The Matrix Cookbook

Kaare Brandt Petersen
Michael Syskind Pedersen

Version: February 16, 2006

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 at cookbook@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: Christian Rishøj, Douglas L. Theobald, Esben Hoegh-Rasmussen, Lars Christiansen, and Vasile Sima. 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
$\mathbf{a}_i$	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. $(\mathrm{diag}(\mathbf{A}))_{ij} = \delta_{ij}A_{ij}$
$\mathrm{vec}(\mathbf{A})$	The vector-version of the matrix $\mathbf{A}$ (see Sec. 9.2.2)
$\|\|\mathbf{A}\|\|$	Matrix norm (subscript if any denotes what norm)
$\mathbf{A}^T$	Transposed matrix
$\mathbf{A}^*$	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

$$
\begin{aligned}
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{ABC}...)^{-1} &= ...\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T \\
(\mathbf{A}+\mathbf{B})^T &= \mathbf{A}^T+\mathbf{B}^T \\
(\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T \\
(\mathbf{ABC}...)^T &= ...\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T \\
(\mathbf{A}^H)^{-1} &= (\mathbf{A}^{-1})^H \\
(\mathbf{A}+\mathbf{B})^H &= \mathbf{A}^H+\mathbf{B}^H \\
(\mathbf{AB})^H &= \mathbf{B}^H\mathbf{A}^H \\
(\mathbf{ABC}...)^H &= ...\mathbf{C}^H\mathbf{B}^H\mathbf{A}^H
\end{aligned}
$$

1.1 Trace and Determinants

$$
\begin{aligned}
\mathrm{Tr}(\mathbf{A}) &= \textstyle\sum_i \mathbf{A}_{ii} \\
\mathrm{Tr}(\mathbf{A}) &= \textstyle\sum_i \lambda_i, \qquad \lambda_i = \mathrm{eig}(\mathbf{A}) \\
\mathrm{Tr}(\mathbf{A}) &= \mathrm{Tr}(\mathbf{A}^T) \\
\mathrm{Tr}(\mathbf{AB}) &= \mathrm{Tr}(\mathbf{BA}) \\
\mathrm{Tr}(\mathbf{A}+\mathbf{B}) &= \mathrm{Tr}(\mathbf{A})+\mathrm{Tr}(\mathbf{B}) \\
\mathrm{Tr}(\mathbf{ABC}) &= \mathrm{Tr}(\mathbf{BCA}) = \mathrm{Tr}(\mathbf{CAB}) \\
\det(\mathbf{A}) &= \textstyle\prod_i \lambda_i \qquad \lambda_i = \mathrm{eig}(\mathbf{A}) \\
\det(\mathbf{AB}) &= \det(\mathbf{A})\det(\mathbf{B}) \\
\det(\mathbf{A}^{-1}) &= 1/\det(\mathbf{A}) \\
\det(\mathbf{I}+\mathbf{u}\mathbf{v}^T) &= 1+\mathbf{u}^T\mathbf{v}
\end{aligned}
$$

1.2 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

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

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

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

$$\mathbf{v}_1 \propto \left[\begin{array}{c} A_{12} \\ \lambda_1 - A_{11} \end{array} \right] \qquad \mathbf{v}_2 \propto \left[\begin{array}{c} A_{12} \\ \lambda_2 - A_{11} \end{array} \right]$$

Inverse

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \left[\begin{array}{cc} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{array} \right]$$

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 2.5 for differentiation of structured matrices. The basic assumptions can be written in a formula as

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

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 ([13]):

$$\begin{aligned}
\partial\mathbf{A} &= 0 \qquad (\mathbf{A} \text{ is a constant}) && (1)\\
\partial(\alpha\mathbf{X}) &= \alpha\partial\mathbf{X} && (2)\\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial\mathbf{X}+\partial\mathbf{Y} && (3)\\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial\mathbf{X}) && (4)\\
\partial(\mathbf{XY}) &= (\partial\mathbf{X})\mathbf{Y}+\mathbf{X}(\partial\mathbf{Y}) && (5)\\
\partial(\mathbf{X}\circ\mathbf{Y}) &= (\partial\mathbf{X})\circ\mathbf{Y}+\mathbf{X}\circ(\partial\mathbf{Y}) && (6)\\
\partial(\mathbf{X}\otimes\mathbf{Y}) &= (\partial\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes(\partial\mathbf{Y}) && (7)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} && (8)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (9)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (10)\\
\partial\mathbf{X}^T &= (\partial\mathbf{X})^T && (11)\\
\partial\mathbf{X}^H &= (\partial\mathbf{X})^H && (12)
\end{aligned}$$

2.1 Derivatives of a Determinant

2.1.1 General form

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

2.1.2 Linear forms

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

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

2.1.3 Square forms

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

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

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

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

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

$$\frac{\partial \det(\mathbf{X}^T\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X}^T\mathbf{A}\mathbf{X})(\mathbf{A}\mathbf{X}(\mathbf{X}^T\mathbf{A}\mathbf{X})^{-1} + \mathbf{A}^T\mathbf{X}(\mathbf{X}^T\mathbf{A}^T\mathbf{X})^{-1}) \quad (13)$$

2.1.4 Other nonlinear forms

Some special cases are (See [8, 7])

$$\begin{aligned}
\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})|}{\partial \mathbf{X}} &= 2(\mathbf{X}^+)^T \\
\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})}{\partial \mathbf{X}^+} &= -2\mathbf{X}^T \\
\frac{\partial \ln |\det(\mathbf{X})|}{\partial \mathbf{X}} &= (\mathbf{X}^{-1})^T = (\mathbf{X}^T)^{-1} \\
\frac{\partial \det(\mathbf{X}^k)}{\partial \mathbf{X}} &= k\det(\mathbf{X}^k)\mathbf{X}^{-T}
\end{aligned}$$

2.2 Derivatives of an Inverse

From [19] we have the basic identity

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

from which it follows

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

2.3 Derivatives of Matrices, Vectors and Scalar Forms

2.3.1 First Order

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

2.3.2 Second Order

$$
\begin{aligned}
\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl}X_{mn} &= 2\sum_{kl} X_{kl} \\
\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} &= \mathbf{X}(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T) \\
\frac{\partial (\mathbf{B}\mathbf{x}+\mathbf{b})^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d})}{\partial \mathbf{x}} &= \mathbf{B}^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d}) + \mathbf{D}^T\mathbf{C}^T(\mathbf{B}\mathbf{x}+\mathbf{b}) \\
\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})_{kl}}{\partial X_{ij}} &= \delta_{lj}(\mathbf{X}^T\mathbf{B})_{ki} + \delta_{kj}(\mathbf{B}\mathbf{X})_{il} \\
\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})}{\partial X_{ij}} &= \mathbf{X}^T\mathbf{B}\mathbf{J}^{ij} + \mathbf{J}^{ji}\mathbf{B}\mathbf{X} \qquad (\mathbf{J}^{ij})_{kl} = \delta_{ik}\delta_{jl}
\end{aligned}
$$

See Sec 8.2 for useful properties of the Single-entry matrix $\mathbf{J}^{ij}$

$$
\begin{aligned}
\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} &= (\mathbf{B}+\mathbf{B}^T)\mathbf{x} \\
\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} &= \mathbf{D}^T\mathbf{X}\mathbf{b}\mathbf{c}^T + \mathbf{D}\mathbf{X}\mathbf{c}\mathbf{b}^T \\
\frac{\partial}{\partial \mathbf{X}}(\mathbf{X}\mathbf{b}+\mathbf{c})^T\mathbf{D}(\mathbf{X}\mathbf{b}+\mathbf{c}) &= (\mathbf{D}+\mathbf{D}^T)(\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T
\end{aligned}
$$

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

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) &= -2\mathbf{A}^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) \\
\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s}) &= -2\mathbf{W}(\mathbf{x}-\mathbf{s}) \\
\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) &= 2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) \\
\frac{\partial}{\partial \mathbf{A}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) &= -2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})\mathbf{s}^T
\end{aligned}$$

2.3.3 Higher order and non-linear

$$\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^T\mathbf{X}^n\mathbf{b} = \sum_{r=0}^{n-1}(\mathbf{X}^r)^T\mathbf{a}\mathbf{b}^T(\mathbf{X}^{n-1-r})^T \tag{14}$$

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^T(\mathbf{X}^n)^T\mathbf{X}^n\mathbf{b} = \sum_{r=0}^{n-1}\Big[&\mathbf{X}^{n-1-r}\mathbf{a}\mathbf{b}^T(\mathbf{X}^n)^T\mathbf{X}^r \\
&+(\mathbf{X}^r)^T\mathbf{X}^n\mathbf{a}\mathbf{b}^T(\mathbf{X}^{n-1-r})^T\Big]
\end{aligned} \tag{15}$$

See B.1.1 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

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

2.3.4 Gradient and Hessian

Using the above we have for the gradient and the hessian

$$\begin{aligned}
f &= \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} \\
\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} &= (\mathbf{A}+\mathbf{A}^T)\mathbf{x} + \mathbf{b} \\
\frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^T} &= \mathbf{A}+\mathbf{A}^T
\end{aligned}$$

2.4 Derivatives of Traces

2.4.1 First Order

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) &= \mathbf{I} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{A}) &= \mathbf{A}^T \qquad (16) \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) &= \mathbf{A}^T\mathbf{B}^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}) &= \mathbf{B}\mathbf{A} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{A}) &= \mathbf{A} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T) &= \mathbf{A}
\end{aligned}$$

2.4.2 Second Order

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2) &= 2\mathbf{X}^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2\mathbf{B}) &= (\mathbf{X}\mathbf{B}+\mathbf{B}\mathbf{X})^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{B}\mathbf{X}) &= \mathbf{B}\mathbf{X}+\mathbf{B}^T\mathbf{X} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{B}\mathbf{X}^T) &= \mathbf{X}\mathbf{B}^T+\mathbf{X}\mathbf{B} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X}) &= \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T+\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) &= 2\mathbf{X} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}\mathbf{X}\mathbf{X}^T) &= (\mathbf{B}+\mathbf{B}^T)\mathbf{X} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{B}) &= \mathbf{C}^T\mathbf{X}\mathbf{B}\mathbf{B}^T+\mathbf{C}\mathbf{X}\mathbf{B}\mathbf{B}^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{X}^T\mathbf{B}\mathbf{X}\mathbf{C}\right] &= \mathbf{B}\mathbf{X}\mathbf{C}+\mathbf{B}^T\mathbf{X}\mathbf{C}^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{C}) &= \mathbf{A}^T\mathbf{C}^T\mathbf{X}\mathbf{B}^T+\mathbf{C}\mathbf{A}\mathbf{X}\mathbf{B} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{A}\mathbf{X}\mathbf{b}+\mathbf{c})(\mathbf{A}\mathbf{X}\mathbf{b}+\mathbf{c})^T\Big] &= 2\mathbf{A}^T(\mathbf{A}\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T
\end{aligned}$$

See [7].

2.4.3 Higher Order

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^k) &= k(\mathbf{X}^{k-1})^T \\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^k) &= \sum_{r=0}^{k-1} (\mathbf{X}^r \mathbf{A}\mathbf{X}^{k-r-1})^T \\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\left[\mathbf{B}^T\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{B}\right] &= \mathbf{C}\mathbf{X}\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{B}\mathbf{B}^T \\
&\quad + \mathbf{C}^T\mathbf{X}\mathbf{B}\mathbf{B}^T\mathbf{X}^T\mathbf{C}^T\mathbf{X} \\
&\quad + \mathbf{C}\mathbf{X}\mathbf{B}\mathbf{B}^T\mathbf{X}^T\mathbf{C}\mathbf{X} \\
&\quad + \mathbf{C}^T\mathbf{X}\mathbf{X}^T\mathbf{C}^T\mathbf{X}\mathbf{B}\mathbf{B}^T
\end{aligned}$$

2.4.4 Other

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

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

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

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] &= -2\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \\
&\quad + 2\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}
\end{aligned}$$

See [7].

2.5 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 $f(\mathbf{A})$

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

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

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

If $\mathbf{A}$ has no special structure we have simply $\mathbf{S}^{ij} = \mathbf{J}^{ij}$, that is, the structure matrix is simply the singleentry matrix. Many structures have a representation in singleentry matrices, see Sec. 8.2.6 for more examples of structure matrices.

2.5.1 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 $\mathbf{X}$:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}} \tag{17}$$

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

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M} \sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}} \tag{18}$$

Using matrix notation, this can be written as:

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \mathrm{Tr}\Big[(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \frac{\partial \mathbf{U}}{\partial X_{ij}} \Big]. \tag{19}$$

2.5.2 Symmetric

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

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

That is, e.g., ([5], [20]):

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A} \circ \mathbf{I}), \text{ see (23)} \tag{20}$$

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})) \tag{21}$$

$$\frac{\partial \ln \det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I}) \tag{22}$$

2.5.3 Diagonal

If $\mathbf{X}$ is diagonal, then ([13]):

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} \circ \mathbf{I} \tag{23}$$

2.5.4 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.

$$
\begin{aligned}
& \frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} \\
= \; & \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} \\
= \; & \begin{bmatrix}
\mathrm{Tr}(\mathbf{A}) & \mathrm{Tr}([\mathbf{A}^T]_{n1}) & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) & \cdots & A_{n1} \\
\mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A}) & \ddots & \ddots & \vdots \\
\mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \ddots & \ddots & \ddots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) \\
\vdots & \ddots & \ddots & \ddots & \mathrm{Tr}([\mathbf{A}^T]_{n1}) \\
A_{1n} & \cdots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A})
\end{bmatrix} \\
\equiv \; & \boldsymbol{\alpha}(\mathbf{A})
\end{aligned} \tag{24}
$$

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

$$
\begin{aligned}
\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} &= \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} \\
&= \boldsymbol{\alpha}(\mathbf{A}) + \boldsymbol{\alpha}(\mathbf{A})^T - \boldsymbol{\alpha}(\mathbf{A}) \circ \mathbf{I}
\end{aligned} \tag{25}
$$

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

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}, \tag{26}$$

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. [9]).

