Matricks

Michael Syskind Pedersen*

Revision 1.2

January 6, 2005

1 Matrix Notation and Basic Operations

A matrix is defined the following way:

$$\mathbf{A} \in \mathbb{R}^{m \times n} \Leftrightarrow \mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \qquad a_{ij} \in \mathbb{R} \tag{1}$$

Some calculations may involve complex matrices. A complex matrix is defined as

$$\mathbf{A} = \mathbf{B} + i\mathbf{C} \in \mathbb{C}^{m \times n}, \tag{2}$$

where the real part is given by $\Re\mathbf{A} = \mathbf{B} \in \mathbb{R}^{m \times n}$ and the imaginary part is given by $\Im\mathbf{A} = \mathbf{C} \in \mathbb{R}^{m \times n}$.
The *complex conjugation* of a matrix is denoted as

$$\mathbf{A}^* = (\mathbf{B} + i\mathbf{C})^* = \mathbf{B} - i\mathbf{C} \tag{3}$$

A vector is defined the following way:

$$\mathbf{a} \in \mathbb{R}^{m \times 1} \Leftrightarrow \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \qquad a_i \in \mathbb{R} \tag{4}$$

*Thanks to the Oticon Foundation for funding

In the following the matrices and vectors is allowed to be complex unless other is stated.
The *transposition* of a matrix is denoted as

$$\mathbf{A}^T \Leftrightarrow (a_{ij})^T = a_{ji} \tag{5}$$

The *conjugate transposition* (or *Hermitian transposition*) of a matrix is denoted as

$$\mathbf{A}^H \Leftrightarrow (a_{ij})^H = (a_{ij}^*)^T = a_{ji}^* \tag{6}$$

A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is called *Hermitian* if $\mathbf{A} = \mathbf{A}^H$. Notice, if $\mathbf{A} \in \mathbb{R}^{m\times n}$ then $\mathbf{A}^H = \mathbf{A}^T$ since $\mathbf{A}^* = \mathbf{A}$.
If $\mathbf{A} \in \mathbb{C}^{m\times p}$ and $\mathbf{B} \in \mathbb{C}^{p\times n}$ then the *product* between $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\mathbf{AB} = \mathbf{C} \in \mathbb{C}^{m\times n} \Leftrightarrow c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj} \tag{7}$$

The *trace* of a matrix is defined as

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}, \qquad \mathbf{A} \in \mathbb{C}^{n\times n} \tag{8}$$

Combining (7) and (8) yields

$$\text{tr}(\mathbf{AB}) = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik} b_{ki}, \qquad \mathbf{A} \in \mathbb{C}^{n\times n} \tag{9}$$

The *submatrix* of a matrix $\mathbf{A}$, denoted by $(\mathbf{A})_{ij}$ is a $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and the jth column of $\mathbf{A}$.
The *cofactor* of a submatrix $(\mathbf{A})_{ij}$ can be found as

$$\text{cof}(\mathbf{A})_{ij} = (-1)^{i+j} \det(\mathbf{A})_{ij}, \tag{10}$$

where $\det$ is the *determinant* (see (15)). The *matrix of cofactors* can be created from the cofactors

$$\text{cof}(\mathbf{A}) = \begin{bmatrix} \text{cof}(\mathbf{A})_{11} & \cdots & \text{cof}(\mathbf{A})_{1n} \\ \vdots & \text{cof}(\mathbf{A})_{ij} & \vdots \\ \text{cof}(\mathbf{A})_{n1} & \cdots & \text{cof}(\mathbf{A})_{nn} \end{bmatrix} \tag{11}$$

The *adjoint* matrix is the transpose of the cofactor matrix

$$\text{adj}(\mathbf{A}) = (\text{cof}(\mathbf{A}))^T, \tag{12}$$

and

$$\text{adj}(\mathbf{A})_{ij} = \text{adj}(\mathbf{A})_{ji} = (-1)^{i+j} \det(\mathbf{A})_{ji} \tag{13}$$

The *determinant* of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is defined (Golub and van Loan [1996]) as

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det\left((\mathbf{A})_{1j}\right) \tag{14}$$

$$= \sum_{j=1}^{n} a_{1j} \mathrm{cof}(\mathbf{A})_{1j}. \tag{15}$$

The *inverse* $\mathbf{A}^{-1}$ of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is defined such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}, \tag{16}$$

where $\mathbf{I}$ is the $n \times n$ *identity matrix*. If $\mathbf{A}^{-1}$ exists, $\mathbf{A}$ is said to be *nonsingular*. Otherwise, $\mathbf{A}$ is said to be *singular* (Golub and van Loan [1996]).
The inverse matrix can be calculated by the following formula

$$\mathbf{A}^{-1} = \frac{\mathrm{adj}(\mathbf{A})}{\det(\mathbf{A})} \tag{17}$$

The *Hadamard product* is defined as the product of corresponding elements, i.e.

$$\mathbf{A} \circ \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{bmatrix} \tag{18}$$

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{p\times q}$. The *Kronecker product* is then defined as (Golub and van Loan [1996])

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \tag{19}$$

The dimension of the resulting matrix is thus $mp \times nq$.

