The Matrix Cookbook

Kaare Brandt Petersen
ISP, IMM, Technical University of Denmark

September 21, 2004

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. Among the few exceptions are the derivatives involving traces and the Petersen-Hao approximation on inverses.

Errors: Very likely there are errors, typos, and mistakes for which I apologize and would be grateful to receive corrections at kbp@imm.dtu.dk or other channels of communication found on my homepage.

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 kbp@imm.dtu.dk.

Notation: Matrices are written in capital bold letters like $\mathbf{A}$, vectors are in bold lower case like $\mathbf{a}$ and scalars as plain letters (both upper and lower) like a or A. Thus, A_{12} denotes the scalar placed at entry $(1, 2)$ in the matrix $\mathbf{A}$, while $\mathbf{A}_{12}$ would denote a matrix with some indices for whatever purpose. Parenthesis around a matrix, however, followed by indices denotes that specific entry of the matrix, i.e. $(\mathbf{A})_{ij} = A_{ij}$.

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

Contents

1 Basics

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$$
$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$
$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$$
$$\mathrm{Tr}(\mathbf{A}) = \sum_i \mathbf{A}_{ii} = \sum_i \lambda_i, \qquad \lambda_i = \mathrm{eig}(\mathbf{A})$$
$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{BCA}) = \mathrm{Tr}(\mathbf{CAB})$$
$$\det(\mathbf{A}) = |\mathbf{A}| = \prod_i \lambda_i \qquad \lambda_i = \mathrm{eig}(\mathbf{A})$$
$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|, \qquad \text{if } \mathbf{A} \text{ and } \mathbf{B} \text{ are invertible}$$
$$|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$$

2 Derivatives

2.1 Derivatives of a Determinant

2.1.1 General form

$$\frac{\partial|\mathbf{A}|}{\partial x} = |\mathbf{A}|\mathrm{Tr}\left[\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right]$$

2.1.2 Linear forms

$$\frac{\partial|\mathbf{A}|}{\partial \mathbf{A}} = |\mathbf{A}|(\mathbf{A}^{-1})^T$$
$$\frac{\partial|\mathbf{BAC}|}{\partial \mathbf{A}} = |\mathbf{BAC}|(\mathbf{A}^{-1})^T = |\mathbf{BAC}|(\mathbf{A}^T)^{-1}$$

2.1.3 Square forms

Assume **B** to be square and symmetric. Then

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

Note that **A** does *not* have to be square.

$$\frac{\partial \ln|\mathbf{A}^T\mathbf{A}|}{\partial \mathbf{A}} = 2(\mathbf{A}^+)^T$$
$$\frac{\partial \ln|\mathbf{A}^T\mathbf{A}|}{\partial \mathbf{A}^+} = -2\mathbf{A}^T$$

See [4].

2.1.4 Other nonlinear forms

$$\frac{\partial \ln|\mathbf{A}|}{\partial \mathbf{A}} = (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$$

$$\frac{\partial |\mathbf{X}^k|}{\partial \mathbf{X}} = k|\mathbf{X}^k|\mathbf{X}^{-T}$$

See [3].

2.2 Derivatives of an Inverse

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

See [8]. If the entries of $\mathbf{A}$ are independent (i.e. not symmetric, Toeplitz or with other kinds of structure), then

$$\frac{\partial (\mathbf{A}^{-1})_{kl}}{\partial A_{ij}} = -(\mathbf{A}^{-1})_{ki}(\mathbf{A}^{-1})_{jl}$$

$$\frac{\partial \mathbf{b}^T\mathbf{A}^{-1}\mathbf{c}}{\partial \mathbf{A}} = -\mathbf{A}^{-T}\mathbf{b}\mathbf{c}^T\mathbf{A}^{-T}$$

$$\frac{\partial |\mathbf{A}^{-1}|}{\partial \mathbf{A}} = -|\mathbf{A}^{-1}|(\mathbf{A}^{-1})^T$$

2.3 Derivatives of Matrices, Vectors and Scalar Forms

2.3.1 First Order

$$\frac{\partial \mathbf{a}^T\mathbf{b}}{\partial \mathbf{a}} = \frac{\partial \mathbf{b}^T\mathbf{a}}{\partial \mathbf{a}} = \mathbf{b}$$

$$\frac{\partial \mathbf{b}^T\mathbf{A}\mathbf{c}}{\partial \mathbf{A}} = \mathbf{b}\mathbf{c}^T$$

$$\frac{\partial \mathbf{b}^T\mathbf{A}^T\mathbf{c}}{\partial \mathbf{A}} = \mathbf{c}\mathbf{b}^T$$

$$\frac{\partial \mathbf{b}^T\mathbf{A}\mathbf{b}}{\partial \mathbf{A}} = \frac{\partial \mathbf{b}^T\mathbf{A}^T\mathbf{b}}{\partial \mathbf{A}} = \mathbf{b}\mathbf{b}^T$$

