Linear Gaussian Estimation

assumption: discrete, linear, time varying

- motion model:

$$x_k = A_{k-1} x_{k-1} + v_k + w_k$$

(random; input; process noise; random; deterministic; random)

$$y_k = C_k x_k + n_k$$

$$x_0 \in \mathbb{R}^N \sim \mathcal{N}(\check{x}_0, \check{P}_0)$$
$$w_k \in \mathbb{R}^N \sim \mathcal{N}(0, Q_k)$$
$$n_k \in \mathbb{R}^M \sim \mathcal{N}(0, R_k)$$

- batch linear-Gaussian estimation problem
  - Bayesian
  - Maximum a Posteriori

Maximum a Posteriori

- $\hat{x} = \arg\max_x P(x \mid v, y)$

$x = x_{0:K} = (x_0, x_1, \ldots, x_K)$

$v = (\check{x}_0, v_{1:K}) = (\check{x}_0, v_1, \ldots, v_K)$ for prior

$y = y_{0:K} = (y_0, \ldots, y_K)$ for posterior

- Bayes' rule:

$$\hat{x} = \arg\max_x P(x \mid v, y)$$
$$= \arg\max_x \frac{P(y \mid x, v) P(x \mid v)}{P(y \mid v)}$$

(v does not affect y; $P(y \mid v)$ does not affect x)

$$= \arg\max_x P(y \mid x) P(x \mid v)$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$
$$\therefore P(A \cap B) = P(B \mid A) P(A)$$
$$\therefore P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

$$P(A \mid BC) = \frac{P(A \cap B \cap C)}{P(B \cap C)}$$
$$P(C \mid AB) = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$
$$\therefore P(A \cap B \cap C) = P(A \cap B) P(C \mid A \cap B) = P(A \mid B) P(B) P(C \mid A \cap B)$$
$$\therefore P(A \mid BC) = \frac{P(A \mid B) P(B) P(C \mid A \cap B)}{P(B) P(C \mid B)} = \frac{P(A \mid B) P(C \mid A \cap B)}{P(C \mid B)}$$

- assume each w_k, n_k are NOT correlated:

$$P(y \mid x) = \prod_{k=0}^{K} P(y_k \mid x_k)$$

- Bayes' Rule:

$$P(x \mid v) = P(x_0 \mid \check{x}_0) \prod_{k=0}^{K} P(x_k \mid x_{k-1}, v_k)$$

$$\therefore P(x_0 \mid \check{x}_0) = \frac{1}{\sqrt{(2\pi)^N \det \check{P}_0}} \times \exp\left(-\frac{1}{2}(x_0 - \check{x}_0)^T \check{P}_0^{-1} (x_0 - \check{x}_0)\right)$$

$$\therefore P(x_k \mid x_{k-1}, v_k) = \frac{1}{\sqrt{(2\pi)^N \det Q_k}} \times \exp\left(-\frac{1}{2}(x_k - (A_{k-1}x_{k-1} + v_k))^T Q_k^{-1} (x_k - (A_{k-1}x_{k-1} + v_k))\right)$$

$$\therefore P(y_k \mid x_k) = \frac{1}{\sqrt{(2\pi)^M \det R_k}} \times \exp\left(-\frac{1}{2}(y_k - C_k x_k)^T R_k^{-1} (y_k - C_k x_k)\right)$$

- logarithm

$$\ln(P(y \mid x) P(x \mid v)) = \ln P(x_0 \mid \check{x}_0) + \sum_{k=1}^{K} \ln P(x_k \mid x_{k-1}, v_k) + \sum_{k=1}^{K} \ln P(y_k \mid x_k)$$

where

$$\ln P(x_0 \mid \check{x}_0) = -\frac{1}{2}(x_0 - \check{x}_0)^T \check{P}_0^{-1}(x_0 - \check{x}_0) + -\frac{1}{2}\ln((2\pi)^N \det \check{P}_0)$$

$$\ln P(x_k \mid x_{k-1}, v_k) = -\frac{1}{2}(x_k - A_{k-1}x_{k-1} - v_k)^T Q_k^{-1}(x_k - A_{k-1}x_{k-1} - v_k) + -\frac{1}{2}\ln((2\pi)^N \det Q_k)$$

$$\ln P(y_k \mid x_k) = -\frac{1}{2}(y_k - C_k x_k)^T R_k^{-1}(y_k - C_k x_k) + -\frac{1}{2}\ln((2\pi)^M \det R_k)$$

- J, cost

$$J_{v,k}(x) = \begin{cases} \frac{1}{2}(x_0 - \check{x}_0)^T \check{P}_0^{-1}(x_0 - \check{x}_0) \\ \frac{1}{2}(x_k - A_{k-1}x_{k-1} - v_k)^T Q_k^{-1}(x_k - A_{k-1}x_{k-1} - v_k) \end{cases}$$

$$J_{y,k}(x) = \frac{1}{2}(y_k - C_k x_k)^T R_k^{-1}(y_k - C_k x_k)$$

Mahalanobis distance

$$J(x) = \sum_{k=0}^{K}(J_{v,k}(x) + J_{y,k}(x))$$

$$\therefore x = \arg\min_x J(x)$$

$k=$

x_1 x_{10}

- from previous

$$J_{v,k}(x) = \begin{cases} \frac{1}{2}(x_0 - \check{x}_0)^T \check{P}_0^{-1}(x_0 - \check{x}_0) \\ \frac{1}{2}(x_k - A_{k-1}x_{k-1} - v_k)^T Q_k^{-1}(x_k - A_{k-1}x_{k-1} - v_k) \end{cases}$$

$$J_{y,k}(x) = \frac{1}{2}(y_k - C_k x_k)^T R_k^{-1}(y_k - C_k x_k)$$

let

$$z = \begin{bmatrix} \check{x}_0 \\ v_1 \\ \vdots \\ v_K \\ y_0 \\ y_1 \\ \vdots \\ y_K \end{bmatrix}, \quad x = \begin{bmatrix} x_0 \\ \vdots \\ x_K \end{bmatrix}$$

