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 $\mathbf{X} = \mathbf{X}^R + j\mathbf{X}^I$
  - $(\mathbf{XY})^R = \mathbf{X}^R\mathbf{Y}^R - \mathbf{X}^I\mathbf{Y}^I$
  - $(\mathbf{XY})^I = \mathbf{X}^R\mathbf{Y}^I + \mathbf{X}^I\mathbf{Y}^R$
- $\mathbf{X}^C = \mathbf{X}^R - j\mathbf{X}^I$ is the complex conjugate of $\mathbf{X}$
- $\mathbf{X}^H=(\mathbf{X}^R)^T=(\mathbf{X}^T)^C$ is the Hermitian transpose of $\mathbf{X}$
- **X:** denotes the long column vector formed by concatenating the columns of **X** (see vectorization).
- $\mathbf{A} \otimes \mathbf{B}$ = **KRON**(**A**,**B**), the kroneker product
- $\mathbf{A} \bullet \mathbf{B}$ the Hadamard or elementwise product
- matrices and vectors **A**, **B**, **C** do not depend on **X**
- $\mathbf{I}_n = \mathbf{I}_{[n\#n]}$ the $n\#n$ identity matrix
- $\mathbf{T}_{m,n}$ = **TVEC**(m,n) is the vectorized transpose matrix, i.e. $\mathbf{X}^T\text{:}=\mathbf{T}_{m,n}\mathbf{X}\text{:}$ for $\mathbf{X}_{[m,n]}$
- $\partial\mathbf{Y}/\partial\mathbf{X}$ and $\partial\mathbf{Y}/\partial\mathbf{X}^C$ are partial derivatives with $\mathbf{X}^C$ and $\mathbf{X}$ respectively held constant (note that $\mathbf{X}^H=(\mathbf{X}^C)^T$)
- $\partial\mathbf{Y}/\partial\mathbf{X}^R$ and $\partial\mathbf{Y}/\partial\mathbf{X}^I$ are partial derivatives with $\mathbf{X}^I$ and $\mathbf{X}^R$ respectively held constant

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 $d\mathbf{Y}\text{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\text{:}$ 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 $d\mathbf{Y}\text{:} = d\mathbf{Y}/d\mathbf{X}\ d\mathbf{X}\text{:}$ 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 generally true iff $\mathbf{Y}(\mathbf{X})$ is an analytic function. This normally implies that $\mathbf{Y}(\mathbf{X})$ does not depend explicitly on $\mathbf{X}^C$ or $\mathbf{X}^H$.

Even for non-analytic functions we can treat $\mathbf{X}$ and $\mathbf{X}^C$ (with $\mathbf{X}^H=(\mathbf{X}^C)^T$) as distinct variables and write uniquely $d\mathbf{Y}\mathbf{:} = \partial\mathbf{Y}/\partial\mathbf{X}\ d\mathbf{X}\mathbf{:} + \partial\mathbf{Y}/\partial\mathbf{X}^C\ d\mathbf{X}^C\mathbf{:}$ provided that $\mathbf{Y}$ 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). $\partial\mathbf{Y}/\partial\mathbf{X}$ is the *Generalized Complex Derivative* and $\partial\mathbf{Y}/\partial\mathbf{X}^C$ is the *Complex Conjugate Derivative* [R.4, R.9]; their properties are studied in *Wirtinger Calculus*.

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

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

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

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

Complex Constrained Minimization

Suppose $f(\mathbf{X})$ is a scalar real function of a complex matrix (or vector), $\mathbf{X}$, and $\mathbf{G}(\mathbf{X})$ is a complex-valued matrix (or vector or scalar) function of $\mathbf{X}$. To minimize $f(\mathbf{X})$ subject to $\mathbf{G}(\mathbf{X})=\mathbf{0}$, we use complex Lagrange multipliers and minimize $f(\mathbf{X})+\mathrm{tr}(\mathbf{K}^H\mathbf{G}(\mathbf{X}))+\mathrm{tr}(\mathbf{K}^T\mathbf{G}(\mathbf{X})^C)$ subject to $\mathbf{G}(\mathbf{X})=\mathbf{0}$. Hence we solve $\partial f/\partial\mathbf{X}+\partial\mathrm{tr}(\mathbf{K}^H\mathbf{G})/\partial\mathbf{X}+\partial\mathrm{tr}(\mathbf{K}^T\mathbf{G}^C)/\partial\mathbf{X} = \mathbf{0}^T$ subject to $\mathbf{G}(\mathbf{X})=\mathbf{0}$. If $\mathbf{g}(\mathbf{X})$ is a vector, this becomes $\partial f/\partial\mathbf{X}+\mathbf{k}^H\partial\mathbf{g}/\partial\mathbf{X}+\mathbf{k}^T\partial\mathbf{g}^C/\partial\mathbf{X} = \mathbf{0}^T$. If $g(\mathbf{X})$ is a scalar, this becomes $\partial f/\partial\mathbf{X}+k^C\partial g/\partial\mathbf{x}+k\partial g^C/\partial\mathbf{x} = \mathbf{0}^T$.

