Matrix Calculus

Go to: Introduction, Notation, Index

Contents of Calculus Section

Notation

- j is the square root of -1
- $\mathbf{X}^R$ and $\mathbf{X}^I$ are the real and imaginary parts of $\mathbf{X} = \mathbf{X}^R + j\mathbf{X}^I$
- $\mathbf{X}^C$ is the complex conjugate of $\mathbf{X}$
- **X:** denotes the long column vector formed by concatenating the columns of **X** (see vectorization).
- **A** ¤ **B** = **KRON**(**A**,**B**), the kroneker product
- **A** • **B** the Hadamard or elementwise product
- matrices and vectors **A**, **B**, **C** do not depend on **X**

Derivatives

In the main part of this page we express results in terms of differentials rather than derivatives for two reasons: they avoid notational disagreements and they cope easily with the complex case. In most cases however, the differentials have been written in the form $d\mathbf{Y}\mathbf{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:}$ so that the corresponding derivative may be easily extracted.

Derivatives with respect to a real matrix

If **X** is $p\#q$ and **Y** is $m\#n$, then $d\mathbf{Y}\mathbf{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:}$ where the derivative $d\mathbf{Y}/d\mathbf{X}$ is a large $mn\#pq$ matrix. If **X** and/or **Y** are column vectors or scalars, then the vectorization operator **:** has no effect and may be omitted. $d\mathbf{Y}/d\mathbf{X}$ is also called the *Jacobian Matrix* of **Y:** with respect to **X:** and det($d\mathbf{Y}/d\mathbf{X}$) is the corresponding *Jacobian*. The Jacobian occurs when changing variables in an integration: Integral(f(**Y**)d**Y:**)=Integral(f(**Y**(**X**)) det($d\mathbf{Y}/d\mathbf{X}$) d**X:**).

Although they do not generalise so well, other authors use alternative notations for the cases when **X** and **Y** are both vectors or when one is a scalar. In particular:

- $dy/d\mathbf{x}$ is sometimes written as a column vector rather than a row vector
- $d\mathbf{y}/d\mathbf{x}$ is sometimes transposed from the above definition or else is sometimes written $d\mathbf{y}/d\mathbf{x}^T$ to emphasise the correspondence between the columns of the derivative and those of $\mathbf{x}^T$.
- $d\mathbf{Y}/dx$ and $dy/d\mathbf{X}$ are often written as matrices rather than, as here, a column vector and row vector respectively. The matrix form may be converted to the form used here by appending **:** or $\mathbf{:}^T$ respectively.

Derivatives with respect to a complex matrix

If $\mathbf{X}$ is complex then $d\mathbf{Y}\mathbf{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:}$ can only be true iff $\mathbf{Y}(\mathbf{X})$ is an analytic function which normally implies that $\mathbf{Y}(\mathbf{X})$ does not depend on $\mathbf{X}^C$ or $\mathbf{X}^H$.

Even for non-analytic functions we can write uniquely $d\mathbf{Y}\mathbf{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:} + d\mathbf{Y}/d\mathbf{X}^C\ d\mathbf{X}^C\mathbf{:}$ provided that is analytic with respect to $\mathbf{X}$ and $\mathbf{X}^C$ individually (or equivalently with respect to $\mathbf{X}^R$ and $\mathbf{X}^I$ individually). $d\mathbf{Y}/d\mathbf{X}$ is the *Generalized Complex Derivative* and $d\mathbf{Y}/d\mathbf{X}^C$ is the *Complex Conjugate Derivative* [R.4, R.9].

We define the generalized derivatives in terms of partial derivatives with respect to $\mathbf{X}^R$ and $\mathbf{X}^I$:

- $d\mathbf{Y}/d\mathbf{X} = ½\ (d\mathbf{Y}/d\mathbf{X}^R - \mathrm{j}\ d\mathbf{Y}/d\mathbf{X}^I)$
- $d\mathbf{Y}/d\mathbf{X}^C = (d\mathbf{Y}^C/d\mathbf{X})^C = ½\ (d\mathbf{Y}/d\mathbf{X}^R + \mathrm{j}\ d\mathbf{Y}/d\mathbf{X}^I)$

We have the following relationships for both analytic and non-analytic functions $\mathbf{Y}(\mathbf{X})$:

- *Cauchy Riemann equations*: The following are equivalent:
  - $\mathbf{Y}(\mathbf{X})$ is an analytic function of $\mathbf{X}$
  - $d\mathbf{Y}\mathbf{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:}$
  - $d\mathbf{Y}/d\mathbf{X}^C = \mathbf{0}$ for all $\mathbf{X}$
  - $d\mathbf{Y}/d\mathbf{X}^R + \mathrm{j}\ d\mathbf{Y}/d\mathbf{X}^I = \mathbf{0}$ for all $\mathbf{X}$
- $d\mathbf{Y}\mathbf{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:} + d\mathbf{Y}/d\mathbf{X}^C\ d\mathbf{X}^C\mathbf{:}$
- $d\mathbf{Y}/d\mathbf{X}^R = d\mathbf{Y}/d\mathbf{X} + d\mathbf{Y}/d\mathbf{X}^C$
- $d\mathbf{Y}/d\mathbf{X}^I = \mathrm{j}\ (d\mathbf{Y}/d\mathbf{X} - d\mathbf{Y}/d\mathbf{X}^C)$
- *Chain rule*: If $\mathbf{Z}$ is a function of $\mathbf{Y}$ which is itself a function of $\mathbf{X}$, then $d\mathbf{Z}/d\mathbf{X} = d\mathbf{Z}/d\mathbf{Y}\ d\mathbf{Y}/d\mathbf{X}$. This is the same as for real derivatives.
- *Real-valued*: If $\mathbf{Y}(\mathbf{X})$ is real for all complex $\mathbf{X}$, then
  - $d\mathbf{Y}/d\mathbf{X}^C = (d\mathbf{Y}/d\mathbf{X})^C$
  - $d\mathbf{Y}\mathbf{:} = 2(d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:})^R$
  - If $\mathbf{W}(\mathbf{X})$ is analytic with $\mathbf{W}(\mathbf{X})=\mathbf{Y}(\mathbf{X})$ for all real $\mathbf{X}$, then $d\mathbf{W}/d\mathbf{X} = 2\ (d\mathbf{Y}/d\mathbf{X})^R$ for all real $\mathbf{X}$
    - Example: If $\mathbf{C}=\mathbf{C}^H$, $y(\mathbf{x})=\mathbf{x}^H\mathbf{C}\mathbf{x}$ and $w(\mathbf{x})=\mathbf{x}^T\mathbf{C}\mathbf{x}$, then $\mathrm{d}y/\mathrm{d}\mathbf{x} = \mathbf{x}^H\mathbf{C}$ and $\mathrm{d}w/\mathrm{d}\mathbf{x} = 2\mathbf{x}^T\mathbf{C}^R$

Complex Gradient Vector

If $f(\mathbf{x})$ is a real function of a complex vector then $df/d\mathbf{x}^C = (df/d\mathbf{x})^C$ and we can define $\mathbf{grad}(f(\mathbf{x})) = 2(df/d\mathbf{x})^H = (df/d\mathbf{x}^R + j\ df/d\mathbf{x}^I)^T$ as the *Complex Gradient Vector* [R.9] with the following properties:

- $\mathbf{grad}(f(\mathbf{x}))$ is zero at an extreme value of f.
- $\mathbf{grad}(f(\mathbf{x}))$ points in the direction of steepest slope of $f(\mathbf{x})$
- The magnitude of the steepest slope is equal to $|\mathbf{grad}(f(\mathbf{x}))|$. Specifically, if $\mathbf{g}(\mathbf{x}) = \mathbf{grad}(f(\mathbf{x}))$, then $\lim_{a\to 0} a^{-1}(\ f(\mathbf{x}+a\mathbf{g}(\mathbf{x})) - f(\mathbf{x})\ ) = |\ \mathbf{g}(\mathbf{x})\ |^2$
- $\mathbf{grad}(f(\mathbf{x}))$ is normal to the surface $f(\mathbf{x})$ = constant which means that it can be used for gradient ascent/descent algorithms.