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

$$\mathrm{cof}(\mathbf{A}, i, j) = (-1)^{i+j}\det([\mathbf{A}]_{ij}), \tag{27}$$

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

$$\mathrm{cof}(\mathbf{A}) = \begin{bmatrix} \mathrm{cof}(\mathbf{A},1,1) & \cdots & \mathrm{cof}(\mathbf{A},1,n) \\ \vdots & \mathrm{cof}(\mathbf{A},i,j) & \vdots \\ \mathrm{cof}(\mathbf{A},n,1) & \cdots & \mathrm{cof}(\mathbf{A},n,n) \end{bmatrix} \tag{28}$$

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

$$\mathrm{adj}(\mathbf{A}) = (\mathrm{cof}(\mathbf{A}))^T, \tag{29}$$

3.1.3 Determinant

The *determinant* of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is defined as (see [9])

$$\begin{aligned} \det(\mathbf{A}) &= \sum_{j=1}^{n}(-1)^{j+1}A_{1j}\det([\mathbf{A}]_{1j}) \\ &= \sum_{j=1}^{n} A_{1j}\mathrm{cof}(\mathbf{A},1,j). \end{aligned} \tag{30}$$

3.1.4 Construction

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

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\cdot \mathrm{adj}(\mathbf{A}) \tag{31}$$

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 5.2 on singular values),

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

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

$$c(\mathbf{A}) = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\|, \tag{32}$$

where $\|\cdot\|$ is a norm such as e.g the 1-norm, the 2-norm, the ∞-norm or the Frobenius norm (see Sec 9.4 for more on matrix norms).

3.2 Exact Relations

3.2.1 The Woodbury identity

$$(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{A}^{-1}$$

If $\mathbf{P}, \mathbf{R}$ are positive definite, then (see [22])

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

3.2.2 The Kailath Variant

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

See [4] page 153.

3.2.3 The Searle Set of Identities

The following set of identities, can be found in [17], page 151,

$$(\mathbf{I} + \mathbf{A}^{-1})^{-1} = \mathbf{A}(\mathbf{A} + \mathbf{I})^{-1}$$

$$(\mathbf{A} + \mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}$$

$$(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}$$

$$\mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} = \mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$$

$$\mathbf{A}^{-1} + \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{A} + \mathbf{B})\mathbf{B}^{-1}$$

$$(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}\mathbf{B}$$

$$(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}$$

3.3 Implication on Inverses

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1} \quad \Rightarrow \quad \mathbf{A}\mathbf{B}^{-1}\mathbf{A} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}$$

See [17].

3.3.1 A PosDef identity

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

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

See [22].

3.4 Approximations

$$(\mathbf{I}+\mathbf{A})^{-1} = \mathbf{I} - \mathbf{A} + \mathbf{A}^2 - \mathbf{A}^3 + \ldots$$

$$\mathbf{A} - \mathbf{A}(\mathbf{I}+\mathbf{A})^{-1}\mathbf{A} \cong \mathbf{I} - \mathbf{A}^{-1} \qquad \text{if } \mathbf{A} \text{ large and symmetric}$$

If σ^2 is small then

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

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 [18])

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

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

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

The matrix $\mathbf{A}^+$ is unique and does always exist.

3.6.2 Properties

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

$$\begin{aligned}
(\mathbf{A}^+)^+ &= \mathbf{A} \\
(\mathbf{A}^T)^+ &= (\mathbf{A}^+)^T \\
(c\mathbf{A})^+ &= (1/c)\mathbf{A}^+ \\
(\mathbf{A}^T\mathbf{A})^+ &= \mathbf{A}^+(\mathbf{A}^T)^+ \\
(\mathbf{A}\mathbf{A}^T)^+ &= (\mathbf{A}^T)^+\mathbf{A}^+
\end{aligned}$$

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

$$\begin{aligned}
(\mathbf{A}\mathbf{A}^+)(\mathbf{A}\mathbf{A}^+) &= \mathbf{A}\mathbf{A}^+ \\
(\mathbf{A}^+\mathbf{A})(\mathbf{A}^+\mathbf{A}) &= \mathbf{A}^+\mathbf{A} \\
\mathrm{Tr}(\mathbf{A}\mathbf{A}^+) &= \mathrm{rank}(\mathbf{A}\mathbf{A}^+) \qquad \text{(See [18])} \\
\mathrm{Tr}(\mathbf{A}^+\mathbf{A}) &= \mathrm{rank}(\mathbf{A}^+\mathbf{A}) \qquad \text{(See [18])}
\end{aligned}$$

3.6.3 Construction

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

$$\begin{array}{llll}
\mathbf{A}\ n\times n & \text{Square} & \mathrm{rank}(\mathbf{A}) = n & \Rightarrow \quad \mathbf{A}^+ = \mathbf{A}^{-1} \\
\mathbf{A}\ n\times m & \text{Broad} & \mathrm{rank}(\mathbf{A}) = n & \Rightarrow \quad \mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \\
\mathbf{A}\ n\times m & \text{Tall} & \mathrm{rank}(\mathbf{A}) = m & \Rightarrow \quad \mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T
\end{array}$$

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

$$\mathbf{A}^+ = \mathbf{V}\mathbf{D}^+\mathbf{U}^T$$

A different way is this: There does always exists 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

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

See [3].

4 Complex Matrices

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]):

$$\frac{df(z)}{dz} = \frac{\partial\Re(f(z))}{\partial\Re z} + i\frac{\partial\Im(f(z))}{\partial\Re z} \tag{33}$$

and

$$\frac{df(z)}{dz} = -i\frac{\partial\Re(f(z))}{\partial\Im z} + \frac{\partial\Im(f(z))}{\partial\Im z} \tag{34}$$

or in a more compact form:

$$\frac{\partial f(z)}{\partial\Im z} = i\frac{\partial f(z)}{\partial\Re z}. \tag{35}$$

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 ([16], [6]):

- Generalized Complex Derivative:

$$\frac{df(z)}{dz} = \frac{1}{2}\Big(\frac{\partial f(z)}{\partial\Re z} - i\frac{\partial f(z)}{\partial\Im z}\Big). \tag{36}$$

- Conjugate Complex Derivative

$$\frac{df(z)}{dz^*} = \frac{1}{2}\Big(\frac{\partial f(z)}{\partial\Re z} + i\frac{\partial f(z)}{\partial\Im z}\Big). \tag{37}$$

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 [14] when deriving a complex gradient.
Notice:

$$\frac{df(z)}{dz} \neq \frac{\partial f(z)}{\partial\Re z} + i\frac{\partial f(z)}{\partial\Im z}. \tag{38}$$

- Complex Gradient Vector: If f is a real function of a complex vector $\mathbf{z}$, then the complex gradient vector is given by ([11, p. 798])

$$\begin{aligned} \nabla f(\mathbf{z}) &= 2\frac{df(\mathbf{z})}{d\mathbf{z}^*} \\ &= \frac{\partial f(\mathbf{z})}{\partial\Re\mathbf{z}} + i\frac{\partial f(\mathbf{z})}{\partial\Im\mathbf{z}}. \end{aligned} \tag{39}$$

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

$$\begin{aligned} \nabla f(\mathbf{Z}) &= 2\frac{df(\mathbf{Z})}{d\mathbf{Z}^*} \\ &= \frac{\partial f(\mathbf{Z})}{\partial \Re \mathbf{Z}} + i\frac{\partial f(\mathbf{Z})}{\partial \Im \mathbf{Z}}. \end{aligned} \tag{40}$$

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])

$$\begin{aligned} \frac{\partial g(u)}{\partial x} &= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial u^*}\frac{\partial u^*}{\partial x} \\ &= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \Big(\frac{\partial g^*}{\partial u}\Big)^*\frac{\partial u^*}{\partial x} \end{aligned} \tag{41}$$

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:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\mathrm{Tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \partial \mathbf{U})}{\partial \mathbf{X}} + \frac{\mathrm{Tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}^*})^T \partial \mathbf{U}^*)}{\partial \mathbf{X}}. \tag{42}$$

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:

$$\frac{\partial \mathrm{Tr}(\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{43}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{X}^*)}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{I} \tag{44}$$

Since the two results have the same sign, the conjugate complex derivative (37) should be used.

$$\frac{\partial \mathrm{Tr}(\mathbf{X})}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^T)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{45}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{X})}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^T)}{\partial \Im \mathbf{X}} = -\mathbf{I} \tag{46}$$

Here, the two results have different signs, the generalized complex derivative (36) should be used. Hereby, it can be seen that (16) holds even if $\mathbf{X}$ is a complex number.

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{A} \tag{47}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{A} \tag{48}$$

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \mathbf{A}^T \tag{49}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Im \mathbf{X}} = \mathbf{A}^T \tag{50}$$

$$\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^H\mathbf{X})}{\partial \Re \mathbf{X}} = 2\Re \mathbf{X} \tag{51}$$

$$i\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^H\mathbf{X})}{\partial \Im \mathbf{X}} = i2\Im \mathbf{X} \tag{52}$$

By inserting (51) and (52) in (36) and (37), it can be seen that

$$\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}} = \mathbf{X}^* \tag{53}$$

$$\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = \mathbf{X} \tag{54}$$

Since the function $\mathrm{Tr}(\mathbf{X}\mathbf{X}^H)$ is a real function of the complex matrix $\mathbf{X}$, the complex gradient matrix (40) is given by

$$\nabla \mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = 2\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = 2\mathbf{X} \tag{55}$$

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 (9) and (5), $\det(\mathbf{X}^H\mathbf{A}\mathbf{X})$ can be calculated as (see Sec. B.1.2 for details)

$$\begin{aligned}\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} - i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}}\Big)\\ &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\big((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\big)^T\end{aligned} \tag{56}$$

and the complex conjugate derivative yields

$$\begin{aligned}\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}^*} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} + i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}}\Big)\\ &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\end{aligned} \tag{57}$$

5 Decompositions

5.1 Eigenvalues and Eigenvectors

5.1.1 Definition

The eigenvectors $\mathbf{v}$ and eigenvalues λ are the ones satisfying

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{D}, \qquad (\mathbf{D})_{ij} = \delta_{ij}\lambda_i$$

where the columns of $\mathbf{V}$ are the vectors $\mathbf{v}_i$

5.1.2 General Properties

$$\begin{array}{rcl} \mathrm{eig}(\mathbf{A}\mathbf{B}) & = & \mathrm{eig}(\mathbf{B}\mathbf{A}) \\ \mathbf{A} \text{ is } n \times m & \Rightarrow & \text{At most } \min(n,m) \text{ distinct } \lambda_i \\ \mathrm{rank}(\mathbf{A}) = r & \Rightarrow & \text{At most } r \text{ non-zero } \lambda_i \end{array}$$

5.1.3 Symmetric

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

$$\begin{array}{rcl} \mathbf{V}\mathbf{V}^T & = & \mathbf{I} \qquad \text{(i.e. } \mathbf{V} \text{ is orthogonal)} \\ \lambda_i & \in & \mathbb{R} \qquad \text{(i.e. } \lambda_i \text{ is real)} \\ \mathrm{Tr}(\mathbf{A}^p) & = & \sum_i \lambda_i^p \\ \mathrm{eig}(\mathbf{I} + c\mathbf{A}) & = & 1 + c\lambda_i \\ \mathrm{eig}(\mathbf{A} - c\mathbf{I}) & = & \lambda_i - c \\ \mathrm{eig}(\mathbf{A}^{-1}) & = & \lambda_i^{-1} \end{array}$$

For a symmetric, positive matrix $\mathbf{A}$,

$$\mathrm{eig}(\mathbf{A}^T\mathbf{A}) = \mathrm{eig}(\mathbf{A}\mathbf{A}^T) = \mathrm{eig}(\mathbf{A}) \circ \mathrm{eig}(\mathbf{A}) \tag{58}$$

5.2 Singular Value Decomposition

Any $n \times m$ matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

where

$$\begin{array}{rcll} \mathbf{U} & = & \text{eigenvectors of } \mathbf{A}\mathbf{A}^T & n \times n \\ \mathbf{D} & = & \sqrt{\mathrm{diag}(\mathrm{eig}(\mathbf{A}\mathbf{A}^T))} & n \times m \\ \mathbf{V} & = & \text{eigenvectors of } \mathbf{A}^T\mathbf{A} & m \times m \end{array}$$

5.2.1 Symmetric Square decomposed into squares

Assume $\mathbf{A}$ to be $n \times n$ and symmetric. Then

$$\begin{bmatrix} \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \end{bmatrix}$$

where $\mathbf{D}$ is diagonal with the eigenvalues of $\mathbf{A}$ and $\mathbf{V}$ is orthogonal and the eigenvectors of $\mathbf{A}$.

