Probabilistic Robotics Course

Least Squares and Uncertainty

Giorgio Grisetti
grisetti@diag.uniroma1.it

Department of Computer, Control and Management Engineering
Sapienza University of Rome

Maximum Likelihood Estimation

Using

$$\mathbf{x}^* = \operatorname*{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{z})$$

- Bayes' Rule

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{z})}$$

$$\propto p(\mathbf{z}|\mathbf{x})$$

- Independence,

$$= \prod_i p(\mathbf{z}^{[i]}|\mathbf{x})$$

We can further simplify the task

Gaussian Assumption

Measurements affected by Gaussian noise

$$
\begin{aligned}
p(\mathbf{z}_i|\mathbf{x}) &= \mathcal{N}(\underbrace{\mathbf{z}^{[i]}}_{\text{measurement}};\underbrace{\mathbf{h}^{[i]}}_{\text{prediction}}(\underbrace{\mathbf{x}}_{\text{state}}),\mathbf{\Sigma}^{[i]}) \\
&\propto \exp\left(-\underbrace{(\mathbf{h}^{[i]}(\mathbf{x})-\mathbf{z}^{[i]})}_{\mathbf{e}^{[i]}(\mathbf{x})}{}^T\underbrace{\mathbf{\Sigma}^{[i]-1}}_{\mathbf{\Omega}^{[i]}}(\mathbf{h}^{[i]}(\mathbf{x})-\mathbf{z}^{[i]})\right)
\end{aligned}
$$

error function

Gaussian Assumption

Through Gaussian assumption

- Maximization becomes minimization
- Product turns into sum

$$
\begin{aligned}
\mathbf{x}^* &= \operatorname*{argmax}_x \prod_i p(\mathbf{z}^{[i]} | \mathbf{x}) \\
&= \operatorname*{argmax}_x \prod_i \exp(-\mathbf{e}^{[i]}(\mathbf{x})^T \boldsymbol{\Omega}^{[i]} \mathbf{e}^{[i]}(\mathbf{x})) \\
&= \operatorname*{argmin}_x \sum_i \mathbf{e}^{[i]}(\mathbf{x})^T \boldsymbol{\Omega}^{[i]} \mathbf{e}^{[i]}(\mathbf{x})
\end{aligned}
$$

Goal of the Lesson

Use our skills with Gaussian distribution to derive an estimate of the uncertainty of the solution, around the optimum reported by Gauss-Newton

- Chain rule to get $p(x,z)$
- Conditioning on z

Around the Optimum…

The measurement functions can be approximated by their Taylor expansion

$$\mathbf{z}^{[i]^*} = \mathbf{h}^{[i]}(\mathbf{x}^*)$$

$$\mathbf{h}^{[i]}(\boldsymbol{\Delta}\mathbf{x} + \mathbf{x}^*) \simeq \mathbf{z}^{[i]^*} + \underbrace{\left.\frac{\partial \mathbf{h}^{[i]}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*}}_{\mathbf{J}^{[i]}} \boldsymbol{\Delta}\mathbf{x}$$

$$p(\mathbf{z}^{[i]} | \boldsymbol{\Delta}\mathbf{x} + \mathbf{x}^*) \sim \mathcal{N}(\mathbf{J}^{[i]} \boldsymbol{\Delta}\mathbf{x} + \mathbf{z}^{[i]*}, \boldsymbol{\Omega}^{[i]^{-1}})$$

Consequences of Taylor

The conditional over ***all*** measurements is again a multivariate Gaussian

$$p(\mathbf{z}|\mathbf{\Delta x} + \mathbf{x}^*) \sim \mathcal{N}(\mu_{\mathbf{z}}, \mathbf{\Omega}_{\mathbf{z}}^{-1})$$

$$\mu_{\mathbf{z}} = \begin{pmatrix} \mathbf{J}^{[1]}\mathbf{\Delta x} + \mathbf{z}^{[1]*} \\ \mathbf{J}^{[2]}\mathbf{\Delta x} + \mathbf{z}^{[2]*} \\ \mathbf{J}^{[K]}\mathbf{\Delta x} + \mathbf{z}^{[K]*} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{J}^{[1]} \\ \mathbf{J}^{[2]} \\ \vdots \\ \mathbf{J}^{[K]} \end{pmatrix}}_{\mathbf{J}} \mathbf{\Delta x} + \underbrace{\begin{pmatrix} \mathbf{z}^{[1]*} \\ \mathbf{z}^{[2]*} \\ \vdots \\ \mathbf{z}^{[K]*} \end{pmatrix}}_{\mathbf{z}^*}$$

$$\mathbf{\Omega}_{\mathbf{z}} = \begin{pmatrix} \mathbf{\Omega}^{[1]} & & & \\ & \mathbf{\Omega}^{[2]} & & \\ & & \ddots & \\ & & & \mathbf{\Omega}^{[K]} \end{pmatrix}$$

Chain Rule

We know

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{x}^*, \mathbf{\Sigma}_\mathbf{x}).$$

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \mathbf{J}\underbrace{(\mathbf{x} - \mathbf{x}^*)}_{\mathbf{\Delta x}} + \mathbf{z}^*, \mathbf{\Sigma}_\mathbf{z})$$

We want to compute

$$p(\mathbf{x}, \mathbf{z}) = \mathcal{N}\left(\mu_{\mathbf{x},\mathbf{z}}, \mathbf{\Sigma}_{\mathbf{x},\mathbf{z}}\right)$$