Basic Properties

- We may write the following differentials unambiguously without parentheses:
  - *Transpose:* $d\mathbf{Y}^T=d(\mathbf{Y}^T)=(d\mathbf{Y})^T$
  - *Hermitian Transpose:* $d\mathbf{Y}^H=d(\mathbf{Y}^H)=(d\mathbf{Y})^H$
  - *Conjugate:* $d\mathbf{Y}^C=d(\mathbf{Y}^C)=(d\mathbf{Y})^C$
- *Linearity:* $d(\mathbf{Y}+\mathbf{Z})=d\mathbf{Y}+d\mathbf{Z}$
- *Chain Rule:* If $\mathbf{Z}$ is a function of $\mathbf{Y}$ which is itself a function of $\mathbf{X}$, then for both the normal and the generalized complex derivative: $d\mathbf{Z}\mathbf{:} = d\mathbf{Z}/d\mathbf{Y}\ d\mathbf{Y}\mathbf{:} = d\mathbf{Z}/d\mathbf{Y}\ d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\mathbf{:}$
- *Product Rule:* $d(\mathbf{YZ}) = \mathbf{Y}\ d\mathbf{Z} + d\mathbf{Y}\ \ \mathbf{Z}$
  - $d(\mathbf{YZ})\mathbf{:} = (\mathbf{I} \otimes \mathbf{Y})\ d\mathbf{Z}\mathbf{:} + (\mathbf{Z}^T \otimes \mathbf{I})\ d\mathbf{Y}\mathbf{:}\ = ((\mathbf{I} \otimes \mathbf{Y})\ d\mathbf{Z}/d\mathbf{X} + (\mathbf{Z}^T \otimes \mathbf{I})\ d\mathbf{Y}/d\mathbf{X}\ )\ d\mathbf{X}\mathbf{:}$
- Hadamard Product: $d(\mathbf{Y} \bullet \mathbf{Z}) = \mathbf{Y} \bullet d\mathbf{Z} + d\mathbf{Y} \bullet \mathbf{Z}$
- Kroneker Product: $d(\mathbf{Y} \otimes \mathbf{Z}) = \mathbf{Y} \otimes d\mathbf{Z} + d\mathbf{Y} \otimes \mathbf{Z}$

Differentials of Linear Functions

- $d(\mathbf{Ax}) = d(\mathbf{x}^T\mathbf{A})\mathbf{:} = \mathbf{A}\ d\mathbf{x}$
  - $d(\mathbf{x}^T\mathbf{a}) = d(\mathbf{a}^T\mathbf{x}) = \mathbf{a}^T\ d\mathbf{x}$
- $d(\mathbf{A}^T\mathbf{XB})\mathbf{:} = (\mathbf{A}^T\ d\mathbf{X}\ \mathbf{B})\mathbf{:} = (\mathbf{B} \otimes \mathbf{A})^T\ d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{a}^T\mathbf{Xb}) = (\mathbf{b} \otimes \mathbf{a})^T\ d\mathbf{X}\mathbf{:}\ = (\mathbf{ab}^T)\mathbf{:}^T\ d\mathbf{X}\mathbf{:}$
    - $d(\mathbf{a}^T\mathbf{Xa}) = d(\mathbf{a}^T\mathbf{X}^T\mathbf{a}) = (\mathbf{a} \otimes \mathbf{a})^T\ d\mathbf{X}\mathbf{:}\ = (\mathbf{aa}^T)\mathbf{:}^T\ d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{XB})\mathbf{:} = (d\mathbf{X}\ \mathbf{B})\mathbf{:} = (\mathbf{B}^T \otimes \mathbf{I})\ d\mathbf{X}\mathbf{:}$
    - $d(\mathbf{xb}^T)\mathbf{:} = (d\mathbf{x}\ \mathbf{b}^T)\mathbf{:} = (\mathbf{b} \otimes \mathbf{I})\ d\mathbf{x}$
  - $d(\mathbf{a}^T\mathbf{X}^T\mathbf{b}) = (\mathbf{a} \otimes \mathbf{b})^T\ d\mathbf{X}\mathbf{:}\ = (\mathbf{ba}^T)\mathbf{:}^T\ d\mathbf{X}\mathbf{:}$
- [$\mathbf{x}$: Complex]
  - $d\ (\mathbf{x}^H\mathbf{A})\mathbf{:} = \mathbf{A}^T\ d\mathbf{x}^C$
- Writing $\mathbf{I}_n\ = \mathbf{I}_{[n\#n]}$ and $\mathbf{T}_{q,m} =$ **TVEC**(q,m),
  - $d(\mathbf{X}_{[m\#n]} \otimes \mathbf{A}_{[p\#q]})\mathbf{:} = (\mathbf{I}_n \otimes \mathbf{T}_{q,m} \otimes \mathbf{I}_p)(\mathbf{I}_{mn} \otimes \mathbf{A}\mathbf{:})\ d\mathbf{X}\mathbf{:} = (\mathbf{I}_{nq} \otimes \mathbf{T}_{m,p}\ )(\mathbf{I}_n \otimes \mathbf{A}\mathbf{:} \otimes \mathbf{I}_m)\ d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{A}_{[p\#q]} \otimes \mathbf{X}_{[m\#n]})\mathbf{:} = (\mathbf{I}_q \otimes \mathbf{T}_{n,p} \otimes \mathbf{I}_m)(\mathbf{A}\mathbf{:} \otimes \mathbf{I}_{mn})\ d\mathbf{X}\mathbf{:} = (\mathbf{T}_{m,n} \otimes \mathbf{I}_{pq}\ )(\mathbf{I}_n \otimes \mathbf{A}\mathbf{:} \otimes \mathbf{I}_m)\ d\mathbf{X}\mathbf{:}$

