Ocean Uncertainty Quantification Summer School

Intro to Data Assimilation

Sarah Williamson

July 16, 2024

Outline

- Introduction and notation
- One-dimensional linear problem: a simple example
    - Extension to multi-dimensional problems
- Data assimilation techniques
    - Sequential methods
    - Variational (smoother) methods

What is data assimilation?

Data assimilation aims to combine available information (data) with a numerical model while taking into account any possible uncertainties

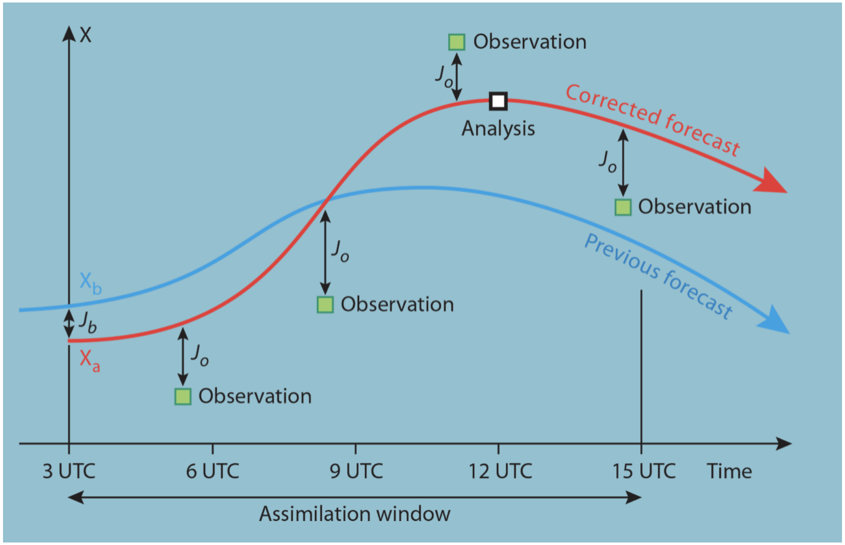
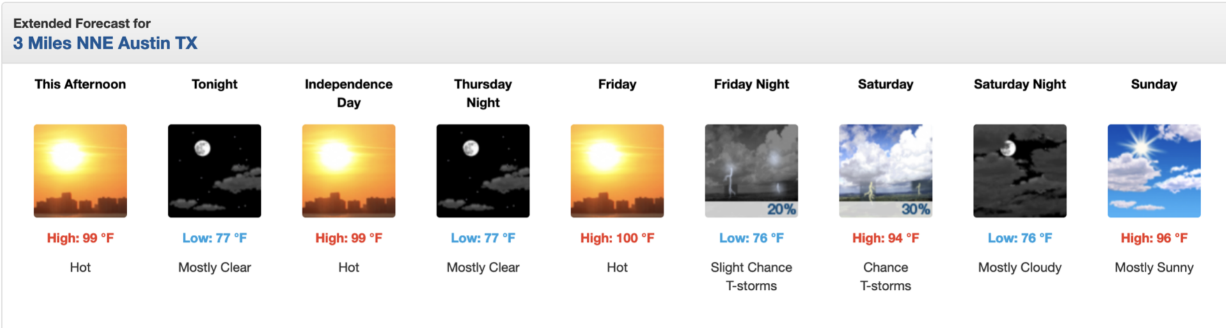
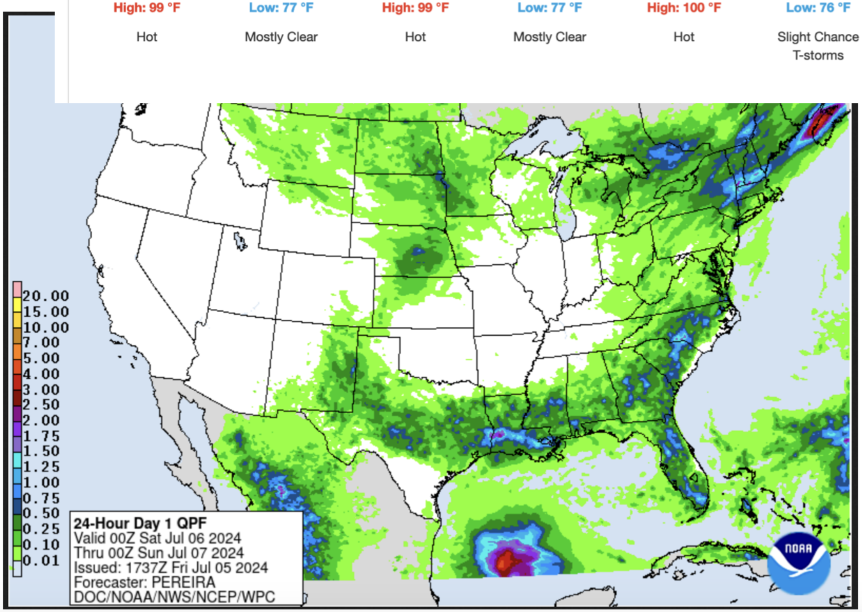


