Some Matrix and Probability Identities

Perry Moerland
Dalle Molle Institute of Perceptual Artificial Intelligence
C.P. 592, 1920 Martigny, Switzerland
`Perry.Moerland@idiap.ch`

This cheat sheet might come in handy for the readers of (Moerland 2000); it has been left out of the original paper because of space restrictions and lists some formulas useful when dealing with matrices and probabilities.

We assume that matrices are non-singular when taking the inverse. Matrices and vectors are also supposed to have the right size for the operations applied to them. Lower-case bold letters $\mathbf{a}$, $\mathbf{b}$ etc., denote vectors and upper-case bold letters $\mathbf{X}$, $\mathbf{Y}$ etc., denote matrices.

Another useful list (the intersection of which with this list is non-empty) has been made available by Sam Roweis at `http://www.gatsby.ucl.ac.uk/~roweis/notes/{matrixid.ps.gz,gaussid.ps.gz}`.

Trace Rotation

$$\mathrm{tr}(\mathbf{X}\mathbf{Y}\cdots\mathbf{Z}) = \mathrm{tr}(\mathbf{Z}\mathbf{X}\mathbf{Y}\cdots) = \cdots = \mathrm{tr}(\mathbf{Z}\cdots\mathbf{X}\mathbf{Y}) \tag{1}$$

Derivatives

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

$$\frac{\partial(\mathbf{a}^T\mathbf{X}\mathbf{a})}{\partial\mathbf{a}} = (\mathbf{X}+\mathbf{X}^T)\mathbf{a} \tag{3}$$

$$\frac{\partial\mathrm{tr}(\mathbf{X}^T\mathbf{Y}\mathbf{X})}{\partial\mathbf{X}} = (\mathbf{Y}+\mathbf{Y}^T)\mathbf{X} \tag{4}$$

$$\frac{\partial\ln|\mathbf{X}|}{\partial\mathbf{X}} = (\mathbf{X}^T)^{-1} \tag{5}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}\mathbf{b})}{\partial\mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{6}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}^T\mathbf{b})}{\partial\mathbf{X}} = \mathbf{b}^T\mathbf{a} \tag{7}$$

Matrix inversion lemma This lemma is useful for calculating the inverse of certain matrices in a more efficient way:

$$(\mathbf{X}+\mathbf{Y}\mathbf{Z}\mathbf{Y}^T)^{-1} = \mathbf{X}^{-1} - \mathbf{X}^{-1}\mathbf{Y}(\mathbf{Z}^{-1}+\mathbf{Y}^T\mathbf{X}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{X}^{-1}, \tag{8}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are square. If $\mathbf{X}$ is easy to invert (for example, diagonal) and $\mathbf{Y}$ has many rows and few columns, this lemma can make taking the inverse computationally efficient (Golub and Van Loan 1996, page 50).

Determinant factoring lemma For diagonal $\ell \times \ell$ matrix $\mathbf{R}$, $\ell \times k$ matrix $\mathbf{X}$ and $k \times \ell$ matrix $\mathbf{Y}$:

$$|\mathbf{R} + \mathbf{XY}| = |\mathbf{R}||\mathbf{I}_k + \mathbf{YR}^{-1}\mathbf{X}|. \tag{9}$$

If $\ell \gg k$ and using that the inverse and determinant of diagonal matrix $\mathbf{R}$ are easy to calculate, this lemma can considerably simplify the calculation of the determinant.

Vec operator Function that stacks the columns of a matrix into one vector:

$$\text{vec}\left[\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right] = (a_{11}\ a_{21}\ a_{12}\ a_{22})^T. \tag{10}$$

Diag operator Function from vectors to matrices that puts the vector on the main diagonal:

$$\text{diag}\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right] = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}. \tag{11}$$

We also use it as a function from matrices to matrices that sets the off-diagonal elements to zero.

Kronecker product The Kronecker product of two matrices is defined as:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} \end{pmatrix}. \tag{12}$$

Then with the "vec" operator (10), we have (Golub and Van Loan 1996, page 180):

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}) \tag{13}$$

Element-wise product The element-wise product of two matrices is:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \circ \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix}. \tag{14}$$

With the "vec" (10) and "diag" (11) operators , the element-wise product can be removed from an expression:

$$\text{vec}(\mathbf{A} \circ \mathbf{B}) = \text{diag}\{\text{vec}(\mathbf{A})\}\text{vec}(\mathbf{B}). \tag{15}$$

Basic Probability Rules

$$P(\mathbf{x}) = \sum_{\mathbf{y}} P(\mathbf{x}, \mathbf{y}) \qquad \text{(Sum rule)} \tag{16}$$

$$P(\mathbf{x}, \mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \qquad \text{(Product rule)} \tag{17}$$

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \qquad \text{(Bayes' theorem)} \tag{18}$$

Linear functions of a Gaussian vector

$$\mathbf{a} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \Rightarrow \quad \mathbf{Xa} + \mathbf{b} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\mu} + \mathbf{b}, \mathbf{X\Sigma X}^T). \tag{19}$$

Marginals and Conditionals of a Gaussian

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{Z}^T & \mathbf{Y} \end{bmatrix} \right),$$

with $\mathbf{Z}$ the cross-variance matrix between $\mathbf{x}$ and $\mathbf{y}$. The Gaussian marginal distributions are:

$$\mathbf{x} \sim \mathcal{N}(\mathbf{a}, \mathbf{X}) \tag{20}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{b}, \mathbf{Y}), \tag{21}$$

and Gaussian conditional distributions:

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mathbf{a} + \mathbf{Z}\mathbf{Y}^{-1}(\mathbf{y} - \mathbf{b}), \mathbf{X} - \mathbf{Z}\mathbf{Y}^{-1}\mathbf{Z}^T) \tag{22}$$

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{b} + \mathbf{Z}^T\mathbf{X}^{-1}(\mathbf{x} - \mathbf{a}), \mathbf{Y} - \mathbf{Z}^T\mathbf{X}^{-1}\mathbf{Z}). \tag{23}$$

References

Golub, G. H. and C. F. Van Loan (1996). *Matrix Computations* (3rd ed.). Baltimore: The Johns Hopkins University Press.

Moerland, P. (2000). Mixtures of latent variable models for density estimation and classification. IDIAP-RR 25, IDIAP, `ftp://ftp.idiap.ch/pub/reports/2000/rr00-25.ps.gz`. Submitted for publication.