5.2.2 Square decomposed into squares

Assume $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then

$$\begin{bmatrix} \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{U}^T \end{bmatrix}$$

where $\mathbf{D}$ is diagonal with the square root of the eigenvalues of $\mathbf{A}\mathbf{A}^T$, $\mathbf{V}$ is the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{U}^T$ is the eigenvectors of $\mathbf{A}^T\mathbf{A}$.

5.2.3 Square decomposed into rectangular

Assume $\mathbf{V}_*\mathbf{D}_*\mathbf{U}_*^T = \mathbf{0}$ then we can expand the SVD of $\mathbf{A}$ into

$$\begin{bmatrix} \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \mid \mathbf{V}_* \end{bmatrix} \left[\begin{array}{c|c} \mathbf{D} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{D}_* \end{array}\right] \left[\begin{array}{c} \mathbf{U}^T \\ \hline \mathbf{U}_*^T \end{array}\right]$$

where the SVD of $\mathbf{A}$ is $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{U}^T$.

5.2.4 Rectangular decomposition I

Assume $\mathbf{A}$ is $n \times m$

$$\begin{bmatrix} \quad \mathbf{A} \quad \end{bmatrix} = \begin{bmatrix} \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{D} \end{bmatrix} \begin{bmatrix} \quad \mathbf{U}^T \quad \end{bmatrix}$$

where $\mathbf{D}$ is diagonal with the square root of the eigenvalues of $\mathbf{A}\mathbf{A}^T$, $\mathbf{V}$ is the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{U}^T$ is the eigenvectors of $\mathbf{A}^T\mathbf{A}$.

5.2.5 Rectangular decomposition II

Assume $\mathbf{A}$ is $n \times m$

$$\begin{bmatrix} \quad \mathbf{A} \quad \end{bmatrix} = \begin{bmatrix} \quad \mathbf{V} \quad \end{bmatrix} \begin{bmatrix} \\ \quad \mathbf{D} \quad \\ \\ \end{bmatrix} \begin{bmatrix} \\ \quad \mathbf{U}^T \quad \\ \\ \end{bmatrix}$$

5.2.6 Rectangular decomposition III

Assume $\mathbf{A}$ is $n \times m$

$$\begin{bmatrix} \quad \mathbf{A} \quad \end{bmatrix} = \begin{bmatrix} \mathbf{V} \end{bmatrix} \begin{bmatrix} \quad \mathbf{D} \quad \end{bmatrix} \begin{bmatrix} \\ \quad \mathbf{U}^T \quad \\ \\ \end{bmatrix}$$

where $\mathbf{D}$ is diagonal with the square root of the eigenvalues of $\mathbf{A}\mathbf{A}^T$, $\mathbf{V}$ is the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{U}^T$ is the eigenvectors of $\mathbf{A}^T\mathbf{A}$.

5.3 Triangular Decomposition

5.3.1 Cholesky-decomposition

Assume $\mathbf{A}$ is positive definite, then

$$\mathbf{A} = \mathbf{B}^T\mathbf{B}$$

where $\mathbf{B}$ is a unique upper triangular matrix.

6 Statistics and Probability

6.1 Definition of Moments

Assume $\mathbf{x} \in \mathbb{R}^{n\times 1}$ is a random variable

6.1.1 Mean

The vector of means, $\mathbf{m}$, is defined by

$$(\mathbf{m})_i = \langle x_i \rangle$$

6.1.2 Covariance

The matrix of covariance $\mathbf{M}$ is defined by

$$(\mathbf{M})_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

or alternatively as

$$\mathbf{M} = \langle (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \rangle$$

6.1.3 Third moments

The matrix of third centralized moments – in some contexts referred to as coskewness – is defined using the notation

$$m^{(3)}_{ijk} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)(x_k - \langle x_k \rangle) \rangle$$

as

$$\mathbf{M}_3 = \left[m^{(3)}_{::1} m^{(3)}_{::2} ... m^{(3)}_{::n} \right]$$

where ':' denotes all elements within the given index. $\mathbf{M}_3$ can alternatively be expressed as

$$\mathbf{M}_3 = \langle (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \otimes (\mathbf{x} - \mathbf{m})^T \rangle$$

6.1.4 Fourth moments

The matrix of fourth centralized moments – in some contexts referred to as cokurtosis – is defined using the notation

$$m^{(4)}_{ijkl} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)(x_k - \langle x_k \rangle)(x_l - \langle x_l \rangle) \rangle$$

as

$$\mathbf{M}_4 = \left[m^{(4)}_{::11} m^{(4)}_{::21} ... m^{(4)}_{::n1} | m^{(4)}_{::12} m^{(4)}_{::22} ... m^{(4)}_{::n2} | ... | m^{(4)}_{::1n} m^{(4)}_{::2n} ... m^{(4)}_{::nn} \right]$$

or alternatively as

$$\mathbf{M}_4 = \langle (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \otimes (\mathbf{x} - \mathbf{m})^T \otimes (\mathbf{x} - \mathbf{m})^T \rangle$$

6.2 Expectation of Linear Combinations

6.2.1 Linear Forms

Assume $\mathbf{X}$ and $\mathbf{x}$ to be a matrix and a vector of random variables. Then (see See [18])

$$E[\mathbf{AXB}+\mathbf{C}] = \mathbf{A}E[\mathbf{X}]\mathbf{B}+\mathbf{C}$$

$$\mathrm{Var}[\mathbf{Ax}] = \mathbf{A}\mathrm{Var}[\mathbf{x}]\mathbf{A}^T$$

$$\mathrm{Cov}[\mathbf{Ax},\mathbf{By}] = \mathbf{A}\mathrm{Cov}[\mathbf{x},\mathbf{y}]\mathbf{B}^T$$

Assume $\mathbf{x}$ to be a stochastic vector with mean $\mathbf{m}$, then (see [7])

$$\begin{aligned} E[\mathbf{Ax}+\mathbf{b}] &= \mathbf{Am}+\mathbf{b} \\ E[\mathbf{Ax}] &= \mathbf{Am} \\ E[\mathbf{x}+\mathbf{b}] &= \mathbf{m}+\mathbf{b} \end{aligned}$$

6.2.2 Quadratic Forms

Assume $\mathbf{A}$ is symmetric, $\mathbf{c} = E[\mathbf{x}]$ and $\mathbf{\Sigma} = \mathrm{Var}[\mathbf{x}]$. Assume also that all coordinates x_i are independent, have the same central moments $\mu_1, \mu_2, \mu_3, \mu_4$ and denote $\mathbf{a} = \mathrm{diag}(\mathbf{A})$. Then (See [18])

$$E[\mathbf{x}^T\mathbf{Ax}] = \mathrm{Tr}(\mathbf{A\Sigma}) + \mathbf{c}^T\mathbf{Ac}$$

$$\mathrm{Var}[\mathbf{x}^T\mathbf{Ax}] = 2\mu_2^2\mathrm{Tr}(\mathbf{A}^2) + 4\mu_2\mathbf{c}^T\mathbf{A}^2\mathbf{c} + 4\mu_3\mathbf{c}^T\mathbf{Aa} + (\mu_4 - 3\mu_2^2)\mathbf{a}^T\mathbf{a}$$

Also, assume $\mathbf{x}$ to be a stochastic vector with mean $\mathbf{m}$, and covariance $\mathbf{M}$. Then (see [7])

$$\begin{aligned} E[(\mathbf{Ax}+\mathbf{a})(\mathbf{Bx}+\mathbf{b})^T] &= \mathbf{AMB}^T + (\mathbf{Am}+\mathbf{a})(\mathbf{Bm}+\mathbf{b})^T \\ E[\mathbf{xx}^T] &= \mathbf{M}+\mathbf{mm}^T \\ E[\mathbf{xa}^T\mathbf{x}] &= (\mathbf{M}+\mathbf{mm}^T)\mathbf{a} \\ E[\mathbf{x}^T\mathbf{ax}^T] &= \mathbf{a}^T(\mathbf{M}+\mathbf{mm}^T) \\ E[(\mathbf{Ax})(\mathbf{Ax})^T] &= \mathbf{A}(\mathbf{M}+\mathbf{mm}^T)\mathbf{A}^T \\ E[(\mathbf{x}+\mathbf{a})(\mathbf{x}+\mathbf{a})^T] &= \mathbf{M}+(\mathbf{m}+\mathbf{a})(\mathbf{m}+\mathbf{a})^T \end{aligned}$$

$$\begin{aligned} E[(\mathbf{Ax}+\mathbf{a})^T(\mathbf{Bx}+\mathbf{b})] &= \mathrm{Tr}(\mathbf{AMB}^T) + (\mathbf{Am}+\mathbf{a})^T(\mathbf{Bm}+\mathbf{b}) \\ E[\mathbf{x}^T\mathbf{x}] &= \mathrm{Tr}(\mathbf{M}) + \mathbf{m}^T\mathbf{m} \\ E[\mathbf{x}^T\mathbf{Ax}] &= \mathrm{Tr}(\mathbf{AM}) + \mathbf{m}^T\mathbf{Am} \\ E[(\mathbf{Ax})^T(\mathbf{Ax})] &= \mathrm{Tr}(\mathbf{AMA}^T) + (\mathbf{Am})^T(\mathbf{Am}) \\ E[(\mathbf{x}+\mathbf{a})^T(\mathbf{x}+\mathbf{a})] &= \mathrm{Tr}(\mathbf{M}) + (\mathbf{m}+\mathbf{a})^T(\mathbf{m}+\mathbf{a}) \end{aligned}$$

See [7].

6.2.3 Cubic Forms

Assume $\mathbf{x}$ to be a stochastic vector with independent coordinates, mean $\mathbf{m}$, covariance $\mathbf{M}$ and central moments $\mathbf{v}_3 = E[(\mathbf{x}-\mathbf{m})^3]$. Then (see [7])

$$\begin{aligned}
E[(\mathbf{Ax}+\mathbf{a})(\mathbf{Bx}+\mathbf{b})^T(\mathbf{Cx}+\mathbf{c})] &= \mathbf{A}\mathrm{diag}(\mathbf{B}^T\mathbf{C})\mathbf{v}_3 \\
&\quad +\mathrm{Tr}(\mathbf{BMC}^T)(\mathbf{Am}+\mathbf{a}) \\
&\quad +\mathbf{AMC}^T(\mathbf{Bm}+\mathbf{b}) \\
&\quad +(\mathbf{AMB}^T+(\mathbf{Am}+\mathbf{a})(\mathbf{Bm}+\mathbf{b})^T)(\mathbf{Cm}+\mathbf{c}) \\
E[\mathbf{xx}^T\mathbf{x}] &= \mathbf{v}_3 + 2\mathbf{Mm} + (\mathrm{Tr}(\mathbf{M})+\mathbf{m}^T\mathbf{m})\mathbf{m} \\
E[(\mathbf{Ax}+\mathbf{a})(\mathbf{Ax}+\mathbf{a})^T(\mathbf{Ax}+\mathbf{a})] &= \mathbf{A}\mathrm{diag}(\mathbf{A}^T\mathbf{A})\mathbf{v}_3 \\
&\quad +[2\mathbf{AMA}^T+(\mathbf{Ax}+\mathbf{a})(\mathbf{Ax}+\mathbf{a})^T](\mathbf{Am}+\mathbf{a}) \\
&\quad +\mathrm{Tr}(\mathbf{AMA}^T)(\mathbf{Am}+\mathbf{a})
\end{aligned}$$

$$\begin{aligned}
E[(\mathbf{Ax}+\mathbf{a})\mathbf{b}^T(\mathbf{Cx}+\mathbf{c})(\mathbf{Dx}+\mathbf{d})^T] &= (\mathbf{Ax}+\mathbf{a})\mathbf{b}^T(\mathbf{CMD}^T+(\mathbf{Cm}+\mathbf{c})(\mathbf{Dm}+\mathbf{d})^T) \\
&\quad +(\mathbf{AMC}^T+(\mathbf{Am}+\mathbf{a})(\mathbf{Cm}+\mathbf{c})^T)\mathbf{b}(\mathbf{Dm}+\mathbf{d})^T \\
&\quad +\mathbf{b}^T(\mathbf{Cm}+\mathbf{c})(\mathbf{AMD}^T-(\mathbf{Am}+\mathbf{a})(\mathbf{Dm}+\mathbf{d})^T)
\end{aligned}$$

6.3 Weighted Scalar Variable

Assume $\mathbf{x} \in \mathbb{R}^{n\times 1}$ is a random variable, $\mathbf{w} \in \mathbb{R}^{n\times 1}$ is a vector of constants and y is the linear combination $y = \mathbf{w}^T\mathbf{x}$. Assume further that $\mathbf{m}, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$ denotes the mean, covariance, and central third and fourth moment matrix of the variable $\mathbf{x}$. Then it holds that

$$\begin{aligned}
\langle y\rangle &= \mathbf{w}^T\mathbf{m} \\
\langle (y-\langle y\rangle)^2\rangle &= \mathbf{w}^T\mathbf{M}_2\mathbf{w} \\
\langle (y-\langle y\rangle)^3\rangle &= \mathbf{w}^T\mathbf{M}_3\mathbf{w}\otimes\mathbf{w} \\
\langle (y-\langle y\rangle)^4\rangle &= \mathbf{w}^T\mathbf{M}_4\mathbf{w}\otimes\mathbf{w}\otimes\mathbf{w}
\end{aligned}$$

