086761 - Homework 2

Yuri Feldman, 309467801 yurif@cs.technion.ac.il
Alexander Shender, 328626114 aka.sova@gmail.com

November 18, 2017

1

1.1

Substitute $\Lambda^{-1} = \Sigma$, $\eta = \Sigma^{-1}\mu$:

$$p(x) = \frac{\exp\left(-\frac{1}{2}\mu^T\Sigma^{-T}\Sigma\Sigma^{-1}\mu\right)}{\sqrt{\det 2\pi\Sigma}} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\right) = \tag{1.1}$$

$\Sigma^{-1} = \Sigma^{-T}$ since it is symmetric

$$= \frac{\exp\left(-\frac{1}{2}\left(x^T\Sigma^{-1}x - 2\mu^T\Sigma^{-1} + \mu^T\Sigma^{-1}\mu\right)\right)}{\sqrt{\det 2\pi\Sigma}} = \tag{1.2}$$

$$= \frac{1}{\sqrt{\det 2\pi\Sigma}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} \tag{1.3}$$

the latter is the original multivariate Gaussian.

2

2.1

$$p(x) = N(\hat{x_0}, \Sigma_0) = \frac{1}{\sqrt{\det(2\pi\Sigma_0)}} e^{-\frac{1}{2}\|x-x_0\|^2_{\Sigma_0}} \tag{2.1}$$

$$p(z \mid x) = N(h(x), \Sigma_v) = \frac{1}{\sqrt{\det(2\pi\Sigma_v)}} e^{-\frac{1}{2}\|z-h(x)\|^2_{\Sigma_v}} \tag{2.2}$$

2.2

$$p(x \mid z_1) = \frac{p(z_1 \mid x)p(x)}{\int_x p(z_1 \mid x)p(x)dx} \tag{2.3}$$

2.3

We focus on the exponent (omitting the $-\frac{1}{2}$ factor) in the above product, since all the rest is normalization factor (does not depend on x).

$$\| h(x) - z_1 \|^2_{\Sigma_v} + \| x - \hat{x_0} \|^2_{\Sigma_0} \approx \tag{2.4}$$

linearization around $\hat{x_0}$, H the Jacobian of h at $\hat{x}_0$:

$$\approx \|h(\hat{x_0}) + H(x - \hat{x_0}) - z_1\|^2_{\Sigma_v} + \|x - \hat{x}_0\|^2_{\Sigma_0} = \tag{2.5}$$

$$= x^T \left(H^T \Sigma_v^{-1} H + \Sigma_0^{-1}\right) x + 2\left((h(\hat{x}_0) - H\hat{x}_0 - z_1)^T \Sigma_v^{-1} H - \hat{x}_0^T \Sigma_0^{-1}\right) x + \dots \tag{2.6}$$

equating terms with the exponent of a MVN, we get:

$$\Sigma_1 = \left(H^T \Sigma_v^{-1} H + \Sigma_0^{-1}\right)^{-1} \tag{2.7}$$

$$\hat{x}_1 = \Sigma_1 \left(\Sigma_0^{-1}\hat{x}_0 - H^T \Sigma_v^{-1}(h(\hat{x}_0) - H\hat{x}_0 - z_1)\right) \tag{2.8}$$

Can iteratively use obtained $\hat{x}_1$ as the new linearization point (re-calculate jacobian H) to improve estimate until convergence (Gauss Newton).

2.4

$$p(x \mid z_1, z_2) \propto p(z_1, z_2 \mid x)p(x) = \tag{2.9}$$

assume noise v_1, v_2 independent

$$= p(z_1 \mid x)p(z_2 \mid x)p(x) \tag{2.10}$$

We again get a Gaussian, can obtain mean and variance by minimizing the exponent (again omitting the $-\frac{1}{2}$ factor)

$$\|h(x) - z_2\|^2_{\Sigma_v} + \|h(x) - z_1\|^2_{\Sigma_v} + \|x - \hat{x_0}\|^2_{\Sigma_0} \tag{2.11}$$

Again need to pick a linearization point (e.g. $\hat{x}_1$ is a sensible one), denote $\bar{x}$ as in the lecture, then

$$\approx \|Hx + h(\bar{x}) - H\bar{x} - z_2\|^2_{\Sigma_v} + \|Hx + h(\bar{x}) - H\bar{x} - z_1\|^2_{\Sigma_v} + \|x - \hat{x_0}\|^2_{\Sigma_0} \tag{2.12}$$

Following the same derivation as before, it's easy to see that

$$\Sigma_2 = \left(2H^T\Sigma_v^{-1}H + \Sigma_0^{-1}\right)^{-1} \tag{2.13}$$

$$\hat{x}_2 = \Sigma_2 \left(\Sigma_0^{-1}\hat{x}_0 - 2H^T\Sigma_v^{-1}(h(\hat{x}_0) - H\hat{x}_0 - \frac{z_1 + z_2}{2})\right) \tag{2.14}$$

And again as in the previous clause, we can re-linearize to get better esimates with Gauss-Newton iterations.