$z - Hx$

&

$$H = \begin{bmatrix} 1 & & & \\ -A_0 & 1 & & \\ & \ddots & \ddots & \\ & & -A_{K-1} & 1 \\ \hline C_0 & & & \\ & C_1 & & \\ & & \ddots & \\ & & & C_K \end{bmatrix}$$

(faint: $-A_1 x_0 + x_1 = B_u$, $x_1 =$)

$$W = \begin{bmatrix} \check{P}_0 & & & \\ & Q_1 & & \\ & & \ddots & \\ & & & Q_K \\ \hline & & & & R_0 \\ & & & & & R_1 \\ & & & & & & \ddots \\ & & & & & & & R_K \end{bmatrix}$$

$$\therefore J(x) = \frac{1}{2}(z - Hx)^T W^{-1}(z - Hx)$$

$$P(z \mid x) = \eta \exp\left(-\frac{1}{2}(z - Hx)^T W^{-1}(z - Hx)\right)$$

(η: normalization constant)

$$\therefore \left.\frac{\partial J(x)}{\partial x^T}\right|_{\hat{x}} = -H^T W^{-1}(z - H\hat{x}) = 0$$

$$\Rightarrow (H^T W^{-1} H)\hat{x} = H^T W^{-1} z$$

Bayesian Inference

compute full Bayesian Posterior

$P(x \mid v, y)$

- $x_k = A_{k-1}x_{k-1} + v_k + w_k$

lift → $x = A(v + w)$

where

$$A = \begin{bmatrix} 1 & & & & \\ A_0 & 1 & & & \\ A_1 A_0 & A_1 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ A_{K-2}\cdots A_0 & A_{K-2}\cdots A_1 & A_{K-2}\cdots A_2 & \cdots & 1 \\ A_{K-1}\cdots A_0 & A_{K-1}\cdots A_1 & A_{K-1}\cdots A_2 & \cdots\ A_{K-1} & 1 \end{bmatrix}$$

- $\check{x} = E[x] = E[A(v + w)] = E[Av + Aw] = Av$
- $\check{P} = E[(x - E[x])(x - E[x])^T] = AQA^T$

$Q = E[ww^T] = \mathrm{diag}(\check{P}_0, Q_1, \ldots, Q_K)$

- prior $\therefore \Rightarrow$

$$P(x \mid v) = \mathcal{N}(\check{x}, \check{P}) = \mathcal{N}(Av, AQA^T)$$

- $y_k = C_k x_k + n_k$

lift → $y = Cx + n$

where

$$C = \mathrm{diag}(C_0, C_1, \ldots, C_K)$$

- joint density of prior lifted state & measurements is now:

$$P(x, y \mid v) = \mathcal{N}\left(\begin{bmatrix} \check{x} \\ C\check{x} \end{bmatrix}, \begin{bmatrix} \check{P} & \check{P}C^T \\ C\check{P} & C\check{P}C^T + R \end{bmatrix}\right)$$

$$R = E[nn^T] = \mathrm{diag}(R_0, R_1, \ldots, R_K)$$

- $P(x, y \mid v) = P(x \mid y, v) P(y \mid v)$ — we only care about this ($P(x \mid y, v)$)

- $P(x, y) = \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right)$
- $P(x, y) = P(x \mid y) P(y)$

w/ schur complement

$$\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} = \begin{bmatrix} I & \Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} & 0 \\ 0 & \Sigma_{yy} \end{bmatrix}\begin{bmatrix} I & 0 \\ \Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix}\begin{bmatrix} (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix}\begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix}$$

- PDF of $P(x, y)$

recall:

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

$$\Rightarrow P(x, y) = (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(Z)$$

$$Z = \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)^T \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}^{-1} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)$$

$$= \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)^T \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix}\begin{bmatrix} (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix}\begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)$$

$$= [x - \mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)]^T (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} \times [x - \mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)] + (y - \mu_y)^T \Sigma_{yy}^{-1}(y - \mu_y)$$

sum of two quadratic terms

$$P(x, y) = P(x \mid y) P(y)$$
$$P(x \mid y) = \mathcal{N}(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y),\ \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$$
$$P(y) = \mathcal{N}(\mu_y, \Sigma_{yy})$$

on $(H^T W^{-1} H)\hat{x} = H^T W^{-1} z$

- unique solution:

$$\hat{x} = (H^T W^{-1} H)^{-1} H^T W^{-1} z$$

(need to be invertible)

w/ $\hat{x} \in \mathbb{R}^{K+1}$

$$\mathrm{rank}(H^T W^{-1} H) = K+1$$
$$\Rightarrow \mathrm{rank}(H^T H) = \mathrm{rank}(H^T)$$

- $\mathrm{rank}(H^T)$: we hv $\check{x}_0$ / we don't hv $\check{x}_0$

✓ - we hv $\check{x}_0$

$$\mathrm{rank}\ H^T = \mathrm{rank}\begin{bmatrix} 1 & -A_0^T & & & C_0^T & & & \\ & 1 & -A_1^T & & & C_1^T & & \\ & & \ddots & -A_{K-1}^T & & & C_2^T & \\ & & & 1 & & & & C_K^T \end{bmatrix}$$

$\Rightarrow$ full rank. $\therefore$ $Q_k \succ 0$, $R_k \succ 0 \Rightarrow$ hv sol.