7 Gaussians

7.1 Basics

7.1.1 Density and normalization

The density of $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ is

$$p(\mathbf{x}) = \frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

Note that if $\mathbf{x}$ is d-dimensional, then $\det(2\pi\mathbf{\Sigma}) = (2\pi)^d \det(\mathbf{\Sigma})$.
Integration and normalization

$$\int \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right] d\mathbf{x} = \sqrt{\det(2\pi\mathbf{\Sigma})}$$

$$\int \exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x}\right] d\mathbf{x} = \sqrt{\det(2\pi\mathbf{A}^{-1})} \exp\left[\frac{1}{2}\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}\right]$$

$$\int \exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbf{S}^T\mathbf{A}\mathbf{S}) + \mathrm{Tr}(\mathbf{B}^T\mathbf{S})\right] d\mathbf{S} = \sqrt{\det(2\pi\mathbf{A}^{-1})} \exp\left[\frac{1}{2}\mathrm{Tr}(\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})\right]$$

The derivatives of the density are

$$\frac{\partial p(\mathbf{x})}{\partial \mathbf{x}} = -p(\mathbf{x})\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})$$

$$\frac{\partial^2 p}{\partial \mathbf{x}\partial \mathbf{x}^T} = p(\mathbf{x})\left(\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\right)$$

7.1.2 Marginal Distribution

Assume $\mathbf{x} \sim \mathcal{N}_{\mathbf{x}}(\mu, \mathbf{\Sigma})$ where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix} \qquad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_a & \mathbf{\Sigma}_c \\ \mathbf{\Sigma}_c^T & \mathbf{\Sigma}_b \end{bmatrix}$$

then

$$\begin{aligned} p(\mathbf{x}_a) &= \mathcal{N}_{\mathbf{x}_a}(\boldsymbol{\mu}_a, \mathbf{\Sigma}_a) \\ p(\mathbf{x}_b) &= \mathcal{N}_{\mathbf{x}_b}(\boldsymbol{\mu}_b, \mathbf{\Sigma}_b) \end{aligned}$$

7.1.3 Conditional Distribution

Assume $\mathbf{x} \sim \mathcal{N}_{\mathbf{x}}(\mu, \mathbf{\Sigma})$ where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix} \qquad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_a & \mathbf{\Sigma}_c \\ \mathbf{\Sigma}_c^T & \mathbf{\Sigma}_b \end{bmatrix}$$

then

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}_{\mathbf{x}_a}(\hat{\mu}_a, \hat{\mathbf{\Sigma}}_a) \qquad \left\{ \begin{array}{rcl} \hat{\boldsymbol{\mu}}_a & = & \boldsymbol{\mu}_a + \mathbf{\Sigma}_c \mathbf{\Sigma}_b^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\ \hat{\mathbf{\Sigma}}_a & = & \mathbf{\Sigma}_a - \mathbf{\Sigma}_c \mathbf{\Sigma}_b^{-1} \mathbf{\Sigma}_c^T \end{array} \right.$$

$$p(\mathbf{x}_b|\mathbf{x}_a) = \mathcal{N}_{\mathbf{x}_b}(\hat{\mu}_b, \hat{\mathbf{\Sigma}}_b) \qquad \left\{ \begin{array}{rcl} \hat{\boldsymbol{\mu}}_b & = & \boldsymbol{\mu}_b + \mathbf{\Sigma}_c^T \mathbf{\Sigma}_a^{-1}(\mathbf{x}_a - \boldsymbol{\mu}_a) \\ \hat{\mathbf{\Sigma}}_b & = & \mathbf{\Sigma}_b - \mathbf{\Sigma}_c^T \mathbf{\Sigma}_a^{-1} \mathbf{\Sigma}_c \end{array} \right.$$

7.1.4 Linear combination

Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{m}_x, \mathbf{\Sigma}_x)$ and $\mathbf{y} \sim \mathcal{N}(\mathbf{m}_y, \mathbf{\Sigma}_y)$ then

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{c} \sim \mathcal{N}(\mathbf{A}\mathbf{m}_x + \mathbf{B}\mathbf{m}_y + \mathbf{c}, \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T + \mathbf{B}\mathbf{\Sigma}_y\mathbf{B}^T)$$

7.1.5 Rearranging Means

$$\mathcal{N}_{\mathbf{A}\mathbf{x}}[\mathbf{m}, \mathbf{\Sigma}] = \frac{\sqrt{\det(2\pi(\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1})}}{\sqrt{\det(2\pi\mathbf{\Sigma})}} \mathcal{N}_{\mathbf{x}}[\mathbf{A}^{-1}\mathbf{m}, (\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}]$$

7.1.6 Rearranging into squared form

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

$$-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} = -\frac{1}{2}(\mathbf{x} - \mathbf{A}^{-1}\mathbf{b})^T\mathbf{A}(\mathbf{x} - \mathbf{A}^{-1}\mathbf{b}) + \frac{1}{2}\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}$$

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

7.1.7 Sum of two squared forms

In vector formulation (assuming $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ are symmetric)

$$\begin{array}{rl} & -\frac{1}{2}(\mathbf{x} - \mathbf{m}_1)^T\mathbf{\Sigma}_1^{-1}(\mathbf{x} - \mathbf{m}_1) \\ & -\frac{1}{2}(\mathbf{x} - \mathbf{m}_2)^T\mathbf{\Sigma}_2^{-1}(\mathbf{x} - \mathbf{m}_2) \\ = & -\frac{1}{2}(\mathbf{x} - \mathbf{m}_c)^T\mathbf{\Sigma}_c^{-1}(\mathbf{x} - \mathbf{m}_c) + C \end{array}$$

$$\begin{array}{rcl} \mathbf{\Sigma}_c^{-1} & = & \mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1} \\ \mathbf{m}_c & = & (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2) \\ C & = & \frac{1}{2}(\mathbf{m}_1^T\mathbf{\Sigma}_1^{-1} + \mathbf{m}_2^T\mathbf{\Sigma}_2^{-1})(\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2) \\ & & -\frac{1}{2}\left(\mathbf{m}_1^T\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{m}_2^T\mathbf{\Sigma}_2^{-1}\mathbf{m}_2\right) \end{array}$$

In a trace formulation (assuming $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ are symmetric)

$$\begin{aligned} & -\frac{1}{2}\mathrm{Tr}((\mathbf{X}-\mathbf{M}_1)^T\mathbf{\Sigma}_1^{-1}(\mathbf{X}-\mathbf{M}_1)) \\ & -\frac{1}{2}\mathrm{Tr}((\mathbf{X}-\mathbf{M}_2)^T\mathbf{\Sigma}_2^{-1}(\mathbf{X}-\mathbf{M}_2)) \\ = & -\frac{1}{2}\mathrm{Tr}[(\mathbf{X}-\mathbf{M}_c)^T\mathbf{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M}_c)] + C \end{aligned}$$

$$\begin{aligned} \mathbf{\Sigma}_c^{-1} &= \mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1} \\ \mathbf{M}_c &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{M}_2) \\ C &= \frac{1}{2}\mathrm{Tr}\Big[(\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{M}_2)^T(\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{M}_2)\Big] \\ & \quad -\frac{1}{2}\mathrm{Tr}(\mathbf{M}_1^T\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{M}_2^T\mathbf{\Sigma}_2^{-1}\mathbf{M}_2) \end{aligned}$$

7.1.8 Product of gaussian densities

Let $\mathcal{N}_\mathbf{x}(\mathbf{m}, \mathbf{\Sigma})$ denote a density of $\mathbf{x}$, then

$$\mathcal{N}_\mathbf{x}(\mathbf{m}_1, \mathbf{\Sigma}_1) \cdot \mathcal{N}_\mathbf{x}(\mathbf{m}_2, \mathbf{\Sigma}_2) = c_c \mathcal{N}_\mathbf{x}(\mathbf{m}_c, \mathbf{\Sigma}_c)$$

$$\begin{aligned} c_c &= \mathcal{N}_{\mathbf{m}_1}(\mathbf{m}_2, (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)) \\ &= \frac{1}{\sqrt{\det(2\pi(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2))}} \exp\left[-\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2)\right] \\ \mathbf{m}_c &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2) \\ \mathbf{\Sigma}_c &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1} \end{aligned}$$

but note that the product is not normalized as a density of $\mathbf{x}$.

7.2 Moments

7.2.1 Mean and covariance of linear forms

First and second moments. Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$

$$E(\mathbf{x}) = \mathbf{m}$$

$$\mathrm{Cov}(\mathbf{x}, \mathbf{x}) = \mathrm{Var}(\mathbf{x}) = \mathbf{\Sigma} = E(\mathbf{x}\mathbf{x}^T) - E(\mathbf{x})E(\mathbf{x}^T) = E(\mathbf{x}\mathbf{x}^T) - \mathbf{m}\mathbf{m}^T$$

As for any other distribution is holds for gaussians that

$$E[\mathbf{A}\mathbf{x}] = \mathbf{A}E[\mathbf{x}]$$

$$\mathrm{Var}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathrm{Var}[\mathbf{x}]\mathbf{A}^T$$

$$\mathrm{Cov}[\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}] = \mathbf{A}\mathrm{Cov}[\mathbf{x}, \mathbf{y}]\mathbf{B}^T$$

7.2.2 Mean and variance of square forms

Mean and variance of square forms: Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$

$$\begin{aligned}
E(\mathbf{x}\mathbf{x}^T) &= \mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T \\
E[\mathbf{x}^T\mathbf{A}\mathbf{x}] &= \mathrm{Tr}(\mathbf{A}\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{A}\mathbf{m} \\
\mathrm{Var}(\mathbf{x}^T\mathbf{A}\mathbf{x}) &= 2\sigma^4\mathrm{Tr}(\mathbf{A}^2) + 4\sigma^2\mathbf{m}^T\mathbf{A}^2\mathbf{m} \\
E[(\mathbf{x}-\mathbf{m}')^T\mathbf{A}(\mathbf{x}-\mathbf{m}')] &= (\mathbf{m}-\mathbf{m}')^T\mathbf{A}(\mathbf{m}-\mathbf{m}') + \mathrm{Tr}(\mathbf{A}\mathbf{\Sigma})
\end{aligned}$$

Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and $\mathbf{A}$ and $\mathbf{B}$ to be symmetric, then

$$\mathrm{Cov}(\mathbf{x}^T\mathbf{A}\mathbf{x}, \mathbf{x}^T\mathbf{B}\mathbf{x}) = 2\sigma^4\mathrm{Tr}(\mathbf{A}\mathbf{B})$$

7.2.3 Cubic forms

$$\begin{aligned}
E[\mathbf{x}\mathbf{b}^T\mathbf{x}\mathbf{x}^T] &= \mathbf{m}\mathbf{b}^T(\mathbf{M} + \mathbf{m}\mathbf{m}^T) + (\mathbf{M} + \mathbf{m}\mathbf{m}^T)\mathbf{b}\mathbf{m}^T \\
&\quad + \mathbf{b}^T\mathbf{m}(\mathbf{M} - \mathbf{m}\mathbf{m}^T)
\end{aligned}$$

