③

An Informal Introduction to Least Squares

★ Least Squares in General

⇒ Approach for computing a solution for an overdetermined system.

{More equations than unknowns}

⇒ Minimizes the sum of the squared error in the equation.

⇒ Standard approach to a large set of problems.

⇒ Often used to estimate model parameters given observations.

★ Our Problem

⇒ Given a system described by a set of n observation function $\{f_i(X)\}_{i=1:n}$

⇒ Let

- X be the state vector
- Z_i be a measurement of the state X.
- $\hat{Z}_i = f_i(X)$ be a function which maps X to a predicted measurement $\hat{Z}_i$.

⇒ Given n noisy measurements $z_{1:n}$ about the state x.

⇒ **Goal:** Estimate the state x which best explains the measurements $z_{1:n}$

$$x \to f_1(x) = \hat{z}_1 \qquad z_1$$
$$x \to f_2(x) = \hat{z}_2 \qquad z_2$$
$$\vdots \qquad\qquad \vdots$$
$$x \to f_n(x) = \hat{z}_n \qquad z_n$$

State (Unknown)	predicted measurements	real measurements

★ Error Function

⇒ Error e_i is typically the difference between actual and predicted measurement:

$$e_i(x) = z_i - f_i(x)$$

⇒ We assume that the error has zero mean and is normally distributed.
↳ Gaussian error with information matrix Ω_i

{Inverse of Covariance matrix}

⇒ The squared error of a measurement depends only on the state and is a scalar.

$$e_i(x) = e_i(x)^T \Omega_i e_i(x)$$

★ Goal: Find the minimum

⇒ Find the state x^* which minimizes the error given all measurements

$$x^* = \underset{x}{\operatorname{argmin}} \; F(x)$$

global error (scalar) ← $F(x)$

$$F(x) = \sum_i e_i(x)$$

Squared error terms (scalar) ← $e_i(x)$

$$e_i(x) = e_i^T(x)\,\Omega_i\,e_i(x)$$

error term (vector) ← $e_i(x)$

⇒ In general complex and no closed form solution:

↳ Numerical approaches

★ Assumption

⇒ A "good" initial guess is available.

⇒ The error function are "smooth" in the neighborhood of the minima.

⇒ We can solve the problem by iterative local linearizations.

★ Solve via Iterative Local Linearizations

1. Linearize the error terms around the current solution/initial guess.

2. Compute the first derivative of the squared error function.

3. Set it to zero and solve linear system

4. Obtain the new state (that is hopefully closer to the minimum)

5. Iterate

★ Linearizing the Error function

$$e_i(x) = z_i - f_i(x)$$

$$\boxed{e_i(x+\Delta x) = e_i(x) + J_i(x)\Delta x}$$

→ Linearized about x

→ $J_i(x)$ is the jacobian of $e_i(x)$.

★ Squared Error

⇒ With the previous linearization, we can fix x and carry out the minimization in the increments Δx.

$$e_i(x+\Delta x) = e_i^T(x+\Delta x)\,\Omega_i\, e_i(x+\Delta x)$$

$$\Rightarrow (e_i + J_i\Delta x)^T \Omega_i (e_i + J_i \Delta x)$$

$$\Rightarrow \underbrace{e_i^T\Omega e_i}_{c_i} + \underbrace{e_i^T\Omega_i J_i\Delta x + \Delta x^T J_i^T \Omega e_i}_{2\underbrace{e_i^T\Omega_i J_i}_{b_i^T}\Delta x} + \Delta x^T \underbrace{J_i^T\Omega J_i}_{H_i}\Delta x$$

$$\Rightarrow c_i + 2b_i^T\Delta x + \Delta x^T H_i \Delta x$$

★ Global Error

$$F(x+\Delta x) = \sum_i \left(c_i + 2b_i^T\Delta x + \Delta x^T H_i \Delta x\right)$$

$$\Rightarrow \underbrace{\sum_i c_i}_{c} + 2\underbrace{\left(\sum_i b_i^T\right)}_{b}\Delta x + \Delta x^T \underbrace{\left(\sum_i H_i\right)}_{H}\Delta x$$

$$\Rightarrow c + 2b^T\Delta x + \Delta x^T H \Delta x$$

→ {Quadratic form}

★ Deriving the Quadratic form

$$F(x+\Delta x) = \Delta x^T H \Delta x + 2b^T \Delta x + c$$

$$\frac{\delta F}{\delta \Delta x} = (H+H^T)\Delta x + 2b = 0$$

$$\Rightarrow 2H\Delta x + 2b = 0 \quad \{H \text{ is Symmetric}\}$$

$$\Rightarrow \boxed{\Delta x = -H^{-1} b}$$

Δx

★ Gauss-Newton Solution

⇒ Iterate the following steps:

1. Linearize around x and compute for each measurement

$$e_i(x+\Delta x) \simeq e_i(x) + J_i \Delta x$$

2. Compute the terms of the linear System

$$b^T = \sum_i e_i^T \Omega_i J_i \qquad H = \sum_i J_i^T \Omega J_i$$

3. Solve the linear System

$$\Delta x^* = -H^{-1} b$$

4. Updating state

$$x \leftarrow x + \Delta x$$

* Example: Odometry Calibration

⇒ Odometry error can be classified as:

→ Systematic error

{Variation in wheel diameter and Uncertainity in the effective wheel base}

→ Non-Systematic error

{wheel slippage}

⇒ Odometry measurements u_i.

⇒ Ground truth odometry u_i^* is available.

⇒ Eliminate Systematic error through Calibration.

⇒ There is a function $f_i(x)$ which, given some bias parameters x, returns an unbiased (corrected) odometry for the reading u_i' as follows.

$$u_i' = f_i(x) = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} u_i$$

⇒ To obtain the correction function $f(x)$, we need to find the parameters x.

⇒ The state vector is:

$$X = (x_{11}\ x_{12}\ x_{13}\ x_{21}\ x_{22}\ x_{23}\ x_{31}\ x_{32}\ x_{33})^T$$

⇒ The error function is:

$$e_i(X) = u_i^* - \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} u_i$$

⇒ Its derivative is:

$$J_i = \frac{\delta e_i(X)}{\delta X} = -\begin{pmatrix} u_{ix} & u_{iy} & u_{i\theta} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{ix} & u_{iy} & u_{i\theta} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & u_{ix} & u_{iy} & u_{i\theta} \end{pmatrix}$$

★ How to Efficiently solve the Linear System

⇒ Linear system $H \Delta x = -b$

⇒ Can be solved by matrix inversion (in theory)

⇒ In practice:

- → Cholesky factorization
- → QR decomposition
- → Iterative methods such as Conjugate gradients (for large systems)

★ Cholesky decomposition for solving a Linear System

⇒ A Symmetric and positive definite

⇒ System to Solve $Ax = b$

⇒ Cholesky leads to $A = LL^T$ with L being a lower triangular matrix

$$L(L^T x) = b$$

⇒ Solve first: $Ly = b$ | ⇒ and then $L^T x = y$

★ Least Squares vs Probablistic State Estimation

⇒ Bayes rule, independence and Markov assumptions allows us to write

$$P(X_{0:t} \mid Z_{1:t}, u_{1:t})$$
$$= \eta P(X_0) \prod_t P(X_t \mid X_{t-1}, u_t) P(Z_t \mid x_t)$$

★ Log likelyhood

$$\log P(X_{0:t} \mid Z_{1:t}, u_{1:t})$$
$$= \text{Const} + \log P(X_0) + \sum_t \left[\log P(X_t \mid X_{t-1}, u_t) + \log P(Z_t \mid X_t) \right]$$

⇒ Assuming gaussian distribution.

$$\log N(x, \mu, \Sigma)$$

$$= \text{Const} - \frac{1}{2}\underbrace{\underbrace{(x-\mu)^T}_{e(x)^T}\underbrace{\Sigma^{-1}}_{\Omega}\underbrace{(x-\mu)}_{e(x)}}_{e(x)}$$

⇒ $\log P(x_{0:t} \mid z_{1:t}, u_{1:t})$

$$= \text{Const} - \frac{1}{2} e_p(x) - \frac{1}{2}\sum_t \left[e_{u_t}(x) + e_{z_t}(x)\right]$$

⇒ Maximizing the log likelihood leads to

$$\arg\max \log P(x_{0:t} \mid z_{1:t}, u_{1:t})$$

$$\Rightarrow \arg\min \; e_p(x) + \sum_t \left[e_{u_t}(x) + e_{z_t}(x)\right]$$

⇒ Minimizing the squared error is equivalent to maximizing the log likelihood of independent gaussian distributions.