✓ - we don't hv $\check{x}_0$

$$\mathrm{rank}\ H^T = \mathrm{rank}\begin{bmatrix} -A_0^T & & & C_0^T & & \\ \hline 1 & -A_1^T & & C_1^T & & \\ & 1 & \ddots\ -A_{K-1}^T & & C_2^T & \\ & & 1 & & & C_K^T \end{bmatrix}$$

$$\Rightarrow \mathrm{rank}\begin{bmatrix} 1 & -A_1^T & & C_1^T & & \\ & 1 & \ddots\ -A_{K-1}^T & & C_2^T & \\ & & 1 & & & C_K^T \\ \hline -A_0^T & & & C_0^T & & \end{bmatrix}$$

$$\Rightarrow \mathrm{rank}\begin{bmatrix} 1 & -A_1^T & & C_1^T & & \\ & 1 & \ddots\ -A_{K-1}^T & & C_2^T & \\ & & 1 & & & C_K^T \\ \hline & & & C_0^T & A_0^T C_1^T & A_0^T A_1^T C_2^T \cdots A_0^T \cdots A_{K-1}^T C_K^T \end{bmatrix}$$

→ make sure this CUR

- $\mathrm{rank}[C_0^T\ A_0^T C_1^T\ A_0^T A_1^T C_2^T \cdots A_0^T \cdots A_{K-1}^T C_K^T] = N$

from Cayley-Hamilton theorem: any power of $A \in \mathbb{R}^{N \times N}$, power $k \geq N$

$$\Rightarrow (A^T)^{k-1} = a_0 I^T + a_1 A^T + a_2 A^T A^T + \cdots + a_{N-1}(A^T)^{N-1}$$

assume $A_0 = A_1 = A_2 = \cdots = A_{K-1} = A$ & $k \gg N$

we first hv

$$(A^T)^{K-1} C^T = a_0 I C^T + a_1 A^T C^T + a_2 A^T A^T C^T + \cdots + a_{N-1}(A^T)^{N-1} C^T$$

$$\mathrm{rank}[C_0^T\ A_0^T C_1^T\ A_0^T A_1^T C_2^T \cdots A_0^T \cdots A_{K-1}^T C_K^T]$$
$$\Rightarrow \mathrm{rank}[C^T\ A^T C^T\ A^T A^T C^T \cdots (A^T)^K C^T]$$
$$= \mathrm{rank}[C^T\ A^T C^T \cdots (A^T)^{N-1} C^T]$$

- $\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{(N-1)} \end{bmatrix}$

- $\mathrm{rank}\ \mathcal{O} = N$ (observability)

$\therefore$ $Q_k \succ 0$, $R_k \succ 0$, $\mathrm{rank}\ \mathcal{O} = N$ $\Rightarrow$ hv sol.

$$P(x, y \mid v) = P(x \mid y, v) P(y \mid v)$$

$$\therefore P(x \mid y, v)$$

recall $P(x \mid y) = \mathcal{N}(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y),\ \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$

&

$$P(x, y \mid v) = \mathcal{N}\left(\begin{bmatrix} \check{x} \\ C\check{x} \end{bmatrix}, \begin{bmatrix} \check{P} & \check{P}C^T \\ C\check{P} & C\check{P}C^T + R \end{bmatrix}\right)$$

$$= \mathcal{N}\left(\check{x} + \check{P}C^T(C\check{P}C^T + R)^{-1}(y - C\check{x}),\ \check{P} - \check{P}C^T(C\check{P}C^T + R)^{-1}(C\check{P})\right)$$

thru Sherman-Morrison-Woodbury Identity

$$= \mathcal{N}\left((\check{P}^{-1} - C^T R^{-1} C)^{-1}(\check{P}^{-1}\check{x} + C^T R^{-1} y),\ (\check{P}^{-1} + C^T R^{-1} C)^{-1}\right)$$

i.e.,

$$\hat{x} = (\check{P}^{-1} - C^T R^{-1} C)^{-1}(\check{P}^{-1}\check{x} + C^T R^{-1} y)$$
$$\Rightarrow (\check{P}^{-1} - C^T R^{-1} C)\hat{x} = \check{P}^{-1}\check{x} + C^T R^{-1} y$$
$$\Rightarrow ((AQA^T)^{-1} - C^T R^{-1} C)\hat{x} = (AQA^T)^{-1} Av + C^T R^{-1} y$$
$$\Rightarrow (A^{-T} Q^{-1} A^{-1} - C^T R^{-1} C)\hat{x} = A^{-T} Q^{-1} A^{-1} Av + C^T R^{-1} y$$
$$\Rightarrow (A^{-T} Q^{-1} A^{-1} - C^T R^{-1} C)\hat{x} = A^{-T} Q^{-1} v + C^T R^{-1} y$$

$$A^{-1} = \begin{bmatrix} 1 & & & & \\ -A_0 & 1 & & & \\ & -A_1 & 1 & & \\ & & -A_2 & \ddots & \\ & & & & 1 \\ & & & -A_{K-1} & 1 \end{bmatrix}, \quad z = \begin{bmatrix} v \\ y \end{bmatrix}, \quad H = \begin{bmatrix} A^{-1} \\ C \end{bmatrix}, \quad W = \begin{bmatrix} Q & \\ & R \end{bmatrix}$$

$$\Rightarrow (H^T W^{-1} H)\hat{x} = H^T W^{-1} z$$

---

solving $(H^T W^{-1} H)\hat{x} = H^T W^{-1} z$

- Cholesky Decomposition

$H^T W^{-1} H = LL^T$ — Cholesky Decomp.