Figure 1: Courtesy of Aneesh Subramanian

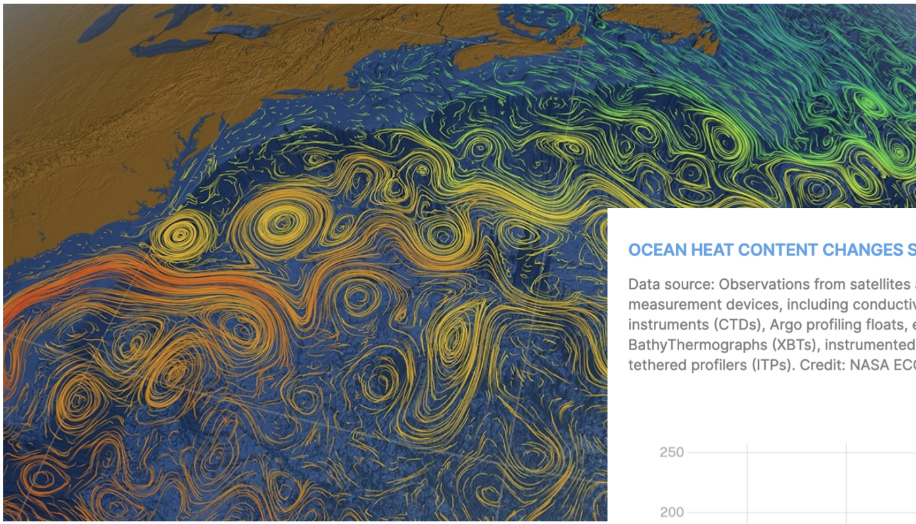
What can one do with data assimilation?

