1 Gaussian distribution is normalized

(1) Σ symetric $\rightarrow \Sigma^{-1}$ symetric
Proof:

$$\begin{aligned}
\Sigma\Sigma^{-1} &= I \\
(\Sigma\Sigma^{-1})^T &= I \\
(\Sigma^{-1})^T\Sigma^T &= I \\
\Sigma = \Sigma^T \rightarrow \Sigma^{-1})^T\Sigma &= I \\
\Sigma^{-1}\Sigma = I \rightarrow \Sigma^{-1} &= (\Sigma^{-1})^T
\end{aligned}$$

Gaussian distribution:

$$N(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}(x-\mu)^T\Sigma^{-1}(x-\mu)$$

$$\begin{aligned}
\Delta^2 &= (x-\mu)^T\Sigma^{-1}(x-\mu) \\
&= x^{T-1}x - \mu^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu + \mu^T\Sigma^-1\mu
\end{aligned}$$

$$\begin{aligned}
\mu^T\Sigma^{-1}x &= a \\
(\mu^T\Sigma^{-1}x)^T &= a^T \\
x^T\Sigma^{-1}\mu &= a \\
\mu^T\Sigma^{-1}x &= x^T\Sigma^{-1}\mu
\end{aligned}$$

$\rightarrow \Delta^2 = x^{T-1}x - 2x^T\Sigma^{-1}\mu + const$

(2) $\Sigma\ symetric \rightarrow eigenvectors\ orthogonal$
$proof:$
$\lambda_i u_i = \Sigma u_i \rightarrow \lambda_i u_j^T u_i = u_j^T\Sigma u_i = (u_j\Sigma)^T u_i = (u_j\lambda_j)^T u_i = \lambda_j u_j^T u_i \rightarrow u_i\ and\ u_j\ orthogonal$

$$\begin{aligned}
\Sigma &= \sum_{i=1}^{D}\lambda_i u_i u_i^T = U\Lambda U^T \\
\Sigma^{-1} &= (U\Lambda U^T)^{-1} \\
\Sigma^{-1} &= U\Lambda^{-1}U^T \\
\Sigma^{-1} &= \sum_{i=1}^{D}\frac{1}{\lambda_i}u_i u_i^T
\end{aligned}$$

$$\begin{aligned}\Delta^2 &= (x-\mu)^T\Sigma^{-1}(x-\mu)\\ &= (x-\mu)^T\sum_{i=1}^{D}\frac{1}{\lambda_i}u_iu_i^T(x-\mu)\\ &= \sum_{i=1}^{D}\frac{1}{\lambda_i}(x-\mu)^Tu_iu_i^T(x-\mu)\end{aligned}$$

Let $y_i = u_i^T(x-\mu) \rightarrow \Delta^2 = \sum_{i=1}^{D}\frac{y_i}{\lambda_i}$

We have: $|\Sigma|^{1/2} = \prod_{i=1}^{D}\lambda_i^{1/2} \rightarrow p(y) = \prod_{i=1}^{D}\frac{1}{(2\pi\lambda_i)^{1/2}}e^{\frac{y_i^2}{-2\lambda_i}}$

$$\begin{aligned}\int_{-\infty}^{\infty}p(y)dy &= \int_{-\infty}^{\infty}\prod_{i=1}^{D}\frac{1}{(2\pi\lambda_i)^{1/2}}e^{\frac{y_i^2}{-2\lambda_i}}dy\\ &= \prod_{i=1}^{D}\frac{1}{(2\pi\lambda_i)^{1/2}}\int_{-\infty}^{\infty}e^{\frac{y_i^2}{-2\lambda_i}}dy\\ &= 1\end{aligned}$$

$\rightarrow$ *Gaussian distribution is normalized*

2 Conditional Gaussian distribution

$$x = \begin{bmatrix}x_a\\ x_b\end{bmatrix} \qquad \mu = \begin{bmatrix}\mu_a\\ \mu_b\end{bmatrix} \qquad \Sigma = \begin{bmatrix}\Sigma_{aa} & \Sigma_{ab}\\ \Sigma_{ba} & \Sigma_{bb}\end{bmatrix}$$

Let $\Lambda \equiv \Sigma^{-1} \rightarrow \Lambda = \begin{bmatrix}\Lambda_{aa} & \Lambda_{ab}\\ \Lambda_{ba} & \Lambda_{bb}\end{bmatrix}$

