Lecture 8: Handling Nonidealities in Estimation

AER1513: State Estimation

Timothy D. Barfoot

University of Toronto

Copyright © 2023

UNIVERSITY OF TORONTO

Outline

Lecture 8: Handling Nonidealities in Estimation

- Motivation
- Biases
- Correspondences
- Outliers

UNIVERSITY OF TORONTO

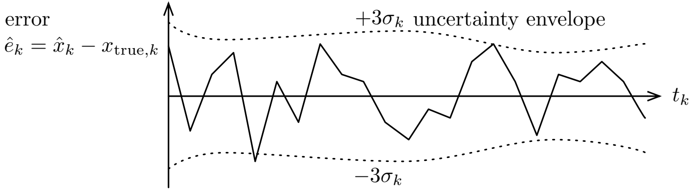
Motivation

- this lecture is a mixed bag of topics that are tied together by a theme: what to do when our data does not conform to the assumptions we've made to derive our estimators?
- a big assumption we have made throughout is that our noise variables are zero-mean – in reality there could be a bias and in some cases we can deal with this (and some we can't)
- sometimes we are faced with associating a particular measurement with a source (e.g., a range measurement with a particular landmark) – this is the problem of data association or determining correspondences
- finally, most of our estimators require us to provide the covariance associated with the noise variables, but sometimes we can be off in our determination of these covariances and so we have outliers (i.e., measurements that are not very probable given our noise models) – we'll discuss some ways of dealing with outliers

UNIVERSITY OF TORONTO

Is our estimator healthy?

- before dealing with nonidealities, we need a notion of what properties a healthy estimator should have
- we can plot the error of the estimate relative to ground truth along with an uncertainty envelope:



- qualitatively, we'd like the error to hover around zero (unbiased) and the stay mostly inside the uncertainty envelope (consistent)

UNIVERSITY OF TORONTO

Unbiased

- quantitatively, for an unbiased estimator we would like $E[\hat{e}_k] = 0$, with the expectation over many trials
- in practice, we employ the ergodic hypothesis and compute the sample mean over a large number of timesteps:

$$\hat{e}_{\text{mean}} = \frac{1}{K}\sum_{k=1}^{K} \hat{e}_k \tag{1}$$

- naturally, this never be exactly zero but we can perform a statistical hypothesis test to see if it is likely that the true mean error is zero or not:

$$Q_{\mathcal{N}(0,\sigma^2/K)}(\ell) \leq \hat{e}_{\text{mean}} \leq Q_{\mathcal{N}(0,\sigma^2/K)}(u) \tag{2}$$

where Q is the quantile function for the zero-mean sample mean and ℓ and u define a confidence interval (e.g., 95% two-sided)

UNIVERSITY OF TORONTO

Consistent

- quantitatively, for an consistent estimator we would like $E[\hat{e}_k^2/\hat{P}_k] = 1$, with the expectation over many trials
- in practice, we employ the ergodic hypothesis and compute the quantity $\sum_{k=1}^{K} \frac{\hat{e}_k^2}{\hat{P}_k}$ over a large number of timesteps, which is distributed according to a chi-squared distribution with K degrees of freedom
- a statistical hypothesis test for consistency is then

$$Q_{\chi^2(K)}(\ell) \leq \sum_{k=1}^{K} \frac{\hat{e}_k^2}{\hat{P}_k^2} \leq Q_{\chi^2(K)}(u), \tag{3}$$

- in higher dimensions, we use the more general Normalized Estimation Error Squared (NEES)
- with no groundtruth, we can use the Normalized Innovation Squared (NIS) test instead

UNIVERSITY OF TORONTO

Biased system model

- imagine we now have a system like this:

$$\begin{aligned} \mathbf{x}_k &= \boldsymbol{A}\mathbf{x}_{k-1} + \boldsymbol{B}(\mathbf{u}_k + \bar{\mathbf{u}}) + \mathbf{w}_k && (4a) \\ \mathbf{y}_k &= \boldsymbol{C}\mathbf{x}_k + \bar{\mathbf{y}} + \mathbf{n}_k && (4b) \end{aligned}$$

  where $\bar{\mathbf{u}}$ is an input bias and $\bar{\mathbf{y}}$ a measurement bias

- we will continue to assume that all measurements are corrupted with zero-mean Gaussian noise:

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}), \quad \mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}) \tag{5}$$

UNIVERSITY OF TORONTO

Effect on Kalman filter

– we previously showed that the Kalman filter was unbiased and consistent by examining the error dynamics

– as before, the estimation errors are

$$
\begin{aligned}
\check{\mathbf{e}}_k &= \check{\mathbf{x}}_k - \mathbf{x}_k && (6a)\\
\hat{\mathbf{e}}_k &= \hat{\mathbf{x}}_k - \mathbf{x}_k && (6b)
\end{aligned}
$$