Weather forecasting

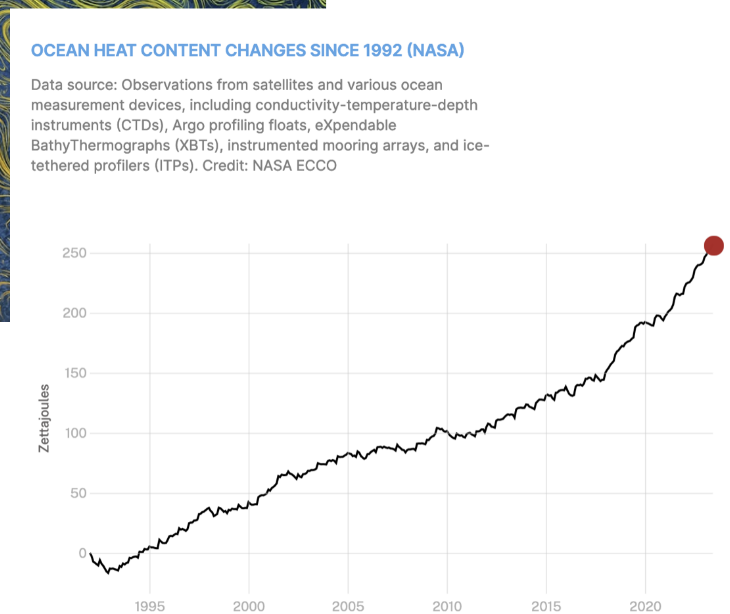




Improved ocean state reconstruction

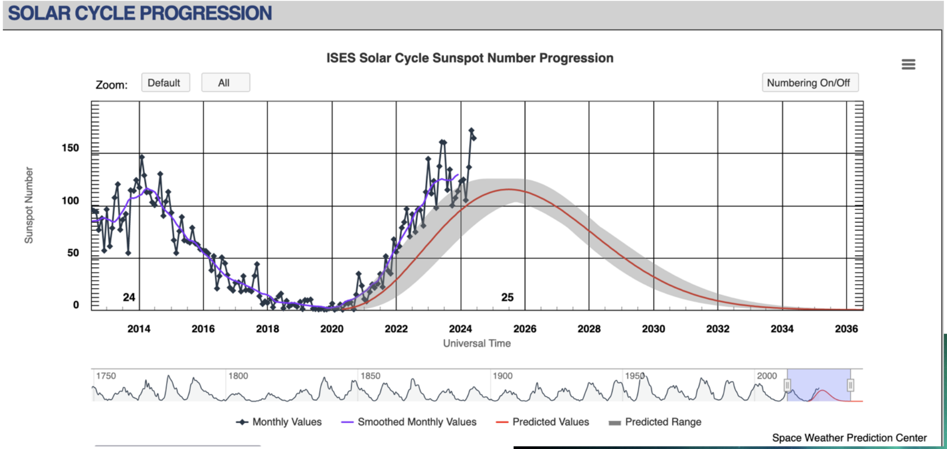


Above: NASA ocean visualization, from https://www.whoi.edu/know-your-ocean/ocean-topics/how-the-ocean-works/ocean-circulation/currents-gyres-eddies/.

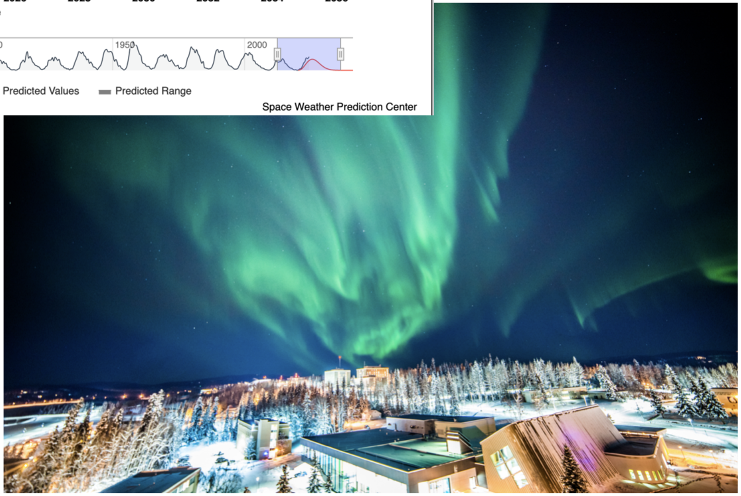


Right: NASA ECCO heat content reconstruction, from https://climate.nasa.gov/vital-signs/ocean-warming/?intent=121

Sunspot cycle predictions



Sunspot prediction from https://www.swpc.noaa.gov/products/solar-cycle-progression



Aurora photo from https://www.uaf.edu/news/archives/news-archives-2010-2021/featured-photo-week-101.php

Three key pieces to data assimilation:

Data,

Three key pieces to data assimilation:

Data, numerical model,

Three key pieces to data assimilation:

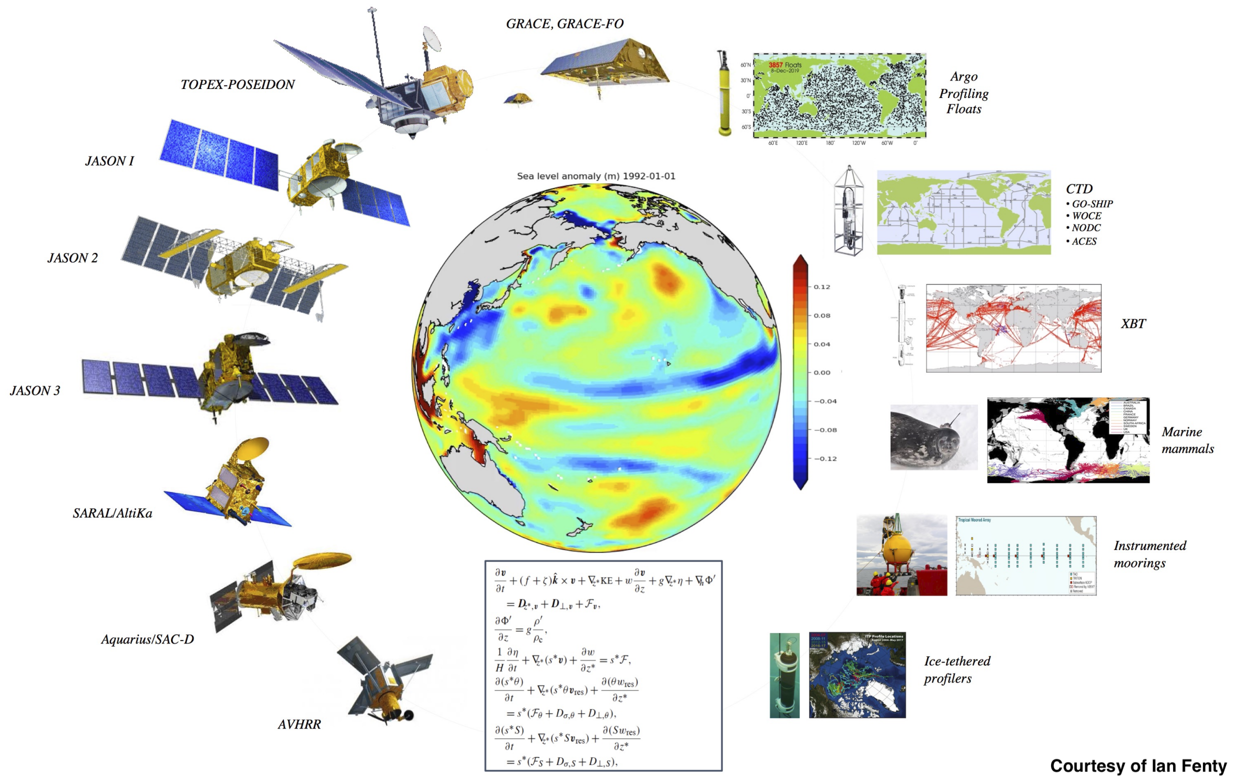
Data, numerical model, uncertainties

Ingredient 1: Data

Where do we get data?

From NASA JPL: https://www.jpl.nasa.gov/news/seal-takes-ocean-heat-transport-data-to-new-depths

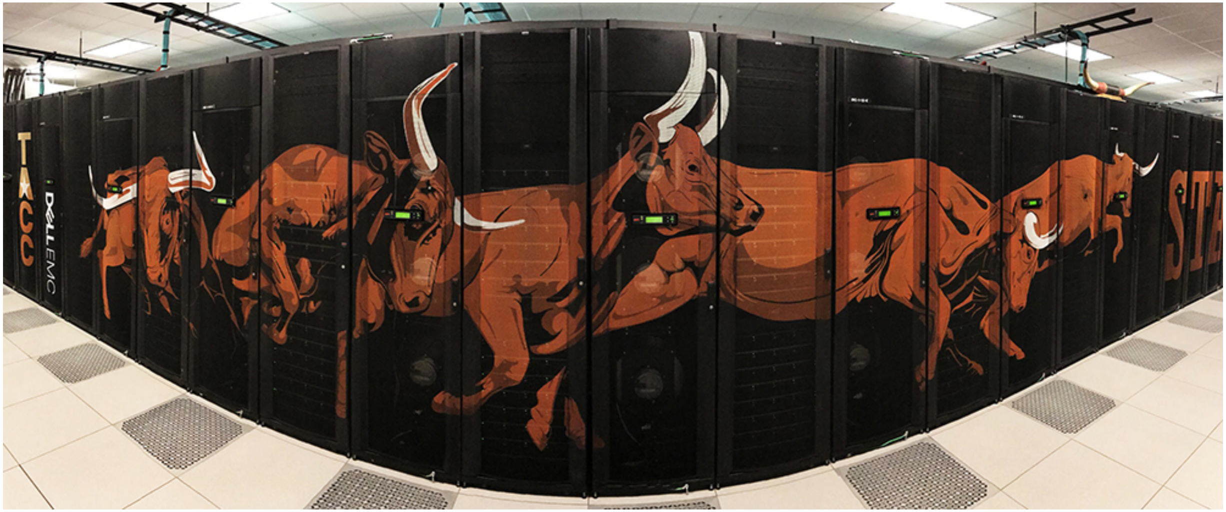
Seals are just one example...



