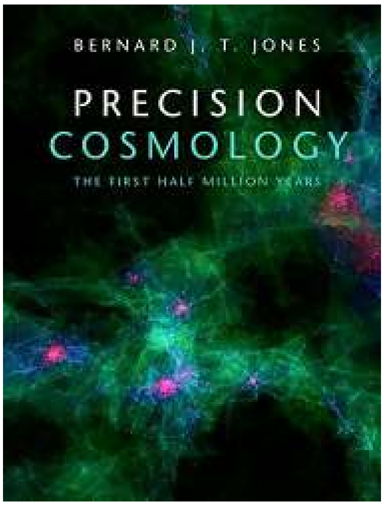
Matrices

A Supplement to “Precision Cosmology”

Bernard Jones



Simply a collection of useful formulae involving matrices of different kinds, and in particular block structured matrices. This is particularly useful in dealing with multidimensional Gaussian Likelihood functions and other aspects of statistics.

This is one of a set of Supplementary Notes and Chapters to “Precision Cosmology”.
Some of these Supplements might have been a chapter in the book itself, but were regarded either as being somewhat more specialised than the material elsewhere in the book, or somewhat tangential to the main subject matter.
The are mostly early drafts and have not been fully proof-read.
Please send comments on errors or ambiguities to “`PrecisionCosmology(at)gmail.com`”.

Contents

1 Matrices *page* 1
1.1 Cholesky representation 1
1.2 Block matrices 2
1.2.1 Determinants 3
1.2.2 Multiplying block matrices 3
1.2.3 Block diagonal matrices 3
1.2.4 Block triangular matrices 4
1.2.5 Inverses 4
1.3 Matrices in statistics 5
1.3.1 Conditional variance 5
1.3.2 Marginal distributions 5
1.4 Sherman-Morrison-Woodbury inversion 6
References 7

1 Matrices

1.1 Cholesky representation

refer to the two matrix books

Dealing with large covariance matrices beyond 2×2 or 3×3 in size requires computational resources, and for very large matrices some matrix algorithms that make handling the matrix more efficient. The Cholesky decomposition of a real [1] positive definite square matrix $\mathbf{A}$ is a factorisation of the matrix into the product of a lower triangular matrix $\mathbf{L}$ and its transpose $\mathbf{L}^T$:

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T \tag{1.1}$$

A lower triangular matrix is a matrix is a matrix that, when expressed in *echelon* form [2], has only zeros above the main diagonal. In other words, the $n \times n$ matrix $\mathbf{A}$ is lower triangular when its elements a_{ij} are such that $a_{ij} = 0 for \ \leq i < j \leq n$. We can see the process easily for a the covariance matrix $\mathbf{\Sigma}$ of a bivariate Gaussian distribution:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_x\sigma_Y \\ \rho\sigma_x\sigma_Y & \sigma_Y^2 \end{pmatrix} \tag{1.2}$$

If our matrix $\mathbf{L}$ is lower triangular, it will have three components, a, b, c and

$$\mathbf{\Sigma} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_x\sigma_Y \\ \rho\sigma_x\sigma_Y & \sigma_Y^2 \end{pmatrix} \tag{1.3}$$

We have seen this elsewhere: cross reference

So we have to solve the equations

$$a^2 = \sigma_X^2, \quad ab = \rho\sigma_X\sigma_Y, \quad b^2 + c^2 = \sigma_Y^2 \tag{1.4}$$

The solutions are

$$a = \sigma_X, \quad b = \rho\sigma_Y, \quad c = \sigma_Y\sqrt{(1 - \rho^2)} \tag{1.5}$$

The numerical method used for reducing a matrix to its Cholesky form is Gaussian elimination.

Why is this interesting? Given two Gaussian random variables Z_1 and Z_2 which are

[1] This also applies to Hermitian Matrices, i.e. matrices $\mathbf{H} = \{h_{ij}\}$ having entries that are complex numbers and which are equal to the transpose of their complex conjugate: $\mathbf{H} = \mathbf{H}^{*T}$. The complex conjugate transposed of a matrix is often denotes by the symbol $\mathbf{H}^\dagger$. The ij element of $\mathbf{H}^\dagger$ is h_{ji}^*, where * denotes complex conjugation. Hermitian matrices have real eigenvalues.

