

Robot Mapping

Least Squares

Cyrill Stachniss



AIS Autonomous Intelligent Systems

1

Three Main SLAM Paradigms

Kalman
filter

Particle
filter

Graph-
based



**least squares
approach to SLAM**

2

Least Squares in General

- Approach for computing a solution for an **overdetermined system**
- "More equations than unknowns"
- Minimizes the **sum of the squared errors** in the equations
- Standard approach to a large set of problems

3

Least Squares History

- Method developed by Carl Friedrich Gauss in 1795 (he was 18 years old)
- First showcase: predicting the future location of the asteroid Ceres in 1801



Courtesy:
Astronomische
Nachrichten, 1828

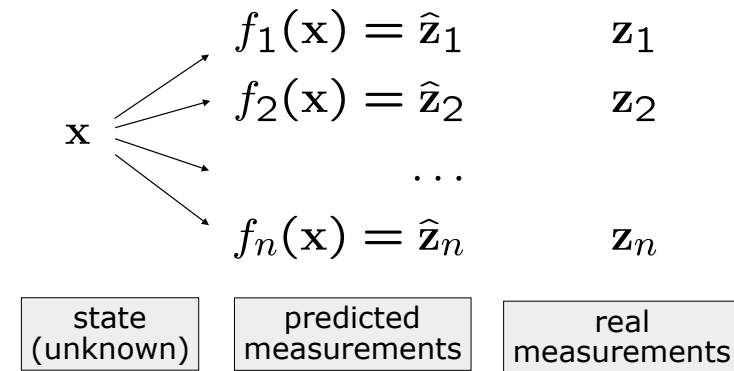
4

Problem

- Given a system described by a set of n observation functions $\{f_i(\mathbf{x})\}_{i=1:n}$
 - Let
 - \mathbf{x} be the state vector
 - \mathbf{z}_i be a measurement of the state \mathbf{x}
 - $\hat{\mathbf{z}}_i = f_i(\mathbf{x})$ be a function which maps \mathbf{x} to a predicted measurement $\hat{\mathbf{z}}_i$
 - Given n noisy measurements $\mathbf{z}_{1:n}$ about the state \mathbf{x}
- ➔ **Goal:** Estimate the state \mathbf{x} which best explains the measurements $\mathbf{z}_{1:n}$

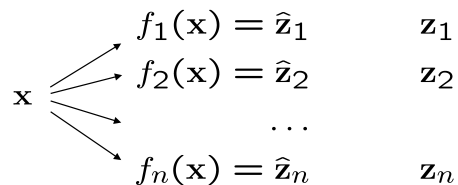
5

Graphical Explanation



6

Example



- \mathbf{x} position of 3D features
- \mathbf{z}_i coordinates of the 3D features projected on camera images
- Estimate the most likely 3D position of the features based on the image projections (given the camera poses)

7

Error Function

- Error \mathbf{e}_i is typically the **difference** between the **predicted and actual** measurement

$$\mathbf{e}_i(\mathbf{x}) = \mathbf{z}_i - f_i(\mathbf{x})$$

- We assume that the error has **zero mean** and is **normally distributed**
- Gaussian error with information matrix $\mathbf{\Omega}_i$
- The squared error of a measurement depends only on the state and is a scalar

$$e_i(\mathbf{x}) = \mathbf{e}_i(\mathbf{x})^T \mathbf{\Omega}_i \mathbf{e}_i(\mathbf{x})$$

8

Goal: Find the Minimum

- Find the state \mathbf{x}^* which minimizes the error given all measurements

$$\begin{aligned}\mathbf{x}^* &= \underset{\mathbf{x}}{\operatorname{argmin}} F(\mathbf{x}) \longleftarrow \text{global error (scalar)} \\ &= \underset{\mathbf{x}}{\operatorname{argmin}} \sum_i e_i(\mathbf{x}) \longleftarrow \text{squared error terms (scalar)} \\ &= \underset{\mathbf{x}}{\operatorname{argmin}} \sum_i \mathbf{e}_i^T(\mathbf{x}) \Omega_i \mathbf{e}_i(\mathbf{x}) \\ &\quad \uparrow \text{error terms (vector)}\end{aligned}$$

9

Goal: Find the Minimum

- Find the state \mathbf{x}^* which minimizes the error given all measurements

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_i \mathbf{e}_i^T(\mathbf{x}) \Omega_i \mathbf{e}_i(\mathbf{x})$$

- A general solution is to derive the global error function and find its nulls
- In general complex and no closed form solution
- ➡ Numerical approaches

10

Assumption

- A “good” initial guess is available
- The error functions are “smooth” in the neighborhood of the (hopefully global) minima
- Then, we can solve the problem by iterative local linearizations

