~~2018~~ 2019 16833 ~~L11.3~~ ~~L12.2~~ L11.2

Inference

SLAM as a Maximum Likelihood Problem

given: measurements Z

unknown: variables Θ (e.g. state)

$p(\Theta|Z)$, inference problem: find best Θ, ie. Θ^* that maximizes $p(\Theta|Z)$

Objective: $\Theta^* = \underset{\Theta}{\operatorname{argmax}}\, p(\Theta|Z)$

generative sensor models $p(z_i|\Theta)$

$$\underset{\text{posterior}}{p(\Theta|Z)} \overset{\text{Bayes rule}}{=} \frac{\overset{\text{likelihood}}{p(Z|\Theta)}\ \overset{\text{prior}}{p(\Theta)}}{\underset{\text{evidence}}{p(Z)}}$$

evidence independent of Θ

$$\underbrace{\Theta^* = \underset{\Theta}{\operatorname{argmax}}\ p(Z|\Theta)\, p(\Theta)}_{\text{maximum a posteriori (MAP)}}$$

in the absence of prior information, we can simply maximize the likelihood

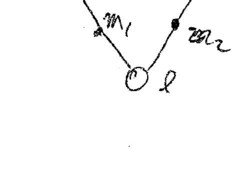
$$\Theta^* = \underset{\Theta}{\operatorname{argmax}}\ p(Z|\Theta) \quad \text{maximum likelihood (ML)}$$

$p(Z|\Theta) \rightarrow L(\Theta;Z)$
- to emphasize Θ is the variable
- does not need to be normalized

$p(Z|\Theta) = \prod p(z_i|\Theta)$ ~~factorization~~

because of conditional independence of measurements given Θ

this factorization corresponds to the factor graph



$p(Z|\Theta) = p(p|x_0)\, p(u_1|x_0,x_1)\, p(m_1|l,x_0)\, p(m_2|l,x_1)$

$Z = \{p, u_1, m_1, m_2\}$ $\Theta = \{x_0, x_1, l\}$

Least-Squares Problem

$$\Theta^* = \arg\max_\theta \prod p(z_i|\theta)$$

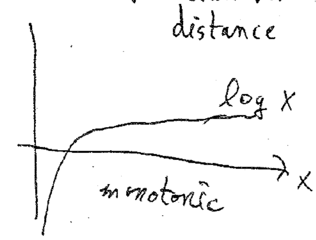
$$= \arg\max \prod \underbrace{\cancel{\frac{1}{\sqrt{(2\pi)^n|\Sigma_i|}}}}_{\text{const}} e^{-\frac{1}{2}\|h_i(\theta)-z_i\|^2_{\Sigma_i}}$$

$|\Sigma_i|$ determinant

$$(x-z)^T\Sigma^{-1}(x-z) = \|x-z\|^2_\Sigma$$

Mahalanobis distance

$$= \arg\max_\theta \log \prod e^{-\frac{1}{2}\|h_i(\theta)-z_i\|^2_{\Sigma_i}}$$



$$= \arg\max_\theta \sum \cancel{\log} \cancel{e}^{-\frac{1}{2}\|h_i(\theta)-z_i\|^2_{\Sigma_i}}$$

$$= \arg\min_\theta \sum \cancel{\tfrac{1}{2}} \|h_i(\theta)-z_i\|^2_{\Sigma_i} \quad \text{least squares}$$

if $h(\theta)$ linear: $h(x) = Hx + h_0$

$$\arg\min_x \sum \|H_i x + \underbrace{h_{0,i} - z_i}_{-d}\|^2_{\Sigma_i}$$

$$(Hx-d)^T\Sigma^{-1}(Hx-d) = (Hx-d)^T \Sigma^{-\frac{T}{2}}\Sigma^{-\frac{1}{2}}(Hx-d)$$

$$\underbrace{\Sigma^{-\frac{1}{2}}H}_{A_i} x - \underbrace{\Sigma^{-\frac{1}{2}}d}_{b_i}$$

$$\arg\min_x \sum_i \|A_i x - b_i\|^2 \leftarrow \text{L2 norm}$$

$$= \arg\min_x \|Ax-b\|^2$$

$A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$

$$\overset{n}{\underset{m}{\begin{bmatrix} A_0 \\ \hline A_1 \\ \hline \vdots \end{bmatrix}}} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \end{bmatrix}$$