– the error dynamics (difference of estimator and system model) are

$$
\begin{aligned}
\check{\mathbf{e}}_k &= \boldsymbol{A}\hat{\mathbf{e}}_{k-1} - \left(\boldsymbol{B}\bar{\mathbf{u}} + \mathbf{w}_k\right) && (7a)\\
\hat{\mathbf{e}}_k &= \left(\mathbf{1} - \mathbf{K}_k\boldsymbol{C}\right)\check{\mathbf{e}}_k + \mathbf{K}_k\left(\bar{\mathbf{y}} + \mathbf{n}_k\right) && (7b)
\end{aligned}
$$

where $\hat{\mathbf{e}}_0 = \hat{\mathbf{x}}_0 - \mathbf{x}_0$

– we see new bias terms in the error dynamics due to the modified system model

UNIVERSITY OF TORONTO

Effect on Kalman filter

- we will still assume our initial state knowledge is unbiased and consistent:

$$E\left[\hat{\mathbf{e}}_0\right] = \mathbf{0}, \quad E\left[\hat{\mathbf{e}}_0 \hat{\mathbf{e}}_0^T\right] = \hat{\mathbf{P}}_0 \tag{8}$$

- at $k = 1$, the mean of our errors are

$$\begin{aligned} E\left[\check{\mathbf{e}}_1\right] &= \boldsymbol{A} \underbrace{E\left[\hat{\mathbf{e}}_0\right]}_{\mathbf{0}} - \left(\boldsymbol{B}\bar{\mathbf{u}} + \underbrace{E\left[\mathbf{w}_1\right]}_{\mathbf{0}}\right) = -\boldsymbol{B}\bar{\mathbf{u}} \neq \mathbf{0} && \text{(9a)} \\ E\left[\hat{\mathbf{e}}_1\right] &= \left(\mathbf{1} - \mathbf{K}_1 \boldsymbol{C}\right) \underbrace{E\left[\check{\mathbf{e}}_1\right]}_{-\boldsymbol{B}\bar{\mathbf{u}}} + \mathbf{K}_1 \left(\bar{\mathbf{y}} + \underbrace{E\left[\mathbf{n}_1\right]}_{\mathbf{0}}\right) \\ &= -\left(\mathbf{1} - \mathbf{K}_1 \boldsymbol{C}\right) \boldsymbol{B}\bar{\mathbf{u}} + \mathbf{K}_1 \bar{\mathbf{y}} \neq \mathbf{0} && \text{(9b)} \end{aligned}$$

which are already biased in the case that $\bar{\mathbf{u}} \neq \mathbf{0}$ and/or $\bar{\mathbf{y}} \neq \mathbf{0}$

UNIVERSITY OF TORONTO

Effect on Kalman filter

– for the covariance of the predicted error we have

$$
\begin{aligned}
E\left[\check{\mathbf{e}}_1\check{\mathbf{e}}_1^T\right] &= E\left[\left(\boldsymbol{A}\hat{\mathbf{e}}_0 - (\boldsymbol{B}\bar{\mathbf{u}} + \mathbf{w}_1)\right)\left(\boldsymbol{A}\hat{\mathbf{e}}_0 - (\boldsymbol{B}\bar{\mathbf{u}} + \mathbf{w}_1)\right)^T\right] \\
&= \underbrace{E\left[(\boldsymbol{A}\hat{\mathbf{e}}_0 - \mathbf{w}_1)(\boldsymbol{A}\hat{\mathbf{e}}_0 - \mathbf{w}_1)^T\right]}_{\check{\mathbf{P}}_1} + (-\boldsymbol{B}\bar{\mathbf{u}})\,\underbrace{E\left[(\boldsymbol{A}\hat{\mathbf{e}}_0 - \mathbf{w}_1)^T\right]}_{\mathbf{0}} \\
&\qquad + \underbrace{E\left[(\boldsymbol{A}\hat{\mathbf{e}}_0 - \mathbf{w}_1)\right]}_{\mathbf{0}}(-\boldsymbol{B}\bar{\mathbf{u}})^T + (-\boldsymbol{B}\bar{\mathbf{u}})(-\boldsymbol{B}\bar{\mathbf{u}})^T \\
&= \check{\mathbf{P}}_1 + (-\boldsymbol{B}\bar{\mathbf{u}})(-\boldsymbol{B}\bar{\mathbf{u}})^T \qquad (10a)
\end{aligned}
$$

– rearranging we see that

$$
\check{\mathbf{P}}_1 = E\left[\check{\mathbf{e}}_1\check{\mathbf{e}}_1^T\right] - \underbrace{E[\check{\mathbf{e}}_1]\,E[\check{\mathbf{e}}_1]^T}_{\text{bias effect}} \qquad (11)
$$

and therefore the KF will underestimate the true uncertainty in the error and become inconsistent

