* Conditioning a joint Gaussian distribution.

$$p(x_a, x_b) = \alpha \exp \underbrace{\left\{ -\frac{1}{2} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix} \right\}}_{\Delta}$$

Problem: $p(x_a \mid x_b)$ $\qquad x_a \in \mathbb{R}^n, \quad x_b \in \mathbb{R}^m$

$$\Sigma = \begin{bmatrix} \Sigma_a & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_b \end{bmatrix} \qquad \Sigma_{ab} = \Sigma_{ba}^T \quad (\Sigma \text{ symetric})$$

Information matrix $\Lambda = \Sigma^{-1}$, $\quad \Lambda = \begin{bmatrix} \Lambda_a & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_b \end{bmatrix}$

Expand the exponent Δ:

$$\Delta = -\frac{1}{2}(x_a-\mu_a)^T \Lambda_a (x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T \Lambda_{ab} (x_b-\mu_b)$$
$$- \frac{1}{2}(x_b-\mu_b)^T \Lambda_{ba} (x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T \Lambda_b (x_b-\mu_b).$$

Solution: Completing the square

intuition $\rightarrow$ we want an exponent only dependent on x_a since x_b is conditioning ("given").

$$\Delta = -\frac{1}{2} x_a^T \Sigma_{a|b}^{-1} x_a + x_a^T \Sigma_{a|b}^{-1} m - \frac{1}{2} m^T \Sigma_{a|b}^{-1} m + \text{ct}.$$

Q: what happens to x_b and constant terms?

* $\boxed{\Sigma_{a|b}^{-1} = \Lambda_a}$ I (2nd order term) $x_a^T \Sigma_{a|b}^{-1} x_a$

* 1st order terms:

$$x_a^T \cdot \underbrace{\left(\Lambda_a \mu_a - \Lambda_{ab}(x_b - \mu_b) \right)}_{\Sigma_{a|b}^{-1} \mu_{a|b}}$$

$$\Rightarrow \quad \mu_{a|b} = \Sigma_{a|b} \left(\Lambda_a \mu_a - \Lambda_{ab}(x_b - \mu_b) \right)$$

I

$$= \boxed{\mu_a - \Sigma_{a|b} \Lambda_{ab}(x_b - \mu_b)} \quad \mathrm{II}$$

We'll use the foll. matrix equality:

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

$$M = (A - BD^{-1}C)^{-1}$$ (M^{-1} Schur complement)

$$\begin{bmatrix} \Sigma_a & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_b \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_a & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_b \end{bmatrix}$$

$$\rightarrow \Lambda_a = (\overbrace{\Sigma_a - \Sigma_{ab}\Sigma_b^{-1}\Sigma_{ba}}^{\Sigma_{a|b}})^{-1}$$ (Schur complement)

$$\rightarrow \Lambda_{ab} = -\Lambda_a \cdot \Sigma_{ab}\Sigma_b^{-1}$$

$$\Rightarrow \mu_{a|b} = \mu_a - \underbrace{\cancel{\Sigma_{a|b}}(-\cancel{\Lambda_a}}_{I}\Sigma_{ab}\Sigma_b^{-1})(x_b - \mu_b)$$

$$= \mu_a + \Sigma_{ab}\Sigma_b^{-1}(x_b - \mu_b)$$

$$\boxed{p(x_a|x_b) = N(x_a; \mu_a + \Sigma_{ab}\Sigma_b^{-1}(x_b-\mu_b), \Sigma_a - \Sigma_{ab}\Sigma_b^{-1}\Sigma_{ba})}$$

* Marginalizing a joint Gaussian distribution.

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

(idea) solution: same as before, we will expand the exponent and complete the square, now twice

$$\int e^{\Delta} dx_b = e^{\Delta(x_a)} \underbrace{\int e^{\Delta(x_b)} dx_b}_{\eta} \cdot \underbrace{e^{\Delta(ctant)}}_{\eta'}$$

3.7

$$\Delta = -\frac{1}{2}(x_a-\mu_a)^T\Lambda_a(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b)$$

$$-\frac{1}{2}(x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_b) - \frac{1}{2}(x_b-\mu_b)^T\Lambda_b(x_b-\mu_b)$$

$$\Delta(x_b)\left\{\begin{aligned} = &-\frac{1}{2}x_b^T\Lambda_b x_b \qquad 1\\ &+ x_b^T\big(\underbrace{\Lambda_b\mu_b - \Lambda_{ba}(x_a-\mu_a)}_{\Lambda_b\cdot m}\big) \qquad 2\\ &-\frac{1}{2}m^T\Lambda_b\cdot m \end{aligned}\right.$$

$$\Lambda_b\cdot m \Rightarrow m = \mu_b - \Lambda_b^{-1}\Lambda_{ba}(x_a-\mu_a)$$

$$\Delta(x_a)\left\{\begin{aligned} &+\frac{1}{2}m^T\Lambda_b\cdot m\\ &-\frac{1}{2}x_a^T\Lambda_a x_a \qquad 3\\ &+x_a^T(\Lambda_a\mu_a+\Lambda_{ab}\mu_b) \qquad 4\\ &+\text{ctant} \end{aligned}\right.$$

$$(\Lambda_b^{-1})^T = \Lambda_b^{-1}$$

- $$\frac{1}{2}m^T\Lambda_b m = \frac{1}{2}\left(\mu_b - \Lambda_b^{-1}\Lambda_{ba}(x_a-\mu_a)\right)^T\cdot\Lambda_b\cdot(\cdot) \overset{?}{=}$$

$$= \frac{1}{2}x_a^T\Lambda_{ab}\cancel{\Lambda_b^{-1}}\cdot\cancel{\Lambda_b}\cdot\Lambda_b^{-1}\Lambda_{ba}x_a$$

$$- x_a^T\Lambda_{ab}\cancel{\Lambda_b^{-1}}\cdot\cancel{\Lambda_b}\left(\mu_b+\Lambda_b^{-1}\Lambda_{ba}\mu_a\right)$$

$$+\text{ctant.}$$

$$\Delta(x_a) = -\frac{1}{2} x_a^T \Lambda_a x_a + \frac{1}{2} x_a^T \Lambda_{ab} \Lambda_b^{-1} \Lambda_{ba} x_a$$

$$+ x_a^T \left(\Lambda_a \mu_a + \cancel{\Lambda_{ab} \mu_b} \right) - x_a^T \Lambda_{ab} \left(\cancel{\mu_b} + \Lambda_b^{-1} \Lambda_{ba} \mu_a \right)$$

$$+ \text{c'tant}$$

$$= -\frac{1}{2} x_a^T \Big(\underbrace{\Lambda_a - \Lambda_{ab} \Lambda_b^{-1} \Lambda_{ba}}_{\Sigma_a} \Big) x_a \qquad \text{(Schur complement!)}$$

$$+ x_a^T \left(\Lambda_a - \Lambda_{ab} \Lambda_b^{-1} \Lambda_{ba} \right) \mu_a$$

$$+ \text{c'tant.}$$

$$\boxed{p(x_a) = \int p(x_a, x_b)\, dx_b = \mathcal{N}(x_a; \mu_a, \Sigma_a)}$$

Gaussian again!

Marginalizing a Gaussian is as simple as selecting the submatrix inside Σ and the corresponding mean!

Gaussians are their self conjugate priors.

Prob Rob 2.4, 3.1, 3.2