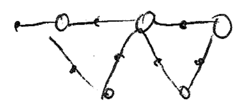
$$\|Ax-b\|^2 = (Ax-b)^T(Ax-b)$$
$$= x^TA^TAx - 2A^Tb\,x + b^Tb$$

how to find minimum? set derivative to 0

$$\frac{\partial \|Ax-b\|^2}{\partial x} \stackrel{!}{=} 0 : \quad 2A^TAx - 2A^Tb = 0$$
$$\boxed{A^TAx = A^Tb}$$ normal equations

solution: $x = \underbrace{(A^TA)^{-1}A^T}_{\text{pseudo inverse}}b$

recap:



$\underset{\theta}{\operatorname{argmax}} \prod p(z|\theta)$

$\underset{\theta}{\operatorname{argmin}} \sum \|h_i(\theta)\|^2$

$\underset{\theta}{\operatorname{argmin}} \|A\theta - b\|^2$

⚠ note: we have not solved $Ax=b$

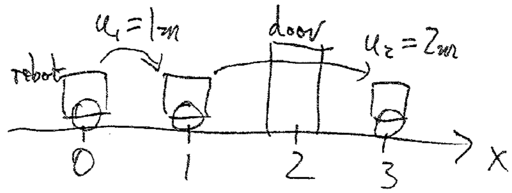
in general no exact solution because of noise!

instead we have solved $\underset{x}{\operatorname{argmin}} \|Ax-b\|^2$

why is $(A^TA)^{-1}$ a bad idea? $O(n^3)$ $O(n^{2.373})$ but with a huge constant

in SLAM $n \gg 1000$

1D Example

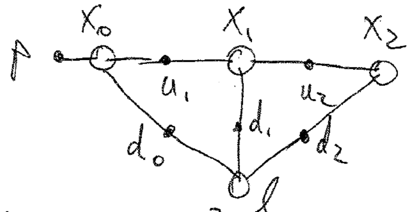


$t=0$ $t=1$ $t=2$

$d_0 = 2m$ $d_1 = 1m$ $d_2 = -1m$ measurement: distance, signed

$$\Big(\underbrace{\frac{h_u(x_0,x_1)}{\sigma_u}}_{A_i\theta} - \underbrace{\frac{u_1}{\sigma_u}}_{b_i}\Big)^2$$

factor graph:



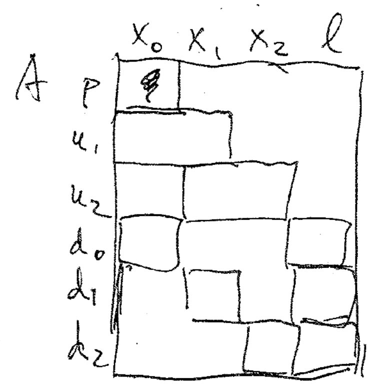
scalar: $\frac{1}{\sigma_u^2}\left(h_u(x_0,x_1) - u_1\right)^2$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum \| h_i(\theta) - z_i \|^2_{\Sigma_i} = \| h_p(x_0) - p\|^2_{\Sigma_p} + \| h_u(x_0,x_1) - u_1\|^2_{\Sigma_u}$$
$$+ \| h_u(x_1,x_2) - u_2\|^2_{\Sigma_u} + \| h_d(x_0,l) - d_0\|^2_{\Sigma_d} + \| h_d(x_1,l) - d_1\|^2_{\Sigma_d}$$
$$+ \| h_d(x_2,l) - d_2\|^2_{\Sigma_d}$$

$h_p(x) = x$, $p=0$, $\sigma_u = \cancel{0.1m}\ 0.01m$

$h_u(x_A, x_B) = x_B - x_A$, $\sigma_u = \cancel{0.01m}\ 0.1m$

$h_d(x, l) = l - x$, $\sigma_d = 0.01m$



A is sparse! $\Updownarrow$ structure of the factor graph

A

	x_0	x_1	x_2	l
p	100			
u_1	-10	+10		
u_2		-10	10	
d_1	-100			100
d_2		-100		100
d_3			-100	100

b
0
10
20
200
100
-100

normal equations $A^TA x = A^Tb$

$$x = (A^TA)^{-1} A^Tb$$

$$= \begin{array}{c} \begin{bmatrix} 0 \\ 1 \\ 3 \\ 2 \end{bmatrix} \end{array} \begin{array}{c} x_0 \\ x_1 \\ x_2 \\ l \end{array}$$

how to avoid inverse?