- first decompose this
- $Ax = b$, $LL^T x = b$
  $\Rightarrow Ly = b$
- $L^T x = y$
- solve x!

$$\therefore (H^T W^{-1} H)\hat{x} = H^T W^{-1} z$$
$$\Rightarrow LL^T\hat{x} = H^T W^{-1} z$$
$$\Rightarrow Ld = H^T W^{-1} z$$
$$\Rightarrow L^T\hat{x} = d$$

- let LL^T

$$L = \begin{bmatrix} L_0 & & & & & \\ L_{10} & L_1 & & & & \\ & L_{21} & L_2 & & & \\ & & \ddots & \ddots & & \\ & & & L_{K-1,K-2} & L_{K-1} & \\ & & & & L_{K,K-1} & L_K \end{bmatrix}$$

$$L_0 L_0^T = \underbrace{\check{P}_0^{-1} + C_0^T R_0^{-1} C_0}_{I_0} + A_0^T Q_1^{-1} A_0$$

$$L_{10} L_0^T = -Q_1^{-1} A_0$$

$$L_1 L_1^T = -L_{10} L_{10}^T + Q_1^{-1} + C_1^T R_1^{-1} C_1 + A_1^T Q_2^{-1} A_1$$

$$L_{21} L_1^T = -Q_2^{-1} A_1$$

$$\vdots$$

$$L_{K-1} L_{K-1}^T = \underbrace{-L_{K-1,K-2} L_{K-1,K-2}^T + Q_{K-1}^{-1} + C_{K-1}^T R_{K-1}^{-1} C_{K-1}}_{I_{K-1}} + A_{K-1}^T Q_K^{-1} A_{K-1}$$

$$L_{K,K-1} L_{K-1}^T = -Q_K^{-1} A_{K-1}$$

$$L_K L_K^T = \underbrace{-L_{K,K-1} L_{K,K-1}^T + Q_K^{-1} + C_K^T R_K^{-1} C_K}_{I_K}$$

$I_0, I_1, \ldots, I_{K-1}, I_K \Rightarrow$ Information Matrix

Basically solve L_0 → solve L_{10} → solve L_1 … $O(N(K+1))$

forward

- $Ld = H^T W^{-1} z$, get $d = [d_0\ d_1 \ldots d_K]^T$

$$L_0 d_0 = \check{P}_0^{-1}\check{x}_0 + C_0^T R_0^{-1} y_0 - A_0^T Q_1^{-1} v_1$$

$$L_1 d_1 = \underbrace{-L_{10} d_0 + Q_1^{-1} v_1 + C_1^T R_1^{-1} y_1}_{q_1} - A_1^T Q_2^{-1} v_2$$

$$\vdots$$

$$L_{K-1} d_{K-1} = \underbrace{-L_{K-1,K-2} d_{K-2} + Q_{K-1}^{-1} v_{K-1} + C_{K-1}^T R_{K-1}^{-1} y_{K-1}}_{q_{K-1}} - A_{K-1}^T Q_K^{-1} v_K$$

$$L_K d_K = \underbrace{-L_{K,K-1} d_{K-1} + Q_K^{-1} v_K + C_K^T R_K^{-1} y_K}_{q_K}$$

backward

- $L^T\hat{x} = d$

$$L_K^T \hat{x}_K = d_K$$
$$L_{K-1}^T \hat{x}_{K-1} = -L_{K,K-1}^T \hat{x}_K + d_{K-1}$$
$$\vdots$$
$$L_1^T \hat{x}_1 = -L_{21}^T \hat{x}_2 + d_1$$
$$L_0^T \hat{x}_0 = -L_{10}^T \hat{x}_1 + d_0$$

Similarly: solve $\hat{x}_K$ → solve $\hat{x}_{K-1}$ → solve $\hat{x}_{K-2}$ … $O(N(K+1))$

Similarly, solve d_0 → solve d_1 → solve d_2 … $O(N(K+1))$

Forward

$$L_{K-1} L_{K-1}^T = I_{K-1} + A_{K-1}^T Q_K^{-1} A_{K-1}$$
$$L_{K-1} d_{K-1} = q_{K-1} - A_{K-1}^T Q_K^{-1} v_K$$
$$L_{K,K-1} L_{K-1}^T = -Q_K^{-1} A_{K-1}$$
$$I_K = -L_{K,K-1} L_{K,K-1}^T + Q_K^{-1} + C_K^T R_K^{-1} C_K$$
$$q_K = -L_{K,K-1} d_{K-1} + Q_K^{-1} v_K + C_K^T R_K^{-1} y_K$$

Backward

$$L_{K-1}^T \hat{x}_{K-1} = -L_{K,K-1}^T \hat{x}_K + d_{K-1}$$

→ Kalman filter

△ solving $(H^T W^{-1} H)\hat{x} = H^T W^{-1} z$

• Rauch–Tung–Striebel

from above:

$L_{k+1}\Lambda_{k+1}^T = J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1}$

$L_{k+1}\Lambda_{k+1}^T = -Q_{k+1}^{-1} A_{k+1}$

$J_k = -L_{k+1}\Lambda_{k+1}^T + Q_k^{-1} + C_k^T R_k^{-1} C_k$

$\therefore L_{k,k+1} = -Q_{k+1}^{-1} A_{k+1} L^{-T}_{k+1}$

$L_{k,k+1}\Lambda_{k+1}^T = Q_{k+1}^{-1} A_{k+1} L_{k+1}^{-1} L_{k+1}^{-T} A_{k+1}^T Q_{k+1}^{-T}$