1.1 Eigenvalues and Eigenvectors

Consider the linear system

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \tag{20}$$

where $\mathbf{A} \in \mathbb{C}^{n\times n}$, $\mathbf{x} \in \mathbb{C}^{n\times 1}$, $\lambda \in \mathbb{C}$. If λ satisfies (20), λ is referred to as an *eigenvalue*. The n eigenvectors $\lambda_1, \ldots, \lambda_n$ are defined by solving the *characteristic equation*:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{21}$$

The left hand side of the characteristic equation is called the *characteristic polynomial*. The corresponding vectors $\mathbf{x}$ that satisfy (20) are called *eigenvectors*. In general, $\mathbf{x}$ is referred to as the *right eigenvector*, if it satisfies $\mathbf{Ax} = \lambda\mathbf{x}$ and the *left eigenvector* if it satisfies $\mathbf{x}^H\mathbf{A} = \lambda\mathbf{x}^H$ (Golub and van Loan [1996]). When the eigenvalues are known, the eigenvectors are found by solving the linear system

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \tag{22}$$

2 Properties of Products, Transposition, Determinant, Trace and Inverse

The formulas from are taken from Roweis [1999], Minka [2000], Stainvas [2002] and Golub and van Loan [1996]. In these references though, it is not always clear whether the formulas hold if the matrices contain complex numbers.

$$\begin{aligned}
\mathbf{A}(\mathbf{B}+\mathbf{C}) &= \mathbf{AB}+\mathbf{AC}, && \mathbf{A}\in\mathbb{C}^{m\times p}, \mathbf{B},\mathbf{C}\in\mathbb{C}^{p\times n} && (23)\\
(\mathbf{A}+\mathbf{B})^H &= \mathbf{A}^H+\mathbf{B}^H, && \mathbf{A},\mathbf{B}\in\mathbb{C}^{m\times n} && (24)\\
(\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T, && \mathbf{A}\in\mathbb{C}^{m\times p}, \mathbf{B}\in\mathbb{C}^{p\times n} && (25)\\
(\mathbf{ABC}\ldots)^T &= \ldots\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T, && \mathbf{ABC}\ldots\in\mathbb{C}^{m\times n} && (26)\\
(\mathbf{AB})^H &= \mathbf{B}^H\mathbf{A}^H, && \mathbf{A}\in\mathbb{C}^{m\times p}, \mathbf{B}\in\mathbb{C}^{p\times n} && (27)\\
(\mathbf{ABC}\ldots)^H &= \ldots\mathbf{C}^H\mathbf{B}^H\mathbf{A}^H, && \mathbf{ABC}\ldots\in\mathbb{C}^{m\times n} && (28)\\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1}, && \mathbf{A}\in\mathbb{C}^{n\times n}, \mathbf{B}\in\mathbb{C}^{n\times n} && (29)\\
(\mathbf{A}^{-1})^T &= (\mathbf{A}^T)^{-1}, && \mathbf{A}\in\mathbb{C}^{n\times n} && (30)\\
(\mathbf{A}^{-1})^H &= (\mathbf{A}^H)^{-1}, && \mathbf{A}\in\mathbb{C}^{n\times n} && (31)
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}(\mathbf{A}) &= \sum_{i=1}^{n}\lambda_i, && \mathbf{A}\in\mathbb{C}^{n\times n} && (32)\\
\mathrm{tr}(\mathbf{A}) &= \sum_{i=1}^{n}a_{ii}, && \mathbf{A}\in\mathbb{C}^{n\times n} && (33)\\
\mathrm{tr}(\mathbf{A}) &= \mathrm{tr}(\mathbf{A}^T), && \mathbf{A}\in\mathbb{C}^{n\times n} && (34)\\
\mathrm{tr}(\mathbf{A}) &= \mathrm{tr}((\mathbf{A}^*)^H) = \mathrm{tr}(\mathbf{A}^H)^*, && \mathbf{A}\in\mathbb{C}^{n\times n} && (35)\\
\mathrm{tr}(\mathbf{A}+\mathbf{B}) &= \mathrm{tr}(\mathbf{A})+\mathrm{tr}(\mathbf{B}), && \mathbf{A},\mathbf{B}\in\mathbb{C}^{n\times n} && (36)\\
\mathrm{tr}(\mathbf{AB}) &= \mathrm{tr}(\mathbf{BA}), && \mathbf{A},\mathbf{B}\in\mathbb{C}^{n\times n} && (37)
\end{aligned}$$

