Matrix Reference Manual

Matrix Calculus

Go to: Introduction, Notation, Index

Contents

Notation

- $d/dx\ (\mathbf{y})$ is a vector whose (i) element is $dy(i)/dx$
- $d/d\mathbf{x}\ (y)$ is a vector whose (i) element is $dy/dx(i)$
- $d/d\mathbf{x}\ (\mathbf{y}^T)$ is a matrix whose (i,j) element is $dy(j)/dx(i)$
- $d/dx\ (\mathbf{Y})$ is a matrix whose (i,j) element is $dy(i,j)/dx$
- $d/d\mathbf{X}\ (y)$ is a matrix whose (i,j) element is $dy/dx(i,j)$

Note that the Hermitian transpose is not used because complex conjugates are not analytic.

In the expressions below matrices and vectors **A**, **B**, **C** do not depend on **X**.

Derivatives of Linear Products

- $d/dx\ (\mathbf{AYB}) = \mathbf{A} * d/dx\ (\mathbf{Y}) * \mathbf{B}$
  - $d/dx\ (\mathbf{Ay}) = \mathbf{A} * d/dx\ (\mathbf{y})$
- $d/d\mathbf{x}\ (\mathbf{x}^T\mathbf{A}) = \mathbf{A}$
  - $d/d\mathbf{x}\ (\mathbf{x}^T) = \mathbf{I}$
  - $d/d\mathbf{x}\ (\mathbf{x}^T\mathbf{a}) = d/d\mathbf{x}\ (\mathbf{a}^T\mathbf{x}) = \mathbf{a}$
- $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}\mathbf{b}) = \mathbf{a}\mathbf{b}^T$
  - $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}\mathbf{a}) = d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}^T\mathbf{a}) = \mathbf{a}\mathbf{a}^T$
- $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}^T\mathbf{b}) = \mathbf{b}\mathbf{a}^T$
- $d/dx\ (\mathbf{YZ}) = \mathbf{Y} * d/dx\ (\mathbf{Z}) + d/dx\ (\mathbf{Y}) * \mathbf{Z}$

Derivatives of Quadratic Products

- $d/d\mathbf{x}\ (\mathbf{Ax}+\mathbf{b})^T\mathbf{C}(\mathbf{D}x+\mathbf{e}) = \mathbf{A}^T\mathbf{C}(\mathbf{Dx}+\mathbf{e}) + \mathbf{D}^T\mathbf{C}^T(\mathbf{Ax}+\mathbf{b})$
  - $d/d\mathbf{x}\ (\mathbf{x}^T\mathbf{Cx}) = (\mathbf{C}+\mathbf{C}^T)\mathbf{x}$
    - [C: symmetric]: $d/d\mathbf{x}\ (\mathbf{x}^T\mathbf{Cx}) = 2\mathbf{Cx}$
    - $d/d\mathbf{x}\ (\mathbf{x}^T\mathbf{x}) = 2\mathbf{x}$
  - $d/d\mathbf{x}\ (\mathbf{Ax}+\mathbf{b})^T\ (\mathbf{D}x+\mathbf{e}) = \mathbf{A}^T\ (\mathbf{Dx}+\mathbf{e}) + \mathbf{D}^T\ (\mathbf{Ax}+\mathbf{b})$

        - d/dx $(\mathbf{Ax}+\mathbf{b})^T(\mathbf{Ax}+\mathbf{b}) = 2\mathbf{A}^T(\mathbf{Ax}+\mathbf{b})$
    - [C: symmetric]: d/dx $(\mathbf{Ax}+\mathbf{b})^T\mathbf{C}(\mathbf{Ax}+\mathbf{b}) = 2\mathbf{A}^T\mathbf{C}(\mathbf{Ax}+\mathbf{b})$
- $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}^T\mathbf{Xb}) = \mathbf{X}(\mathbf{ab}^T + \mathbf{ba}^T)$
    - $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}^T\mathbf{Xa}) = 2\mathbf{Xaa}^T$
- $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}^T\mathbf{CXb}) = \mathbf{C}^T\mathbf{Xab}^T + \mathbf{CXba}^T$
    - $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}^T\mathbf{CXa}) = (\mathbf{C} + \mathbf{C}^T)\mathbf{Xaa}^T$
    - [C:*Symmetric*] $d/d\mathbf{X}\ (\mathbf{a}^T\mathbf{X}^T\mathbf{CXa}) = 2\mathbf{CXaa}^T$
- $d/d\mathbf{X}\ ((\mathbf{Xa}+\mathbf{b})^T\mathbf{C}(\mathbf{Xa}+\mathbf{b})) = (\mathbf{C}+\mathbf{C}^T)(\mathbf{Xa}+\mathbf{b})\mathbf{a}^T$

