1 KL divergence

$$D_{KL}(P\|Q) = \mathop{\mathbb{E}}_{x\sim P}\left[\log\left(\frac{p(x)}{q(x)}\right)\right]$$

2 Multivariate Normal

- **KL divergence**

$$\begin{aligned} D_{KL}\left(\mathcal{N}_a\|\mathcal{N}_b\right) &= \frac{1}{2}\left(\operatorname{Tr}\left(\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\Sigma}_a\right) + \left(\boldsymbol{\mu}_b-\boldsymbol{\mu}_a\right)\boldsymbol{\Sigma}_b^{-1}\left(\boldsymbol{\mu}_b-\boldsymbol{\mu}_a\right) - k + \log\frac{|\boldsymbol{\Sigma}_b|}{|\boldsymbol{\Sigma}_a|}\right) \\ &= \frac{1}{2}\left(\|\boldsymbol{L}_b \setminus \boldsymbol{L}_a\|_F^2 + \|\boldsymbol{L}_b \setminus (\boldsymbol{\mu}_b-\boldsymbol{\mu}_a)\|_2^2 - k + 2\left(\log\operatorname{diag}\left(\boldsymbol{L}_b\right)^T\underline{\mathbf{1}} - \log\operatorname{diag}\left(\boldsymbol{L}_a\right)^T\underline{\mathbf{1}}\right)\right) \end{aligned}$$

- **Linear Gaussian systems** Given a linear system:

$$\begin{aligned} p(\boldsymbol{x}) &= \mathcal{N}\left(x \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x\right) \\ p(\boldsymbol{y}\mid\boldsymbol{x}) &= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x}+b, \boldsymbol{\Sigma}_y\right) \end{aligned}$$

We have the following:

$$\begin{aligned} p(\boldsymbol{x}\mid\boldsymbol{y}) &= \mathcal{N}\left(x \mid \boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y}\right) \\ \boldsymbol{\mu}_{x|y} &= \boldsymbol{\Sigma}_{x|y}\left[\boldsymbol{A}^T\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}-\boldsymbol{b}) + \boldsymbol{\Sigma}_x^{-1}\boldsymbol{\mu}_x\right] \\ \boldsymbol{\Sigma}_{x|y} &= \boldsymbol{\Sigma}_x^{-1} + \boldsymbol{A}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{A} \end{aligned}$$

$$p(\boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{\mu}_x + \boldsymbol{b}, \boldsymbol{\Sigma}_y + \boldsymbol{A}\boldsymbol{\Sigma}_x\boldsymbol{A}^T\right)$$

- **quadratic relations**
  - $\mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\right] = \operatorname{Tr}\left(\boldsymbol{A}\boldsymbol{\Sigma}\right) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}$
  -
$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{b},\boldsymbol{B})}\left[\left(\boldsymbol{a}-\boldsymbol{A}\boldsymbol{x}\right)^T\Sigma^{-1}\left(\boldsymbol{a}-\boldsymbol{A}\boldsymbol{x}\right)\right] = \left(\boldsymbol{a}-\boldsymbol{A}\boldsymbol{b}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{a}-\boldsymbol{A}\boldsymbol{b}\right) + \operatorname{Tr}\left(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{B}\right)$$