge (u, v) in which u has higher index than v. This edge is not in E_f (by the definition of E_f), in contradition to the assumption that G_f has a cycle. Thus G_f is acyclic.

The sequence $\langle v_1, v_2, \ldots, v_{|V|} \rangle$ is a topological sort for G_f, because from the definition of E_f we know that all edges are directed from smaller indices to larger indices.

The proof for E_b is similar.

b. For all vertices $v \in V$, we know that either $\delta(s, v) = \infty$ or $\delta(s, v)$ is finite. If $\delta(s, v) = \infty$, then $v.d$ will be ∞. Thus, we need to consider only the case where $v.d$ is finite. There must be some shortest path from s to v. Let $p = \langle v_0, v_1, \ldots, v_{k-1}, v_k \rangle$ be that path, where $v_0 = s$ and $v_k = v$. Let us now consider how many times there is a change in direction in p, that is, a situation in which $(v_{i-1}, v_i) \in E_f$ and $(v_i, v_{i+1}) \in E_b$ or vice versa. There can be at most $|V|-1$ edges in p, so there can be at most $|V|-2$ changes in direction. Any portion of the path where there is no change in direction is computed with the correct d values in the first or second half of a single pass once the vertex that begins the no-change-in-direction sequence has the correct d value, because the edges are relaxed in the order of the direction of the sequence. Each change in

direction requires a half pass in the new direction of the path. The following table shows the maximum number of passes needed depending on the parity of $|V| - 1$ and the direction of the first edge:

$\lvert V\rvert - 1$	first edge direction	passes
even	forward	$(\lvert V\rvert - 1)/2$
even	backward	$(\lvert V\rvert - 1)/2 + 1$
odd	forward	$\lvert V\rvert/2$
odd	backward	$\lvert V\rvert/2$

In any case, the maximum number of passes that we will need is $\lceil |V|/2 \rceil$.

c. This scheme does not affect the asymptotic running time of the algorithm because even though we perform only $\lceil |V|/2 \rceil$ passes instead of $|V| - 1$ passes, it is still $O(V)$ passes. Each pass still takes $\Theta(E)$ time, so the running time remains $O(VE)$.

Solution to Problem 24-2

a. Consider boxes with dimensions $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d)$, and $z = (z_1, \ldots, z_d)$. Suppose there exists a permutation π such that $x_{\pi(i)} < y_i$ for $i = 1, \ldots, d$ and there exists a permutation π' such that $y_{\pi'(i)} < z_i$ for $i = 1, \ldots, d$, so that x nests inside y and y nests inside z. Construct a permutation π'', where $\pi''(i) = \pi'(\pi(i))$. Then for $i = 1, \ldots, d$, we have $x_{\pi''(i)} = x_{\pi'(\pi(i))} < y_{\pi'(i)} < z_i$, and so x nests inside z.

b. Sort the dimensions of each box from longest to shortest. A box X with sorted dimensions $(x_1, x_2, \ldots, x_d)$ nests inside a box Y with sorted dimensions $(y_1, y_2, \ldots, y_d)$ if and only if $x_i < y_i$ for $i = 1, 2, \ldots, d$. The sorting can be done in $O(d \lg d)$ time, and the test for nesting can be done in $O(d)$ time, and so the algorithm runs in $O(d \lg d)$ time. This algorithm works because a d-dimensional box can be oriented so that every permutation of its dimensions is possible. (Experiment with a 3-dimensional box if you are unsure of this).

c. Construct a dag $G = (V, E)$, where each vertex v_i corresponds to box B_i, and $(v_i, v_j) \in E$ if and only if box B_i nests inside box B_j. Graph G is indeed a dag, because nesting is transitive and antireflexive (i.e., no box nests inside itself). The time to construct the dag is $O(dn^2 + dn \lg d)$, from comparing each of the $\binom{n}{2}$ pairs of boxes after sorting the dimensions of each.

Add a supersource vertex s and a supersink vertex t to G, and add edges (s, v_i) for all vertices v_i with in-degree 0 and (v_j, t) for all vertices v_j with out-degree 0. Call the resulting dag G'. The time to do so is $O(n)$.

Find a longest path from s to t in G'. (Section 24.2 discusses how to find a longest path in a dag.) This path corresponds to a longest sequence of nesting boxes. The time to find a longest path is $O(n^2)$, since G' has $n + 2$ vertices and $O(n^2)$ edges.

Overall, this algorithm runs in $O(dn^2 + dn \lg d)$ time.

Solution to Problem 24-3

This solution is also posted publicly

a. We can use the Bellman-Ford algorithm on a suitable weighted, directed graph $G = (V, E)$, which we form as follows. There is one vertex in V for each currency, and for each pair of currencies c_i and c_j, there are directed edges (v_i, v_j) and (v_j, v_i). (Thus, $|V| = n$ and $|E| = n(n-1)$.)

We are looking for a cycle $\langle i_1, i_2, i_3, \ldots, i_k, i_1 \rangle$ such that

$$R[i_1, i_2] \cdot R[i_2, i_3] \cdots R[i_{k-1}, i_k] \cdot R[i_k, i_1] > 1 .$$

Taking logarithms of both sides of this inequality gives

$$\lg R[i_1, i_2] + \lg R[i_2, i_3] + \cdots + \lg R[i_{k-1}, i_k] + \lg R[i_k, i_1] > 0 .$$

If we negate both sides, we get

$$(-\lg R[i_1, i_2]) + (-\lg R[i_2, i_3]) + \cdots + (_{\lg} R[i_{k-1}, i_k]) + (-\lg R[i_k, i_1]) < 0 ,$$

and so we want to determine whether G contains a negative-weight cycle with these edge weights.

We can determine whether there exists a negative-weight cycle in G by adding an extra vertex v_0 with 0-weight edges (v_0, v_i) for all $v_i \in V$, running BELLMAN-FORD from v_0, and using the boolean result of BELLMAN-FORD (which is TRUE if there are no negative-weight cycles and FALSE if there is a negative-weight cycle) to guide our answer. That is, we invert the boolean result of BELLMAN-FORD.

This method works because adding the new vertex v_0 with 0-weight edges from v_0 to all other vertices cannot introduce any new cycles, yet it ensures that all negative-weight cycles are reachable from v_0.

It takes $\Theta(n^2)$ time to create G, which has $\Theta(n^2)$ edges. Then it takes $O(n^3)$ time to run BELLMAN-FORD. Thus, the total time is $O(n^3)$.

Another way to determine whether a negative-weight cycle exists is to create G and, without adding v_0 and its incident edges, run either of the all-pairs shortest-paths algorithms. If the resulting shortest-path distance matrix has any negative values on the diagonal, then there is a negative-weight cycle.

b. Note: The solution to this part also serves as a solution to Exercise 24.1-6.

Assuming that we ran BELLMAN-FORD to solve part (a), we only need to find the vertices of a negative-weight cycle. We can do so as follows. Go through the edges once again. Once we find an edge (u, v) for which $u.d + w(u, v) < v.d$, then we know that either vertex v is on a negative-weight cycle or is reachable from one. We can find a vertex on the negative-weight cycle by tracing back the π values from v, keeping track of which vertices we've visited until we reach a vertex x that we've visited before. Then we can trace back π values from x until we get back to x, and all vertices in between, along with x, will constitute a negative-weight cycle. We can use the recursive method given by the PRINT-PATH procedure of Section 22.2, but stop it when it returns to vertex x.

The running time is $O(n^3)$ to run BELLMAN-FORD, plus $O(m)$ to check all the edges and $O(n)$ to print the vertices of the cycle, for a total of $O(n^3)$ time.

Solution to Problem 24-4

a. Since all weights are nonnegative, use Dijkstra's algorithm. Implement the priority queue as an array $Q[0..|E|+1]$, where $Q[i]$ is a list of vertices v for which $v.d = i$. Initialize $v.d$ for $v \neq s$ to $|E|+1$ instead of to ∞, so that all vertices have a place in Q. (Any initial $v.d > \delta(s,v)$ works in the algorithm, since $v.d$ decreases until it reaches $\delta(s,v)$.)

The $|V|$ EXTRACT-MINs can be done in $O(E)$ total time, and decreasing a d value during relaxation can be done in $O(1)$ time, for a total running time of $O(E)$.

- When $v.d$ decreases, just add v to the front of the list in $Q[v.d]$.
- EXTRACT-MIN removes the head of the list in the first nonempty slot of Q. To do EXTRACT-MIN without scanning all of Q, keep track of the smallest i for which $Q[i]$ is not empty. The key point is that when $v.d$ decreases due to relaxation of edge (u,v), $v.d$ remains $\geq u.d$, so it never moves to an earlier slot of Q than the one that had u, the previous minimum. Thus EXTRACT-MIN can always scan upward in the array, taking a total of $O(E)$ time for all EXTRACT-MINs.

b. For all $(u,v) \in E$, we have $w_1(u,v) \in \{0,1\}$, so $\delta_1(s,v) \leq |V|-1 \leq |E|$. Use part (a) to get the $O(E)$ time bound.

c. To show that $w_i(u,v) = 2w_{i-1}(u,v)$ or $w_i(u,v) = 2w_{i-1}(u,v)+1$, observe that the i bits of $w_i(u,v)$ consist of the $i-1$ bits of $w_{i-1}(u,v)$ followed by one more bit. If that low-order bit is 0, then $w_i(u,v) = 2w_{i-1}(u,v)$; if it is 1, then $w_i(u,v) = 2w_{i-1}(u,v)+1$.

Notice the following two properties of shortest paths:

1. If all edge weights are multiplied by a factor of c, then all shortest-path weights are multiplied by c.
2. If all edge weights are increased by at most c, then all shortest-path weights are increased by at most $c(|V|-1)$, since all shortest paths have at most $|V|-1$ edges.

The lowest possible value for $w_i(u,v)$ is $2w_{i-1}(u,v)$, so by the first observation, the lowest possible value for $\delta_i(s,v)$ is $2\delta_{i-1}(s,v)$.

The highest possible value for $w_i(u,v)$ is $2w_{i-1}(u,v)+1$. Therefore, using the two observations together, the highest possible value for $\delta_i(s,v)$ is $2\delta_{i-1}(s,v)+|V|-1$.

d. We have

$$\begin{aligned}\widehat{w}_i(u,v) &= w_i(u,v) + 2\delta_{i-1}(s,u) - 2\delta_{i-1}(s,v) \\ &\geq 2w_{i-1}(u,v) + 2\delta_{i-1}(s,u) - 2\delta_{i-1}(s,v) \\ &\geq 0\,.\end{aligned}$$

The second line follows from part (c), and the third line follows from Lemma 24.10: $\delta_{i-1}(s,v) \le \delta_{i-1}(s,u) + w_{i-1}(u,v)$.

e. Observe that if we compute $\widehat{w}_i(p)$ for any path $p : u \rightsquigarrow v$, the terms $\delta_{i-1}(s,t)$ cancel for every intermediate vertex t on the path. Thus,

$$\widehat{w}_i(p) = w_i(p) + 2\delta_{i-1}(s,u) - 2\delta_{i-1}(s,v) \ .$$

(This relationship will be shown in detail in equation (25.10) within the proof of Lemma 25.1.) The δ_{i-1} terms depend only on u, v, and s, but not on the path p; therefore the same paths will be of minimum w_i weight and of minimum $\widehat{w}_i$ weight between u and v. Letting $u = s$, we get

$$\begin{aligned}\widehat{\delta}_i(s,v) &= \delta_i(s,v) + 2\delta_{i-1}(s,s) - 2\delta_{i-1}(s,v) \\ &= \delta_i(s,v) - 2\delta_{i-1}(s,v) \ .\end{aligned}$$

Rewriting this result as $\delta_i(s,v) = \widehat{\delta}_i(s,v) + 2\delta_{i-1}(s,v)$ and combining it with $\delta_i(s,v) \le 2\delta_{i-1}(s,v) + |V| - 1$ (from part (c)) gives us $\widehat{\delta}_i(s,v) \le |V| - 1 \le |E|$.

f. To compute $\delta_i(s,v)$ from $\delta_{i-1}(s,v)$ for all $v \in V$ in $O(E)$ time:

1. Compute the weights $\widehat{w}_i(u,v)$ in $O(E)$ time, as shown in part (d).
2. By part (e), $\widehat{\delta}_i(s,v) \le |E|$, so use part (a) to compute all $\widehat{\delta}_i(s,v)$ in $O(E)$ time.
3. Compute all $\delta_i(s,v)$ from $\widehat{\delta}_i(s,v)$ and $\delta_{i-1}(s,v)$ as shown in part (e), in $O(V)$ time.

To compute all $\delta(s,v)$ in $O(E \lg W)$ time:

1. Compute $\delta_1(s,v)$ for all $v \in V$. As shown in part (b), this takes $O(E)$ time.
2. For each $i = 2, 3, \ldots, k$, compute all $\delta_i(s,v)$ from $\delta_{i-1}(s,v)$ in $O(E)$ time as shown above. This procedure computes $\delta(s,v) = \delta_k(u,v)$ in time $O(Ek) = O(E \lg W)$.

Solution to Problem 24-6

Observe that a bitonic sequence can increase, then decrease, then increase, or it can decrease, then increase, then decrease. That is, there can be at most two changes of direction in a bitonic sequence. Any sequence that increases, then decreases, then increases, then decreases has a bitonic sequence as a subsequence.

Now, let us suppose that we had an even stronger condition than the bitonic property given in the problem: for each vertex $v \in V$, the weights of the edges along any shortest path from s to v are increasing. Then we could call INITIALIZE-SINGLE-SOURCE and then just relax all edges one time, going in increasing order of weight. Then the edges along every shortest path would be relaxed in order of their appearance on the path. (We rely on the uniqueness of edge weights to ensure that the ordering is correct.) The path-relaxation property (Lemma 24.15) would guarantee that we would have computed correct shortest paths from s to each vertex.

If we weaken the condition so that the weights of the edges along any shortest path increase and then decrease, we could relax all edges one time, in increasing order of weight, and then one more time, in decreasing order of weight. That order, along with uniqueness of edge weights, would ensure that we had relaxed the edges of every shortest path in order, and again the path-relaxation property would guarantee that we would have computed correct shortest paths.

To make sure that we handle all bitonic sequences, we do as suggested above. That is, we perform four passes, relaxing each edge once in each pass. The first and third passes relax edges in increasing order of weight, and the second and fourth passes in decreasing order. Again, by the path-relaxation property and the uniqueness of edge weights, we have computed correct shortest paths.

The total time is $O(V + E \lg V)$, as follows. The time to sort $|E|$ edges by weight is $O(E \lg E) = O(E \lg V)$ (since $|E| = O(V^2)$). INITIALIZE-SINGLE-SOURCE takes $O(V)$ time. Each of the four passes takes $O(E)$ time. Thus, the total time is $O(E \lg V + V + E) = O(V + E \lg V)$.

Lecture Notes for Chapter 25: All-Pairs Shortest Paths

Chapter 25 overview

Given a directed graph $G = (V, E)$, weight function $w : E \to \mathbb{R}$, $|V| = n$. Assume that we can number the vertices $1, 2, \ldots, n$.

Goal: create an $n \times n$ matrix $D = (d_{ij})$ of shortest-path distances, so that $d_{ij} = \delta(i, j)$ for all vertices i and j.

Could run BELLMAN-FORD once from each vertex:

- $O(V^2E)$—which is $O(V^4)$ if the graph is ***dense*** ($E = \Theta(V^2)$).

If no negative-weight edges, could run Dijkstra's algorithm once from each vertex:

- $O(VE \lg V)$ with binary heap—$O(V^3 \lg V)$ if dense,
- $O(V^2 \lg V + VE)$ with Fibonacci heap—$O(V^3)$ if dense.

We'll see how to do in $O(V^3)$ in all cases, with no fancy data structure.

Shortest paths and matrix multiplication

Assume that G is given as adjacency matrix of weights: $W = (w_{ij})$, with vertices numbered 1 to n.

$$w_{ij} = \begin{cases} 0 & \text{if } i = j \ , \\ \text{weight of } (i, j) & \text{if } i \neq j, (i, j) \in E \ , \\ \infty & \text{if } i \neq j, (i, j) \notin E \ . \end{cases}$$

Won't worry about predecessors—see book.

Will use dynamic programming at first.

Optimal substructure

Recall: subpaths of shortest paths are shortest paths.

Recursive solution

Let $l_{ij}^{(m)}$ = weight of shortest path $i \rightsquigarrow j$ that contains $\le m$ edges.

- $m = 0$
  $\Rightarrow$ there is a shortest path $i \rightsquigarrow j$ with $\le m$ edges if and only if $i = j$
  $\Rightarrow l_{ij}^{(0)} = \begin{cases} 0 & \text{if } i = j \ , \\ \infty & \text{if } i \ne j \ . \end{cases}$
- $m \ge 1$
  $$\begin{aligned} \Rightarrow l_{ij}^{(m)} &= \min\left(l_{ij}^{(m-1)} \min_{1 \le k \le n} \left\{l_{ik}^{(m-1)} + w_{kj}\right\}\right) && (k \text{ ranges over all possible predecessors of } j) \\ &= \min_{1 \le k \le n} \left\{l_{ik}^{(m-1)} + w_{kj}\right\} && (\text{since } w_{jj} = 0 \text{ for all } j) \ . \end{aligned}$$
- Observe that when $m = 1$, must have $l_{ij}^{(1)} = w_{ij}$.

  Conceptually, when the path is restricted to at most 1 edge, the weight of the shortest path $i \rightsquigarrow j$ must be w_{ij}.

  And the math works out, too:
  $$\begin{aligned} l_{ij}^{(1)} &= \min_{1 \le k \le n} \left\{l_{ik}^{(0)} + w_{kj}\right\} \\ &= l_{ii}^{(0)} + w_{ij} && (l_{ii}^{(0)} \text{ is the only non-}\infty \text{ among } l_{ik}^{(0)}) \\ &= w_{ij} \ . \end{aligned}$$

All simple shortest paths contain $\le n - 1$ edges
$\Rightarrow \delta(i, j) = l_{ij}^{(n-1)} = l_{ij}^{(n)} = l_{ij}^{(n+1)} = \ldots$

Compute a solution bottom-up

Compute $L^{(1)}, L^{(2)}, \ldots, L^{(n-1)}$.

Start with $L^{(1)} = W$, since $l_{ij}^{(1)} = w_{ij}$.

Go from $L^{(m-1)}$ to $L^{(m)}$:

```
EXTEND(L, W, n)
  let L' = (l'_ij) be a new n × n matrix
  for i = 1 to n
      for j = 1 to n
          l'_ij = ∞
          for k = 1 to n
              l'_ij = min(l'_ij, l_ik + w_kj)
  return L'
```

Compute each $L^{(m)}$:

```
SLOW-APSP(W, n)
  L^(1) = W
  for m = 2 to n − 1
      let L^(m) be a new n × n matrix
      L^(m) = EXTEND(L^(m−1), W, n)
  return L^(n−1)
```

Time

- EXTEND: $\Theta(n^3)$.
- SLOW-APSP: $\Theta(n^4)$.

Observation

EXTEND is like matrix multiplication:

$L \rightarrow A$
$W \rightarrow B$
$L' \rightarrow C$
$\min \rightarrow +$
$+ \rightarrow \cdot$
$\infty \rightarrow 0$

```
let C be an n × n matrix
for i = 1 to n
    for j = 1 to n
        c_ij = 0
        for k = 1 to n
            c_ij = c_ij + a_ik · b_kj
return C
```

So, we can view EXTEND as just like matrix multiplication!

Why do we care?

Because our goal is to compute $L^{(n-1)}$ as fast as we can. Don't need to compute *all* the intermediate $L^{(1)}, L^{(2)}, L^{(3)}, \ldots, L^{(n-2)}$.

Suppose we had a matrix A and we wanted to compute A^{n-1} (like calling EXTEND $n-1$ times).

