Lecture 12: Pose-and-Point Estimation Problems

AER1513: State Estimation

Timothy D. Barfoot

University of Toronto

UNIVERSITY OF TORONTO

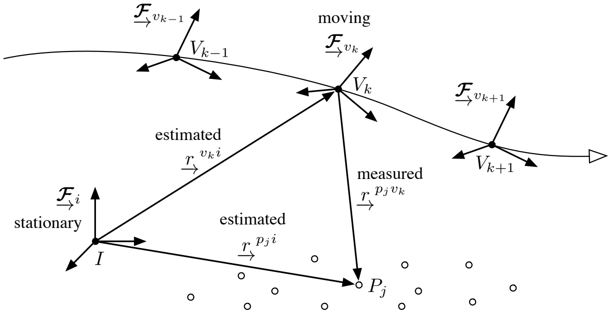
Outline

UNIVERSITY OF TORONTO

Motivation

- in this lecture, we will address one of the most fundamental problems in robotics, estimating the trajectory of a robot and the structure of the world around it (i.e., point landmarks) together
- in robotics this is called the simultaneous localization and mapping (SLAM) problem
- in computer vision an almost identical problem came to prominence through the application of aligning aerial photographs into a mosaic; the classic solution to this problem is called bundle adjustment (BA)
- we will look at BA through the lens of our $SE(3)$ estimation techniques.

Setup



– the setup is very similar to the pose estimation setup from the last lecture

– the main difference is that the landmark positions are not known and must also be estimated from nonlinear observations

Bundle adjustment

- the state that we wish to estimate is

$$
\begin{aligned}
\mathbf{T}_k = \mathbf{T}_{v_k i} \quad &: \quad \text{pose of vehicle at time } k \\
\mathbf{p}_j = \begin{bmatrix} \mathbf{r}_i^{p_j i} \\ 1 \end{bmatrix} \quad &: \quad \text{position of landmark } j
\end{aligned}
$$

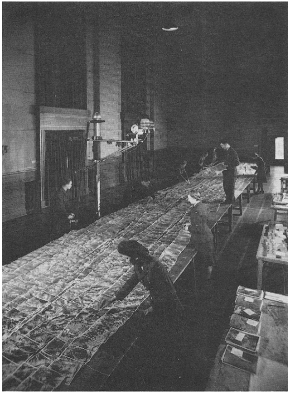
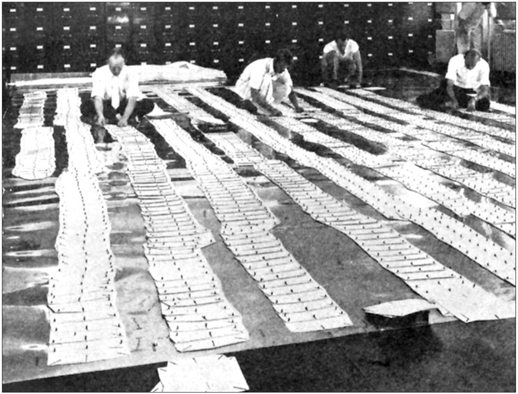
  where $k = 1 \ldots K$ and $j = 1 \ldots M$
- we will use the shorthand,

$$\mathbf{x} = \{\mathbf{T}_1, \ldots, \mathbf{T}_K, \mathbf{p}_1, \ldots, \mathbf{p}_M\} \tag{1}$$

  as well as $\mathbf{x}_{jk} = \{\mathbf{T}_k, \mathbf{p}_j\}$ to indicate the subset of the state including the kth pose and jth landmark
- notably, we exclude $\mathbf{T}_0$ from the state to be estimated as the system is otherwise unobservable

UNIVERSITY OF TORONTO

Be glad you weren't a grad student back then

- this problem originated from stitching together aerial images
- the original solution was pretty slow