UNIVERSITY OF TORONTO

Effect on Kalman filter

- for the covariance of the corrected error we have

$$\begin{aligned}
E\left[\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^T\right] &= E\left[\left(\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\check{\mathbf{e}}_1+\mathbf{K}_1(\bar{\mathbf{y}}+\mathbf{n}_1)\right)\left(\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\check{\mathbf{e}}_1+\mathbf{K}_1(\bar{\mathbf{y}}+\mathbf{n}_1)\right)^T\right] \\
&= \underbrace{E\left[\left(\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\check{\mathbf{e}}_1+\mathbf{K}_1\mathbf{n}_1\right)\left(\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\check{\mathbf{e}}_1+\mathbf{K}_1\mathbf{n}_1\right)^T\right]}_{\hat{\mathbf{P}}_1+(\mathbf{1}-\mathbf{K}_1\boldsymbol{C})\mathbf{B}\bar{\mathbf{u}}\bar{\mathbf{u}}^T\mathbf{B}^T(\mathbf{1}-\mathbf{K}_1\boldsymbol{C})^T} \\
&\quad +(\mathbf{K}_1\bar{\mathbf{y}})\underbrace{E\left[\left(\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\check{\mathbf{e}}_1+\mathbf{K}_1\mathbf{n}_1\right)^T\right]}_{(-(\mathbf{1}-\mathbf{K}_1\boldsymbol{C})\boldsymbol{B}\bar{\mathbf{u}})^T} \\
&\quad +\underbrace{E\left[\left(\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\check{\mathbf{e}}_1+\mathbf{K}_1\mathbf{n}_1\right)\right]}_{-(\mathbf{1}-\mathbf{K}_1\boldsymbol{C})\boldsymbol{B}\bar{\mathbf{u}}}(\mathbf{K}_1\bar{\mathbf{y}})^T+(\mathbf{K}_1\bar{\mathbf{y}})(\mathbf{K}_1\bar{\mathbf{y}})^T \\
&= \hat{\mathbf{P}}_1+\left(-\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\boldsymbol{B}\bar{\mathbf{u}}+\mathbf{K}_1\bar{\mathbf{y}}\right)\left(-\left(\mathbf{1}-\mathbf{K}_1\boldsymbol{C}\right)\boldsymbol{B}\bar{\mathbf{u}}+\mathbf{K}_1\bar{\mathbf{y}}\right)^T \quad \text{(12a)}
\end{aligned}$$

- rearranging we again see the covariance is inconsistent:

$$\hat{\mathbf{P}}_1 = E\left[\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^T\right] - \underbrace{E[\hat{\mathbf{e}}_1]\,E[\hat{\mathbf{e}}_1]^T}_{\text{bias effect}} \quad (13)$$

UNIVERSITY OF TORONTO

Fix for known biases

- if we knew the values of the biases, we could just modify the KF to cancel them out:

predictor:

$$\check{\mathbf{P}}_k = \boldsymbol{A}\hat{\mathbf{P}}_{k-1}\boldsymbol{A}^T + \boldsymbol{Q} \tag{14a}$$

$$\check{\mathbf{x}}_k = \boldsymbol{A}\hat{\mathbf{x}}_{k-1} + \boldsymbol{B}\mathbf{u}_k + \underbrace{\boldsymbol{B}\bar{\mathbf{u}}}_{\text{bias}} \tag{14b}$$

Kalman gain:

$$\mathbf{K}_k = \check{\mathbf{P}}_k\boldsymbol{C}^T\left(\boldsymbol{C}\check{\mathbf{P}}_k\boldsymbol{C}^T + \boldsymbol{R}\right)^{-1} \tag{14c}$$

corrector:

$$\hat{\mathbf{P}}_k = \left(\mathbf{1} - \mathbf{K}_k\boldsymbol{C}\right)\check{\mathbf{P}}_k \tag{14d}$$

$$\hat{\mathbf{x}}_k = \check{\mathbf{x}}_k + \mathbf{K}_k\Big(\mathbf{y}_k - \boldsymbol{C}\check{\mathbf{x}}_k - \underbrace{\bar{\mathbf{y}}}_{\text{bias}}\Big) \tag{14e}$$

- this results in an unbiased and consistent filter once again
- unfortunately, we usually do not know the values of the biases and sometimes they change over time