7.2.4 Mean of Quartic Forms

$$\begin{aligned}
E[\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{x}^T] &= 2(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)^2 + \mathbf{m}^T\mathbf{m}(\mathbf{\Sigma} - \mathbf{m}\mathbf{m}^T) \\
&\quad + \mathrm{Tr}(\mathbf{\Sigma})(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T) \\
E[\mathbf{x}\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T] &= (\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)(\mathbf{A} + \mathbf{A}^T)(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T) \\
&\quad + \mathbf{m}^T\mathbf{A}\mathbf{m}(\mathbf{\Sigma} - \mathbf{m}\mathbf{m}^T) + \mathrm{Tr}[\mathbf{A}\mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)] \\
E[\mathbf{x}^T\mathbf{x}\mathbf{x}^T\mathbf{x}] &= 2\mathrm{Tr}(\mathbf{\Sigma}^2) + 4\mathbf{m}^T\mathbf{\Sigma}\mathbf{m} + (\mathrm{Tr}(\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{m})^2 \\
E[\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T\mathbf{B}\mathbf{x}] &= \mathrm{Tr}[\mathbf{A}\mathbf{\Sigma}(\mathbf{B} + \mathbf{B}^T)\mathbf{\Sigma}] + \mathbf{m}^T(\mathbf{A} + \mathbf{A}^T)\mathbf{\Sigma}(\mathbf{B} + \mathbf{B}^T)\mathbf{m} \\
&\quad + (\mathrm{Tr}(\mathbf{A}\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{A}\mathbf{m})(\mathrm{Tr}(\mathbf{B}\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{B}\mathbf{m})
\end{aligned}$$

$$\begin{aligned}
& E[\mathbf{a}^T\mathbf{x}\mathbf{b}^T\mathbf{x}\mathbf{c}^T\mathbf{x}\mathbf{d}^T\mathbf{x}] \\
= \; & (\mathbf{a}^T(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)\mathbf{b})(\mathbf{c}^T(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)\mathbf{d}) \\
& + (\mathbf{a}^T(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)\mathbf{c})(\mathbf{b}^T(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)\mathbf{d}) \\
& + (\mathbf{a}^T(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)\mathbf{d})(\mathbf{b}^T(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T)\mathbf{c}) - 2\mathbf{a}^T\mathbf{m}\mathbf{b}^T\mathbf{m}\mathbf{c}^T\mathbf{m}\mathbf{d}^T\mathbf{m}
\end{aligned}$$

$$\begin{aligned}
& E[(\mathbf{A}\mathbf{x} + \mathbf{a})(\mathbf{B}\mathbf{x} + \mathbf{b})^T(\mathbf{C}\mathbf{x} + \mathbf{c})(\mathbf{D}\mathbf{x} + \mathbf{d})^T] \\
= \; & [\mathbf{A}\mathbf{\Sigma}\mathbf{B}^T + (\mathbf{A}\mathbf{m} + \mathbf{a})(\mathbf{B}\mathbf{m} + \mathbf{b})^T][\mathbf{C}\mathbf{\Sigma}\mathbf{D}^T + (\mathbf{C}\mathbf{m} + \mathbf{c})(\mathbf{D}\mathbf{m} + \mathbf{d})^T] \\
& + [\mathbf{A}\mathbf{\Sigma}\mathbf{C}^T + (\mathbf{A}\mathbf{m} + \mathbf{a})(\mathbf{C}\mathbf{m} + \mathbf{c})^T][\mathbf{B}\mathbf{\Sigma}\mathbf{D}^T + (\mathbf{B}\mathbf{m} + \mathbf{b})(\mathbf{D}\mathbf{m} + \mathbf{d})^T] \\
& + (\mathbf{B}\mathbf{m} + \mathbf{b})^T(\mathbf{C}\mathbf{m} + \mathbf{c})[\mathbf{A}\mathbf{\Sigma}\mathbf{D}^T - (\mathbf{A}\mathbf{m} + \mathbf{a})(\mathbf{D}\mathbf{m} + \mathbf{d})^T] \\
& + \mathrm{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{C}^T)[\mathbf{A}\mathbf{\Sigma}\mathbf{D}^T + (\mathbf{A}\mathbf{m} + \mathbf{a})(\mathbf{D}\mathbf{m} + \mathbf{d})^T]
\end{aligned}$$

$$
\begin{aligned}
& E[(\mathbf{A}\mathbf{x}+\mathbf{a})^T(\mathbf{B}\mathbf{x}+\mathbf{b})(\mathbf{C}\mathbf{x}+\mathbf{c})^T(\mathbf{D}\mathbf{x}+\mathbf{d})] \\
= \; & \mathrm{Tr}[\mathbf{A}\boldsymbol{\Sigma}(\mathbf{C}^T\mathbf{D}+\mathbf{D}^T\mathbf{C})\boldsymbol{\Sigma}\mathbf{B}^T] \\
& +[(\mathbf{A}\mathbf{m}+\mathbf{a})^T\mathbf{B}+(\mathbf{B}\mathbf{m}+\mathbf{b})^T\mathbf{A}]\boldsymbol{\Sigma}[\mathbf{C}^T(\mathbf{D}\mathbf{m}+\mathbf{d})+\mathbf{D}^T(\mathbf{C}\mathbf{m}+\mathbf{c})] \\
& +[\mathrm{Tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^T)+(\mathbf{A}\mathbf{m}+\mathbf{a})^T(\mathbf{B}\mathbf{m}+\mathbf{b})][\mathrm{Tr}(\mathbf{C}\boldsymbol{\Sigma}\mathbf{D}^T)+(\mathbf{C}\mathbf{m}+\mathbf{c})^T(\mathbf{D}\mathbf{m}+\mathbf{d})]
\end{aligned}
$$

See [7].

7.2.5 Moments

$$E[\mathbf{x}] = \sum_k \rho_k \mathbf{m}_k$$

$$\mathrm{Cov}(\mathbf{x}) = \sum_k \sum_{k'} \rho_k \rho_{k'} (\boldsymbol{\Sigma}_k + \mathbf{m}_k \mathbf{m}_k^T - \mathbf{m}_k \mathbf{m}_{k'}^T)$$

7.3 Miscellaneous

7.3.1 Whitening

Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ then

$$\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\mathbf{m}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Conversely having $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ one can generate data $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ by setting

$$\mathbf{x} = \boldsymbol{\Sigma}^{1/2}\mathbf{z} + \mathbf{m} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$$

Note that $\boldsymbol{\Sigma}^{1/2}$ means the matrix which fulfils $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$, and that it exists and is unique since $\boldsymbol{\Sigma}$ is positive definite.

7.3.2 The Chi-Square connection

Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ and $\mathbf{x}$ to be n dimensional, then

$$z = (\mathbf{x}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\mathbf{m}) \sim \chi^2_n$$

where χ^2_n denotes the Chi square distribution with n degrees of freedom.

7.3.3 Entropy

Entropy of a D-dimensional gaussian

$$H(\mathbf{x}) = -\int \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma}) \ln \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma}) d\mathbf{x} = \ln \sqrt{\det(2\pi\boldsymbol{\Sigma})} + \frac{D}{2}$$

7.4 Mixture of Gaussians

7.4.1 Density

The variable $\mathbf{x}$ is distributed as a mixture of gaussians if it has the density

$$p(\mathbf{x}) = \sum_{k=1}^{K} \rho_k \frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma}_k)}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1}(\mathbf{x}-\mathbf{m}_k)\right]$$

where ρ_k sum to 1 and the $\mathbf{\Sigma}_k$ all are positive definite.

7.4.2 Derivatives

Defining $p(\mathbf{s}) = \sum_k \rho_k \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$ one get

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{s})}{\partial \rho_j} &= \frac{\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)} \frac{\partial}{\partial \rho_j} \ln[\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)] \\
&= \frac{\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)} \frac{1}{\rho_j} \\
\frac{\partial \ln p(\mathbf{s})}{\partial \boldsymbol{\mu}_j} &= \frac{\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\mu}_j} \ln[\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)] \\
&= \frac{\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)} \left[-\mathbf{\Sigma}_k^{-1}(\mathbf{s}-\boldsymbol{\mu}_k)\right] \\
\frac{\partial \ln p(\mathbf{s})}{\partial \mathbf{\Sigma}_j} &= \frac{\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)} \frac{\partial}{\partial \mathbf{\Sigma}_j} \ln[\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)] \\
&= \frac{\rho_j \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_{\mathbf{s}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)} \frac{1}{2} \left[\mathbf{\Sigma}_j^{-T} + \mathbf{\Sigma}_j^{-T}(\mathbf{s}-\boldsymbol{\mu}_j)(\mathbf{s}-\boldsymbol{\mu}_j)^T \mathbf{\Sigma}_j^{-T}\right]
\end{aligned}$$

But ρ_k and $\mathbf{\Sigma}_k$ needs to be constrained.

8 Special Matrices

8.1 Units, Permutation and Shift

8.1.1 Unit vector

Let $\mathbf{e}_i \in \mathbb{R}^{n\times 1}$ be the ith unit vector, i.e. the vector which is zero in all entries except the ith at which it is 1.

8.1.2 Rows and Columns

$$\begin{aligned} i.th \text{ row of } \mathbf{A} &= \mathbf{e}_i^T\mathbf{A} \\ j.th \text{ column of } \mathbf{A} &= \mathbf{A}\mathbf{e}_j \end{aligned}$$

8.1.3 Permutations

Let $\mathbf{P}$ be some permutation matrix, e.g.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_2 & \mathbf{e}_1 & \mathbf{e}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_2^T \\ \mathbf{e}_1^T \\ \mathbf{e}_3^T \end{bmatrix}$$

For permutation matrices it holds that

$$\mathbf{P}\mathbf{P}^T = \mathbf{I}$$

and that

$$\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{A}\mathbf{e}_2 & \mathbf{A}\mathbf{e}_1 & \mathbf{A}\mathbf{e}_3 \end{bmatrix} \qquad \mathbf{P}\mathbf{A} = \begin{bmatrix} \mathbf{e}_2^T\mathbf{A} \\ \mathbf{e}_1^T\mathbf{A} \\ \mathbf{e}_3^T\mathbf{A} \end{bmatrix}$$

That is, the first is a matrix which has columns of $\mathbf{A}$ but in permuted sequence and the second is a matrix which has the rows of $\mathbf{A}$ but in the permuted sequence.

8.1.4 Translation, Shift or Lag Operators

Let $\mathbf{L}$ denote the lag (or 'translation' or 'shift') operator defined on a 4×4 example by

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

i.e. a matrix of zeros with one on the sub-diagonal, $(\mathbf{L})_{ij} = \delta_{i,j+1}$. With some signal x_t for $t = 1, ..., N$, the n.th power of the lag operator shifts the indices, i.e.

$$(\mathbf{L}^n\mathbf{x})_t = \begin{cases} 0 & \text{for} \quad t = 1, .., n \\ x_{t-n} & \text{for} \quad t = n+1, ..., N \end{cases}$$

A related but slightly different matrix is the 'recurrent shifted' operator defined on a 4x4 example by

$$\hat{\mathbf{L}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

i.e. a matrix defined by $(\hat{\mathbf{L}})_{ij} = \delta_{i,j+1} + \delta_{i,1}\delta_{j,dim(\mathbf{L})}$. On a signal $\mathbf{x}$ it has the effect

$$(\hat{\mathbf{L}}^n \mathbf{x})_t = x_{t'}, \quad t' = [(t-n) \mod N] + 1$$

That is, $\hat{\mathbf{L}}$ is like the shift operator $\mathbf{L}$ except that it 'wraps' the signal as if it was periodic and shifted (substituting the zeros with the rear end of the signal). Note that $\hat{\mathbf{L}}$ is invertible and orthogonal, i.e.

$$\hat{\mathbf{L}}^{-1} = \hat{\mathbf{L}}^T$$

8.2 The Singleentry Matrix

8.2.1 Definition

The single-entry matrix $\mathbf{J}^{ij} \in \mathbb{R}^{n \times n}$ is defined as the matrix which is zero everywhere except in the entry (i, j) in which it is 1. In a 4×4 example one might have

$$\mathbf{J}^{23} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The single-entry matrix is very useful when working with derivatives of expressions involving matrices.

8.2.2 Swap and Zeros

Assume $\mathbf{A}$ to be $n \times m$ and $\mathbf{J}^{ij}$ to be $m \times p$

$$\mathbf{A}\mathbf{J}^{ij} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A}_i & \ldots & \mathbf{0} \end{bmatrix}$$

i.e. an $n \times p$ matrix of zeros with the i.th column of $\mathbf{A}$ in place of the j.th column. Assume $\mathbf{A}$ to be $n \times m$ and $\mathbf{J}^{ij}$ to be $p \times n$

$$\mathbf{J}^{ij}\mathbf{A} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{A}_j \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

i.e. an $p \times m$ matrix of zeros with the j.th row of $\mathbf{A}$ in the placed of the i.th row.

8.2.3 Rewriting product of elements

$$A_{ki}B_{jl} = (\mathbf{A}\mathbf{e}_i\mathbf{e}_j^T\mathbf{B})_{kl} = (\mathbf{A}\mathbf{J}^{ij}\mathbf{B})_{kl}$$
$$A_{ik}B_{lj} = (\mathbf{A}^T\mathbf{e}_i\mathbf{e}_j^T\mathbf{B}^T)_{kl} = (\mathbf{A}^T\mathbf{J}^{ij}\mathbf{B}^T)_{kl}$$
$$A_{ik}B_{jl} = (\mathbf{A}^T\mathbf{e}_i\mathbf{e}_j^T\mathbf{B})_{kl} = (\mathbf{A}^T\mathbf{J}^{ij}\mathbf{B})_{kl}$$
$$A_{ki}B_{lj} = (\mathbf{A}\mathbf{e}_i\mathbf{e}_j^T\mathbf{B}^T)_{kl} = (\mathbf{A}\mathbf{J}^{ij}\mathbf{B}^T)_{kl}$$

8.2.4 Properties of the Singleentry Matrix

If $i = j$

$$\mathbf{J}^{ij}\mathbf{J}^{ij} = \mathbf{J}^{ij} \qquad (\mathbf{J}^{ij})^T(\mathbf{J}^{ij})^T = \mathbf{J}^{ij}$$
$$\mathbf{J}^{ij}(\mathbf{J}^{ij})^T = \mathbf{J}^{ij} \qquad (\mathbf{J}^{ij})^T\mathbf{J}^{ij} = \mathbf{J}^{ij}$$

If $i \neq j$

$$\mathbf{J}^{ij}\mathbf{J}^{ij} = \mathbf{0} \qquad (\mathbf{J}^{ij})^T(\mathbf{J}^{ij})^T = \mathbf{0}$$
$$\mathbf{J}^{ij}(\mathbf{J}^{ij})^T = \mathbf{J}^{ii} \qquad (\mathbf{J}^{ij})^T\mathbf{J}^{ij} = \mathbf{J}^{jj}$$