3

3.1

$$p(x_k \mid x_{k-1}, u_{k-1}) = N(f(x_{k-1}, u_{k-1}), \Sigma_w) = \frac{1}{\sqrt{\det(2\pi\Sigma_w)}} e^{-\frac{1}{2}\|x_k - f(x_{k-1}, u_{k-1})\|^2_{\Sigma_w}} \tag{3.1}$$

3.2

$$p(x_1 \mid z_1, u_0) = \frac{p(z_1 \mid x_1) \int_{x_0} p(x_1 \mid x_0, u_0) p(x_0) dx_0}{\int_{x_1, x_0} p(z_1 \mid x_1) p(x_1 \mid x_0, u_0) p(x_0) dx_1 dx_0} \tag{3.2}$$

3.3

From previous clause:

$$x_1^{MAP} \doteq \arg\max_{x_1} p(x_1 \mid z_1, u_0) = \arg\min_{x_1} - \log(p(z_1 \mid x_1) p(x_1)) = \tag{3.3}$$

$$\arg\min_{x_1} \|z_1 - h(x_1)\|^2_{\Sigma_v} - \log \int_{x_0} p(x_1 \mid x_0, u_0) p(x_0) dx_0 \tag{3.4}$$

The integrand has the form of a gaussian, however with the left term exponent with generally nonlinear dependence on x_0:

$$p(x_1 \mid x_0, u_0) p(x_0) \propto \exp\{-\frac{1}{2}\left(\|\Sigma_w^{-\frac{1}{2}}(x_1 - f(x_0, u_0))\|^2 + \|\Sigma_0^{-\frac{1}{2}}(x_0 - \hat{x}_0)\|^2\right)\} \tag{3.5}$$

Linearizing f around $\hat{x}_0$, h around $\hat{x}_1 = f(\hat{x}_0, u_0)$:

$$\approx \exp\{-\frac{1}{2}\left(\|\Sigma_w^{-\frac{1}{2}}(\hat{x}_1 + H_f(x_0 - \hat{x}_0) - x_1)\|^2 + \|\Sigma_0^{-\frac{1}{2}}(x_0 - \hat{x}_0)\|^2\right)\} = \tag{3.6}$$

$$= \exp\{-\frac{1}{2}\left(\|\Sigma_w^{-\frac{1}{2}}\begin{pmatrix} H_f & -I \end{pmatrix}\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} + \Sigma_w^{-\frac{1}{2}}(\hat{x}_1 - H_f \hat{x}_0)\|^2\right. \tag{3.7}$$

$$\left. + \|\begin{pmatrix} \Sigma_0^{-\frac{1}{2}} & 0 \end{pmatrix}\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - \Sigma_0^{-\frac{1}{2}}\hat{x}_0\|^2\right)\} \tag{3.8}$$

this term is a Gaussian over $x_{0:1}$, and its marginalization w.r.t. x_0 yields a Gaussian in x_1 only, and everything collapses to a (linear, if we also marginalize h around $\hat{x}_1$)

least-squares problem, of which the solution may be refined by repeated re-linearization as in the previous problem.

Could probably have a somewhat easier derivation if had marginalized the entire posterior w.r.t. x_0.

3.4

Joint covariance $\Sigma_{0:1}$ is of the same size as $x_{0:1}x_{0:1}^T$ (that is, square, with side the sum of vector dimensions). If $x_0 \in \mathbb{R}^n$ and $x_1 \in \mathbb{R}^m$ then $\Sigma_{0:1} \in \mathbb{R}^{n+m\times n+m}$, $\Sigma_{00} \in \mathbb{R}^{n\times n}$, $\Sigma_{11} \in \mathbb{R}^{m\times m}$, $\Sigma_{01} \in \mathbb{R}^{n\times m}$.

We are interested in

$$p(x_1 \mid u_0, z_1) = \int_{x_0} p(x_{0:1} \mid u_0, z_1) dx_0 \tag{3.9}$$

Marginalization in covariance form: $\Sigma_1' = \Sigma_{11}$. Marginalization in information form: $I_1' = I_{11} - I_{01}^T I_{00}^{-1} I_{01}$

We hence continue from Eq. (3.8). Denoting the integrand's information matrix as $I_{0:1}^-$, we get

$$I_{0:1}^- = \begin{pmatrix} H_f^T \Sigma_w^{-1} H_f + \Sigma_0^{-1} & -H_f^T \Sigma_w^{-1} \\ -\Sigma_w^{-1} H_f & \Sigma_w^{-1} \end{pmatrix} \tag{3.10}$$

Linearizing the "innovation" term around $\hat{x}_1$ we get (denote H_h the Jacobian of h at $\hat{x}_1$)

$$\|z_1 - h(x_1)\|_{\Sigma_v}^2 \approx \|h(\hat{x}_1) + H_h(x_1 - \hat{x}_1) - z_1\|_{\Sigma_v}^2 = \tag{3.11}$$

$$= \|\Sigma_v^{-\frac{1}{2}} \begin{pmatrix} 0 & H_h \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} + \Sigma_v^{-\frac{1}{2}} (h(\hat{x}_1) - H_h \hat{x}_1 - z_1)\|^2 \tag{3.12}$$

This means that the information matrix of the joint distribution is:

$$I_{0:1} = \begin{pmatrix} H_f^T \Sigma_w^{-1} H_f + \Sigma_0^{-1} & -H_f^T \Sigma_w^{-1} \\ -\Sigma_w^{-1} H_f & H_h^T \Sigma_v^{-1} H_h + \Sigma_w^{-1} \end{pmatrix} \tag{3.13}$$

From here, using marginalization formula:

$$I_1' = H_h^T \Sigma_v^{-1} H_h + \Sigma_w^{-1} - \Sigma_w^{-1} H_f \left(H_f^T \Sigma_w^{-1} H_f + \Sigma_0^{-1}\right)^{-1} H_f^T \Sigma_w^{-1} \tag{3.14}$$

And Σ_1' is the inverse of the above.