UNIVERSITY OF TORONTO

Measurement model

- BA does not use a motion model (no prior), so we set it up as a maximum likelihood problem
- the measurement, $\mathbf{y}_{jk}$, will correspond to some observation of point j from pose k (i.e., some function of $\mathbf{r}_{v_k}^{p_j v_k}$)
- the measurement model for this problem will be of the form

$$\mathbf{y}_{jk} = \mathbf{g}\left(\mathbf{x}_{jk}\right) + \mathbf{n}_{jk} \tag{2}$$

  where $\mathbf{g}(\cdot)$ is the nonlinear model and $\mathbf{n}_{jk} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_{jk}\right)$ is additive Gaussian noise
- we can use the shorthand

$$\mathbf{y} = \{\mathbf{y}_{10}, \dots \mathbf{y}_{M0}, \dots, \mathbf{y}_{1K}, \dots, \mathbf{y}_{MK}\} \tag{3}$$

  to capture all the measurements that we have available

UNIVERSITY OF TORONTO

Two nonlinearities

- we can think of the overall observation model as the composition of two nonlinearities: one to transform the point into the vehicle frame and one turn that point into the actual sensor measurement through a camera (or other sensor) model
- letting

$$\mathbf{z}(\mathbf{x}_{jk}) = \mathbf{T}_k \mathbf{p}_j \tag{4}$$

  we can write

$$\mathbf{g}\left(\mathbf{x}_{jk}\right) = \mathbf{s}\left(\mathbf{z}(\mathbf{x}_{jk})\right) \tag{5}$$

  where $\mathbf{s}(\cdot)$ is the nonlinear camera (or sensor) model
- in other words, we have

$$\mathbf{g} = \mathbf{s} \circ \mathbf{z} \tag{6}$$

  in terms of the composition of functions

UNIVERSITY OF TORONTO

Perturbations

- we define the following perturbations to our state variables:

$$
\begin{aligned}
\mathbf{T}_k &= \exp\left(\boldsymbol{\epsilon}_k^\wedge\right)\mathbf{T}_{\mathrm{op},k} \approx \left(\mathbf{1} + \boldsymbol{\epsilon}_k^\wedge\right)\mathbf{T}_{\mathrm{op},k} && \text{(7a)} \\
\mathbf{p}_j &= \mathbf{p}_{\mathrm{op},j} + \mathbf{D}\,\boldsymbol{\zeta}_j && \text{(7b)}
\end{aligned}
$$

  where

$$
\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{8}
$$

  is a dilation matrix so that our landmark perturbation, $\boldsymbol{\zeta}_j$, is 3×1

- the book works things out to second order but we'll stick to first order in the interest of simplicity

Shorthands

- we will use the shorthands

$$\mathbf{x}_{\text{op}} = \left\{ \mathbf{T}_{\text{op},1}, \ldots, \mathbf{T}_{\text{op},K}, \mathbf{p}_{\text{op},1}, \ldots, \mathbf{p}_{\text{op},M} \right\}, \quad \delta\mathbf{x} = \left[\begin{array}{c} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_K \\ \hline \boldsymbol{\zeta}_1 \\ \vdots \\ \boldsymbol{\zeta}_M \end{array} \right] \tag{9}$$

for the operating point and perturbation to the whole trajectory

- we will also use

$$\mathbf{x}_{\text{op},jk} = \left\{ \mathbf{T}_{\text{op},k}, \mathbf{p}_{\text{op},j} \right\}, \quad \delta\mathbf{x}_{jk} = \begin{bmatrix} \boldsymbol{\epsilon}_k \\ \boldsymbol{\zeta}_j \end{bmatrix} \tag{10}$$

for the parts associated with kth pose and the jth landmark

First nonlinearity

– using the perturbation schemes above, we have for the linearization of the first nonlinearity in measurement model that

$$\begin{aligned} \mathbf{z}(\mathbf{x}_{jk}) &= \mathbf{T}_k \mathbf{p}_j \\ &\approx \left(\mathbf{1} + \boldsymbol{\epsilon}_k^\wedge\right) \mathbf{T}_{\mathrm{op},k} \left(\mathbf{p}_{\mathrm{op},j} + \mathbf{D}\,\boldsymbol{\zeta}_j\right) \\ &\approx \mathbf{T}_{\mathrm{op},k}\mathbf{p}_{\mathrm{op},j} + \boldsymbol{\epsilon}_k^\wedge \mathbf{T}_{\mathrm{op},k}\mathbf{p}_{\mathrm{op},j} + \mathbf{T}_{\mathrm{op},k}\mathbf{D}\,\boldsymbol{\zeta}_j \\ &= \mathbf{z}(\mathbf{x}_{\mathrm{op},jk}) + \mathbf{Z}_{jk}\,\delta\mathbf{x}_{jk} \end{aligned} \tag{11}$$