8.2.5 The Singleentry Matrix in Scalar Expressions

Assume $\mathbf{A}$ is $n \times m$ and $\mathbf{J}$ is $m \times n$, then

$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ij}) = \mathrm{Tr}(\mathbf{J}^{ij}\mathbf{A}) = (\mathbf{A}^T)_{ij}$$

Assume $\mathbf{A}$ is $n \times n$, $\mathbf{J}$ is $n \times m$ and $\mathbf{B}$ is $m \times n$, then

$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ij}\mathbf{B}) = (\mathbf{A}^T\mathbf{B}^T)_{ij}$$
$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ji}\mathbf{B}) = (\mathbf{B}\mathbf{A})_{ij}$$
$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ij}\mathbf{J}^{ij}\mathbf{B}) = \mathrm{diag}(\mathbf{A}^T\mathbf{B}^T)_{ij}$$

Assume $\mathbf{A}$ is $n \times n$, $\mathbf{J}^{ij}$ is $n \times m$ $\mathbf{B}$ is $m \times n$, then

$$\mathbf{x}^T\mathbf{A}\mathbf{J}^{ij}\mathbf{B}\mathbf{x} = (\mathbf{A}^T\mathbf{x}\mathbf{x}^T\mathbf{B}^T)_{ij}$$
$$\mathbf{x}^T\mathbf{A}\mathbf{J}^{ij}\mathbf{J}^{ij}\mathbf{B}\mathbf{x} = \mathrm{diag}(\mathbf{A}^T\mathbf{x}\mathbf{x}^T\mathbf{B}^T)_{ij}$$

8.2.6 Structure Matrices

The structure matrix is defined by

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

If $\mathbf{A}$ has no special structure then

$$\mathbf{S}^{ij} = \mathbf{J}^{ij}$$

If $\mathbf{A}$ is symmetric then

$$\mathbf{S}^{ij} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij}\mathbf{J}^{ij}$$

8.3 Symmetric and Antisymmetric

8.3.1 Symmetric

The matrix $\mathbf{A}$ is said to be *symmetric* if

$$\mathbf{A} = \mathbf{A}^T$$

Symmetric matrices have many important properties, e.g. that their eigenvalues are real and eigenvalues orthogonal.

8.3.2 Antisymmetric

The *antisymmetric* matrix is also known as the *skew* symmetric matrix. It has the following property from which it is defined

$$\mathbf{A} = -\mathbf{A}^T$$

Hereby, it can be seen that the antisymmetric matrices always have a zero diagonal. The $n \times n$ antisymmetric matrices also have the following properties.

$$\begin{aligned} \det(\mathbf{A}^T) &= \det(-\mathbf{A}) = (-1)^n \det(\mathbf{A}) \\ -\det(\mathbf{A}) &= \det(-\mathbf{A}) = 0, \quad \text{if } n \text{ is odd} \end{aligned}$$

8.4 Vandermonde Matrices

A Vandermonde matrix has the form [12]

$$\mathbf{V} = \begin{bmatrix} 1 & v_1 & v_1^2 & \cdots & v_1^{n-1} \\ 1 & v_2 & v_2^2 & \cdots & v_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & v_n & v_n^2 & \cdots & v_n^{n-1} \end{bmatrix}. \tag{59}$$

The transpose of $\mathbf{V}$ is also said to a Vandermonde matrix. The determinant is given by [21]

$$\det \mathbf{V} = \prod_i > j(v_i - v_j) \tag{60}$$

8.5 Toeplitz Matrices

A Toeplitz matrix $\mathbf{T}$ is a matrix where the elements of each diagonal is the same. In the $n \times n$ square case, it has the following structure:

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{12} \\ t_{n1} & \cdots & t_{21} & t_{11} \end{bmatrix} = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ t_{-(n-1)} & \cdots & t_{-1} & t_0 \end{bmatrix} \quad (61)$$

A Toeplitz matrix is *persymmetric*. If a matrix is persymmetric (or orthosymmetric), it means that the matrix is symmetric about its northeast-southwest diagonal (anti-diagonal) [9]. Persymmetric matrices is a larger class of matrices, since a persymmetric matrix not necessarily has a Toeplitz structure. There are some special cases of Toeplitz matrices. The *symmetric* Toeplitz matrix is given by:

$$\mathbf{T} = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ t_{-(n-1)} & \cdots & t_1 & t_0 \end{bmatrix} \quad (62)$$

The circular Toeplitz matrix:

$$\mathbf{T}_C = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_n & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ t_1 & \cdots & t_{n-1} & t_0 \end{bmatrix} \quad (63)$$

The upper triangular Toeplitz matrix:

$$\mathbf{T}_U = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ 0 & \cdots & 0 & t_0 \end{bmatrix}, \quad (64)$$

and the lower triangular Toeplitz matrix:

$$\mathbf{T}_L = \begin{bmatrix} t_0 & 0 & \cdots & 0 \\ t_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ t_{-(n-1)} & \cdots & t_{-1} & t_0 \end{bmatrix} \quad (65)$$

8.5.1 Properties of Toeplitz Matrices

The Toeplitz matrix has some computational advantages. The addition of two Toeplitz matrices can be done with $\mathcal{O}(n)$ flops, multiplication of two Toeplitz matrices can be done in $\mathcal{O}(n \ln n)$ flops. Toeplitz equation systems can be solved in $\mathcal{O}(n^2)$ flops. The inverse of a positive definite Toeplitz matrix can be found in $\mathcal{O}(n^2)$ flops too. The inverse of a Toeplitz matrix is persymmetric. The product of two lower triangular Toeplitz matrices is a Toeplitz matrix. More information on Toeplitz matrices and circulant matrices can be found in [10, 7].

8.6 The DFT Matrix

The DFT matrix is an $N \times N$ symmetric matrix $\mathbf{W}_N$, where the k,nth element is given by

$$W_N^{kn} = e^{\frac{-j2\pi kn}{N}} \tag{66}$$

Thus the discrete Fourier transform (DFT) can be expressed as

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn}. \tag{67}$$

Likewise the inverse discrete Fourier transform (IDFT) can be expressed as

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W_N^{-kn}. \tag{68}$$

The DFT of the vector $\mathbf{x} = [x(0), x(1), \cdots, x(N-1)]^T$ can be written in matrix form as

$$\mathbf{X} = \mathbf{W}_N \mathbf{x}, \tag{69}$$

where $\mathbf{X} = [X(0), X(1), \cdots, x(N-1)]^T$. The IDFT is similarly given as

$$\mathbf{x} = \mathbf{W}_N^{-1} \mathbf{X}. \tag{70}$$

Some properties of $\mathbf{W}_N$ exist:

$$\mathbf{W}_N^{-1} = \frac{1}{N} \mathbf{W}_N^* \tag{71}$$

$$\mathbf{W}_N \mathbf{W}_N^* = N\mathbf{I} \tag{72}$$

$$\mathbf{W}_N^* = \mathbf{W}_N^H \tag{73}$$

If $W_N = e^{\frac{-j2\pi}{N}}$, then [15]

$$W_N^{m+N/2} = -W_N^m \tag{74}$$

Notice, the DFT matrix is a Vandermonde Matrix.

The following important relation between the circulant matrix and the discrete Fourier transform (DFT) exists

$$\mathbf{T}_C = \mathbf{W}_N^{-1} (\mathbf{I} \circ (\mathbf{W}_N \mathbf{t})) \mathbf{W}_N, \tag{75}$$

where $\mathbf{t} = [t_0, t_1, \cdots, t_{n-1}]^T$ is the first row of $\mathbf{T}_C$.

8.7 Positive Definite and Semi-definite Matrices

8.7.1 Definitions

A matrix $\mathbf{A}$ is positive definite if and only if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > \mathbf{0}, \quad \forall \mathbf{x}$$

A matrix $\mathbf{A}$ is positive semi-definite if and only if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \geq \mathbf{0}, \quad \forall \mathbf{x}$$

Note that if $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is also positive semi-definite.

8.7.2 Eigenvalues

The following holds with respect to the eigenvalues:

$$\begin{array}{rcl} \mathbf{A} \text{ pos. def.} & \Leftrightarrow & \text{eig}(\mathbf{A}) > 0 \\ \mathbf{A} \text{ pos. semi-def.} & \Leftrightarrow & \text{eig}(\mathbf{A}) \geq 0 \end{array}$$

8.7.3 Trace

The following holds with respect to the trace:

$$\begin{array}{rcl} \mathbf{A} \text{ pos. def.} & \Rightarrow & \text{Tr}(\mathbf{A}) > 0 \\ \mathbf{A} \text{ pos. semi-def.} & \Rightarrow & \text{Tr}(\mathbf{A}) \geq 0 \end{array}$$

8.7.4 Inverse

If $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is invertible and $\mathbf{A}^{-1}$ is also positive definite.

8.7.5 Diagonal

If $\mathbf{A}$ is positive definite, then $A_{ii} > 0, \forall i$

8.7.6 Decomposition I

The matrix $\mathbf{A}$ is positive semi-definite of rank r $\Leftrightarrow$ there exists a matrix $\mathbf{B}$ of rank r such that $\mathbf{A} = \mathbf{B}\mathbf{B}^T$

The matrix $\mathbf{A}$ is positive definite $\Leftrightarrow$ there exists an invertible matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{B}\mathbf{B}^T$

8.7.7 Decomposition II

Assume $\mathbf{A}$ is an $n \times n$ positive semi-definite, then there exists an $n \times r$ matrix $\mathbf{B}$ of rank r such that $\mathbf{B}^T\mathbf{A}\mathbf{B} = \mathbf{I}$.

8.7.8 Equation with zeros

Assume $\mathbf{A}$ is positive semi-definite, then $\mathbf{X}^T\mathbf{A}\mathbf{X} = \mathbf{0} \quad \Rightarrow \quad \mathbf{A}\mathbf{X} = \mathbf{0}$

8.7.9 Rank of product

Assume $\mathbf{A}$ is positive definite, then $\text{rank}(\mathbf{B}\mathbf{A}\mathbf{B}^T) = \text{rank}(\mathbf{B})$

8.7.10 Positive definite property

If $\mathbf{A}$ is $n \times n$ positive definite and $\mathbf{B}$ is $r \times n$ of rank r, then $\mathbf{B}\mathbf{A}\mathbf{B}^T$ is positive definite.

8.7.11 Outer Product

If $\mathbf{X}$ is $n \times r$, where $n \leq r$ and $\text{rank}(\mathbf{X}) = n$, then $\mathbf{X}\mathbf{X}^T$ is positive definite.

8.7.12 Small pertubations

If $\mathbf{A}$ is positive definite and $\mathbf{B}$ is symmetric, then $\mathbf{A} - t\mathbf{B}$ is positive definite for sufficiently small t.

8.8 Block matrices

Let $\mathbf{A}_{ij}$ denote the ijth block of $\mathbf{A}$.

8.8.1 Multiplication

Assuming the dimensions of the blocks matches we have

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \hline \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{array}\right]$$

8.8.2 The Determinant

The determinant can be expressed as by the use of

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\ \mathbf{C}_2 &= \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{aligned}$$

as

$$\det\left(\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]\right) = \det(\mathbf{A}_{22}) \cdot \det(\mathbf{C}_1) = \det(\mathbf{A}_{11}) \cdot \det(\mathbf{C}_2)$$

8.8.3 The Inverse

The inverse can be expressed as by the use of

$$\begin{aligned}
\mathbf{C}_1 &= \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\
\mathbf{C}_2 &= \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}
\end{aligned}$$

as

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \mathbf{C}_1^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1} \\ \hline -\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{C}_2^{-1} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{C}_1^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \hline -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{C}_1^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{C}_1^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{array}\right]$$

8.8.4 Block diagonal

For block diagonal matrices we have

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} (\mathbf{A}_{11})^{-1} & \mathbf{0} \\ \hline \mathbf{0} & (\mathbf{A}_{22})^{-1} \end{array}\right]$$

$$\det\left(\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_{22} \end{array}\right]\right) = \det(\mathbf{A}_{11}) \cdot \det(\mathbf{A}_{22})$$

9 Functions and Operators

9.1 Functions and Series

9.1.1 Finite Series

$$(\mathbf{X}^n - \mathbf{I})(\mathbf{X} - \mathbf{I})^{-1} = \mathbf{I} + \mathbf{X} + \mathbf{X}^2 + ... + \mathbf{X}^{n-1}$$

9.1.2 Taylor Expansion of Scalar Function

Consider some scalar function $f(\mathbf{x})$ which takes the vector $\mathbf{x}$ as an argument. This we can Taylor expand around $\mathbf{x}_0$

$$f(\mathbf{x}) \cong f(\mathbf{x}_0) + \mathbf{g}(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T\mathbf{H}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

where

$$\mathbf{g}(\mathbf{x}_0) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\Big|_{\mathbf{x}_0} \qquad \mathbf{H}(\mathbf{x}_0) = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T}\Big|_{\mathbf{x}_0}$$