The parameters are

$$\begin{aligned}
\mu_{\mathbf{x},\mathbf{z}} &= \begin{pmatrix} \mathbf{x}^* \\ \mathbf{z}^* \end{pmatrix} \\
\mathbf{\Omega}_{\mathbf{x},\mathbf{z}} &= \begin{pmatrix} \mathbf{J}^T \mathbf{\Omega}_\mathbf{z} \mathbf{J} + \mathbf{\Omega}_\mathbf{x} & -\mathbf{J}^T \mathbf{\Omega}_\mathbf{z} \\ -\mathbf{\Omega}_\mathbf{z} \mathbf{J}^T & \mathbf{\Omega}_\mathbf{z} \end{pmatrix}
\end{aligned}$$

Applying Chain Rule

$$p(\mathbf{\Delta x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_x})$$

$$\mathbf{\Sigma_x} = \infty \Rightarrow \mathbf{\Omega_x} = \mathbf{0}$$

$$p(\mathbf{\Delta x}, \mathbf{z}) \sim \mathcal{N}((0, \mathbf{z}^*)^T, \mathbf{\Omega}_{\mathbf{x},\mathbf{z}}^{-1})$$

Instead of working in x, we work on Delta x, that has the same covariance but it is centered in 0

We know nothing a priori about x, so the information matrix is set to 0

$$\mathbf{\Omega_{x,z}} = \begin{pmatrix} \mathbf{J}^T \mathbf{\Omega_z} \mathbf{J} & -\mathbf{J}^T \mathbf{\Omega_z} \\ -\mathbf{\Omega_z} \mathbf{J} & \mathbf{\Omega_z} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{\Omega_{xx}} & \mathbf{\Omega_{xz}} \\ \mathbf{\Omega_{zx}} & \mathbf{\Omega_{zz}} \end{pmatrix}$$

This is the joint information matrix of measurements and states

Applying Chain Rule

$$
\begin{aligned}
\boldsymbol{\Omega}_{\mathbf{xx}} &= \mathbf{J}^T \boldsymbol{\Omega}_{\mathbf{z}} \mathbf{J} \\
&= \begin{pmatrix} \mathbf{J}^{[1]T} & \mathbf{J}^{[2]T} & \cdots & \mathbf{J}^{[K]T} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}^{[1]} & & & \\ & \boldsymbol{\Omega}^{[2]} & & \\ & & \ddots & \\ & & & \boldsymbol{\Omega}^{[K]} \end{pmatrix} \begin{pmatrix} \mathbf{J}^{[1]} \\ \mathbf{J}^{[2]} \\ \vdots \\ \mathbf{J}^{[K]} \end{pmatrix} \\
&= \mathbf{J}^{[1]T} \boldsymbol{\Omega}^{[1]} \mathbf{J}^{[1]} + \cdots + \underbrace{\mathbf{J}^{[k]T} \boldsymbol{\Omega}^{[k]} \mathbf{J}^{[k]}}_{\mathbf{H}^{[k]}} + \cdots + \mathbf{J}^{[k]T} \boldsymbol{\Omega}^{[k]} \mathbf{J}^{[K]} \\
&= \sum_{k=1}^{K} \mathbf{H}^{[k]} \\
&= \mathbf{H}
\end{aligned}
$$

Conditioning

Let $(\mathbf{x}^T \ \mathbf{z}^T)$ be a Gaussian random variable such that

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_z \end{pmatrix}, \boldsymbol{\Sigma}_{\mathbf{x},\mathbf{z}}\right)$$

The conditional

$p(\mathbf{x} \mid \mathbf{z})$

is Gaussian with parameters

$\mathrm{N}(\mathrm{x};\ \mu_{\mathbf{x}|\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}})$

The farther is the measurement from the prediction, the more the mean changes

$$\mu_{\mathbf{x}|\mathbf{z}} = \mu_{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{xz}}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}(\mathbf{z} - \mu_{\mathbf{z}})$$

$$\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} = \boldsymbol{\Sigma}_{\mathbf{xx}} - \boldsymbol{\Sigma}_{\mathbf{xz}}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}\boldsymbol{\Sigma}_{\mathbf{zx}}$$

Conditioning

Let $(\mathbf{x}^T\ \mathbf{z}^T)$ be a Gaussian random variable such that

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_z \end{pmatrix}, \boldsymbol{\Sigma}_{\mathbf{x},\mathbf{z}}\right)$$

The conditional
$p(\mathbf{x} \mid \mathbf{z})$

is Gaussian with parameters

$$p(\mathbf{x}) = \mathcal{N}^{-1}(\mathbf{x}; \nu_{\mathbf{x}|\mathbf{z}}, \boldsymbol{\Omega}_{\mathbf{x}|\mathbf{z}})$$

$$\nu_{\mathbf{x}|\mathbf{z}} = \nu_{\mathbf{x}} - \boldsymbol{\Omega}_{\mathbf{x}\mathbf{z}}\mathbf{z}$$

$$\boldsymbol{\Omega}_{\mathbf{x}|\mathbf{z}} = \boldsymbol{\Omega}_{\mathbf{x}\mathbf{x}}$$

The information of the conditional is just the xx block!

Concluding

$$\mu_{\mathbf{x}|\mathbf{z}} = \mathbf{x}^* + \boldsymbol{\Sigma}_{\mathbf{xz}}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}(\mathbf{z} - \mathbf{z}^*)$$

The farther is the measurement from the prediction, the more the mean changes

$$\mathbf{z} \simeq \mathbf{z}^*$$

If least squares converged well, the prediction at the optimum is close to the measurements

$$\mu_{\mathbf{x}|\mathbf{z}} \simeq \mathbf{x}^*$$

The mean of the Gaussian approximation of the solution is the optimum found by LS

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} &= \boldsymbol{\Omega}_{\mathbf{xx}}^{-1} \\ &\simeq \mathbf{H}^{-1}\end{aligned}$$

The covariance matrix is the inverse of the H matrix