Derivatives of Cubic Products

- $d/d\mathbf{x}\ (\mathbf{x}^T\mathbf{Axx}^T) = (\mathbf{A}+\mathbf{A}^T)\mathbf{xx}^T+\mathbf{x}^T\mathbf{AxI}$

Derivatives of Inverses

- $d/dx\ (\mathbf{Y}^{-1}) = -\mathbf{Y}^{-1}d/dx\ (\mathbf{Y})\mathbf{Y}^{-1}$

Derivative of Trace

Note: matrix dimensions must result in an $n*n$ argument for tr().

- $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{X})) = \mathbf{I}$
- $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{X}^k)) = k(\mathbf{X}^{k-1})^T$
- $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{AX}^k)) = \mathbf{SUM}_{r=0:k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T$
- $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{AX}^{-1}\mathbf{B})) = -(\mathbf{X}^{-1}\mathbf{BAX}^{-1})^T$
    - $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{AX}^{-1})) = d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A})) = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{X}^{-T}$
- $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{A}^T\mathbf{XB}^T)) = d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{BX}^T\mathbf{A})) = \mathbf{AB}$
    - $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{XA}^T)) = d/d\mathbf{X}\ (\mathbf{tr}(\mathbf{A}^T\mathbf{X})) = d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{X}^T\mathbf{A})) = d/d\mathbf{X}\ (\mathbf{tr}(\mathbf{AX}^T)) = \mathbf{A}$
- $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{AXBX}^T)) = \mathbf{A}^T\mathbf{XB}^T + \mathbf{AXB}$
    - $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{XAX}^T)) = \mathbf{X}(\mathbf{A}+\mathbf{A}^T)$
    - $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{X}^T\mathbf{AX})) = \mathbf{X}^T(\mathbf{A}+\mathbf{A}^T)$
    - $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{AX}^T\mathbf{X})) = (\mathbf{A}+\mathbf{A}^T)\mathbf{X}$
- $d/d\mathbf{X}\ (\mathrm{tr}(\mathbf{AXBX})) = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T + \mathbf{B}^T\mathbf{X}^T\mathbf{A}^T$
-
- [C:*symmetric*] $d/d\mathbf{X}\ (\mathbf{tr}((\mathbf{X}^T\mathbf{CX})^{-1}\mathbf{A}) = d/d\mathbf{X}\ (\mathbf{tr}(\mathbf{A}\ (\mathbf{X}^T\mathbf{CX})^{-1}) = -(\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})(\mathbf{A}+\mathbf{A}^T)(\mathbf{X}^T\mathbf{CX})^{-1}$
- [B,C:*symmetric*] $d/d\mathbf{X}\ (\mathbf{tr}((\mathbf{X}^T\mathbf{CX})^{-1}(\mathbf{X}^T\mathbf{BX})) = d/d\mathbf{X}\ (\mathbf{tr}(\ (\mathbf{X}^T\mathbf{BX})(\mathbf{X}^T\mathbf{CX})^{-1}) =$ $-2(\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{X}^T\mathbf{BX}(\mathbf{X}^T\mathbf{CX})^{-1} + 2\mathbf{BX}(\mathbf{X}^T\mathbf{CX})^{-1}$
-

Derivative of Determinant

Note: matrix dimensions must result in an $n*n$ argument for det().