$= Q_{k+1}^{-1} A_{k+1} (L_{k+1}\Lambda_{k+1}^T)^{-1} A_{k+1}^T Q_{k+1}^{-T}$

$\therefore J_k = -L_{k,k+1} L_{k,k+1}^T + Q_k^{-1} + C_k^T R_k^{-1} C_k$

$= Q_{k+1}^{-1} - Q_{k+1}^{-1} A_{k+1} (L_{k+1}\Lambda_{k+1}^T)^{-1} A_{k+1}^T Q_{k+1}^{-T} + C_k^T R_k^{-1} C_k$

$= (A_{k+1} L_{k+1}^{-1} A_{k+1} + Q_{k+1})^{-1} + C_k^T R_k^{-1} C_k$

(from Sherman–Morrison–Woodbury)

$J_k = (A_{k+1} L_{k+1}^{-1} A_{k+1} + Q_{k+1})^{-1} + C_k^T R_k^{-1} C_k$

$\check{P}_{k+1} = J_k^{-1}$

$\check{P}_{k,k+1} = A_{k+1} L_{k+1}^{-1} A_{k+1}^T Q_{k+1}$

$\check{P}_{k+1}^{-1} = \check{P}_{k+1}^{-1} + C_k^T R_k^{-1} C_k$

△ inducing KF

let us define

$K_k = \hat{P}_{k+1}^- C_k^T R_k^{-1}$

$= (\hat{P}_{k+1}^- + C_k^T R_k^{-1} C_k)^{-1} C_k^T R_k^{-1}$ (SMW)

$= \hat{P}_{k+1}^- C_k^T (C_k \hat{P}_{k+1}^- C_k^T + R_k)^{-1}$

which is the Kalman gain that we are familiar of

$\hat{P}_{k+1}^{-1} = \check{P}_{k+1}^{-1} - C_k^T R_k^{-1} C_k$

$= \check{P}_{k+1}^{-1} [I - \check{P}_{k+1}^- C_k^T R_k^{-1} C_k)$

$= \check{P}_{k+1}^{-1} (I - K_k C_k)$

$\Rightarrow \hat{P}_{k+1} = (I - K_k C_k) \check{P}_{k+1}$

• now, from above,

$\begin{cases} L_{k,k+1}\Lambda_{k+1}^T = -Q_{k+1}^{-1} A_{k+1} \\ L_{k+1} d_{k+1} = g_{k+1} - A_{k+1}^T Q_{k+1}^{-1} v_k \\ L_{k+1} L_{k+1}^T = J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1} \end{cases}$

$L_{k,k+1} = (-Q_{k+1}^{-1} A_{k+1}) L^{-T}_{k+1}$

$d_{k+1} = L_{k+1}^{-1} (g_{k+1} - A_{k+1}^T Q_{k+1}^{-1} v_k)$

$\Rightarrow L_{k,k+1} d_{k+1} = (-Q_{k+1}^{-1} A_{k+1}) L_{k+1}^{-1} L_{k+1}^{-T} \times (g_{k+1} - A_{k+1}^T Q_{k+1}^{-1} v_k)$

$\Rightarrow L_{k,k+1} d_{k+1} = (-Q_{k+1}^{-1} A_{k+1})(L_{k+1}\Lambda_{k+1}^T)^{-1} (L_{k+1} d_{k+1} = -Q_{k+1}^{-1} v_k + Q_{k+1}^{-1} A_{k+1} L_{k+1}^{-1} d_{k+1})$ (1)

$\because g_k = -L_{k,k+1} d_{k+1} + Q_k^{-1} v_k + C_k^T R_k^{-1} y_k$ (2)

combining (1) & (2)

$g_k = Q_{k+1}^{-1} A_{k+1} (J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1})^{-1} g_{k+1} + (Q_{k+1}^{-1} - Q_{k+1}^{-1} A_{k+1} (J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1})^{-1} A_{k+1}^T Q_{k+1}^{-1}) v_k + C_k^T R_k^{-1} y_k$ (SMW)

$= (A_{k+1} J_{k+1}^{-1} A_{k+1}^T + Q_{k+1})^{-1} A_{k+1} J_{k+1}^{-1} g_{k+1} + (A_{k+1} J_{k+1}^{-1} A_{k+1}^T + Q_{k+1})^{-1} v_k + C_k^T R_k^{-1} y_k$

$= (A_{k+1} \check{P}_{k+1} A_{k+1}^T + Q_{k+1})^{-1} (A_{k+1} J_{k+1}^{-1} g_{k+1} + v_k) + C_k^T R_k^{-1} y_k$

$= (\hat{P}_{k,k+1}^{-1})^{-1} (A_{k+1} \hat{P}_{k+1,k} \hat{P}_{k+1,k}^{-1} \hat{x}_{k+1,k} + v_k) + C_k^T R_k^{-1} y_k$ (let $g_0 = \hat{P}_{1,0}^{-1} \hat{x}_{1,0}$)

$\Rightarrow \hat{P}_{k+1,k}^{-1} \hat{x}_{k+1,k} = \hat{P}_{k+1,k}^{-1} (A_{k+1} \hat{x}_{k+1,k} + v_k) + C_k^T R_k^{-1} y_k$

$\Rightarrow \hat{P}_{k+1}^{-1} \hat{x}_{k+1,k} = \check{P}_{k+1}^{-1} \check{x}_{k+1,k} + C_k^T R_k^{-1} y_k$

$\Rightarrow \hat{x}_{k+1,k} = \hat{P}_{k+1,k} \check{P}_{k+1}^{-1} \check{x}_{k+1,k} + \hat{P}_{k+1} C_k^T R_k^{-1} y_k$

$= (I - K_k C_k) \hat{x}_{k+1,k} + K_k y_k$

$\Rightarrow \hat{x}_{k+1} = \check{x}_{k+1,k} + K_k (y_k - C_k \check{x}_{k+1})$

end of "forward"