- Example: If $f(\mathbf{x})=\mathbf{x}^H\mathbf{S}\mathbf{x}$ where $\mathbf{S}=\mathbf{S}^H$ and $g(\mathbf{x})=\mathbf{a}^H\mathbf{x}-1$, then $\partial f/\partial\mathbf{x}+k^H\partial g/\partial\mathbf{x}+k^T\partial g^C/\partial\mathbf{x}=\mathbf{x}^H\mathbf{S}+k\mathbf{a}^H+\mathbf{0}^T=\mathbf{0}^T$ which implies $\mathbf{S}\mathbf{x}+k^C\mathbf{a}=\mathbf{0}$ from which $\mathbf{x}=-k^C\mathbf{S}^{-1}\mathbf{a}$. Substituting this into the constraint, $g(\mathbf{x})=\mathbf{a}^H\mathbf{x}-1=0$, gives $-k^C\mathbf{a}^H\mathbf{S}^{-1}\mathbf{a} = 1$ from which $k=-(\mathbf{a}^H\mathbf{S}^{-1}\mathbf{a})^{-1}$. Substituting this back into the expression for $\mathbf{x}$ gives $\mathbf{x} = \mathbf{S}^{-1}\mathbf{a}(\mathbf{a}^H\mathbf{S}^{-1}\mathbf{a})^{-1}$.

Complex Gradient Vector

If $f(\mathbf{X})$ is a real function of a complex matrix (or vector), $\mathbf{X}$, then $\partial f/\partial\mathbf{X}^C= (\partial f/\partial\mathbf{X})^C$ and we can define the complex-valued column vector $\mathbf{grad}(f(\mathbf{X})) = 2\,(\partial f/\partial\mathbf{X})^H = (\partial f/\partial\mathbf{X}^R+j\,\partial f/\partial\mathbf{X}^I)^T$ as the *Complex Gradient Vector* [R.9] with the properties listed below. If we use <-> to represent the vector mapping associated with the Complex-to-Real isomporphism, and $\mathbf{X}_{[m\#n]}$**:** <-> $\mathbf{y}_{[2mn]}$ where $\mathbf{y}$ is real, then $\mathbf{grad}(f(\mathbf{X}))$ <-> $\mathbf{grad}(f(\mathbf{y}))$ where the latter is the conventional **grad** function from vector calculus.

- $\mathbf{grad}(f(\mathbf{X}))$ 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.
- If $f(\mathbf{X})=\mathbf{y}^H\mathbf{y}$, then $\mathbf{grad}(f(\mathbf{X}))=2(\partial\mathbf{y}/\partial\mathbf{X})^H\mathbf{y}+2(\partial\mathbf{y}/\partial\mathbf{X}^C)^T\mathbf{y}^C$

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:* $d(\mathbf{Y}\mathbf{Z}) =\mathbf{Y}\, d\mathbf{Z} + d\mathbf{Y}\;\; \mathbf{Z}$
  - $d(\mathbf{Y}\mathbf{Z})\mathbf{:} = (\mathbf{I}\otimes\mathbf{Y})\, d\mathbf{Z}\mathbf{:} + (\mathbf{Z}^T\otimes\mathbf{I})\, d\mathbf{Y}\mathbf{:}\; = ((\mathbf{I}\otimes\mathbf{Y})\, d\mathbf{Z}/d\mathbf{X} + (\mathbf{Z}^T\otimes\mathbf{I})\, d\mathbf{Y}/d\mathbf{X}\,)\, d\mathbf{X}\mathbf{:}$
- 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

- $d(\mathbf{A}\mathbf{x}) = d(\mathbf{x}^T\mathbf{A}^T)\mathbf{:} =\mathbf{A}\, d\mathbf{x}$
  - $d(\mathbf{x}^T\mathbf{a}) = d(\mathbf{a}^T\mathbf{x}) = \mathbf{a}^T\, d\mathbf{x}$
  - $d(\mathbf{b}\mathbf{x}^T\mathbf{a}) = \mathbf{b}\mathbf{a}^T\, d\mathbf{x}$
- $d(\mathbf{A}\mathbf{X}\mathbf{B})\mathbf{:} = (\mathbf{A}\, d\mathbf{X}\, \mathbf{B})\mathbf{:} = (\mathbf{B}^T\otimes\mathbf{A})\, d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{a}^T\mathbf{X}\mathbf{b}) = (\mathbf{b}\otimes\mathbf{a})^T\, d\mathbf{X}\mathbf{:}\; = (\mathbf{a}\mathbf{b}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
    - $d(\mathbf{a}^T\mathbf{X}\mathbf{a}) = d(\mathbf{a}^T\mathbf{X}^T\mathbf{a}) = (\mathbf{a}\otimes\mathbf{a})^T\, d\mathbf{X}\mathbf{:}\; = (\mathbf{a}\mathbf{a}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
  - $[\mathbf{X}_{[m\#n]}]$ $d(\mathbf{A}\mathbf{X})\mathbf{:} = (\mathbf{I}_n\otimes\mathbf{A})\, d\mathbf{X}\mathbf{:}$
  - $[\mathbf{X}_{[m\#n]}]$ $d(\mathbf{X}\mathbf{B})\mathbf{:} = (d\mathbf{X}\,\mathbf{B})\mathbf{:} = (\mathbf{B}^T\otimes\mathbf{I}_m)\, d\mathbf{X}\mathbf{:}$
    - $[\mathrm{x}_{[n]}]$ $d(\mathbf{x}\mathbf{b}^T)\mathbf{:} = (d\mathbf{x}\,\mathbf{b}^T)\mathbf{:} = (\mathbf{b}\otimes\mathbf{I}_n)\, d\mathbf{x}$

