Completing Square

Pengfei Diao

I. Technique of Completing Square

The **completing square** is a technique to rearrange the second order polynomial of the form ax^2+bx+c into the form $a(x+h)^2+k$. All you have to bear in mind for completing the square is $(x+h)^2 = x^2+2h+h^2$, and the rests are just elementary manipulations of constants. It generally takes 3 steps to complete square for ax^2+bx+c :
1. Pull out the constant a from ax^2+bx+c by multiplying $\frac{1}{a}$ to each term, so we have

$$\begin{aligned} & a(\frac{1}{a}ax^2+\frac{1}{a}bx+\frac{1}{a}c) \\ = \quad & a(x^2+\frac{b}{a}x+\frac{c}{a}) \end{aligned} \tag{1}$$

2. Introduce two new terms $(\frac{b}{2a})^2-(\frac{b}{2a})^2$ into the polynomial from (1) and collect terms as

$$\begin{aligned} & a[x^2+\frac{b}{a}x+\frac{c}{a}+(\frac{b}{2a})^2-(\frac{b}{2a})^2] \\ = \quad & a\{[x^2+2\frac{b}{2a}x+(\frac{b}{2a})^2]+[\frac{c}{a}-(\frac{b}{2a})^2]\} \end{aligned} \tag{2}$$

3. Since the first three terms in the polynomial of (2) has already the form x^2+2h+h^2, we can rewrite them in the form $(x+h)^2$ and pull out the last two terms, and we have

$$a(x+\frac{b}{2a})^2+a[\frac{c}{a}-(\frac{b}{2a})^2] \tag{3}$$

which finishes completing the square. In fact as you might notice that when $\frac{c}{a}=(\frac{b}{2a})^2$, we can direct jump from step 1 to step 3.

As an example here we show how to complete the square for polynomial $-\frac{x^2}{4}-\frac{3}{2}x-4$. Following the steps above, we first pull the multiplier of x^2 out of the polynomial as

$$\begin{aligned} & -\frac{1}{4}\{(-4)\cdot(-\frac{x^2}{4})+(-4)\cdot(-\frac{3}{2}x)+(-4)\cdot -4)\} \\ = \quad & -\frac{1}{4}(x^2+6x+16) \end{aligned} \tag{4}$$

Then we compute the added terms in step 2 and insert them into the proper place of the polynomial from (4) as

$$\begin{aligned} & -\frac{1}{4}(x^2+6x+16) \\ = & -\frac{1}{4}(x^2+2\cdot 3x+3^2+16-3^2) \\ = & -\frac{1}{4}(x^2+2\cdot 3x+3^2+7) \end{aligned} \tag{5}$$

Finally we rewrite the first three terms of the polynomial from (5) in form $(x+h)^2$ and pull out the last term we obtain

$$\begin{aligned} & -\frac{1}{4}(x^2+2\cdot 3x+3^2+7) \\ = & -\frac{1}{4}(x+3)^2-\frac{7}{4} \end{aligned}$$

II. Marginalize Gaussian Distribution

You will find completing square happens quite often when manipulating Gaussian functions. As a illustration we will show how completing square can benefit computing the marginal distribution of the 2-D Gaussian density function. The idea can be extend to computing the marginal distribution of Gaussian density of any dimension. Assume of that we have a 2-D Gaussian density $N(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ with $\boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\boldsymbol{\mu}=\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, and $\boldsymbol{\Sigma}=\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$. We want to marginalize it over x_2 to get the marginal distribution of x_1 which is defined as

$$\begin{aligned} p(x_1) &= \int_{x_2} N(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})dx_2 \\ &= \int_{x_2} \frac{1}{2\pi^{D/2}|\boldsymbol{\Sigma}|^{1/2}} exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma^{-1}}(\boldsymbol{x}-\boldsymbol{\mu})\}dx_2 \\ &= \int_{x_2} Cexp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma^{-1}}(\boldsymbol{x}-\boldsymbol{\mu})\}dx_2 \end{aligned} \tag{6}$$

where $C=\dfrac{1}{2\pi^{D/2}|\boldsymbol{\Sigma}|^{1/2}}$

We shall eventually show that the right part of equation (6) can be reduced to an univariate Gaussian in the form $Cexp(-\frac{1}{2\sigma^2}(x_1-\mu)^2)$.