We have

$$\begin{aligned}-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) &= -\frac{1}{2}(x-\mu)^T\Lambda(x-\mu)\\ &= -\frac{1}{2}(x_a-\mu_a)^T\Lambda_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b)\\ &\quad - \frac{1}{2}(x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T\Lambda_{bb}(x_b-\mu_b)\end{aligned}$$

We calculate $p(x_a|x_b) \rightarrow x_b$ is regarded as a constant

$$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = -\frac{1}{2}x_a^T\Lambda_{aa}x_a + x_a^T(\Lambda_{aa}\mu_a - \Lambda_{ab}(x_b-\mu_b)) + const$$

It is quadratic form of xa hence conditional distribution $p(x_a|x_b)$ will be Gaussian, because this distribution is characterized by its mean and its variance. Compare with Gaussian distribution

$$-\frac{1}{2}\Delta^2 = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + const$$

$\rightarrow \Sigma_{a|b} = \Lambda_{aa}^{-1}$
$\rightarrow \Sigma_{a|b}^{-1}\mu_{a|b} = \Lambda_{aa}\mu_a - \Lambda_{ab}(x_b - \mu_b)$
$\rightarrow \mu_{a|b} = \Sigma_{a|b}(\Lambda_{aa}\mu_a - \Lambda_{ab}(x_b - \mu_b)) = \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b)$

By using Schur complement, with $M = (A - BD^{-1}C)^{-1}$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & -D^{-1}CMBD^{-1} \end{bmatrix}$$

$\rightarrow \Lambda_{aa} = (\Sigma_{aa}^{-1}\Sigma_{aa}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$
$\Lambda_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}$

As a result

$$\begin{aligned} \mu_{a|b} &= \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - x_a) \\ \Sigma_{a|b} &= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \\ p(a|b) &= N(x_{ab}|\mu_{a|b}, \Sigma_{a|b}) \end{aligned}$$

3 Marginal Gaussian distribution

$$p(x_a) = \int p(x_a, x_b)dx_b$$

We need to integrate out x_b by looking the quadratic form related to x_b
with $m = \Lambda_{bb}\mu_b - \Lambda_{ba}(x_a - \mu_a)$

$$\begin{aligned} -\frac{1}{2}(x-\mu)^T\Lambda(x-\mu) &= -\frac{1}{2}x_b^T\Lambda_{bb}x_b + x_b^Tm - \frac{1}{2}x_a^T\Lambda_{aa}x_a + x_a^T(\Lambda_{aa}\mu_a + \Lambda_{ab}\mu_b) + const \\ -\frac{1}{2}x_b^T\Lambda_{bb}x_b + x_b^Tm &= -\frac{1}{2}(x_b - \Lambda_{bb}^{-1}m)^T\Lambda_{bb}(x_b - \Lambda_{bb}^{-1}m) + \frac{1}{2}m^T\Lambda_{bb}^{-1}m \\ \rightarrow -\frac{1}{2}(x-\mu)^T\Lambda(x-\mu) &= -\frac{1}{2}(x_b - \Lambda_{bb}^{-1}m)^T\Lambda_{bb}(x_b - \Lambda_{bb}^{-1}m) \\ &\quad + \frac{1}{2}m^T\Lambda_{bb}^{-1}m - \frac{1}{2}x_a^T\Lambda_{aa}x_a + x_a^T(\Lambda_{aa}\mu_a + \Lambda_{ab}\mu_b) + const \\ &= -\frac{1}{2}(x_b - \Lambda_{bb}^{-1}m)^T\Lambda_{bb}(x_b - \Lambda_{bb}^{-1}m) \\ &\quad - \frac{1}{2}x_a^T(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})x_a + x_a^T(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})^{-1}\mu_a + const \end{aligned}$$

We can integrate over unnormalized Gaussian:

$$\int e^{-\frac{1}{2}(x_b-\Lambda_{bb}^{-1}m)^T\Lambda_{bb}(x_b-\Lambda_{bb}^{-1}m)}dx_b$$

The remaining term:

$$-\frac{1}{2}x_a^T(\Lambda_{aa}-\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})x_a+x_a^T(\Lambda_{aa}-\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})^{-1}\mu_a+const$$

Compare with $-\frac{1}{2}\Delta^2=-\frac{1}{2}x^T\Sigma^{-1}x+x^T\Sigma^{-1}\mu+const$
We have:

$$\begin{aligned} E[x_a] &= \mu_a \\ cov[x_a] &= \Sigma_{aa} \\ p(x_a) &= N(x_a|\mu_a,\Sigma_{aa}) \end{aligned}$$