11

Solve Via Iterative Local Linearizations

- Linearize the error terms around the current solution/initial guess
- Compute the first derivative of the squared error function
- Set it to zero and solve linear system
- Obtain the new state (that is hopefully closer to the minimum)
- Iterate

12

Linearizing the Error Function

- Approximate the error functions around an initial guess \mathbf{x} via Taylor expansion

$$e_i(\mathbf{x} + \Delta\mathbf{x}) \simeq \underbrace{e_i(\mathbf{x})}_{e_i} + \mathbf{J}_i(\mathbf{x})\Delta\mathbf{x}$$

- Reminder: Jacobian

$$\mathbf{J}_f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \frac{\partial f_m(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

13

Squared Error

- With the previous linearization, we can fix \mathbf{x} and carry out the minimization in the increments $\Delta\mathbf{x}$
- We replace the Taylor expansion in the squared error terms:

$$e_i(\mathbf{x} + \Delta\mathbf{x}) = \dots$$

14

Squared Error

- With the previous linearization, we can fix \mathbf{x} and carry out the minimization in the increments $\Delta\mathbf{x}$
- We replace the Taylor expansion in the squared error terms:

$$\begin{aligned} e_i(\mathbf{x} + \Delta\mathbf{x}) &= \mathbf{e}_i^T(\mathbf{x} + \Delta\mathbf{x})\Omega_i e_i(\mathbf{x} + \Delta\mathbf{x}) \\ &\simeq (\mathbf{e}_i + \mathbf{J}_i\Delta\mathbf{x})^T \Omega_i (\mathbf{e}_i + \mathbf{J}_i\Delta\mathbf{x}) \\ &= \mathbf{e}_i^T \Omega_i \mathbf{e}_i + \\ &\quad \mathbf{e}_i^T \Omega_i \mathbf{J}_i \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{J}_i^T \Omega_i \mathbf{e}_i + \\ &\quad \Delta\mathbf{x}^T \mathbf{J}_i^T \Omega_i \mathbf{J}_i \Delta\mathbf{x} \end{aligned}$$

15

Squared Error (cont.)

- All summands are scalar so the transposition has no effect
- By grouping similar terms, we obtain:

$$\begin{aligned} e_i(\mathbf{x} + \Delta\mathbf{x}) &\simeq \mathbf{e}_i^T \Omega_i \mathbf{e}_i + \\ &\quad \mathbf{e}_i^T \Omega_i \mathbf{J}_i \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{J}_i^T \Omega_i \mathbf{e}_i + \\ &\quad \Delta\mathbf{x}^T \mathbf{J}_i^T \Omega_i \mathbf{J}_i \Delta\mathbf{x} \\ &= \underbrace{\mathbf{e}_i^T \Omega_i \mathbf{e}_i}_{c_i} + 2 \underbrace{\mathbf{e}_i^T \Omega_i \mathbf{J}_i}_{\mathbf{b}_i^T} \Delta\mathbf{x} + \Delta\mathbf{x}^T \underbrace{\mathbf{J}_i^T \Omega_i \mathbf{J}_i}_{\mathbf{H}_i} \Delta\mathbf{x} \\ &= c_i + 2\mathbf{b}_i^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{H}_i \Delta\mathbf{x} \end{aligned}$$

16

Global Error

- The global error is the sum of the squared errors terms corresponding to the individual measurements
- Form a new expression which approximates the global error in the neighborhood of the current solution \mathbf{x}

$$\begin{aligned} F(\mathbf{x} + \Delta\mathbf{x}) &\simeq \sum_i (c_i + \mathbf{b}_i^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{H}_i \Delta\mathbf{x}) \\ &= \sum_i c_i + 2(\sum_i \mathbf{b}_i^T) \Delta\mathbf{x} + \Delta\mathbf{x}^T (\sum_i \mathbf{H}_i) \Delta\mathbf{x} \end{aligned}$$

17

Global Error (cont.)

$$\begin{aligned} F(\mathbf{x} + \Delta\mathbf{x}) &\simeq \sum_i (c_i + \mathbf{b}_i^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{H}_i \Delta\mathbf{x}) \\ &= \underbrace{\sum_i c_i}_c + 2 \underbrace{(\sum_i \mathbf{b}_i^T)}_{\mathbf{b}^T} \Delta\mathbf{x} + \Delta\mathbf{x}^T \underbrace{(\sum_i \mathbf{H}_i)}_{\mathbf{H}} \Delta\mathbf{x} \\ &= c + 2\mathbf{b}^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{H} \Delta\mathbf{x} \end{aligned}$$

with

$$\begin{aligned} \mathbf{b}^T &= \sum_i \mathbf{e}_i^T \Omega_i \mathbf{J}_i \\ \mathbf{H} &= \sum_i \mathbf{J}_i^T \Omega_i \mathbf{J}_i \end{aligned}$$