(37) can be extended to the more general case (Roweis [1999]):

$$\mathrm{tr}(\mathbf{ABC}\ldots) = \mathrm{tr}(\mathbf{BC}\ldots\mathbf{A}) = \mathrm{tr}(\mathbf{C}\ldots\mathbf{AB}) = \ldots, \qquad \mathbf{A},\mathbf{B},\mathbf{C},\ldots\in\mathbb{C}^{n\times n} \tag{38}$$

$$\det(\mathbf{A}) = \prod_{i=1}^{n}\lambda_i, \qquad \mathbf{A}\in\mathbb{C}^{n\times n} \tag{39}$$

$$\begin{aligned}
\det(\mathbf{A}^T) &= \det(\mathbf{A}), & \mathbf{A} \in \mathbb{C}^{n\times n} \quad (40)\\
\det(\mathbf{A}^H) &= \det(\mathbf{A}^*) = \det(\mathbf{A})^*, & \mathbf{A} \in \mathbb{C}^{n\times n} \quad (41)\\
\det(\mathbf{A}^{-1}) &= \frac{1}{\det(\mathbf{A})}, & \mathbf{A} \in \mathbb{C}^{n\times n} \quad (42)\\
\det(\mathbf{AB}) &= \det(\mathbf{A})\det(\mathbf{B}), & \mathbf{A},\mathbf{B} \in \mathbb{C}^{n\times n} \quad (43)\\
\det(\alpha\mathbf{A}) &= \alpha^n \det(\mathbf{A}), & \mathbf{A} \in \mathbb{C}^{n\times n} \quad (44)
\end{aligned}$$

2.1 Matrix Inversions

Several matrix inversion formulas exist.
Sherman-Morrison Formula: Given the matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and the vectors $\mathbf{u}$ and $\mathbf{v}$, where $\mathbf{u} \otimes \mathbf{v} \in \mathbb{R}^{n\times n}$. Then (Press et al. [2002])

$$(\mathbf{A} + \mathbf{u}\otimes\mathbf{v})^{-1} = \mathbf{A}^{-1} + \frac{(\mathbf{A}^{-1}\mathbf{u})\otimes(\mathbf{v}\mathbf{A}^{-1})}{1+\lambda}, \tag{45}$$

where

$$\lambda = \mathbf{v}\mathbf{A}^{-1}\mathbf{u}. \tag{46}$$

The following type of matrix inversion formulas are derived by considering block matrices.
Woodbury Formula ([Bishop, 1995, p. 153])

$$(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \tag{47}$$

(III [2003], Roweis [1999]))

$$(\mathbf{A} + \mathbf{B}\mathbf{X}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{X} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \tag{48}$$

$$(\mathbf{A} + \mathbf{X}\mathbf{B}\mathbf{X}^{T})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{B}^{-1} + \mathbf{X}^T\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{A}^{-1} \tag{49}$$

Matrices can as well be inverted by partitioning the matrix (Press et al. [2002]). Consider the following matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{R} & \mathbf{S} \end{bmatrix}, \tag{50}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are square matrices. If

$$\mathbf{A}^{-1} = \begin{bmatrix} \tilde{\mathbf{P}} & \tilde{\mathbf{Q}} \\ \tilde{\mathbf{R}} & \tilde{\mathbf{S}} \end{bmatrix}, \tag{51}$$

then $\tilde{\mathbf{P}}$, $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{S}}$ can be written as

$$\begin{aligned}
\tilde{\mathbf{P}} &= (\mathbf{P} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{R})^{-1} && (52)\\
&= \mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{Q}(\mathbf{S} - \mathbf{R}\mathbf{P}^{-1}\mathbf{Q})^{-1}\mathbf{R}\mathbf{P}^{-1} && (53)\\
\tilde{\mathbf{Q}} &= -(\mathbf{P} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{R})^{-1}\mathbf{Q}\mathbf{S}^{-1} && (54)\\
&= -\mathbf{P}^{-1}\mathbf{Q}(\mathbf{S} - \mathbf{R}\mathbf{P}^{-1}\mathbf{Q})^{-1} && (55)\\
\tilde{\mathbf{R}} &= -(\mathbf{S} - \mathbf{R}\mathbf{P}^{-1}\mathbf{Q})^{-1}\mathbf{Q}\mathbf{P}^{-1} && (56)\\
&= -\mathbf{S}^{-1}\mathbf{R}(\mathbf{P} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{R})^{-1} && (57)\\
\tilde{\mathbf{S}} &= (\mathbf{S} - \mathbf{R}\mathbf{P}^{-1}\mathbf{Q})^{-1} && (58)\\
&= \mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{R}(\mathbf{P} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{R})^{-1}\mathbf{Q}\mathbf{S}^{-1}. && (59)
\end{aligned}$$

The determinant of the partitioned matrix $\mathbf{A}$ can be written as (Press et al. [2002])

$$\begin{aligned}
\det \mathbf{A} &= \det \mathbf{P} \det(\mathbf{S} - \mathbf{R}\mathbf{P}^{-1}\mathbf{Q}) && (60)\\
&= \det \mathbf{S} \det(\mathbf{P} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{R}). && (61)
\end{aligned}$$

3 Matrix Derivatives

Several kinds of derivatives can be expressed as scalars, vectors or matrices:

- Scalar differentiated with respect to a scalar: $\frac{dy}{dx}$
- Scalar differentiated with respect to a vector: $\frac{dy}{d\mathbf{x}} = \frac{\partial y}{\partial x_j}$
- Scalar differentiated with respect to a matrix: $\frac{dy}{d\mathbf{X}} = \frac{\partial y}{\partial x_{ij}}$
- Vector differentiated with respect to a scalar: $\frac{d\mathbf{y}}{dx} = \frac{\partial y_i}{\partial x}$
- Vector differentiated with respect to a vector: $\frac{d\mathbf{y}}{d\mathbf{x}} = \frac{\partial y_i}{\partial x_j}$
- Matrix differentiated with respect to a scalar: $\frac{d\mathbf{Y}}{dx} = \frac{\partial y_{ij}}{\partial x}$