- $d/d\mathbf{X}\ (\det(\mathbf{X})) = d/d\mathbf{X}\ (\det(\mathbf{X}^T)) = \det(\mathbf{X})*\mathbf{X}^{-T}$
    - $d/d\mathbf{X}\ (\det(\mathbf{AXB})) = \det(\mathbf{AXB})*\mathbf{X}^{-T}$

- $d/d\mathbf{X}$ (ln(det(**AXB**))) = $\mathbf{X}^{-T}$

- $d/d\mathbf{X}$ (det($\mathbf{X}^k$)) = k*det($\mathbf{X}^k$)*$\mathbf{X}^{-T}$
  - $d/d\mathbf{X}$ (ln(det($\mathbf{X}^k$))) = $k\mathbf{X}^{-T}$
- [*Real*] $d/d\mathbf{X}$ (det($\mathbf{X}^T\mathbf{C}\mathbf{X}$)) = det($\mathbf{X}^T\mathbf{C}\mathbf{X}$)*($\mathbf{C}+\mathbf{C}^T$)$\mathbf{X}$($\mathbf{X}^T\mathbf{C}\mathbf{X}$)$^{-1}$
  - [**C**: *Real,Symmetric*] $d/d\mathbf{X}$ (det($\mathbf{X}^T\mathbf{C}\mathbf{X}$)) = 2det($\mathbf{X}^T\mathbf{C}\mathbf{X}$)* $\mathbf{C}\mathbf{X}$($\mathbf{X}^T\mathbf{C}\mathbf{X}$)$^{-1}$
- [**C**: *Real,Symmetricc*] $d/d\mathbf{X}$ (ln(det($\mathbf{X}^T\mathbf{C}\mathbf{X}$))) = 2$\mathbf{C}\mathbf{X}$($\mathbf{X}^T\mathbf{C}\mathbf{X}$)$^{-1}$

Jacobian

If **y** is a function of **x**, then $d\mathbf{y}^T/d\mathbf{x}$ is the Jacobian matrix of **y** with respect to **x**.

Its determinant, $|d\mathbf{y}^T/d\mathbf{x}|$, is the *Jacobian* of **y** with respect to **x** and represents the ratio of the hyper-volumes $d\mathbf{y}$ and $d\mathbf{x}$. The Jacobian occurs when changing variables in an integration: Integral(f(**y**)d**y**)=Integral(f(**y**(**x**)) $|d\mathbf{y}^T/d\mathbf{x}|$ d**x**).

Hessian matrix

If f is a function of **x** then the symmetric matrix $d^2f/d\mathbf{x}^2 = \mathbf{d}/d\mathbf{x}^T(df/d\mathbf{x})$ is the *Hessian* matrix of f(**x**). A value of **x** for which df/d**x** = **0** corresponds to a minimum, maximum or saddle point according to whether the Hessian is positive definite, negative definite or indefinite.

- $d^2/d\mathbf{x}^2$ ($\mathbf{a}^T\mathbf{x}$) = 0
- $d^2/d\mathbf{x}^2$ $(\mathbf{A}\mathbf{x}+\mathbf{b})^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{e}) = \mathbf{A}^T\mathbf{C}\mathbf{D} + \mathbf{D}^T\mathbf{C}^T\mathbf{A}$
  - $d^2/d\mathbf{x}^2$ ($\mathbf{x}^T\mathbf{C}\mathbf{x}$) = $\mathbf{C}+\mathbf{C}^T$
    - $d^2/d\mathbf{x}^2$ ($\mathbf{x}^T\mathbf{x}$) = 2**I**
  - $d^2/d\mathbf{x}^2$ $(\mathbf{A}\mathbf{x}+\mathbf{b})^T(\mathbf{D}\mathbf{x}+\mathbf{e}) = \mathbf{A}^T\mathbf{D} + \mathbf{D}^T\mathbf{A}$
    - $d^2/d\mathbf{x}^2$ $(\mathbf{A}\mathbf{x}+\mathbf{b})^T(\mathbf{A}\mathbf{x}+\mathbf{b}) = 2\mathbf{A}^T\mathbf{A}$
  - [C: symmetric]: $d^2/d\mathbf{x}^2$ $(\mathbf{A}\mathbf{x}+\mathbf{b})^T\mathbf{C}(\mathbf{A}\mathbf{x}+\mathbf{b}) = 2\mathbf{A}^T\mathbf{C}\mathbf{A}$

The Matrix Reference Manual is written by Mike Brookes, Imperial College, London, UK. Please send any comments or suggestions to mike.brookes@ic.ac.uk

---