Could compute $A, A^2, A^4, A^8, \ldots$

If we knew $A^m = A^{n-1}$ for all $m \geq n-1$, could just finish with A^r, where r is the smallest power of 2 that's $\geq n-1$. ($r = 2^{\lceil \lg(n-1) \rceil}$)

FASTER-APSP(W, n)

```
L^(1) = W
m = 1
while m < n − 1
    let L^(2m) be a new n × n matrix
    L^(2m) = EXTEND(L^(m), L^(m), n)
    m = 2m
return L^(m)
```

OK to overshoot, since products don't change after $L^{(n-1)}$.

Time

$\Theta(n^3 \lg n)$.

Floyd-Warshall algorithm

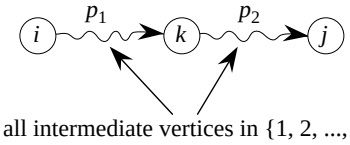
A different dynamic-programming approach.

For path $p = \langle v_1, v_2, \ldots, v_l \rangle$, an ***intermediate vertex*** is any vertex of p other than v_1 or v_l.

Let $d_{ij}^{(k)}$ = shortest-path weight of any path $i \rightsquigarrow j$ with all intermediate vertices in $\{1, 2, \ldots, k\}$.

Consider a shortest path $i \overset{p}{\rightsquigarrow} j$ with all intermediate vertices in $\{1, 2, \ldots, k\}$:

- If k is not an intermediate vertex, then all intermediate vertices of p are in $\{1, 2, \ldots, k-1\}$.
- If k is an intermediate vertex:



Recursive formulation

$$d_{ij}^{(k)} = \begin{cases} w_{ij} & \text{if } k = 0\,, \\ \min\left(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}\right) & \text{if } k \geq 1\,. \end{cases}$$

(Have $d_{ij}^{(0)} = w_{ij}$ because can't have intermediate vertices $\Rightarrow \leq 1$ edge.)

Want $D^{(n)} = \left(d_{ij}^{(n)}\right)$, since all vertices numbered $\leq n$.

Compute bottom-up

Compute in increasing order of k:

FLOYD-WARSHALL(W, n)
  $D^{(0)} = W$
  for $k = 1$ **to** n
      let $D^{(k)} = \left(d_{ij}^{(k)}\right)$ be a new $n \times n$ matrix
      for $i = 1$ **to** n
          for $j = 1$ **to** n
              $d_{ij}^{(k)} = \min\left(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}\right)$
  return $D^{(n)}$

Can drop superscripts. (See Exercise 25.2-4 in text.)

Time

$\Theta(n^3)$.

Transitive closure

Given $G = (V, E)$, directed.

Compute $G^* = (V, E^*)$.

- $E^* = \{(i, j) : \text{there is a path } i \rightsquigarrow j \text{ in } G\}$.

Could assign weight of 1 to each edge, then run FLOYD-WARSHALL.

- If $d_{ij} < n$, then there is a path $i \rightsquigarrow j$.
- Otherwise, $d_{ij} = \infty$ and there is no path.

Simpler way

Substitute other values and operators in FLOYD-WARSHALL.

- Use unweighted adjacency matrix
- min $\rightarrow \vee$ (OR)
- $+ \rightarrow \wedge$ (AND)
- $t_{ij}^{(k)} = \begin{cases} 1 & \text{if there is path } i \rightsquigarrow j \text{ with all intermediate vertices in } \{1, 2, \ldots, k\}, \\ 0 & \text{otherwise.} \end{cases}$
- $t_{ij}^{(0)} = \begin{cases} 0 & \text{if } i \neq j \text{ and } (i, j) \notin E, \\ 1 & \text{if } i = j \text{ or } (i, j) \in E. \end{cases}$
- $t_{ij}^{(k)} = t_{ij}^{(k-1)} \vee \left(t_{ik}^{(k-1)} \wedge t_{kj}^{(k-1)}\right)$.

TRANSITIVE-CLOSURE(G, n)

```
n = |G.V|
let T^(0) = (t_ij^(0)) be a new n × n matrix
for i = 1 to n
    for j = 1 to n
        if i == j or (i, j) ∈ G.E
            t_ij^(0) = 1
        else t_ij^(0) = 0
for k = 1 to n
    let T^(k) = (t_ij^(k)) be a new n × n matrix
    for i = 1 to n
        for j = 1 to n
            t_ij^(k) = t_ij^(k-1) ∨ (t_ik^(k-1) ∧ t_kj^(k-1))
return T^(n)
```

Time

$\Theta(n^3)$, but simpler operations than FLOYD-WARSHALL.

Johnson's algorithm

Idea

If the graph is sparse, it pays to run Dijkstra's algorithm once from each vertex.

If we use a Fibonacci heap for the priority queue, the running time is down to $O(V^2 \lg V + VE)$, which is better than FLOYD-WARSHALL's $\Theta(V^3)$ time if $E = o(V^2)$.

But Dijkstra's algorithm requires that all edge weights be nonnegative.

Donald Johnson figured out how to make an equivalent graph that *does* have all edge weights ≥ 0.

Reweighting

Compute a new weight function $\hat{w}$ such that

1. For all $u, v \in V$, p is a shortest path $u \rightsquigarrow v$ using w if and only if p is a shortest path $u \rightsquigarrow v$ using $\hat{w}$.
2. For all $(u, v) \in E$, $\hat{w}(u, v) \geq 0$.

Property (1) says that it suffices to find shortest paths with $\hat{w}$. Property (2) says we can do so by running Dijkstra's algorithm from each vertex.

How to come up with $\hat{w}$?

Lemma shows it's easy to get property (1):

Lemma (Reweighting doesn't change shortest paths)
Given a directed, weighted graph $G = (V, E)$, $w : E \to \mathbb{R}$. Let h be any function such that $h : V \to \mathbb{R}$. For all $(u, v) \in E$, define

$$\hat{w}(u, v) = w(u, v) + h(u) - h(v) \ .$$

Let $p = \langle v_0, v_1, \ldots, v_k \rangle$ be any path $v_0 \rightsquigarrow v_k$.

Then p is a shortest path $v_0 \rightsquigarrow v_k$ with w if and only if p is a shortest path $v_0 \rightsquigarrow v_k$ with $\hat{w}$. (Formally, $w(p) = \delta(v_0, v_k)$ if and only if $\hat{w} = \hat{\delta}(v_0, v_k)$, where $\hat{\delta}$ is the shortest-path weight with $\hat{w}$.)

Also, G has a negative-weight cycle with w if and only if G has a negative-weight cycle with $\hat{w}$.

Proof First, we'll show that $\hat{w}(p) = w(p) + h(v_0) - h(v_k)$:

$$\begin{aligned}
\hat{w}(p) &= \sum_{i=1}^{k} \hat{w}(v_{i-1}, v_i) \\
&= \sum_{i=1}^{k} (w(v_{i-1}, v_i) + h(v_{i-1}) - h(v_i)) \\
&= \sum_{i=1}^{k} w(v_{i-1}, v_i) + h(v_0) - h(v_k) \qquad \text{(sum telescopes)} \\
&= w(p) + h(v_0) - h(v_k) \ .
\end{aligned}$$

Therefore, any path $v_0 \overset{p}{\rightsquigarrow} v_k$ has $\hat{w}(p) = w(p) + h(v_0) - h(v_k)$. Since $h(v_0)$ and $h(v_k)$ don't depend on the path from v_0 to v_k, if one path $v_0 \rightsquigarrow v_k$ is shorter than another with w, it's also shorter with $\hat{w}$.

Now show there exists a negative-weight cycle with w if and only if there exists a negative-weight cycle with $\hat{w}$:

- Let cycle $c = \langle v_0, v_1, \ldots, v_k \rangle$, where $v_0 = v_k$.
- Then

$$\begin{aligned} \hat{w}(c) &= w(c) + h(v_0) - h(v_k) \\ &= w(c) \qquad\qquad (\text{since } v_0 = v_k)\,. \end{aligned}$$

Therefore, c has a negative-weight cycle with w if and only if it has a negative-weight cycle with $\hat{w}$. ■ (lemma)

So, now to get property (2), we just need to come up with a function $h : V \to \mathbb{R}$ such that when we compute $\hat{w}(u, v) = w(u, v) + h(u) - h(v)$, it's ≥ 0.

Do what we did for difference constraints:

- $G' = (V', E')$
  - $V' = V \cup \{s\}$, where s is a new vertex.
  - $E' = E \cup \{(s, v) : v \in V\}$.
  - $w(s, v) = 0$ for all $v \in V$.
- Since no edges enter s, G' has the same set of cycles as G. In particular, G' has a negative-weight cycle if and only if G does.

Define $h(v) = \delta(s, v)$ for all $v \in V$.

Claim

$\hat{w}(u, v) = w(u, v) + h(u) - h(v) \geq 0$.

Proof By the triangle inequality,

$$\begin{aligned} \delta(s, v) &\leq \delta(s, u) + w(u, v) \\ h(v) &\leq h(u) + w(u, v)\,. \end{aligned}$$

Therefore, $w(u, v) + h(u) - h(v) \geq 0$. ■ (claim)

Johnson's algorithm

form G'
run BELLMAN-FORD on G' to compute $\delta(s, v)$ for all $v \in G'.V$
if BELLMAN-FORD returns FALSE
    G has a negative-weight cycle
else compute $\hat{w}(u, v) = w(u, v) + \delta(s, u) - \delta(s, v)$ for all $(u, v) \in E$
    let $D = (d_{uv})$ be a new $n \times n$ matrix
    for each vertex $u \in G.V$
        run Dijkstra's algorithm from u using weight function $\hat{w}$
            to compute $\hat{\delta}(u, v)$ for all $v \in V$
        for each vertex $v \in G.V$
            // Compute entry d_{uv} in matrix D.
            $d_{uv} = \underbrace{\hat{\delta}(u, v) + \delta(s, v) - \delta(s, u)}$
                because if p is a path $u \leadsto v$, then $\hat{w}(p) = w(p) + h(u) - h(v)$
    return D

Time

- $\Theta(V + E)$ to compute G'.
- $O(VE)$ to run BELLMAN-FORD.
- $\Theta(E)$ to compute $\hat{w}$.
- $O(V^2 \lg V + VE)$ to run Dijkstra's algorithm $|V|$ times (using Fibonacci heap).
- $\Theta(V^2)$ to compute D matrix.

Total: $O(V^2 \lg V + VE)$.

Solutions for Chapter 25: All-Pairs Shortest Paths

Solution to Exercise 25.1-3

This solution is also posted publicly

The matrix $L^{(0)}$ corresponds to the identity matrix

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

of regular matrix multiplication. Substitute 0 (the identity for $+$) for ∞ (the identity for min), and 1 (the identity for $\cdot$) for 0 (the identity for $+$).

Solution to Exercise 25.1-5

This solution is also posted publicly

The all-pairs shortest-paths algorithm in Section 25.1 computes

$L^{(n-1)} = W^{n-1} = L^{(0)} \cdot W^{n-1}$,

where $l_{ij}^{(n-1)} = \delta(i, j)$ and $L^{(0)}$ is the identity matrix. That is, the entry in the ith row and jth column of the matrix "product" is the shortest-path distance from vertex i to vertex j, and row i of the product is the solution to the single-source shortest-paths problem for vertex i.

Notice that in a matrix "product" $C = A \cdot B$, the ith row of C is the ith row of A "multiplied" by B. Since all we want is the ith row of C, we never need more than the ith row of A.

Thus the solution to the single-source shortest-paths from vertex i is $L_i^{(0)} \cdot W^{n-1}$, where $L_i^{(0)}$ is the ith row of $L^{(0)}$—a vector whose ith entry is 0 and whose other entries are ∞.

Doing the above "multiplications" starting from the left is essentially the same as the BELLMAN-FORD algorithm. The vector corresponds to the d values in BELLMAN-FORD—the shortest-path estimates from the source to each vertex.

- The vector is initially 0 for the source and ∞ for all other vertices, the same as the values set up for d by INITIALIZE-SINGLE-SOURCE.
- Each "multiplication" of the current vector by W relaxes all edges just as BELLMAN-FORD does. That is, a distance estimate in the row, say the distance to v, is updated to a smaller estimate, if any, formed by adding some $w(u, v)$ to the current estimate of the distance to u.
- The relaxation/multiplication is done $n - 1$ times.

Solution to Exercise 25.1-10

Run SLOW-ALL-PAIRS-SHORTEST-PATHS on the graph. Look at the diagonal elements of $L^{(m)}$. Return the first value of m for which one (or more) of the diagonal elements ($l_{ii}^{(m)}$) is negative. If m reaches $n + 1$, then stop and declare that there are no negative-weight cycles.

Let the number of edges in a minimum-length negative-weight cycle be m^*, where $m^* = \infty$ if the graph has no negative-weight cycles.

Correctness

Let's assume that for some value $m^* \leq n$ and some value of i, we find that $l_{ii}^{(m^*)} < 0$. Then the graph has a cycle with m^* edges that goes from vertex i to itself, and this cycle has negative weight (stored in $l_{ii}^{(m^*)}$). This is the minimum-length negative-weight cycle because SLOW-ALL-PAIRS-SHORTEST-PATHS computes all paths of 1 edge, then all paths of 2 edges, and so on, and all cycles shorter than m^* edges were checked before and did not have negative weight. Now assume that for all $m \leq n$, there is no negative $l_{ii}^{(m)}$ element. Then, there is no negative-weight cycle in the graph, because all cycles have length at most n.

Time

$O(n^4)$. More precisely, $\Theta(n^3 \cdot \min(n, m^*))$.

Faster solution

Run FASTER-ALL-PAIRS-SHORTEST-PATHS on the graph until the first time that the matrix $L^{(m)}$ has one or more negative values on the diagonal, or until we have computed $L^{(m)}$ for some $m > n$. If we find any negative entries on the diagonal, we know that the minimum-length negative-weight cycle has more than $m/2$ edges and at most m edges. We just need to binary search for the value of m^* in the range $m/2 < m^* \leq m$. The key observation is that on our way to computing $L^{(m)}$, we computed $L^{(1)}$, $L^{(2)}$, $L^{(4)}$, $L^{(8)}$, ..., $L^{(m/2)}$, and these matrices suffice to compute every matrix we'll need. Here's pseudocode:

```
FIND-MIN-LENGTH-NEG-WEIGHT-CYCLE(W)
  n = W.rows
  L^(1) = W
  m = 1
  while m ≤ n and no diagonal entries of L^(m) are negative
      L^(2m) = EXTEND-SHORTEST-PATHS(L^(m), L^(m))
      m = 2m
  if m > n and no diagonal entries of L^(m) are negative
      return "no negative-weight cycles"
  elseif m ≤ 2
      return m
  else
      low = m/2
      high = m
      d = m/4
      while d ≥ 1
          s = low + d
          L^(s) = EXTEND-SHORTEST-PATHS(L^(low), L^(d))
          if L^(s) has any negative entries on the diagonal
              high = s
          else low = s
          d = d/2
      return high
```

Correctness

If, after the first **while** loop, $m > n$ and no diagonal entries of $L^{(m)}$ are negative, then there is no negative-weight cycle. Otherwise, if $m \le 2$, then either $m = 1$ or $m = 2$, and $L^{(m)}$ is the first matrix with a negative entry on the diagonal. Thus, the correct value to return is m.

If $m > 2$, then we maintain an interval bracketed by the values *low* and *high*, such that the correct value m^* is in the range $low < m^* \le high$. We use the following loop invariant:

Loop invariant: At the start of each iteration of the "**while** $d \ge 1$" loop,

1. $d = 2^p$ for some integer $p \ge -1$,
2. $d = (high - low)/2$,
3. $low < m^* \le high$.

Initialization: Initially, m is an integer power of 2 and $m > 2$. Since $d = m/4$, we have that d is an integer power of 2 and $d > 1/2$, so that $d = 2^p$ for some integer $p \ge 0$. We also have $(high - low)/2 = (m - (m/2))/2 = m/4 = d$. Finally, $L^{(m)}$ has a negative entry on the diagonal and $L^{(m/2)}$ does not. Since $low = m/2$ and $high = m$, we have that $low < m^* \le high$.

Maintenance: We use *high*, *low*, and d to denote variable values in a given iteration, and $high'$, low', and d' to denote the same variable values in the next iteration. Thus, we wish to show that $d = 2^p$ for some integer $p \ge -1$ implies $d' = 2^{p'}$ for some integer $p' \ge -1$, that $d = (high - low)/2$ implies $d' = (high' - low')/2$, and that $low < m^* \le high$ implies $low' < m^* \le high'$.

To see that $d' = 2^{p'}$, note that $d' = d/2$, and so $d = 2^{p-1}$. The condition that $d \geq 1$ implies that $p \geq 0$, and so $p' \geq -1$.

Within each iteration, s is set to $low + d$, and one of the following actions occurs:

- If $L^{(s)}$ has any negative entries on the diagonal, then $high'$ is set to s and d' is set to $d/2$. Upon entering the next iteration, $(high' - low')/2 = (s - low')/2 = ((low + d) - low)/2 = d/2 = d'$. Since $L^{(s)}$ has a negative diagonal entry, we know that $m^* \leq s$. Because $high' = s$ and $low' = low$, we have that $low' < m^* \leq high'$.
- If $L^{(s)}$ has no negative entries on the diagonal, then low' is set to s, and d' is set to $d/2$. Upon entering the next iteration, $(high' - low')/2 = (high' - s)/2 = (high-(low+d))/2 = (high-low)/2-d/2 = d-d/2 = d/2 = d'$. Since $L^{(s)}$ has no negative diagonal entries, we know that $m^* > s$. Because $low' = s$ and $high' = high$, we have that $low' < m^* \leq high'$.

Termination: At termination, $d < 1$. Since $d = 2^p$ for some integer $p \geq -1$, we must have $p = -1$, so that $d = 1/2$. By the second part of the loop invariant, if we multiply both sides by 2, we get that $high - low = 2d = 1$. By the third part of the loop invariant, we know that $low < m^* \leq high$. Since $high - low = 2d = 1$ and $m^* > low$, the only possible value for m^* is $high$, which the procedure returns.

Time

If there is no negative-weight cycle, the first **while** loop iterates $\Theta(\lg n)$ times, and the total time is $\Theta(n^3 \lg n)$.

Now suppose that there is a negative-weight cycle. We claim that each time we call EXTEND-SHORTEST-PATHS$(L^{(low)}, L^{(d)})$, we have already computed $L^{(low)}$ and $L^{(d)}$. Initially, since $low = m/2$, we had already computed $L^{(low)}$ in the first **while** loop. In succeeding iterations of the second **while** loop, the only way that low changes is when it gets the value of s, and we have just computed $L^{(s)}$. As for $L^{(d)}$, observe that d takes on the values $m/4, m/8, m/16, \ldots, 1$, and again, we computed all of these L matrices in the first **while** loop. Thus, the claim is proven. Each of the two **while** loops iterates $\Theta(\lg m^*)$ times. Since we have already computed the parameters to each call of EXTEND-SHORTEST-PATHS, each iteration is dominated by the $\Theta(n^3)$-time call to EXTEND-SHORTEST-PATHS. Thus, the total time is $\Theta(n^3 \lg m^*)$.

In general, therefore, the running time is $\Theta(n^3 \lg \min(n, m^*))$.

Space

The slower algorithm needs to keep only three matrices at any time, and so its space requirement is $\Theta(n^3)$. This faster algorithm needs to maintain $\Theta(\lg \min(n, m^*))$ matrices, and so the space requirement increases to $\Theta(n^3 \lg \min(n, m^*))$.

Solution to Exercise 25.2-4

This solution is also posted publicly

With the superscripts, the computation is $d_{ij}^{(k)} = \min\left(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}\right)$. If, having dropped the superscripts, we were to compute and store d_{ik} or d_{kj} before using these values to compute d_{ij}, we might be computing one of the following:

$$\begin{aligned} d_{ij}^{(k)} &= \min\left(d_{ij}^{(k-1)}, d_{ik}^{(k)} + d_{kj}^{(k-1)}\right) , \\ d_{ij}^{(k)} &= \min\left(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k)}\right) , \\ d_{ij}^{(k)} &= \min\left(d_{ij}^{(k-1)}, d_{ik}^{(k)} + d_{kj}^{(k)}\right) . \end{aligned}$$

In any of these scenarios, we're computing the weight of a shortest path from i to j with all intermediate vertices in $\{1, 2, \ldots, k\}$. If we use $d_{ik}^{(k)}$, rather than $d_{ik}^{(k-1)}$, in the computation, then we're using a subpath from i to k with all intermediate vertices in $\{1, 2, \ldots, k\}$. But k cannot be an *intermediate* vertex on a shortest path from i to k, since otherwise there would be a cycle on this shortest path. Thus, $d_{ik}^{(k)} = d_{ik}^{(k-1)}$. A similar argument applies to show that $d_{kj}^{(k)} = d_{kj}^{(k-1)}$. Hence, we can drop the superscripts in the computation.

