The Matrix Cookbook

[http://matrixcookbook.com]

Kaare Brandt Petersen
Michael Syskind Pedersen

Version: November 15, 2012

Introduction

What is this? These pages are a collection of facts (identities, approximations, inequalities, relations, ...) about matrices and matters relating to them. It is collected in this form for the convenience of anyone who wants a quick desktop reference .

Disclaimer: The identities, approximations and relations presented here were obviously not invented but collected, borrowed and copied from a large amount of sources. These sources include similar but shorter notes found on the internet and appendices in books - see the references for a full list.

Errors: Very likely there are errors, typos, and mistakes for which we apologize and would be grateful to receive corrections at cookbook@2302.dk.

Its ongoing: The project of keeping a large repository of relations involving matrices is naturally ongoing and the version will be apparent from the date in the header.

Suggestions: Your suggestion for additional content or elaboration of some topics is most welcome acookbook@2302.dk.

Keywords: Matrix algebra, matrix relations, matrix identities, derivative of determinant, derivative of inverse matrix, differentiate a matrix.

Acknowledgements: We would like to thank the following for contributions and suggestions: Bill Baxter, Brian Templeton, Christian Rishøj, Christian Schröppel, Dan Boley, Douglas L. Theobald, Esben Hoegh-Rasmussen, Evripidis Karseras, Georg Martius, Glynne Casteel, Jan Larsen, Jun Bin Gao, Jürgen Struckmeier, Kamil Dedecius, Karim T. Abou-Moustafa, Korbinian Strimmer, Lars Christiansen, Lars Kai Hansen, Leland Wilkinson, Liguo He, Loic Thibaut, Markus Froeb, Michael Hubatka, Miguel Barão, Ole Winther, Pavel Sakov, Stephan Hattinger, Troels Pedersen, Vasile Sima, Vincent Rabaud, Zhaoshui He. We would also like thank The Oticon Foundation for funding our PhD studies.

Contents

Notation and Nomenclature

$\mathbf{A}$	Matrix
$\mathbf{A}_{ij}$	Matrix indexed for some purpose
$\mathbf{A}_i$	Matrix indexed for some purpose
$\mathbf{A}^{ij}$	Matrix indexed for some purpose
$\mathbf{A}^n$	Matrix indexed for some purpose **or**
	The n.th power of a square matrix
$\mathbf{A}^{-1}$	The inverse matrix of the matrix $\mathbf{A}$
$\mathbf{A}^+$	The pseudo inverse matrix of the matrix $\mathbf{A}$ (see Sec. 3.6)
$\mathbf{A}^{1/2}$	The square root of a matrix (if unique), not elementwise
$(\mathbf{A})_{ij}$	The (i,j).th entry of the matrix $\mathbf{A}$
A_{ij}	The (i,j).th entry of the matrix $\mathbf{A}$
$[\mathbf{A}]_{ij}$	The ij-submatrix, i.e. $\mathbf{A}$ with i.th row and j.th column deleted
$\mathbf{a}$	Vector (column-vector)
$\mathbf{a}_i$	Vector indexed for some purpose
a_i	The i.th element of the vector $\mathbf{a}$
a	Scalar
$\Re z$	Real part of a scalar
$\Re \mathbf{z}$	Real part of a vector
$\Re \mathbf{Z}$	Real part of a matrix
$\Im z$	Imaginary part of a scalar
$\Im \mathbf{z}$	Imaginary part of a vector
$\Im \mathbf{Z}$	Imaginary part of a matrix
$\det(\mathbf{A})$	Determinant of $\mathbf{A}$
$\mathrm{Tr}(\mathbf{A})$	Trace of the matrix $\mathbf{A}$
$\mathrm{diag}(\mathbf{A})$	Diagonal matrix of the matrix $\mathbf{A}$, i.e. $(\mathrm{diag}(\mathbf{A}))_{ij} = \delta_{ij} A_{ij}$
$\mathrm{eig}(\mathbf{A})$	Eigenvalues of the matrix $\mathbf{A}$
$\mathrm{vec}(\mathbf{A})$	The vector-version of the matrix $\mathbf{A}$ (see Sec. 10.2.2)
$\sup$	Supremum of a set
$\|\|\mathbf{A}\|\|$	Matrix norm (subscript if any denotes what norm)
$\mathbf{A}^T$	Transposed matrix
$\mathbf{A}^{-T}$	The inverse of the transposed and vice versa, $\mathbf{A}^{-T} = (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$.
$\mathbf{A}^*$	Complex conjugated matrix
$\mathbf{A}^H$	Transposed and complex conjugated matrix (Hermitian)
$\mathbf{A} \circ \mathbf{B}$	Hadamard (elementwise) product
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product
$\mathbf{0}$	The null matrix. Zero in all entries.
$\mathbf{I}$	The identity matrix
$\mathbf{J}^{ij}$	The single-entry matrix, 1 at (i,j) and zero elsewhere
$\mathbf{\Sigma}$	A positive definite matrix
$\mathbf{\Lambda}$	A diagonal matrix

1 Basics

$$\begin{aligned}
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} && (1)\\
(\mathbf{ABC}...)^{-1} &= ...\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} && (2)\\
(\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T && (3)\\
(\mathbf{A}+\mathbf{B})^T &= \mathbf{A}^T+\mathbf{B}^T && (4)\\
(\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T && (5)\\
(\mathbf{ABC}...)^T &= ...\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T && (6)\\
(\mathbf{A}^H)^{-1} &= (\mathbf{A}^{-1})^H && (7)\\
(\mathbf{A}+\mathbf{B})^H &= \mathbf{A}^H+\mathbf{B}^H && (8)\\
(\mathbf{AB})^H &= \mathbf{B}^H\mathbf{A}^H && (9)\\
(\mathbf{ABC}...)^H &= ...\mathbf{C}^H\mathbf{B}^H\mathbf{A}^H && (10)
\end{aligned}$$

1.1 Trace

$$\begin{aligned}
\mathrm{Tr}(\mathbf{A}) &= \textstyle\sum_i A_{ii} && (11)\\
\mathrm{Tr}(\mathbf{A}) &= \textstyle\sum_i \lambda_i, \qquad \lambda_i = \mathrm{eig}(\mathbf{A}) && (12)\\
\mathrm{Tr}(\mathbf{A}) &= \mathrm{Tr}(\mathbf{A}^T) && (13)\\
\mathrm{Tr}(\mathbf{AB}) &= \mathrm{Tr}(\mathbf{BA}) && (14)\\
\mathrm{Tr}(\mathbf{A}+\mathbf{B}) &= \mathrm{Tr}(\mathbf{A})+\mathrm{Tr}(\mathbf{B}) && (15)\\
\mathrm{Tr}(\mathbf{ABC}) &= \mathrm{Tr}(\mathbf{BCA}) = \mathrm{Tr}(\mathbf{CAB}) && (16)\\
\mathbf{a}^T\mathbf{a} &= \mathrm{Tr}(\mathbf{a}\mathbf{a}^T) && (17)
\end{aligned}$$

1.2 Determinant

Let $\mathbf{A}$ be an $n \times n$ matrix.

$$\begin{aligned}
\det(\mathbf{A}) &= \textstyle\prod_i \lambda_i \qquad \lambda_i = \mathrm{eig}(\mathbf{A}) && (18)\\
\det(c\mathbf{A}) &= c^n \det(\mathbf{A}), \qquad \text{if } \mathbf{A} \in \mathbb{R}^{n\times n} && (19)\\
\det(\mathbf{A}^T) &= \det(\mathbf{A}) && (20)\\
\det(\mathbf{AB}) &= \det(\mathbf{A})\det(\mathbf{B}) && (21)\\
\det(\mathbf{A}^{-1}) &= 1/\det(\mathbf{A}) && (22)\\
\det(\mathbf{A}^n) &= \det(\mathbf{A})^n && (23)\\
\det(\mathbf{I}+\mathbf{u}\mathbf{v}^T) &= 1+\mathbf{u}^T\mathbf{v} && (24)
\end{aligned}$$

For $n = 2$:

$$\det(\mathbf{I}+\mathbf{A}) = 1+\det(\mathbf{A})+\mathrm{Tr}(\mathbf{A}) \tag{25}$$

For $n = 3$:

$$\det(\mathbf{I}+\mathbf{A}) = 1+\det(\mathbf{A})+\mathrm{Tr}(\mathbf{A})+\frac{1}{2}\mathrm{Tr}(\mathbf{A})^2-\frac{1}{2}\mathrm{Tr}(\mathbf{A}^2) \tag{26}$$