UNIVERSITY OF TORONTO

Input bias estimation

- consider the case where we have just an input bias
- since we are already estimating the state, what if we just augment it with the bias and try to estimate them both together?
- let the augmented state be

$$\mathbf{x}'_k = \begin{bmatrix} \mathbf{x}_k \\ \bar{\mathbf{u}}_k \end{bmatrix} \tag{15}$$

  where we have made the bias now a function of time
- we need to define a motion model for the bias:

$$\bar{\mathbf{u}}_k = \bar{\mathbf{u}}_{k-1} + \mathbf{s}_k \tag{16}$$

  where $\mathbf{s}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{W})$; this is just a random walk

UNIVERSITY OF TORONTO

Input bias estimation

– the motion model for our augmented-state system is now

$$\mathbf{x}'_k = \underbrace{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}}_{\boldsymbol{A}'} \mathbf{x}'_{k-1} + \underbrace{\begin{bmatrix} \boldsymbol{B} \\ \mathbf{0} \end{bmatrix}}_{\boldsymbol{B}'} \mathbf{u}_k + \underbrace{\begin{bmatrix} \mathbf{w}_k \\ \mathbf{s}_k \end{bmatrix}}_{\mathbf{w}'_k} \tag{17}$$

with

$$\mathbf{w}'_k \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{Q}'\right), \quad \boldsymbol{Q}' = \begin{bmatrix} \boldsymbol{Q} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{W} \end{bmatrix} \tag{18}$$

– the observation model is simply

$$\mathbf{y}_k = \underbrace{\begin{bmatrix} \boldsymbol{C} & \mathbf{0} \end{bmatrix}}_{\boldsymbol{C}'} \mathbf{x}'_k + \mathbf{n}_k \tag{19}$$

– we're back to an unbiased system

UNIVERSITY OF TORONTO

Input bias estimation

- this seems too easy – will the augmented state trick really work?
- assume in the case of no bias we have

$$\boldsymbol{Q} > 0, \quad \boldsymbol{R} > 0, \quad \text{rank } \mathcal{O} = N \tag{20}$$

  and so that system has a unique estimate (batch and KF)
- for the augmented system, we are thus required to show

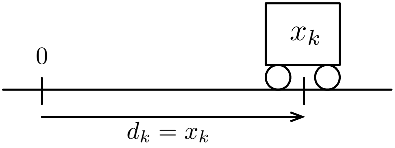
$$\underbrace{\boldsymbol{Q}' > 0, \quad \boldsymbol{R} > 0}_{\text{true by definitions}}, \quad \text{rank } \mathcal{O}' = N + U \tag{21}$$

  with $U = \dim \bar{\mathbf{u}}_k$ and

$$\mathcal{O}' = \begin{bmatrix} \boldsymbol{C}' \\ \boldsymbol{C}'\boldsymbol{A}' \\ \vdots \\ \boldsymbol{C}'\boldsymbol{A}'^{(N+U-1)} \end{bmatrix} \tag{22}$$

UNIVERSITY OF TORONTO

Input bias estimation: good case



$$x_k = x_{k-1} + v_{k-1}$$
$$v_k = v_{k-1} + a_k + \bar{a}$$

interoceptive bias

- this is an example of input bias
- the input is acceleration (i.e., force) but the actual force that gets applied differs from our intended one by an unknown bias

UNIVERSITY OF TORONTO

Input bias estimation: good case

- the system matrices in this case are

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{23}$$

  such that $N = 2$ and $U = 1$

- the unbiased system is observable:

$$\mathcal{O} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \Rightarrow \quad \text{rank } \mathcal{O} = 2 = N \tag{24}$$

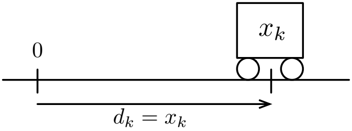
- the augmented-state system is observable, too:

$$\mathcal{O}' = \begin{bmatrix} \boldsymbol{C}' \\ \boldsymbol{C}'\boldsymbol{A}' \\ \boldsymbol{C}'\boldsymbol{A}'^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{C}\boldsymbol{A} & \boldsymbol{C}\boldsymbol{B} \\ \boldsymbol{C}\boldsymbol{A}^2 & \boldsymbol{C}\boldsymbol{A}\boldsymbol{B} + \boldsymbol{C}\boldsymbol{B} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$\Rightarrow \quad \text{rank } \mathcal{O}' = 3 = N + U \tag{25}$$

UNIVERSITY OF TORONTO

Input bias estimation: bad case



$$x_k = x_{k-1} + v_{k-1} + \bar{v}$$
$$v_k = v_{k-1} + a_k + \bar{a}$$

interoceptive biases

- this is also an example of input bias
- this is a strange system wherein the command to the system is a function of both speed and acceleration, and we have unknown biases on both of these quantities

Input bias estimation: bad case

- the system matrices in this case are

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{26}$$

  such that $N = 2$ and $U = 2$

- the unbiased system is observable:

$$\mathcal{O} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \Rightarrow \quad \text{rank } \mathcal{O} = 2 = N \tag{27}$$