[2] A set of vectors $\mathbf{x}_i$ is said to be in echelon order when they can be ordered in such a way that the number of consecutive zero components in $\mathbf{x}_i$, starting from the first component, increases *strictly* with i. If we can permute the rows and the columns of a matrix so that, viewed as vectors, they are row-wise and column-wise in echelon order, the matrix is triangular.

normally distributed with zero mean and unit variance, i.e. $N(0,1)$, then we can generate pairs of correlated random variables X, Y with given variances σ_X, σ_Y and covariance ρ, i/e/ taken from $N(X, Y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$. This is accomplished by calculating

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + \begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \tag{1.6}$$

which gives the transformation

$$X = \mu_X + \sigma_X Z_1 \tag{1.7}$$

$$Y = \mu_Y + \sigma_Y [\rho Z_1 + (1-\rho^2)^{1/2} Z_2] \tag{1.8}$$

Positive definite matrices have eigenvalues $\lambda \geq 0$. One of the issues in dealing with such matrices is that numerical issues may create a matrix that is not positive definite. Computation with the Cholesky decomposition of a matrix is often more stable than working with the matrix itself.

1.2 Block matrices

A matrix $\mathbf{M}$ having m rows and n columns is said to be an $m \times n$ matrix:

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ m_{m1} & m_{m2} & \dots & m_{mn} \end{pmatrix} \quad \text{i.e.} \quad \underset{m\times n}{\mathbf{M}} = (m_{ij})$$

The *transpose* of the $m \times n$ matrix $\mathbf{M} = (m_{ij})$ is the $n \times m$ matrix $\mathbf{M}^T = (m_{ji})$ in which rows and columns have been interchanged. The rows $(m_{i1}, m_{i2}, \dots, m_{in}), i = 1, \dots, n$ are $1 \times n$ matrices or *row vectors*. The columns are $m \times 1$ matrices which can be written as $(m_{1i}, m_{2i}, \dots, m_{ni})^T, i = 1, \dots, m$.

Expressing matrices in the abbreviated *block form* and manipulating the blocks is extremely useful in many respects. The basic idea best shown using the example

$$\mathbf{M} = \left(\begin{array}{ccc:c} a & b & c & x \\ d & e & f & y \\ \hdashline u & v & w & t \end{array}\right) = \begin{pmatrix} \underset{2\times3}{\mathbf{P}} & \underset{2\times1}{\mathbf{Q}} \\ \underset{1\times3}{\mathbf{R}} & \underset{1\times1}{\mathbf{S}} \end{pmatrix}$$

where

$$\mathbf{P} = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \quad \mathbf{Q} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \mathbf{R} = \begin{pmatrix} u & v & w \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} t \end{pmatrix}$$

The blocks have to be *coherent* both in the block layout and when manipulating block matrices. Block matrices can be added and multiplied provided their structures are coherent.

1.2.1 Determinants

Determinants are used when determining eigenvalues of matrices. A block structured matrix $\mathbf{M}$ consisting of a regular (*i.e.* non-singular) $m \times m$ matrix $\mathbf{A}$, an $m \times n$ matrix $\mathbf{B}$, an $n \times m$ matrix $\mathbf{C}$ and an $n \times n$ matrix $\mathbf{D}$ has determinant

$$\det \mathbf{M} = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}| \cdot |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| \tag{1.9}$$

$$= |\mathbf{D}\mathbf{A} - \mathbf{C}\mathbf{B}| \quad \text{if and only if } m = n,\ \mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} \tag{1.10}$$

with a corresponding result if $\mathbf{A}$ and $\mathbf{C}$ commute.[3]

When $\mathbf{A}$ is 1×1 and non-zero, this reduces the dimensionality of the determinant calculation by one.