18

Quadratic Form

- We can write the global error terms as a quadratic form in $\Delta\mathbf{x}$

$$F(\mathbf{x} + \Delta\mathbf{x}) \simeq c + 2\mathbf{b}^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{H} \Delta\mathbf{x}$$

- We need to compute the derivative of $F(\mathbf{x} + \Delta\mathbf{x})$ w.r.t. $\Delta\mathbf{x}$ (given \mathbf{x})

19

Deriving a Quadratic Form

- Assume a quadratic form

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{b}^T \mathbf{x}$$

- The first derivative is

$$\frac{\partial f}{\partial \mathbf{x}} = (\mathbf{H} + \mathbf{H}^T) \mathbf{x} + \mathbf{b}$$

See: The Matrix Cookbook, Section 2.2.4

20

Quadratic Form

- We can write the global error terms as a quadratic form in $\Delta \mathbf{x}$

$$F(\mathbf{x} + \Delta \mathbf{x}) \simeq c + 2\mathbf{b}^T \Delta \mathbf{x} + \Delta \mathbf{x}^T \mathbf{H} \Delta \mathbf{x}$$

- The derivative of the approximated $F(\mathbf{x} + \Delta \mathbf{x})$ w.r.t. $\Delta \mathbf{x}$ is then:

$$\frac{\partial F(\mathbf{x} + \Delta \mathbf{x})}{\partial \Delta \mathbf{x}} \simeq 2\mathbf{b} + 2\mathbf{H}\Delta \mathbf{x}$$

21

Minimizing the Quadratic Form

- Derivative of $F(\mathbf{x} + \Delta \mathbf{x})$

$$\frac{\partial F(\mathbf{x} + \Delta \mathbf{x})}{\partial \Delta \mathbf{x}} \simeq 2\mathbf{b} + 2\mathbf{H}\Delta \mathbf{x}$$

- Setting it to zero leads to

$$0 = 2\mathbf{b} + 2\mathbf{H}\Delta \mathbf{x}$$

- Which leads to the linear system

$$\mathbf{H}\Delta \mathbf{x} = -\mathbf{b}$$

- The solution for the increment $\Delta \mathbf{x}^*$ is

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

22

Gauss-Newton Solution

Iterate the following steps:

- Linearize around \mathbf{x} and compute for each measurement

$$\mathbf{e}_i(\mathbf{x} + \Delta \mathbf{x}) \simeq \mathbf{e}_i(\mathbf{x}) + \mathbf{J}_i \Delta \mathbf{x}$$

- Compute the terms for the linear system $\mathbf{b}^T = \sum_i \mathbf{e}_i^T \Omega_i \mathbf{J}_i$ $\mathbf{H} = \sum_i \mathbf{J}_i^T \Omega_i \mathbf{J}_i$

- Solve the linear system

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

- Updating state $\mathbf{x} \leftarrow \mathbf{x} + \Delta \mathbf{x}^*$

23

Example: Odometry Calibration

- Odometry measurements \mathbf{u}_i
- Eliminate systematic error through calibration
- Assumption: Ground truth odometry \mathbf{u}_i^* is available
- Ground truth by motion capture, scan-matching, or a SLAM system

24

Example: Odometry Calibration

- There is a function $f_i(\mathbf{x})$ which, given some bias parameters \mathbf{x} , returns an unbiased (corrected) odometry for the reading \mathbf{u}'_i as follows

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

- To obtain the correction function $f(\mathbf{x})$, we need to find the parameters \mathbf{x}

25

Odometry Calibration (cont.)

- The state vector is

$$\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{21} & x_{22} & x_{23} & x_{31} & x_{32} & x_{33} \end{pmatrix}^T$$

- The error function is

$$\mathbf{e}_i(\mathbf{x}) = \mathbf{u}_i^* - \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \mathbf{u}_i$$

- Its derivative is:

$$\mathbf{J}_i = \frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial \mathbf{x}} = - \begin{pmatrix} u_{i,x} & u_{i,y} & u_{i,\theta} & & & & & & \\ & & & u_{i,x} & u_{i,y} & u_{i,\theta} & & & \\ & & & & & & u_{i,x} & u_{i,y} & u_{i,\theta} \end{pmatrix}$$

Does not depend on \mathbf{x} , why? What are the consequences?



\mathbf{e} is linear, no need to iterate!

Questions

- How do the parameters look like if the odometry is perfect?
- How many measurements (at least) are needed to find a solution for the calibration problem?
- \mathbf{H} is symmetric. Why?
- How does the structure of the measurement function affect the structure of \mathbf{H} ?