correct to first order in $\delta\mathbf{x}_{jk}$, where

$$\mathbf{z}(\mathbf{x}_{\mathrm{op},jk}) = \mathbf{T}_{\mathrm{op},k}\mathbf{p}_{\mathrm{op},j} \tag{12a}$$

$$\mathbf{Z}_{jk} = \begin{bmatrix} (\mathbf{T}_{\mathrm{op},k}\mathbf{p}_{\mathrm{op},j})^{\odot} & \mathbf{T}_{\mathrm{op},k}\mathbf{D} \end{bmatrix} \tag{12b}$$

$$\delta\mathbf{x}_{jk} = \begin{bmatrix} \boldsymbol{\epsilon}_k \\ \boldsymbol{\zeta}_j \end{bmatrix} \tag{12c}$$

UNIVERSITY OF TORONTO

Second nonlinearity

- we insert the linearization of the first function into the second to work out the chain rule:

$$
\begin{aligned}
\mathbf{g}(\mathbf{x}_{jk}) &= \mathbf{s}\left(\mathbf{z}(\mathbf{x}_{jk})\right) \\
&\approx \mathbf{s}\Big(\mathbf{z}(\mathbf{x}_{\mathrm{op},jk}) + \mathbf{Z}_{jk}\,\delta\mathbf{x}_{jk}\Big) \\
&\approx \mathbf{s}(\mathbf{z}(\mathbf{x}_{\mathrm{op},jk})) + \mathbf{S}_{jk}\mathbf{Z}_{jk}\,\delta\mathbf{x}_{jk} \\
&\approx \mathbf{g}(\mathbf{x}_{\mathrm{op},jk}) + \mathbf{G}_{jk}\,\delta\mathbf{x}_{jk} \qquad (13)
\end{aligned}
$$

correct to first order, where

$$
\begin{aligned}
\mathbf{g}(\mathbf{x}_{\mathrm{op},jk}) &= \mathbf{s}(\mathbf{z}(\mathbf{x}_{\mathrm{op},jk})) \qquad (14a) \\
\mathbf{G}_{jk} &= \mathbf{S}_{jk}\mathbf{Z}_{jk} \qquad (14b) \\
\mathbf{S}_{jk} &= \left.\frac{\partial \mathbf{s}}{\partial \mathbf{z}}\right|_{\mathbf{z}(\mathbf{x}_{\mathrm{op},jk})} \qquad (14c)
\end{aligned}
$$

UNIVERSITY OF TORONTO

Error terms

- for each observation of a point from a pose, we define an error term as

$$\mathbf{e}_{y,jk}(\mathbf{x}) = \mathbf{y}_{jk} - \mathbf{g}\left(\mathbf{x}_{jk}\right) \tag{15}$$

  where $\mathbf{y}_{jk}$ is the measured quantity and $\mathbf{g}$ is our observation model described above
- approximating the error function, we have

$$\mathbf{e}_{y,jk}(\mathbf{x}) \approx \underbrace{\mathbf{y}_{jk} - \mathbf{g}(\mathbf{x}_{\mathrm{op},jk})}_{\mathbf{e}_{y,jk}(\mathbf{x}_{\mathrm{op}})} - \mathbf{G}_{jk}\delta\mathbf{x}_{jk} \tag{16}$$

- we can now form an objective function

UNIVERSITY OF TORONTO

Objective function

– we seek to find the values of $\mathbf{x}$ to minimize the following objective function:

$$J(\mathbf{x}) = \frac{1}{2} \sum_{j,k} \mathbf{e}_{y,jk}(\mathbf{x})^T \mathbf{R}_{jk}^{-1} \mathbf{e}_{y,jk}(\mathbf{x}) \quad (17)$$

where $\mathbf{x}$ is the full state that we wish to estimate (all poses and landmarks) and $\mathbf{R}_{jk}$ is the symmetric, positive-definite covariance matrix associated with the jkth measurement

– if a particular landmark is not actually observed from a particular pose, we can simply delete the appropriate term from the objective function