For $n = 4$:

$$\begin{aligned}
\det(\mathbf{I}+\mathbf{A}) &= 1 + \det(\mathbf{A}) + \mathrm{Tr}(\mathbf{A}) + \frac{1}{2} \\
&\quad + \mathrm{Tr}(\mathbf{A})^2 - \frac{1}{2}\mathrm{Tr}(\mathbf{A}^2) \\
&\quad + \frac{1}{6}\mathrm{Tr}(\mathbf{A})^3 - \frac{1}{2}\mathrm{Tr}(\mathbf{A})\mathrm{Tr}(\mathbf{A}^2) + \frac{1}{3}\mathrm{Tr}(\mathbf{A}^3) \qquad (27)
\end{aligned}$$

For small ε, the following approximation holds

$$\det(\mathbf{I}+\varepsilon\mathbf{A}) \cong 1 + \det(\mathbf{A}) + \varepsilon\mathrm{Tr}(\mathbf{A}) + \frac{1}{2}\varepsilon^2\mathrm{Tr}(\mathbf{A})^2 - \frac{1}{2}\varepsilon^2\mathrm{Tr}(\mathbf{A}^2) \qquad (28)$$

1.3 The Special Case 2x2

Consider the matrix $\mathbf{A}$

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Determinant and trace

$$\det(\mathbf{A}) = A_{11}A_{22} - A_{12}A_{21} \qquad (29)$$

$$\mathrm{Tr}(\mathbf{A}) = A_{11} + A_{22} \qquad (30)$$

Eigenvalues

$$\lambda^2 - \lambda \cdot \mathrm{Tr}(\mathbf{A}) + \det(\mathbf{A}) = 0$$

$$\lambda_1 = \frac{\mathrm{Tr}(\mathbf{A}) + \sqrt{\mathrm{Tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2} \qquad \lambda_2 = \frac{\mathrm{Tr}(\mathbf{A}) - \sqrt{\mathrm{Tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2}$$

$$\lambda_1 + \lambda_2 = \mathrm{Tr}(\mathbf{A}) \qquad \lambda_1\lambda_2 = \det(\mathbf{A})$$

Eigenvectors

$$\mathbf{v}_1 \propto \begin{bmatrix} A_{12} \\ \lambda_1 - A_{11} \end{bmatrix} \qquad \mathbf{v}_2 \propto \begin{bmatrix} A_{12} \\ \lambda_2 - A_{11} \end{bmatrix}$$

Inverse

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix} \qquad (31)$$

2 Derivatives

This section is covering differentiation of a number of expressions with respect to a matrix $\mathbf{X}$. Note that it is always assumed that $\mathbf{X}$ has *no special structure*, i.e. that the elements of $\mathbf{X}$ are independent (e.g. not symmetric, Toeplitz, positive definite). See section 2.8 for differentiation of structured matrices. The basic assumptions can be written in a formula as

$$\frac{\partial X_{kl}}{\partial X_{ij}} = \delta_{ik}\delta_{lj} \tag{32}$$

that is for e.g. vector forms,

$$\left[\frac{\partial \mathbf{x}}{\partial y}\right]_i = \frac{\partial x_i}{\partial y} \qquad \left[\frac{\partial x}{\partial \mathbf{y}}\right]_i = \frac{\partial x}{\partial y_i} \qquad \left[\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right]_{ij} = \frac{\partial x_i}{\partial y_j}$$

The following rules are general and very useful when deriving the differential of an expression ([19]):

$$\begin{aligned}
\partial\mathbf{A} &= 0 \qquad (\mathbf{A} \text{ is a constant}) && (33)\\
\partial(\alpha\mathbf{X}) &= \alpha\partial\mathbf{X} && (34)\\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial\mathbf{X}+\partial\mathbf{Y} && (35)\\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial\mathbf{X}) && (36)\\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial\mathbf{X})\mathbf{Y}+\mathbf{X}(\partial\mathbf{Y}) && (37)\\
\partial(\mathbf{X}\circ\mathbf{Y}) &= (\partial\mathbf{X})\circ\mathbf{Y}+\mathbf{X}\circ(\partial\mathbf{Y}) && (38)\\
\partial(\mathbf{X}\otimes\mathbf{Y}) &= (\partial\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes(\partial\mathbf{Y}) && (39)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} && (40)\\
\partial(\det(\mathbf{X})) &= \mathrm{Tr}(\mathrm{adj}(\mathbf{X})\partial\mathbf{X}) && (41)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (42)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (43)\\
\partial\mathbf{X}^T &= (\partial\mathbf{X})^T && (44)\\
\partial\mathbf{X}^H &= (\partial\mathbf{X})^H && (45)
\end{aligned}$$

2.1 Derivatives of a Determinant

2.1.1 General form

$$\frac{\partial \det(\mathbf{Y})}{\partial x} = \det(\mathbf{Y})\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right] \tag{46}$$

$$\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} = \delta_{ij}\det(\mathbf{X}) \tag{47}$$

$$\begin{aligned}
\frac{\partial^2 \det(\mathbf{Y})}{\partial x^2} = \det(\mathbf{Y})\Bigg[&\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\frac{\partial\mathbf{Y}}{\partial x}}{\partial x}\right]\\
&+\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right]\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right]\\
&-\mathrm{Tr}\left[\left(\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right)\left(\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right)\right]\Bigg]
\end{aligned} \tag{48}$$

2.1.2 Linear forms

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(\mathbf{X}^{-1})^T \tag{49}$$

$$\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} = \delta_{ij} \det(\mathbf{X}) \tag{50}$$

$$\frac{\partial \det(\mathbf{AXB})}{\partial \mathbf{X}} = \det(\mathbf{AXB})(\mathbf{X}^{-1})^T = \det(\mathbf{AXB})(\mathbf{X}^T)^{-1} \tag{51}$$

2.1.3 Square forms

If $\mathbf{X}$ is square and invertible, then

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{X}^{-T} \tag{52}$$

If $\mathbf{X}$ is not square but $\mathbf{A}$ is symmetric, then

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} \tag{53}$$

If $\mathbf{X}$ is not square and $\mathbf{A}$ is not symmetric, then

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = \det(\mathbf{X}^T\mathbf{AX})(\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} + \mathbf{A}^T\mathbf{X}(\mathbf{X}^T\mathbf{A}^T\mathbf{X})^{-1}) \tag{54}$$

2.1.4 Other nonlinear forms

Some special cases are (See [9, 7])

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})|}{\partial \mathbf{X}} = 2(\mathbf{X}^+)^T \tag{55}$$

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})}{\partial \mathbf{X}^+} = -2\mathbf{X}^T \tag{56}$$

$$\frac{\partial \ln |\det(\mathbf{X})|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^T = (\mathbf{X}^T)^{-1} \tag{57}$$

$$\frac{\partial \det(\mathbf{X}^k)}{\partial \mathbf{X}} = k\det(\mathbf{X}^k)\mathbf{X}^{-T} \tag{58}$$

2.2 Derivatives of an Inverse

From [27] we have the basic identity

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x} = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\mathbf{Y}^{-1} \tag{59}$$

from which it follows

$$\begin{aligned}
\frac{\partial(\mathbf{X}^{-1})_{kl}}{\partial X_{ij}} &= -(\mathbf{X}^{-1})_{ki}(\mathbf{X}^{-1})_{jl} && (60)\\
\frac{\partial \mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial \mathbf{X}} &= -\mathbf{X}^{-T}\mathbf{a}\mathbf{b}^T\mathbf{X}^{-T} && (61)\\
\frac{\partial \det(\mathbf{X}^{-1})}{\partial \mathbf{X}} &= -\det(\mathbf{X}^{-1})(\mathbf{X}^{-1})^T && (62)\\
\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} &= -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T && (63)\\
\frac{\partial \mathrm{Tr}((\mathbf{X}+\mathbf{A})^{-1})}{\partial \mathbf{X}} &= -((\mathbf{X}+\mathbf{A})^{-1}(\mathbf{X}+\mathbf{A})^{-1})^T && (64)
\end{aligned}$$

From [32] we have the following result: Let $\mathbf{A}$ be an $n \times n$ invertible square matrix, $\mathbf{W}$ be the inverse of $\mathbf{A}$, and $J(\mathbf{A})$ is an $n \times n$ -variate and differentiable function with respect to $\mathbf{A}$, then the partial differentials of J with respect to $\mathbf{A}$ and $\mathbf{W}$ satisfy

$$\frac{\partial J}{\partial \mathbf{A}} = -\mathbf{A}^{-T}\frac{\partial J}{\partial \mathbf{W}}\mathbf{A}^{-T}$$