The following rules are very useful when deriving the differential of an expression (Minka [2000]):

$$\begin{aligned}
\partial \mathbf{A} &= 0 && (\mathbf{A} \text{ is a constant}) && (62)\\
\partial(\alpha \mathbf{X}) &= \alpha \partial \mathbf{X} && && (63)\\
\partial(\mathbf{X} + \mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} && && (64)\\
\partial(\mathrm{tr}(\mathbf{X})) &= \mathrm{tr}(\partial \mathbf{X}) && && (65)\\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) && && (66)\\
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) && && (67)
\end{aligned}$$

$$\begin{aligned}
\partial(\mathbf{X}\otimes\mathbf{Y}) &= (\partial\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes(\partial\mathbf{Y}) && (68)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} && (69)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (70)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (71)\\
\partial\mathbf{X}^T &= (\partial\mathbf{X})^T && (72)\\
\partial\mathbf{X}^H &= (\partial\mathbf{X})^H && (73)
\end{aligned}$$

3.1 Complex Derivatives

In order to differentiate an expression $f(z)$ with respect to a complex z, the Cauchy-Riemann equations have to be satisfied (Brookes [2003]):

$$\frac{df(z)}{dz}=\frac{\partial\Re(f(z))}{\partial\Re z}+i\frac{\partial\Im(f(z))}{\partial\Re z} \tag{74}$$

and

$$\frac{df(z)}{dz}=-i\frac{\partial\Re(f(z))}{\partial\Im z}+\frac{\partial\Im(f(z))}{\partial\Im z} \tag{75}$$

or in a more compact form:

$$\frac{\partial f(z)}{\partial\Im z}=i\frac{\partial f(z)}{\partial\Re z}. \tag{76}$$

A complex function that satisfies the Cauchy-Riemann equations for points in a region R is said yo be *analytic* in this region R. In general, expressions involving complex conjugate or conjugate transpose do not satisfy the Cauchy-Riemann equations. In order to avoid this problem, a more generalized definition of complex derivative is used (Schwartz [1967], Brandwood [1983]):

- Generalized Complex Derivative:

$$\frac{df(z)}{dz}=\frac{1}{2}\Big(\frac{\partial f(z)}{\partial\Re z}-i\frac{\partial f(z)}{\partial\Im z}\Big) \tag{77}$$

- Conjugate Complex Derivative

$$\frac{df(z)}{dz^*}=\frac{1}{2}\Big(\frac{\partial f(z)}{\partial\Re z}+i\frac{\partial f(z)}{\partial\Im z}\Big) \tag{78}$$

The Generalized Complex Derivative equals the normal derivative, when f is an analytic function. For a non-analytic function such as $f(z)=z^*$, the derivative equals zero. The Conjugate Complex Derivative equals zero, when f is an analytic function. The Conjugate Complex Derivative has e.g been used by Parra and Spence [2000] when deriving a complex gradient.
Notice:

$$\frac{df(z)}{dz}\neq\frac{\partial f(z)}{\partial\Re z}+i\frac{\partial f(z)}{\partial\Im z} \tag{79}$$

- Complex Gradient Vector: If f is a real function of a complex vector $\mathbf{z}$, then the complex gradient vector is given by ([Haykin, 2002, p. 798])

$$
\begin{aligned}
\nabla f(\mathbf{z}) &= 2\frac{df(\mathbf{z})}{d\mathbf{z}^*} \qquad (80)\\
&= \frac{\partial f(\mathbf{z})}{\partial \Re \mathbf{z}} + i\frac{\partial f(\mathbf{z})}{\partial \Im \mathbf{z}}
\end{aligned}
$$

- Complex Gradient Matrix: If f is a real function of a complex matrix $\mathbf{Z}$, then the complex gradient matrix is given by (Anemüller et al. [2003])

$$
\begin{aligned}
\nabla f(\mathbf{Z}) &= 2\frac{df(\mathbf{Z})}{d\mathbf{Z}^*} \qquad (81)\\
&= \frac{\partial f(\mathbf{Z})}{\partial \Re \mathbf{Z}} + i\frac{\partial f(\mathbf{Z})}{\partial \Im \mathbf{Z}}
\end{aligned}
$$

  These expressions can be used for gradient descent algorithms.

3.2 Scalar differentiated with respect to matrices

Let $y = f(\mathbf{X})$ be a scalar function of $\mathbf{X} \in \mathbb{C}^{m \times n}$. Then the derivative of y is defined as [1]

$$
\frac{dy}{d\mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} \qquad (82)
$$

This matrix is known as the *gradient matrix* (Felippa [2003]).