- $d(\mathbf{AX}^T\mathbf{B})\mathbf{:} = (\mathbf{B}^T \otimes \mathbf{A})\, d\mathbf{X}^T\mathbf{:}$
  - $d(\mathbf{a}^T\mathbf{X}^T\mathbf{b}) = (\mathbf{a} \otimes \mathbf{b})^T\, d\mathbf{X}\mathbf{:} = (\mathbf{ab}^T)\mathbf{:}^T\, d\mathbf{X}^T\mathbf{:} = (\mathbf{ba}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
- $d(|\mathbf{x}|) = |\mathbf{x}|^{-1}\mathbf{x}^T\, d\mathbf{x}$
- [x: Complex] $d\,(\mathbf{x}^H\mathbf{A})\mathbf{:} = \mathbf{A}^T\, d\mathbf{x}^C$
- $d(\mathbf{X}_{[m\#n]} \otimes \mathbf{A}_{[p\#q]})\mathbf{:} = (\mathbf{I}_n \otimes \mathbf{T}_{q,m} \otimes \mathbf{I}_p)(\mathbf{I}_{mn} \otimes \mathbf{A:})\, d\mathbf{X}\mathbf{:} = (\mathbf{I}_{nq} \otimes \mathbf{T}_{m,p})(\mathbf{I}_n \otimes \mathbf{A:} \otimes \mathbf{I}_m)\, d\mathbf{X}\mathbf{:}$
- $d(\mathbf{A}_{[p\#q]} \otimes \mathbf{X}_{[m\#n]})\mathbf{:} = (\mathbf{I}_q \otimes \mathbf{T}_{n,p} \otimes \mathbf{I}_m)(\mathbf{A:} \otimes \mathbf{I}_{mn})\, d\mathbf{X}\mathbf{:} = (\mathbf{T}_{m,n} \otimes \mathbf{I}_{pq})(\mathbf{I}_n \otimes \mathbf{A:} \otimes \mathbf{I}_m)\, d\mathbf{X}\mathbf{:}$

