

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

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

- 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

The diagram illustrates the Gaussian Assumption equation for the probability of a measurement \mathbf{z}_i given a state \mathbf{x} . The equation is presented with color-coded terms and labels:

- prediction** (purple box) points to $\mathbf{h}^{[i]}$.
- measurement** (blue box) points to $\mathbf{z}^{[i]}$.
- state** (red box) points to \mathbf{x} .
- error function** (purple box) points to $\mathbf{e}^{[i]}(\mathbf{x})$.

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

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

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

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

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

Consequences of Taylor

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

$$p(\mathbf{z} | \Delta \mathbf{x} + \mathbf{x}^*) \sim \mathcal{N}(\mu_{\mathbf{z}}, \Omega_{\mathbf{z}}^{-1})$$
$$\mu_{\mathbf{z}} = \begin{pmatrix} \mathbf{J}^{[1]} \Delta \mathbf{x} + \mathbf{z}^{[1]*} \\ \mathbf{J}^{[2]} \Delta \mathbf{x} + \mathbf{z}^{[2]*} \\ \vdots \\ \mathbf{J}^{[K]} \Delta \mathbf{x} + \mathbf{z}^{[K]*} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{J}^{[1]} \\ \mathbf{J}^{[2]} \\ \vdots \\ \mathbf{J}^{[K]} \end{pmatrix}}_{\mathbf{J}} \Delta \mathbf{x} + \underbrace{\begin{pmatrix} \mathbf{z}^{[1]*} \\ \mathbf{z}^{[2]*} \\ \vdots \\ \mathbf{z}^{[K]*} \end{pmatrix}}_{\mathbf{z}^*}$$
$$\Omega_{\mathbf{z}} = \begin{pmatrix} \Omega^{[1]} & & & \\ & \Omega^{[2]} & & \\ & & \ddots & \\ & & & \Omega^{[K]} \end{pmatrix}$$

Chain Rule

We know

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

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

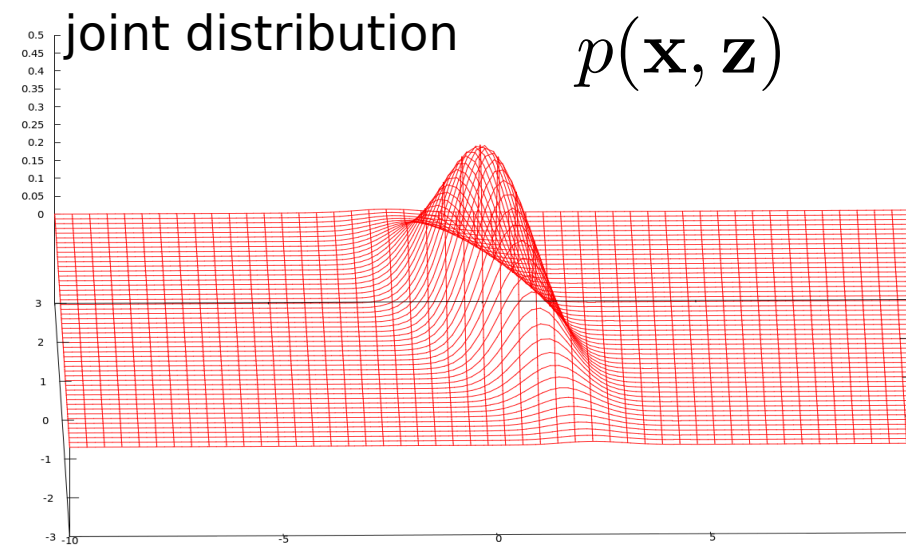
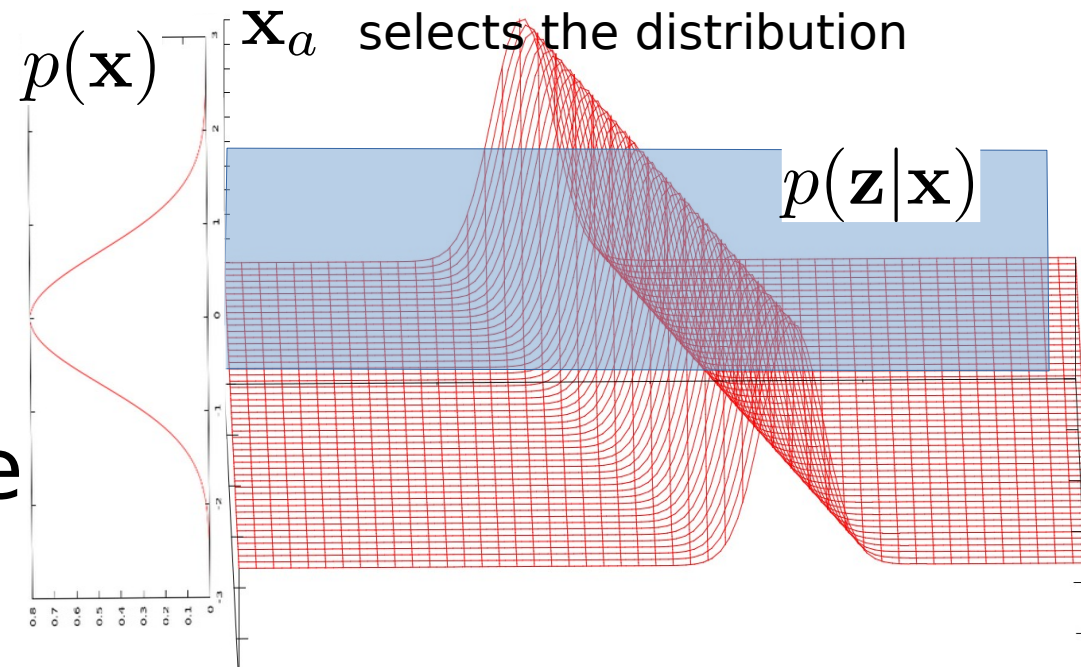
We want to compute