We start with working on the exponential part of (6), and denote the inverse covariance as $\boldsymbol{\Sigma^{-1}}=\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$. The exponential part could be written as

$$-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^T \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \tag{7}$$

Let $x_1 - \mu_1 = u$ and $x_2 - \mu_2 = v$ and substitute them into (7) we have

$$-\frac{1}{2}\begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{8}$$

Expanding (8) through vector-matrix multiplication we have

$$\begin{aligned} & -\frac{1}{2}\begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ = & -\frac{1}{2}\begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} \lambda_{11}u + \lambda_{12}v \\ \lambda_{21}u + \lambda_{22}v \end{bmatrix} \\ = & -\frac{1}{2}(\lambda_{11}u^2 + \lambda_{12}uv + \lambda_{21}uv + \lambda_{22}v^2) \end{aligned} \tag{9}$$

Since λ_{12} and λ_{21} are equal due to the symmetry of $\mathbf{\Sigma}^{-1}$, we can further reduce (9) to

$$-\frac{1}{2}(\lambda_{11}u^2 + 2\lambda_{12}uv + \lambda_{22}v^2) \tag{10}$$

We want integrate out x_2, therefore any term which does not involve v (recall that $v = x_2 - \mu_2$) can be, for the moment, put aside, and we complete square for terms involving v as

$$\begin{aligned} & -\frac{1}{2}\lambda_{11}u^2 - \frac{1}{2}(2\lambda_{12}uv + \lambda_{22}v^2) \\ = & -\frac{1}{2}\lambda_{11}u^2 - \frac{\lambda_{22}}{2}(v^2 + 2\frac{\lambda_{12}}{\lambda_{22}}uv) \\ = & -\frac{1}{2}\lambda_{11}u^2 - \frac{\lambda_{22}}{2}\{v^2 + 2\frac{\lambda_{12}}{\lambda_{22}}uv + (\frac{\lambda_{12}}{\lambda_{22}})^2u^2 - (\frac{\lambda_{12}}{\lambda_{22}})^2u^2\} \\ = & -\frac{1}{2}\lambda_{11}u^2 - \frac{\lambda_{22}}{2}\{\left[v - (\frac{\lambda_{12}}{\lambda_{22}})u\right]^2 - (\frac{\lambda_{12}}{\lambda_{22}})^2u^2\} \\ = & -\frac{1}{2}\lambda_{11}u^2 - \frac{\lambda_{22}}{2}\left[v - (\frac{\lambda_{12}}{\lambda_{22}})u\right]^2 - \frac{1}{2}\cdot(-\frac{\lambda_{12}^2}{\lambda_{22}})u^2 \\ = & -\frac{1}{2}(\lambda_{11} - \frac{\lambda_{12}^2}{\lambda_{22}})u^2 - \frac{\lambda_{22}}{2}\left[v - (\frac{\lambda_{12}}{\lambda_{22}})u\right]^2 \end{aligned} \tag{11}$$

Plugging x_1 and x_2 back into (11) we have

$$\begin{aligned} & -\frac{1}{2}(\lambda_{11} - \frac{\lambda_{12}^2}{\lambda_{22}})(x_1 - \mu_1)^2 - \frac{\lambda_{22}}{2}\left[x_2 - \mu_2 - (\frac{\lambda_{12}}{\lambda_{22}})(x_1 - \mu_1)\right]^2 \\ = & -\frac{1}{2}(\lambda_{11} - \frac{\lambda_{12}^2}{\lambda_{22}})(x_1 - \mu_1)^2 - \frac{\lambda_{22}}{2}\left[x_2 - (\frac{\lambda_{12}}{\lambda_{22}}x_1 - \frac{\lambda_{12}}{\lambda_{22}}\mu_1 + \mu_2)\right]^2 \end{aligned} \tag{12}$$

Now we come back to the integral (6) and we obtain

$$\begin{aligned}
& C\int_{x_2} exp\left\{-\frac{1}{2}(\lambda_{11}-\frac{\lambda_{12}^2}{\lambda_{22}})(x_1-\mu_1)^2-\frac{\lambda_{22}}{2}\left[x_2-(\frac{\lambda_{12}}{\lambda_{22}}x_1-\frac{\lambda_{12}}{\lambda_{22}}\mu_1+\mu_2)\right]^2\right\}dx_2 \\
= \;& C\int_{x_2} exp\left[-\frac{1}{2}(\lambda_{11}-\frac{\lambda_{12}^2}{\lambda_{22}})(x_1-\mu_1)^2\right] exp\{-\frac{\lambda_{22}}{2}\left[x_2-(\frac{\lambda_{12}}{\lambda_{22}}x_1-\frac{\lambda_{12}}{\lambda_{22}}\mu_1+\mu_2)\right]^2\}dx_2
\end{aligned}$$