Courtesy of Ian Fenty

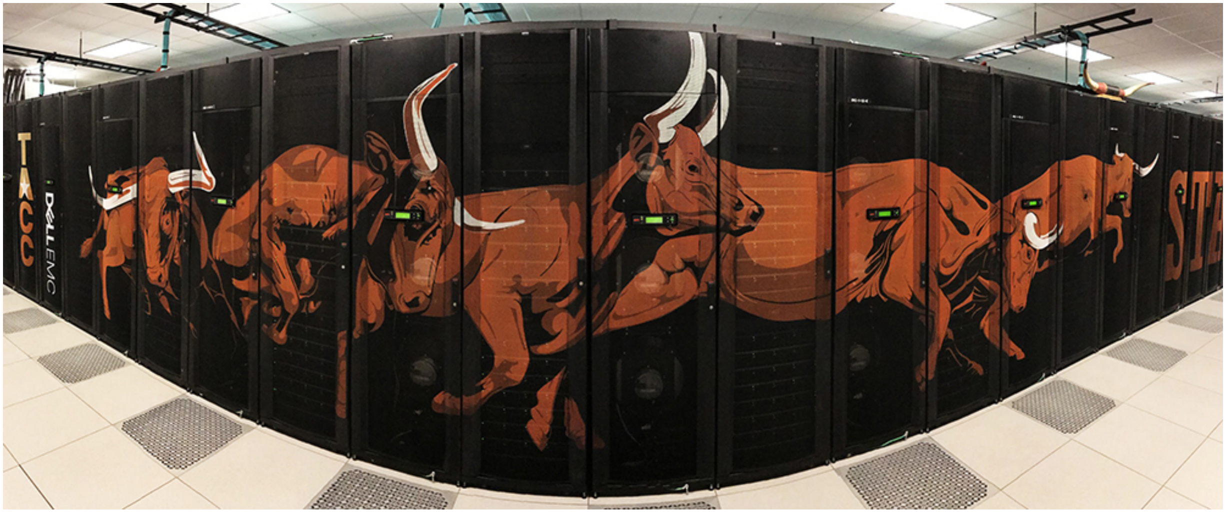
Ingredient 2: Numerical model

Numerical models



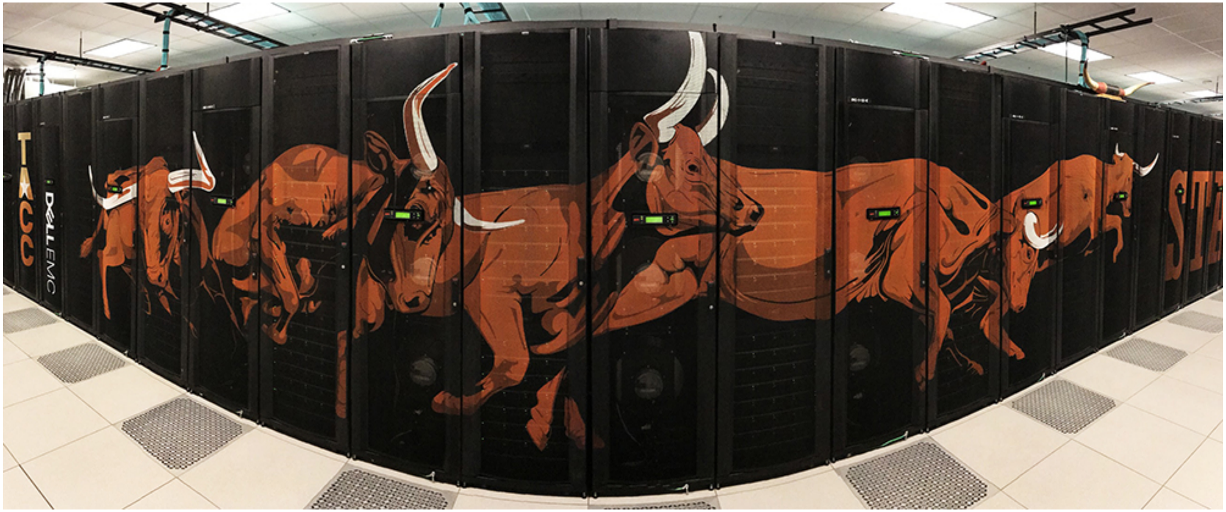
- The numerical model represents a time-evolving state via the equations of motion, conservation laws, etc.