- the augmented-state system is unobservable:

$$\mathcal{O}' = \begin{bmatrix} \boldsymbol{C}' \\ \boldsymbol{C}'\boldsymbol{A}' \\ \boldsymbol{C}'\boldsymbol{A}'^2 \\ \boldsymbol{C}'\boldsymbol{A}'^3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{C}\boldsymbol{A} & \boldsymbol{C}\boldsymbol{B} \\ \boldsymbol{C}\boldsymbol{A}^2 & \boldsymbol{C}(\boldsymbol{A}+\boldsymbol{1})\boldsymbol{B} \\ \boldsymbol{C}\boldsymbol{A}^3 & \boldsymbol{C}(\boldsymbol{A}^2+\boldsymbol{A}+\boldsymbol{1})\boldsymbol{B} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 \\ 1 & 3 & 3 & 3 \end{bmatrix}$$

$$\Rightarrow \quad \text{rank } \mathcal{O}' = 3 < 4 = N + U \tag{28}$$

UNIVERSITY OF TORONTO

Measurement bias estimation

- we can try to also use the augmented-state trick when we have a measurement bias
- the augmented state is

$$\mathbf{x}'_k = \begin{bmatrix} \mathbf{x}_k \\ \bar{\mathbf{y}}_k \end{bmatrix} \tag{29}$$

  where we have again made the bias a function of time
- we again assume a random-walk motion model for the bias:

$$\bar{\mathbf{y}}_k = \bar{\mathbf{y}}_{k-1} + \mathbf{s}_k \tag{30}$$

  where $\mathbf{s}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{W})$

UNIVERSITY OF TORONTO

Measurement bias estimation

– the motion model for our augmented-state system is

$$\mathbf{x}'_k = \underbrace{\begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}}_{\boldsymbol{A}'} \mathbf{x}'_{k-1} + \underbrace{\begin{bmatrix} \boldsymbol{B} \\ \mathbf{0} \end{bmatrix}}_{\boldsymbol{B}'} \mathbf{u}_k + \underbrace{\begin{bmatrix} \mathbf{w}_k \\ \mathbf{s}_k \end{bmatrix}}_{\mathbf{w}'_k} \tag{31}$$

with

$$\mathbf{w}'_k \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{Q}'\right), \quad \boldsymbol{Q}' = \begin{bmatrix} \boldsymbol{Q} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{W} \end{bmatrix} \tag{32}$$

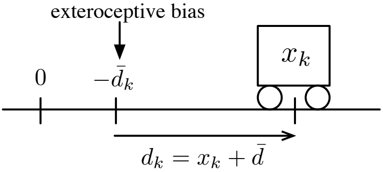
– the observation model is

$$\mathbf{y}_k = \underbrace{\begin{bmatrix} \boldsymbol{C} & \mathbf{1} \end{bmatrix}}_{\boldsymbol{C}'} \mathbf{x}'_k + \mathbf{n}_k \tag{33}$$

– we're again back to an unbiased system

UNIVERSITY OF TORONTO

Measurement bias estimation: bad case



$$x_k = x_{k-1} + v_{k-1}$$
$$v_k = v_{k-1} + a_k$$

- this is also an example of measurement bias
- the measurement we receive is corrupted by an unknown bias
- for the roboticists in the crowd, this is also the world's simplest SLAM problem

Measurement bias estimation: bad case

– the system matrices in this case are

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{34}$$

such that $N = 2$ and $U = 1$

– the unbiased system is observable:

$$\mathcal{O} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \Rightarrow \quad \text{rank } \mathcal{O} = 2 = N \tag{35}$$

– the augmented-state system is unobservable:

$$\mathcal{O}' = \begin{bmatrix} \boldsymbol{C}' \\ \boldsymbol{C}'\boldsymbol{A}' \\ \boldsymbol{C}'\boldsymbol{A}'^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C} & \mathbf{1} \\ \boldsymbol{C}\boldsymbol{A} & \mathbf{1} \\ \boldsymbol{C}\boldsymbol{A}^2 & \mathbf{1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$$

$$\Rightarrow \quad \text{rank } \mathcal{O}' = 2 < 3 = N + U \tag{36}$$

UNIVERSITY OF TORONTO

Measurement bias estimation: bad case

- since we are rank-deficient by 1, this means that $\dim(\text{ null } \mathcal{O}') = 1$; the nullspace of the observability matrix corresponds to those vectors that produce outputs of zero
- here we see that

$$\text{null } \mathcal{O}' = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\} \tag{37}$$

  which means that we can shift the cart and landmark together (left or right) and the measurement will not change
- does this mean our estimator will fail? not if we interpret the solution properly
  - batch: we have infinitely many solutions, not one unique one
  - KF: the bias will stay at the initial guess and not change
- roboticists: this shows why SLAM is unobservable (in general)

