Notes [1]

Gabriel Maher

January 30, 2018

[1] This document contains my notes on miscellaneous topics.

Part I

Mathematics

Chapter 1

Linear Algebra

1.1 Matrix Inverse Formulas

This chapter contains formulas for the inverses of various sums of matrices.

1.1.1 Inverse of $A + B$

Let $C + D$ be the inverse of $A + B$, then

$$\begin{aligned} (C+D)(A+B) &= I \\ \Rightarrow CA + CB + DA + DB &= I. \end{aligned} \tag{1.1}$$

Now let $C = A^{-1}$, then

$$\begin{aligned} A^{-1}A + A^{-1}B + DA + DB &= I \\ \Rightarrow A^{-1}B + DA + DB &= 0. \end{aligned} \tag{1.2}$$

Now let $D = A^{-1}MA^{-1}$, then

$$\begin{aligned} A^{-1}B + A^{-1}M + A^{-1}MA^{-1}B &= 0 \\ \Rightarrow A^{-1}\left(B + M + MA^{-1}B\right) &= 0 \\ \Rightarrow B + M + MA^{-1}B &= 0. \end{aligned} \tag{1.3}$$

Now let $M = BFB$, then

$$\begin{aligned} B + BFB + BFBA^{-1}B &= 0 \\ \Rightarrow B\left(I + FB + FBA^{-1}B\right) &= 0 \\ \Rightarrow I + FB + FBA^{-1}B &= 0 \\ \Rightarrow F(B + BA^{-1}B) &= -I \\ \Rightarrow F &= (B + BA^{-1}B)^{-1}. \end{aligned} \tag{1.4}$$

Therefore $D = A^{-1}B(B + BA^{-1}B)^{-1}BA^{-1}$. Substituting for C and D, the inverse of $A + B$ is then

$$(A + B)^{-1} = A^{-1} - A^{-1}B(B + BA^{-1}B)^{-1}BA^{-1} \tag{1.5}$$

Note that (1.5) only requires A to be invertible. If B is invertible (1.5) can be simplified to

$$(A + B)^{-1} = A^{-1} - A^{-1}B(I + A^{-1}B)^{-1}A^{-1} \tag{1.6}$$

or

$$(A + B)^{-1} = A^{-1} - A^{-1}(I + BA^{-1})BA^{-1} \tag{1.7}$$

1.1.2 Inverse of $A + BCD$

Using (1.5)

$$\begin{aligned}(A + BCD)^{-1} &= A^{-1} - A^{-1}BCD(BCD + BCDA^{-1}BCD)^{-1}BCDA^{-1} \\ &= A^{-1} + A^{-1}BCD(BC(D + DA^{-1}BCD))^{-1}BCDA^{-1} \\ &= A^{-1} + A^{-1}BCD(D + DA^{-1}BCD)DA^{-1} \\ &= A^{-1} + A^{-1}BCD((D(CD)^{-1} + DA^{-1}B)CD)^{-1}DA^{-1} \\ &= A^{-1} + A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \qquad (1.8)\end{aligned}$$

1.1.3 Inverse of $AB + CD$

Using (1.5)

$$\begin{aligned}(AB + CD)^{-1} &= B^{-1}A^{-1} - (AB)^{-1}CD(CD + CD(AB)^{-1}CD)^{-1}CD(AB)^{-1} \\ &= B^{-1}A^{-1} - B^{-1}A^{-1}CD(CD(I + (AB)^{-1}CD))^{-1}CDB^{-1}A^{-1} \\ &= B^{-1}A^{-1} - B^{-1}A^{-1}CD(I + B^{-1}A^{-1}CD)^{-1}B^{-1}A^{-1} \qquad (1.9)\end{aligned}$$

Chapter 2

Probability

2.1 Gaussian Distributions

2.1.1 Conditional Multivariate Gaussian Distribution

The PDF of a multivariate Gaussian vector is given by

$$P(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{2.1}$$

Let

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \tag{2.2}$$

and let $\boldsymbol{x}_2$ be given. We then wish to find $P(\boldsymbol{x}_1|\boldsymbol{x}_2)$, the conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2$. To do this we need to find the block form of Σ^{-1}.