2.3 Derivatives of Eigenvalues

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\sum \mathrm{eig}(\mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} && (65)\\
\frac{\partial}{\partial \mathbf{X}}\prod \mathrm{eig}(\mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}}\det(\mathbf{X}) = \det(\mathbf{X})\mathbf{X}^{-T} && (66)
\end{aligned}$$

If $\mathbf{A}$ is real and symmetric, λ_i and $\mathbf{v}_i$ are distinct eigenvalues and eigenvectors of $\mathbf{A}$ (see (276)) with $\mathbf{v}_i^T\mathbf{v}_i = 1$, then [33]

$$\begin{aligned}
\partial\lambda_i &= \mathbf{v}_i^T\partial(\mathbf{A})\mathbf{v}_i && (67)\\
\partial\mathbf{v}_i &= (\lambda_i\mathbf{I}-\mathbf{A})^{+}\partial(\mathbf{A})\mathbf{v}_i && (68)
\end{aligned}$$

2.4 Derivatives of Matrices, Vectors and Scalar Forms

2.4.1 First Order

$$\begin{aligned}
\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} && (69)\\
\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} &= \mathbf{a}\mathbf{b}^T && (70)\\
\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} &= \mathbf{b}\mathbf{a}^T && (71)\\
\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} &= \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T && (72)\\
\frac{\partial \mathbf{X}}{\partial X_{ij}} &= \mathbf{J}^{ij} && (73)\\
\frac{\partial(\mathbf{X}\mathbf{A})_{ij}}{\partial X_{mn}} &= \delta_{im}(\mathbf{A})_{nj} = (\mathbf{J}^{mn}\mathbf{A})_{ij} && (74)\\
\frac{\partial(\mathbf{X}^T\mathbf{A})_{ij}}{\partial X_{mn}} &= \delta_{in}(\mathbf{A})_{mj} = (\mathbf{J}^{nm}\mathbf{A})_{ij} && (75)
\end{aligned}$$

2.4.2 Second Order

$$\frac{\partial}{\partial X_{ij}}\sum_{klmn} X_{kl}X_{mn} = 2\sum_{kl} X_{kl} \quad (76)$$

$$\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T) \quad (77)$$

$$\frac{\partial (\mathbf{B}\mathbf{x}+\mathbf{b})^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d}) + \mathbf{D}^T\mathbf{C}^T(\mathbf{B}\mathbf{x}+\mathbf{b}) \quad (78)$$

$$\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})_{kl}}{\partial X_{ij}} = \delta_{lj}(\mathbf{X}^T\mathbf{B})_{ki} + \delta_{kj}(\mathbf{B}\mathbf{X})_{il} \quad (79)$$

$$\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})}{\partial X_{ij}} = \mathbf{X}^T\mathbf{B}\mathbf{J}^{ij} + \mathbf{J}^{ji}\mathbf{B}\mathbf{X} \qquad (\mathbf{J}^{ij})_{kl} = \delta_{ik}\delta_{jl} \quad (80)$$

See Sec 9.7 for useful properties of the Single-entry matrix $\mathbf{J}^{ij}$

$$\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B}+\mathbf{B}^T)\mathbf{x} \quad (81)$$

$$\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} = \mathbf{D}^T\mathbf{X}\mathbf{b}\mathbf{c}^T + \mathbf{D}\mathbf{X}\mathbf{c}\mathbf{b}^T \quad (82)$$

$$\frac{\partial}{\partial \mathbf{X}}(\mathbf{X}\mathbf{b}+\mathbf{c})^T\mathbf{D}(\mathbf{X}\mathbf{b}+\mathbf{c}) = (\mathbf{D}+\mathbf{D}^T)(\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T \quad (83)$$

Assume $\mathbf{W}$ is symmetric, then

$$\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) = -2\mathbf{A}^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) \quad (84)$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s}) = 2\mathbf{W}(\mathbf{x}-\mathbf{s}) \quad (85)$$

$$\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s}) = -2\mathbf{W}(\mathbf{x}-\mathbf{s}) \quad (86)$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) = 2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) \quad (87)$$

$$\frac{\partial}{\partial \mathbf{A}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) = -2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})\mathbf{s}^T \quad (88)$$

As a case with complex values the following holds

$$\frac{\partial (a-\mathbf{x}^H\mathbf{b})^2}{\partial \mathbf{x}} = -2\mathbf{b}(a-\mathbf{x}^H\mathbf{b})^* \quad (89)$$

This formula is also known from the LMS algorithm [14]

2.4.3 Higher-order and non-linear

$$\frac{\partial (\mathbf{X}^n)_{kl}}{\partial X_{ij}} = \sum_{r=0}^{n-1}(\mathbf{X}^r\mathbf{J}^{ij}\mathbf{X}^{n-1-r})_{kl} \quad (90)$$

For proof of the above, see B.1.3.

$$\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^T\mathbf{X}^n\mathbf{b} = \sum_{r=0}^{n-1}(\mathbf{X}^r)^T\mathbf{a}\mathbf{b}^T(\mathbf{X}^{n-1-r})^T \quad (91)$$

$$\frac{\partial}{\partial \mathbf{X}} \mathbf{a}^T (\mathbf{X}^n)^T \mathbf{X}^n \mathbf{b} = \sum_{r=0}^{n-1} \Big[\mathbf{X}^{n-1-r} \mathbf{a}\mathbf{b}^T (\mathbf{X}^n)^T \mathbf{X}^r + (\mathbf{X}^r)^T \mathbf{X}^n \mathbf{a}\mathbf{b}^T (\mathbf{X}^{n-1-r})^T \Big] \quad (92)$$

See B.1.3 for a proof.
Assume $\mathbf{s}$ and $\mathbf{r}$ are functions of $\mathbf{x}$, i.e. $\mathbf{s} = \mathbf{s}(\mathbf{x}), \mathbf{r} = \mathbf{r}(\mathbf{x})$, and that $\mathbf{A}$ is a constant, then

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{s}^T \mathbf{A} \mathbf{r} = \left[\frac{\partial \mathbf{s}}{\partial \mathbf{x}} \right]^T \mathbf{A} \mathbf{r} + \left[\frac{\partial \mathbf{r}}{\partial \mathbf{x}} \right]^T \mathbf{A}^T \mathbf{s} \quad (93)$$

$$\frac{\partial}{\partial \mathbf{x}} \frac{(\mathbf{A}\mathbf{x})^T (\mathbf{A}\mathbf{x})}{(\mathbf{B}\mathbf{x})^T (\mathbf{B}\mathbf{x})} = \frac{\partial}{\partial \mathbf{x}} \frac{\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{B}^T \mathbf{B} \mathbf{x}} \quad (94)$$

$$= 2 \frac{\mathbf{A}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{B} \mathbf{B} \mathbf{x}} - 2 \frac{\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} \mathbf{B}^T \mathbf{B} \mathbf{x}}{(\mathbf{x}^T \mathbf{B}^T \mathbf{B} \mathbf{x})^2} \quad (95)$$

2.4.4 Gradient and Hessian

Using the above we have for the gradient and the Hessian

$$f = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} \quad (96)$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} + \mathbf{b} \quad (97)$$

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^T} = \mathbf{A} + \mathbf{A}^T \quad (98)$$

2.5 Derivatives of Traces

Assume $F(\mathbf{X})$ to be a differentiable function of each of the elements of X. It then holds that

$$\frac{\partial \mathrm{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$$

where $f(\cdot)$ is the scalar derivative of $F(\cdot)$.