If the elements of $\mathbf{A}$ are independent variables, then

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

$$\frac{\partial (\mathbf{A}\mathbf{B})_{ij}}{\partial A_{mn}} = \delta_{im}(\mathbf{B})_{nj} = (\mathbf{J}^{mn}\mathbf{B})_{ij}$$

$$\frac{\partial (\mathbf{A}^T\mathbf{B})_{ij}}{\partial A_{mn}} = \delta_{in}(\mathbf{B})_{mj} = (\mathbf{J}^{nm}\mathbf{B})_{ij}$$

2.3.2 Second Order

$$\frac{\partial}{\partial A_{ij}} \sum_{klmn} A_{kl} A_{mn} = 2 \sum_{kl} A_{kl}$$

$$\frac{\partial \mathbf{b}^T \mathbf{A}^T \mathbf{A} \mathbf{c}}{\partial \mathbf{A}} = \mathbf{A}(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T)$$

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

$$\frac{\partial (\mathbf{A}^T \mathbf{B} \mathbf{A})_{kl}}{\partial A_{ij}} = \delta_{lj} (\mathbf{A}^T \mathbf{B})_{ki} + \delta_{kj} (\mathbf{B}\mathbf{A})_{il}$$

$$\frac{\partial (\mathbf{A}^T \mathbf{B} \mathbf{A})}{\partial A_{ij}} = \mathbf{A}^T \mathbf{B} \mathbf{J}^{ij} + \mathbf{J}^{ji} \mathbf{B} \mathbf{A} \qquad (\mathbf{J}^{ij})_{kl} = \delta_{ik} \delta_{jl}$$

See Sec 7.2 for useful properties of the Single-entry matrix $\mathbf{J}^{ij}$

$$\frac{\partial \mathbf{a}^T \mathbf{B} \mathbf{a}}{\partial \mathbf{a}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{a}$$

$$\frac{\partial \mathbf{b}^T \mathbf{A}^T \mathbf{D} \mathbf{A} \mathbf{c}}{\partial \mathbf{A}} = \mathbf{D}^T \mathbf{A} \mathbf{b} \mathbf{c}^T + \mathbf{D} \mathbf{A} \mathbf{c} \mathbf{b}^T$$

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

2.4 Derivatives of Traces

2.4.1 First Order

$$\frac{\partial}{\partial \mathbf{A}} \mathrm{Tr}(\mathbf{A}) = \mathbf{I}$$

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

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

$$\frac{\partial}{\partial \mathbf{A}} \mathrm{Tr}(\mathbf{B}\mathbf{A}\mathbf{C}) = \mathbf{C}\mathbf{B}$$

2.4.2 Second Order

$$\frac{\partial}{\partial \mathbf{A}} \mathrm{Tr}(\mathbf{A}^2) = 2\mathbf{A}$$

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

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

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr}(\mathbf{A}^T\mathbf{A}) = 2\mathbf{A}$$

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr}(\mathbf{B}^T\mathbf{A}^T\mathbf{C}\mathbf{A}\mathbf{B}) = \mathbf{C}^T\mathbf{A}\mathbf{B}\mathbf{B}^T + \mathbf{C}\mathbf{A}\mathbf{B}\mathbf{B}^T$$

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr}\left[\mathbf{A}^T\mathbf{B}\mathbf{A}\mathbf{C}\right] = \mathbf{B}\mathbf{A}\mathbf{C} + \mathbf{B}^T\mathbf{A}\mathbf{C}^T$$

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

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr}\Big[(\mathbf{A}\mathbf{X}\mathbf{b}+\mathbf{c})(\mathbf{A}\mathbf{X}\mathbf{b}+\mathbf{c})^T\Big] = 2\mathbf{A}^T(\mathbf{A}\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T$$

See [3].