9.1.3 Matrix Functions by Infinite Series

As for analytical functions in one dimension, one can define a matrix function for square matrices $\mathbf{X}$ by an infinite series

$$\mathbf{f}(\mathbf{X}) = \sum_{n=0}^{\infty} c_n \mathbf{X}^n$$

assuming the limit exists and is finite. If the coefficients c_n fulfils $\sum_n c_n x^n < \infty$, then one can prove that the above series exists and is finite, see [1]. Thus for any analytical function $f(x)$ there exists a corresponding matrix function $\mathbf{f}(\mathbf{x})$ constructed by the Taylor expansion. Using this one can prove the following results:

1) A matrix $\mathbf{A}$ is a zero of its own characteristic polynomium [1]:

$$p(\lambda) = \det(\mathbf{I}\lambda - \mathbf{A}) = \sum_n c_n \lambda^n \qquad \Rightarrow \qquad p(\mathbf{A}) = \mathbf{0}$$

2) If $\mathbf{A}$ is square it holds that [1]

$$\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^{-1} \quad \Rightarrow \quad \mathbf{f}(\mathbf{A}) = \mathbf{U}\mathbf{f}(\mathbf{B})\mathbf{U}^{-1}$$

3) A useful fact when using power series is that

$$\mathbf{A}^n \to \mathbf{0} \text{for} n \to \infty \qquad \text{if} \qquad |\mathbf{A}| < 1$$

9.1.4 Exponential Matrix Function

In analogy to the ordinary scalar exponential function, one can define exponential and logarithmic matrix functions:

$$
\begin{aligned}
e^{\mathbf{A}} &\equiv \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n = \mathbf{I} + \mathbf{A} + \frac{1}{2}\mathbf{A}^2 + ... \\
e^{-\mathbf{A}} &\equiv \sum_{n=0}^{\infty} \frac{1}{n!}(-1)^n\mathbf{A}^n = \mathbf{I} - \mathbf{A} + \frac{1}{2}\mathbf{A}^2 - ... \\
e^{t\mathbf{A}} &\equiv \sum_{n=0}^{\infty} \frac{1}{n!}(t\mathbf{A})^n = \mathbf{I} + t\mathbf{A} + \frac{1}{2}t^2\mathbf{A}^2 + ... \\
\ln(\mathbf{I}+\mathbf{A}) &\equiv \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\mathbf{A}^n = \mathbf{A} - \frac{1}{2}\mathbf{A}^2 + \frac{1}{3}\mathbf{A}^3 - ...
\end{aligned}
$$

Some of the properties of the exponential function are [1]

$$
\begin{gathered}
e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{A}+\mathbf{B}} \qquad \text{if} \qquad \mathbf{AB} = \mathbf{BA} \\
(e^{\mathbf{A}})^{-1} = e^{-\mathbf{A}} \\
\frac{d}{dt}e^{t\mathbf{A}} = \mathbf{A}e^{t\mathbf{A}} = e^{t\mathbf{A}}\mathbf{A}, \qquad t \in \mathbb{R} \\
\frac{d}{dt}\mathrm{Tr}(e^{t\mathbf{A}}) = \mathrm{Tr}(\mathbf{A}e^{t\mathbf{A}}) \\
\det(e^{\mathbf{A}}) = e^{\mathrm{Tr}(\mathbf{A})}
\end{gathered}
$$

9.1.5 Trigonometric Functions

$$
\begin{aligned}
\sin(\mathbf{A}) &\equiv \sum_{n=0}^{\infty} \frac{(-1)^n\mathbf{A}^{2n+1}}{(2n+1)!} = \mathbf{A} - \frac{1}{3!}\mathbf{A}^3 + \frac{1}{5!}\mathbf{A}^5 - ... \\
\cos(\mathbf{A}) &\equiv \sum_{n=0}^{\infty} \frac{(-1)^n\mathbf{A}^{2n}}{(2n)!} = \mathbf{I} - \frac{1}{2!}\mathbf{A}^2 + \frac{1}{4!}\mathbf{A}^4 - ...
\end{aligned}
$$

9.2 Kronecker and Vec Operator

9.2.1 The Kronecker Product

The Kronecker product of an $m \times n$ matrix $\mathbf{A}$ and an $r \times q$ matrix $\mathbf{B}$, is an $mr \times nq$ matrix, $\mathbf{A} \otimes \mathbf{B}$ defined as

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

The Kronecker product has the following properties (see [13])

$$
\begin{aligned}
\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) &= \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C} \\
\mathbf{A} \otimes \mathbf{B} &\neq \mathbf{B} \otimes \mathbf{A} \\
\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) &= (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} \\
(\alpha_A \mathbf{A} \otimes \alpha_B \mathbf{B}) &= \alpha_A \alpha_B (\mathbf{A} \otimes \mathbf{B}) \\
(\mathbf{A} \otimes \mathbf{B})^T &= \mathbf{A}^T \otimes \mathbf{B}^T \\
(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= \mathbf{AC} \otimes \mathbf{BD} \\
(\mathbf{A} \otimes \mathbf{B})^{-1} &= \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \\
\mathrm{rank}(\mathbf{A} \otimes \mathbf{B}) &= \mathrm{rank}(\mathbf{A})\mathrm{rank}(\mathbf{B}) \\
\mathrm{Tr}(\mathbf{A} \otimes \mathbf{B}) &= \mathrm{Tr}(\mathbf{A})\mathrm{Tr}(\mathbf{B}) \\
\det(\mathbf{A} \otimes \mathbf{B}) &= \det(\mathbf{A})^{\mathrm{rank}(\mathbf{B})} \det(\mathbf{B})^{\mathrm{rank}(\mathbf{A})}
\end{aligned}
$$

9.2.2 The Vec Operator

The vec-operator applied on a matrix $\mathbf{A}$ stacks the columns into a vector, i.e. for a 2×2 matrix

$$
\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad \mathrm{vec}(\mathbf{A}) = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{12} \\ A_{22} \end{bmatrix}
$$

Properties of the vec-operator include (see [13])

$$
\begin{aligned}
\mathrm{vec}(\mathbf{AXB}) &= (\mathbf{B}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{X}) \\
\mathrm{Tr}(\mathbf{A}^T\mathbf{B}) &= \mathrm{vec}(\mathbf{A})^T \mathrm{vec}(\mathbf{B}) \\
\mathrm{vec}(\mathbf{A} + \mathbf{B}) &= \mathrm{vec}(\mathbf{A}) + \mathrm{vec}(\mathbf{B}) \\
\mathrm{vec}(\alpha\mathbf{A}) &= \alpha \cdot \mathrm{vec}(\mathbf{A})
\end{aligned}
$$

9.3 Solutions to Systems of Equations

9.3.1 Existence in Linear Systems

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

$$\mathbf{Ax} = \mathbf{b}$$

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

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

9.3.2 Standard Square

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

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

9.3.3 Degenerated Square

9.3.4 Over-determined Rectangular

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

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

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

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

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].

9.3.5 Under-determined Rectangular

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

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

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

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

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

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

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 $\mathbf{B}_1$.

9.3.6 Linear form and zeros

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

9.3.7 Square form and zeros

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

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

9.3.8 The Lyapunov Equation

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} = \mathbf{C}$$
$$\text{vec}(\mathbf{X}) = (\mathbf{I} \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I})^{-1}\text{vec}(\mathbf{C})$$

See Sec 9.2.1 and 9.2.2 for details on the Kronecker product and the vec operator.

9.3.9 Encapsulating Sum

$$\textstyle\sum_n \mathbf{A}_n \mathbf{X} \mathbf{B}_n = \mathbf{C}$$
$$\text{vec}(\mathbf{X}) = \left(\textstyle\sum_n \mathbf{B}_n^T \otimes \mathbf{A}_n\right)^{-1} \text{vec}(\mathbf{C})$$

See Sec 9.2.1 and 9.2.2 for details on the Kronecker product and the vec operator.

9.4 Matrix Norms

9.4.1 Definitions

A matrix norm is a mapping which fulfils

$$\begin{aligned} ||\mathbf{A}|| &\geq 0 \qquad ||\mathbf{A}|| = 0 \Leftrightarrow \mathbf{A} = \mathbf{0} \\ ||c\mathbf{A}|| &= |c|||\mathbf{A}||, \qquad c \in \mathbb{R} \\ ||\mathbf{A} + \mathbf{B}|| &\leq ||\mathbf{A}|| + ||\mathbf{B}|| \end{aligned}$$

9.4.2 Examples

$$\begin{aligned} ||\mathbf{A}||_1 &= \max_j \textstyle\sum_i |A_{ij}| \\ ||\mathbf{A}||_2 &= \sqrt{\max \text{eig}(\mathbf{A}^T\mathbf{A})} \\ ||\mathbf{A}||_p &= (\max_{||\mathbf{x}||_p=1} ||\mathbf{A}\mathbf{x}||_p)^{1/p} \\ ||\mathbf{A}||_\infty &= \max_i \textstyle\sum_j |A_{ij}| \\ ||\mathbf{A}||_F &= \sqrt{\textstyle\sum_{ij} |A_{ij}|^2} = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}^H)} \qquad \text{(Frobenius)} \\ ||\mathbf{A}||_{max} &= \max_{ij} |A_{ij}| \\ ||\mathbf{A}||_{KF} &= ||\text{sing}(\mathbf{A})||_1 \qquad \text{(Ky Fan)} \end{aligned}$$

where $\text{sing}(\mathbf{A})$ is the vector of singular values of the matrix $\mathbf{A}$.

9.4.3 Inequalities

E. H. Rasmussen has in yet unpublished material derived and collected the following inequalities. They are collected in a table as below, assuming $\mathbf{A}$ is an $m \times n$, and $d = \min\{m, n\}$

	$\|\|\mathbf{A}\|\|_{max}$	$\|\|\mathbf{A}\|\|_1$	$\|\|\mathbf{A}\|\|_\infty$	$\|\|\mathbf{A}\|\|_2$	$\|\|\mathbf{A}\|\|_F$	$\|\|\mathbf{A}\|\|_{KF}$
$\|\|\mathbf{A}\|\|_{max}$		1	1	1	1	1
$\|\|\mathbf{A}\|\|_1$	m		m	$\sqrt{m}$	$\sqrt{m}$	$\sqrt{m}$
$\|\|\mathbf{A}\|\|_\infty$	n	n		$\sqrt{n}$	$\sqrt{n}$	$\sqrt{n}$
$\|\|\mathbf{A}\|\|_2$	$\sqrt{mn}$	$\sqrt{n}$	$\sqrt{m}$		1	1
$\|\|\mathbf{A}\|\|_F$	$\sqrt{mn}$	$\sqrt{n}$	$\sqrt{m}$	$\sqrt{d}$		1
$\|\|\mathbf{A}\|\|_{KF}$	$\sqrt{mnd}$	$\sqrt{nd}$	$\sqrt{md}$	d	$\sqrt{d}$	

which are to be read as, e.g.

$$||\mathbf{A}||_2 \leq \sqrt{m} \cdot ||\mathbf{A}||_\infty$$

9.4.4 Condition Number

The 2-norm of $\mathbf{A}$ equals $\sqrt{(\max(\text{eig}(\mathbf{A}^T\mathbf{A})))}$ [9, p.57]. For a symmetric, positive definite matrix, this reduces to $\max(\text{eig}(\mathbf{A}))$ The condition number based on the 2-norm thus reduces to

$$\|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2 = \max(\text{eig}(\mathbf{A}))\max(\text{eig}(\mathbf{A}^{-1})) = \frac{\max(\text{eig}(\mathbf{A}))}{\min(\text{eig}(\mathbf{A}))}. \tag{76}$$

9.5 Rank

9.5.1 Sylvester's Inequality

If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times r$, then

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n \leq \text{rank}(\mathbf{AB}) \leq min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$$

9.6 Integral Involving Dirac Delta Functions

Assuming $\mathbf{A}$ to be square, then

$$\int p(\mathbf{s})\delta(\mathbf{x} - \mathbf{A}\mathbf{s})d\mathbf{s} = \frac{1}{\det(\mathbf{A})}p(\mathbf{A}^{-1}\mathbf{x})$$

Assuming $\mathbf{A}$ to be "underdetermined", i.e. "tall", then

$$\int p(\mathbf{s})\delta(\mathbf{x} - \mathbf{A}\mathbf{s})d\mathbf{s} = \left\{ \begin{array}{ll} \frac{1}{\sqrt{\det(\mathbf{A}^T\mathbf{A})}}p(\mathbf{A}^+\mathbf{x}) & \text{if } \mathbf{x} = \mathbf{A}\mathbf{A}^+\mathbf{x} \\ 0 & \text{elsewhere} \end{array} \right\}$$

See [8].

9.7 Miscellaneous

For any $\mathbf{A}$ it holds that

$$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}^T) = \mathrm{rank}(\mathbf{A}\mathbf{A}^T) = \mathrm{rank}(\mathbf{A}^T\mathbf{A})$$

It holds that

$$\mathbf{A} \text{ is positive definite} \quad \Leftrightarrow \quad \exists \mathbf{B} \text{ invertible, such that } \mathbf{A} = \mathbf{B}\mathbf{B}^T$$

9.7.1 Orthogonal matrix

If $\mathbf{A}$ is orthogonal, then $\det(\mathbf{A}) = \pm 1$.