Numerical models



- The numerical model represents a time-evolving state via the equations of motion, conservation laws, etc.
- State consists of coupled variables

Numerical models



- The numerical model represents a time-evolving state via the equations of motion, conservation laws, etc.
- State consists of coupled variables
- For the ocean, these can be,
    - Temperature, salinity, velocities...

Ingredient 3: Uncertainties

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

- Data sparsity, either spatially (low density of data), temporally (infrequent data) or both (usually)

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

- Data sparsity, either spatially (low density of data), temporally (infrequent data) or both (usually)
- Incomplete data (e.g., data on temperature, but no data on salinity)

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

- Data sparsity, either spatially (low density of data), temporally (infrequent data) or both (usually)
- Incomplete data (e.g., data on temperature, but no data on salinity)
- Measurement error (e.g. noisy data)

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

- Data sparsity, either spatially (low density of data), temporally (infrequent data) or both (usually)
- Incomplete data (e.g., data on temperature, but no data on salinity)
- Measurement error (e.g. noisy data)

Sources of uncertainty in models:

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

- Data sparsity, either spatially (low density of data), temporally (infrequent data) or both (usually)
- Incomplete data (e.g., data on temperature, but no data on salinity)
- Measurement error (e.g. noisy data)

Sources of uncertainty in models:

- Model discretization (discrete solution is only ever an approximation)

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

- Data sparsity, either spatially (low density of data), temporally (infrequent data) or both (usually)
- Incomplete data (e.g., data on temperature, but no data on salinity)
- Measurement error (e.g. noisy data)

Sources of uncertainty in models:

- Model discretization (discrete solution is only ever an approximation)
- Boundary conditions (e.g. ice sheet bottom topography)

Alas, nothing is perfect...

We'll (almost) always have uncertainties in both the data and model

Sources of uncertainty in data:

- Data sparsity, either spatially (low density of data), temporally (infrequent data) or both (usually)
- Incomplete data (e.g., data on temperature, but no data on salinity)
- Measurement error (e.g. noisy data)

Sources of uncertainty in models:

- Model discretization (discrete solution is only ever an approximation)
- Boundary conditions (e.g. ice sheet bottom topography)
- Model parameters

What can we do?

Importantly: data assimilation *takes into account* the model and data uncertainties!

What can we do?

Importantly: data assimilation *takes into account* the model and data uncertainties!

The best model estimate will be one that's computed with statistical info on sources of uncertainties

Notation and concepts

Variables we'll use

Let's define:

- $\boldsymbol{x}$ - model state
- $\tilde{\boldsymbol{x}}$ - model prediction
- $\boldsymbol{y}$ - data/observation
- $\boldsymbol{n}$ - observational error

Variables we'll use

Let's define:

- $\boldsymbol{x}$ - model state
- $\tilde{\boldsymbol{x}}$ - model prediction
- $\boldsymbol{y}$ - data/observation
- $\boldsymbol{n}$ - observational error

Generally, these variables are multi-dimensional, but we'll also (briefly) look at a one-dimensional problem

A couple definitions from statistics

Let $p(x)$ denote a probability distribution, $y = x + n$ an observation, and let $E(.)$ denote the *expectation value*

A couple definitions from statistics

Let $p(x)$ denote a probability distribution, $y = x + n$ an observation, and let $E(.)$ denote the *expectation value*

Expectation value

$E(n)$ is the expected value of n; this is the average value n will be given p

A couple definitions from statistics

Let $p(x)$ denote a probability distribution, $y = x + n$ an observation, and let $E(.)$ denote the *expectation value*

Expectation value

$E(n)$ is the expected value of n; this is the average value n will be given p

Variance

$E\left((n - E(n))^2\right)$ is the variance of n; this is the amount n will vary around the mean

Example: Least-squares linear estimation

Goal:

Estimate unknown quantity x given observations

$$y_1 = x + n_1, \quad y_2 = x + n_2$$

with errors n_1, n_2

Example: Least-squares linear estimation

Goal:

Estimate unknown quantity x given observations

$$y_1 = x + n_1, \quad y_2 = x + n_2$$

with errors n_1, n_2

Assumptions:

$$E(n_1) = E(n_2) = 0 \implies \text{Unbiased observations}$$
$$E(n_1^2) = \sigma_1^2, \quad E(n_2^2) = \sigma_2^2$$
$$E(n_1 n_2) = 0 \implies \text{Observation errors are uncorrelated}$$

Example: Least-squares linear estimation

A linear approximation would be

$$\tilde{x} = a_1 y_1 + a_2 y_2$$

for (currently) unknown weights a_1, a_2.