• backward

from above

$L_{k+1}^T \hat{x}_{k+1} = -L_{k,k+1}^T \hat{x}_k + d_{k+1}$

$\Rightarrow \hat{x}_{k+1} = (L_{k+1}^T)^{-1} (-L_{k,k+1}^T \hat{x}_k + d_{k+1})$

$= L_{k+1}^{-T} L_{k+1}^{-1} (-L_{k,k+1}^T \hat{x}_k + d_{k+1})$

$= (L_{k+1} L_{k+1}^T)^{-1} L_{k+1} (-L_{k,k+1}^T \hat{x}_k + d_{k+1})$

$\Rightarrow$ from

$L_{k+1} L_{k+1}^T = J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1}$

$L_{k+1} d_{k+1} = g_{k+1} - A_{k+1}^T Q_{k+1}^{-1} v_k$

$L_{k+1} L_{k+1}^T = -Q_{k+1}^{-1} A_{k+1}$

$= (L_{k+1} L_{k+1}^T)^{-1} (-L_{k+1} L_{k,k+1}^T \hat{x}_k + L_{k+1} d_{k+1})$

$= (J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1})^{-1} \times (-(L_{k,k+1} L_{k+1}^T) \hat{x}_k + g_{k+1} - A_{k+1}^T Q_{k+1}^{-1} v_k)$

$= (J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1})^{-1} \times (A_{k+1}^T Q_{k+1}^{-1} \hat{x}_k + g_{k+1} - A_{k+1}^T Q_{k+1}^{-1} v_k)$

$= (J_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1})^{-1} \times (A_{k+1}^T Q_{k+1}^{-1} (\hat{x}_k - v_k) + g_{k+1})$

$= (\check{P}_{k+1} + A_{k+1}^T Q_{k+1}^{-1} A_{k+1})^{-1} A_{k+1}^T Q_{k+1}^{-1} (\hat{x}_k - v_k)$ (SMW)

$+ (I_{k+1} + A_{k+1}^T (Q_{k+1}^{-1} A_{k+1}))^{-1} g_{k+1}$

$= \hat{P}_{k+1,k} A_{k+1}^T \check{P}_{k+1}^{-1} (\hat{x}_k - v_k)$

$+ (\hat{P}_{k+1} - \hat{P}_{k+1,k} A_{k+1}^T \check{P}_{k+1}^{-1} A_{k+1} \hat{P}_{k+1}) \check{P}_{k+1}^{-1} \check{x}_{k+1}$

$= \hat{P}_{k+1,k} A_{k+1}^T \check{P}_{k+1}^{-1} (\hat{x}_k - v_k)$

$+ \check{x}_{k+1,k} - \hat{P}_{k+1,k} A_{k+1}^T \check{P}_{k+1}^{-1} A_{k+1} \check{x}_{k+1,k}$

$= \hat{P}_{k+1,k} A_{k+1}^T \check{P}_{k+1}^{-1} (\hat{x}_k - \check{x}_{k+1,k} - v_k) + \check{x}_{k+1}$

$= \hat{P}_{k+1,k} A_{k+1}^T \check{P}_{k+1}^{-1} (\hat{x}_k - \check{x}_{k+1,k}) + \check{x}_{k+1}$ (with $\hat{x}_{k+1,k}^0$)

Non-Linear Estimation

Recall

J, cost

$J_{0,k}(x) = \begin{cases} \frac{1}{2}(x - \check{x}_0)^T \check{P}_0^{-1} (x - \check{x}_0) \\ \frac{1}{2}(x_k - A_k x_{k-1} - v_k)^T Q_k^{-1} (x_k - A_k x_{k-1} - v_k) \end{cases}$

$J_{0,k}(x) = \frac{1}{2}(y_k - C_k x_k)^T R_k^{-1} (y_k - C_k x_k)$

Mahalanobis distance

$J(x) = \sum_{k=0}^{K} (J_{0,k}(x) + J_{y,k}(x))$

$\therefore x = \underset{x}{\operatorname{argmin}}\, J(x)$

• here the system, i.e.,
- the prediction model
- the measurement model
is no longer linear

$\begin{cases} e_{v,k}(x) = \begin{cases} \check{x}_0 - x_0 & , k=0 \\ f(x_{k-1}, v_k, 0) - x_k & , k=1\ldots K \end{cases} \\ e_{y,k}(x) = y_k - g(x_k, 0) \quad , k=0\ldots K \end{cases}$

$\begin{cases} J_{v,k}(x) = \frac{1}{2} e_{v,k}(x)^T W_{v,k}^{-1} e_{v,k}(x) \\ J_{y,k}(x) = \frac{1}{2} e_{y,k}(x)^T W_{y,k}^{-1} e_{y,k}(x) \end{cases}$