UNIVERSITY OF TORONTO

Data association

- data association (a.k.a., finding correspondences) has to do with figuring out to which parts of a model measurements correspond
- example: positioning using GPS satellites
  - the positions of the satellites are assumed to be known (as a function of time)
  - a receiver on the ground measures range to the satellites (e.g., using time of flight)
  - it is easy to know which range measurement is associated with which satellite
- example: attitude determination using celestial observation
  - a map (or chart) of all the brightest stars in the sky is preconstructed
  - live observations are matched to the map
  - knowing which star you are looking at, is much more difficult than the GPS case
- external vs. internal data association

UNIVERSITY OF TORONTO

External data association

- in external data association, specialized knowledge of the model/measurements is used for association
- this knowledge is external to the estimation problem
- this is sometimes called known data association because from the perspective of the estimation problem, the job has been done
- examples:
  - GPS satellites with encoded identification
  - targets with unique, known colours
  - visual bar codes

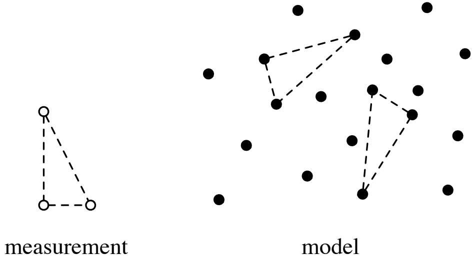
UNIVERSITY OF TORONTO

Internal data association

- in internal data association, only the measurements/model are used to do data association
- this is sometimes called unknown data association
- typically, association is based on the likelihood of a given measurement, given the model
- in the simplest version, the most likely association is accepted and the other possibilities are ignored
- more sophisticated techniques allow multiple data association hypotheses to be carried forward into the estimation problem
- example: imagine a bunch of identical landmarks and we get a range/bearing measurement – which one did it come from?

UNIVERSITY OF TORONTO

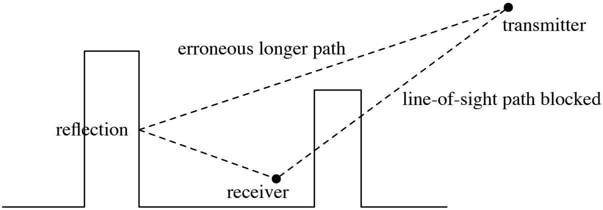
Internal data association



- sometimes constellations of landmarks are used to help perform data association
- the data-aligned rigidity-constrained exhaustive search (DARCES) algorithm is an example
- the idea is that the distances between pairs of points in a constellation can be used as a type of unique identifier for data association

UNIVERSITY OF TORONTO

Outliers



- outlier measurements we define to be those that are highly improbable according to our noise model
- example: multipath reflections in GPS - we receive a range measurement and assume it came from a line-of-sight observation even though it might have reflected
- example: misassociations during the data association step can lead to outliers

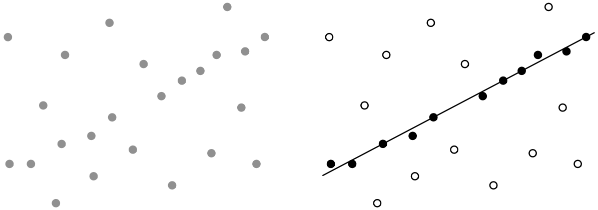
UNIVERSITY OF TORONTO

Outlier rejection

- if we accept that a portion (possibly large) of our measurements could be outliers, we need to devise a means to detect and reduce/remove the influence of outliers on our estimators
- we will discuss the two most common techniques to handle outliers:
  (i) random sample consensus (RANSAC)
  (ii) M-estimation (a.k.a., robust estimation)
- these can be used separately or in tandem

UNIVERSITY OF TORONTO

RANSAC



example dataset with inliers and outliers RANSAC finds line with most inliers

- imagine fitting a line to some data using least squares
- some of the data points clearly were generated by a line model (inliers) and some were not (outliers)
- RANSAC can be used to identify which points are which
- is a probabilistic algorithm in the sense that its ability to find a reasonable answer can only be guaranteed to occur with a probability that improves with more time spent in the search

UNIVERSITY OF TORONTO

RANSAC

- RANSAC proceeds in an iterative manner:
  1. select a (small) random subset of the original data to be hypothesized inliers (e.g., pick two points if fitting a line to xy-data)
  2. fit a model to the hypothesized inliers (e.g., a line is fit to two points)
  3. test the rest of the original data against the fitted model and classify as either inliers or outliers
  4. refit the model using both the hypothesized and classified inliers
  5. evaluate the refit model in terms of the residual error of all the inlier data
- this is repeated for a large number of iterations and the hypothesis with the lowest residual error is selected as the best
- there are many variants to the plain-vanilla algorithm presented