2.5.1 First Order

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}) = \mathbf{I} \quad (99)$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}\mathbf{A}) = \mathbf{A}^T \quad (100)$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A}^T \mathbf{B}^T \quad (101)$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}) = \mathbf{B}\mathbf{A} \quad (102)$$

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

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^T) = \mathbf{A} \quad (104)$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A} \otimes \mathbf{X}) = \mathrm{Tr}(\mathbf{A})\mathbf{I} \quad (105)$$

2.5.2 Second Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2) = 2\mathbf{X}^T \quad (106)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2\mathbf{B}) = (\mathbf{X}\mathbf{B}+\mathbf{B}\mathbf{X})^T \quad (107)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{B}\mathbf{X}) = \mathbf{B}\mathbf{X}+\mathbf{B}^T\mathbf{X} \quad (108)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}\mathbf{X}\mathbf{X}^T) = \mathbf{B}\mathbf{X}+\mathbf{B}^T\mathbf{X} \quad (109)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{X}^T\mathbf{B}) = \mathbf{B}\mathbf{X}+\mathbf{B}^T\mathbf{X} \quad (110)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{B}\mathbf{X}^T) = \mathbf{X}\mathbf{B}^T+\mathbf{X}\mathbf{B} \quad (111)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}\mathbf{X}^T\mathbf{X}) = \mathbf{X}\mathbf{B}^T+\mathbf{X}\mathbf{B} \quad (112)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{X}\mathbf{B}) = \mathbf{X}\mathbf{B}^T+\mathbf{X}\mathbf{B} \quad (113)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X}) = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T+\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T \quad (114)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{X}^T) = 2\mathbf{X} \quad (115)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{B}) = \mathbf{C}^T\mathbf{X}\mathbf{B}\mathbf{B}^T+\mathbf{C}\mathbf{X}\mathbf{B}\mathbf{B}^T \quad (116)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{X}^T\mathbf{B}\mathbf{X}\mathbf{C}\right] = \mathbf{B}\mathbf{X}\mathbf{C}+\mathbf{B}^T\mathbf{X}\mathbf{C}^T \quad (117)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{C}) = \mathbf{A}^T\mathbf{C}^T\mathbf{X}\mathbf{B}^T+\mathbf{C}\mathbf{A}\mathbf{X}\mathbf{B} \quad (118)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{A}\mathbf{X}\mathbf{B}+\mathbf{C})(\mathbf{A}\mathbf{X}\mathbf{B}+\mathbf{C})^T\Big] = 2\mathbf{A}^T(\mathbf{A}\mathbf{X}\mathbf{B}+\mathbf{C})\mathbf{B}^T \quad (119)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\otimes\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X})\mathrm{Tr}(\mathbf{X}) = 2\mathrm{Tr}(\mathbf{X})\mathbf{I} \quad (120)$$

See [7].

2.5.3 Higher Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^k) = k(\mathbf{X}^{k-1})^T \quad (121)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^k) = \sum_{r=0}^{k-1}(\mathbf{X}^r\mathbf{A}\mathbf{X}^{k-r-1})^T \quad (122)$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{B}^T\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{B}\right] = \;& \mathbf{C}\mathbf{X}\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{B}\mathbf{B}^T \\ & +\mathbf{C}^T\mathbf{X}\mathbf{B}\mathbf{B}^T\mathbf{X}^T\mathbf{C}^T\mathbf{X} \\ & +\mathbf{C}\mathbf{X}\mathbf{B}\mathbf{B}^T\mathbf{X}^T\mathbf{C}\mathbf{X} \\ & +\mathbf{C}^T\mathbf{X}\mathbf{X}^T\mathbf{C}^T\mathbf{X}\mathbf{B}\mathbf{B}^T\end{aligned} \quad (123)$$

2.5.4 Other

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}) = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T} \tag{124}$$

Assume $\mathbf{B}$ and $\mathbf{C}$ to be symmetric, then

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{A}\Big] = -(\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1})(\mathbf{A}+\mathbf{A}^T)(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{125}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] = -2\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} + 2\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{126}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] = -2\mathbf{C}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} + 2\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{127}$$

See [7].

$$\frac{\partial \mathrm{Tr}(\sin(\mathbf{X}))}{\partial \mathbf{X}} = \cos(\mathbf{X})^T \tag{128}$$

2.6 Derivatives of vector norms

2.6.1 Two-norm

$$\frac{\partial}{\partial \mathbf{x}}||\mathbf{x}-\mathbf{a}||_2 = \frac{\mathbf{x}-\mathbf{a}}{||\mathbf{x}-\mathbf{a}||_2} \tag{129}$$

$$\frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} = \frac{\mathbf{I}}{\|\mathbf{x}-\mathbf{a}\|_2} - \frac{(\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{a})^T}{\|\mathbf{x}-\mathbf{a}\|_2^3} \tag{130}$$

$$\frac{\partial||\mathbf{x}||_2^2}{\partial \mathbf{x}} = \frac{\partial||\mathbf{x}^T\mathbf{x}||_2}{\partial \mathbf{x}} = 2\mathbf{x} \tag{131}$$

2.7 Derivatives of matrix norms

For more on matrix norms, see Sec. 10.4.

2.7.1 Frobenius norm

$$\frac{\partial}{\partial \mathbf{X}}||\mathbf{X}||_{\mathrm{F}}^2 = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = 2\mathbf{X} \tag{132}$$

See (248). Note that this is also a special case of the result in equation 119.

2.8 Derivatives of Structured Matrices

Assume that the matrix $\mathbf{A}$ has some structure, i.e. symmetric, toeplitz, etc. In that case the derivatives of the previous section does not apply in general. Instead, consider the following general rule for differentiating a scalar function $f(\mathbf{A})$

$$\frac{df}{dA_{ij}} = \sum_{kl}\frac{\partial f}{\partial A_{kl}}\frac{\partial A_{kl}}{\partial A_{ij}} = \mathrm{Tr}\left[\left[\frac{\partial f}{\partial \mathbf{A}}\right]^T\frac{\partial \mathbf{A}}{\partial A_{ij}}\right] \tag{133}$$

The matrix differentiated with respect to itself is in this document referred to as the *structure matrix* of $\mathbf{A}$ and is defined simply by

$$\frac{\partial \mathbf{A}}{\partial A_{ij}} = \mathbf{S}^{ij} \tag{134}$$

If $\mathbf{A}$ has no special structure we have simply $\mathbf{S}^{ij} = \mathbf{J}^{ij}$, that is, the structure matrix is simply the single-entry matrix. Many structures have a representation in singleentry matrices, see Sec. 9.7.6 for more examples of structure matrices.

2.8.1 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 $\mathbf{X}$:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}} \tag{135}$$

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

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M} \sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}} \tag{136}$$

Using matrix notation, this can be written as:

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \mathrm{Tr}\Big[(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \frac{\partial \mathbf{U}}{\partial X_{ij}} \Big]. \tag{137}$$

2.8.2 Symmetric

If $\mathbf{A}$ is symmetric, then $\mathbf{S}^{ij} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij}\mathbf{J}^{ij}$ and therefore

$$\frac{df}{d\mathbf{A}} = \left[\frac{\partial f}{\partial \mathbf{A}}\right] + \left[\frac{\partial f}{\partial \mathbf{A}}\right]^T - \mathrm{diag}\left[\frac{\partial f}{\partial \mathbf{A}}\right] \tag{138}$$

That is, e.g., ([5]):

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A} \circ \mathbf{I}), \text{ see (142)} \tag{139}$$

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})) \tag{140}$$

$$\frac{\partial \ln \det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I}) \tag{141}$$

2.8.3 Diagonal

If $\mathbf{X}$ is diagonal, then ([19]):

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A} \circ \mathbf{I} \tag{142}$$

2.8.4 Toeplitz

Like symmetric matrices and diagonal matrices also Toeplitz matrices has a special structure which should be taken into account when the derivative with respect to a matrix with Toeplitz structure.

$$\begin{aligned}
&\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} \qquad (143)\\
=\;& \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}}\\
=\;& \begin{bmatrix}
\mathrm{Tr}(\mathbf{A}) & \mathrm{Tr}([\mathbf{A}^T]_{n1}) & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) & \cdots & A_{n1} \\
\mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A}) & \ddots & \ddots & \vdots \\
\mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \ddots & \ddots & \ddots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) \\
\vdots & \ddots & \ddots & \ddots & \mathrm{Tr}([\mathbf{A}^T]_{n1}) \\
A_{1n} & \cdots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A})
\end{bmatrix}\\
\equiv\;& \boldsymbol{\alpha}(\mathbf{A})
\end{aligned}$$

As it can be seen, the derivative $\boldsymbol{\alpha}(\mathbf{A})$ also has a Toeplitz structure. Each value in the diagonal is the sum of all the diagonal valued in $\mathbf{A}$, the values in the diagonals next to the main diagonal equal the sum of the diagonal next to the main diagonal in $\mathbf{A}^T$. This result is only valid for the unconstrained Toeplitz matrix. If the Toeplitz matrix also is symmetric, the same derivative yields

$$\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} = \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} = \boldsymbol{\alpha}(\mathbf{A}) + \boldsymbol{\alpha}(\mathbf{A})^T - \boldsymbol{\alpha}(\mathbf{A}) \circ \mathbf{I} \tag{144}$$

3 Inverses

3.1 Basic

3.1.1 Definition

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

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}, \tag{145}$$

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* (see e.g. [12]).