Solution to Exercise 25.2-6

Here are two ways to detect negative-weight cycles:

1. Check the main-diagonal entries of the result matrix for a negative value. There is a negative weight cycle if and only if $d_{ii}^{(n)} < 0$ for some vertex i:
   - $d_{ii}^{(n)}$ is a path weight from i to itself; so if it is negative, there is a path from i to itself (i.e., a cycle), with negative weight.
   - If there is a negative-weight cycle, consider the one with the fewest vertices.
     - If it has just one vertex, then some $w_{ii} < 0$, so d_{ii} starts out negative, and since d values are never increased, it is also negative when the algorithm terminates.
     - If it has at least two vertices, let k be the highest-numbered vertex in the cycle, and let i be some other vertex in the cycle. $d_{ik}^{(k-1)}$ and $d_{ki}^{(k-1)}$ have correct shortest-path weights, because they are not based on negative-weight cycles. (Neither $d_{ik}^{(k-1)}$ nor $d_{ki}^{(k-1)}$ can include k as an intermediate vertex, and i and k are on the negative-weight cycle with the fewest vertices.) Since $i \rightsquigarrow k \rightsquigarrow i$ is a negative-weight cycle, the sum of those two weights is negative, so $d_{ii}^{(k)}$ will be set to a negative value. Since d values are never increased, it is also negative when the algorithm terminates.

   In fact, it suffices to check whether $d_{ii}^{(n-1)} < 0$ for some vertex i. Here's why. A negative-weight cycle containing vertex i either contains vertex n or it does not. If it does not, then clearly $d_{ii}^{(n-1)} < 0$. If the negative-weight cycle contains

vertex n, then consider $d_{nn}^{(n-1)}$. This value must be negative, since the cycle, starting and ending at vertex n, does not include vertex n as an intermediate vertex.

2. Alternatively, one could just run the normal FLOYD-WARSHALL algorithm one extra iteration to see if any of the d values change. If there are negative cycles, then some shortest-path cost will be cheaper. If there are no such cycles, then no d values will change because the algorithm gives the correct shortest paths.

Solution to Exercise 25.3-4

This solution is also posted publicly

It changes shortest paths. Consider the following graph. $V = \{s, x, y, z\}$, and there are 4 edges: $w(s, x) = 2$, $w(x, y) = 2$, $w(s, y) = 5$, and $w(s, z) = -10$. So we'd add 10 to every weight to make $\widehat{w}$. With w, the shortest path from s to y is $s \to x \to y$, with weight 4. With $\widehat{w}$, the shortest path from s to y is $s \to y$, with weight 15. (The path $s \to x \to y$ has weight 24.) The problem is that by just adding the same amount to every edge, you penalize paths with more edges, even if their weights are low.

Solution to Exercise 25.3-6

In this solution, we assume that $\infty - \infty$ is undefined; in particular, it's not 0.

Let $G = (V, E)$, where $V = \{s, u\}$, $E = \{(u, s)\}$, and $w(u, s) = 0$. There is only one edge, and it enters s. When we run Bellman-Ford from s, we get $h(s) = \delta(s, s) = 0$ and $h(u) = \delta(s, u) = \infty$. When we reweight, we get $\widehat{w}(u, s) = 0 + \infty - 0 = \infty$. We compute $\widehat{\delta}(u, s) = \infty$, and so we compute $d_{us} = \infty + 0 - \infty \neq 0$. Since $\delta(u, s) = 0$, we get an incorrect answer.

If the graph G is strongly connected, then we get $h(v) = \delta(s, v) < \infty$ for all vertices $v \in V$. Thus, the triangle inequality says that $h(v) \leq h(u) + w(u, v)$ for all edges $(u, v) \in E$, and so $\widehat{w}(u, v) = w(u, v) + h(u) - h(v) \geq 0$. Moreover, all edge weights $\widehat{w}(u, v)$ used in Lemma 25.1 are finite, and so the lemma holds. Therefore, the conditions we need in order to use Johnson's algorithm hold: that reweighting does not change shortest paths, and that all edge weights $\widehat{w}(u, v)$ are nonnegative. Again relying on G being strongly connected, we get that $\widehat{\delta}(u, v) < \infty$ for all edges $(u, v) \in E$, which means that $d_{uv} = \widehat{\delta}(u, v) + h(v) - h(u)$ is finite and correct.

Solution to Problem 25-1

a. Let $T = (t_{ij})$ be the $|V| \times |V|$ matrix representing the transitive closure, such that t_{ij} is 1 if there is a path from i to j, and 0 otherwise.

Initialize T (when there are no edges in G) as follows:

$$t_{ij} = \begin{cases} 1 & \text{if } i = j \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

We update T as follows when an edge (u, v) is added to G:

```
TRANSITIVE-CLOSURE-UPDATE(T, u, v)
  let T be |V| × |V|
  for i = 1 to |V|
      for j = 1 to |V|
          if t_iu == 1 and t_vj == 1
              t_ij = 1
```

- With this procedure, the effect of adding edge (u, v) is to create a path (via the new edge) from every vertex that could already reach u to every vertex that could already be reached from v.
- Note that the procedure sets $t_{uv} = 1$, because both t_{uu} and t_{vv} are initialized to 1.
- This procedure takes $\Theta(V^2)$ time because of the two nested loops.

b. Consider inserting the edge $(v_{|V|}, v_1)$ into the straight-line graph $v_1 \to v_2 \to \cdots \to v_{|V|}$.

Before this edge is inserted, only $|V|\,(|V|+1)/2$ entries in T are 1 (the entries on and above the main diagonal). After the edge is inserted, the graph is a cycle in which every vertex can reach every other vertex, so all $|V|^2$ entries in T are 1. Hence $|V|^2 - (|V|\,(|V|+1)/2) = \Theta(V^2)$ entries must be changed in T, so any algorithm to update the transitive closure must take $\Omega(V^2)$ time on this graph.

c. The algorithm in part (a) would take $\Theta(V^4)$ time to insert all possible $\Theta(V^2)$ edges, so we need a more efficient algorithm in order for any sequence of insertions to take only $O(V^3)$ total time.

To improve the algorithm, notice that the loop over j is pointless when $t_{iv} = 1$. That is, if there is already a path $i \rightsquigarrow v$, then adding the edge (u, v) cannot make any new vertices reachable from i. The loop to set t_{ij} to 1 for j such that there exists a path $v \rightsquigarrow j$ is just setting entries that are already 1. Eliminate this redundant processing as follows:

```
TRANSITIVE-CLOSURE-UPDATE(T, u, v)
  let T be |V| × |V|
  for i = 1 to |V|
      if t_iu == 1 and t_iv == 0
          for j = 1 to |V|
              if t_vj == 1
                  t_ij = 1
```

We show that this procedure takes $O(V^3)$ time to update the transitive closure for any sequence of n insertions:

- There cannot be more than $|V|^2$ edges in G, so $n \le |V|^2$.

- Summed over n insertions, the time for the outer **for** loop header and the test for $t_{iu} == 1$ and $t_{iv} == 0$ is $O(nV) = O(V^3)$.
- The last three lines, which take $\Theta(V)$ time, are executed only $O(V^2)$ times for n insertions. To see why, notice that the last three lines are executed only when t_{iv} equals 0, and in that case, the last line sets $t_{iv} = 1$. Thus, the number of 0 entries in T is reduced by at least 1 each time the last three lines run. Since there are only $|V|^2$ entries in T, these lines can run at most $|V|^2$ times.
- Hence, the total running time over n insertions is $O(V^3)$.

Lecture Notes for Chapter 26: Maximum Flow

Chapter 26 overview

Network flow

[The third edition treats flow networks differently from the first two editions. The concept of net flow is gone, except that we do discuss net flow across a cut. Skew symmetry is also gone, as is implicit summation notation. The third edition counts flows on edges directly. We find that although the mathematics is not quite as slick as in the first two editions, the approach in the third edition matches intuition more closely, and therefore students tend to pick it up more quickly.]

Use a graph to model material that flows through conduits.

Each edge represents one conduit, and has a ***capacity***, which is an upper bound on the ***flow rate*** = units/time.

Can think of edges as pipes of different sizes. But flows don't have to be of liquids. Book has an example where a flow is how many trucks per day can ship hockey pucks between cities.

Want to compute max rate that we can ship material from a designated ***source*** to a designated ***sink***.

Flow networks

$G = (V, E)$ directed.

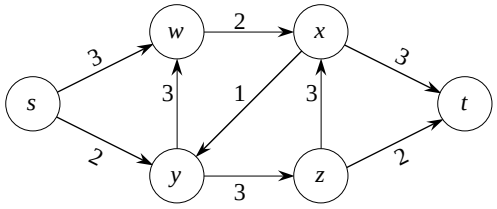
Each edge (u, v) has a ***capacity*** $c(u, v) \geq 0$.

If $(u, v) \notin E$, then $c(u, v) = 0$.

If $(u, v) \in E$, then reverse edge $(v, u) \notin E$. (Can work around this restriction.)

Source vertex s, ***sink*** vertex t, assume $s \rightsquigarrow v \rightsquigarrow t$ for all $v \in V$, so that each vertex lies on a path from source to sink.

Example: *[Edges are labeled with capacities.]*



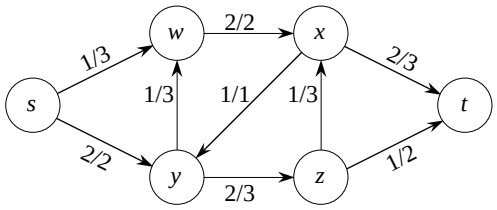
Flow

A function $f : V \times V \to \mathbb{R}$ satisfying

- ***Capacity constraint:*** For all $u, v \in V, 0 \le f(u,v) \le c(u,v)$,
- ***Flow conservation:*** For all $u \in V - \{s,t\}$, $\underbrace{\sum_{v \in V} f(v,u)}_{\text{flow into } u} = \underbrace{\sum_{v \in V} f(u,v)}_{\text{flow out of } u}$.

  Equivalently, $\sum_{v \in V} f(u,v) - \sum_{v \in V} f(v,u) = 0$.

[Add flows to previous example. Edges here are labeled as flow/capacity. Leave on board.]



- Note that all flows are $\le$ capacities.
- Verify flow conservation by adding up flows at a couple of vertices.
- Note that all flows $= 0$ is legitimate.

Value of flow f
$$\begin{aligned} &= |f| \\ &= \sum_{v \in V} f(s,v) - \sum_{v \in V} f(v,s) \\ &= \text{flow out of source} - \text{flow into source}\ . \end{aligned}$$

In the example above, value of flow $f = |f| = 3$.

Maximum-flow problem

Given G, s, t, and c, find a flow whose value is maximum.

Antiparallel edges

Definition of flow network does not allow both (u,v) and (v,u) to be edges. These edges would be ***antiparallel***.

What if we really need antiparallel edges?

- Choose one of them, say (u, v).
- Create a new vertex v'.
- Replace (u, v) by two new edges (u, v') and (v', v), with $c(u, v') = c(v', v) = c(u, v)$.
- Get an equivalent flow network with no antiparallel edges.

Cuts

A ***cut*** (S, T) of flow network $G = (V, E)$ is a partition of V into S and $T = V - S$ such that $s \in S$ and $t \in T$.

- Similar to cut used in minimum spanning trees, except that here the graph is directed, and we require $s \in S$ and $t \in T$.

For flow f, the ***net flow*** across cut (S, T) is

$$f(S, T) = \sum_{u \in S} \sum_{v \in T} f(u, v) - \sum_{u \in S} \sum_{v \in T} f(v, u) \ .$$

Capacity of cut (S, T) is

$$c(S, T) = \sum_{u \in S} \sum_{v \in T} c(u, v) \ .$$

A ***minimum cut*** of G is a cut whose capacity is minimum over all cuts of G.

Asymmetry between net flow across a cut and capacity of a cut: For capacity, count only capacities of edges going from S to T. Ignore edges going in the reverse direction. For net flow, count flow on all edges across the cut: flow on edges going from S to T minus flow on edges going from T to S.

In previous example, consider the cut $S = \{s, w, y\}, T = \{x, z, t\}$.

$$\begin{aligned} f(S, T) &= \underbrace{f(w, x) + f(y, z)}_{\text{from } S \text{ to } T} - \underbrace{f(x, y)}_{\text{from } T \text{ to } S} \\ &= 2 + 2 - 1 \\ &= 3 \ . \\ c(S, T) &= \underbrace{c(w, x) + c(y, z)}_{\text{from } S \text{ to } T} \\ &= 2 + 3 \\ &= 5 \ . \end{aligned}$$

Now consider the cut $S = \{s, w, x, y\}, T = \{z, t\}$.

$$\begin{aligned} f(S, T) &= \underbrace{f(x, t) + f(y, z)}_{\text{from } S \text{ to } T} - \underbrace{f(z, x)}_{\text{from } T \text{ to } S} \\ &= 2 + 2 - 1 \\ &= 3 \ . \\ c(S, T) &= \underbrace{c(x, t) + c(y, z)}_{\text{from } S \text{ to } T} \\ &= 3 + 3 \\ &= 6 \ . \end{aligned}$$

Same flow as previous cut, higher capacity.

Lemma
For any cut (S, T), $f(S, T) = |f|$.
(Net flow across the cut equals value of the flow.)
[Leave on board.]

[This proof is much more involved than the proof in the first two editions. You might want to omit it, or just give the intuition that no matter where you cut the pipes in a network, you'll see the same flow volume coming out of the openings.]

Proof Rewrite flow conservation: for any $u \in V - \{s, t\}$,

$$\sum_{v \in V} f(u, v) - \sum_{v \in V} f(v, u) = 0 \, .$$

Take definition of $|f|$ and add in left-hand side of above equation, summed over all vertices in $S - \{s\}$. Above equation applies to each vertex in $S - \{s\}$ (since $t \notin S$ and obviously $s \notin S - \{s\}$), so just adding in lots of 0s:

$$|f| = \sum_{v \in V} f(s, v) - \sum_{v \in V} f(v, s) + \sum_{u \in S - \{s\}} \left(\sum_{v \in V} f(u, v) - \sum_{v \in V} f(v, u) \right) \, .$$

Expand right-hand summation and regroup terms:

$$\begin{aligned}
|f| &= \sum_{v \in V} f(s, v) - \sum_{v \in V} f(v, s) + \sum_{u \in S - \{s\}} \sum_{v \in V} f(u, v) - \sum_{u \in S - \{s\}} \sum_{v \in V} f(v, u) \\
&= \sum_{v \in V} \left(f(s, v) + \sum_{u \in S - \{s\}} f(u, v) \right) - \sum_{v \in V} \left(f(v, s) + \sum_{u \in S - \{s\}} f(v, u) \right) \\
&= \sum_{v \in V} \sum_{u \in S} f(u, v) - \sum_{v \in V} \sum_{u \in S} f(v, u) \, .
\end{aligned}$$

Partition V into $S \cup T$ and split each summation over V into summations over S and T:

$$\begin{aligned}
|f| &= \sum_{v \in S} \sum_{u \in S} f(u, v) + \sum_{v \in T} \sum_{u \in S} f(u, v) - \sum_{v \in S} \sum_{u \in S} f(v, u) - \sum_{v \in T} \sum_{u \in S} f(v, u) \\
&= \sum_{v \in T} \sum_{u \in S} f(u, v) - \sum_{v \in T} \sum_{u \in S} f(v, u) \\
&\qquad + \left(\sum_{v \in S} \sum_{u \in S} f(u, v) - \sum_{v \in S} \sum_{u \in S} f(v, u) \right) \, .
\end{aligned}$$

Summations within parentheses are the same, since $f(x, y)$ appears once in each summation, for any $x, y \in V$. These summations cancel:

$$\begin{aligned}
|f| &= \sum_{u \in S} \sum_{v \in T} f(u, v) - \sum_{u \in S} \sum_{v \in T} f(v, u) \\
&= f(S, T) \, .
\end{aligned}$$

■ (lemma)

Corollary
The value of any flow $\leq$ capacity of any cut.
[Leave on board.]

Proof Let (S, T) be any cut, f be any flow.

$$\begin{aligned}
|f| &= f(S,T) && \text{(lemma)}\\
&= \sum_{u\in S}\sum_{v\in T} f(u,v) - \sum_{u\in S}\sum_{v\in T} f(v,u) && \text{(definition of } f(S,T)\text{)}\\
&\le \sum_{u\in S}\sum_{v\in T} f(u,v) && (f(v,u) \ge 0)\\
&\le \sum_{u\in S}\sum_{v\in T} c(u,v) && \text{(capacity constraint)}\\
&= c(S,T)\ . && \text{(definition of } c(S,T)\text{)} \quad \blacksquare \text{ (corollary)}
\end{aligned}$$

Therefore, maximum flow $\le$ capacity of minimum cut.

Will see a little later that this is in fact an equality.

The Ford-Fulkerson method

Residual network

Given a flow f in network $G = (V, E)$.

Consider a pair of vertices $u, v \in V$.

How much additional flow can we push directly from u to v?
That's the ***residual capacity***,

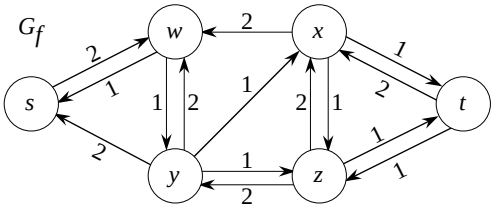
$$c_f(u,v) = \begin{cases} c(u,v) - f(u,v) & \text{if } (u,v) \in E\ , \\ f(v,u) & \text{if } (v,u) \in E\ , \\ 0 & \text{otherwise (i.e., } (u,v), (v,u) \notin E)\ . \end{cases}$$

The ***residual network*** is $G_f = (V, E_f)$, where

$E_f = \{(u,v) \in V \times V : c_f(u,v) > 0\}$.

Each edge of the residual network can admit a positive flow.

For our example:



Every edge $(u, v) \in E_f$ corresponds to an edge $(u, v) \in E$ or $(v, u) \in E$ (or both).

Therefore, $|E_f| \le 2\,|E|$.

Residual network is similar to a flow network, except that it may contain antiparallel edges ((u, v) and (v, u)). Can define a flow in a residual network that satisfies the definition of a flow, but with respect to capacities c_f in G_f.

Given flows f in G and f' in G_f, define $(f \uparrow f')$, the ***augmentation*** of f by f', as a function $V \times V \to \mathbb{R}$:

$$(f \uparrow f')(u,v) = \begin{cases} f(u,v) + f'(u,v) - f'(v,u) & \text{if } (u,v) \in E\ , \\ 0 & \text{otherwise} \end{cases}$$

for all $u, v \in V$.

Intuition: Increase the flow on (u,v) by $f'(u,v)$ but decrease it by $f'(v,u)$ because pushing flow on the reverse edge in the residual network decreases the flow in the original network. Also known as ***cancellation***.

Lemma
Given a flow network G, a flow f in G, and the residual network G_f, let f' be a flow in G_f. Then $f \uparrow f'$ is a flow in G with value $|f \uparrow f'| = |f| + |f'|$.

[See book for proof. It has a lot of summations in it. Probably not worth writing on the board.]

Augmenting path

A simple path $s \rightsquigarrow t$ in G_f.

- Admits more flow along each edge.
- Like a sequence of pipes through which we can squirt more flow from s to t.