Differentials of Quadratic Products

- $d(\mathbf{Ax}+\mathbf{b})^T\mathbf{C}(\mathbf{Dx}+\mathbf{e}) = ((\mathbf{Ax}+\mathbf{b})^T\mathbf{CD} + (\mathbf{Dx}+\mathbf{e})^T\mathbf{C}^T\mathbf{A})\, d\mathbf{x}$
  - $d(\mathbf{x}^T\mathbf{Cx}) = \mathbf{x}^T(\mathbf{C}+\mathbf{C}^T)d\mathbf{x}$ = [$\mathbf{C}=\mathbf{C}^T$] $2\mathbf{x}^T\mathbf{C}d\mathbf{x}$
    - $d(\mathbf{x}^T\mathbf{x}) = 2\mathbf{x}^Td\mathbf{x}$
  - $d(\mathbf{Ax}+\mathbf{b})^T(\mathbf{Dx}+\mathbf{e}) = ((\mathbf{Ax}+\mathbf{b})^T\mathbf{D} + (\mathbf{Dx}+\mathbf{e})^T\mathbf{A})d\mathbf{x}$
    - $d(\mathbf{Ax}+\mathbf{b})^T(\mathbf{Ax}+\mathbf{b}) = 2(\mathbf{Ax}+\mathbf{b})^T\mathbf{A}d\mathbf{x}$
  - $d(\mathbf{Ax}+\mathbf{b})^T\mathbf{C}(\mathbf{Ax}+\mathbf{b})$ = [$\mathbf{C}=\mathbf{C}^T$] $2(\mathbf{Ax}+\mathbf{b})^T\mathbf{CA}\, d\mathbf{x}$
- $d(\mathbf{Ax}+\mathbf{b})^H\mathbf{C}(\mathbf{Dx}+\mathbf{e}) = (\mathbf{Ax}+\mathbf{b})^H\mathbf{CD}\, d\mathbf{x} + (\mathbf{Dx}+\mathbf{e})^T\mathbf{C}^T\mathbf{A}^C\, d\mathbf{x}^C$
  - $d(\mathbf{Ax}+\mathbf{b})^H\mathbf{C}(\mathbf{Ax}+\mathbf{b}) = (\mathbf{Ax}+\mathbf{b})^H\mathbf{CA}\, d\mathbf{x} + (\mathbf{Ax}+\mathbf{b})^T\mathbf{C}^T\mathbf{A}^C\, d\mathbf{x}^C$ = [$\mathbf{C}=\mathbf{C}^H$] $2((\mathbf{Ax}+\mathbf{b})^H\mathbf{CA}\, d\mathbf{x})^R$
  - $d(\mathbf{Ax}+\mathbf{b})^H(\mathbf{Ax}+\mathbf{b}) = 2((\mathbf{Ax}+\mathbf{b})^H\mathbf{A}\, d\mathbf{x})^R$
  - $d\,(\mathbf{x}^H\mathbf{Cx}) = \mathbf{x}^H\mathbf{C}\, d\mathbf{x} + \mathbf{x}^T\mathbf{C}^T\, d\mathbf{x}^C$ = [$\mathbf{C}=\mathbf{C}^H$] $2(\mathbf{x}^H\mathbf{C}\, d\mathbf{x})^R$
  - $d\,(\mathbf{x}^H\mathbf{x}) = 2(\mathbf{x}^H\, d\mathbf{x})^R$
- $d(\mathbf{a}^T\mathbf{X}^T\mathbf{Xb}) = \mathbf{X}(\mathbf{ab}^T + \mathbf{ba}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{a}^T\mathbf{X}^T\mathbf{Xa}) = 2(\mathbf{Xaa}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
- $d(\mathbf{a}^T\mathbf{X}^T\mathbf{CXb}) = (\mathbf{C}^T\mathbf{Xab}^T + \mathbf{CXba}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
  - $d(\mathbf{a}^T\mathbf{X}^T\mathbf{CXa}) = ((\mathbf{C} + \mathbf{C}^T)\mathbf{Xaa}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$ = [$\mathbf{C}=\mathbf{C}^T$] $2(\mathbf{CXaa}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
- $d((\mathbf{Xa}+\mathbf{b})^T\mathbf{C}(\mathbf{Xa}+\mathbf{b})) = ((\mathbf{C}+\mathbf{C}^T)(\mathbf{Xa}+\mathbf{b})\mathbf{a}^T)\mathbf{:}^T\, d\mathbf{X}\mathbf{:}$
- [$\mathbf{X}_{[n\#n]}$] $d(\mathbf{X}^2)\mathbf{:} = (\mathbf{X}d\mathbf{X} + d\mathbf{X}\,\mathbf{X})\mathbf{:} = (\mathbf{I}_n \otimes \mathbf{X} + \mathbf{X}^T \otimes \mathbf{I}_n)\, d\mathbf{X}\mathbf{:}$
- [$\mathbf{X}_{[m\#n]}$] $d(\mathbf{X}^T\mathbf{CX})\mathbf{:} = (\mathbf{I}_n \otimes \mathbf{X}^T\mathbf{C})\, d\mathbf{X}\mathbf{:} + (\mathbf{X}^T\mathbf{C}^T \otimes \mathbf{I}_n)\, d\mathbf{X}^T\mathbf{:} = (\mathbf{I}_n \otimes \mathbf{X}^T\mathbf{C} + \mathbf{T}_{n,n}(\mathbf{I}_n \otimes \mathbf{X}^T\mathbf{C}^T))\, d\mathbf{X}\mathbf{:}$
  - [$\mathbf{X}_{[m\#n]}$, $\mathbf{C}_{[m\#m]}=\mathbf{C}^T$] $d(\mathbf{X}^T\mathbf{CX})\mathbf{:} = (\mathbf{I}_{n\times n}+\mathbf{T}_{n,n})(\mathbf{I}_n \otimes \mathbf{X}^T\mathbf{C})\, d\mathbf{X}\mathbf{:}$
  - [$\mathbf{X}_{[m\#n]}$] $d(\mathbf{X}^T\mathbf{X})\mathbf{:} = (\mathbf{I}_n \otimes \mathbf{X}^T)\, d\mathbf{X}\mathbf{:} + (\mathbf{X}^T \otimes \mathbf{I}_n)\, d\mathbf{X}^T\mathbf{:} = (\mathbf{I}_{n\times n} + \mathbf{T}_{n,n})(\mathbf{I}_n \otimes \mathbf{X}^T)\, d\mathbf{X}\mathbf{:}$
- [$\mathbf{X}_{[m\#n]}$] $d(\mathbf{X}^H\mathbf{CX})\mathbf{:} = (\mathbf{X}^H\mathbf{C}d\mathbf{X})\mathbf{:} + (d(\mathbf{X}^H)\,\mathbf{CX})\mathbf{:} = (\mathbf{I}_n \otimes \mathbf{X}^H\mathbf{C})\, d\mathbf{X}\mathbf{:} + (\mathbf{X}^T\mathbf{C}^T \otimes \mathbf{I}_n)\, d\mathbf{X}^H\mathbf{:}$
- $\mathbf{grad}((\mathbf{Ax}+\mathbf{b})^H(\mathbf{Ax}+\mathbf{b})) = 2\mathbf{A}^H(\mathbf{Ax}+\mathbf{b})$
  - $\mathbf{grad}(\mathbf{x}^H\mathbf{x}) = 2\mathbf{x}$

Differentials of Cubic Products

- $d(\mathbf{xx}^T\mathbf{Ax}) = (\mathbf{xx}^T(\mathbf{A}+\mathbf{A}^T)+\mathbf{x}^T\mathbf{Ax}\times\mathbf{I})d\mathbf{x}$
  - $d(\mathbf{xx}^T\mathbf{x}) = (2\mathbf{xx}^T+\mathbf{x}^T\mathbf{x}\times\mathbf{I})d\mathbf{x}$
- [$\mathbf{X}_{[m\#n]}$] $d(\mathbf{XAX}^T\mathbf{BX})\mathbf{:} = (\mathbf{X}^T\mathbf{B}^T\mathbf{XA}^T \otimes \mathbf{I}_m + \mathbf{I}_n \otimes \mathbf{XAX}^T\mathbf{B})\, d\mathbf{X}\mathbf{:} + (\mathbf{X}^T\mathbf{B} \otimes \mathbf{XA})\, d\mathbf{X}^T\mathbf{:} = (\mathbf{X}^T\mathbf{B}^T\mathbf{XA}^T \otimes \mathbf{I}_m + \mathbf{T}_{n,m}(\mathbf{XA} \otimes \mathbf{X}^T\mathbf{B}) + \mathbf{I}_n \otimes \mathbf{XAX}^T\mathbf{B})\, d\mathbf{X}\mathbf{:}$
  - [$\mathbf{X}_{[m\#n]}$] $d(\mathbf{XX}^T\mathbf{X})\mathbf{:} = (\mathbf{X}^T\mathbf{X} \otimes \mathbf{I}_m + \mathbf{I}_n \otimes \mathbf{XX}^T)\, d\mathbf{X}\mathbf{:} + (\mathbf{X}^T \otimes \mathbf{X})\, d\mathbf{X}^T\mathbf{:} = (\mathbf{X}^T\mathbf{X} \otimes \mathbf{I}_m + \mathbf{T}_{n,m}(\mathbf{X} \otimes \mathbf{X}^T) + \mathbf{I}_n \otimes \mathbf{XX}^T)\, d\mathbf{X}\mathbf{:}$
- [$\mathbf{X}_{[m\#n]}$] $d(\mathbf{XAXBX})\mathbf{:} = (\mathbf{X}^T\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T \otimes \mathbf{I}_m + \mathbf{X}^T\mathbf{B}^T \otimes \mathbf{XA} + \mathbf{I}_n \otimes \mathbf{XAXB})\, d\mathbf{X}\mathbf{:}$