Example: Least-squares linear estimation

A linear approximation would be

$$\tilde{x} = a_1 y_1 + a_2 y_2$$

for (currently) unknown weights a_1, a_2.

To figure out the weights, we'll minimize a model-data misfit:

$$J(x) = \frac{1}{\sigma_1^2}(x - y_1)^2 + \frac{1}{\sigma_2^2}(x - y_2)^2$$

Minimizing J

We begin by taking a derivative:

$$\frac{\partial J}{\partial x} = \frac{1}{\sigma_1^2}\frac{\partial}{\partial x}(x - y_1)^2 + \frac{1}{\sigma_2^2}\frac{\partial}{\partial x}(x - y_2)^2$$

Minimizing J

We begin by taking a derivative:

$$\begin{aligned}\frac{\partial J}{\partial x} &= \frac{1}{\sigma_1^2}\frac{\partial}{\partial x}(x - y_1)^2 + \frac{1}{\sigma_2^2}\frac{\partial}{\partial x}(x - y_2)^2 \\ &= \frac{2}{\sigma_1^2}(x - y_1) + \frac{2}{\sigma_2^2}(x - y_2)\end{aligned}$$

Minimizing J

We begin by taking a derivative:

$$\begin{aligned}\frac{\partial J}{\partial x} &= \frac{1}{\sigma_1^2}\frac{\partial}{\partial x}(x - y_1)^2 + \frac{1}{\sigma_2^2}\frac{\partial}{\partial x}(x - y_2)^2 \\ &= \frac{2}{\sigma_1^2}(x - y_1) + \frac{2}{\sigma_2^2}(x - y_2)\end{aligned}$$

Next, we set it equal to zero and solve for x:

$$\implies x = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y_2$$

Minimizing J

We begin by taking a derivative:

$$\begin{aligned}\frac{\partial J}{\partial x} &= \frac{1}{\sigma_1^2}\frac{\partial}{\partial x}(x-y_1)^2 + \frac{1}{\sigma_2^2}\frac{\partial}{\partial x}(x-y_2)^2 \\ &= \frac{2}{\sigma_1^2}(x-y_1) + \frac{2}{\sigma_2^2}(x-y_2)\end{aligned}$$

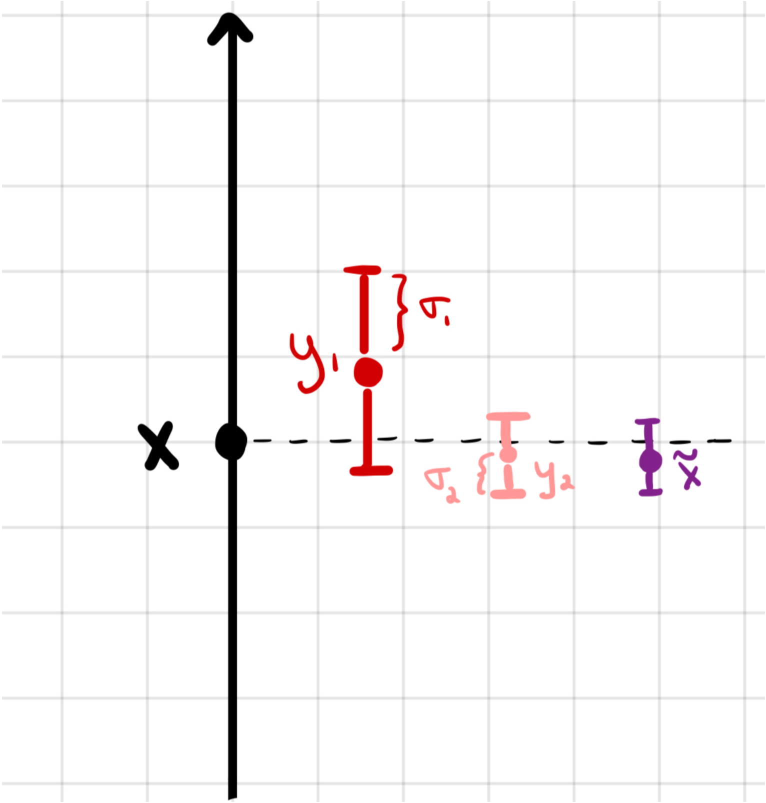
Next, we set it equal to zero and solve for x:

$$\implies x = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}y_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}y_2$$