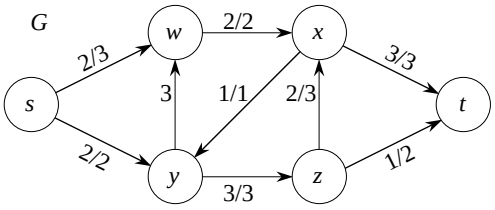
How much more flow can we push from s to t along augmenting path p?

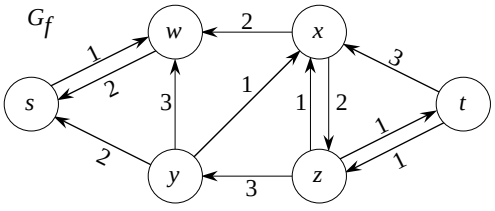
$c_f(p) = \min\{c_f(u,v) : (u,v) \text{ is on } p\}$.

For our example, consider the augmenting path $p = \langle s, w, y, z, x, t \rangle$.

Minimum residual capacity is 1.

After we push 1 additional unit along p: *[Continue from G left on board from before. Edge (y,w) has $f(y,w) = 0$, which we omit, showing only $c(y,w) = 3$.]*





Observe that G_f now has no augmenting path. Why? No edges cross the cut $(\{s,w\}, \{x,y,z,t\})$ in the forward direction in G_f. So no path can get from s to t.

Claim that the flow shown in G is a maximum flow.

Lemma

Given flow network G, flow f in G, residual network G_f. Let p be an augmenting path in G_f. Define $f_p : V \times V \to \mathbb{R}$:

$$f_p(u,v) = \begin{cases} c_f(p) & \text{if } (u,v) \text{ is on } p \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Then f_p is a flow in G_f with value $|f_p| = c_f(p) > 0$.

Corollary

Given flow network G, flow f in G, and an augmenting path p in G_f, define f_p as in lemma. Then $f \uparrow f_p$ is a flow in G with value $|f \uparrow f_p| = |f| + |f_p| > |f|$.

Theorem (Max-flow min-cut theorem)

The following are equivalent:

1. f is a maximum flow.
2. G_f has no augmenting path.
3. $|f| = c(S,T)$ for some cut (S,T).

Proof

(1) $\Rightarrow$ (2): Show the contrapositive: if G_f has an augmenting path, then f is not a maximum flow. If G_f has augmenting path p, then by the above corollary, $f \uparrow f_p$ is a flow in G with value $|f| + |f_p| > |f|$, so that f was not a maximum flow.

(2) $\Rightarrow$ (3): Suppose G_f has no augmenting path. Define

$$\begin{aligned} S &= \{v \in V : \text{there exists a path } s \rightsquigarrow v \text{ in } G_f\} \ , \\ T &= V - S \ . \end{aligned}$$

Must have $t \in T$; otherwise there is an augmenting path.
Therefore, (S,T) is a cut.
Consider $u \in S$ and $v \in T$:

- If $(u,v) \in E$, must have $f(u,v) = c(u,v)$; otherwise, $(u,v) \in E_f \Rightarrow v \in S$.
- If $(v,u) \in E$, must have $f(v,u) = 0$; otherwise, $c_f(u,v) = f(v,u) > 0 \Rightarrow (u,v) \in E_f \Rightarrow v \in S$.
- If $(u,v), (v,u) \notin E$, must have $f(u,v) = f(v,u) = 0$.

Then,

$$\begin{aligned} f(S,T) &= \sum_{u \in S} \sum_{v \in T} f(u,v) - \sum_{v \in T} \sum_{u \in S} f(v,u) \\ &= \sum_{u \in S} \sum_{v \in T} c(u,v) - \sum_{v \in T} \sum_{u \in S} 0 \\ &= c(S,T) \ . \end{aligned}$$

By lemma, $|f| = f(S,T) = c(S,T)$.

(3) $\Rightarrow$ (1): By corollary, $|f| \le c(S,T)$.
Therefore, $|f| = c(S,T) \Rightarrow f$ is a max flow. ■ (theorem)

Ford-Fulkerson algorithm

Keep augmenting flow along an augmenting path until there is no augmenting path.

Represent the flow attribute using the usual dot-notation, but on an edge: $(u, v).f$.

FORD-FULKERSON(G, s, t)
  for all $(u, v) \in G.E$
    $(u, v).f = 0$
  while there is an augmenting path p in G_f
    augment f by $c_f(p)$

Analysis

If capacities are all integer, then each augmenting path raises $|f|$ by ≥ 1. If max flow is f^*, then need $\leq |f^*|$ iterations $\Rightarrow$ time is $O(E\,|f^*|)$.
[Handwaving—see book for better explanation.]

Note that this running time is *not* polynomial in input size. It depends on $|f^*|$, which is not a function of $|V|$ and $|E|$.

If capacities are rational, can scale them to integers.

If irrational, FORD-FULKERSON might never terminate!

Edmonds-Karp algorithm

Do FORD-FULKERSON, but compute augmenting paths by BFS of G_f. Augmenting paths are shortest paths $s \rightsquigarrow t$ in G_f, with all edge weights $= 1$.

Edmonds-Karp runs in $O(VE^2)$ time.

To prove, need to look at distances to vertices in G_f.

Let $\delta_f(u, v) =$ shortest path distance u to v in G_f, with unit edge weights.

Lemma

For all $v \in V - \{s, t\}$, $\delta_f(s, v)$ increases monotonically with each flow augmentation.

Proof Suppose there exists $v \in V - \{s, t\}$ such that some flow augmentation causes $\delta_f(s, v)$ to decrease. Will derive a contradiction.

Let f be the flow before the first augmentation that causes a shortest-path distance to decrease, f' be the flow afterward.

Let v be a vertex with minimum $\delta_{f'}(s, v)$ whose distance was decreased by the augmentation, so $\delta_{f'}(s, v) < \delta_f(s, v)$.

Let a shortest path s to v in $G_{f'}$ be $s \rightsquigarrow u \to v$, so $(u, v) \in E_{f'}$ and $\delta_{f'}(s, v) = \delta_{f'}(s, u) + 1$. (Or $\delta_{f'}(s, u) = \delta_{f'}(s, v) - 1$.)

Since $\delta_{f'}(s, u) < \delta_{f'}(s, v)$ and how we chose v, we have $\delta_{f'}(s, u) \geq \delta_f(s, u)$.

Claim

$(u, v) \notin E_f$.

Proof If $(u, v) \in E_f$, then

$$\begin{aligned}\delta_f(s, v) &\leq \delta_f(s, u) + 1 \quad \text{(triangle inequality)}\\ &\leq \delta_{f'}(s, u) + 1\\ &= \delta_{f'}(s, v)\ ,\end{aligned}$$

contradicting $\delta_{f'}(s, v) < \delta_f(s, v)$. ■ (claim)

How can $(u, v) \notin E_f$ and $(u, v) \in E_{f'}$?

The augmentation must increase flow v to u.

Since Edmonds-Karp augments along shortest paths, the shortest path s to u in G_f has (v, u) as its last edge.

Therefore,

$$\begin{aligned}\delta_f(s, v) &= \delta_f(s, u) - 1\\ &\leq \delta_{f'}(s, u) - 1\\ &= \delta_{f'}(s, v) - 2\ ,\end{aligned}$$

contradicting $\delta_{f'}(s, v) < \delta_f(s, v)$.

Therefore, v cannot exist. ■ (lemma)

Theorem

Edmonds-Karp performs $O(VE)$ augmentations.

Proof Suppose p is an augmenting path and $c_f(u, v) = c_f(p)$. Then call (u, v) a ***critical*** edge in G_f, and it disappears from the residual network after augmenting along p.

≥ 1 edge on any augmenting path is critical.

Will show that each of the $|E|$ edges can become critical $\leq |V|/2$ times.

Consider $u, v \in V$ such that either $(u, v) \in E$ or $(v, u) \in E$ or both. Since augmenting paths are shortest paths, when (u, v) becomes critical first time, $\delta_f(s, v) = \delta_f(s, u) + 1$.

Augment flow, so that (u, v) disppears from the residual network. This edge cannot reappear in the residual network until flow from u to v decreases, which happens only if (v, u) is on an augmenting path in $G_{f'}$: $\delta_{f'}(s, u) = \delta_{f'}(s, v) + 1$. ($f'$ is flow when this occurs.)

By lemma, $\delta_f(s, v) \leq \delta_{f'}(s, v) \Rightarrow$

$$\begin{aligned}\delta_{f'}(s, u) &= \delta_{f'}(s, v) + 1\\ &\geq \delta_f(s, v) + 1\\ &= \delta_f(s, u) + 2\ .\end{aligned}$$

Therefore, from the time (u, v) becomes critical to the next time, distance of u from s increases by ≥ 2. Initially, distance to u is ≥ 0, and augmenting path can't have s, u, and t as intermediate vertices.

Therefore, until u becomes unreachable from source, its distance is $\leq |V| - 2 \Rightarrow$ after (u, v) becomes critical the first time, it can become critical $\leq (|V| - 2)/2 = |V|/2 - 1$ times more $\Rightarrow$ (u, v) can become critical $\leq |V|/2$ times.

Since $O(E)$ pairs of vertices can have an edge between them in residual network, total # of critical edges during execution of Edmonds-Karp is $O(VE)$. Since each augmenting path has ≥ 1 critical edge, have $O(VE)$ augmentations. ■ (theorem)

Use BFS to find each augmenting path in $O(E)$ time $\Rightarrow O(VE^2)$ time.

Can get better bounds.

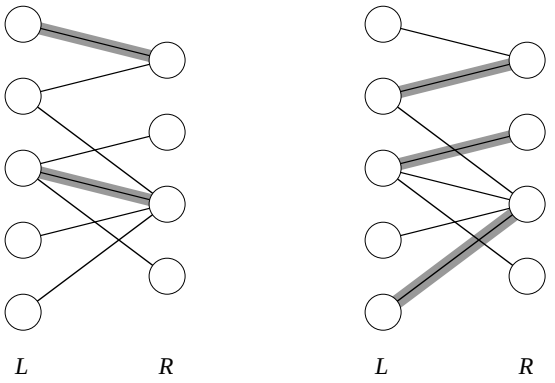
Push-relabel algorithms in Sections 26.4–26.5 give $O(V^3)$.

Can do even better.

Maximum bipartite matching

Example of a problem that can be solved by turning it into a flow problem.

$G = (V, E)$ (undirected) is ***bipartite*** if we can partition $V = L \cup R$ such that all edges in E go between L and R.



matching maximum matching

A ***matching*** is a subset of edges $M \subseteq E$ such that for all $v \in V$, ≤ 1 edge of M is incident on v. (Vertex v is ***matched*** if an edge of M is incident on it; otherwise ***unmatched***).

Maximum matching: a matching of maximum cardinality. (M is a maximum matching if $|M| \geq |M'|$ for all matchings M'.)

Problem

Given a bipartite graph (with the partition), find a maximum matching.

Application

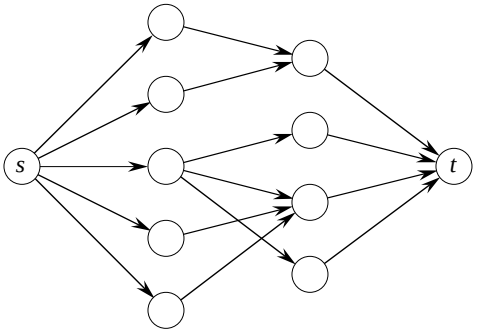
Matching planes to routes.

- $L =$ set of planes.
- $R =$ set of routes.
- $(u, v) \in E$ if plane u can fly route v.

- Want maximum # of routes to be served by planes.

Given G, define flow network $G' = (V', E')$.

- $V' = V \cup \{s, t\}$.
- $E' = \{(s, u) : u \in L\} \cup \{(u, v) : (u, v) \in E\} \cup \{(v, t) : v \in R\}$.
- $c(u, v) = 1$ for all $(u, v) \in E'$.



Each vertex in V has ≥ 1 incident edge $\Rightarrow |E| \geq |V|/2$.

Therefore, $|E| \leq |E'| = |E| + |V| \leq 3|E|$.

Therefore, $|E'| = \Theta(E)$.

Find a max flow in G'. Book shows that it will have integer values for all (u, v).

Use edges that carry flow of 1 in matching.

Book proves that this method produces a maximum matching.

Solutions for Chapter 26: Maximum Flow

Solution to Exercise 26.1-1

We will prove that for every flow in $G = (V, E)$, we can construct a flow in $G' = (V', E')$ that has the same value as that of the flow in G. The required result follows since a maximum flow in G is also a flow. Let f be a flow in G. By construction, $V' = V \cup \{x\}$ and $E' = (E - \{(u, v)\}) \cup \{(u, x), (x, v)\}$. Construct f' in G' as follows:

$$f'(y, z) = \begin{cases} f(y, z) & \text{if } (y, z) \neq (u, x) \text{ and } (y, z) \neq (x, v) , \\ f(u, v) & \text{if } (y, z) = (u, x) \text{ or } (y, z) = (x, v) . \end{cases}$$

Informally, f' is the same as f, except that the flow $f(u, v)$ now passes through an intermediate vertex x. The vertex x has incoming flow (if any) only from u, and has outgoing flow (if any) only to vertex v.

We first prove that f' satisfies the required properties of a flow. It is obvious that the capacity constraint is satisfied for every edge in E' and that every vertex in $V' - \{u, v, x\}$ obeys flow conservation.

To show that edges (u, x) and (x, v) obey the capacity constraint, we have

$$\begin{aligned} f(u, x) &= f(u, v) \leq c(u, v) = c(u, x) , \\ f(x, v) &= f(u, v) \leq c(u, v) = c(x, v) . \end{aligned}$$

We now prove flow conservation for u. Assuming that $u \notin \{s, t\}$, we have

$$\begin{aligned} \sum_{y \in V'} f'(u, y) &= \sum_{y \in V' - \{x\}} f'(u, y) + f'(u, x) \\ &= \sum_{y \in V - \{v\}} f(u, y) + f(u, v) \\ &= \sum_{y \in V} f(u, y) \\ &= \sum_{y \in V} f(y, u) \qquad \text{(because } f \text{ obeys flow conservation)} \\ &= \sum_{y \in V'} f'(y, u) . \end{aligned}$$

For vertex v, a symmetric argument proves flow conservation.

For vertex x, we have

$$\begin{aligned}\sum_{y \in V'} f'(y,x) &= f'(u,x) \\ &= f'(x,v) \\ &= \sum_{y \in V'} f'(x,y) \ .\end{aligned}$$

Thus, f' is a valid flow in G'.

We now prove that the values of the flow in both cases are equal. If the source s is not in $\{u, v\}$, the proof is trivial, since our construction assigns the same flows to incoming and outgoing edges of s. If $s = u$, then

$$\begin{aligned}|f'| &= \sum_{y \in V'} f'(u,y) - \sum_{y \in V'} f'(y,u) \\ &= \sum_{y \in V' - \{x\}} f'(u,y) - \sum_{y \in V'} f'(y,u) + f'(u,x) \\ &= \sum_{y \in V - \{v\}} f(u,y) - \sum_{y \in V} f(y,u) + f(u,v) \\ &= \sum_{y \in V} f(u,y) - \sum_{y \in V} f(y,u) \\ &= |f| \ .\end{aligned}$$

The case when $s = v$ is symmetric. We conclude that f' is a valid flow in G' with $|f'| = |f|$.

Solution to Exercise 26.1-3

We show that, given any flow f' in the flow network $G = (V, E)$, we can construct a flow f as stated in the exercise. The result will follow when f' is a maximum flow. The idea is that even if there is a path from s to the connected component of u, no flow can enter the component, since the flow has no path to reach t. Thus, all the flow inside the component must be cyclic, which can be made zero without affecting the net value of the flow.

Two cases are possible: where u is not connected to t, and where u is not connected to s. We only analyze the former case. The analysis for the latter case is similar.

Let Y be the set of all vertices that have no path to t. Our roadmap will be to first prove that no flow can leave Y. We use this result and flow conservation to prove that no flow can enter Y. We shall then constuct the flow f, which has the required properties, and prove that $|f| = |f'|$.

The first step is to prove that there can be no flow from a vertex $y \in Y$ to a vertex $v \in V - Y$. That is, $f'(y,v) = 0$. This is so, because there are no edges (y,v) in E. If there were an edge $(y,v) \in E$, then there would be a path from y to t, which contradicts how we defined the set Y.

We will now prove that $f'(v,y) = 0$, too. We will do so by applying flow conservation to each vertex in Y and taking the sum over Y. By flow conservation, we have

$$\sum_{y \in Y} \sum_{v \in V} f'(y, v) = \sum_{y \in Y} \sum_{v \in V} f'(v, y) \ .$$

Partitioning V into Y and $V - Y$ gives

$$\sum_{y \in Y} \sum_{v \in V - Y} f'(y, v) + \sum_{y \in Y} \sum_{v \in Y} f'(y, v)$$
$$= \sum_{y \in Y} \sum_{v \in V - Y} f'(v, y) + \sum_{y \in Y} \sum_{v \in Y} f'(v, y) \ . \qquad (*)$$

But we also have

$$\sum_{y \in Y} \sum_{v \in Y} f'(y, v) = \sum_{y \in Y} \sum_{v \in Y} f'(v, y) \ ,$$

since the left-hand side is the same as the right-hand side, except for a change of variable names v and y. We also have

$$\sum_{y \in Y} \sum_{v \in V - Y} f'(y, v) = 0 \ ,$$

since $f'(y, v) = 0$ for each $y \in Y$ and $v \in V - Y$. Thus, equation $(*)$ simplifies to

$$\sum_{y \in Y} \sum_{v \in V - Y} f'(v, y) = 0 \ .$$

Because the flow function is nonnegative, $f(v, y) = 0$ for each $v \in V$ and $y \in Y$. We conclude that there can be no flow between any vertex in Y and any vertex in $V - Y$.

The same technique can show that if there is a path from u to t but not from s to u, and we define Z as the set of vertices that do not have have a path from s to u, then there can be no flow between any vertex in Z and any vertex in $V - Z$. Let $X = Y \cup Z$. We thus have $f'(v, x) = f'(x, v) = 0$ if $x \in X$ and $v \notin X$.

We are now ready to construct flow f:

$$f(u, v) = \begin{cases} f'(u, v) & \text{if } u, v \notin X \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

We note that f satisfies the requirements of the exercise. We now prove that f also satisfies the requirements of a flow function.

The capacity constraint is satisfied, since whenever $f(u, v) = f'(u, v)$, we have $f(u, v) = f'(u, v) \le c(u, v)$ and whenever $f(u, v) = 0$, we have $f(u, v) = 0 \le c(u, v)$.

For flow conservation, let x be some vertex other than s or t. If $x \in X$, then from the construction of f, we have

$$\sum_{v \in V} f(x, v) = \sum_{v \in V} f(v, x) = 0 \ .$$

Otherwise, if $x \notin X$, note that $f(x,v) = f'(x,v)$ and $f(v,x) = f'(v,x)$ for all vertices $v \in V$. Thus,

$$\begin{aligned}\sum_{v \in V} f(x,v) &= \sum_{v \in V} f'(x,v) \\ &= \sum_{v \in V} f'(v,x) \quad \text{(because } f' \text{ obeys flow conservation)} \\ &= \sum_{v \in V} f(v,x)\,.\end{aligned}$$

Finally, we prove that the value of the flow remains the same. Since $s \notin X$, we have $f(s,v) = f'(s,v)$ and $f(v,x) = f'(v,x)$ for all vertices $v \in V$, and so

$$\begin{aligned}|f| &= \sum_{v \in V} f(s,v) - \sum_{v \in V} f(v,s) \\ &= \sum_{v \in V} f'(s,v) - \sum_{v \in V} f'(v,s) \\ &= |f'|\,.\end{aligned}$$

Solution to Exercise 26.1-4

To see that the flows form a convex set, we show that if f_1 and f_2 are flows, then so is $\alpha f_1 + (1-\alpha) f_2$ for all α such that $0 \le \alpha \le 1$.