- [$\mathbf{X}_{[n\#n]}$] $d(\mathbf{X}^3)\mathbf{:} = ((\mathbf{X}^T)^2 \otimes \mathbf{I}_n + \mathbf{X}^T \otimes \mathbf{X} + \mathbf{I}_n \otimes \mathbf{X}^2)\, d\mathbf{X:}$

Differentials of Inverses

- $d(\mathbf{X}^{-1}) = -\mathbf{X}^{-1} d\mathbf{X}\, \mathbf{X}^{-1}$ [2.1]
  - $d(\mathbf{X}^{-1})\mathbf{:} = -(\mathbf{X}^{-T} \otimes \mathbf{X}^{-1})\, d\mathbf{X:}$
- $d(\mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}) = -(\mathbf{X}^{-T}\mathbf{ab}^T\mathbf{X}^{-T})\mathbf{:}^T d\mathbf{X:} = -(\mathbf{ab}^T)\mathbf{:}^T (\mathbf{X}^{-T} \otimes \mathbf{X}^{-1})\, d\mathbf{X:}$ [2.9]
- $d(\mathrm{tr}(\mathbf{A}^T\mathbf{X}^{-1}\mathbf{B})) = d(\mathrm{tr}(\mathbf{B}^T\mathbf{X}^{T}\mathbf{A})) = -(\mathbf{X}^{-T}\mathbf{AB}^T\mathbf{X}^{-T})\mathbf{:}^T d\mathbf{X:} = -(\mathbf{AB}^T)\mathbf{:}^T (\mathbf{X}^{-T} \otimes \mathbf{X}^{-1})\, d\mathbf{X:}$