3.2.1 The Chain Rule

Sometimes the objective is to find the derivative of a matrix which is a function of another matrix. Let $\mathbf{U} = f(\mathbf{X})$, the goal is to find the derivative of the function g($\mathbf{U}$) with respect to $\mathbf{X}$:

$$
\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}} \qquad (83)
$$

[1] The definition of matrix derivative agrees with the definitions in Roweis [1999], Stainvas [2002], Felippa [2003]. The matrix derivative in Minka [2000] is defined in a different way:

$$
\frac{dy}{d\mathbf{X}} = \frac{dy}{d\mathbf{X}^T}
$$

Then the Chain Rule can then be written the following way:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}} \tag{84}$$

Using matrix notation, this can be written as:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\mathrm{tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \partial \mathbf{U})}{\partial \mathbf{X}}. \tag{85}$$

3.2.2 The Chain Rule for complex numbers

The chain rule is a little more complicated when the function of a complex $u = f(x)$ is non-analytic. For a non-analytic function, the following chain rule can be applied (Brookes [2003])

$$\begin{aligned} \frac{\partial g(u)}{\partial x} &= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial u^*}\frac{\partial u^*}{\partial x} \\ &= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \left(\frac{\partial g^*}{\partial u}\right)^* \frac{\partial u^*}{\partial x} \end{aligned} \tag{86}$$

Notice, if the function is analytic, the second term reduces to zero, and the function is reduced to the normal well-known chain rule. For the matrix derivative of a scalar function $g(\mathbf{U})$, the chain rule can be written the following way:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\mathrm{tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \partial \mathbf{U})}{\partial \mathbf{X}} + \frac{\mathrm{tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}^*})^T \partial \mathbf{U}^*)}{\partial \mathbf{X}}. \tag{87}$$

3.2.3 Basic derivatives

Assume that all the arguments are real and the argument inside the trace is square. Further, assume that the elements of $\mathbf{X}$ are functionally independent.

$$\frac{\partial(\mathbf{a}^T\mathbf{X}\mathbf{b})}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{88}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^T\mathbf{b})}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T \tag{89}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{b})}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{a}\mathbf{b}^T + \mathbf{b}\mathbf{a}^T) \tag{90}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{a})}{\partial \mathbf{X}} = 2\mathbf{X}\mathbf{a}\mathbf{a}^T \tag{91}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{b})}{\partial \mathbf{X}} = \mathbf{C}^T\mathbf{X}\mathbf{a}\mathbf{b}^T + \mathbf{C}\mathbf{X}\mathbf{b}\mathbf{a}^T \tag{92}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{a})}{\partial \mathbf{X}} = (\mathbf{C} + \mathbf{C}^T)\mathbf{X}\mathbf{a}\mathbf{a}^T \tag{93}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^T\mathbf{\Sigma}\mathbf{X}\mathbf{a})}{\partial\mathbf{X}} = 2\mathbf{\Sigma}\mathbf{X}\mathbf{a}\mathbf{a}^T, \qquad \mathbf{\Sigma} = \mathbf{\Sigma}^T \tag{94}$$

$$\frac{\partial((\mathbf{X}\mathbf{a}+\mathbf{b})^T\mathbf{C}(\mathbf{X}\mathbf{a}+\mathbf{b}))}{\partial\mathbf{X}} = (\mathbf{C}+\mathbf{C}^T)(\mathbf{X}\mathbf{a}+\mathbf{b})\mathbf{a}^T \tag{95}$$

Provided $\mathbf{X}^{-1}$ exists:

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^{-1}\mathbf{b})}{\partial\mathbf{X}} = -(\mathbf{X}^T)^{-1}\mathbf{a}\mathbf{b}^T(\mathbf{X}^T)^{-1} \tag{96}$$

Useful formulas involving derivatives of traces and determinants exist.

3.2.4 Derivatives of Traces

One of the most common matrix norms is the *Frobenius* norm. The Frobenius norm is defined as (Golub and van Loan [1996])

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{m}\sum_{i=j}^{n}|a_{ij}|^2}. \tag{97}$$

Using matrix notation, the *squared Frobenius* norm can be written as

$$\sum_{i=1}^{m}\sum_{i=j}^{n}|a_{ij}|^2 = \mathrm{tr}(\mathbf{A}^H\mathbf{A}) \tag{98}$$

Therefore, matrix derivatives of traces are a common problem. Then the following formulas exists:

$$\frac{\partial \mathrm{tr}(\mathbf{X})}{\partial\mathbf{X}} = \frac{\partial \mathrm{tr}(\mathbf{X}^T)}{\partial\mathbf{X}} = \mathbf{I} \tag{99}$$