For the capacity constraint, first observe that $\alpha \le 1$ implies that $1 - \alpha \ge 0$. Thus, for any $u, v \in V$, we have

$$\begin{aligned}\alpha f_1(u,v) + (1-\alpha) f_2(u,v) &\ge 0 \cdot f_1(u,v) + 0 \cdot (1-\alpha) f_2(u,v) \\ &= 0\,.\end{aligned}$$

Since $f_1(u,v) \le c(u,v)$ and $f_2(u,v) \le c(u,v)$, we also have

$$\begin{aligned}\alpha f_1(u,v) + (1-\alpha) f_2(u,v) &\le \alpha c(u,v) + (1-\alpha) c(u,v) \\ &= (\alpha + (1-\alpha)) c(u,v) \\ &= c(u,v)\,.\end{aligned}$$

For flow conservation, observe that since f_1 and f_2 obey flow conservation, we have $\sum_{v \in V} f_1(v,u) = \sum_{v \in V} f_1(u,v)$ and $\sum_{v \in V} f_1(v,u) = \sum_{v \in V} f_1(u,v)$ for any $u \in V - \{s,t\}$. We need to show that

$$\sum_{v \in V} (\alpha f_1(v,u) + (1-\alpha) f_2(v,u)) = \sum_{v \in V} (\alpha f_1(u,v) + (1-\alpha) f_2(u,v))$$

for any $u \in V - \{s,t\}$. We multiply both sides of the equality for f_1 by α, multiply both sides of the equality for f_2 by $1-\alpha$, and add the left-hand and right-hand sides of the resulting equalities to get

$$\alpha \sum_{v \in V} f_1(v,u) + (1-\alpha) \sum_{v \in V} f_2(v,u) = \alpha \sum_{v \in V} f_1(u,v) + (1-\alpha) \sum_{v \in V} f_2(u,v)\,.$$

Observing that

$$\begin{aligned}\alpha \sum_{v \in V} f_1(v,u) + (1-\alpha) \sum_{v \in V} f_2(v,u) &= \sum_{v \in V} \alpha f_1(v,u) + \sum_{v \in V} (1-\alpha) f_2(v,u) \\ &= \sum_{v \in V} (\alpha f_1(v,u) + (1-\alpha) f_2(v,u))\end{aligned}$$

and, likewise, that

$$\alpha \sum_{v \in V} f_1(u,v) + (1-\alpha) \sum_{v \in V} f_2(u,v) = \sum_{v \in V} (\alpha f_1(u,v) + (1-\alpha) f_2(u,v))$$

completes the proof that flow conservation holds, and thus that flows form a convex set.

Solution to Exercise 26.1-6

Create a vertex for each corner, and if there is a street between corners u and v, create directed edges (u,v) and (v,u). Set the capacity of each edge to 1. Let the source be corner on which the professor's house sits, and let the sink be the corner on which the school is located. We wish to find a flow of value 2 that also has the property that $f(u,v)$ is an integer for all vertices u and v. Such a flow represents two edge-disjoint paths from the house to the school.

Solution to Exercise 26.1-7

We will construct G' by splitting each vertex v of G into two vertices v_1, v_2, joined by an edge of capacity $l(v)$. All incoming edges of v are now incoming edges to v_1. All outgoing edges from v are now outgoing edges from v_2.

More formally, construct $G' = (V', E')$ with capacity function c' as follows. For every $v \in V$, create two vertices v_1, v_2 in V'. Add an edge (v_1, v_2) in E' with $c'(v_1, v_2) = l(v)$. For every edge $(u,v) \in E$, create an edge (u_2, v_1) in E' with capacity $c'(u_2, v_1) = c(u,v)$. Make s_1 and t_2 as the new source and target vertices in G'. Clearly, $|V'| = 2|V|$ and $|E'| = |E| + |V|$.

Let f be a flow in G that respects vertex capacities. Create a flow function f' in G' as follows. For each edge $(u,v) \in G$, let $f'(u_2, v_1) = f(u,v)$. For each vertex $u \in V - \{t\}$, let $f'(u_1, u_2) = \sum_{v \in V} f(u,v)$. Let $f'(t_1, t_2) = \sum_{v \in V} f(v,t)$.

We readily see that there is a one-to-one correspondence between flows that respect vertex capacities in G and flows in G'. For the capacity constraint, every edge in G' of the form (u_2, v_1) has a corresponding edge in G with a corresponding capacity and flow and thus satisfies the capacity constraint. For edges in E' of the form (u_1, u_2), the capacities reflect the vertex capacities in G. Therefore, for $u \in V - \{s,t\}$, we have $f'(u_1, u_2) = \sum_{v \in V} f(u,v) \le l(u) = c'(u_1, u_2)$. We also have $f'(t_1, t_2) = \sum_{v \in V} f(v,t) \le l(t) = c'(t_1, t_2)$. Note that this constraint also enforces the vertex capacities in G.

Now, we prove flow conservation. By construction, every vertex of the form u_1 in G' has exactly one outgoing edge (u_1, u_2), and every incoming edge to u_1 corresponds to an incoming edge of $u \in G$. Thus, for all vertices $u \in V - \{s, t\}$, we have

$$\begin{aligned}
\text{incoming flow to } u_1 &= \sum_{v \in V'} f'(v, u_1) \\
&= \sum_{v \in V} f(v, u) \\
&= \sum_{v \in V} f(u, v) \qquad \text{(because } f \text{ obeys flow conservation)} \\
&= f'(u_1, u_2) \\
&= \text{outgoing flow from } u_1 \, .
\end{aligned}$$

For t_1, we have

$$\begin{aligned}
\text{incoming flow} &= \sum_{v \in V'} f'(v, t_1) \\
&= \sum_{v \in V} f(v, t) \\
&= f'(t_1, t_2) \\
&= \text{outgoing flow} \, .
\end{aligned}$$

Vertices of the form u_2 have exactly one incoming edge (u_1, u_2), and every outgoing edge of u_2 corresponds to an outgoing edge of $u \in G$. Thus, for $u_2 \neq t_2$,

$$\begin{aligned}
\text{incoming flow} &= f'(u_1, u_2) \\
&= \sum_{v \in V} f(u, v) \\
&= \sum_{v \in V'} f'(u_2, v) \\
&= \text{outgoing flow} \, .
\end{aligned}$$

Finally, we prove that $|f'| = |f|$:

$$\begin{aligned}
|f'| &= \sum_{v \in V'} f'(s_1, v) \\
&= f'(s_1, s_2) \qquad \text{(because there are no other outgoing edges from } s_1) \\
&= \sum_{v \in V} f(s, v) \\
&= |f| \, .
\end{aligned}$$

Solution to Exercise 26.2-1

Lemma

1. If $v \notin V_1$, then $f(s, v) = 0$.
2. If $v \notin V_2$, then $f(v, s) = 0$.
3. If $v \notin V_1 \cup V_2$, then $f'(s, v) = 0$.
4. If $v \notin V_1 \cup V_2$, then $f'(v, s) = 0$.

Proof

1. Let $v \notin V_1$ be some vertex. From the definition of V_1, there is no edge from s to v. Thus, $f(s, v) = 0$.
2. Let $v \notin V_2$ be some vertex. From the definition of V_2, there is no edge from v to s. Thus, $f(v, s) = 0$.
3. Let $v \notin V_1 \cup V_2$ be some vertex. From the definition of V_1 and V_2, neither (s, v) nor (v, s) exists. Therefore, the third condition of the definition of residual capacity (equation (26.2)) applies, and $c_f(s, v) = 0$. Thus, $f'(s, v) = 0$.
4. Let $v \notin V_1 \cup V_2$ be some vertex. By equation (26.2), we have that $c_f(v, s) = 0$ and thus $f'(v, s) = 0$. ■ (lemma)

We conclude that the summations in equation (26.6) equal the summations in equation (26.7).

Solution to Exercise 26.2-8

Let G_f be the residual network just before an iteration of the **while** loop of FORD-FULKERSON, and let E_s be the set of residual edges of G_f into s. We'll show that the augmenting path p chosen by FORD-FULKERSON does not include an edge in E_s. Thus, even if we redefine G_f to disallow edges in E_s, the path p still remains an augmenting path in the redefined network. Since p remains unchanged, an iteration of the **while** loop of FORD-FULKERSON updates the flow in the same way as before the redefinition. Furthermore, by disallowing some edges, we do not introduce any new augmenting paths. Thus, FORD-FULKERSON still correctly computes a maximum flow.

Now, we prove that FORD-FULKERSON never chooses an augmenting path p that includes an edge $(v, s) \in E_s$. Why? The path p always starts from s, and if p included an edge (v, s), the vertex s would be repeated twice in the path. Thus, p would no longer be a *simple* path. Since FORD-FULKERSON chooses only simple paths, p cannot include (v, s).

Solution to Exercise 26.2-9

The augmented flow $f \uparrow f'$ satisfies the flow conservation property but not the capacity constraint property.

First, we prove that $f \uparrow f'$ satisfies the flow conservation property. We note that if edge $(u, v) \in E$, then $(v, u) \notin E$ and $f'(v, u) = 0$. Thus, we can rewrite the definition of flow augmentation (equation (26.4)), when applied to two flows, as

$$(f \uparrow f')(u, v) = \begin{cases} f(u, v) + f'(u, v) & \text{if } (u, v) \in E\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

The definition implies that the new flow on each edge is simply the sum of the two flows on that edge. We now prove that in $f \uparrow f'$, the net incoming flow for each

vertex equals the net outgoing flow. Let $u \notin \{s, t\}$ be any vertex of G. We have

$$\begin{aligned}
&\sum_{v \in V} (f \uparrow f')(v, u) \\
&\quad = \sum_{v \in V} (f(v, u) + f'(v, u)) \\
&\quad = \sum_{v \in V} f(v, u) + \sum_{v \in V} f'(v, u) \\
&\quad = \sum_{v \in V} f(u, v) + \sum_{v \in V} f'(u, v) \quad \text{(because } f, f' \text{ obey flow conservation)} \\
&\quad = \sum_{v \in V} (f(u, v) + f'(u, v)) \\
&\quad = \sum_{v \in V} (f \uparrow f')(u, v) \,.
\end{aligned}$$

We conclude that $f \uparrow f'$ satisfies flow conservation.

We now show that $f \uparrow f'$ need not satisfy the capacity constraint by giving a simple counterexample. Let the flow network G have just a source and a target vertex, with a single edge (s, t) having $c(s, t) = 1$. Define the flows f and f' to have $f(s, t) = f'(s, t) = 1$. Then, we have $(f \uparrow f')(s, t) = 2 > c(s, t)$. We conclude that $f \uparrow f'$ need not satisfy the capacity constraint.

Solution to Exercise 26.2-11

This solution is also posted publicly

For any two vertices u and v in G, we can define a flow network G_{uv} consisting of the directed version of G with $s = u$, $t = v$, and all edge capacities set to 1. (The flow network G_{uv} has V vertices and $2|E|$ edges, so that it has $O(V)$ vertices and $O(E)$ edges, as required. We want all capacities to be 1 so that the number of edges of G crossing a cut equals the capacity of the cut in G_{uv}.) Let f_{uv} denote a maximum flow in G_{uv}.

We claim that for any $u \in V$, the edge connectivity k equals $\min_{v \in V - \{u\}} \{|f_{uv}|\}$. We'll show below that this claim holds. Assuming that it holds, we can find k as follows:

```
EDGE-CONNECTIVITY(G)
  k = ∞
  select any vertex u ∈ G.V
  for each vertex v ∈ G.V − {u}
      set up the flow network G_uv as described above
      find the maximum flow f_uv on G_uv
      k = min(k, |f_uv|)
  return k
```

The claim follows from the max-flow min-cut theorem and how we chose capacities so that the capacity of a cut is the number of edges crossing it. We prove

that $k = \min_{v \in V - \{u\}} \{|f_{uv}|\}$, for any $u \in V$ by showing separately that k is at least this minimum and that k is at most this minimum.

- Proof that $k \geq \min_{v \in V - \{u\}} \{|f_{uv}|\}$:

  Let $m = \min_{v \in V - \{u\}} \{|f_{uv}|\}$. Suppose we remove only $m - 1$ edges from G. For any vertex v, by the max-flow min-cut theorem, u and v are still connected. (The max flow from u to v is at least m, hence any cut separating u from v has capacity at least m, which means at least m edges cross any such cut. Thus at least one edge is left crossing the cut when we remove $m-1$ edges.) Thus every vertex is connected to u, which implies that the graph is still connected. So at least m edges must be removed to disconnect the graph—i.e., $k \geq \min_{v \in V - \{u\}} \{|f_{uv}|\}$.

- Proof that $k \leq \min_{v \in V - \{u\}} \{|f_{uv}|\}$:

  Consider a vertex v with the minimum $|f_{uv}|$. By the max-flow min-cut theorem, there is a cut of capacity $|f_{uv}|$ separating u and v. Since all edge capacities are 1, exactly $|f_{uv}|$ edges cross this cut. If these edges are removed, there is no path from u to v, and so our graph becomes disconnected. Hence $k \leq \min_{v \in V - \{u\}} \{|f_{uv}|\}$.

- Thus, the claim that $k = \min_{v \in V - \{u\}} \{|f_{uv}|\}$, for any $u \in V$ is true.

Solution to Exercise 26.2-12

The idea of the proof is that if $f(v, s) = 1$, then there must exist a cycle containing the edge (v, s) and for which each edge carries one unit of flow. If we reduce the flow on each edge in the cycle by one unit, we can reduce $f(v, s)$ to 0 without affecting the value of the flow.

Given the flow network G and the flow f, we say that vertex y is *flow-connected* to vertex z if there exists a path p from y to z such that each edge of p has a positive flow on it. We also define y to be flow-connected to itself. In particular, s is flow-connected to s.

We start by proving the following lemma:

Lemma

Let $G = (V, E)$ be a flow network and f be a flow in G. If s is not flow-connected to v, then $f(v, s) = 0$.

Proof The idea is that since s is not flow-connected to v, there cannot be any flow from s to v. By using flow conservation, we will prove that there cannot be any flow from v to s either, and thus that $f(v, s) = 0$.

Let Y be the set of all vertices y such that s is flow-connected to y. By applying flow conservation to vertices in $V - Y$ and taking the sum, we obtain

$$\sum_{z \in V - Y} \sum_{x \in V} f(x, z) = \sum_{z \in V - Y} \sum_{x \in V} f(z, x) \, .$$

Partitioning V into Y and $V - Y$ gives

$$\sum_{z \in V-Y} \sum_{x \in V-Y} f(x,z) + \sum_{z \in V-Y} \sum_{x \in Y} f(x,z) = \sum_{z \in V-Y} \sum_{x \in V-Y} f(z,x) + \sum_{z \in V-Y} \sum_{x \in Y} f(z,x) \,. \qquad (\dagger)$$

But we have

$$\sum_{z \in V-Y} \sum_{x \in V-Y} f(x,z) = \sum_{z \in V-Y} \sum_{x \in V-Y} f(z,x) \,,$$

since the left-hand side is the same as the right-hand side, except for a change of variable names x and z. We also have

$$\sum_{z \in V-Y} \sum_{x \in Y} f(x,z) = 0 \,,$$

since the flow from any vertex in Y to any vertex in $V - Y$ must be 0. Thus, equation (†) simplifies to

$$\sum_{z \in V-Y} \sum_{x \in Y} f(z,x) = 0 \,.$$

The above equation implies that $f(z,x) = 0$ for each $z \in V - Y$ and $x \in Y$. In particular, since $v \in V - Y$ and $s \in Y$, we have that $f(v,s) = 0$. ∎

Now, we show how to construct the required flow f'. By the contrapositive of the lemma, $f(v,s) > 0$ implies that s is flow-connected to v through some path p. Let path p' be the path $s \overset{p}{\rightsquigarrow} v \to s$. Path p' is a cycle that has positive flow on each edge. Because we assume that all edge capacities are integers, the flow on each edge of p' is at least 1. If we subtract 1 from each edge of the cycle to obtain a flow f', then f' still satisfies the properties of a flow network and has the same value as $|f|$. Because edge (v,s) is in the cycle, we have that $f'(v,s) = f(v,s) - 1 = 0$.

Solution to Exercise 26.2-13

Let (S,T) and (X,Y) be two cuts in G (and G'). Let c' be the capacity function of G'. One way to define c' is to add a small amount δ to the capacity of each edge in G. That is, if u and v are two vertices, we set

$$c'(u,v) = c(u,v) + \delta \,.$$

Thus, if $c(S,T) = c(X,Y)$ and (S,T) has fewer edges than (X,Y), then we would have $c'(S,T) < c'(X,Y)$. We have to be careful and choose a small δ, lest we change the relative ordering of two unequal capacities. That is, if $c(S,T) < c(X,Y)$, then no matter many more edges (S,T) has than (X,Y), we still need to have $c'(S,T) < c'(X,Y)$. With this definition of c', a minimum cut in G' will be a minimum cut in G that has the minimum number of edges.

How should we choose the value of δ? Let m be the minimum difference between capacities of two unequal-capacity cuts in G. Choose $\delta = m/(2|E|)$. For any cut (S,T), since the cut can have at most $|E|$ edges, we can bound $c'(S,T)$ by

$c(S,T) \le c'(S,T) \le c(S,T) + |E| \cdot \delta$.

Let $c(S,T) < c(X,Y)$. We need to prove that $c'(S,T) < c'(X,Y)$. We have

$$\begin{aligned} c'(S,T) &\le c(S,T) + |E| \cdot \delta \\ &= c(S,T) + m/2 \\ &< c(X,Y) \qquad (\text{since } c(X,Y) - c(S,T) \ge m) \\ &\le c'(X,Y) \ . \end{aligned}$$

Because all capacities are integral, we can choose $m = 1$, obtaining $\delta = 1/2|E|$. To avoid dealing with fractional values, we can scale all capacities by $2|E|$ to obtain

$c'(u,v) = 2|E| \cdot c(u,v) + 1$.

Solution to Exercise 26.3-3

This solution is also posted publicly

By definition, an augmenting path is a simple path $s \rightsquigarrow t$ in the residual network G'_f. Since G has no edges between vertices in L and no edges between vertices in R, neither does the flow network G' and hence neither does G'_f. Also, the only edges involving s or t connect s to L and R to t. Note that although edges in G' can go only from L to R, edges in G'_f can also go from R to L.

Thus any augmenting path must go

$s \to L \to R \to \cdots \to L \to R \to t$,

crossing back and forth between L and R at most as many times as it can do so without using a vertex twice. It contains s, t, and equal numbers of distinct vertices from L and R—at most $2 + 2 \cdot \min(|L|, |R|)$ vertices in all. The length of an augmenting path (i.e., its number of edges) is thus bounded above by $2 \cdot \min(|L|, |R|) + 1$.

Solution to Exercise 26.4-1

We apply the definition of excess flow (equation (26.14)) to the initial preflow f created by INITIALIZE-PREFLOW (equation (26.15)) to obtain

$$\begin{aligned} e(s) &= \sum_{v \in V} f(v,s) - \sum_{v \in V} f(s,v) \\ &= 0 - \sum_{v \in V} c(s,v) \\ &= -\sum_{v \in V} c(s,v) \ . \end{aligned}$$

Now,

$$-|f^*| = \sum_{v \in V} f^*(v,s) - \sum_{v \in V} f^*(s,v)$$

$$\begin{aligned} &\geq 0 - \sum_{v \in V} c(s, v) \qquad \text{(since } f^*(v, s) \geq 0 \text{ and } f^*(s, v) \leq c(s, v)\text{)} \\ &= e(s) \,. \end{aligned}$$

Solution to Exercise 26.4-3

Each time we call RELABEL(u), we examine all edges $(u, v) \in E_f$. Since the number of relabel operations is at most $2|V| - 1$ per vertex, edge (u, v) will be examined during relabel operations at most $4|V| - 2 = O(V)$ times (at most $2|V| - 1$ times during calls to RELABEL(u) and at most $2|V| - 1$ times during calls to RELABEL(v)). Summing up over all the possible residual edges, of which there are at most $2|E| = O(E)$, we see that the total time spent relabeling vertices is $O(VE)$.