Differentials of Quadratic Products

- $d(\mathbf{Ax}+\mathbf{b})^T\mathbf{C}(\mathbf{Dx}+\mathbf{e}) = ((\mathbf{Ax}+\mathbf{b})^T\mathbf{CD} + (\mathbf{Dx}+\mathbf{e})^T\mathbf{C}^T\mathbf{A})\ d\mathbf{x}$
  - $d(\mathbf{x}^T\mathbf{Cx}) = \mathbf{x}^T(\mathbf{C}+\mathbf{C}^T)d\mathbf{x}\ =$ [$\mathbf{C}=\mathbf{C}^T$] $2\mathbf{x}^T\mathbf{C}d\mathbf{x}$
    - $d(\mathbf{x}^T\mathbf{x}) = 2\mathbf{x}^Td\mathbf{x}$
  - $d(\mathbf{Ax}+\mathbf{b})^T\ (\mathbf{Dx}+\mathbf{e}) =\ (\ (\mathbf{Ax}+\mathbf{b})^T\mathbf{D} + (\mathbf{Dx}+\mathbf{e})^T\mathbf{A})d\mathbf{x}$
    - $d(\mathbf{Ax}+\mathbf{b})^T\ (\mathbf{Ax}+\mathbf{b}) = 2(\mathbf{Ax}+\mathbf{b})^T\mathbf{A}d\mathbf{x}$
  - $d(\mathbf{Ax}+\mathbf{b})^T\mathbf{C}(\mathbf{Ax}+\mathbf{b}) =$ [$\mathbf{C}=\mathbf{C}^T$] $2(\mathbf{Ax}+\mathbf{b})^T\mathbf{CA}\ d\mathbf{x}$
- $d(\mathbf{Ax}+\mathbf{b})^H\mathbf{C}(\mathbf{Dx}+\mathbf{e}) = (\mathbf{Ax}+\mathbf{b})^H\mathbf{CD}\ d\mathbf{x} + (\mathbf{Dx}+\mathbf{e})^T\mathbf{C}^T\mathbf{A}^C\ d\mathbf{x}^C$
  - $d\ (\mathbf{x}^H\mathbf{Cx}) = \mathbf{x}^H\mathbf{C}\ d\mathbf{x} + \mathbf{x}^T\mathbf{C}^T\ d\mathbf{x}^C =$ [$\mathbf{C}=\mathbf{C}^H$] $2(\mathbf{x}^H\mathbf{C}\ d\mathbf{x})^R$
  - $d\ (\mathbf{x}^H\mathbf{x}) = 2(\mathbf{x}^H\ d\mathbf{x})^R$
- $d(\mathbf{a}^T\mathbf{X}^T\mathbf{Xb}) = \mathbf{X}(\mathbf{ab}^T + \mathbf{ba}^T)\mathbf{:}^T\ d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{a}^T\mathbf{X}^T\mathbf{Xa}) = 2(\mathbf{Xaa}^T\ )\mathbf{:}^T\ d\mathbf{X}\mathbf{:}$
- $d(\mathbf{a}^T\mathbf{X}^T\mathbf{CXb}) = (\mathbf{C}^T\mathbf{Xab}^T + \mathbf{CXba}^T)\mathbf{:}^T\ d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{a}^T\mathbf{X}^T\mathbf{CXa}) = ((\mathbf{C} + \mathbf{C}^T)\mathbf{Xaa}^T\ )\mathbf{:}^T\ d\mathbf{X}\mathbf{:} =$ [$\mathbf{C}=\mathbf{C}^T$] $2(\mathbf{CXaa}^T)\mathbf{:}^T\ d\mathbf{X}\mathbf{:}$