$$p(\mathbf{x}, \mathbf{z}) = \mathcal{N}(\mu_{\mathbf{x}, \mathbf{z}}, \Sigma_{\mathbf{x}, \mathbf{z}})$$

The parameters are

$$\mu_{\mathbf{x}, \mathbf{z}} = \begin{pmatrix} \mathbf{x}^* \\ \mathbf{z}^* \end{pmatrix}$$

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



Applying Chain Rule

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

$$\Sigma_{\mathbf{x}} = \infty \Rightarrow \Omega_{\mathbf{x}} = \mathbf{0}$$

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

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

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

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

This is the joint information matrix of measurements and states

Applying Chain Rule

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

Conditioning

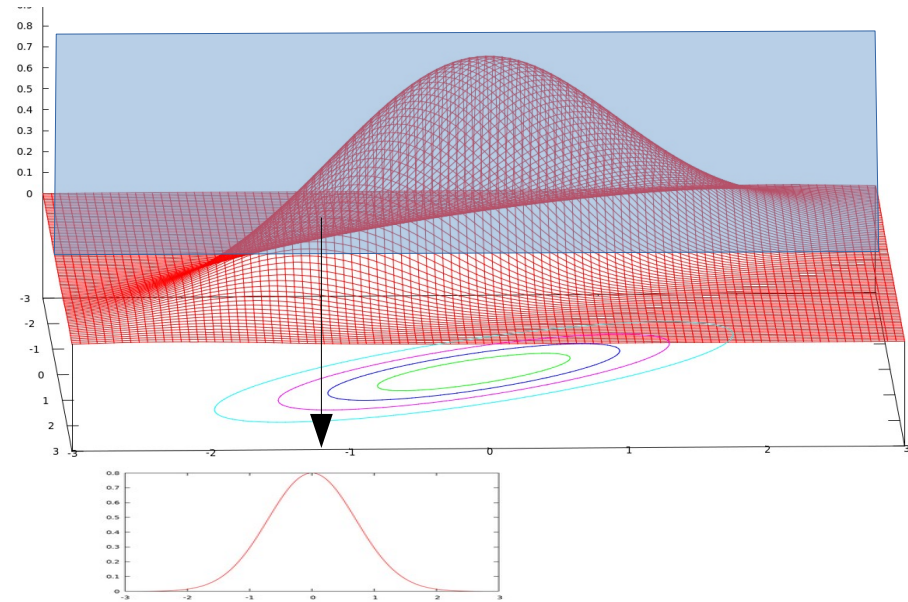
Let $(\mathbf{x}^T \mathbf{z}^T)$ be a Gaussian random variable such that

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

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

is Gaussian with parameters

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



$$\nu_{\mathbf{x}|\mathbf{z}} = \nu_{\mathbf{x}} - \Omega_{\mathbf{xz}}\mathbf{z}$$

$$\Omega_{\mathbf{x}|\mathbf{z}} = \Omega_{\mathbf{xx}}$$

The information of the conditional is just the xx block!

Concluding

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

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

$$\begin{aligned} \Sigma_{\mathbf{x}|\mathbf{z}} &= \Omega_{\mathbf{xx}}^{-1} \\ &\simeq \mathbf{H}^{-1} \end{aligned}$$

The covariance matrix is the inverse of the H matrix