~~2018~~ 2019 16833 ~~L13.3~~ ~~L14.1~~ L12.3 L13.1

back to normal equations : $\underbrace{A^TA}_{\text{information matrix, Hessian}} x = A^Tb$

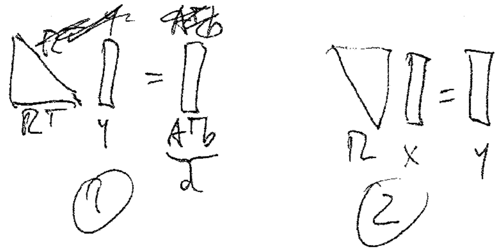
Solve by sparse matrix factorization

Cholesky : $A^TA = R^TR$, where R is upper triangular

$$R^TR\,x = A^Tb$$

solve by forward-/backsubstitution

$$R^T y = A^Tb \,, \quad Rx = y$$



$R_{1,1}\,y_1 = d_1 \quad y_1 = \frac{d_1}{R_{1,1}}$

$R_{2,1}\,y_1 + R_{2,2}\,y_2 = d_2 \quad y_2 = \dots$

matlab $(A^TA)\backslash(A^Tb)$ — square = exact

$A\backslash b$ — rectangular → least-sq. ⇕ normal eq.? vs A, b

LDL^T : faster than Cholesky, avoids n square roots

QR : more numerically stable, but slower
works directly on A

~~$A = QR$~~ $\quad A = Q\begin{bmatrix} R \\ 0 \end{bmatrix}$ (Q: square, orthonormal) $\quad \underset{x}{\operatorname{argmin}} \|Ax-b\|^2$

~~$Q\begin{bmatrix} R \\ 0 \end{bmatrix}x = b$~~

$$\|Ax-b\|^2 = \left\| Q\begin{bmatrix} R \\ 0 \end{bmatrix}x - b \right\|^2$$

$$\overset{Q \text{ orthonormal}}{=} \left\| Q^TQ\begin{bmatrix} R \\ 0 \end{bmatrix}x - Q^Tb \right\|^2$$

$$= \left\| \begin{bmatrix} R \\ 0 \end{bmatrix}x - \begin{bmatrix} d \\ e \end{bmatrix} \right\|^2$$

$$= \underbrace{\|Rx-d\|^2}_{\text{exact solution}} + \underbrace{\|e\|^2}_{\text{residual}}$$

Q is never explicitly formed
instead, we maintain Q^Tb

$Rx=d$, solve by backsub
$d=y$ in Cholesky !!

⚠ note: $A \neq QR$, $Q \in \mathbb{R}^{m\times m}$, $R \in \mathbb{R}^{n\times n}$

Nonlinear Least-Squares

$$\operatorname{argmin} \sum_i \| h_i(X) - z_i \|^2_{\Sigma_i}$$

linear case: $h(X) = h_0 + HX$

solution directly by solving normal equations

nonlinear case: no direct solution possible, have to iterate

need: initial estimate X^0

linearize h with X^0 as linearization point

Taylor expansion:

- scalar valued function with one variable

$$h(X) = h(X^0+\Delta) = h(X^0) + h'(X^0)\,\Delta + \frac{1}{2} h''(X^0)\Delta^2 + \underbrace{\frac{1}{6} h'''(X^0)\Delta^3 + \dots}_{\text{often small}}$$

Δ is now the variable, X^0 is a constant

- general case, $X \in \mathbb{R}^n$, $h(X) \in \mathbb{R}^m$

$$h(X) = h(X^0+\Delta) \approx h(X^0) + H\Delta$$

drop higher order terms
linear approx.
first order Taylor approx.