RANSAC

- how many iterations, k, are needed to ensure a subset is selected comprised solely of inliers, with probability p?
- if we assume that each measurement is selected independently, and each has probability w of being an inlier, then the following relation holds:

$$1 - p = (1 - w^n)^k \tag{38}$$

  where n is the number of data points in the random subset and k is the number of iterations
- solving for k gives

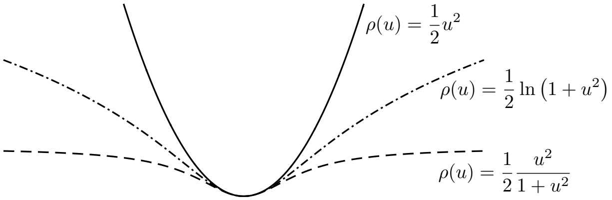
$$k = \frac{\ln(1 - p)}{\ln(1 - w^n)} \tag{39}$$

- in reality this can be thought of as an upper bound as the data points are typically selected sequentially, not independently
- there can also be constraints between the data points that complicate the selection of random subsets

UNIVERSITY OF TORONTO

M-estimation

- many of our earlier estimation techniques were shown to be minimizing a sum-of-squared-error cost function, which can be highly sensitive to outliers
- a single large outlier can exercise a huge influence on the estimate because it dominates the quadratic cost
- M-estimation (a.k.a., robust estimation) modifies the shape of the cost function from quadratic so that outliers do not dominate the solution



M-estimation

– recall the batch-MAP objective function:

$$J(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{N} \mathbf{e}_i(\mathbf{x})^T \mathbf{W}_i^{-1} \mathbf{e}_i(\mathbf{x}) \tag{40}$$

– we can generalize this objective function to be

$$J'(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \, \rho \left(u_i(\mathbf{x}) \right) \tag{41}$$

where $\alpha_i > 0$ is a scalar weight and

$$u_i(\mathbf{x}) = \sqrt{\mathbf{e}_i(\mathbf{x})^T \mathbf{W}_i^{-1} \mathbf{e}_i(\mathbf{x})} \tag{42}$$

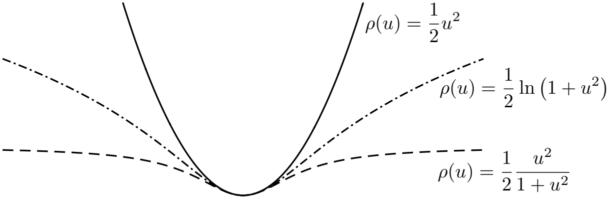
and $\rho(\cdot)$ is some nonlinear robust cost function

UNIVERSITY OF TORONTO

M-estimation

- robust cost functions should be bounded, positive definite, have a unique zero at zero input, and increase more slowly than squared
- this means that large errors will not carry as much weight and have little power on the solution due to a reduced gradient
- there are many possible cost functions including

$$\underbrace{\rho(u) = \frac{1}{2}u^2}_{\text{quadratic}}, \quad \underbrace{\rho(u) = \frac{1}{2}\ln\left(1+u^2\right)}_{\text{Cauchy}}, \quad \underbrace{\rho(u) = \frac{1}{2}\frac{u^2}{1+u^2}}_{\text{Geman-McClure}} \tag{43}$$



UNIVERSITY OF TORONTO

M-estimation

- at a minimum, the gradient of the objective function should be zero
- consider the original (non-robust) objective function:

$$J(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{N} \mathbf{e}_i(\mathbf{x})^T \mathbf{W}_i^{-1} \mathbf{e}_i(\mathbf{x}), \tag{44}$$

- the gradient of this objective function is

$$\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = \sum_{i=1}^{N} \mathbf{e}_i(\mathbf{x})^T \mathbf{W}_i^{-1} \frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial \mathbf{x}} \tag{45}$$

M-estimation

- using the chain rule, the gradient of our new objective function is

$$\frac{\partial J'(\mathbf{x})}{\partial \mathbf{x}} = \sum_{i=1}^{N} \alpha_i \frac{\partial \rho}{\partial u_i} \frac{\partial u_i}{\partial \mathbf{e}_i} \frac{\partial \mathbf{e}_i}{\partial \mathbf{x}} \tag{46}$$