1.2.2 Multiplying block matrices

Block coherent matrices can be multiplied in the obvious way. The general idea is illustrated by the following examples:

$$\mathbf{A}\begin{pmatrix} \mathbf{X} & \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{AX} & \mathbf{AY} \end{pmatrix}, \qquad \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}\mathbf{A} = \begin{pmatrix} \mathbf{XA} \\ \mathbf{YA} \end{pmatrix}\mathbf{A} \tag{1.12}$$

and of course the "obvious"

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}\begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{R} & \mathbf{S} \end{pmatrix} = \begin{pmatrix} \mathbf{AP} + \mathbf{BR} & \mathbf{AQ} + \mathbf{BS} \\ \mathbf{CP} + \mathbf{DR} & \mathbf{CQ} + \mathbf{DS} \end{pmatrix}. \tag{1.13}$$

1.2.3 Block diagonal matrices

A matrix $\mathbf{M}$ made up of k square matrices $\mathbf{A}_{ii}$, $i = 1, \ldots, k$ not necessarily all of the same size, in a layout of the form

$$\mathbf{M} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A}_{kk} \end{pmatrix} \tag{1.14}$$

is said to be *block diagonal*. The alternate notation for this is

$$\mathbf{M} = \mathbf{A}_{11} \oplus \mathbf{A}_{22} + \cdots \oplus \mathbf{A}_{kk} = \bigoplus_{i=1}^{k} \mathbf{A}_{ii} \tag{1.15}$$

[3] Equation (1.9) is proved using the identity

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}\begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix} = \begin{pmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C}\mathbf{A}^{-1} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{pmatrix} \tag{1.11}$$

which displays this as the *direct sum* of the matrices $\mathbf{A}_{ii}$. As with all block matrices they can be added and multiplied provided that they are block-coherent. The determinant of block diagonal matrix is the product of the determinants of the individual blocks:

$$|\,\mathbf{A}_{11} \oplus \cdots \oplus \mathbf{A}_{kk}\,| = \Pi_{i=1}^{k} |\,\mathbf{A}_{ii}\,| \tag{1.16}$$

1.2.4 Block triangular matrices

A *block upper triangular matrix* is a matrix of the form

$$\mathbf{M} = \begin{pmatrix} \mathbf{A}_{11} & * & \dots & * \\ \mathbf{0} & \mathbf{A}_{22} & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_{kk} \end{pmatrix} \tag{1.17}$$

where $*$ denotes any matrix, and the matrices $\mathbf{A}_{ii}, i = 1, \dots, k$ are square, but not necessarily of the same size. A matrix is *block lower triangular* if its transpose is block upper triangular.

1.2.5 Inverses

If matrices $\mathbf{P}$ and $\mathbf{R}$ are square non-singular, but not necessarily of the same size, then

$$\begin{pmatrix} \underset{m\times m}{\mathbf{P}} & \underset{m\times n}{\mathbf{Q}} \\ \underset{n\times m}{\mathbf{0}} & \underset{n\times n}{\mathbf{R}} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{P}^{-1} & -\mathbf{P}^{-1}\mathbf{Q}\mathbf{R}^{-1} \\ \mathbf{0} & \mathbf{R}^{-1} \end{pmatrix} \tag{1.18}$$

For a 3×3 block structure:

$$\begin{pmatrix} \mathbf{A} & \mathbf{H} & \mathbf{G} \\ \mathbf{0} & \mathbf{B} & \mathbf{F} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{H}\mathbf{B}^{-1} & \mathbf{A}^{-1}\mathbf{H}\mathbf{B}^{-1}\mathbf{F}\mathbf{C}^{-1} - \mathbf{A}^{-1}\mathbf{G}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1} & -\mathbf{B}^{-1}\mathbf{F}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix} \tag{1.19}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are square and non-singular, but not necessarily of the same size.

The general 2×2 block matrix can be inverted as

$$\begin{pmatrix} \underset{m\times m}{\mathbf{A}} & \underset{m\times n}{\mathbf{B}} \\ \underset{n\times m}{\mathbf{C}} & \underset{n\times n}{\mathbf{D}} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix} \tag{1.20}$$

The matrices $(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})$ and $(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})$ are called the *Shur complements* of $\mathbf{A}$ and $\mathbf{D}$, respectively.