Solution to Exercise 26.4-4

We can find a minimum cut, given a maximum flow found in $G = (V, E)$ by a push-relabel algorithm, in $O(V)$ time. First, find a height $\widehat{h}$ such that $0 < \widehat{h} < |V|$ and there is no vertex whose height equals $\widehat{h}$ at termination of the algorithm. We need consider only $|V| - 2$ vertices, since $s.h = |V|$ and $t.h = 0$. Because $\widehat{h}$ can be one of at most $|V| - 1$ possible values, we know that for at least one number in $1, 2, \ldots, |V| - 1$, there will be no vertex of that height. Hence, $\widehat{h}$ is well defined, and it is easy to find in $O(V)$ time by using a simple boolean array indexed by heights $1, 2, \ldots, |V| - 1$.

Let $S = \{u \in V : u.h > \widehat{h}\}$ and $T = \{v \in V : v.h < \widehat{h}\}$. Because we know that $s.h = |V| > \widehat{h}$, we have $s \in S$, and because $t.h = 0 < \widehat{h}$, we have $t \in T$, as required for a cut.

We need to show that $f(u, v) = c(u, v)$, i.e., that $(u, v) \notin E_f$, for all $u \in S$ and $v \in T$. Once we do that, we have that $f(S, T) = c(S, T)$, and by Corollary 26.5, (S, T) is a minimum cut.

Suppose for the purpose of contradiction that there exist vertices $u \in S$ and $v \in T$ such that $(u, v) \in E_f$. Because h is always maintained as a height function (Lemma 26.16), we have that $u.h \leq v.h + 1$. But we also have $v.h < \widehat{h} < u.h$, and because all values are integer, $v.h \leq u.h - 2$. Thus, we have $u.h \leq v.h + 1 \leq u.h - 2 + 1 = u.h - 1$, which gives the contradiction that $u.height \leq u.height - 1$. Thus, (S, T) is a minimum cut.

Solution to Exercise 26.4-7

If we set $s.h = |V| - 2$, we have to change our definition of a height function to allow $s.h = |V| - 2$, rather than $s.h = |V|$. The only change we need to make to

the proof of correctness is to update the proof of Lemma 26.17. The original proof derives the contradiction that $s.h \le k < |V|$, which is at odds with $s.h = |V|$. When $s.h = |V| - 2$, there is no contradiction.

As in the original proof, let us suppose that we have a simple augmenting path $\langle v_0, v_1, \ldots, v_k \rangle$, where $v_0 = s$ and $v_k = t$, so that $k < |V|$. How could (s, v_1) be a residual edge? It had been saturated in INITIALIZE-PREFLOW, which means that we had to have pushed some flow from v_1 to s. In order for that to have happened, we must have had $v_1.h = s.h + 1$. If we set $s.h = |V| - 2$, then $v_1.h$ was $|V| - 1$ at the time. Since then, $v_1.h$ did not decrease, and so we have $v_1.h \ge |V| - 1$. Working backwards over our augmenting path, we have $v_{k-i}.h \le t.h + i$ for $i = 0, 1, \ldots, k$. As before, because the augmenting path is simple, $k < |V|$. Letting $i = k - 1$, we have $v_1.h \le t.h + k - 1 < 0 + |V| - 1$. We now have the contradiction that $v_1.h \ge |V| - 1$ and $v_1.h < |V| - 1$, which shows that Lemma 26.17 still holds.

Nothing in the analysis changes asymptotically.

Solution to Problem 26-2

a. The idea is to use a maximum-flow algorithm to find a maximum bipartite matching that selects the edges to use in a minimum path cover. We must show how to formulate the max-flow problem and how to construct the path cover from the resulting matching, and we must prove that the algorithm indeed finds a minimum path cover.

Define G' as suggested, with directed edges. Make G' into a flow network with source x_0 and sink y_0 by defining all edge capacities to be 1. G' is the flow network corresponding to a bipartite graph G'' in which $L = \{x_1, \ldots, x_n\}$, $R = \{y_1, \ldots, y_n\}$, and the edges are the (undirected version of the) subset of E' that doesn't involve x_0 or y_0.

The relationship of G to the bipartite graph G'' is that every vertex i in G is represented by two vertices, x_i and y_i, in G''. Edge (i, j) in G corresponds to edge (x_i, y_j) in G''. That is, an edge (x_i, y_j) in G'' means that an edge in G leaves i and enters j. Vertex x_i tells us about edges leaving i, and y_i tells us about edges entering i.

The edges in a bipartite matching in G'' can be used in a path cover of G, for the following reasons:

- In a bipartite matching, no vertex is used more than once. In a bipartite matching in G'', since no x_i is used more than once, at most one edge in the matching leaves any vertex i in G. Similarly, since no y_j is used more than once, at most one edge in the matching enters any vertex j in G.
- In a path cover, since no vertex appears in more than one path, at most one path edge enters each vertex and at most one path edge leaves each vertex.

We can construct a path cover P from any bipartite matching M (not just a maximum matching) by moving from some x_i to its matching y_j (if any), then from x_j to its matching y_k, and so on, as follows:

1. Start a new path containing a vertex i that has not yet been placed in a path.
2. If x_i is unmatched, the path can't go any farther; just add it to P.
3. If x_i is matched to some y_j, add j to the current path. If j has already been placed in a path (i.e., though we've just entered j by processing y_j, we've already built a path that leaves j by processing x_j), combine this path with that one and go back to step 1. Otherwise go to step 2 to process x_j.

This algorithm constructs a path cover, for the following reasons:

- Every vertex is put into some path, because we keep picking an unused vertex from which to start a path until there are no unused vertices.
- No vertex is put into two paths, because every x_i is matched to at most one y_j, and vice versa. That is, at most one candidate edge leaves each vertex, and at most one candidate edge enters each vertex. When building a path, we start or enter a vertex and then leave it, building a single path. If we ever enter a vertex that was left earlier, it must have been the start of another path, since there are no cycles, and we combine those paths so that the vertex is entered and left on a single path.

Every edge in M is used in some path because we visit every x_i, and we incorporate the single edge, if any, from each visited x_i. Thus, there is a one-to-one correspondence between edges in the matching and edges in the constructed path cover.

We now show that the path cover P constructed above has the fewest possible paths when the matching is maximum.

Let f be the flow corresonding to the bipartite matching M.

$$\begin{aligned}
|V| &= \sum_{p \in P} (\text{\# vertices in } p) && \text{(every vertex is on exactly 1 path)} \\
&= \sum_{p \in P} (1 + \text{\# edges in } p) \\
&= \sum_{p \in P} 1 + \sum_{p \in P} (\text{\# edges in } p) \\
&= |P| + |M| && \text{(by 1-to-1 correspondence)} \\
&= |P| + |f| && \text{(by Lemma 26.9) .}
\end{aligned}$$

Thus, for the fixed set V in our graph G, $|P|$ (the number of paths) is minimized when the flow f is maximized.

The overall algorithm is as follows:

- Use FORD-FULKERSON to find a maximum flow in G' and hence a maximum bipartite matching M in G''.
- Construct the path cover as described above.

Time

$O(VE)$ total:

- $O(V + E)$ to set up G',
- $O(VE)$ to find the maximum bipartite matching,

- $O(E)$ to trace the paths, because each edge $\in M$ is traversed only once and there are $O(E)$ edges in M.

b. The algorithm does not work if there are cycles.

Consider a graph G with 4 vertices, consisting of a directed triangle and an edge pointing to the triangle:

$E = \{(1,2),(2,3),(3,1),(4,1)\}$

G can be covered with a single path: $4 \to 1 \to 2 \to 3$, but our algorithm might find only a 2-path cover.

In the bipartite graph G', the edges (x_i, y_j) are

$(x_1, y_2), (x_2, y_3), (x_3, y_1), (x_4, y_1)$.

There are 4 edges from an x_i to a y_j, but 2 of them lead to y_1, so a maximum bipartite matching can have only 3 edges (and the maximum flow in G' has value 3). In fact, there are 2 possible maximum matchings. It is always possible to match (x_1, y_2) and (x_2, y_3), and then either (x_3, y_1) or (x_4, y_1) can be chosen, but not both.

The maximum flow found by one of our max-flow algorithms could find the flow corresponding to either of these matchings, since both are maximal. If it finds the matching with edge (x_3, x_1), then the matching would not contain (x_4, x_1); given that matching, our path algorithm is forced to produce 2 paths, one of which contains just the vertex 4.

Solution to Problem 26-3

a. Assume for the sake of contradiction that $A_k \notin T$ for some $A_k \in R_i$. Since $A_k \notin T$, we must have $A_k \in S$. On the other hand, we have $J_i \in T$. Thus, the edge (A_k, J_i) crosses the cut (S, T). But $c(A_k, J_i) = \infty$ by construction, which contradicts the assumption that (S, T) is a *finite*-capacity cut.

b. Let us define a *project-plan* as a set of jobs to accept and experts to hire. Let P be a project-plan. We assume that P has two attributes. The attribute $P.J$ denotes the set of accepted jobs, and $P.A$ denotes the set of hired experts.

A *valid* project-plan is one in which we have hired all experts that are required by the accepted jobs. Specifically, let P be a valid project plan. If $J_i \in P.J$, then $A_k \in P.A$ for each $A_k \in R_i$. Note that Professor Gore might decide to hire more experts than those that are actually required.

We define the *revenue* of a project-plan as the total profit from the accepted jobs minus the total cost of the hired experts. The problem asks us to find a valid project plan with maximum revenue.

We start by proving the following lemma, which establishes the relationship between the capacity of a cut in flow network G and the revenue of a valid project-plan.

Lemma (Min-cut max-revenue)
There exists a finite-capacity cut (S, T) of G with capacity $c(S, T)$ if and only if there exists a valid project-plan with net revenue $\left(\sum_{J_i \in J} p_i\right) - c(S, T)$.

Proof Let (S, T) be a finite-capacity cut of G with capacity $c(S, T)$. We prove one direction of the lemma by constructing the required project-plan.

Construct the project-plan P by including J_i in $P.J$ if and only if $J_i \in T$ and including A_k in $P.A$ if and only if $A_k \in T$. From part (a), P is a valid project-plan, since, for every $J_i \in P.J$, we have $A_k \in P.A$ for each $A_k \in R_i$.

Since the capacity of the cut is finite, there cannot be any edges of the form (A_k, J_i) crossing the cut, where $A_k \in S$ and $J_i \in T$. All edges going from a vertex in S to a vertex in T must be either of the form (s, A_k) or of the form (J_i, t). Let E_A be the set of edges of the form (s, A_k) that cross the cut, and let E_J be the set of edges of the form (J_i, t) that cross the cut, so that

$$c(S, T) = \sum_{(s, A_k) \in E_A} c(s, A_k) + \sum_{(J_i, t) \in E_J} c(J_i, t) \; .$$

Consider edges of the form (s, A_k). We have

$$\begin{aligned} (s, A_k) \in E_A \quad & \text{if and only if} \quad A_k \in T \\ & \text{if and only if} \quad A_k \in P.A \; . \end{aligned}$$

By construction, $c(s, A_k) = c_k$. Taking summations over E_A and over $P.A$, we obtain

$$\sum_{(s, A_k) \in E_A} c(s, A_k) = \sum_{A_k \in P.A} c_k \; .$$

Similarly, consider edges of the form (J_i, t). We have

$$\begin{aligned} (J_i, t) \in E_J \quad & \text{if and only if} \quad J_i \in S \\ & \text{if and only if} \quad J_i \notin T \\ & \text{if and only if} \quad J_i \notin P.J \; . \end{aligned}$$

By construction, $c(J_i, t) = p_i$. Taking summations over E_J and over $P.J$, we obtain

$$\sum_{(J_i, t) \in E_J} c(J_i, t) = \sum_{J_i \notin P.J} p_i \; .$$

Let v be the net revenue of P. Then, we have

$$\begin{aligned}
\nu &= \sum_{J_i \in P.J} p_i - \sum_{A_k \in P.A} c_k \\
&= \left(\sum_{J_i \in J} p_i - \sum_{J_i \notin P.J} p_i \right) - \sum_{A_k \in P.A} c_k \\
&= \sum_{J_i \in J} p_i - \left(\sum_{J_i \notin P.J} p_i + \sum_{A_k \in P.A} c_k \right) \\
&= \sum_{J_i \in J} p_i - \left(\sum_{(J_i,t) \in E_J} c(J_i, t) + \sum_{(s,A_k) \in E_A} c(s, A_k) \right) \\
&= \left(\sum_{J_i \in J} p_i \right) - c(S, T) \; .
\end{aligned}$$

Now, we prove the other direction of the lemma by constructing the required cut from a valid project-plan.

Construct the cut (S, T) as follows. For every $J_i \in P.J$, let $J_i \in T$. For every $A_k \in P.A$, let $A_k \in T$.

First, we prove that the cut (S, T) is a finite-capacity cut. Since edges of the form (A_k, J_i) are the only infinite-capacity edges, it suffices to prove that there are no edges (A_k, J_i) such that $A_k \in S$ and $J_i \in T$.

For the purpose of contradiction, assume there is an edge (A_k, J_i) such that $A_k \in S$ and $J_i \in T$. By our constuction, we must have $J_i \in P.J$ and $A_k \notin P.A$. But since the edge (A_k, J_i) exists, we have $A_k \in R_i$. Since P is a valid project-plan, we derive the contradiction that A_k must have been in $P.A$.

From here on, the analysis is the same as the previous direction. In particular, the last equation from the previous analysis holds: the net revenue ν equals $\left(\sum_{J_i \in J} p_i\right) - c(S, T)$. ■

We conclude that the problem of finding a maximum-revenue project-plan reduces to the problem of finding a minimum cut in G. Let (S, T) be a minimum cut. From the lemma, the maximum net revenue is given by

$$\left(\sum_{j_i \in J} p_i \right) - c(S, T) \; .$$

c. Construct the flow network G as shown in the problem statement. Obtain a minimum cut (S, T) by running any of the maximum-flow algorithms (say, Edmonds-Karp). Construct the project plan P as follows: add J_i to $P.J$ if and only if $J_i \in T$. Add A_k to $P.A$ if and only if $A_k \in T$.

First, we note that the number of vertices in G is $|V| = m + n + 2$, and the number of edges in G is $|E| = r + m + n$. Constructing G and recovering the project-plan from the minimum cut are clearly linear-time operations. The running time of our algorithm is thus asymptotically the same as the running time of the algorithm used to find the minimum cut. If we use Edmonds-Karp to find the minimum cut, the running time is $O(VE^2)$.

Solution to Problem 26-4

This solution is also posted publicly

a. Just execute one iteration of the Ford-Fulkerson algorithm. The edge (u, v) in E with increased capacity ensures that the edge (u, v) is in the residual network. So look for an augmenting path and update the flow if a path is found.

Time

$O(V + E) = O(E)$ if we find the augmenting path with either depth-first or breadth-first search.

To see that only one iteration is needed, consider separately the cases in which (u, v) is or is not an edge that crosses a minimum cut. If (u, v) does not cross a minimum cut, then increasing its capacity does not change the capacity of any minimum cut, and hence the value of the maximum flow does not change. If (u, v) does cross a minimum cut, then increasing its capacity by 1 increases the capacity of that minimum cut by 1, and hence possibly the value of the maximum flow by 1. In this case, there is either no augmenting path (in which case there was some other minimum cut that (u, v) does not cross), or the augmenting path increases flow by 1. No matter what, one iteration of Ford-Fulkerson suffices.

b. Let f be the maximum flow before reducing $c(u, v)$.

If $f(u, v) = 0$, we don't need to do anything.

If $f(u, v) > 0$, we will need to update the maximum flow. Assume from now on that $f(u, v) > 0$, which in turn implies that $f(u, v) \geq 1$.

Define $f'(x, y) = f(x, y)$ for all $x, y \in V$, except that $f'(u, v) = f(u, v)-1$. Although f' obeys all capacity contraints, even after $c(u, v)$ has been reduced, it is not a legal flow, as it violates flow conservation at u (unless $u = s$) and v (unless $v = t$). f' has one more unit of flow entering u than leaving u, and it has one more unit of flow leaving v than entering v.

The idea is to try to reroute this unit of flow so that it goes out of u and into v via some other path. If that is not possible, we must reduce the flow from s to u and from v to t by one unit.

Look for an augmenting path from u to v (note: *not* from s to t).

- If there is such a path, augment the flow along that path.
- If there is no such path, reduce the flow from s to u by augmenting the flow from u to s. That is, find an augmenting path $u \rightsquigarrow s$ and augment the flow along that path. (There definitely is such a path, because there is flow from s to u.) Similarly, reduce the flow from v to t by finding an augmenting path $t \rightsquigarrow v$ and augmenting the flow along that path.

Time

$O(V + E) = O(E)$ if we find the paths with either DFS or BFS.

Solution to Problem 26-5

a. The capacity of a cut is defined to be the sum of the capacities of the edges crossing it. Since the number of such edges is at most $|E|$, and the capacity of each edge is at most C, the capacity of *any* cut of G is at most $C\,|E|$.

b. The capacity of an augmenting path is the minimum capacity of any edge on the path, so we are looking for an augmenting path whose edges *all* have capacity at least K. Do a breadth-first search or depth-first-search as usual to find the path, considering only edges with residual capacity at least K. (Treat lower-capacity edges as though they don't exist.) This search takes $O(V + E) = O(E)$ time. (Note that $|V| = O(E)$ in a flow network.)

c. MAX-FLOW-BY-SCALING uses the Ford-Fulkerson method. It repeatedly augments the flow along an augmenting path until there are no augmenting paths with capacity at least 1. Since all the capacities are integers, and the capacity of an augmenting path is positive, when there are no augmenting paths with capacity at least 1, there must be no augmenting paths whatsoever in the residual network. Thus, by the max-flow min-cut theorem, MAX-FLOW-BY-SCALING returns a maximum flow.

d. • The first time line 4 is executed, the capacity of any edge in G_f equals its capacity in G, and by part (a) the capacity of a minimum cut of G is at most $C\,|E|$. Initially $K = 2^{\lfloor \lg C \rfloor}$, and so $2K = 2 \cdot 2^{\lfloor \lg C \rfloor} = 2^{\lfloor \lg C \rfloor + 1} > 2^{\lg C} = C$. Thus, the capacity of a minimum cut of G_f is initially less than $2K\,|E|$.

• The other times line 4 is executed, K has just been halved, and so the capacity of a cut of G_f is at most $2K\,|E|$ at line 4 if and only if that capacity was at most $K\,|E|$ when the **while** loop of lines 5–6 last terminated. Thus, we want to show that when line 7 is reached, the capacity of a minimum cut of G_f is at most $K\,|E|$.

Let G_f be the residual network when line 7 is reached. When we reach line 7, G_f contains no augmenting path with capacity at least K. Therefore, a maximum flow f' in G_f has value $|f'| < K\,|E|$. Then, by the max-flow min-cut theorem, a minimum cut in G_f has capacity less than $K\,|E|$.

e. By part (d), when line 4 is reached, the capacity of a minimum cut of G_f is at most $2K\,|E|$, and thus the maximum flow in G_f is at most $2K\,|E|$. The following lemma shows that the value of a maximum flow in G equals the value of the current flow f in G plus the value of a maximum flow in G_f.

Lemma

Let f be a flow in flow network G, and f' be a maximum flow in the residual network G_f. Then $f \uparrow f'$ is a maximum flow in G.

Proof By the max-flow min-cut theorem, $|f'| = c_f(S, T)$ for some cut (S, T) of G_f, which is also a cut of G. By Lemma 26.4, $|f| = f(S, T)$. By Lemma 26.1, $f \uparrow f'$ is a flow in G with value $|f \uparrow f'| = |f| + |f'|$. We

will show that $|f| + |f'| = c(S,T)$ which, by the max-flow min-cut theorem, will prove that $f \uparrow f'$ is a maximum flow in G.