- $d((\mathbf{Xa+b})^T\mathbf{C}(\mathbf{Xa+b})) = ((\mathbf{C+C}^T)(\mathbf{Xa+b})\mathbf{a}^T):^T d\mathbf{X}:$
- $d(\mathbf{X}^2): = (\mathbf{X}d\mathbf{X} + d\mathbf{X}\,\mathbf{X}): = (\mathbf{I} ¤ \mathbf{X} + \mathbf{X}^T ¤ \mathbf{I})\, d\mathbf{X}:$
- $d(\mathbf{X}^T\mathbf{CX}): = (\mathbf{X}^T\mathbf{C}d\mathbf{X}): + (d(\mathbf{X}^T)\,\mathbf{CX}): = (\mathbf{I} ¤ \mathbf{X}^T\mathbf{C})\, d\mathbf{X}: + (\mathbf{X}^T\mathbf{C}^T ¤ \mathbf{I})\, d\mathbf{X}^T:$
- $d(\mathbf{X}^H\mathbf{CX}): = (\mathbf{X}^H\mathbf{C}d\mathbf{X}): + (d(\mathbf{X}^H)\,\mathbf{CX}): = (\mathbf{I} ¤ \mathbf{X}^H\mathbf{C})\, d\mathbf{X}: + (\mathbf{X}^T\mathbf{C}^T ¤ \mathbf{I})\, d\mathbf{X}^H:$

Differentials of Cubic Products

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

Differentials of Inverses

- $d(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}d\mathbf{X}\,\mathbf{X}^{-1}$ [2.1]
  - $d(\mathbf{X}^{-1}): = -(\mathbf{X}^{-T} ¤ \mathbf{X}^{-1})\, d\mathbf{X}:$
- $d(\mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}) = -(\mathbf{X}^{-T}\mathbf{ab}^T\mathbf{X}^{-T}):^T d\mathbf{X}: = -(\mathbf{ab}^T):^T (\mathbf{X}^{-T} ¤ \mathbf{X}^{-1})\, d\mathbf{X}:$ [2.6]
- $d(\mathrm{tr}(\mathbf{A}^T\mathbf{X}^{-1}\mathbf{B})) = d(\mathrm{tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{A})) = -(\mathbf{X}^{-T}\mathbf{AB}^T\mathbf{X}^{-T}):^T d\mathbf{X}: = -(\mathbf{AB}^T):^T (\mathbf{X}^{-T} ¤ \mathbf{X}^{-1})\, d\mathbf{X}:$

