PRELIMINARIES FOR LECTURE 3

M-V Gaussian characterization

$x \sim N_D(\mu, \Sigma) \qquad \Sigma = E \Lambda E^T \qquad E^T E = I \qquad \Lambda = diag(\lambda_1, \dots, \lambda_D)$

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

$y = \Lambda^{-\frac{1}{2}} E^T (x - \mu) \qquad g(y)\,dy = f(x)\,dx$

$$\boxed{g(y) = \underbrace{f(x)}_{\text{in terms of } y} \overbrace{\left|\frac{dx}{dy}\right|}^{\text{Jacobian}}}$$

Formula for multivariate transformation

$x = E \Lambda^{\frac{1}{2}} y + \mu \qquad \frac{dx}{dy} = E \Lambda^{\frac{1}{2}}$

$$\underline{\underline{g(y)}} = f(E\Lambda^{\frac{1}{2}} y + \mu) \times |J| = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-\frac{1}{2}} \, e^{-\frac{1}{2}(E\Lambda^{\frac{1}{2}} y)^T \Sigma^{-1} E \Lambda^{\frac{1}{2}} y} \, |E \Lambda^{\frac{1}{2}}|$$

$$= (2\pi)^{-\frac{D}{2}} \cancel{|E|^{-\frac{1}{2}}} \cancel{|\Lambda|^{-\frac{1}{2}}} \cancel{|E^T|^{-\frac{1}{2}}} \, e^{-\frac{1}{2} y^T \Lambda^{\frac{1}{2}} E^T (E^T)^{-1} \Lambda^{-1} E^{-1} E \Lambda^{\frac{1}{2}} y} \, \cancel{|E|} \cancel{|\Lambda|^{\frac{1}{2}}}$$

$$= (2\pi)^{-\frac{D}{2}} e^{-\frac{1}{2} y^T y} \equiv \underline{\underline{N_D(y \mid 0, I)}}$$

Completing the square

Claim: $\frac{1}{2} x^T A x - b^T x = \frac{1}{2}(x - A^{-1}b)^T A (x - A^{-1} b) - \frac{1}{2} b^T A^{-1} b$

Proof: $\frac{1}{2}(x - A^{-1}b)^T A (x - A^{-1}b) = \frac{1}{2} x^T A x - \frac{1}{2} x^T A A^{-1} b - \frac{1}{2} b^T (A^T)^{-1} A x + \frac{1}{2} b^T (A^T)^{-1} A A^{-1} b$

$= \frac{1}{2} x^T A x - b^T x + \frac{1}{2} b^T A^{-1} b$

> Note: this requires that $A^T = A$
> i.e. that A is symmetric.

Further: $\int \exp\left(-\frac{1}{2} x^T A x + b^T x\right) dx = \int \exp\left(-\frac{1}{2}(x - A^{-1}b)^T A (x - A^{-1}b)\right) \exp\left(\frac{1}{2} b^T A^{-1} b\right) dx$

$= \exp\left(\frac{1}{2} b^T A^{-1} b\right) (2\pi)^{\frac{D}{2}} |A|^{-\frac{1}{2}} \int (2\pi)^{-\frac{D}{2}} |A^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - A^{-1}b)^T A (x - A^{-1}b)\right) dx$

$= \exp\left(\frac{1}{2} b^T A^{-1} b\right) (2\pi)^{\frac{D}{2}} |A|^{-\frac{1}{2}} = \exp\left(\frac{1}{2} b^T A^{-1} b\right) |2\pi A^{-1}|^{\frac{1}{2}}$

Linear Gaussian systems

$$p(x) = N(x \mid \mu_x, \Sigma_x)$$

$$p(y \mid x) = N(y \mid Ax + b, \Sigma_y)$$

Then

$$\boxed{\begin{aligned} p(y) &= N(y \mid A\mu_x + b, \Sigma_y + A\Sigma_x A^T) \quad (*) \\ p(x \mid y) &= N(x \mid \mu_{x|y}, \Sigma_{x|y}) \\ \Sigma_{x|y}^{-1} &= \Sigma_x^{-1} + A^T \Sigma_y^{-1} A \\ \mu_{x|y} &= \Sigma_{x|y}\left[A^T \Sigma_y^{-1}(y-b) + \Sigma_x^{-1}\mu_x\right] \end{aligned}}$$

Proof of (*): $\log p(x,y) = -\frac{1}{2}(x-\mu_x)\Sigma_x^{-1}(x-\mu_x) - \frac{1}{2}(y-Ax-b)^T\Sigma_y^{-1}(y-Ax-b) + \text{constant}$

Because this is a quadratic form of x and y, this is a joint Gaussian.
Hence, we know that y has a Gaussian distribution (marginally).

$E(y) = E(E(y|x)) = E(Ax+b) = A\mu_x + b$ "law of total expectation"

$Var(y) = E(Var(y|x)) + Var(E(y|x))$ "law of total variance"

$= \Sigma_y + Var(Ax+b) = \Sigma_y + Var(Ax)$ "Wikipedia: $cov(AX+a) = A\, cov(X) A^T$"

$= \Sigma_y + A\, Var(x) A^T = \Sigma_y + A\Sigma_x A^T$ $\square$