3.1.2 Cofactors and Adjoint

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 (i,j) *cofactor* of a matrix is defined as

$$\text{cof}(\mathbf{A}, i, j) = (-1)^{i+j}\det([\mathbf{A}]_{ij}), \tag{146}$$

The *matrix of cofactors* can be created from the cofactors

$$\text{cof}(\mathbf{A}) = \begin{bmatrix} \text{cof}(\mathbf{A},1,1) & \cdots & \text{cof}(\mathbf{A},1,n) \\ \vdots & \text{cof}(\mathbf{A},i,j) & \vdots \\ \text{cof}(\mathbf{A},n,1) & \cdots & \text{cof}(\mathbf{A},n,n) \end{bmatrix} \tag{147}$$

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

$$\text{adj}(\mathbf{A}) = (\text{cof}(\mathbf{A}))^T, \tag{148}$$

3.1.3 Determinant

The *determinant* of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is defined as (see [12])

$$\det(\mathbf{A}) = \sum_{j=1}^{n}(-1)^{j+1}A_{1j}\det\left([\mathbf{A}]_{1j}\right) \tag{149}$$

$$= \sum_{j=1}^{n} A_{1j}\text{cof}(\mathbf{A},1,j). \tag{150}$$

3.1.4 Construction

The inverse matrix can be constructed, using the adjoint matrix, by

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\cdot\text{adj}(\mathbf{A}) \tag{151}$$

For the case of 2×2 matrices, see section 1.3.

3.1.5 Condition number

The condition number of a matrix $c(\mathbf{A})$ is the ratio between the largest and the smallest singular value of a matrix (see Section 5.3 on singular values),

$$c(\mathbf{A}) = \frac{d_+}{d_-} \tag{152}$$

The condition number can be used to measure how singular a matrix is. If the condition number is large, it indicates that the matrix is nearly singular. The condition number can also be estimated from the matrix norms. Here

$$c(\mathbf{A}) = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\|, \tag{153}$$

where $\|\cdot\|$ is a norm such as e.g the 1-norm, the 2-norm, the ∞-norm or the Frobenius norm (see Sec 10.4 for more on matrix norms).

The 2-norm of $\mathbf{A}$ equals $\sqrt{(\max(\text{eig}(\mathbf{A}^H\mathbf{A})))}$ [12, p.57]. For a symmetric matrix, this reduces to $||\mathbf{A}||_2 = \max(|\text{eig}(\mathbf{A})|)$ [12, p.394]. If the matrix is symmetric and positive definite, $||\mathbf{A}||_2 = \max(\text{eig}(\mathbf{A}))$. The condition number based on the 2-norm thus reduces to

$$\|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2 = \max(\text{eig}(\mathbf{A}))\max(\text{eig}(\mathbf{A}^{-1})) = \frac{\max(\text{eig}(\mathbf{A}))}{\min(\text{eig}(\mathbf{A}))}. \tag{154}$$

3.2 Exact Relations

3.2.1 Basic

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{155}$$

3.2.2 The Woodbury identity

The Woodbury identity comes in many variants. The latter of the two can be found in [12]

$$(\mathbf{A} + \mathbf{CBC}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{A}^{-1} \tag{156}$$
$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} \tag{157}$$

If $\mathbf{P}, \mathbf{R}$ are positive definite, then (see [30])

$$(\mathbf{P}^{-1} + \mathbf{B}^T\mathbf{R}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{R}^{-1} = \mathbf{PB}^T(\mathbf{BPB}^T + \mathbf{R})^{-1} \tag{158}$$

3.2.3 The Kailath Variant

$$(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \tag{159}$$

See [4, page 153].

3.2.4 Sherman-Morrison

$$(\mathbf{A} + \mathbf{bc}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{bc}^T\mathbf{A}^{-1}}{1 + \mathbf{c}^T\mathbf{A}^{-1}\mathbf{b}} \tag{160}$$

3.2.5 The Searle Set of Identities

The following set of identities, can be found in [25, page 151],

$$\begin{aligned}
(\mathbf{I}+\mathbf{A}^{-1})^{-1} &= \mathbf{A}(\mathbf{A}+\mathbf{I})^{-1} && (161)\\
(\mathbf{A}+\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B} &= \mathbf{A}^{-1}\mathbf{B}(\mathbf{I}+\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1} && (162)\\
(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1} &= \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A} && (163)\\
\mathbf{A}-\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A} &= \mathbf{B}-\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} && (164)\\
\mathbf{A}^{-1}+\mathbf{B}^{-1} &= \mathbf{A}^{-1}(\mathbf{A}+\mathbf{B})\mathbf{B}^{-1} && (165)\\
(\mathbf{I}+\mathbf{A}\mathbf{B})^{-1} &= \mathbf{I}-\mathbf{A}(\mathbf{I}+\mathbf{B}\mathbf{A})^{-1}\mathbf{B} && (166)\\
(\mathbf{I}+\mathbf{A}\mathbf{B})^{-1}\mathbf{A} &= \mathbf{A}(\mathbf{I}+\mathbf{B}\mathbf{A})^{-1} && (167)
\end{aligned}$$

3.2.6 Rank-1 update of inverse of inner product

Denote $\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}$ and that $\mathbf{X}$ is extended to include a new column vector in the end $\tilde{\mathbf{X}} = [\mathbf{X} \quad \mathbf{v}]$. Then [34]

$$(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1} = \begin{bmatrix} \mathbf{A} + \frac{\mathbf{A}\mathbf{X}^T\mathbf{v}\mathbf{v}^T\mathbf{X}\mathbf{A}^T}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} & \frac{-\mathbf{A}\mathbf{X}^T\mathbf{v}}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} \\ \frac{-\mathbf{v}^T\mathbf{X}\mathbf{A}^T}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} & \frac{1}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} \end{bmatrix}$$

3.2.7 Rank-1 update of Moore-Penrose Inverse

The following is a rank-1 update for the Moore-Penrose pseudo-inverse of real valued matrices and proof can be found in [18]. The matrix $\mathbf{G}$ is defined below:

$$(\mathbf{A}+\mathbf{c}\mathbf{d}^T)^+ = \mathbf{A}^+ + \mathbf{G} \qquad (168)$$

Using the the notation

$$\begin{aligned}
\beta &= 1+\mathbf{d}^T\mathbf{A}^+\mathbf{c} && (169)\\
\mathbf{v} &= \mathbf{A}^+\mathbf{c} && (170)\\
\mathbf{n} &= (\mathbf{A}^+)^T\mathbf{d} && (171)\\
\mathbf{w} &= (\mathbf{I}-\mathbf{A}\mathbf{A}^+)\mathbf{c} && (172)\\
\mathbf{m} &= (\mathbf{I}-\mathbf{A}^+\mathbf{A})^T\mathbf{d} && (173)
\end{aligned}$$

the solution is given as six different cases, depending on the entities $||\mathbf{w}||$, $||\mathbf{m}||$, and β. Please note, that for any (column) vector $\mathbf{v}$ it holds that $\mathbf{v}^+ = \mathbf{v}^T(\mathbf{v}^T\mathbf{v})^{-1} = \frac{\mathbf{v}^T}{||\mathbf{v}||^2}$. The solution is:

Case 1 of 6: If $||\mathbf{w}|| \neq 0$ and $||\mathbf{m}|| \neq 0$. Then

$$\begin{aligned}
\mathbf{G} &= -\mathbf{v}\mathbf{w}^+ - (\mathbf{m}^+)^T\mathbf{n}^T + \beta(\mathbf{m}^+)^T\mathbf{w}^+ && (174)\\
&= -\frac{1}{||\mathbf{w}||^2}\mathbf{v}\mathbf{w}^T - \frac{1}{||\mathbf{m}||^2}\mathbf{m}\mathbf{n}^T + \frac{\beta}{||\mathbf{m}||^2||\mathbf{w}||^2}\mathbf{m}\mathbf{w}^T && (175)
\end{aligned}$$

Case 2 of 6: If $||\mathbf{w}|| = 0$ and $||\mathbf{m}|| \neq 0$ and $\beta = 0$. Then

$$\begin{aligned}
\mathbf{G} &= -\mathbf{v}\mathbf{v}^+\mathbf{A}^+ - (\mathbf{m}^+)^T\mathbf{n}^T && (176)\\
&= -\frac{1}{||\mathbf{v}||^2}\mathbf{v}\mathbf{v}^T\mathbf{A}^+ - \frac{1}{||\mathbf{m}||^2}\mathbf{m}\mathbf{n}^T && (177)
\end{aligned}$$

Case 3 of 6: If $||\mathbf{w}|| = 0$ and $\beta \neq 0$. Then

$$\mathbf{G} = \frac{1}{\beta}\mathbf{m}\mathbf{v}^T\mathbf{A}^+ - \frac{\beta}{||\mathbf{v}||^2||\mathbf{m}||^2 + |\beta|^2}\left(\frac{||\mathbf{v}||^2}{\beta}\mathbf{m} + \mathbf{v}\right)\left(\frac{||\mathbf{m}||^2}{\beta}(\mathbf{A}^+)^T\mathbf{v} + \mathbf{n}\right)^T \tag{178}$$