2.4.3 Higher Order

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

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

$$\begin{aligned}\frac{\partial}{\partial \mathbf{A}} \text{Tr}\left[\mathbf{B}^T\mathbf{A}^T\mathbf{C}\mathbf{A}\mathbf{A}^T\mathbf{C}\mathbf{A}\mathbf{B}\right] \quad = \quad & \mathbf{C}\mathbf{A}\mathbf{A}^T\mathbf{C}\mathbf{A}\mathbf{B}\mathbf{B}^T + \mathbf{C}^T\mathbf{A}\mathbf{B}\mathbf{B}^T\mathbf{A}^T\mathbf{C}^T\mathbf{A} \\ & + \mathbf{C}\mathbf{A}\mathbf{B}\mathbf{B}^T\mathbf{A}^T\mathbf{C}\mathbf{A} + \mathbf{C}^T\mathbf{A}\mathbf{A}^T\mathbf{C}^T\mathbf{A}\mathbf{B}\mathbf{B}^T\end{aligned}$$

2.4.4 Other

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

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

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

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

See [3].

3 Inverses

3.1 Exact Relations

3.1.1 The Woodbury identity

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

3.1.2 The Kailath Variant

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

See [2] page 153.

3.1.3 A PosDef identity

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

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

See [9].

3.2 Approximations

$$(\mathbf{I}+\mathbf{A})^{-1} \cong \mathbf{I}-\mathbf{A}, \qquad \text{if } \mathbf{A} \text{ small}$$

3.2.1 The Petersen-Hao approximation

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

3.3 Generalized Inverse

3.3.1 Definition

A generalized inverse matrix of the matrix $\mathbf{A}$ is any matrix $\mathbf{A}^-$ such that

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

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

3.4 Pseudo Inverse

3.4.1 Definition

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

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

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

3.4.2 Basic Properties

Assume $\mathbf{A}^+$ to be the pseudo-inverse of $\mathbf{A}$, then

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

See [1].

3.4.3 Construction

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

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

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

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

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 [1].

4 Decompositions

4.1 Eigenvalues and Eigenvectors

4.1.1 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 $\mathbf{v}_i$

4.1.2 General Properties

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

4.1.3 Symmetric

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

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

4.2 Singular Value Decomposition

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

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

where

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

4.2.1 Symmetric Square decomposed into squares

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

$$\left[\ \mathbf{A}\ \right] = \left[\ \mathbf{V}\ \right]\left[\ \mathbf{D}\ \right]\left[\ \mathbf{V}^T\ \right]$$

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

4.2.2 Square decomposed into squares

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

$$\left[\ \mathbf{A}\ \right] = \left[\ \mathbf{V}\ \right]\left[\ \mathbf{D}\ \right]\left[\ \mathbf{U}^T\ \right]$$

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

4.2.3 Square decomposed into rectangular

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

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

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

4.2.4 Rectangular decomposition I

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

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

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

4.2.5 Rectangular decomposition II

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

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

4.2.6 Rectangular decomposition III

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

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

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

4.3 Triangular Decomposition

4.3.1 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.

5 General Statistics and Probability

5.1 Moments of any distribution

5.1.1 Mean and covariance of linear forms

Assume $\mathbf{X}$ and $\mathbf{x}$ to be a matrix and a vector of random variables. Then

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

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

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

See [7].

5.1.2 Mean and Variance of Square Forms

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

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

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

See [7]

5.2 Expectations

Assume $\mathbf{x}$ to be a stochastic vector with mean $\mathbf{m}$, covariance $\mathbf{M}$ and central moments $\mathbf{v}_r = E[(\mathbf{x} - \mathbf{m})^r]$.

5.2.1 Linear Forms

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

5.2.2 Quadratic Forms

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

$$\begin{aligned}
E[(\mathbf{A}\mathbf{x} + \mathbf{a})^T(\mathbf{B}\mathbf{x} + \mathbf{b})] &= \text{Tr}(\mathbf{A}\mathbf{M}\mathbf{B}^T) + (\mathbf{A}\mathbf{m} + \mathbf{a})^T(\mathbf{B}\mathbf{m} + \mathbf{b}) \\
E[\mathbf{x}^T\mathbf{x}] &= \text{Tr}(\mathbf{M}) + \mathbf{m}^T\mathbf{m} \\
E[\mathbf{x}^T\mathbf{A}\mathbf{x}] &= \text{Tr}(\mathbf{A}\mathbf{M}) + \mathbf{m}^T\mathbf{A}\mathbf{m} \\
E[(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x})] &= \text{Tr}(\mathbf{A}\mathbf{M}\mathbf{A}^T) + (\mathbf{A}\mathbf{m})^T(\mathbf{A}\mathbf{m}) \\
E[(\mathbf{x} + \mathbf{a})^T(\mathbf{x} + \mathbf{a})] &= \text{Tr}(\mathbf{M}) + (\mathbf{m} + \mathbf{a})^T(\mathbf{m} + \mathbf{a})
\end{aligned}$$

See [3].