As we can see the first exponential function is a function of x_1, and since we only integrate over x_2 we can pull it out of the integral as

$$= \;Cexp\left[-\frac{1}{2}(\lambda_{11}-\frac{\lambda_{12}^2}{\lambda_{22}})(x_1-\mu_1)^2\right]\int_{x_2} exp\{-\frac{\lambda_{22}}{2}\left[x_2-(\frac{\lambda_{12}}{\lambda_{22}}x_1-\frac{\lambda_{12}}{\lambda_{22}}\mu_1+\mu_2)\right]^2\}dx_2 \tag{13}$$

Now if we take a look at the second exponential function, it is obvious that for a fixed x_1 this function is an unnormalized univariate Gaussian(recall the general form of univariate Gaussian) whose mean is given as

$$\frac{\lambda_{12}}{\lambda_{22}}x_1-\frac{\lambda_{12}}{\lambda_{22}}\mu_1+\mu_2$$

It is also easy to see that changing the mean of a Gaussian does not change its shape, therefore the integral over this unnormalized Gaussian is constant that is

$$\begin{aligned}
f(x_1) &= \int_{x_2} exp\{-\frac{\lambda_{22}}{2}\left[x_2-(\frac{\lambda_{12}}{\lambda_{22}}x_1-\frac{\lambda_{12}}{\lambda_{22}}\mu_1+\mu_2)\right]^2\}dx_2 \\
&= K
\end{aligned} \tag{14}$$

We could therefore rewrite Expression (13) as

$$CK\cdot exp\left[-\frac{1}{2}(\lambda_{11}-\frac{\lambda_{12}^2}{\lambda_{22}})(x_1-\mu_1)^2\right]=p(x_1) \tag{15}$$

which is the formula for marginal distribution $p(x_1)$.
It is obvious that $p(x_1)$ is a Gaussian function whose mean μ is u_1 and variance σ^2 is $(\lambda_{11}-\frac{\lambda_{12}^2}{\lambda_{22}})^{-1}$. Recall our linear algebra text book we know that the inverse of a 2x2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is computed as

$$\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Hence the inverse covariance matrix Σ^{-1} of the original Gaussian density function is given as

$$\Sigma^{-1}=\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}=\frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21}}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix} \tag{16}$$

Due to the symmetry of covariance matrix $\mathbf{\Sigma}$, λ_{12} and λ_{21} are equal. We can rewrite (16) as

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \tag{17}$$

Let $M = \sigma_{11}\sigma_{22} - \sigma_{12}^2$ we then write down the variance σ^2 of the Gaussian derived from equation (15) in terms of σ from original covariance matrix as

$$\begin{aligned} \sigma^2 &= (\lambda_{11} - \frac{\lambda_{12}^2}{\lambda_{22}})^{-1} \\ &= \left[\frac{\sigma_{22}}{M} - (\frac{-\sigma_{12}}{M})^2 \frac{M}{\sigma_{11}}\right]^{-1} \\ &= (\frac{\sigma_{11}\sigma_{22} - \sigma 12^2}{M\sigma_{11}})^{-1} \\ &= (\frac{M}{M\sigma_{11}})^{-1} \\ &= \sigma_{11} \end{aligned} \tag{18}$$

Therefore we have derived the marginal distribution $p(x_1)$ as

$$p(x_1) = CK \cdot exp\left[-\frac{1}{2\sigma_{11}}(x_1 - \mu_1)^2\right] \tag{19}$$

where CK is a constant normalizer and, according to probability law, it must make the marginal distribution $p(x_1)$ fulfilling the condition

$$\int_{x_1} CK \cdot exp\left[-\frac{1}{2\sigma_{11}}(x_1 - \mu_1)^2\right] = 1 \tag{20}$$

Therefore CK equal to $\frac{1}{\sqrt{2\pi\sigma_{11}}}$, and $p(x_1)$ becomes again a Gaussian distribution $N(x_1|\mu = \mu_1, \sigma^2 = \sigma_{11})$. In fact, it is not hard to verify $CK = \frac{1}{\sqrt{2\pi\sigma_{11}}}$ by explicitly computing C and K in a similar fashion as how we derive (18).