Case 4 of 6: If $||\mathbf{w}|| \neq 0$ and $||\mathbf{m}|| = 0$ and $\beta = 0$. Then

$$\begin{aligned} \mathbf{G} &= -\mathbf{A}^+\mathbf{n}\mathbf{n}^+ - \mathbf{v}\mathbf{w}^+ && (179) \\ &= -\frac{1}{||\mathbf{n}||^2}\mathbf{A}^+\mathbf{n}\mathbf{n}^T - \frac{1}{||\mathbf{w}||^2}\mathbf{v}\mathbf{w}^T && (180) \end{aligned}$$

Case 5 of 6: If $||\mathbf{m}|| = 0$ and $\beta \neq 0$. Then

$$\mathbf{G} = \frac{1}{\beta}\mathbf{A}^+\mathbf{n}\mathbf{w}^T - \frac{\beta}{||\mathbf{n}||^2||\mathbf{w}||^2 + |\beta|^2}\left(\frac{||\mathbf{w}||^2}{\beta}\mathbf{A}^+\mathbf{n} + \mathbf{v}\right)\left(\frac{||\mathbf{n}||^2}{\beta}\mathbf{w} + \mathbf{n}\right)^T \tag{181}$$

Case 6 of 6: If $||\mathbf{w}|| = 0$ and $||\mathbf{m}|| = 0$ and $\beta = 0$. Then

$$\begin{aligned} \mathbf{G} &= -\mathbf{v}\mathbf{v}^+\mathbf{A}^+ - \mathbf{A}^+\mathbf{n}\mathbf{n}^+ + \mathbf{v}^+\mathbf{A}^+\mathbf{n}\mathbf{v}\mathbf{n}^+ && (182) \\ &= -\frac{1}{||\mathbf{v}||^2}\mathbf{v}\mathbf{v}^T\mathbf{A}^+ - \frac{1}{||\mathbf{n}||^2}\mathbf{A}^+\mathbf{n}\mathbf{n}^T + \frac{\mathbf{v}^T\mathbf{A}^+\mathbf{n}}{||\mathbf{v}||^2||\mathbf{n}||^2}\mathbf{v}\mathbf{n}^T && (183) \end{aligned}$$

3.3 Implication on Inverses

$$\text{If} \qquad (\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1} \qquad \text{then} \qquad \mathbf{A}\mathbf{B}^{-1}\mathbf{A} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B} \tag{184}$$

See [25].

3.3.1 A PosDef identity

Assume $\mathbf{P}, \mathbf{R}$ to be positive definite and invertible, then

$$(\mathbf{P}^{-1} + \mathbf{B}^T\mathbf{R}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^T(\mathbf{B}\mathbf{P}\mathbf{B}^T + \mathbf{R})^{-1} \tag{185}$$

See [30].

3.4 Approximations

The following identity is known as the *Neuman series* of a matrix, which holds when $|\lambda_i| < 1$ for all eigenvalues λ_i

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{n=0}^{\infty}\mathbf{A}^n \tag{186}$$

which is equivalent to

$$(\mathbf{I} + \mathbf{A})^{-1} = \sum_{n=0}^{\infty}(-1)^n\mathbf{A}^n \tag{187}$$

When $|\lambda_i| < 1$ for all eigenvalues λ_i, it holds that $\mathbf{A} \to 0$ for $n \to \infty$, and the following approximations holds

$$\begin{aligned} (\mathbf{I} - \mathbf{A})^{-1} &\cong \mathbf{I} + \mathbf{A} + \mathbf{A}^2 && (188) \\ (\mathbf{I} + \mathbf{A})^{-1} &\cong \mathbf{I} - \mathbf{A} + \mathbf{A}^2 && (189) \end{aligned}$$

The following approximation is from [22] and holds when $\mathbf{A}$ large and symmetric

$$\mathbf{A} - \mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}\mathbf{A} \cong \mathbf{I} - \mathbf{A}^{-1} \tag{190}$$

If σ^2 is small compared to $\mathbf{Q}$ and $\mathbf{M}$ then

$$(\mathbf{Q} + \sigma^2\mathbf{M})^{-1} \cong \mathbf{Q}^{-1} - \sigma^2\mathbf{Q}^{-1}\mathbf{M}\mathbf{Q}^{-1} \tag{191}$$

Proof:

$$\begin{aligned}
(\mathbf{Q} + \sigma^2\mathbf{M})^{-1} &= &(192)\\
(\mathbf{Q}\mathbf{Q}^{-1}\mathbf{Q} + \sigma^2\mathbf{M}\mathbf{Q}^{-1}\mathbf{Q})^{-1} &= &(193)\\
((\mathbf{I} + \sigma^2\mathbf{M}\mathbf{Q}^{-1})\mathbf{Q})^{-1} &= &(194)\\
\mathbf{Q}^{-1}(\mathbf{I} + \sigma^2\mathbf{M}\mathbf{Q}^{-1})^{-1} & &(195)
\end{aligned}$$

This can be rewritten using the Taylor expansion:

$$\begin{aligned}
\mathbf{Q}^{-1}(\mathbf{I} + \sigma^2\mathbf{M}\mathbf{Q}^{-1})^{-1} &= &(196)\\
\mathbf{Q}^{-1}(\mathbf{I} - \sigma^2\mathbf{M}\mathbf{Q}^{-1} + (\sigma^2\mathbf{M}\mathbf{Q}^{-1})^2 - ...) &\cong \mathbf{Q}^{-1} - \sigma^2\mathbf{Q}^{-1}\mathbf{M}\mathbf{Q}^{-1} &(197)
\end{aligned}$$

3.5 Generalized Inverse

3.5.1 Definition

A generalized inverse matrix of the matrix $\mathbf{A}$ is any matrix $\mathbf{A}^-$ such that (see [26])

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A} \tag{198}$$

The matrix $\mathbf{A}^-$ is not unique.

3.6 Pseudo Inverse

3.6.1 Definition

The pseudo inverse (or Moore-Penrose inverse) of a matrix $\mathbf{A}$ is the matrix $\mathbf{A}^+$ that fulfils

$$\begin{array}{ll}
\text{I} & \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}\\
\text{II} & \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+\\
\text{III} & \mathbf{A}\mathbf{A}^+ \text{ symmetric}\\
\text{IV} & \mathbf{A}^+\mathbf{A} \text{ symmetric}
\end{array}$$

The matrix $\mathbf{A}^+$ is unique and does always exist. Note that in case of complex matrices, the symmetric condition is substituted by a condition of being Hermitian.

3.6.2 Properties

Assume $\mathbf{A}^+$ to be the pseudo-inverse of $\mathbf{A}$, then (See [3] for some of them)

$$\begin{aligned}
(\mathbf{A}^+)^+ &= \mathbf{A} && (199)\\
(\mathbf{A}^T)^+ &= (\mathbf{A}^+)^T && (200)\\
(\mathbf{A}^H)^+ &= (\mathbf{A}^+)^H && (201)\\
(\mathbf{A}^*)^+ &= (A^+)^* && (202)\\
(\mathbf{A}^+\mathbf{A})\mathbf{A}^H &= \mathbf{A}^H && (203)\\
(\mathbf{A}^+\mathbf{A})\mathbf{A}^T &\neq \mathbf{A}^T && (204)\\
(c\mathbf{A})^+ &= (1/c)\mathbf{A}^+ && (205)\\
\mathbf{A}^+ &= (\mathbf{A}^T\mathbf{A})^+\mathbf{A}^T && (206)\\
\mathbf{A}^+ &= \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ && (207)\\
(\mathbf{A}^T\mathbf{A})^+ &= \mathbf{A}^+(\mathbf{A}^T)^+ && (208)\\
(\mathbf{A}\mathbf{A}^T)^+ &= (\mathbf{A}^T)^+\mathbf{A}^+ && (209)\\
\mathbf{A}^+ &= (\mathbf{A}^H\mathbf{A})^+\mathbf{A}^H && (210)\\
\mathbf{A}^+ &= \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^+ && (211)\\
(\mathbf{A}^H\mathbf{A})^+ &= \mathbf{A}^+(\mathbf{A}^H)^+ && (212)\\
(\mathbf{A}\mathbf{A}^H)^+ &= (\mathbf{A}^H)^+\mathbf{A}^+ && (213)\\
(\mathbf{A}\mathbf{B})^+ &= (\mathbf{A}^+\mathbf{A}\mathbf{B})^+(\mathbf{A}\mathbf{B}\mathbf{B}^+)^+ && (214)\\
f(\mathbf{A}^H\mathbf{A}) - f(0)\mathbf{I} &= \mathbf{A}^+[f(\mathbf{A}\mathbf{A}^H) - f(0)\mathbf{I}]\mathbf{A} && (215)\\
f(\mathbf{A}\mathbf{A}^H) - f(0)\mathbf{I} &= \mathbf{A}[f(\mathbf{A}^H\mathbf{A}) - f(0)\mathbf{I}]\mathbf{A}^+ && (216)
\end{aligned}$$

where $\mathbf{A} \in \mathbb{C}^{n\times m}$.