$$\frac{\partial \mathrm{tr}(\mathbf{X}^k)}{\partial\mathbf{X}} = k(\mathbf{X}^{k-1})^T \tag{100}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X}^k)}{\partial\mathbf{X}} = \sum_{r=0}^{k-1}(\mathbf{X}^r\mathbf{A}\mathbf{X}^{k-r-1})^T \tag{101}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X})}{\partial\mathbf{X}} = \frac{\partial \mathrm{tr}(\mathbf{X}\mathbf{A})}{\partial\mathbf{X}} = \mathbf{A}^T, \text{see (143)} \tag{102}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X}^T)}{\partial\mathbf{X}} = \frac{\partial \mathrm{tr}(\mathbf{X}^T\mathbf{A})}{\partial\mathbf{X}} = \mathbf{A} \tag{103}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial\mathbf{X}} = \frac{\partial \mathrm{tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T)}{\partial\mathbf{X}} = \frac{\partial \mathrm{tr}(\mathbf{B}\mathbf{A}\mathbf{X})}{\partial\mathbf{X}} = (\mathbf{B}\mathbf{A})^T \tag{104}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A}^T\mathbf{X}\mathbf{B}^T)}{\partial\mathbf{X}} = \frac{\partial \mathrm{tr}(\mathbf{B}\mathbf{X}^T\mathbf{A})}{\partial\mathbf{X}} = \mathbf{A}\mathbf{B} \tag{105}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X})}{\partial\mathbf{X}} = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T + \mathbf{B}^T\mathbf{X}^T\mathbf{A}^T \tag{106}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}\mathbf{X}^T)}{\partial\mathbf{X}} = \mathbf{A}\mathbf{X}^T\mathbf{B} + \mathbf{B}\mathbf{X}^T\mathbf{A} \tag{107}$$

$$\frac{\partial \text{tr}(\mathbf{AXBX}^T\mathbf{C})}{\partial \mathbf{X}} = \mathbf{A}^T\mathbf{C}^T\mathbf{XB}^T + \mathbf{CAXB} \tag{108}$$

$$\frac{\partial \text{tr}(\mathbf{AX}^T\mathbf{BXC})}{\partial \mathbf{X}} = \mathbf{BXCA} + \mathbf{B}^T\mathbf{XA}^T\mathbf{C}^T \tag{109}$$

$$\frac{\partial \text{tr}(\mathbf{XAX}^T)}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{AX}^T\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}^T\mathbf{XA})}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{A} + \mathbf{A}^T) \tag{110}$$

$$\frac{\partial \text{tr}(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{AXX}^T)}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{XX}^T\mathbf{A})}{\partial \mathbf{X}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{X} \tag{111}$$

Notice, (110) and (111) are special cases of (108) and (109), respectively. Provided that $\mathbf{X}^{-1}$ exists, the following expressions can be proved.

$$\frac{\partial \text{tr}(\mathbf{AX}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1})^T\mathbf{A}^T\mathbf{B}^T(\mathbf{X}^{-1})^T \tag{112}$$

$$\frac{\partial \text{tr}(\mathbf{AX}^{-1})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}^{-1}\mathbf{A})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1})^T\mathbf{A}^T(\mathbf{X}^{-1})^T \tag{113}$$

The following is true provided $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ is symmetric:

$$\frac{\partial \text{tr}((\mathbf{X\Sigma X}^T)^{-1}\mathbf{A})}{\partial \mathbf{X}} = -(\mathbf{\Sigma X}(\mathbf{X}^T\mathbf{\Sigma X})^{-1})(\mathbf{A} + \mathbf{A}^T)(\mathbf{X}^T\mathbf{\Sigma X})^{-1} \tag{114}$$

$$\begin{aligned}\frac{\partial \text{tr}((\mathbf{X\Sigma_1X}^T)^{-1}(\mathbf{X\Sigma_2X}^T))}{\partial \mathbf{X}} &= -2\mathbf{\Sigma_1X}(\mathbf{X}^T\mathbf{\Sigma_1X})^{-1}\mathbf{X}^T\mathbf{\Sigma_2X}(\mathbf{X}^T\mathbf{\Sigma_1X})^{-1} \\ &+ 2\mathbf{\Sigma_2X}(\mathbf{X}^T\mathbf{\Sigma_1X})^{-1}\end{aligned} \tag{115}$$

3.2.5 Complex Derivatives of Traces

If the derivatives involve complex numbers, the conjugate transpose is often involved. The most useful way to show complex derivative is to show the derivative with respect to the real and the imaginary part separately. An easy example is:

$$\frac{\partial \text{tr}(\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{116}$$

$$i\frac{\partial \text{tr}(\mathbf{X}^*)}{\partial \Im \mathbf{X}} = i\frac{\partial \text{tr}(\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{I} \tag{117}$$

Since the two results have the same sign, the conjugate complex derivative (78) should be used.

$$\frac{\partial \text{tr}(\mathbf{X})}{\partial \Re \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}^T)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{118}$$

$$i\frac{\partial \text{tr}(\mathbf{X})}{\partial \Im \mathbf{X}} = i\frac{\partial \text{tr}(\mathbf{X}^T)}{\partial \Im \mathbf{X}} = -\mathbf{I} \tag{119}$$

Here, the two results have different signs, the generalized complex derivative (77) should be used. Hereby, it can be seen that (99) holds even if $\mathbf{X}$ is a complex number.