- substituting

$$\frac{\partial u_i}{\partial \mathbf{e}_i} = \frac{1}{u_i(\mathbf{x})} \mathbf{e}_i(\mathbf{x})^T \mathbf{W}_i^{-1} \tag{47}$$

the gradient can be written as

$$\frac{\partial J'(\mathbf{x})}{\partial \mathbf{x}} = \sum_{i=1}^{N} \mathbf{e}_i(\mathbf{x})^T \mathbf{Y}_i(\mathbf{x})^{-1} \frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial \mathbf{x}} \tag{48}$$

where we have a new (inverse) covariance that depends on $\mathbf{x}$:

$$\mathbf{Y}_i(\mathbf{x})^{-1} = \frac{\alpha_i}{u_i(\mathbf{x})} \left. \frac{\partial \rho}{\partial u_i} \right|_{u_i(\mathbf{x})} \mathbf{W}_i^{-1} \tag{49}$$

M-estimation

- we are already using an iterative solver to minimize the objective function, due to the nonlinear dependence of $\mathbf{e}_i(\mathbf{x})$ on $\mathbf{x}$
- it therefore makes sense to evaluate $\mathbf{Y}_i(\mathbf{x})$ at the value of the state from the previous iteration, $\mathbf{x}_{\text{op}}$
- this means we can simply work with the cost function,

$$J''(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N} \mathbf{e}_i(\mathbf{x})^T \mathbf{Y}_i(\mathbf{x}_{\text{op}})^{-1} \mathbf{e}_i(\mathbf{x}) \tag{50}$$

where

$$\mathbf{Y}_i(\mathbf{x}_{\text{op}})^{-1} = \frac{\alpha_i}{u_i(\mathbf{x}_{\text{op}})} \left.\frac{\partial \rho}{\partial u_i}\right|_{u_i(\mathbf{x}_{\text{op}})} \mathbf{W}_i^{-1} \tag{51}$$

- at each iteration we solve the original least-squares problem, but with a modified covariance matrix that updates as $\mathbf{x}_{\text{op}}$ updates
- this is referred to as iteratively reweighted least squares

UNIVERSITY OF TORONTO

M-estimation

– to see why this iterative scheme works, we can examine the gradient of $J''(\mathbf{x})$:

$$\frac{\partial J''(\mathbf{x})}{\partial \mathbf{x}} = \sum_{i=1}^{N} \mathbf{e}_i(\mathbf{x})^T \mathbf{Y}_i(\mathbf{x}_{\text{op}})^{-1} \frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial \mathbf{x}} \tag{52}$$

– if the iterative scheme converges, we will have $\hat{\mathbf{x}} = \mathbf{x}_{\text{op}}$, so

$$\left.\frac{\partial J'(\mathbf{x})}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}} = \left.\frac{\partial J''(\mathbf{x})}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}} = \mathbf{0} \tag{53}$$

and thus the two systems will have the same minimum (or minima)

– however, the path taken to get to the optimum will differ if we minimize $J''(\mathbf{x})$ rather than $J'(\mathbf{x})$

M-estimation

– with the Cauchy robust cost, the objective function is

$$J'(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{N} \alpha_i \ln \left(1 + \mathbf{e}_i(\mathbf{x})^T \mathbf{W}_i^{-1} \mathbf{e}_i(\mathbf{x})\right) \tag{54}$$

– the robust (inverse) covariance matrix is

$$\mathbf{Y}_i(\mathbf{x}_{\text{op}})^{-1} = \frac{\alpha_i}{u_i(\mathbf{x}_{\text{op}})} \left. \frac{\partial \rho}{\partial u_i} \right|_{u_i(\mathbf{x}_{\text{op}})} \mathbf{W}_i^{-1} = \frac{\alpha_i}{u_i(\mathbf{x}_{\text{op}})} \frac{u_i(\mathbf{x}_{\text{op}})}{1 + u_i(\mathbf{x}_{\text{op}})^2} \mathbf{W}_i^{-1} \tag{55}$$

– inverting, the robust covariance matrix is

$$\mathbf{Y}_i(\mathbf{x}_{\text{op}}) = \frac{1}{\alpha_i} \left(1 + \mathbf{e}_i(\mathbf{x}_{\text{op}})^T \mathbf{W}_i^{-1} \mathbf{e}_i(\mathbf{x}_{\text{op}})\right) \mathbf{W}_i \tag{56}$$

which we see is just an inflated version of $\mathbf{W}_i$

UNIVERSITY OF TORONTO

Summary

- we looked a number of situations that violate some of the assumptions made when formulating our estimation algorithms
- we saw that in the case of input and/or measurement biases, we could try folding the bias into the estimation problem – in some cases this can work (observable) and in others it cannot (unobservable)
- we briefly discussed the idea of data association, which could be done externally to our estimation algorithm or within it
- we acknowledged that sometimes we will have outliers: measurements that are quite improbable according to our noise models
- both RANSAC and M-estimation were presented as means of dealing with outliers – these can be used together or individually

UNIVERSITY OF TORONTO