27

How to Efficiently Solve the Linear System?

- Linear system $\mathbf{H}\Delta\mathbf{x} = -\mathbf{b}$
- Can be solved by matrix inversion (in theory)
- In practice:
 - Cholesky factorization
 - QR decomposition
 - Iterative methods such as conjugate gradients (for large systems)

28

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
- Solve first
$$Ly = b$$
- and then
$$L^T x = y$$

29

Gauss-Newton Summary

Method to minimize a squared error:

- Start with an initial guess
- Linearize the individual error functions
- This leads to a quadratic form
- One obtains a linear system by setting its derivative to zero
- Solving the linear systems leads to a state update
- Iterate

30

Relation to Probabilistic State Estimation

- So far, we minimized an error function
- How does this relate to state estimation in the probabilistic sense?

31

General State Estimation

- Bayes rule, independence and Markov assumptions allow us to write

$$\begin{aligned} p(x_{0:t} \mid z_{1:t}, u_{1:t}) \\ = \eta p(x_0) \prod_t [p(x_t \mid x_{t-1}, u_t) p(z_t \mid x_t)] \end{aligned}$$

32

Log Likelihood

- Written as the log likelihood, leads to

$$\begin{aligned}\log p(x_{0:t} \mid z_{1:t}, u_{1:t}) \\ &= \text{const.} + \log p(x_0) \\ &\quad + \sum_t [\log p(x_t \mid x_{t-1}, u_t) + \log p(z_t \mid x_t)]\end{aligned}$$

33

Gaussian Assumption

- Assuming Gaussian distributions

$$\begin{aligned}\log p(x_{0:t} \mid z_{1:t}, u_{1:t}) \\ &= \text{const.} + \underbrace{\log p(x_0)}_{\mathcal{N}} \\ &\quad + \sum_t \left[\underbrace{\log p(x_t \mid x_{t-1}, u_t)}_{\mathcal{N}} + \underbrace{\log p(z_t \mid x_t)}_{\mathcal{N}} \right]\end{aligned}$$

34

Log of a Gaussian

- Log likelihood of a Gaussian

$$\begin{aligned}\log \mathcal{N}(x, \mu, \Sigma) \\ &= \text{const.} - \frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\end{aligned}$$

35

Error Function as Exponent

- Log likelihood of a Gaussian

$$\begin{aligned}\log \mathcal{N}(x, \mu, \Sigma) \\ &= \text{const.} - \frac{1}{2} \underbrace{\underbrace{(x - \mu)^T}_{\mathbf{e}^T(x)} \underbrace{\Sigma^{-1}}_{\Omega} \underbrace{(x - \mu)}_{\mathbf{e}(x)}}_{e(x)}\end{aligned}$$

- is up to a constant equivalent to the error functions used before

36

Log Likelihood with Error Terms

- Assuming Gaussian distributions

$$\begin{aligned}\log p(x_{0:t} \mid z_{1:t}, u_{1:t}) \\ = \text{const.} - \frac{1}{2}e_p(x) - \frac{1}{2} \sum_t [e_{u_t}(x) + e_{z_t}(x)]\end{aligned}$$

37

Maximizing the Log Likelihood

- Assuming Gaussian distributions

$$\begin{aligned}\log p(x_{0:t} \mid z_{1:t}, u_{1:t}) \\ = \text{const.} - \frac{1}{2}e_p(x) - \frac{1}{2} \sum_t [e_{u_t}(x) + e_{z_t}(x)]\end{aligned}$$

- Maximizing the log likelihood leads to

$$\begin{aligned}\operatorname{argmax} \log p(x_{0:t} \mid z_{1:t}, u_{1:t}) \\ = \operatorname{argmin} e_p(x) + \sum_t [e_{u_t}(x) + e_{z_t}(x)]\end{aligned}$$

38

Minimizing the Squared Error is Equivalent to Maximizing the Log Likelihood of Independent Gaussian Distributions

with individual error terms for the motions, measurements, and prior:

$$\begin{aligned}\operatorname{argmax} \log p(x_{0:t} \mid z_{1:t}, u_{1:t}) \\ = \operatorname{argmin} e_p(x) + \sum_t [e_{u_t}(x) + e_{z_t}(x)]\end{aligned}$$

39

Summary

- Technique to minimize squared error functions
- Gauss-Newton is an iterative approach for non-linear problems
- Uses linearization (approximation!)
- Equivalent to maximizing the log likelihood of independent Gaussians
- Popular method in a lot of disciplines

40

Literature

Least Squares and Gauss-Newton

- Basically every textbook on numeric calculus or optimization
- Wikipedia (for a brief summary)

Relation to Probability Theory

- Thrun et al.: “Probabilistic Robotics”, Chapter 11.4