Assume $\mathbf{A}$ to have full rank, then

$$\begin{aligned}
(\mathbf{A}\mathbf{A}^+)(\mathbf{A}\mathbf{A}^+) &= \mathbf{A}\mathbf{A}^+ && (217)\\
(\mathbf{A}^+\mathbf{A})(\mathbf{A}^+\mathbf{A}) &= \mathbf{A}^+\mathbf{A} && (218)\\
\mathrm{Tr}(\mathbf{A}\mathbf{A}^+) &= \mathrm{rank}(\mathbf{A}\mathbf{A}^+) \quad \text{(See [26])} && (219)\\
\mathrm{Tr}(\mathbf{A}^+\mathbf{A}) &= \mathrm{rank}(\mathbf{A}^+\mathbf{A}) \quad \text{(See [26])} && (220)
\end{aligned}$$

For two matrices it hold that

$$\begin{aligned}
(\mathbf{A}\mathbf{B})^+ &= (\mathbf{A}^+\mathbf{A}\mathbf{B})^+(\mathbf{A}\mathbf{B}\mathbf{B}^+)^+ && (221)\\
(\mathbf{A}\otimes\mathbf{B})^+ &= \mathbf{A}^+\otimes\mathbf{B}^+ && (222)
\end{aligned}$$

3.6.3 Construction

Assume that $\mathbf{A}$ has full rank, then

$\mathbf{A}$ $n\times n$	Square	$\mathrm{rank}(\mathbf{A}) = n$	$\Rightarrow$	$\mathbf{A}^+ = \mathbf{A}^{-1}$
$\mathbf{A}$ $n\times m$	Broad	$\mathrm{rank}(\mathbf{A}) = n$	$\Rightarrow$	$\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$
$\mathbf{A}$ $n\times m$	Tall	$\mathrm{rank}(\mathbf{A}) = m$	$\Rightarrow$	$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$

The so-called "broad version" is also known as *right inverse* and the "tall version" as the *left inverse*.

Assume $\mathbf{A}$ does not have full rank, i.e. $\mathbf{A}$ is $n \times m$ and $\text{rank}(\mathbf{A}) = r < \min(n, m)$. The pseudo inverse $\mathbf{A}^+$ can be constructed from the singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, by

$$\mathbf{A}^+ = \mathbf{V}_r \mathbf{D}_r^{-1} \mathbf{U}_r^T \tag{223}$$

where $\mathbf{U}_r, \mathbf{D}_r$, and $\mathbf{V}_r$ are the matrices with the degenerated rows and columns deleted. A different way is this: There do always exist two matrices $\mathbf{C}$ $n \times r$ and $\mathbf{D}$ $r \times m$ of rank r, such that $\mathbf{A} = \mathbf{C}\mathbf{D}$. Using these matrices it holds that

$$\mathbf{A}^+ = \mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T \tag{224}$$

See [3].

4 Complex Matrices

The complex scalar product $r = pq$ can be written as

$$\begin{bmatrix} \Re r \\ \Im r \end{bmatrix} = \begin{bmatrix} \Re p & -\Im p \\ \Im p & \Re p \end{bmatrix} \begin{bmatrix} \Re q \\ \Im q \end{bmatrix} \tag{225}$$

4.1 Complex Derivatives

In order to differentiate an expression $f(z)$ with respect to a complex z, the Cauchy-Riemann equations have to be satisfied ([7]):

$$\frac{df(z)}{dz} = \frac{\partial \Re(f(z))}{\partial \Re z} + i \frac{\partial \Im(f(z))}{\partial \Re z} \tag{226}$$

and

$$\frac{df(z)}{dz} = -i \frac{\partial \Re(f(z))}{\partial \Im z} + \frac{\partial \Im(f(z))}{\partial \Im z} \tag{227}$$

or in a more compact form:

$$\frac{\partial f(z)}{\partial \Im z} = i \frac{\partial f(z)}{\partial \Re z}. \tag{228}$$

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 ([24], [6]):

- Generalized Complex Derivative:

$$\frac{df(z)}{dz} = \frac{1}{2}\Big(\frac{\partial f(z)}{\partial \Re z} - i \frac{\partial f(z)}{\partial \Im z} \Big). \tag{229}$$

- Conjugate Complex Derivative

$$\frac{df(z)}{dz^*} = \frac{1}{2}\Big(\frac{\partial f(z)}{\partial \Re z} + i \frac{\partial f(z)}{\partial \Im z} \Big). \tag{230}$$

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 [21] when deriving a complex gradient.
Notice:

$$\frac{df(z)}{dz} \neq \frac{\partial f(z)}{\partial \Re z} + i \frac{\partial f(z)}{\partial \Im z}. \tag{231}$$

- Complex Gradient Vector: If f is a real function of a complex vector $\mathbf{z}$, then the complex gradient vector is given by ([14, p. 798])

$$\begin{aligned} \nabla f(\mathbf{z}) &= 2 \frac{df(\mathbf{z})}{d\mathbf{z}^*} \\ &= \frac{\partial f(\mathbf{z})}{\partial \Re \mathbf{z}} + i \frac{\partial f(\mathbf{z})}{\partial \Im \mathbf{z}}. \end{aligned} \tag{232}$$

- Complex Gradient Matrix: If f is a real function of a complex matrix $\mathbf{Z}$, then the complex gradient matrix is given by ([2])

$$\begin{aligned}\nabla f(\mathbf{Z}) &= 2\frac{df(\mathbf{Z})}{d\mathbf{Z}^*} \\ &= \frac{\partial f(\mathbf{Z})}{\partial \Re \mathbf{Z}} + i\frac{\partial f(\mathbf{Z})}{\partial \Im \mathbf{Z}}.\end{aligned} \tag{233}$$

These expressions can be used for gradient descent algorithms.

4.1.1 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 ([7])

$$\begin{aligned}\frac{\partial g(u)}{\partial x} &= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial u^*}\frac{\partial u^*}{\partial x} \\ &= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \left(\frac{\partial g^*}{\partial u}\right)^*\frac{\partial u^*}{\partial x}\end{aligned} \tag{234}$$

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:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\mathrm{Tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \partial \mathbf{U})}{\partial \mathbf{X}} + \frac{\mathrm{Tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}^*})^T \partial \mathbf{U}^*)}{\partial \mathbf{X}}. \tag{235}$$

4.1.2 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:

$$\frac{\partial \mathrm{Tr}(\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{236}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{X}^*)}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{I} \tag{237}$$

Since the two results have the same sign, the conjugate complex derivative (230) should be used.

$$\frac{\partial \mathrm{Tr}(\mathbf{X})}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^T)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{238}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{X})}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^T)}{\partial \Im \mathbf{X}} = -\mathbf{I} \tag{239}$$

Here, the two results have different signs, and the generalized complex derivative (229) should be used. Hereby, it can be seen that (100) holds even if $\mathbf{X}$ is a complex number.