Let

$$\Sigma = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}, \ \Sigma^{-1} = \begin{bmatrix} \tilde{A} & \tilde{B} \\ \tilde{B}^T & \tilde{C} \end{bmatrix} \tag{2.3}$$

. Using $\Sigma\Sigma^{-1} = I$ we then end up with the following four equations:

$$A\tilde{A} + B\tilde{B}^T = I, \tag{2.4}$$
$$A\tilde{B} + B\tilde{C} = 0, \tag{2.5}$$
$$B^T\tilde{A} + C\tilde{B}^T = 0, \tag{2.6}$$
$$B^T\tilde{B} + C\tilde{C} = I. \tag{2.7}$$

From (2.5) and (2.6) we obtain

$$\tilde{B} = -A^{-1}B\tilde{C}, \ \tilde{B}^T = -C^{-1}B^T\tilde{A} \tag{2.8}$$

Then inserting into (2.4) and (2.7) we get

$$(A - BC^{-1}B^T)\tilde{A} = I \Rightarrow \tilde{A} = (A - BC^{-1}B^T)^{-1} \tag{2.9}$$

$$(C - B^T A^{-1} B)\tilde{C} = I \Rightarrow \tilde{C} = (C - B^T A^{-1} B)^{-1} \tag{2.10}$$

Using (1.5) we then have

$$\begin{aligned}\tilde{A} &= A^{-1} - A^{-1}B(C + (-B^T)A^{-1}B)^{-1}(-B^T)A^{-1} \\ &= A^{-1} + A^{-1}B(C - B^T A^{-1} B)^{-1} B^T A^{-1}\end{aligned} \tag{2.11}$$

$$\begin{aligned}\tilde{C} &= C^{-1} - C^{-1}(-B^T)(A + BC^{-1}(-B^T))^{-1}BC^{-1} \\ &= C^{-1} + C^{-1}B^T(A - BC^{-1}B^T)^{-1}BC^{-1}\end{aligned} \tag{2.12}$$

Using the block form Σ^{-1},

$$\begin{aligned}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = {} & (\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T (A - BC^{-1}B^T)^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \\ & - 2(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T (A - BC^{-1}B^T)^{-1} BC^{-1} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \\ & + (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T (C^{-1} + C^{-1}B^T(A - BC^{-1}B^T)BC^{-1})(\boldsymbol{x}_2 - \boldsymbol{\mu}_2).\end{aligned} \tag{2.13}$$

Letting $(A - BC^{-1}B^T) = H$ and $BC^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) = \boldsymbol{b}$ we then have

$$\begin{aligned}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = {} & (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T C^{-1} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \\ & + (\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{b})^T H^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{b}).\end{aligned} \tag{2.14}$$

Using (2.14) we can write the joint density of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as

$$P(\boldsymbol{x}_1, \boldsymbol{x}_1) = \frac{1}{D} e^{(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T C^{-1} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2) + (\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{b})^T H^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{b})}. \tag{2.15}$$

From which it follows that

$$P(\boldsymbol{x}_2) = \int P(\boldsymbol{x}_1, \boldsymbol{x}_1) \, \mathrm{d}\boldsymbol{x}_1 = \frac{M}{D} e^{(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T C^{-1} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)} \tag{2.16}$$

Using Bayes' theorem, the conditional density of $\boldsymbol{x}_1$, given $\boldsymbol{x}_2$, is then

$$P(\boldsymbol{x}_1 | \boldsymbol{x}_2) = \frac{P(\boldsymbol{x}_1, \boldsymbol{x}_2)}{P(\boldsymbol{x}_2)} = \frac{1}{M} e^{(\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{b})^T H^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{b})}. \tag{2.17}$$

So we see that the distribution of $\boldsymbol{x}_1 | \boldsymbol{x}_2$ is

$$(\boldsymbol{x}_1 | \boldsymbol{x}_2) \sim N(\boldsymbol{\mu}_1 + BC^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), (A - BC^{-1}B^T)) \tag{2.18}$$

Note that since Σ is positive definite, then C and C^{-1} are positive definite. Therefore $A - BC^{-1}B$ is smaller than A in some sense and $\boldsymbol{x}_1 | \boldsymbol{x}_2$ will have smaller variance depending on the correlation matrix B. Additionally the maximum likelihood estimate for $\boldsymbol{x}_1 | \boldsymbol{x}_2$ is simply $\boldsymbol{\mu}_1 + BC^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$, the conditional mean.

Bibliography