We have

$$\begin{aligned}
|f| + |f'| &= f(S,T) + c_f(S,T) \\
&= \left(\sum_{u\in S}\sum_{v\in T} f(u,v) - \sum_{u\in S}\sum_{v\in T} f(v,u)\right) + \sum_{u\in S}\sum_{v\in T} c_f(u,v) \\
&= \left(\sum_{u\in S,v\in T} f(u,v) - \sum_{u\in S,v\in T} f(v,u)\right) \\
&\quad + \left(\sum_{\substack{u\in S,v\in T,\\(u,v)\in E}} c(u,v) - \sum_{\substack{u\in S,v\in T,\\(u,v)\in E}} f(u,v) + \sum_{\substack{u\in S,v\in T,\\(v,u)\in E}} f(v,u)\right) .
\end{aligned}$$

Noting that $(u,v) \notin E$ implies $f(u,v) = 0$, we have that

$$\sum_{u\in S,v\in T} f(u,v) = \sum_{\substack{u\in S,v\in T,\\(u,v)\in E}} f(u,v) .$$

Similarly,

$$\sum_{u\in S,v\in T} f(v,u) = \sum_{\substack{u\in S,v\in T,\\(v,u)\in E}} f(v,u) .$$

Thus, the summations of $f(u,v)$ cancel each other out, as do the summations of $f(v,u)$. Therefore,

$$\begin{aligned}
|f| + |f'| &= \sum_{\substack{u\in S,v\in T,\\(u,v)\in E}} c(u,v) \\
&= \sum_{u\in S}\sum_{v\in T} c(u,v) \\
&= c(S,T) .
\end{aligned}$$

■ (lemma)

By this lemma, we see that the value of a maximum flow in G is at most $2K|E|$ more than the value of the current flow f in G. Every time the inner **while** loop finds an augmenting path of capacity at least K, the flow in G increases by at least K. Since the flow cannot increase by more than $2K|E|$, the loop executes at most $(2K|E|)/K = 2|E|$ times.

f. The time complexity is dominated by the **while** loop of lines 4–7. (The lines outside the loop take $O(E)$ time.) The outer **while** loop executes $O(\lg C)$ times, since K is initially $O(C)$ and is halved on each iteration, until $K < 1$. By part (e), the inner **while** loop executes $O(E)$ times for each value of K, and by part (b), each iteration takes $O(E)$ time. Thus, the total time is $O(E^2 \lg C)$.

Solutions for Chapter 27: Multithreaded Algorithms

Solution to Exercise 27.1-1

There will be no change in the asymptotic work, span, or parallelism of P-FIB even if we were to spawn the recursive call to P-FIB$(n-2)$. The serialization of P-FIB under consideration would yield the same recurrence as that for FIB; we can, therefore, calculate the work as $T_1(n) = \Theta(\phi^n)$. Similarly, because the spawned calls to P-FIB$(n-1)$ and P-FIB$(n-2)$ can run in parallel, we can calculate the span in exactly the same way as in the text, $T_\infty(n) = \Theta(n)$, resulting in $\Theta(\phi^n/n)$ parallelism.

Solution to Exercise 27.1-5

By the work law for $P = 4$, we have $80 = T_4 \geq T_1/4$, or $T_1 \leq 320$. By the span law for $P = 64$, we have $T_\infty \leq T_{64} = 10$. Now we will use inequality (27.5) from Exercise 27.1-3 to derive a contradiction. For $P = 10$, we have

$$\begin{aligned} 42 &= T_{10} \\ &\leq \frac{320 - T_\infty}{10} + T_\infty \\ &= 32 + \frac{9}{10} T_\infty \end{aligned}$$

or, equivalently,

$$\begin{aligned} T_\infty &\geq \frac{10}{9} \cdot 10 \\ &> 10 , \end{aligned}$$

which contradicts $T_\infty \leq 10$.

Therefore, the running times reported by the professor are suspicious.

Solution to Exercise 27.1-6

```
FAST-MAT-VEC(A, x)
  n = A.rows
  let y be a new vector of length n
  parallel for i = 1 to n
      y_i = 0
  parallel for i = 1 to n
      y_i = MAT-SUB-LOOP(A, x, i, 1, n)
  return y

MAT-SUB-LOOP(A, x, i, j, j')
  if j == j'
      return a_ij x_j
  else mid = ⌊(j + j')/2⌋
      lhalf = spawn MAT-SUB-LOOP(A, x, i, j, mid)
      uhalf = MAT-SUB-LOOP(A, x, i, mid + 1, j')
      sync
      return lhalf + uhalf
```

We calculate the work $T_1(n)$ of FAST-MAT-VEC by computing the running time of its serialization, i.e., by replacing the **parallel for** loop by an ordinary **for** loop. Therefore, we have $T_1(n) = n\,T_1'(n)$, where $T_1'(n)$ denotes the work of MAT-SUB-LOOP to compute a given output entry y_i. The work of MAT-SUB-LOOP is given by the recurrence

$$T_1'(n) = 2T_1'(n/2) + \Theta(1) .$$

By applying case 1 of the master theorem, we have $T_1'(n) = \Theta(n)$. Therefore, $T_1(n) = \Theta(n^2)$.

To calculate the span, we use

$$T_\infty(n) = \Theta(\lg n) + \max_{1 \le i \le n} iter_\infty(i) .$$

Note that each iteration of the second **parallel for** loop calls procedure MAT-SUB-LOOP with the same parameters, except for the index i. Because MAT-SUB-LOOP recursively halves the space between its last two parameters (1 and n), does constant-time work in the base case, and spawns one of the recursive calls in parallel with the other, it has span $\Theta(\lg n)$. The procedure FAST-MAT-VEC, therefore, has a span of $\Theta(\lg n)$ and $\Theta(n^2/\lg n)$ parallelism.

Solution to Exercise 27.1-7

We analyze the work of P-TRANSPOSE, as usual, by computing the running time of its serialization, where we replace both the **parallel for** loops with simple **for**

loops. We can compute the work of P-TRANSPOSE using the summation

$$\begin{aligned} T_1(n) &= \Theta\left(\sum_{j=2}^{n}(j-1)\right) \\ &= \Theta\left(\sum_{j=1}^{n-1} j\right) \\ &= \Theta(n^2) \, . \end{aligned}$$

The span of P-TRANSPOSE is determined by the span of the doubly nested **parallel for** loops. Although the number of iterations of the inner loop depends on the value of the variable j of the outer loop, each iteration of the inner loop does constant work. Let $iter_\infty(j)$ denote the span of the jth iteration of the outer loop and $iter'_\infty(i)$ denote the span of the ith iteration of the inner loop. We characterize the span $T_\infty(n)$ of P-TRANSPOSE as

$$T_\infty(n) = \Theta(\lg n) + \max_{2 \le j \le n} iter_\infty(j) \, .$$

The maximum occurs when $j = n$, and in this case,

$$iter_\infty(n) = \Theta(\lg n) + \max_{1 \le i \le n-1} iter'_\infty(i) \, .$$

As we noted, each iteration of the inner loop does constant work, and therefore $iter'_\infty(i) = \Theta(1)$ for all i. Thus, we have

$$\begin{aligned} T_\infty(n) &= \Theta(\lg n) + \Theta(\lg n) + \Theta(1) \\ &= \Theta(\lg n) \, . \end{aligned}$$

Since the work P-TRANSPOSE is $\Theta(n^2)$ and its span is $\Theta(\lg n)$, the parallelism of P-TRANSPOSE is $\Theta(n^2/\lg n)$.

Solution to Exercise 27.1-8

If we were to replace the inner **parallel for** loop of P-TRANSPOSE with an ordinary **for** loop, the work would still remain $\Theta(n^2)$. The span, however, would increase to $\Theta(n)$ because the last iteration of the **parallel for** loop, which dominates the span of the computation, would lead to $(n-1)$ iterations of the inner, serial **for** loop. The parallelism, therefore, would reduce to $\Theta(n^2)/\Theta(n) = \Theta(n)$.

Solution to Exercise 27.1-9

Based on the values of work and span given for the two versions of the chess program, we solve for P using

$$\frac{2048}{P} + 1 = \frac{1024}{P} + 8 \, .$$

The solution gives P between 146 and 147.

Solution to Exercise 27.2-3

```
P-FAST-MATRIX-MULTIPLY(A, B)
  n = A.rows
  let C be a new n × n matrix
  parallel for i = 1 to n
      parallel for j = 1 to n
          c_ij = MATRIX-MULT-SUBLOOP(A, B, i, j, 1, n)
  return C

MATRIX-MULT-SUBLOOP(A, B, i, j, k, k')
  if k == k'
      return a_ik b_kj
  else mid = ⌊(k + k')/2⌋
      lhalf = spawn MATRIX-MULT-SUBLOOP(A, B, i, j, k, mid)
      uhalf = MATRIX-MULT-SUBLOOP(A, B, i, j, mid + 1, k')
      sync
      return lhalf + uhalf
```

We calculate the work $T_1(n)$ of P-FAST-MATRIX-MULTIPLY by computing the running time of its serialization, i.e., by replacing the **parallel for** loops by ordinary **for** loops. Therefore, we have $T_1(n) = n^2\,T_1'(n)$, where $T_1'(n)$ denotes the work of MATRIX-MULT-SUBLOOP to compute a given output entry c_{ij}. The work of MATRIX-MULT-SUBLOOP is given by the recurrence

$$T_1'(n) = 2T_1'(n/2) + \Theta(1) .$$

By applying case 1 of the master theorem, we have $T_1'(n) = \Theta(n)$. Therefore, $T_1(n) = \Theta(n^3)$.

To calculate the span, we use

$$T_\infty(n) = \Theta(\lg n) + \max_{1 \le i \le n} iter_\infty(i) .$$

Note that each iteration of the outer **parallel for** loop does the same amount of work: it calls the inner **parallel for** loop. Similarly, each iteration of the inner **parallel for** loop calls procedure MATRIX-MULT-SUBLOOP with the same parameters, except for the indices i and j. Because MATRIX-MULT-SUBLOOP recursively halves the space between its last two parameters (1 and n), does constant-time work in the base case, and spawns one of the recursive calls in parallel with the other, it has span $\Theta(\lg n)$. Since each iteration of the inner **parallel for** loop, which has n iterations, has span $\Theta(\lg n)$, the inner **parallel for** loop has span $\Theta(\lg n)$. By similar logic, the outer **parallel for** loop, and hence procedure P-FAST-MATRIX-MULTIPLY, has span $\Theta(\lg n)$ and $\Theta(n^3/\lg n)$ parallelism.

Solution to Exercise 27.2-4

We can efficiently multiply a $p \times q$ matrix by a $q \times r$ matrix in parallel by using the solution to Exercise 27.2-3 as a base. We will replace the upper limits of the

nested **parallel for** loops with p and r respectively and we will pass q as the last argument to the call of MATRIX-MULT-SUBLOOP. We present the pseudocode for a multithreaded algorithm for multiplying a $p \times q$ matrix by a $q \times r$ matrix in procedure P-GEN-MATRIX-MULTIPLY below. Because the pseudocode for procedure MATRIX-MULT-SUBLOOP (which P-GEN-MATRIX-MULTIPLY calls) remains the same as was presented in the solution to Exercise 27.2-3, we do not repeat it here.

```
P-GEN-MATRIX-MULTIPLY(A, B)
  p = A.rows
  q = A.columns
  r = B.columns
  let C be a new p × r matrix
  parallel for i = 1 to p
      parallel for j = 1 to r
          c_ij = MATRIX-MULT-SUBLOOP(A, B, i, j, 1, q)
  return C
```

To calculate the work for P-GEN-MATRIX-MULTIPLY, we replace the **parallel for** loops with ordinary **for** loops. As before, we can calculate the work of MATRIX-MULT-SUBLOOP to be $\Theta(q)$ (because the input size to the procedure is q here). Therefore, the work of P-GEN-MATRIX-MULTIPLY is $T_1 = \Theta(pqr)$.

We can analyze the span of P-GEN-MATRIX-MULTIPLY as we did in the solution to Exercise 27.2-3, but we must take into account the different number of loop iterations. Each of the p iterations of the outer **parallel for** loop executes the inner **parallel for** loop, and each of the r iterations of the inner **parallel for** loop calls MATRIX-MULT-SUBLOOP, whose span is given by $\Theta(\lg q)$. We know that, in general, the span of a **parallel for** loop with n iterations, where the ith iteration has span $iter_\infty(i)$ is given by

$$T_\infty = \Theta(\lg n) + \max_{1 \le i \le n} iter_\infty(i) \; .$$

Based on the above observations, we can calculate the span of P-GEN-MATRIX-MULTIPLY as

$$\begin{aligned} T_\infty &= \Theta(\lg p) + \Theta(\lg r) + \Theta(\lg q) \\ &= \Theta(\lg(pqr)) \; . \end{aligned}$$

The parallelism of the procedure is, therefore, given by $\Theta(pqr/\lg(pqr))$. To check whether this analysis is consistent with Exercise 27.2-3, we note that if $p = q = r = n$, then the parallelism of P-GEN-MATRIX-MULTIPLY would be $\Theta(n^3/\lg n^3) = \Theta(n^3/\lg n)$.

Solution to Exercise 27.2-5

```
P-MATRIX-TRANSPOSE-RECURSIVE(A, r, c, s)
  // Transpose the s × s submatrix starting at a_rc.
  if s == 1
      return
  else s' = ⌊s/2⌋
      spawn P-MATRIX-TRANSPOSE-RECURSIVE(A, r, c, s')
      spawn P-MATRIX-TRANSPOSE-RECURSIVE(A, r + s', c + s', s − s')
      P-MATRIX-TRANSPOSE-SWAP(A, r, c + s', r + s', c, s', s − s')
      sync

P-MATRIX-TRANSPOSE-SWAP(A, r_1, c_1, r_2, c_2, s_1, s_2)
  // Transpose the s_1 × s_2 submatrix starting at a_{r_1 c_1} with the s_2 × s_1 submatrix
  // starting at a_{r_2 c_2}.
  if s_1 < s_2
      P-MATRIX-TRANSPOSE-SWAP(A, r_2, c_2, r_1, c_1, s_2, s_1)
  elseif s_1 == 1      // since s_1 ≥ s_2, must have that s_2 equals 1
      exchange a_{r_1,c_1} with a_{r_2,c_2}
  else s' = ⌊s_1/2⌋
      spawn P-MATRIX-TRANSPOSE-SWAP(A, r_2, c_2, r_1, c_1, s_2, s')
      P-MATRIX-TRANSPOSE-SWAP(A, r_2, c_2 + s', r_1 + s', c_1, s_2, s_1 − s')
      sync
```

In order to transpose an $n \times n$ matrix A, we call P-MATRIX-TRANSPOSE-RECURSIVE$(A, 1, 1, n)$.

Let us first calculate the work and span of P-MATRIX-TRANSPOSE-SWAP so that we can plug in these values into the work and span calculations of P-MATRIX-TRANSPOSE-RECURSIVE. The work $T_1'(N)$ of P-MATRIX-TRANSPOSE-SWAP on an N-element matrix is the running time of its serialization. We have the recurrence

$$\begin{aligned} T_1'(N) &= 2T_1'(N/2) + \Theta(1) \\ &= \Theta(N) . \end{aligned}$$

The span $T_\infty'(N)$ is similarly described by the recurrence

$$\begin{aligned} T_\infty'(N) &= T_\infty'(N/2) + \Theta(1) \\ &= \Theta(\lg N) . \end{aligned}$$

In order to calculate the work of P-MATRIX-TRANSPOSE-RECURSIVE, we calculate the running time of its serialization. Let $T_1(N)$ be the work of the algorithm on an N-element matrix, where $N = n^2$, and assume for simplicity that n is an exact power of 2. Because the procedure makes two recursive calls with square submatrices of sizes $n/2 \times n/2 = N/4$ and because it does $\Theta(n^2) = \Theta(N)$ work to swap all the elements of the other two submatrices of size $n/2 \times n/2$, its work is given by the recurrence

$$\begin{aligned} T_1(N) &= 2T_1(N/4) + \Theta(N) \\ &= \Theta(N) . \end{aligned}$$

The two parallel recursive calls in P-MATRIX-TRANSPOSE-RECURSIVE execute on matrices of size $n/2 \times n/2$. The span of the procedure is given by maximum of the span of one of these two recursive calls and the $\Theta(\lg N)$ span of P-MATRIX-TRANSPOSE-SWAP, plus $\Theta(1)$. Since the recurrence

$$T_\infty(N) = T_\infty(N/4) + \Theta(1)$$

has the solution $T_\infty(N) = \Theta(\lg N)$ by case 2 of Theorem 4.1, the span of the recursive call is asymptotically the same as the span of P-MATRIX-TRANSPOSE-SWAP, and hence the span of P-MATRIX-TRANSPOSE-RECURSIVE is $\Theta(\lg N)$.

Thus, P-MATRIX-TRANSPOSE-RECURSIVE has parallelism $\Theta(N/\lg N) = \Theta(n^2/\lg n)$.

Solution to Exercise 27.2-6

```
P-FLOYD-WARSHALL(W)
 n = W.rows
 parallel for i = 1 to n
     parallel for j = 1 to n
         d_ij = w_ij
 for k = 1 to n
     parallel for i = 1 to n
         parallel for j = 1 to n
             d_ij = min(d_ij, d_ik + d_kj)
 return D
```

By Exercise 25.2-4, we can compute all the d_{ij} values in parallel.

The work of P-FLOYD-WARSHALL is the same as the running time of its serialization, which we computed as $\Theta(n^3)$ in Section 25.2. The span of the doubly nested **parallel for** loops, which do constant work inside, is $\Theta(\lg n)$. Note, however, that the second set of doubly nested **parallel for** loops is executed within each of the n iterations of the outermost serial **for** loop. Therefore, P-FLOYD-WARSHALL has span $\Theta(n \lg n)$ and $\Theta(n^2/\lg n)$ parallelism.

Solution to Problem 27-1

a. Similar to MAT-VEC-MAIN-LOOP, the required procedure, which we name NESTED-SUM-ARRAYS, will take parameters i and j to specify the range of the array that is being computed in parallel. In order to perform the pairwise addition of two n-element arrays A and B and store the result into array C, we call NESTED-SUM-ARRAYS(A, B, C, 1, $A.length$).

```
NESTED-SUM-ARRAYS(A, B, C, i, j)
  if i == j
      C[i] = A[i] + B[i]
  else k = ⌊(i + j)/2⌋ spawn NESTED-SUM-ARRAYS(A, B, C, i, k)
      NESTED-SUM-ARRAYS(A, B, C, k + 1, j)
      sync
```

The work of NESTED-SUM-ARRAYS is given by the recurrence

$$\begin{aligned} T_1(n) &= 2T_1(n/2) + \Theta(1) \\ &= \Theta(n) , \end{aligned}$$

by case 1 of the master theorem. The span of the procedure is given by the recurrence

$$\begin{aligned} T_\infty(n) &= T_\infty(n/2) + \Theta(1) \\ &= \Theta(\lg n) , \end{aligned}$$

by case 2 of the master theorem. Therefore, the above algorithm has $\Theta(n/\lg n)$ parallelism.

b. Because ADD-SUBARRAY is serial, we can calculate both its work and span to be $\Theta(j - i + 1)$, which based on the arguments from the call in SUM-ARRAYS′ is $\Theta(\textit{grain-size})$, for all but the last call (which is $O(\textit{grain-size})$).

If $\textit{grain-size} = 1$, the procedure SUM-ARRAYS′ calculates r to be n, and each of the n iterations of the serial **for** loop spawns ADD-SUBARRAY with the same value, $k + 1$, for the last two arguments. For example, when $k = 0$, the last two arguments to ADD-SUBARRAY are 1, when $k = 1$, the last two arguments are 2, and so on. That is, in each call to ADD-SUBARRAY, its **for** loop iterates once and calculates a single value in the array C. When $\textit{grain-size} = 1$, the **for** loop in SUM-ARRAYS′ iterates n times and each iteration takes $\Theta(1)$ time, resulting in $\Theta(n)$ work.

Although the **for** loop in SUM-ARRAYS′ looks serial, note that each iteration spawns the call to ADD-SUBARRAY and the procedure waits for all its spawned children at the end of the **for** loop. That is, all loop iterations of SUM-ARRAYS′ execute in parallel. Therefore, one might be tempted to say that the span of SUM-ARRAYS′ is equal to the span of a single call to ADD-SUBARRAY plus the constant work done by the first three lines in SUM-ARRAYS′, giving $\Theta(1)$ span and $\Theta(n)$ parallelism. This calculation of span and parallelism would be wrong, however, because there are r spawns of ADD-SUBARRAY in SUM-ARRAYS′, where r is not a constant. Hence, we must add a $\Theta(r)$ term to the span of SUM-ARRAYS′ in order to account for the overhead of spawning r calls to ADD-SUBARRAY.