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]
- $d(\mathrm{tr}(\mathbf{X}^k)) = k(\mathbf{X}^{k-1})^T\mathbf{:}^T d\mathbf{X:}$
- $d(\mathrm{tr}(\mathbf{AX}^k)) = (\mathbf{SUM}_{r=0:k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T\ )\mathbf{:}^T d\mathbf{X:}$
- $d(\mathrm{tr}(\mathbf{AX}^{-1}\mathbf{B})) = -(\mathbf{X}^{-1}\mathbf{BAX}^{-1})^T\mathbf{:}^T d\mathbf{X:} = -(\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T})\mathbf{:}^T d\mathbf{X:}$ [2.5]
  - $d(\mathrm{tr}(\mathbf{AX}^{-1})) = d(\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A})) = -(\mathbf{X}^{-T}\mathbf{A}^T\mathbf{X}^{-T}\ )\mathbf{:}^T d\mathbf{X:}$
- $d(\mathrm{tr}(\mathbf{A}^T\mathbf{XB}^T)) = d(\mathrm{tr}(\mathbf{BX}^T\mathbf{A})) = (\mathbf{AB})\mathbf{:}^T d\mathbf{X:}$ [2.4]
  - $d(\mathrm{tr}(\mathbf{XA}^T)) = d(\mathbf{tr}(\mathbf{A}^T\mathbf{X})) = d(\mathrm{tr}(\mathbf{X}^T\mathbf{A})) = d(\mathbf{tr}(\mathbf{AX}^T)) = \mathbf{A:}^T d\mathbf{X:}$
  - $d(\mathrm{tr}(\mathbf{A}^T\mathbf{X}^{-1}\mathbf{B}^T)) = d(\mathrm{tr}(\mathbf{BX}^T\mathbf{A})) = -(\mathbf{X}^{-T}\mathbf{ABX}^{-T})\mathbf{:}^T d\mathbf{X:} = -(\mathbf{AB})\mathbf{:}^T (\mathbf{X}^{-T} \otimes \mathbf{X}^{-1})\, d\mathbf{X:}$
- $d(\mathrm{tr}(\mathbf{AXBX}^T\mathbf{C})) = (\mathbf{A}^T\mathbf{C}^T\mathbf{XB}^T + \mathbf{CAXB})\mathbf{:}^T d\mathbf{X:}$
  - $d(\mathrm{tr}(\mathbf{XAX}^T)) = d(\mathrm{tr}(\mathbf{AX}^T\mathbf{X})) = d(\mathrm{tr}(\mathbf{X}^T\mathbf{XA})) = (\mathbf{X}(\mathbf{A}+\mathbf{A}^T))\mathbf{:}^T d\mathbf{X:}$
  - $(\mathrm{tr}(\mathbf{X}^T\mathbf{AX})) = d(\mathrm{tr}(\mathbf{AXX}^T)) = d(\mathrm{tr}(\mathbf{XX}^T\mathbf{A})) = ((\mathbf{A}+\mathbf{A}^T)\mathbf{X})\mathbf{:}^T d\mathbf{X:}$
  - $\mathrm{d}(\mathrm{tr}(\mathbf{XX}^{\mathrm{T}})) = d(\mathrm{tr}(\mathbf{X}^T\mathbf{X})) = 2\mathbf{X}\mathbf{:}^{\mathrm{T}}\, \mathrm{d}\mathbf{X:}$
- $d(\mathrm{tr}(\mathbf{AXBX})) = (\mathbf{A}^T\mathbf{X}^T\mathbf{B}^T + \mathbf{B}^T\mathbf{X}^T\mathbf{A}^T\ )\mathbf{:}^T d\mathbf{X:}$
- $d(\mathrm{tr}((\mathbf{AXb}+\mathbf{c})(\mathbf{AXb}+\mathbf{c})^T) = 2(\mathbf{A}^T(\mathbf{AXb}+\mathbf{c})\mathbf{b}^T)\mathbf{:}^T d\mathbf{X:}$
- [$\mathbf{C}=\mathbf{C}^T$] $d(\mathbf{tr}((\mathbf{X}^T\mathbf{CX})^{-1}\mathbf{A}) = d(\mathbf{tr}(\mathbf{A}\ (\mathbf{X}^T\mathbf{CX})^{-1}) = -((\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})(\mathbf{A}+\mathbf{A}^T)(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{:}^T d\mathbf{X:}$
- [$\mathbf{B}=\mathbf{B}^T$, $\mathbf{C}=\mathbf{C}^T$] $d(\mathrm{tr}((\mathbf{X}^T\mathbf{CX})^{-1}(\mathbf{X}^T\mathbf{BX})) = d(\mathbf{tr}(\ (\mathbf{X}^T\mathbf{BX})(\mathbf{X}^T\mathbf{CX})^{-1}) = 2(\mathbf{BX}(\mathbf{X}^T\mathbf{CX})^{-1} - (\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{X}^T\mathbf{BX}(\mathbf{X}^T\mathbf{CX})^{-1}\ )\mathbf{:}^T d\mathbf{X:}$
- [$\mathbf{D}=\mathbf{D}^H$] $d(\mathrm{tr}((\mathbf{AXB}+\mathbf{C})\mathbf{D}(\mathbf{AXB}+\mathbf{C})^H)) = ((2\mathbf{A}^H(\mathbf{AXB}+\mathbf{C})\mathbf{DB}^H)\mathbf{:}^H d\mathbf{X:})^R$ [2.6]
  - $d(\mathrm{tr}((\mathbf{AXB}+\mathbf{C})(\mathbf{AXB}+\mathbf{C})^H)) = ((2\mathbf{A}^H(\mathbf{AXB}+\mathbf{C})\mathbf{B}^H)\mathbf{:}^H d\mathbf{X:})^R$
  - [$\mathbf{D}=\mathbf{D}^H$] $\mathrm{d}(\mathrm{tr}(\mathbf{XDX}^H)) = ((2\mathbf{XD})\mathbf{:}^H d\mathbf{X:})^R$
  - $\mathrm{d}(\mathrm{tr}(\mathbf{XX}^H)) = (2\mathbf{X:}^H d\mathbf{X:})^R$

Trace Minimization

In the following expressions $\mathbf{M}^\#$ denotes the inverse of $\mathbf{M}$ or, if $\mathbf{M}$ is singular, any generalized inverse (including the pseudoinverse).

- [$\mathbf{D}=\mathbf{D}^H$] $\mathrm{argmin}_{\mathbf{X}}\{\mathrm{tr}((\mathbf{AXB}+\mathbf{C})\mathbf{D}(\mathbf{AXB}+\mathbf{C})^H)\} = -(\mathbf{A}^H\mathbf{A})^\#\mathbf{A}^H\mathbf{CDB}^H(\mathbf{BDB}^H)^\#$ [2.7]
  - [$\mathbf{D}=\mathbf{D}^H$] $\mathrm{argmin}_{\mathbf{X}}\{\mathrm{tr}((\mathbf{AX}+\mathbf{C})\mathbf{D}(\mathbf{AX}+\mathbf{C})^H)\} = -(\mathbf{A}^H\mathbf{A})^\#\mathbf{A}^H\mathbf{C}$
- [$\mathbf{D}=\mathbf{D}^H$] $\mathrm{argmin}_{\mathbf{X}}\{\mathrm{tr}((\mathbf{AXB}+\mathbf{C})^H\mathbf{D}(\mathbf{AXB}+\mathbf{C}))\} = -(\mathbf{A}^H\mathbf{DA})^\#\mathbf{A}^H\mathbf{DCB}^H(\mathbf{BB}^H)^\#$

- $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AX}+\mathbf{C})^H\mathbf{D}(\mathbf{AX}+\mathbf{C}))\} = -(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{A}^H\mathbf{DC}$
- $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{x}}\{(\mathbf{Ax}+\mathbf{c})^H\mathbf{D}(\mathbf{Ax}+\mathbf{c})\} = -(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{A}^H\mathbf{Dc}$