1.3 Matrices in statistics

1.3.1 Conditional variance

Consider an n-dimensional random vector $\mathbf{X}$ and an m-dimensional vector $\mathbf{Y}$ (so their realisations $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Suppose the vector $(\mathbf{X}, \mathbf{Y})$ has a multivariate normal distribution with covariance matrix

$$\mathbf{V} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \tag{1.21}$$

where $\mathbf{A}$ and $\mathbf{C}$ are symmetric and non-singular. Then the conditional variance of $\mathbf{X}$ given $\mathbf{Y}$ is

$$\mathrm{Var}\,(\mathbf{X}\,|\,\mathbf{Y}\,) = \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T \tag{1.22}$$

The right hand side is called the *Schur complement* of $\mathbf{C}$ in $\mathbf{V}$.

The Schur complement also appears in solving linear systems of equations for some, rather than all of the unknown variables. Suppose we wish to solve the equation

$$\mathbf{C}\mathbf{x} = \mathbf{b} \tag{1.23}$$

for only some of the unknown components of $\mathbf{x}$. Here we suppose that $\mathbf{C}$ is symmetric, *i.e.* a covariance matrix. We start by partitioning $\mathbf{C}$ and writing the linear system of equations as

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^T & \mathbf{C}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} \tag{1.24}$$

The solution for $\mathbf{x}_1$ is obtained from

$$(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{12}^T)\,\mathbf{x}_1 = \mathbf{b}_1 - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{B}_2 \tag{1.25}$$

where the matrix multiplying $\mathbf{x}_1$ will be recognised as the Shur complement (1.22) of $\mathbf{C}$.

1.3.2 Marginal distributions

There is a multi-variate analogue of equation (**??**) for the posterior likelihood of a normally distributed dataset under a normal prior. To derive this requires a matrix analogue of the result (**??**). If $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ are k-dimensional vectors, and $\mathbf{A}$ and $\mathbf{B}$ are positive definite symmetric $k \times k$ matrices (*i.e.* covariance matrices) then

$$\begin{aligned}(\mathbf{y}-\mathbf{x})^T\mathbf{A}(\mathbf{y}-\mathbf{x}) &+ (\mathbf{x}-\mathbf{z})^T\mathbf{B}(\mathbf{x}-\mathbf{z}) \\ &= (\mathbf{x}-\mathbf{c})^T\mathbf{D}(\mathbf{x}-\mathbf{c}) + (\mathbf{y}-\mathbf{z})^T\mathbf{E}(\mathbf{y}-\mathbf{z})\end{aligned} \tag{1.26}$$

where

$$\mathbf{c} = (\mathbf{A}+\mathbf{B})^{-1}(\mathbf{A}\mathbf{y}+\mathbf{B}\mathbf{z}) \tag{1.27}$$

$$\mathbf{D} = \mathbf{A}+\mathbf{B}, \quad \mathbf{E} = (\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1} \tag{1.28}$$

1.4 Sherman-Morrison-Woodbury inversion

For any set of conformable matrices

$$(\underset{n\times n}{\mathbf{A}} + \underset{n\times k}{\mathbf{U}}\,\underset{k\times k}{\mathbf{B}}\,\underset{k\times n}{\mathbf{V}})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\mathbf{B}(\mathbf{I} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\mathbf{B})^{-1}\mathbf{V}\mathbf{A}^{-1} \tag{1.29}$$

This is also referred to simply as the *Woodbury Formula* (Press et al., 2007, section 2.7). The use of this equation supposes that we already know the inverse of $\mathbf{A}$ and wish to invert $(\mathbf{A} + \mathbf{UBV})$. If $\mathbf{B}$ is of lower dimension than $\mathbf{A}$ this speeds up the inversion.

There are many forms for this equation and also a number of special cases. In the correction step of the Kalman Filter the matrix $\mathbf{B}$ is invertible we can decrease the number of multiplications by writing the Woodbury formula as

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1}, \quad |\,\mathbf{B}\,| \neq 0. \tag{1.30}$$

References

Press, W.H., Teukolsky, S., Vetterling, W.T., and Flannery, B.P. 2007. *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press; 3rd. edition.