UNIVERSITY OF TORONTO

Gauss-Newton optimization

– inserting our linearized error function (to produce our shortcut to Gauss-Newton) we have

$$J(\mathbf{x}) \approx J(\mathbf{x}_{\text{op}}) - \mathbf{b}^T \, \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^T \, \mathbf{A} \, \delta\mathbf{x} \tag{18}$$

correct to first order, where

$$\mathbf{b} = \sum_{j,k} \mathbf{P}_{jk}^T \mathbf{G}_{jk}^T \mathbf{R}_{jk}^{-1} \mathbf{e}_{y,jk}(\mathbf{x}_{\text{op}}) \tag{19a}$$

$$\mathbf{A} = \sum_{j,k} \mathbf{P}_{jk}^T \mathbf{G}_{jk}^T \mathbf{R}_{jk}^{-1} \mathbf{G}_{jk} \mathbf{P}_{jk} \tag{19b}$$

$$\delta\mathbf{x}_{jk} = \mathbf{P}_{jk} \, \delta\mathbf{x} \tag{19c}$$

and where $\mathbf{P}_{jk}$ is an appropriate projection matrix to pick off the jkth components of the overall perturbed state, $\delta\mathbf{x}$

UNIVERSITY OF TORONTO

Gauss-Newton optimization

– we now minimize $J(\mathbf{x})$ with respect to $\delta\mathbf{x}$ by taking the derivative:

$$\frac{\partial J(\mathbf{x})}{\partial\,\delta\mathbf{x}^T} = -\mathbf{b} + \mathbf{A}\,\delta\mathbf{x} \tag{20}$$

– setting this to zero, the optimal perturbation, $\delta\mathbf{x}^\star$, is the solution to the following linear system:

$$\mathbf{A}\,\delta\mathbf{x}^\star = \mathbf{b} \tag{21}$$

– we iterate between solving for the optimal perturbation and updating the nominal quantities using

$$\mathbf{T}_{\mathrm{op},k} \quad \leftarrow \quad \exp\left(\boldsymbol{\epsilon}_k^{\star\wedge}\right)\mathbf{T}_{\mathrm{op},k} \tag{22a}$$

$$\mathbf{p}_{\mathrm{op},j} \quad \leftarrow \quad \mathbf{p}_{\mathrm{op},j} + \mathbf{D}\,\boldsymbol{\zeta}_j^\star \tag{22b}$$

which ensure that $\mathbf{T}_{\mathrm{op},k} \in SE(3)$ and $\mathbf{p}_{\mathrm{op},j}$ keeps its bottom (fourth) entry equal to 1

UNIVERSITY OF TORONTO

Sparsity

- at each iteration of Gauss-Newton, we are faced with solving a system of the following form:

$$\underbrace{\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \delta \mathbf{x}_1^\star \\ \delta \mathbf{x}_2^\star \end{bmatrix}}_{\delta \mathbf{x}^\star} = \underbrace{\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}}_{\mathbf{b}} \tag{23}$$

  where the state, $\delta \mathbf{x}^\star$, has been partitioned into parts corresponding to (1) the pose perturbation, $\delta \mathbf{x}_1^\star = \boldsymbol{\epsilon}^\star$, and (2) the landmark perturbations, $\delta \mathbf{x}_2^\star = \boldsymbol{\zeta}^\star$

- it turns out that $\mathbf{A}$ has a very special sparsity pattern and is sometimes referred to as an arrowhead matrix
- this pattern is due to the presence of the projection matrices, $\mathbf{P}_{jk}$, in each term of $\mathbf{A}$; they embody the fact that each measurement involves just one pose variable and one landmark

UNIVERSITY OF TORONTO

Arrowhead

$$
\mathbf{A} = \begin{array}{cc}
\left[\begin{array}{ccccc|ccccc}
* & & & & & * & \cdots & * & \cdots & * \\
 & \ddots & & & & \vdots & & \vdots & & \vdots \\
 & & * & & & * & \cdots & * & \cdots & * \\
 & & & \ddots & & \vdots & & \vdots & & \vdots \\
 & & & & * & * & \cdots & * & \cdots & * \\ \hline
* & \cdots & * & \cdots & * & * & & & & \\
\vdots & & \vdots & & \vdots & & \ddots & & & \\
* & \cdots & * & \cdots & * & & & * & & \\
\vdots & & \vdots & & \vdots & & & & \ddots & \\
* & \cdots & * & \cdots & * & & & & & *
\end{array}\right]
&
\begin{array}{c}
\boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_k \\ \vdots \\ \boldsymbol{\epsilon}_K \\ \boldsymbol{\zeta}_1 \\ \vdots \\ \boldsymbol{\zeta}_j \\ \vdots \\ \boldsymbol{\zeta}_M
\end{array}
\\
\begin{array}{cccccccccc}
\boldsymbol{\epsilon}_1 & \cdots & \boldsymbol{\epsilon}_k & \cdots & \boldsymbol{\epsilon}_K & \boldsymbol{\zeta}_1 & \cdots & \boldsymbol{\zeta}_j & \cdots & \boldsymbol{\zeta}_M
\end{array}
&
\end{array}
$$