- $[\mathbf{D}=\mathbf{D}^H, \mathbf{R}=\mathbf{R}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AXB}+\mathbf{C})\mathbf{D}(\mathbf{AXB}+\mathbf{C})^H+(\mathbf{AXP}+\mathbf{Q})\mathbf{R}(\mathbf{AXP}+\mathbf{Q})^H)\} = -(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{A}^H(\mathbf{CDB}^H+\mathbf{QRP}^H)(\mathbf{BDB}^H+\mathbf{PRP}^H)^{\#}$
  - $[\mathbf{D}=\mathbf{D}^H, \mathbf{R}=\mathbf{R}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AX}+\mathbf{C})\mathbf{D}(\mathbf{AX}+\mathbf{C})^H+(\mathbf{AX}+\mathbf{Q})\mathbf{R}(\mathbf{AX}+\mathbf{Q})^H)\} = -(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{A}^H(\mathbf{CD}+\mathbf{QR})(\mathbf{D}+\mathbf{R})^{\#}$
  - $[\mathbf{D}=\mathbf{D}^H, \mathbf{R}=\mathbf{R}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AXB}+\mathbf{C})\mathbf{D}(\mathbf{AXB}+\mathbf{C})^H+(\mathbf{AX})\mathbf{R}(\mathbf{AX})^H)\} = -(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{A}^H(\mathbf{CDB}^H)(\mathbf{BDB}^H+\mathbf{R})^{\#}$
  - $[\mathbf{D}=\mathbf{D}^H, \mathbf{R}=\mathbf{R}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{XB}+\mathbf{C})\mathbf{D}(\mathbf{XB}+\mathbf{C})^H+\mathbf{XRX}^H)\} = -(\mathbf{CDB}^H)(\mathbf{BDB}^H+\mathbf{R})^{\#}$
- $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AXB}+\mathbf{C})\mathbf{D}(\mathbf{AXB}+\mathbf{C})^H) \mid \mathbf{EXF}=\mathbf{G}\} = (\mathbf{A}^H\mathbf{A})^{\#}(\mathbf{E}^H\{\mathbf{E}(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{E}^H\}^{\#}\{\mathbf{E}(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{A}^H\mathbf{CDB}^H(\mathbf{BDB}^H)^{\#}\mathbf{F}+\mathbf{G}\}\{\mathbf{F}^H(\mathbf{BDB}^H)^{\#}\mathbf{F}\}^{\#}\mathbf{F}^H - \mathbf{A}^H\mathbf{CDB}^H)(\mathbf{BDB}^H)^{\#}$ [2.8]
  - $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AX}+\mathbf{C})\mathbf{D}(\mathbf{AX}+\mathbf{C})^H) \mid \mathbf{EX}=\mathbf{G}\} = (\mathbf{A}^H\mathbf{A})^{\#}(\mathbf{E}^H\{\mathbf{E}(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{E}^H\}^{\#}\{\mathbf{E}(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{A}^H\mathbf{C}+\mathbf{G}\} - \mathbf{A}^H\mathbf{C})$
  - $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{Ax}+\mathbf{c})\mathbf{D}(\mathbf{Ax}+\mathbf{c})^H) \mid \mathbf{Ex}=\mathbf{g}\} = (\mathbf{A}^H\mathbf{A})^{\#}(\mathbf{E}^H\{\mathbf{E}(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{E}^H\}^{\#}\{\mathbf{E}(\mathbf{A}^H\mathbf{A})^{\#}\mathbf{A}^H\mathbf{c}+\mathbf{g}\} - \mathbf{A}^H\mathbf{c})$
- $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AXB}+\mathbf{C})^H\mathbf{D}(\mathbf{AXB}+\mathbf{C})) \mid \mathbf{EXF}=\mathbf{G}\} = (\mathbf{A}^H\mathbf{DA})^{\#}(\mathbf{E}^H\{\mathbf{E}(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{E}^H\}^{\#}\{\mathbf{E}(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{A}^H\mathbf{DCB}^H(\mathbf{BB}^H)^{\#}\mathbf{F}+\mathbf{G}\}\{\mathbf{F}^H(\mathbf{BB}^H)^{\#}\mathbf{F}\}^{\#}\mathbf{F}^H - \mathbf{A}^H\mathbf{DCB}^H)(\mathbf{BB}^H)^{\#}$
  - $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{X}}\{\text{tr}((\mathbf{AX}+\mathbf{C})^H\mathbf{D}(\mathbf{AX}+\mathbf{C})) \mid \mathbf{EX}=\mathbf{G}\} = (\mathbf{A}^H\mathbf{DA})^{\#}(\mathbf{E}^H\{\mathbf{E}(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{E}^H\}^{\#}\{\mathbf{E}(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{A}^H\mathbf{DC}+\mathbf{G}\}- \mathbf{A}^H\mathbf{DC})$
  - $[\mathbf{D}=\mathbf{D}^H]$ $\text{argmin}_{\mathbf{x}}\{(\mathbf{Ax}+\mathbf{c})^H\mathbf{D}(\mathbf{Ax}+\mathbf{c}) \mid \mathbf{Ex}=\mathbf{g}\} = (\mathbf{A}^H\mathbf{DA})^{\#}(\mathbf{E}^H\{\mathbf{E}(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{E}^H\}^{\#}\{\mathbf{E}(\mathbf{A}^H\mathbf{DA})^{\#}\mathbf{A}^H\mathbf{Dc}+\mathbf{g}\}- \mathbf{A}^H\mathbf{Dc})$

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.