5.2.3 Cubic Forms

Assume $\mathbf{x}$ to be independent, then

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

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

See [3].

6 Gaussians

6.1 One Dimensional

6.1.1 Density and Normalization

The density is

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

Normalization integrals

$$\int e^{-\frac{(s-\mu)^2}{2\sigma^2}}ds=\sqrt{2\pi\sigma^2}$$

$$\int e^{-(ax^2+bx+c)}dx=\sqrt{\frac{\pi}{a}}\exp\left[\frac{b^2-4ac}{4a}\right]$$

$$\int e^{c_2x^2+c_1x+c_0}dx=\sqrt{\frac{\pi}{-c_2}}\exp\left[\frac{c_1^2-4c_2c_0}{-4c_2}\right]$$

6.1.2 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

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

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

6.1.3 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

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

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

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

6.2 One Dimensional Mixture of Gaussians

6.2.1 Density and Normalization

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

6.2.2 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 (obviously)

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

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

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

6.3 Basics

6.3.1 Density and normalization

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

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

Integration and normalization

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

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

The derivative of the density is a vector of the form

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

6.3.2 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

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

6.3.3 Rearranging Means

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

6.3.4 Rearranging into squared form

$$\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b} - \mathbf{c}^T\mathbf{x} + d = (\mathbf{x}-\mathbf{m})^T\mathbf{A}(\mathbf{x}-\mathbf{m}) + \eta$$

$$\mathbf{m} = \mathbf{A}^{-1}\mathbf{c} = \mathbf{A}^{-1}\mathbf{b}$$

$$\eta = d - \mathbf{c}^T\mathbf{A}^{-1}\mathbf{b}$$

A variant is (Assume $\mathbf{A}$ is symmetric)

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

A variant with traces

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

6.3.5 Sum of two squared forms

Rearranging the sum of two

$$(\mathbf{x}-\mathbf{m}_1)^T\mathbf{\Sigma}_1^{-1}(\mathbf{x}-\mathbf{m}_1) + (\mathbf{x}-\mathbf{m}_2)^T\mathbf{\Sigma}_2^{-1}(\mathbf{x}-\mathbf{m}_2) = (\mathbf{x}-\mathbf{m}_c)^T\mathbf{\Sigma}_c^{-1}(\mathbf{x}-\mathbf{m}_c) + C$$

$$\begin{aligned}
\mathbf{\Sigma}_c^{-1} &= \mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1} \\
\mathbf{m}_c &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2) \\
C &= \mathbf{m}_1^T\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{m}_2^T\mathbf{\Sigma}_2^{-1}\mathbf{m}_2 \\
&\quad -(\mathbf{m}_1^T\mathbf{\Sigma}_1^{-1} + \mathbf{m}_2^T\mathbf{\Sigma}_2^{-1})(\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2)
\end{aligned}$$

6.3.6 Product of gaussian densities

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

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

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

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

6.4 Moments

6.4.1 Mean and covariance of linear forms

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

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

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

As for any other distribution is holds for gaussians that

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

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

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

6.4.2 Mean and variance of square forms

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

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

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

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

6.4.3 Cubic forms

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

6.4.4 Mean of Quartic Forms

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

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

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

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

See [3].

6.4.5 Moments

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

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

6.5 Miscellaneous

6.5.1 Whitening

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

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

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

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

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

6.5.2 The Chi-Square connection

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

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

6.5.3 Entropy

Entropy of a D-dimensional gaussian

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

6.6 Mixture of Gaussians

6.6.1 Density

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

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

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

7 Miscellaneous

7.1 Miscellaneous

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

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

Assume $\mathbf{A}$ is positive definite. Then

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

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

7.2 Indices, Entries and Vectors

Let $\mathbf{e}_i$ denote the column vector which is 1 on entry i and zero elsewhere, i.e. $(\mathbf{e}_i)_j = \delta_{ij}$, and let $\mathbf{J}^{ij}$ denote the matrix which is 1 on entry (i, j) and zero elsewhere.

7.2.1 Rows and Columns

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

7.2.2 Permutations

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

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

then

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

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.

7.2.3 Swap and Zeros

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

$$\mathbf{AJ}^{ij} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_i & \dots & \mathbf{0} \end{bmatrix}$$

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

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

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.