$$\frac{\partial \text{tr}(\mathbf{AX}^H)}{\partial \Re \mathbf{X}} = \mathbf{A} \tag{120}$$

$$i\frac{\partial \text{tr}(\mathbf{A}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{A} \quad (121)$$

$$\frac{\partial \text{tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \mathbf{A}^T \quad (122)$$

$$i\frac{\partial \text{tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Im \mathbf{X}} = \mathbf{A}^T \quad (123)$$

$$\frac{\partial \text{tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}^H\mathbf{X})}{\partial \Re \mathbf{X}} = 2\Re \mathbf{X} \quad (124)$$

$$i\frac{\partial \text{tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = i\frac{\partial \text{tr}(\mathbf{X}^H\mathbf{X})}{\partial \Im \mathbf{X}} = i2\Im \mathbf{X} \quad (125)$$

By inserting (124) and (125) in (77) and (78), it can be seen that

$$\frac{\partial \text{tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}} = \mathbf{X}^* \quad (126)$$

$$\frac{\partial \text{tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = \mathbf{X} \quad (127)$$

Since the function $\text{tr}(\mathbf{X}\mathbf{X}^H)$ is a real function of the complex matrix $\mathbf{X}$, the complex gradient matrix (81) is given by

$$\nabla \text{tr}(\mathbf{X}\mathbf{X}^H) = 2\frac{\partial \text{tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = 2\mathbf{X} \quad (128)$$

3.2.6 Derivatives of Determinants

Assume that all the arguments are real and the argument inside the derivative is square. Then the following formulas exists:

$$\frac{\partial \det(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial \mathbf{X}} = \det(\mathbf{A}\mathbf{X}\mathbf{B})(\mathbf{X}^T)^{-1} \quad (129)$$

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \det(\mathbf{X}^T)}{\partial \mathbf{X}} = \det(\mathbf{X})(\mathbf{X}^T)^{-1}, \qquad \text{see (144)} \quad (130)$$

$$\frac{\partial \det(\mathbf{X}^k)}{\partial \mathbf{X}} = k\det(\mathbf{X}^k)(\mathbf{X}^T)^{-1} \quad (131)$$

$$\frac{\partial \ln(\det(\mathbf{A}\mathbf{X}\mathbf{B}))}{\partial \mathbf{X}} = (\mathbf{X}^T)^{-1} \quad (132)$$

$$\frac{\partial \ln(\det(\mathbf{X}))}{\partial \mathbf{X}} = (\mathbf{X}^T)^{-1}, \qquad \text{see (145)} \quad (133)$$

$$\frac{\partial \ln(\det(\mathbf{X}^k))}{\partial \mathbf{X}} = k(\mathbf{X}^T)^{-1} \quad (134)$$

$$\frac{\partial \det(\mathbf{X}^T\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X}^T\mathbf{A}\mathbf{X})(\mathbf{A}\mathbf{X}(\mathbf{X}^T\mathbf{A}\mathbf{X})^{-1} + \mathbf{A}^T\mathbf{X}(\mathbf{X}^T\mathbf{A}^T\mathbf{X})^{-1}) \quad (135)$$

Provided $\mathbf{\Sigma}$ is symmetric:

$$\frac{\partial \det(\mathbf{X}^T\mathbf{\Sigma}\mathbf{X})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{\Sigma}\mathbf{X})\mathbf{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}\mathbf{X})^{-1} \quad (136)$$

$$\frac{\partial \ln(\det(\mathbf{X}^T\mathbf{\Sigma}\mathbf{X}))}{\partial \mathbf{X}} = 2\mathbf{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}\mathbf{X})^{-1} \quad (137)$$

Further, if $\mathbf{X}$ and $\mathbf{\Sigma}$ are square and non-singular:

$$\frac{\partial \det(\mathbf{X}^T\mathbf{\Sigma}\mathbf{X})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{\Sigma}\mathbf{X})(\mathbf{X}^T)^{-1} \quad (138)$$

3.2.7 Complex Derivative Involving Determinants

Here, a calculation example is provided. The objective is to find the derivative of $\det(\mathbf{X}^H\mathbf{A}\mathbf{X})$ with respect to $\mathbf{X} \in \mathbb{C}^{m\times n}$. The derivative is found with respect to the real part and the imaginary part of $\mathbf{X}$, by use of (70) and (66), $\det(\mathbf{X}^H\mathbf{A}\mathbf{X})$ can be rewritten as:

$$\begin{aligned}
\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X}) &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\text{tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H\mathbf{A}\mathbf{X})] \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\text{tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}(\partial(\mathbf{X}^H)\mathbf{A}\mathbf{X} + \mathbf{X}^H\partial(\mathbf{A}\mathbf{X}))] \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\big(\text{tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)\mathbf{A}\mathbf{X}] + \text{tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\partial(\mathbf{A}\mathbf{X})]\big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\big(\text{tr}[\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)] + \text{tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\partial(\mathbf{X})]\big)
\end{aligned}$$

First, the derivative is found with respect to the real part of $\mathbf{X}$

$$\begin{aligned}
\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re\mathbf{X}} &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\frac{\text{tr}[\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)]}{\partial \Re\mathbf{X}} + \frac{\text{tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\partial(\mathbf{X})]}{\partial \Re\mathbf{X}}\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\big(\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} + ((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A})^T\big)
\end{aligned}$$