UNIVERSITY OF TORONTO

Schur complement

- to exploit the arrowhead sparsity we can use the Schur complement
- we premultiply both sides by the invertible matrix

$$\begin{bmatrix} \mathbf{1} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

  so that our equation becomes

$$\begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^T & \mathbf{0} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \delta\mathbf{x}_1^\star \\ \delta\mathbf{x}_2^\star \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{b}_2 \\ \mathbf{b}_2 \end{bmatrix}$$

- we may then easily solve for $\delta\mathbf{x}_1^\star$ and since $\mathbf{A}_{22}$ is block-diagonal, $\mathbf{A}_{22}^{-1}$ is cheap to compute
- finally, $\delta\mathbf{x}_2^\star$ can also be efficiently computed through back-substitution, again owing to the sparsity of $\mathbf{A}_{22}$
- this procedure brings the complexity of each solve down from $O\left((K+M)^3\right)$ without sparsity to $O\left(K^3 + K^2 M\right)$ with sparsity

UNIVERSITY OF TORONTO

Cholesky decomposition

– alternatively, we can exploit the arrowhead via a sparse Cholesky decomposition:

$$\underbrace{\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix}}_{\mathbf{A}} = \underbrace{\begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \mathbf{U}_{11}^T & \mathbf{0} \\ \mathbf{U}_{12}^T & \mathbf{U}_{22}^T \end{bmatrix}}_{\mathbf{U}^T} \tag{24}$$

where $\mathbf{U}$ is an upper-triangular matrix and

$$\begin{aligned} \mathbf{U}_{22}\mathbf{U}_{22}^T = \mathbf{A}_{22} \quad &: \quad \text{cheap to compute } \mathbf{U}_{22} \text{ via Cholesky} \\ & \quad\;\; \text{due to } \mathbf{A}_{22} \text{ block-diagonal} \\ \mathbf{U}_{12}\mathbf{U}_{22}^T = \mathbf{A}_{12} \quad &: \quad \text{cheap to solve for } \mathbf{U}_{12} \\ & \quad\;\; \text{due to } \mathbf{U}_{22} \text{ block-diagonal} \\ \mathbf{U}_{11}\mathbf{U}_{11}^T + \mathbf{U}_{12}\mathbf{U}_{12}^T = \mathbf{A}_{11} \quad &: \quad \text{cheap to compute } \mathbf{U}_{11} \text{ via Cholesky} \\ & \quad\;\; \text{due to small size of } \delta\mathbf{x}_1^\star \end{aligned}$$

so that we have a procedure to very efficiently compute $\mathbf{U}$, owing to the sparsity of $\mathbf{A}_{22}$

Cholesky decomposition

- after computing the Cholesky decomposition we can solve our linear system in two steps
- first, solve

$$\mathbf{U}\mathbf{c} = \mathbf{b} \tag{25}$$

  for a temporary variable, $\mathbf{c}$
- this can be done very quickly since $\mathbf{U}$ is upper-triangular and so can be solved from the bottom to the top through substitution and exploiting the additional known sparsity of $\mathbf{U}$
- second, solve

$$\mathbf{U}^T \delta\mathbf{x}^\star = \mathbf{c} \tag{26}$$

  for $\delta\mathbf{x}^\star$
- again, since $\mathbf{U}^T$ is lower-triangular we can solve quickly from the top to the bottom through substitution and exploiting the sparsity

UNIVERSITY OF TORONTO

Comments and Summary

- setting BA up as a batch problem made the sparsity easy to visualize and exploit
- if we include a prior on motion in addition to our measurements, this is called simultaneous localization and mapping (SLAM) and then the $\mathbf{A}_{11}$ matrix becomes block-tridiagonal instead of block-diagonal
- we exploited the sparsity of $\mathbf{A}_{22}$ in our solution, but we could have instead chosen to exploit the sparsity of $\mathbf{A}_{11}$, even in the SLAM case
- we could turn this into a sliding-window filter by using mini batches that slide along with time
- our $SE(3)$ approach made it easy to handle three-dimensional landmarks even with a nonlinear measurement model

UNIVERSITY OF TORONTO