设有一对服从多元正态分布的变量 (x , y) $(\boldsymbol{x}, \boldsymbol{y})$ (x, y)，可以写出他们的联合概率密度函数：

p (x , y) = N ([μ x μ y] , [Σ x x Σ x y Σ y x Σ y y])
$$p(\boldsymbol{x}, \boldsymbol{y})=\mathcal{N}\left(\left[\begin{array}{l}\boldsymbol{\mu}_{x} \\\boldsymbol{\mu}_{y}\end{array}\right],\left[\begin{array}{ll}\boldsymbol{\Sigma}_{x x} & \boldsymbol{\Sigma}_{x y} \\\boldsymbol{\Sigma}_{y x} & \boldsymbol{\Sigma}_{y y}\end{array}\right]\right)$$
p(x, y) = N ([μxμy], [ΣxxΣyxΣxyΣyy])

其中， Σ y x = Σ x y T $\boldsymbol{\Sigma}_{y x}=\boldsymbol{\Sigma}_{x y}^{\mathrm{T}}$ Σyx = ΣxyT。

由舒尔补有：

[Σ x x Σ x y Σ y x Σ y y] = [1 Σ x y Σ y y − 1 0 1] [Σ x x − Σ x y Σ y y − 1 Σ y x 0 0 Σ y y] [1 0 Σ y y − 1 Σ y x 1]
$$\left[\begin{array}{cc}\boldsymbol{\Sigma}_{x x} & \boldsymbol{\Sigma}_{x y} \\\boldsymbol{\Sigma}_{y x} & \boldsymbol{\Sigma}_{y y}\end{array}\right]=\left[\begin{array}{cc}\mathbf{1} & \boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \\\mathbf{0} & \mathbf{1}\end{array}\right]\left[\begin{array}{cc}\boldsymbol{\Sigma}_{x x}-\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x} & \mathbf{0} \\\mathbf{0} & \boldsymbol{\Sigma}_{y y}\end{array}\right]\left[\begin{array}{cc}\mathbf{1} & \mathbf{0} \\\boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x} & \mathbf{1}\end{array}\right]$$
[ΣxxΣyxΣxyΣyy] = [10ΣxyΣyy−11] [Σxx − ΣxyΣyy−1Σyx00Σyy] [1Σyy−1Σyx01]

对两边同时求逆有：

[Σ x x Σ x y Σ y x Σ y y] − 1 = [1 0 − Σ y y − 1 Σ y x 1] [(Σ x x − Σ x y Σ y y − 1 Σ y x) − 1 0 0 Σ y y − 1] [1 − Σ x y Σ y y − 1 0 1]
$${\left[\begin{array}{cc}\boldsymbol{\Sigma}_{x x} & \boldsymbol{\Sigma}_{x y} \\\boldsymbol{\Sigma}_{y x} & \boldsymbol{\Sigma}_{y y}\end{array}\right]^{-1}}= \left[\begin{array}{cc}\mathbf{1} & \mathbf{0} \\-\boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x} & \mathbf{1}\end{array}\right]} \left[\begin{array}{cc}\left(\boldsymbol{\Sigma}_{x x}-\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x}\right)^{-1} & \boldsymbol{0} \\\boldsymbol{0} & \boldsymbol{\Sigma}_{y y}^{-1}\end{array}\right]\left[\begin{array}{cc}\mathbf{1} & -\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \\\mathbf{0} & \mathbf{1}\end{array}\right]$$
[ΣxxΣyxΣxyΣyy]−1 = [1−Σyy−1Σyx01] [(Σxx − ΣxyΣyy−1Σyx)−100Σyy−1] [10−ΣxyΣyy−11]