$\Rightarrow J(x) = \sum_{k=0}^{K} (J_{v,k}(x) + J_{y,k}(x))$

$\Rightarrow$ • let $e(x) = \begin{bmatrix} e_v(x) \\ e_y(x) \end{bmatrix}$

$e_v(x) = \begin{bmatrix} e_{v,0}(x) \\ \vdots \\ e_{v,K}(x) \end{bmatrix}$

$e_y(x) = \begin{bmatrix} e_{y,0}(x) \\ \vdots \\ e_{y,K}(x) \end{bmatrix}$

• let $W = \operatorname{diag}(W_v, W_y)$

$W_v = \operatorname{diag}(W_{v,0} \ldots W_{v,K})$

$W_y = \operatorname{diag}(W_{y,0} \ldots W_{y,K})$

• $J(x) = \frac{1}{2} e(x)^T W^{-1} e(x)$

$\Rightarrow$ let $u(x) = L e(x)$

$L^T L = W^{-1}$

$J(x) = \frac{1}{2} e(x)^T L^T L e(x)$

$= \frac{1}{2} [Le(x)]^T Le(x)$

$= \frac{1}{2} u(x)^T u(x)$

$\Rightarrow \hat{x} = \underset{x}{\operatorname{argmin}}\, J(x)$

Solving $\hat{x} = \underset{x}{\operatorname{argmin}}\, J(x)$

• Newton's Method

@ update step

$J(x_{op} + \delta x)$

$\approx J(x_{op}) + \left(\frac{\partial J(x)}{\partial x}\Big|_{x_{op}}\right) \delta x + \frac{1}{2} \delta x^T \left(\frac{\partial J(x)}{\partial x^T}\Big|_{x_{op}}\right) \delta x$

$\frac{\partial J(x_{op} + \delta x)}{\partial \delta x}$

$= \left(\frac{\partial J(x)}{\partial x}\Big|_{x_{op}}\right) + \delta x^T \left(\frac{\partial J(x)}{\partial x^2}\Big|_{x_{op}}\right) = 0$

$\Rightarrow \left(\frac{\partial^2 J(x)}{\partial x^2}\Big|_{x_{op}}\right) \delta x^* = -\left(\frac{\partial J(x)}{\partial x}\Big|_{x_{op}}\right)^T$ (update step)

then

$x_{op} \leftarrow x_{op} + \delta x^*$

• Gauss–Newton Method

from above

$J(x) = \frac{1}{2} u(x)^T u(x)$

Jacobian

$\frac{\partial J(x)}{\partial x}\Big|_{x_{op}} = u(x_{op})^T \left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right)$

Hessian

$\frac{\partial^2 J(x)}{\partial x^2}\Big|_{x_{op}}$

$= \left(\frac{\partial u(x)}{\partial x}\right)^T \left(\frac{\partial u(x)}{\partial x}\right) + \sum_{i=0}^{m} u_i(x) \left(\frac{\partial^2 u(x)}{\partial x^2}\right)$

$\approx \left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right)^T \left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right)$ neglect this

$\therefore \left(\frac{\partial J(x)}{\partial x^2}\Big|_{x_{op}}\right) \delta x^* = -\left(\frac{\partial J(x)}{\partial x}\Big|_{x_{op}}\right)^T$

$\Rightarrow \left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right)^T \left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right) \delta x^* = -\left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right)^T u(x_{op})$

△ GN method

recall:

• $\left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right)^T \left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right) \delta x^* = -\left(\frac{\partial u(x)}{\partial x}\Big|_{x_{op}}\right)^T u(x_{op})$

• $J(x) = \frac{1}{2} e(x)^T W^{-1} e(x)$

$= \frac{1}{2} e(x)^T L^T L e(x)$

$= \frac{1}{2} [Le(x)]^T Le(x)$

$= \frac{1}{2} u(x)^T u(x)$

• $\begin{cases} e_{v,k}(x) = \begin{cases} \check{x}_0 - x_0 & , k=0 \\ f(x_{k-1}, v_k, 0) - x_k & , k=1\ldots K \end{cases} \\ e_{y,k}(x) = y_k - g(x_k, 0) \quad , k=0\ldots K \end{cases}$

NOW:

• for any function $f(x)$, perturbed by δx @ a we can write Taylor expansion: (2nd order)

$f(a + \delta x)$

$= f(a) + \frac{\partial f(x)}{\partial x}\Big|_{x=a} \delta x + \frac{1}{2} \delta x^T \frac{\partial^2 f(x)}{\partial x^2}\Big|_{x=a} \delta x$

$\therefore$ w/ 1st order

$e_{v,k}(x_{op} + \delta x) \approx \begin{cases} e_{v,0}(x_{op}) + \frac{\partial e_{v,0}(x)}{\partial x}\Big|_{x_{op}} \delta x & , k=0 \\ e_{v,k}(x_{op}) + \frac{\partial e_{v,k}(x)}{\partial x}\Big|_{x_{op}} \delta x & , k=1\ldots K \end{cases}$

① $\Rightarrow \begin{cases} \check{x}_0 - x_{op,0} - I \cdot \delta x & , k=0 \\ f(x_{op,k-1}, v_k, 0) - x_{op,k} + F_{k-1} \delta x_{k-1} - \delta x_k & , k=1\ldots K \end{cases}$

$F_{k-1} = \frac{\partial f(x_{k-1}, v_k, w_k)}{\partial x_{k-1}}\Big|_{x_{op,k-1}, v_k, 0}$

$e_{y,k}(x_{op} + \delta x) \approx e_{y,k}(x_{op}) + \frac{\partial e_{y,k}(x)}{\partial x}\Big|_{x_{op,k}}$

② $\Rightarrow y_k - g(x_{op,k}, 0) - G_k \delta x_k$

$G_k = \frac{\partial g(x_k, n_k)}{\partial x_k}\Big|_{x_{op,k}, 0}$

THEN

$J(x) = \frac{1}{2} e(x)^T W^{-1} e(x)$

- stack all $e(x)$'s

$e(x_{op}) = \begin{bmatrix} e_{v,0}(x_{op}) \\ e_{v,1}(x_{op}) \\ \vdots \\ e_{v,K}(x_{op}) \\ e_{y,0}(x_{op}) \\ e_{y,1}(x_{op}) \\ \vdots \\ e_{y,K}(x_{op}) \end{bmatrix}$

- WHILE in iterative method, variable is δx

$\delta x = \begin{bmatrix} \delta x_0 \\ \delta x_1 \\ \vdots \\ \delta x_K \end{bmatrix}$