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{A} \tag{240}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{A} \tag{241}$$

$$\frac{\partial \text{Tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \mathbf{A}^T \quad (242)$$

$$i\frac{\partial \text{Tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Im \mathbf{X}} = \mathbf{A}^T \quad (243)$$

$$\frac{\partial \text{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \frac{\partial \text{Tr}(\mathbf{X}^H\mathbf{X})}{\partial \Re \mathbf{X}} = 2\Re \mathbf{X} \quad (244)$$

$$i\frac{\partial \text{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = i\frac{\partial \text{Tr}(\mathbf{X}^H\mathbf{X})}{\partial \Im \mathbf{X}} = i2\Im \mathbf{X} \quad (245)$$

By inserting (244) and (245) in (229) and (230), it can be seen that

$$\frac{\partial \text{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}} = \mathbf{X}^* \quad (246)$$

$$\frac{\partial \text{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = \mathbf{X} \quad (247)$$

Since the function $\text{Tr}(\mathbf{X}\mathbf{X}^H)$ is a real function of the complex matrix $\mathbf{X}$, the complex gradient matrix (233) is given by

$$\nabla \text{Tr}(\mathbf{X}\mathbf{X}^H) = 2\frac{\partial \text{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = 2\mathbf{X} \quad (248)$$

4.1.3 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 (42) and (37), $\det(\mathbf{X}^H\mathbf{A}\mathbf{X})$ can be calculated as (see App. B.1.4 for details)

$$\begin{aligned}\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} - i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}}\Big) \\ &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\big((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\big)^T \quad (249)\end{aligned}$$

and the complex conjugate derivative yields

$$\begin{aligned}\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}^*} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} + i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}}\Big) \\ &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} \quad (250)\end{aligned}$$

4.2 Higher order and non-linear derivatives

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}\frac{(\mathbf{A}\mathbf{x})^H(\mathbf{A}\mathbf{x})}{(\mathbf{B}\mathbf{x})^H(\mathbf{B}\mathbf{x})} &= \frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{B}^H\mathbf{B}\mathbf{x}} \quad (251) \\ &= 2\frac{\mathbf{A}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{B}\mathbf{B}\mathbf{x}} - 2\frac{\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}\mathbf{B}^H\mathbf{B}\mathbf{x}}{(\mathbf{x}^H\mathbf{B}^H\mathbf{B}\mathbf{x})^2} \quad (252)\end{aligned}$$

4.3 Inverse of complex sum

Given real matrices $\mathbf{A}, \mathbf{B}$ find the inverse of the complex sum $\mathbf{A} + i\mathbf{B}$. Form the auxiliary matrices

$$\begin{aligned} \mathbf{E} &= \mathbf{A} + t\mathbf{B} && (253) \\ \mathbf{F} &= \mathbf{B} - t\mathbf{A}, && (254) \end{aligned}$$

and find a value of t such that $\mathbf{E}^{-1}$ exists. Then

$$\begin{aligned} (\mathbf{A} + i\mathbf{B})^{-1} &= (1 - it)(\mathbf{E} + i\mathbf{F})^{-1} && (255) \\ &= (1 - it)((\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1} - i(\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}\mathbf{F}\mathbf{E}^{-1}) && (256) \\ &= (1 - it)(\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}(\mathbf{I} - i\mathbf{F}\mathbf{E}^{-1}) && (257) \\ &= (\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}((\mathbf{I} - t\mathbf{F}\mathbf{E}^{-1}) - i(t\mathbf{I} + \mathbf{F}\mathbf{E}^{-1})) && (258) \\ &= (\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}(\mathbf{I} - t\mathbf{F}\mathbf{E}^{-1}) \\ &\quad -i(\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}(t\mathbf{I} + \mathbf{F}\mathbf{E}^{-1}) && (259) \end{aligned}$$

5 Solutions and Decompositions

5.1 Solutions to linear equations

5.1.1 Simple Linear Regression

Assume we have data (x_n, y_n) for $n = 1, ..., N$ and are seeking the parameters $a, b \in \mathbb{R}$ such that $y_i \cong ax_i + b$. With a least squares error function, the optimal values for a, b can be expressed using the notation

$$\mathbf{x} = (x_1, ..., x_N)^T \qquad \mathbf{y} = (y_1, ..., y_N)^T \qquad \mathbf{1} = (1, ..., 1)^T \in \mathbb{R}^{N\times 1}$$

and

$$\begin{array}{lll} R_{xx} = \mathbf{x}^T\mathbf{x} & R_{x1} = \mathbf{x}^T\mathbf{1} & R_{11} = \mathbf{1}^T\mathbf{1} \\ R_{yx} = \mathbf{y}^T\mathbf{x} & R_{y1} = \mathbf{y}^T\mathbf{1} & \end{array}$$

as

$$\left[\begin{array}{c} a \\ b \end{array}\right] = \left[\begin{array}{cc} R_{xx} & R_{x1} \\ R_{x1} & R_{11} \end{array}\right]^{-1} \left[\begin{array}{c} R_{x,y} \\ R_{y1} \end{array}\right] \tag{260}$$

5.1.2 Existence in Linear Systems

Assume $\mathbf{A}$ is $n \times m$ and consider the linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{261}$$

Construct the augmented matrix $\mathbf{B} = [\mathbf{A}\ \mathbf{b}]$ then

Condition	*Solution*
$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}) = m$	Unique solution $\mathbf{x}$
$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}) < m$	Many solutions $\mathbf{x}$
$\text{rank}(\mathbf{A}) < \text{rank}(\mathbf{B})$	No solutions $\mathbf{x}$

5.1.3 Standard Square

Assume $\mathbf{A}$ is square and invertible, then

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad \Rightarrow \qquad \mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{262}$$

5.1.4 Degenerated Square

Assume $\mathbf{A}$ is $n \times n$ but of rank $r < n$. In that case, the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is solved by

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b}$$

where $\mathbf{A}^{+}$ is the pseudo-inverse of the rank-deficient matrix, constructed as described in section 3.6.3.

5.1.5 Cramer's rule

The equation

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{263}$$

where $\mathbf{A}$ is square has exactly one solution $\mathbf{x}$ if the ith element in x can be found as

$$x_i = \frac{\det \mathbf{B}}{\det \mathbf{A}}, \tag{264}$$

where $\mathbf{B}$ equals $\mathbf{A}$, but the ith column in $\mathbf{A}$ has been substituted by $\mathbf{b}$.

5.1.6 Over-determined Rectangular

Assume $\mathbf{A}$ to be $n \times m$, $n > m$ (tall) and $\text{rank}(\mathbf{A}) = m$, then

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad \Rightarrow \qquad \mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \mathbf{A}^+\mathbf{b} \tag{265}$$

that is *if* there exists a solution $\mathbf{x}$ at all! If there is no solution the following can be useful:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad \Rightarrow \qquad \mathbf{x}_{min} = \mathbf{A}^+\mathbf{b} \tag{266}$$

Now $\mathbf{x}_{min}$ is the vector $\mathbf{x}$ which minimizes $||\mathbf{A}\mathbf{x} - \mathbf{b}||^2$, i.e. the vector which is "least wrong". The matrix $\mathbf{A}^+$ is the pseudo-inverse of $\mathbf{A}$. See [3].

5.1.7 Under-determined Rectangular

Assume $\mathbf{A}$ is $n \times m$ and $n < m$ ("broad") and $\text{rank}(\mathbf{A}) = n$.

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad \Rightarrow \qquad \mathbf{x}_{min} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} \tag{267}$$

The equation have many solutions $\mathbf{x}$. But $\mathbf{x}_{min}$ is the solution which minimizes $||\mathbf{A}\mathbf{x} - \mathbf{b}||^2$ and also the solution with the smallest norm $||\mathbf{x}||^2$. The same holds for a matrix version: Assume $\mathbf{A}$ is $n \times m$, $\mathbf{X}$ is $m \times n$ and $\mathbf{B}$ is $n \times n$, then

$$\mathbf{A}\mathbf{X} = \mathbf{B} \qquad \Rightarrow \qquad \mathbf{X}_{min} = \mathbf{A}^+\mathbf{B} \tag{268}$$

The equation have many solutions $\mathbf{X}$. But $\mathbf{X}_{min}$ is the solution which minimizes $||\mathbf{A}\mathbf{X} - \mathbf{B}||^2$ and also the solution with the smallest norm $||\mathbf{X}||^2$. See [3].

Similar but different: Assume $\mathbf{A}$ is square $n \times n$ and the matrices $\mathbf{B}_0, \mathbf{B}_1$ are $n \times N$, where $N > n$, then if $\mathbf{B}_0$ has maximal rank

$$\mathbf{A}\mathbf{B}_0 = \mathbf{B}_1 \qquad \Rightarrow \qquad \mathbf{A}_{min} = \mathbf{B}_1\mathbf{B}_0^T(\mathbf{B}_0\mathbf{B}_0^T)^{-1} \tag{269}$$

where $\mathbf{A}_{min}$ denotes the matrix which is optimal in a least square sense. An interpretation is that $\mathbf{A}$ is the linear approximation which maps the columns vectors of $\mathbf{B}_0$ into the columns vectors of $\mathbf{B}_1$.

5.1.8 Linear form and zeros

$$\mathbf{A}\mathbf{x} = \mathbf{0}, \quad \forall \mathbf{x} \qquad \Rightarrow \qquad \mathbf{A} = \mathbf{0} \tag{270}$$

5.1.9 Square form and zeros

If $\mathbf{A}$ is symmetric, then

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{0}, \quad \forall \mathbf{x} \qquad \Rightarrow \qquad \mathbf{A} = \mathbf{0} \tag{271}$$