7.2.4 Rewriting product of elements

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

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

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

7.2.5 The Singleentry Matrix in Scalar Expressions

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

$$\text{Tr}(\mathbf{A}\mathbf{J}^{ij}) = \text{Tr}(\mathbf{J}^{ij}\mathbf{A}) = A_{ji}$$

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

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

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

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

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

7.3 Solutions to Systems of Equations

7.3.1 Existence in Linear Systems

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

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

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

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

7.3.2 Standard Square

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

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

7.3.3 Degenerated Square

7.3.4 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 [1].

7.3.5 Under-determined Rectangular

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

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

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 $||\mathbf{x}||^2$. 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 [1].

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 $\mathbf{B}_1$.

7.3.6 Linear form and zeros

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

7.3.7 Square form and zeros

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

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

7.4 Block matrices

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

7.4.1 Multiplication

Assuming the dimensions of the blocks matches we have

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

7.4.2 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

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

7.4.3 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

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

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

7.4.4 Block diagonal

For block diagonal matrices we have

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

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

7.5 Positive Definite and Semi-definite Matrices

7.5.1 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.

7.5.2 Eigenvalues

The following holds with respect to the eigenvalues:

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

7.5.3 Trace

The following holds with respect to the trace:

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

7.5.4 Inverse

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

7.5.5 Diagonal

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

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

7.5.7 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 $\mathbf{B}^T\mathbf{A}\mathbf{B} = \mathbf{I}$.

7.5.8 Equation with zeros

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

7.5.9 Rank of product

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

7.5.10 Positive definite property

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

7.5.11 Outer Product

If $\mathbf{X}$ is $n \times r$ of rank r, then $\mathbf{XX}^T$ is positive definite.

7.5.12 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.

7.6 Integral Involving Dirac Delta Functions

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

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

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

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

See [4].

References

[1] S. Barnet, *Matrices. Methods and Applicatiopns*, Oxford Applied Mathematics and Computin Science Series, Clarendon Press, Oxford, 1990.
Comment: Well-written and reasonably comprehensive.

[2] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[3] M. Brookes, *Matrix Reference Manual*, website available (May 20, 2004) at http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html
Comment: To the point and very comprehensive, but the HTML format makes it less appealing and less easy to read.

[4] M. Dyrholm, "Some Matrix Results" (August 23, 2004) http://www.imm.dtu.dk/ mad/papers/madrix.pdf

[5] M. S. Pedersen, *Matricks*, Note available on the internet (May 20, 2004) at http://www.imm.dtu.dk/pubdb/views/edoc_download.php/2976/pdf/imm2976.pdf

[6] S. Roweis, *Matrix Identities*, Note available on the internet (May 20, 2004) at http://www.cs.toronto.edu/ roweis/notes/matrixid.pdf

[7] G. Seber and A. Lee, *Linear Regression Analysis*, 2nd Ed. Wiley, New York, 2002.

[8] S. M. Selby, "Standard Mathematical Tables", 23. edition, CRC Press, 1974. (First published in 1964).
Comment: Table like desktop reference.

[9] M. Welling, The Kalman Filter, Lecture Note, California Institute of Technology.

I The identity matrix Jij The single-entry matrix, 1 at (i, j) and zero elsewhere Î£ A positive denite matrix Î› A diagonal matrix. Petersen & Pedersen, The Matrix Cookbook, Version: November 15, 2012, Page 5. 1 BASICS. 1 Basics. @TECHREPORT{Petersen06thematrix, author = {Kaare Brandt Petersen and Michael Syskind Pedersen and Jan Larsen and Korbinian Strimmer and Lars Christiansen and Kai Hansen and Liguo He and Loic Thibaut and Miguel BarÃ£o and Stephan Hattinger and Vasile Sima and We The}, title = {The matrix cookbook}, institution = {}, year = {2006} }.Â Disclaimer: The identities, approximations and relations presented here were obviously not invented but collected, borrowed and copied from a large amount of sources. The Matrix Cookbook. Kaare Brandt Petersen ISP, IMM, Technical University of Denmark. September 21, 2004. What is this?Â 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 kbp@imm.dtu.dk. 66 Petersen & Pedersen, The Matrix Cookbook, Version: November 15, 2012, Page 4 CONTENTS CONTENTS Notation and Nomenclature A Matrix Aij Matrix indexed forÂ product 0 The null matrix. Zero in all entries. I The identity matrix Jij The single-entry matrix, 1 at (i, j) and zero elsewhere Î£ A positive definite matrix Î› A diagonal matrix Petersen & Pedersen, The Matrix Cookbook, Version: November 15, 2012, Page 5 1 BASICS 1 Basics (AB)âˆ'1 = Bâˆ'1 Aâˆ'1 (1) âˆ'1 âˆ'1 âˆ'1 âˆ'1 (ABC...) = ...