Differentials of Trace

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

- $d(\mathrm{tr}(\mathbf{Y}))=\mathrm{tr}(d\mathbf{Y})$
- $d(\mathrm{tr}(\mathbf{X})) = d(\mathrm{tr}(\mathbf{X}^T)) = \mathbf{I}:^T d\mathbf{X}:$ [2.4]
- $d(\mathrm{tr}(\mathbf{X}^k)) = k(\mathbf{X}^{k-1})^T:^T d\mathbf{X}:$
- $d(\mathrm{tr}(\mathbf{AX}^k)) = (\mathbf{SUM}_{r=0:k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T\ ):^T d\mathbf{X}:$
- $d(\mathrm{tr}(\mathbf{AX}^{-1}\mathbf{B})) = -(\mathbf{X}^{-1}\mathbf{BAX}^{-1})^T:^T d\mathbf{X}:= -(\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T}):^T d\mathbf{X}:$ [2.5]
  - $d(\mathrm{tr}(\mathbf{AX}^{-1})) = d(\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A})) = -(\mathbf{X}^{-T}\mathbf{A}^T\mathbf{X}^{-T}\ ):^T d\mathbf{X}:$
- $d(\mathrm{tr}(\mathbf{A}^T\mathbf{XB}^T)) = d(\mathrm{tr}(\mathbf{BX}^T\mathbf{A})) = (\mathbf{AB}):^T d\mathbf{X}:$ [2.4]
  - $d(\mathrm{tr}(\mathbf{XA}^T)) = d(\mathrm{tr}(\mathbf{A}^T\mathbf{X})) = d(\mathrm{tr}(\mathbf{X}^T\mathbf{A})) = d(\mathrm{tr}(\mathbf{AX}^T)) = \mathbf{A}:^T d\mathbf{X}:$
  - $d(\mathrm{tr}(\mathbf{A}^T\mathbf{X}^{-1}\mathbf{B}^T)) = d(\mathrm{tr}(\mathbf{BX}^T\mathbf{A})) = -(\mathbf{X}^{-T}\mathbf{ABX}^{-T}):^T d\mathbf{X}: = -(\mathbf{AB}):^T (\mathbf{X}^{-T} ¤ \mathbf{X}^{-1})\, d\mathbf{X}:$
- $d(\mathrm{tr}(\mathbf{AXBX}^T\mathbf{C})) = (\mathbf{A}^T\mathbf{C}^T\mathbf{XB}^T + \mathbf{CAXB}):^T d\mathbf{X}:$
  - $d(\mathrm{tr}(\mathbf{XAX}^T)) = d(\mathrm{tr}(\mathbf{AX}^T\mathbf{X})) = d(\mathrm{tr}(\mathbf{X}^T\mathbf{XA})) = (\,\mathbf{X}(\mathbf{A+A}^T)):^T d\mathbf{X}:$
  - $d(\mathrm{tr}(\mathbf{X}^T\mathbf{AX})) = d(\mathrm{tr}(\mathbf{AXX}^T)) = d(\mathrm{tr}(\mathbf{XX}^T\mathbf{A})) = ((\mathbf{A+A}^T)\mathbf{X}):^T d\mathbf{X}:$
- $d(\mathrm{tr}(\mathbf{AXBX})) = (\mathbf{A}^T\mathbf{X}^T\mathbf{B}^T + \mathbf{B}^T\mathbf{X}^T\mathbf{A}^T\ ):^T d\mathbf{X}:$
- $d(\mathrm{tr}((\mathbf{AXb+c})(\mathbf{AXb+c})^T) = 2(\mathbf{A}^T(\mathbf{AXb+c})\mathbf{b}^T):^T d\mathbf{X}:$
- $d(\mathbf{tr}((\mathbf{X}^T\mathbf{CX})^{-1}\mathbf{A}) =$ [C:symmetric] $d(\mathbf{tr}(\mathbf{A}\,(\mathbf{X}^T\mathbf{CX})^{-1}) = -((\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})(\mathbf{A+A}^T)(\mathbf{X}^T\mathbf{CX})^{-1}):^T d\mathbf{X}:$
- $d(\mathbf{tr}((\mathbf{X}^T\mathbf{CX})^{-1}(\mathbf{X}^T\mathbf{BX})) =$ [B,C:symmetric] $d(\mathbf{tr}(\,(\mathbf{X}^T\mathbf{BX})(\mathbf{X}^T\mathbf{CX})^{-1}) = 2(\mathbf{BX}(\mathbf{X}^T\mathbf{CX})^{-1}- (\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{X}^T\mathbf{BX}(\mathbf{X}^T\mathbf{CX})^{-1}\ ):^T d\mathbf{X}:$

Differentials of Determinant

Note: matrix dimensions must result in an $n\#n$ argument for det(). Some of the expressions below involve inverses: these forms apply only if the quantity being inverted is square and non-singular; alternative forms involving the adjoint, ADJ(), do not have the non-singular requirement.