$$\underset{\text{Jacobian}}{H} = \left.\frac{\partial h(X)}{\partial X}\right|_{X=X^0} = \begin{pmatrix} \left.\frac{\partial h_1(X)}{\partial X_1}\right|_{X_1=X_1^0} & \cdots & \left.\frac{\partial h_1(X)}{\partial X_n}\right|_{X_n=X_n^0} \\ \left.\frac{\partial h_2(X)}{\partial X_1}\right|_{X_1=X_1^0} & \ddots & \vdots \\ \vdots & & \\ \left.\frac{\partial h_m(X)}{\partial X_1}\right|_{X_1=X_1^0} & \cdots & \left.\frac{\partial h_m(X)}{\partial X_n}\right|_{X_n=X_n^0} \end{pmatrix}$$

Hessian: $n \times n \times m$

variables $X \in \mathbb{R}^n$ measurements $Z = \{z_i\} \in \mathbb{R}^m$, initial estimate X^0

least-squares:

$$\text{argmin} \underbrace{\sum_i \| h_i(X) - z_i \|^2_{\Sigma_i}}_{g(X)}$$

Taylor expansion: $h(X^0 + \Delta) \approx h(X^0) + \underbrace{\left.\frac{\partial h(X)}{\partial X}\right|_{X=X^0}}_{H \text{ Jacobian}} \cdot \Delta$ $\quad \Delta = X - X^0$

(higher order terms Hessian/tensor)

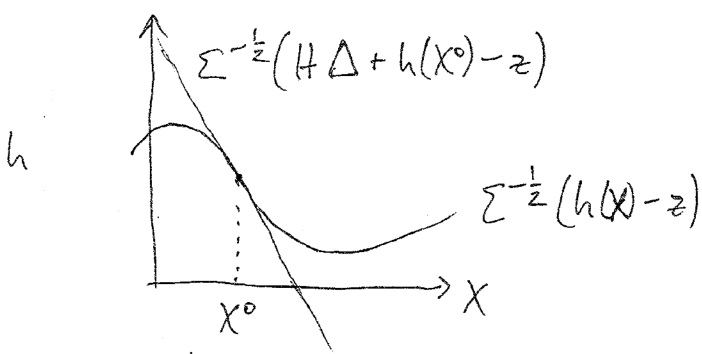
$$g(X) = g(X^0 + \Delta) \approx \sum_i \| H_i \Delta + h(X^0) - z_i \|^2_{\Sigma_i}$$

$$= \sum_i \| A_i \Delta - b_i \|^2$$

$$= \| A \Delta - b \|^2 = L(\Delta)$$

$\Delta^* = \text{argmin} \|A\Delta - b\|^2$

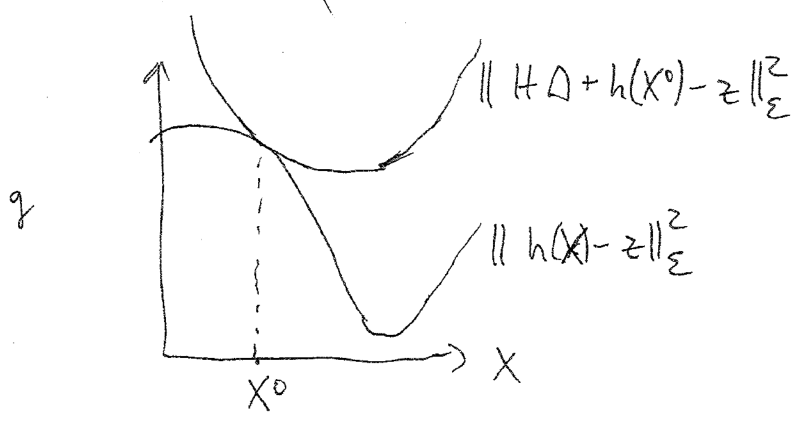
$X^{n+1} = X^n + \Delta^*$



1D: as drawn

$\mathbb{R}^n/\mathbb{R}^m$, m different hyperplanes (tangents) in $\mathbb{R}^n$

$n=2$: plane



1D: as drawn

$\mathbb{R}^n/\mathbb{R}^m$: m different (quadratic) functions in $\mathbb{R}^n$

$n=2$: (quadratic function) elliptic paraboloid

side note: units and role of cov. matrix

$z = \begin{pmatrix} \text{bearing} \\ \text{range} \end{pmatrix}$	$\begin{matrix} ° \\ m \end{matrix}$	$\Sigma^{-\frac{1}{2}}$	$\begin{matrix} \frac{1}{°} \\ \frac{1}{m} \end{matrix}$