A One-dimensional Results

A.1 Gaussian

A.1.1 Density

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

A.1.2 Normalization

$$\begin{aligned}
\int e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds &= \sqrt{2\pi\sigma^2} \\
\int e^{-(ax^2+bx+c)} dx &= \sqrt{\frac{\pi}{a}} \exp\left[\frac{b^2-4ac}{4a}\right] \\
\int e^{c_2x^2+c_1x+c_0} dx &= \sqrt{\frac{\pi}{-c_2}} \exp\left[\frac{c_1^2-4c_2c_0}{-4c_2}\right]
\end{aligned}$$

A.1.3 Derivatives

$$\begin{aligned}
\frac{\partial p(x)}{\partial \mu} &= p(x)\frac{(x-\mu)}{\sigma^2} \\
\frac{\partial \ln p(x)}{\partial \mu} &= \frac{(x-\mu)}{\sigma^2} \\
\frac{\partial p(x)}{\partial \sigma} &= p(x)\frac{1}{\sigma}\left[\frac{(x-\mu)^2}{\sigma^2} - 1\right] \\
\frac{\partial \ln p(x)}{\partial \sigma} &= \frac{1}{\sigma}\left[\frac{(x-\mu)^2}{\sigma^2} - 1\right]
\end{aligned}$$

A.1.4 Completing the Squares

$$c_2x^2 + c_1x + c_0 = -a(x-b)^2 + w$$

$$-a = c_2 \qquad b = \frac{1}{2}\frac{c_1}{c_2} \qquad w = \frac{1}{4}\frac{c_1^2}{c_2} + c_0$$

or

$$c_2x^2 + c_1x + c_0 = -\frac{1}{2\sigma^2}(x-\mu)^2 + d$$

$$\mu = \frac{-c_1}{2c_2} \qquad \sigma^2 = \frac{-1}{2c_2} \qquad d = c_0 - \frac{c_1^2}{4c_2}$$

A.1.5 Moments

If the density is expressed by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(s-\mu)^2}{2\sigma^2}\right] \quad \text{or} \quad p(x) = C\exp(c_2x^2 + c_1x)$$

then the first few basic moments are

$$\begin{aligned}
\langle x \rangle &= \mu &&= \tfrac{-c_1}{2c_2} \\
\langle x^2 \rangle &= \sigma^2 + \mu^2 &&= \tfrac{-1}{2c_2} + \left(\tfrac{-c_1}{2c_2}\right)^2 \\
\langle x^3 \rangle &= 3\sigma^2\mu + \mu^3 &&= \tfrac{c_1}{(2c_2)^2}\left[3 - \tfrac{c_1^2}{2c_2}\right] \\
\langle x^4 \rangle &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 &&= \left(\tfrac{c_1}{2c_2}\right)^4 + 6\left(\tfrac{c_1}{2c_2}\right)^2\left(\tfrac{-1}{2c_2}\right) + 3\left(\tfrac{1}{2c_2}\right)^2
\end{aligned}$$

and the central moments are

$$\begin{aligned}
\langle (x-\mu) \rangle &= 0 &&= 0 \\
\langle (x-\mu)^2 \rangle &= \sigma^2 &&= \left[\tfrac{-1}{2c_2}\right] \\
\langle (x-\mu)^3 \rangle &= 0 &&= 0 \\
\langle (x-\mu)^4 \rangle &= 3\sigma^4 &&= 3\left[\tfrac{1}{2c_2}\right]^2
\end{aligned}$$

A kind of pseudo-moments (un-normalized integrals) can easily be derived as

$$\int \exp(c_2x^2 + c_1x)x^n dx = Z\langle x^n \rangle = \sqrt{\frac{\pi}{-c_2}} \exp\left[\frac{c_1^2}{-4c_2}\right] \langle x^n \rangle$$

From the un-centralized moments one can derive other entities like

$$\begin{aligned}
\langle x^2 \rangle - \langle x \rangle^2 &= \sigma^2 &&= \tfrac{-1}{2c_2} \\
\langle x^3 \rangle - \langle x^2 \rangle\langle x \rangle &= 2\sigma^2\mu &&= \tfrac{2c_1}{(2c_2)^2} \\
\langle x^4 \rangle - \langle x^2 \rangle^2 &= 2\sigma^4 + 4\mu^2\sigma^2 &&= \tfrac{2}{(2c_2)^2}\left[1 - 4\tfrac{c_1^2}{2c_2}\right]
\end{aligned}$$

A.2 One Dimensional Mixture of Gaussians

A.2.1 Density and Normalization

$$p(s) = \sum_k^K \frac{\rho_k}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{1}{2}\frac{(s-\mu_k)^2}{\sigma_k^2}\right]$$

A.2.2 Moments

An useful fact of MoG, is that

$$\langle x^n \rangle = \sum_k \rho_k \langle x^n \rangle_k$$

where $\langle\cdot\rangle_k$ denotes average with respect to the $k.th$ component. We can calculate the first four moments from the densities

$$p(x) = \sum_k \rho_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu_k)^2}{\sigma_k^2}\right]$$

$$p(x) = \sum_k \rho_k C_k \exp\left[c_{k2}x^2 + c_{k1}x\right]$$

as

$$\begin{array}{rclcl}
\langle x\rangle & = & \sum_k \rho_k\mu_k & = & \sum_k \rho_k \left[\frac{-c_{k1}}{2c_{k2}}\right] \\
\langle x^2\rangle & = & \sum_k \rho_k(\sigma_k^2+\mu_k^2) & = & \sum_k \rho_k \left[\frac{-1}{2c_{k2}} + \left(\frac{-c_{k1}}{2c_{k2}}\right)^2\right] \\
\langle x^3\rangle & = & \sum_k \rho_k(3\sigma_k^2\mu_k+\mu_k^3) & = & \sum_k \rho_k \left[\frac{c_{k1}}{(2c_{k2})^2}\left[3-\frac{c_{k1}^2}{2c_{k2}}\right]\right] \\
\langle x^4\rangle & = & \sum_k \rho_k(\mu_k^4+6\mu_k^2\sigma_k^2+3\sigma_k^4) & = & \sum_k \rho_k \left[\left(\frac{1}{2c_{k2}}\right)^2\left[\left(\frac{c_{k1}}{2c_{k2}}\right)^2 - 6\frac{c_{k1}^2}{2c_{k2}}+3\right]\right]
\end{array}$$

If all the gaussians are centered, i.e. $\mu_k = 0$ for all k, then

$$\begin{array}{rclcl}
\langle x\rangle & = & 0 & = & 0 \\
\langle x^2\rangle & = & \sum_k \rho_k\sigma_k^2 & = & \sum_k \rho_k \left[\frac{-1}{2c_{k2}}\right] \\
\langle x^3\rangle & = & 0 & = & 0 \\
\langle x^4\rangle & = & \sum_k \rho_k 3\sigma_k^4 & = & \sum_k \rho_k 3\left[\frac{-1}{2c_{k2}}\right]^2
\end{array}$$

From the un-centralized moments one can derive other entities like

$$\begin{array}{rcl}
\langle x^2\rangle - \langle x\rangle^2 & = & \sum_{k,k'} \rho_k\rho_{k'}\left[\mu_k^2+\sigma_k^2-\mu_k\mu_{k'}\right] \\
\langle x^3\rangle - \langle x^2\rangle\langle x\rangle & = & \sum_{k,k'} \rho_k\rho_{k'}\left[3\sigma_k^2\mu_k+\mu_k^3-(\sigma_k^2+\mu_k^2)\mu_{k'}\right] \\
\langle x^4\rangle - \langle x^2\rangle^2 & = & \sum_{k,k'} \rho_k\rho_{k'}\left[\mu_k^4+6\mu_k^2\sigma_k^2+3\sigma_k^4-(\sigma_k^2+\mu_k^2)(\sigma_{k'}^2+\mu_{k'}^2)\right]
\end{array}$$

A.2.3 Derivatives

Defining $p(s) = \sum_k \rho_k \mathcal{N}_s(\mu_k, \sigma_k^2)$ we get for a parameter θ_j of the j.th component

$$\frac{\partial \ln p(s)}{\partial \theta_j} = \frac{\rho_j \mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k \mathcal{N}_s(\mu_k,\sigma_k^2)} \frac{\partial \ln(\rho_j \mathcal{N}_s(\mu_j,\sigma_j^2))}{\partial \theta_j}$$

that is,

$$\begin{array}{rcl}
\frac{\partial \ln p(s)}{\partial \rho_j} & = & \frac{\rho_j \mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k \mathcal{N}_s(\mu_k,\sigma_k^2)} \frac{1}{\rho_j} \\
\frac{\partial \ln p(s)}{\partial \mu_j} & = & \frac{\rho_j \mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k \mathcal{N}_s(\mu_k,\sigma_k^2)} \frac{(s-\mu_j)}{\sigma_j^2} \\
\frac{\partial \ln p(s)}{\partial \sigma_j} & = & \frac{\rho_j \mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k \mathcal{N}_s(\mu_k,\sigma_k^2)} \frac{1}{\sigma_j}\left[\frac{(s-\mu_j)^2}{\sigma_j^2} - 1\right]
\end{array}$$

Note that ρ_k must be constrained to be proper ratios. Defining the ratios by $\rho_j = e^{r_j}/\sum_k e^{r_k}$, we obtain

$$\frac{\partial \ln p(s)}{\partial r_j} = \sum_l \frac{\partial \ln p(s)}{\partial \rho_l}\frac{\partial \rho_l}{\partial r_j} \quad \text{where} \quad \frac{\partial \rho_l}{\partial r_j} = \rho_l(\delta_{lj} - \rho_j)$$

B Proofs and Details

B.1 Misc Proofs

B.1.1 Proof of Equation 14

Essentially we need to calculate

$$\begin{aligned}
\frac{\partial (\mathbf{X}^n)_{kl}}{\partial X_{ij}} &= \frac{\partial}{\partial X_{ij}} \sum_{u_1,\ldots,u_{n-1}} X_{k,u_1} X_{u_1,u_2} \ldots X_{u_{n-1},l} \\
&= \delta_{k,i}\delta_{u_1,j} X_{u_1,u_2} \ldots X_{u_{n-1},l} \\
&\quad\ + X_{k,u_1}\delta_{u_1,i}\delta_{u_2,j}\ldots X_{u_{n-1},l} \\
&\quad \vdots \\
&\quad\ + X_{k,u_1} X_{u_1,u_2}\ldots \delta_{u_{n-1},i}\delta_{l,j} \\
&= \sum_{r=0}^{n-1} (\mathbf{X}^r)_{ki} (\mathbf{X}^{n-1-r})_{jl} \\
&= \sum_{r=0}^{n-1} (\mathbf{X}^r \mathbf{J}^{ij} \mathbf{X}^{n-1-r})_{kl}
\end{aligned}$$

Using the properties of the single entry matrix found in Sec. 8.2.4, the result follows easily.

B.1.2 Details on Eq. 78

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

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

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

Through the calculations, (16) and (47) were used. In addition, by use of (48), the derivative is found with respect to the imaginary part of $\mathbf{X}$

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

Hence, derivative yields

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

and the complex conjugate derivative yields

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

Notice, for real $\mathbf{X}$, $\mathbf{A}$, the sum of (56) and (57) is reduced to (13).
Similar calculations yield

$$\begin{aligned}\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \mathbf{X}} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Re\mathbf{X}} - i\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Im\mathbf{X}}\Big) \\ &= \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)\big(\mathbf{A}\mathbf{X}^H(\mathbf{X}\mathbf{A}\mathbf{X}^H)^{-1}\big)^T \qquad (77)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \mathbf{X}^*} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Re\mathbf{X}} + i\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Im\mathbf{X}}\Big) \\ &= \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)(\mathbf{X}\mathbf{A}\mathbf{X}^H)^{-1}\mathbf{X}\mathbf{A} \qquad (78)\end{aligned}$$

References

[1] Karl Gustav Andersson and Lars-Christer Boiers. *Ordinaera differentialekvationer*. Studenterlitteratur, 1992.

[2] 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.

[3] S. Barnet. *Matrices. Methods and Applications*. Oxford Applied Mathematics and Computin Science Series. Clarendon Press, 1990.

[4] Christoffer Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[5] Robert J. Boik. Lecture notes: Statistics 550. Online, April 22 2002. Notes.

[6] 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.

[7] M. Brookes. Matrix Reference Manual, 2004. Website May 20, 2004.

[8] Mads Dyrholm. Some matrix results, 2004. Website August 23, 2004.

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

[10] Robert M. Gray. Toeplitz and circulant matrices: A review. Technical report, Information Systems Laboratory, Department of Electrical Engineering,Stanford University, Stanford, California 94305, August 2002.

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

[12] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

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

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

[15] John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing*. Prentice-Hall, 1996.

[16] Laurent Schwartz. *Cours d'Analyse*, volume II. Hermann, Paris, 1967. As referenced in [11].

[17] Shayle R. Searle. *Matrix Algebra Useful for Statistics*. John Wiley and Sons, 1982.

[18] G. Seber and A. Lee. *Linear Regression Analysis*. John Wiley and Sons, 2002.

[19] S. M. Selby. *Standard Mathematical Tables*. CRC Press, 1974.

[20] Inna Stainvas. Matrix algebra in differential calculus. Neural Computing Research Group, Information Engeneering, Aston University, UK, August 2002. Notes.

[21] P. P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice Hall, 1993.

[22] Max Welling. The Kalman Filter. Lecture Note.