- $d(\det(\mathbf{X})) = d(\det(\mathbf{X}^T)) = \mathbf{ADJ}(\mathbf{X}^T)\mathbf{:}^T\, d\mathbf{X:} = \det(\mathbf{X})\, (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.7]
- $d(\det(\mathbf{A}^T\mathbf{XB})) = d(\det(\mathbf{B}^T\mathbf{X}^T\mathbf{A})) = (\mathbf{A}\ \mathbf{ADJ}(\mathbf{A}^T\mathbf{XB})^T\mathbf{B}^T)\mathbf{:}^T\, d\mathbf{X:} =$ [A,B: nonsingular] $\det(\mathbf{A}^T\mathbf{XB}) \times (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.8]
- $d(\ln(\det(\mathbf{A}^T\mathbf{XB}))) =$ [A,B: nonsingular] $(\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.9]
  - $d(\ln(\det(\mathbf{X}))) = (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\det(\mathbf{X}^k)) = d(\det(\mathbf{X})^k) = k \times \det(\mathbf{X}^k) \times (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.10]
- $d(\ln(\det(\mathbf{X}^k))) = k \times (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\det(\mathbf{X}^T\mathbf{CX})) =$ [C=C^T] $2\det(\mathbf{X}^T\mathbf{CX})\times(\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X:}$ [2.11]
  - = [C=C^T, CX: nonsingular] $2\det(\mathbf{X}^T\mathbf{CX})\times(\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\ln(\det(\mathbf{X}^T\mathbf{CX}))) =$ [C=C^T] $2(\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X:}$
  - = [C=C^T, CX: nonsingular] $2(\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\det(\mathbf{X}^H\mathbf{CX})) = \det(\mathbf{X}^H\mathbf{CX}) \times (\mathbf{C}^T\mathbf{X}^C\, (\mathbf{X}^T\mathbf{C}^T\mathbf{X}^C)^{-1})d\mathbf{X:} + (\mathbf{CX}(\mathbf{X}^H\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X}^C\mathbf{:})$ [2.12]
- $d(\ln(\det(\mathbf{X}^H\mathbf{CX}))) = (\mathbf{C}^T\mathbf{X}^C\, (\mathbf{X}^T\mathbf{C}^T\mathbf{X}^C)^{-1})\mathbf{:}^T d\mathbf{X:} + (\mathbf{CX}(\mathbf{X}^H\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X}^C\mathbf{:}$ [2.13]

Jacobian

$d\mathbf{Y}/d\mathbf{X}$ is called the *Jacobian Matrix* of $\mathbf{Y:}$ with respect to $\mathbf{X:}$ and $J_\mathbf{X}(\mathbf{Y})=\det(d\mathbf{Y}/d\mathbf{X})$ is the corresponding *Jacobian*. The Jacobian occurs when changing variables in an integration: Integral($f(\mathbf{Y})d\mathbf{Y:}$)=Integral($f(\mathbf{Y}(\mathbf{X}))\det(d\mathbf{Y}/d\mathbf{X})\, d\mathbf{X:}$).

- $J_\mathbf{X}(\mathbf{X}_{[n\#n]}{}^{-1}) = (-1)^n\det(\mathbf{X})^{-2n}$

Hessian matrix

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

- [Real] $\mathbf{H_x}\, f = d/d\mathbf{x}\, (df/d\mathbf{x})^T$
  - $\mathbf{H_x}\, f$ is symmetric
  - $\mathbf{H_x}\, (\mathbf{a}^T\mathbf{x}) = 0$
  - $\mathbf{H_x}\, (\mathbf{Ax+b})^T\mathbf{C}(\mathbf{Dx+e}) = \mathbf{A}^T\mathbf{CD} + \mathbf{D}^T\mathbf{C}^T\mathbf{A}$
    - $\mathbf{H_x}\, (\mathbf{Ax+b})^T\, (\mathbf{Dx+e}) = \mathbf{A}^T\mathbf{D} + \mathbf{D}^T\mathbf{A}$
    - $\mathbf{H_x}\, (\mathbf{Ax+b})^T\mathbf{C}(\mathbf{Ax+b}) = \mathbf{A}^T(\mathbf{C} + \mathbf{C}^T)\mathbf{A} =$ [C=C^T] $2\mathbf{A}^T\mathbf{CA}$
      - $\mathbf{H_x}\, (\mathbf{Ax+b})^T\, (\mathbf{Ax+b}) = 2\mathbf{A}^T\mathbf{A}$
      - $\mathbf{H_x}\, (\mathbf{x}^T\mathbf{Cx}) = \mathbf{C}+\mathbf{C}^T =$ [C=C^T] $2\mathbf{C}$
      - $\mathbf{H_x}\, (\mathbf{x}^T\mathbf{x}) = 2\mathbf{I}$
- [x: Complex] $\mathbf{H_x}\, f = (d/d\mathbf{x}\, (df/d\mathbf{x})^H)^T = d/d\mathbf{x}^C\, (df/d\mathbf{x})^T$
  - $\mathbf{H_x}\, f$ is hermitian

- $\mathbf{H_x}\ (\mathbf{Ax+b})^H\mathbf{C}(\mathbf{Ax+b}) =$ [$\mathbf{C}=\mathbf{C}^H$] $(\mathbf{A}^H\mathbf{CA})^T$ [2.14]
  - $\mathbf{H_x}\ (\mathbf{x}^H\mathbf{Cx}) =$ [$\mathbf{C}=\mathbf{C}^H$] $\mathbf{C}^T$