- ALSO, recall GN:

$J(x)^T J(x)\, \delta x = -J(x)^T \delta(x)$

HERE

$\because J(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$

"$f(x)$" $= e(x_{op}) = \begin{bmatrix} e_{v,0}(x_{op}) \\ e_{v,1}(x_{op}) \\ \vdots \\ e_{v,K}(x_{op}) \\ e_{y,0}(x_{op}) \\ e_{y,1}(x_{op}) \\ \vdots \\ e_{y,K}(x_{op}) \end{bmatrix}$

"x" $= \delta x = \begin{bmatrix} \delta x_0 \\ \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_K \end{bmatrix}$

"$J(x)$"

$= J$

$= \begin{bmatrix} \frac{\partial e_{v,0}(x_{op})}{\partial \delta x_0} & \frac{\partial e_{v,0}(x_{op})}{\partial \delta x_1} & \cdots \\ \frac{\partial e_{v,1}(x_{op})}{\partial \delta x_0} & \frac{\partial e_{v,1}(x_{op})}{\partial \delta x_1} & \cdots \\ \vdots & \frac{\partial e_{v,2}(x_{op})}{\partial \delta x_2} & \ddots \\ \hline \frac{\partial e_{y,0}(x_{op})}{\partial \delta x_0} & \frac{\partial e_{y,0}(x_{op})}{\partial \delta x_1} & \cdots \\ \frac{\partial e_{y,1}(x_{op})}{\partial \delta x_0} & \frac{\partial e_{y,1}(x_{op})}{\partial \delta x_1} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$

$= \begin{bmatrix} I & & & & \\ F_0 & I & & & \\ & F_1 & \ddots & & \\ & & F_{K-1} & I & \\ \hline -G_0 & & & I & \\ & -G_1 & & & \ddots \end{bmatrix}$

$(J^T W^{-1} J)\, \delta x = -J^T W^{-1} e(x_{op})$

remember $u(x) = Le(x)$

previously, I dropped "L"

@ iteration step i:

solve δx

THEN

$x_{op} \leftarrow x_{op} + \delta x$

★ solving the linear system @ i, has similar formulation w/ LGE

★ NLGE $\longleftrightarrow$ LGE

- iterative linearization
- solve the linear system locally @ i

- solve the linear system globally

★ $\delta x^k = \underset{\delta x}{\operatorname{argmin}} \frac{1}{2} \| e - J \delta x \|_2^2$

Marginalization – Sliding Window

recall the linear system

$H \delta x = b$

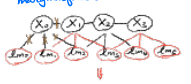
ALSO recall Schur complement

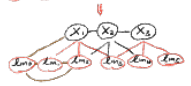
$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$

$S = C - B^T A^{-1} C$

the Schur complement of A in X

• The core idea of marginalization in SWF:





• $H \delta x = b$ (we want this one)

$\Rightarrow \begin{bmatrix} \Lambda_a & \Lambda_b \\ \Lambda_b^T & \Lambda_c \end{bmatrix} \begin{bmatrix} \delta x_a \\ \delta x_b \end{bmatrix} = \begin{bmatrix} g_a \\ g_b \end{bmatrix}$

$\Rightarrow \begin{bmatrix} \Lambda_a - \Lambda_b \Lambda_c^{-1} \Lambda_b^T & 0 \\ \Lambda_b^T & \Lambda_c \end{bmatrix} \begin{bmatrix} \delta x_a \\ \delta x_b \end{bmatrix} = \begin{bmatrix} g_a - \Lambda_b \Lambda_c^{-1} g_b \\ g_b \end{bmatrix}$

$\Rightarrow \begin{bmatrix} \Lambda_a - \Lambda_b \Lambda_c^{-1} \Lambda_b^T & 0 \\ \Lambda_b^T & \Lambda_c \end{bmatrix} \begin{bmatrix} \delta x_a \\ \delta x_b \end{bmatrix} = \begin{bmatrix} g_a - \Lambda_b \Lambda_c^{-1} g_b \\ g_b \end{bmatrix}$

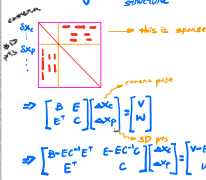
$\Rightarrow (\Lambda_a - \Lambda_b \Lambda_c^{-1} \Lambda_b^T) \delta x_a = g_a - \Lambda_b \Lambda_c^{-1} g_b$

↳ we get marginalized one!

ALSO! sparse $\longrightarrow$ dense

• in pose graph optimization

• Generally H matrix has following structure



$\Rightarrow \begin{bmatrix} B & E \\ E^T & C \end{bmatrix} \begin{bmatrix} \delta x_c \\ \delta x_p \end{bmatrix} = \begin{bmatrix} v \\ w \end{bmatrix}$

$\Rightarrow \begin{bmatrix} B - EC^{-1}E^T & E - EC^{-1}C \\ E^T & C \end{bmatrix} \begin{bmatrix} \delta x_c \\ \delta x_p \end{bmatrix} = \begin{bmatrix} v - EC^{-1}w \\ w \end{bmatrix}$

$\Rightarrow \begin{bmatrix} B - EC^{-1}E^T & 0 \\ E^T & C \end{bmatrix} \begin{bmatrix} \delta x_c \\ \delta x_p \end{bmatrix} = \begin{bmatrix} v - EC^{-1}w \\ w \end{bmatrix}$

$\begin{cases} (B - EC^{-1}E^T)\delta x_c = v - EC^{-1}w \\ E^T \delta x_c + C \delta x_p = w \end{cases}$

C^{-1} (diagonal) $\Rightarrow$ easy to solve then solve δx_p!