- $d(\det(\mathbf{X})) = d(\det(\mathbf{X}^T)) = \mathbf{ADJ}(\mathbf{X}^T)\mathbf{:}^T\, d\mathbf{X:} = \det(\mathbf{X})\ (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.10]
- $d(\det(\mathbf{A}^T\mathbf{XB})) = d(\det(\mathbf{B}^T\mathbf{X}^T\mathbf{A})) = (\mathbf{A}\ \mathbf{ADJ}(\mathbf{A}^T\mathbf{XB})^T\mathbf{B}^T)\mathbf{:}^T\ d\mathbf{X:} =$ $[\mathbf{A},\mathbf{B}\text{: nonsingular}]$ $\det(\mathbf{A}^T\mathbf{XB}) \times (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.11]
- $d(\ln(\det(\mathbf{A}^T\mathbf{XB}))) =$ $[\mathbf{A},\mathbf{B}\text{: nonsingular}]$ $(\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.12]
  - $d(\ln(\det(\mathbf{X}))) = (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\det(\mathbf{X}^k)) = d(\det(\mathbf{X})^k) = k \times \det(\mathbf{X}^k) \times (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$ [2.13]
- $d(\ln(\det(\mathbf{X}^k))) = k \times (\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\det(\mathbf{X}^T\mathbf{CX})) =$ $[\mathbf{C}=\mathbf{C}^T]$ $2(\mathbf{CX}\ \mathbf{ADJ}(\mathbf{X}^T\mathbf{CX}))\mathbf{:}^T\, d\mathbf{X:} = 2\det(\mathbf{X}^T\mathbf{CX})\times(\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X:}$ [2.14]
  - $=$ $[\mathbf{C}=\mathbf{C}^T, \mathbf{CX}\text{: nonsingular}]$ $2\det(\mathbf{X}^T\mathbf{CX})\times(\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\ln(\det(\mathbf{X}^T\mathbf{CX}))) =$ $[\mathbf{C}=\mathbf{C}^T]$ $2(\mathbf{CX}(\mathbf{X}^T\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X:}$
  - $=$ $[\mathbf{C}=\mathbf{C}^T, \mathbf{CX}\text{: nonsingular}]$ $2(\mathbf{X}^{-T})\mathbf{:}^T\, d\mathbf{X:}$
- $d(\det(\mathbf{X}^H\mathbf{CX})) = \det(\mathbf{X}^H\mathbf{CX}) \times (\mathbf{C}^T\mathbf{X}^C\ (\mathbf{X}^T\mathbf{C}^T\mathbf{X}^C)^{-1})\mathbf{:}^T d\mathbf{X:}\ + (\mathbf{CX}(\mathbf{X}^H\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X}^C\mathbf{:})$ [2.15]
- $d(\ln(\det(\mathbf{X}^H\mathbf{CX}))) = (\mathbf{C}^T\mathbf{X}^C\ (\mathbf{X}^T\mathbf{C}^T\mathbf{X}^C)^{-1})\mathbf{:}^T d\mathbf{X:}\ + (\mathbf{CX}(\mathbf{X}^H\mathbf{CX})^{-1})\mathbf{:}^T\, d\mathbf{X}^C\mathbf{:}$ [2.16]

Jacobian

$d\mathbf{Y}/d\mathbf{X}$ is called the *Jacobian Matrix* of **Y:** with respect to **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 **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 **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
  - $\mathbf{H_x}\ (\mathbf{a}^T\mathbf{x}) = 0$
  - $\mathbf{H_x}\ (\mathbf{Ax+b})^T\mathbf{C}(\mathbf{Dx+e}) = \mathbf{A}^T\mathbf{CD} + \mathbf{D}^T\mathbf{C}^T\mathbf{A}$
    - $\mathbf{H_x}\ (\mathbf{Ax+b})^T\ (\mathbf{Dx+e}) = \mathbf{A}^T\mathbf{D} + \mathbf{D}^T\mathbf{A}$
    - $\mathbf{H_x}\ (\mathbf{Ax+b})^T\mathbf{C}(\mathbf{Ax+b}) = \mathbf{A}^T(\mathbf{C} + \mathbf{C}^T)\mathbf{A}$ = [$\mathbf{C}=\mathbf{C}^T$] $2\mathbf{A}^T\mathbf{CA}$
      - $\mathbf{H_x}\ (\mathbf{Ax+b})^T\ (\mathbf{Ax+b}) = 2\mathbf{A}^T\mathbf{A}$
      - $\mathbf{H_x}\ (\mathbf{x}^T\mathbf{Cx}) = \mathbf{C}+\mathbf{C}^T$ = [$\mathbf{C}=\mathbf{C}^T$] $2\mathbf{C}$
      - $\mathbf{H_x}\ (\mathbf{x}^T\mathbf{x}) = 2\mathbf{I}$
- [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
  - $\mathbf{H_x}\ (\mathbf{Ax+b})^H\mathbf{C}(\mathbf{Ax+b})$ = [$\mathbf{C}=\mathbf{C}^H$] $(\mathbf{A}^H\mathbf{CA})^T$ [2.17]
    - $\mathbf{H_x}\ (\mathbf{x}^H\mathbf{Cx})$ = [$\mathbf{C}=\mathbf{C}^H$] $\mathbf{C}^T$