因此，联合概率密度函数 p (x , y) $p(\boldsymbol{x}, \boldsymbol{y})$ p(x, y) 指数部分的二次项为：

([x y] − [μ x μ y]) T [Σ x x Σ x y Σ y x Σ y y] − 1 ([x y] − [μ x μ y]) = ([x y] − [μ x μ y]) T [1 0 − Σ y y − 1 Σ y x 1] [(Σ x x − Σ x y Σ y y − 1 Σ y x) − 1 0 0 Σ y y − 1] × [1 − Σ x y Σ y y − 1 0 1] ([x y] − [μ x μ y]) = (x − μ x − Σ x y Σ y y − 1 (y − μ y)) T (Σ x x − Σ x y Σ y y − 1 Σ y x) − 1 × (x − μ x − Σ x y Σ y y − 1 (y − μ y)) + (y − μ y) T Σ y y − 1 (y − μ y)
$$\begin{aligned}&\left(\left[\begin{array}{l}\boldsymbol{x} \\\boldsymbol{y}\end{array}\right]-\left[\begin{array}{l}\boldsymbol{\mu}_{x} \\\boldsymbol{\mu}_{y}\end{array}\right]\right)^{\mathrm{T}}\left[\begin{array}{ll}\boldsymbol{\Sigma}_{x x} & \boldsymbol{\Sigma}_{x y} \\\boldsymbol{\Sigma}_{y x} & \boldsymbol{\Sigma}_{y y}\end{array}\right]^{-1}\left(\left[\begin{array}{l}\boldsymbol{x} \\\boldsymbol{y}\end{array}\right]-\left[\begin{array}{l}\boldsymbol{\mu}_{x} \\\boldsymbol{\mu}_{y}\end{array}\right]\right) \\=&\left(\left[\begin{array}{l}\boldsymbol{x} \\\boldsymbol{y}\end{array}\right]-\left[\begin{array}{l}\boldsymbol{\mu}_{x} \\\boldsymbol{\mu}_{y}\end{array}\right]\right)^{\mathrm{T}}\left[\begin{array}{cc}\boldsymbol{1} & \boldsymbol{0} \\-\boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x} & \boldsymbol{1}\end{array}\right]\left[\begin{array}{cc}\left(\boldsymbol{\Sigma}_{x x}-\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x}\right)^{-1} & \boldsymbol{0} \\\mathbf{0} & \boldsymbol{\Sigma}_{y y}^{-1}\end{array}\right] \\& \times\left[\begin{array}{cc}\mathbf{1} & -\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \\\mathbf{0} & \mathbf{1}\end{array}\right]\left(\left[\begin{array}{l}\boldsymbol{x} \\\boldsymbol{y}\end{array}\right]-\left[\begin{array}{l}\boldsymbol{\mu}_{x} \\\boldsymbol{\mu}_{y}\end{array}\right]\right) \\=&\left(\boldsymbol{x}-\boldsymbol{\mu}_{x}-\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1}\left(\boldsymbol{y}-\boldsymbol{\mu}_{y}\right)\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{x x}-\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x}\right)^{-1} \\& \times\left(\boldsymbol{x}-\boldsymbol{\mu}_{x}-\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1}\left(\boldsymbol{y}-\boldsymbol{\mu}_{y}\right)\right)+\left(\boldsymbol{y}-\boldsymbol{\mu}_{y}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{y y}^{-1}\left(\boldsymbol{y}-\boldsymbol{\mu}_{y}\right)\end{aligned}$$
==([xy] − [μxμy])T[ΣxxΣyxΣxyΣyy]−1 ([xy] − [μxμy]) ([xy] − [μxμy])T[1−Σyy−1Σyx01] [(Σxx − ΣxyΣyy−1Σyx)−100Σyy−1] × [10−ΣxyΣyy−11] ([xy] − [μxμy]) (x − μx − ΣxyΣyy−1 (y − μy))T(Σxx − ΣxyΣyy−1Σyx)−1 × (x − μx − ΣxyΣyy−1 (y − μy)) + (y − μy)TΣyy−1 (y − μy)

很明显可以看出，这是两个二次项的和。

又由贝叶斯公式有：

p (x , y) = p (x ∣ y) p (y) $p(\boldsymbol{x}, \boldsymbol{y})=p(\boldsymbol{x} \mid \boldsymbol{y}) p(\boldsymbol{y})$
p(x, y) = p(x ∣ y)p(y)

并且：

p (y) = N (μ y , Σ y y) $p(\boldsymbol{y}) =\mathcal{N}\left(\boldsymbol{\mu}_{y}, \boldsymbol{\Sigma}_{y y}\right)$ p(y) = N (μy, Σyy)

因此，由幂运算中同底数幂相乘，底数不变、指数相加的性质，可以得到：

p (x ∣ y) = N (μ x + Σ x y Σ y y − 1 (y − μ y) , Σ x x − Σ x y Σ y y − 1 Σ y x) $p(\boldsymbol{x} \mid \boldsymbol{y})$

=\mathcal{N}\left(\boldsymbol{\mu}_{x}+\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1}\left(\boldsymbol{y}-\boldsymbol{\mu}_{y}\right), \boldsymbol{\Sigma}_{x x}-\boldsymbol{\Sigma}_{x y} \boldsymbol{\Sigma}_{y y}^{-1} \boldsymbol{\Sigma}_{y x}\right) $p(x \mid y) = \mathcal{N}(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$

这便是高斯推断中最重要的部分：从状态的先验概率分布出发，然后基于一些观测值来缩小这个范围。