Based on the above discussion, the span of SUM-ARRAYS′ is $\Theta(r) + \Theta(\textit{grain-size}) + \Theta(1)$. When $\textit{grain-size} = 1$, we get $r = n$; therefore, SUM-ARRAYS′ has $\Theta(n)$ span and $\Theta(1)$ parallelism.

c. For a general *grain-size*, each iteration of the **for** loop in SUM-ARRAYS′ except for the last results in *grain-size* iterations of the **for** loop in ADD-SUBARRAY. In the last iteration of SUM-ARRAYS′, the **for** loop in ADD-SUBARRAY iterates n mod *grain-size* times. Therefore, we can claim that the span of ADD-SUBARRAY is $\Theta(\max(\textit{grain-size}, n \bmod \textit{grain-size})) = \Theta(\textit{grain-size})$.

SUM-ARRAYS′ achieves maximum parallelism when its span, given by $\Theta(r) + \Theta(\textit{grain-size}) + \Theta(1)$, is minimum. Since $r = \lceil n/\textit{grain-size} \rceil$, the minimum occurs when $r \approx \textit{grain-size}$, i.e., when $\textit{grain-size} \approx \sqrt{n}$.

Solution to Problem 27-2

a. We initialize the output matrix C using doubly nested **parallel for** loops and then call P-MATRIX-MULTIPLY-RECURSIVE′, defined below.

```
P-MATRIX-MULTIPLY-LESS-MEM(C, A, B)
n = A.rows
parallel for i = 1 to n
    parallel for j = 1 to n
        c_ij = 0
P-MATRIX-MULTIPLY-RECURSIVE′(C, A, B)
```

```
P-MATRIX-MULTIPLY-RECURSIVE′(C, A, B)
n = A.rows
if n == 1
    c_11 = c_11 + a_11 b_11
else partition A, B, and C into n/2 × n/2 submatrices
        A_11, A_12, A_21, A_22; B_11, B_12, B_21, B_22; and C_11, C_12, C_21, C_22
    spawn P-MATRIX-MULTIPLY-RECURSIVE′(C_11, A_11, B_11)
    spawn P-MATRIX-MULTIPLY-RECURSIVE′(C_12, A_11, B_12)
    spawn P-MATRIX-MULTIPLY-RECURSIVE′(C_21, A_21, B_11)
    P-MATRIX-MULTIPLY-RECURSIVE′(C_22, A_21, B_12)
    sync
    spawn P-MATRIX-MULTIPLY-RECURSIVE′(C_11, A_12, B_21)
    spawn P-MATRIX-MULTIPLY-RECURSIVE′(C_12, A_12, B_22)
    spawn P-MATRIX-MULTIPLY-RECURSIVE′(C_21, A_22, B_21)
    P-MATRIX-MULTIPLY-RECURSIVE′(C_22, A_22, B_22)
    sync
```

b. The procedure P-MATRIX-MULTIPLY-LESS-MEM performs $\Theta(n^2)$ work in the doubly nested **parallel for** loops, and then it calls the procedure P-MATRIX-MULTIPLY-RECURSIVE′. The recurrence for the work $M_1'(n)$ of P-MATRIX-MULTIPLY-RECURSIVE′ is $8M_1'(n/2) + \Theta(1)$, which gives us $M_1'(n) = \Theta(n^3)$. Therefore, $T_1(n) = \Theta(n^3)$.

The span of the doubly nested **parallel for** loops that initialize the output array C is $\Theta(\lg n)$. In P-MATRIX-MULTIPLY-RECURSIVE′, there are two groups of spawned recursive calls; therefore, the span $M_\infty'(n)$ of P-MATRIX-MULTIPLY-RECURSIVE′ is given by the recurrence $M_\infty'(n) = 2M_\infty'(n/2) + \Theta(1)$, which gives us $M_\infty'(n) = \Theta(n)$. Because the span $\Theta(n)$ of P-MATRIX-MULTIPLY-RECURSIVE′ dominates, we have $T_\infty(n) = \Theta(n)$.

c. The parallelism of P-MATRIX-MULTIPLY-LESS-MEM is $\Theta(n^3/n) = \Theta(n^2)$. Ignoring the constants in the Θ-notation, the parallelism for multiplying 1000×1000 matrices is $1000^2 = 10^6$, which is only a factor of 10 less than that of P-MATRIX-MULTIPLY-RECURSIVE. Although the parallelism of the new procedure is much less than that of P-MATRIX-MULTIPLY-RECURSIVE, the algorithm still scales well for a large number of processors.

Solution to Problem 27-4

a. Here is a multithreaded $\otimes$-reduction algorithm:

```
P-REDUCE(x, i, j)
 if i == j
     return x[i]
 else mid = ⌊(i + j)/2⌋
     lh = spawn P-REDUCE(x, i, mid)
     rh = P-REDUCE(x, mid + 1, j)
     sync
     return lh ⊗ rh
```

If we denote the length $j - i + 1$ of the subarray $x[i \,..\, j]$ by n, then the work for the above algorithm is given by the recurrence $T_1(n) = 2T_1(n/2) + \Theta(1) = \Theta(n)$. Because one of the recursive calls to P-REDUCE is spawned and the procedure does constant work following the recursive calls and in the base case, the span is given by the recurrence $T_\infty(n) = T_\infty(n/2) + \Theta(1) = \Theta(\lg n)$.

b. The work and span of P-SCAN-1-AUX dominate the work and span of P-SCAN-1. We can calculate the work of P-SCAN-1-AUX by replacing the **parallel for** loop with an ordinary **for** loop and noting that in each iteration, the running time of P-REDUCE will be equal to $\Theta(l)$. Since P-SCAN-1 calls P-SCAN-1-AUX with 1 and n as the last two arguments, the running time of P-SCAN-1, and hence its work, is $\Theta(1 + 2 + \cdots + n) = \Theta(n^2)$.

As we noted earlier, the **parallel for** loop in P-SCAN-1-AUX undergoes n iterations; therefore, the span of P-SCAN-1-AUX is given by $\Theta(\lg n)$ for the recursive splitting of the loop iterations plus the span of the iteration that has maximum span. Among the loop iterations, the call to P-REDUCE in the last iteration (when $l = n$) has the maximum span, equal to $\Theta(\lg n)$. Thus, P-SCAN-1 has $\Theta(\lg n)$ span and $\Theta(n^2/\lg n)$ parallelism.

c. In P-SCAN-2-AUX, before the **parallel for** loop in lines 7 and 8 executes, the following invariant is satisfied: $y[l] = x[i] \otimes x[i+1] \otimes \cdots \otimes x[l]$ for $l = i, i+1, \ldots, k$ and $y[l] = x[k+1] \otimes x[k+2] \otimes \cdots \otimes x[l]$ for $l = k+1, k+2, \ldots, j$. The **parallel for** loop need not update $y[i], \ldots, y[k]$, since they have the correct values after the call to P-SCAN-2-AUX(x, y, i, k). For $l = k+1, k+2, \ldots, j$, the **parallel for** loop sets

$$\begin{aligned} y[l] &= y[k] \otimes y[l] \\ &= x[i] \otimes \cdots \otimes x[k] \otimes x[k+1] \otimes \cdots \otimes x[l] \end{aligned}$$

$$= \quad x[i] \otimes \cdots \otimes x[l] \,,$$

as desired. We can run this loop in parallel because the lth iteration depends only on the values of $y[k]$, which is the same in all iterations, and $y[l]$. Therefore, when the call to P-SCAN-2-AUX from P-SCAN-2 returns, array y represents the $\otimes$-prefix computation of array x.

Because the work and span of P-SCAN-2-AUX dominate the work and span of P-SCAN-2, we will concentrate on calculating these values for P-SCAN-2-AUX working on an array of size n. The work $PS2A_1(n)$ of P-SCAN-2-AUX is given by the recurrence $PS2A_1(n) = 2PS2A_1(n/2) + \Theta(n)$, which equals $\Theta(n \lg n)$ by case 2 of the master theorem. The span $PS2A_\infty(n)$ of P-SCAN-2-AUX is given by the recurrence $PS2A_\infty(n) = PS2A_\infty(n/2) + \Theta(\lg n)$, which equals $\Theta(\lg^2 n)$ per Exercise 4.6-2. That is, the work, span, and parallelism of P-SCAN-2 are $\Theta(n \lg n)$, $\Theta(\lg^2 n)$, and $\Theta(n/\lg n)$, respectively.

d. The missing expression in line 8 of P-SCAN-UP is $t[k] \otimes \mathit{right}$. The missing expressions in lines 5 and 6 of P-SCAN-DOWN are v and $v \otimes t[k]$, respectively.

As suggested in the hint, we will prove that the value v passed to P-SCAN-DOWN(v, x, t, y, i, j) satisfies $v = x[1] \otimes x[2] \otimes \cdots \otimes x[i-1]$, so that the value $v \otimes x[i]$ stored into $y[i]$ in the base case of P-SCAN-DOWN is correct.

In order to compute the arguments that are passed to P-SCAN-DOWN, we must first understand what $t[k]$ holds as a result of the call to P-SCAN-UP. A call to P-SCAN-UP(x, t, i, j) returns $x[i] \otimes \cdots \otimes x[j]$; because $t[k]$ stores the return value of P-SCAN-UP(x, t, i, k), we can say that $t[k] = x[i] \otimes \cdots \otimes x[k]$.

The value $v = x[1]$ when P-SCAN-DOWN$(x[1], x, t, y, 2, n)$ is called from P-SCAN-3 clearly satisifies $v = x[1] \otimes \cdots \otimes x[i-1]$. Let us suppose that $v = x[1] \otimes x[2] \otimes \cdots \otimes x[i-1]$ in a call of P-SCAN-DOWN(v, x, t, y, i, j). Therefore, v meets the required condition in the first recursive call, with i and k as the last two arguments, in P-SCAN-DOWN. If we can prove that the value $v \otimes t[k]$ passed to the second recursive call in P-SCAN-DOWN equals $x[1] \otimes x[2] \otimes \cdots \otimes x[k]$, we would have proved the required condition on v for all calls to P-SCAN-DOWN. Earlier, we proved that $t[k] = x[i] \otimes \cdots \otimes x[k]$; therefore,

$$\begin{aligned} v \otimes t[k] &= x[1] \otimes x[2] \otimes \cdots \otimes x[i-1] \otimes x[i] \otimes \cdots x[k] \\ &= x[1] \otimes x[2] \otimes \cdots \otimes x[k] \,. \end{aligned}$$

Thus, the value v passed to P-SCAN-DOWN(v, x, t, y, i, j) satisfies $v = x[1] \otimes x[2] \otimes \cdots \otimes x[i-1]$.

e. Let $PSU_1(n)$ and $PSU_\infty(n)$ denote the work and span of P-SCAN-UP and let $PSD_1(n)$ and $PSD_\infty(n)$ denote the work and span of P-SCAN-DOWN. Then the expressions $T_1(n) = PSU_1(n) + PSD_1(n) + \Theta(1)$ and $T_\infty(n) = PSU_\infty(n) + PSD_\infty(n) + \Theta(1)$ characterize the work and span of P-SCAN-3.

The work $PSU_1(n)$ of P-SCAN-UP is given by the recurrence

$$PSU_1(n) = 2PSU_1(n/2) + \Theta(1) \,,$$

and its span is defined by the recurrence

$$PSU_\infty(n) = PSU_\infty(n/2) + \Theta(1) \; .$$

Using the master theorem to solve these recurrences, we get $PSU_1(n) = \Theta(n)$ and $PSU_\infty(n) = \Theta(\lg n)$.

Similarly, the recurrences

$$PSD_1(n) = 2PSD_1(n/2) + \Theta(1) \; , \qquad (*)$$

$$PSD_\infty(n) = PSD_\infty(n/2) + \Theta(1) \qquad (\dagger)$$

define the work and span of P-SCAN-DOWN, and they evaluate to $PSD_1(n) = \Theta(n)$ and $PSD_\infty(n) = \Theta(\lg n)$.

Applying the results for the work and span of P-SCAN-UP and P-SCAN-DOWN obtained above in the expressions for the work and span of P-SCAN-3, we get $T_1(n) = \Theta(n)$ and $T_\infty(n) = \Theta(\lg n)$. Hence, P-SCAN-3 has $\Theta(n/\lg n)$ parallelism. P-SCAN-3 performs less work than P-SCAN-1, but with the same span, and it has the same parallelism as P-SCAN-2 with less work and a lower span.

Solution to Problem 27-5

a. In this part of the problem, we will assume that n is an exact power of 2, so that in a recursive step, when we divide the $n \times n$ matrix A into four $n/2 \times n/2$ matrices, we will be guaranteed that $n/2$ is an integer, for all $n \geq 2$. We make this assumption simply to avoid introducing $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ terms in the pseudocode and the analysis that follow. In the pseudocode below, we assume that we have a procedure BASE-CASE available to us, which calculates the base case of the stencil.

```
SIMPLE-STENCIL(A, i, j, n)
if n == 1
    A[i, j] = BASE-CASE(A, i, j)
else // Calculate submatrix A11.
    SIMPLE-STENCIL(A, i, j, n/2)
    // Calculate submatrices A12 and A21 in parallel.
    spawn SIMPLE-STENCIL(A, i, j + n/2, n/2)
    SIMPLE-STENCIL(A, i + n/2, j, n/2)
    sync
    // Calculate submatrix A22.
    SIMPLE-STENCIL(A, i + n/2, j + n/2, n/2)
```

To perform a simple stencil calculation on an $n \times n$ matrix A, we call SIMPLE-STENCIL$(A, 1, 1, n)$. The recurrence for the work is $T_1(n) = 4T_1(n/2) + \Theta(1) = \Theta(n^2)$. Of the four recursive calls in the algorithm above, only two run in parallel. Therefore, the recurrence for the span is $T_\infty(n) = 3T_\infty(n/2) + \Theta(1) = \Theta(n^{\lg 3})$, and the parallelism is $\Theta(n^{2-\lg 3}) \approx \Theta(n^{0.415})$.

b. Similar to SIMPLE-STENCIL of the previous part, we present P-STENCIL-3, which divides A into nine submatrices, each of size $n/3 \times n/3$, and solves them

recursively. To perform a stencil calculation on an $n \times n$ matrix A, we call P-STENCIL-3$(A, 1, 1, n)$.

```
P-STENCIL-3(A, i, j, n)
 if n == 1
     A[i, j] = BASE-CASE(A, i, j)
 else // Group 1: compute submatrix A_11.
     P-STENCIL-3(A, i, j, n/3)
     // Group 2: compute submatrices A_12 and A_21.
     spawn P-STENCIL-3(A, i, j + n/3, n/3)
     P-STENCIL-3(A, i + n/3, j, n/3)
     sync
     // Group 3: compute submatrices A_13, A_22, and A_31.
     spawn P-STENCIL-3(A, i, j + 2n/3, n/3)
     spawn P-STENCIL-3(A, i + n/3, j + n/3, n/3)
     P-STENCIL-3(A, i + 2n/3, j, n/3)
     sync
     // Group 4: compute submatrices A_23 and A_32.
     spawn P-STENCIL-3(A, i + n/3, j + 2n/3, n/3)
     P-STENCIL-3(A, i + 2n/3, j + n/3, n/3)
     sync
     // Group 5: compute submatrix A_33.
     P-STENCIL-3(A, i + 2n/3, j + 2n/3, n/3)
```

From the pseudocode, we can informally say that we can solve the nine subproblems in five groups, as shown in the following matrix:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}.$$

Each entry in the above matrix specifies the group of the corresponding $n/3 \times n/3$ submatrix of A; we can compute in parallel the entries of all submatrices that fall in the same group. In general, for $i = 2, 3, 4, 5$, we can calculate group i after completing the computation of group $i - 1$.

The recurrence for the work is $T_1(n) = 9T_1(n/3) + \Theta(1) = \Theta(n^2)$. The recurrence for the span is $T_\infty(n) = 5T_\infty(n/3) + \Theta(1) = \Theta(n^{\log_3 5})$. Therefore, the parallelism is $\Theta(n^{2-\log_3 5}) \approx \Theta(n^{0.535})$.

c. Similar to the previous part, we can solve the b^2 subproblems in $2b - 1$ groups:

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & b-2 & b-1 & b \\ 2 & 3 & 4 & \cdots & b-1 & b & b+1 \\ 3 & 4 & 5 & \cdots & b & b+1 & b+2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ b-2 & b-1 & b & \cdots & 2b-5 & 2b-4 & 2b-3 \\ b-1 & b & b+1 & \cdots & 2b-4 & 2b-3 & 2b-2 \\ b & b+1 & b+2 & \cdots & 2b-3 & 2b-2 & 2b-1 \end{pmatrix}.$$

The recurrence for the work is $T_1(n) = b^2 T_1(n/b) + \Theta(1) = \Theta(n^2)$. The recurrence for the span is $T_\infty(n) = (2b-1)T_\infty(n/b) + \Theta(1) = \Theta(n^{\log_b(2b-1)})$. The parallelism is $\Theta(n^{2-\log_b(2b-1)})$.

As the hint suggests, in order to show that the parallelism must be $o(n)$ for any choice of $b \geq 2$, we need to show that $2 - \log_b(2b-1)$, which is the exponent of n in the parallelism, is strictly less than 1 for any choice of $b \geq 2$. Since $b \geq 2$, we know that $2b-1 > b$, which implies that $\log_b(2b-1) > \log_b b = 1$. Hence, $2 - \log_b(2b-1) < 2 - 1 = 1$.

d. The idea behind achieving $\Theta(n/\lg n)$ parallelism is similar to that presented in the previous part, except without recursive division. We will compute $A[1,1]$ serially, which will enable us to compute entries $A[1,2]$ and $A[2,1]$ in parallel, after which we can compute entries $A[1,3]$, $A[2,2]$ and $A[3,1]$ in parallel, and so on. Here is the pseudocode:

```
P-STENCIL(A)
n = A.rows
// Calculate all entries on the antidiagonal and above it.
for i = 1 to n
    parallel for j = 1 to i
        A[i - j + 1, j] = BASE-CASE(A, i - j + 1, j)
// Calculate all entries below the antidiagonal.
for i = 2 to n
    parallel for j = i to n
        A[n + i - j, j] = BASE-CASE(A, n + i - j, j)
```

For each value of index i of the first serial **for** loop, the inner loop iterates i times, doing constant work in each iteration. Because index i ranges from 1 to n in the first **for** loop, we require $\Theta(1 + 2 + \cdots + n) = \Theta(n^2)$ work to calculate all entries on the antidiagonal and above it. For each value of index i of the second serial **for** loop, the inner loop iterates $n - i + 1$ times, doing constant work in each iteration. Because index i ranges from 2 to n in the second **for** loop, we require $\Theta((n-1) + (n-2) + \cdots + 1) = \Theta(n^2)$ work to calculate all entries on the antidiagonal and above it. Therefore, the work of P-STENCIL is $T_1(n) = \Theta(n^2)$.

Note that both **for** loops in P-STENCIL, which execute **parallel for** loops within, are serial. Therefore, in order to calculate the span of P-STENCIL, we must add the spans of all the **parallel for** loops. Given that any **parallel for** loop in P-STENCIL does constant work in each iteration, the span of a **parallel for** loop with n' iterations is $\Theta(\lg n')$. Hence,

$$\begin{aligned} T_\infty(n) &= \Theta((\lg 1 + \lg 2 + \cdots + \lg n) + (\lg(n-1) + \cdots + 1)) \\ &= \Theta(\lg(n!) + \lg(n-1)!) \\ &= \Theta(n \lg n) \,, \end{aligned}$$

giving us $\Theta(n/\lg n)$ parallelism.

Index

This index covers exercises and problems from the textbook that are solved in this manual. The first page in the manual that has the solution is listed here.