Through the calculations, (102) and (120) were used. In addition, by use of (121), the derivative is found with respect to the imaginary part of $\mathbf{X}$

$$\begin{aligned}
i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im\mathbf{X}} &= i\det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\frac{\text{tr}[\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)]}{\partial \Im\mathbf{X}} + \frac{\text{tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\partial(\mathbf{X})]}{\partial \Im\mathbf{X}}\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\big(\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} - ((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A})^T\big)
\end{aligned}$$

Hence, derivative yields

$$\begin{aligned}
\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re\mathbf{X}} - i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im\mathbf{X}}\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\big((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\big)^T \quad (139)
\end{aligned}$$

and the complex conjugate derivative yields

$$\begin{aligned}
\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}^*} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re\mathbf{X}} + i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im\mathbf{X}}\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} \quad (140)
\end{aligned}$$

Notice, for real **X**, **A**, the sum of (139) and (140) is reduced to (135).
Similar calculations yield

$$\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \mathbf{X}} = \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Re\mathbf{X}} - i\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Im\mathbf{X}}\Big) = \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)\big(\mathbf{A}\mathbf{X}^H(\mathbf{X}\mathbf{A}\mathbf{X}^H)^{-1}\big)^T \tag{141}$$

and

$$\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \mathbf{X}^*} = \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Re\mathbf{X}} + i\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Im\mathbf{X}}\Big) = \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)(\mathbf{X}\mathbf{A}\mathbf{X}^H)^{-1}\mathbf{X}\mathbf{A} \tag{142}$$

3.2.8 Special Cases

If the elements of **X** functionally depend on each other, the derivative may yield a result that differs from the case, where the elements of **X** are functionally independent. If **X** is symmetric, then (Boik [2002], Stainvas [2002]):

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A}\circ\mathbf{I}), \text{ see (146)} \tag{143}$$

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X} - (\mathbf{X}\circ\mathbf{I}) \tag{144}$$

$$\frac{\partial \ln\det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - (\mathbf{X}^{-1}\circ\mathbf{I}) \tag{145}$$

If **X** is diagonal, then (Minka [2000]):

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}\circ\mathbf{I} \tag{146}$$

3.3 Other References

Other references not already mentioned, which contain useful information on matrix derivatives are Petersen [2004], Hyvärinen et al. [2001], Dyrholm [2004], Joho [2000], Cichocki and Amari [2002].

4 Correspondence

Please report any errors or suggestions to the author by email to: `msp(a)imm.dtu.dk`. This document will not be changed further. In the future, corrections will be applied into "*The Matrix Cookbook*" available from `http://www.imm.dtu.dk/pubdb/views/publication_details.php?id=3274` .

References

Jörn Anemüller, Terrence J. Sejnowski, and Scott Makeig. Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks*, 16(9):1311–1323, November 2003.

Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford, 1995.

Robert J. Boik. Lecture notes: Statistics 550. Online, April 22 2002. URL `http://www.math.montana.edu/~rjboik/classes/550/notes_550.pdf`. Notes.

D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proceedings*, 130(1):11–16, February 1983. PTS. F and H.

Mike Brookes. Matrix reference manual: Matrix calculus. www, December 2003. URL `http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html`. Online Notes.

Andrezej Cichocki and Shun-ichi Amari. *Adaptive Blind Signal and Image Processing*. Wiley, 2002.

Mads Dyrholm. Matrix results, 2004. URL `http://www.imm.dtu.dk/~mad/papers/madrix.pdf`.

Carlos A. Felippa. Appendix d: Matrix calculus. www, 2003. URL `caswww.colorado.edu/courses.d/IFEM.d/IFEM.AppD.d/IFEM.AppD.pdfV`. Online Notes.

Gene H. Golub and Charles F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, Upper Saddle River, NJ, 4th edition, 2002.

Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. Wiley, 2001.

Julius Orion Smith III. Block matrix decompositions. Online, 2003. URL `http://ccrma-www.stanford.edu/~jos/lattice/Block_matrix_decompositions.%html`.

Marcel Joho. *A Systematic Approach to Adaptive Algorithms for Multichannel System Identification, Inverse Modeling, and Blind Identification*. PhD thesis, Swiss Federal Institute of Technology, Zürich, December 2000. URL `http://www.phonak.uiuc.edu/~joho/research/mj_publications.html`.

Thomas P. Minka. Old and new matrix algebra useful for statistics, December 2000. Notes.

L. Parra and C. Spence. Convolutive blind separation of non-stationary sources. In *IEEE Transactions Speech and Audio Processing*, pages 320–327, May 2000.

K. B. Petersen. The matrix cookbook, 2004. URL `http://www.imm.dtu.dk/pubdb/p.php?3274`. Version 20040914.

William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 2002.

Sam Roweis. Matrix identities. `http://www.cs.toronto.edu/~roweis/notes/matrixid.pdf`, June 1999. Notes.

Laurent Schwartz. *Cours d'Analyse*, volume II. Hermann, Paris, 1967. As referenced in Haykin [2002].

Inna Stainvas. Matrix algebra in differential calculus. Neural Computing Research Group, Information Engeneering, Aston University, UK, August 2002. URL `www.ncrg.aston.ac.uk/~stainvai/probmod/calc.ps`. Notes.