This tells us precisely how to weight the data points:

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}, \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$$

A sketch of this estimation



Comment on this estimate

One can find $\tilde{x}$ through a different minimization

Comment on this estimate

One can find $\tilde{x}$ through a different minimization

- In addition to the assumptions from before, add that $\tilde{x}$ be an unbiased estimate (e.g. $E(\tilde{x}) = E(x)$)

Comment on this estimate

One can find $\tilde{x}$ through a different minimization

- In addition to the assumptions from before, add that $\tilde{x}$ be an unbiased estimate (e.g. $E(\tilde{x}) = E(x)$)
- Minimize

$$\sigma^2 = E((\tilde{x} - x)^2)$$

Comment on this estimate

One can find $\tilde{x}$ through a different minimization

- In addition to the assumptions from before, add that $\tilde{x}$ be an unbiased estimate (e.g. $E(\tilde{x}) = E(x)$)
- Minimize
$$\sigma^2 = E((\tilde{x} - x)^2)$$
- This means $\tilde{x}$ is the estimate that minimizes uncertainty

Extension to multi-dimensions

All of the above can be shifted to multi-dimensions:

Extension to multi-dimensions

All of the above can be shifted to multi-dimensions:

$$\boldsymbol{y} = \boldsymbol{E}\boldsymbol{x} + \boldsymbol{n}$$

$$E(\boldsymbol{n}) = 0$$

$$E(\boldsymbol{n}\boldsymbol{n}^T) = \boldsymbol{R} \text{ (Covariance matrix!)}$$

Extension to multi-dimensions

All of the above can be shifted to multi-dimensions:

$\boldsymbol{y} = \boldsymbol{E}\boldsymbol{x} + \boldsymbol{n}$

$E(\boldsymbol{n}) = 0$

$E(\boldsymbol{n}\boldsymbol{n}^T) = \boldsymbol{R}$ (Covariance matrix!)

Then we want the estimate $\tilde{\boldsymbol{x}}$ to minimize the loss:

$$J(\boldsymbol{x}) = (\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{R}^{-1} (\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})$$

Extension to multi-dimensions

Minimizing J yields $\tilde{\boldsymbol{x}}$:

$$\frac{\partial J}{\partial \boldsymbol{x}} =$$

Extension to multi-dimensions

Minimizing J yields $\tilde{\boldsymbol{x}}$:

$$\frac{\partial J}{\partial \boldsymbol{x}} = \frac{\partial}{\partial x}\left[\boldsymbol{x}^T\boldsymbol{E}^T - \boldsymbol{y}^T\right]\boldsymbol{R}^{-1}(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y}) + (\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})^T\frac{\partial}{\partial x}\left[\boldsymbol{R}^{-1}(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})\right]$$

Extension to multi-dimensions

Minimizing J yields $\tilde{\boldsymbol{x}}$:

$$\frac{\partial J}{\partial \boldsymbol{x}} = \frac{\partial}{\partial x}\left[\boldsymbol{x}^T \boldsymbol{E}^T - \boldsymbol{y}^T\right] \boldsymbol{R}^{-1}(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y}) + (\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})^T \frac{\partial}{\partial x}\left[\boldsymbol{R}^{-1}(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})\right]$$
$$= 2(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{R}^{-1} \boldsymbol{E}$$

Extension to multi-dimensions

Minimizing J yields $\tilde{\boldsymbol{x}}$:

$$\begin{aligned}\frac{\partial J}{\partial \boldsymbol{x}} &= \frac{\partial}{\partial x}\left[\boldsymbol{x}^T\boldsymbol{E}^T - \boldsymbol{y}^T\right]\boldsymbol{R}^{-1}(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y}) + (\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})^T\frac{\partial}{\partial x}\left[\boldsymbol{R}^{-1}(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})\right] \\ &= 2(\boldsymbol{E}\boldsymbol{x} - \boldsymbol{y})^T\boldsymbol{R}^{-1}\boldsymbol{E}\end{aligned}$$

Setting this equal to zero and solving for $\boldsymbol{x}$ gives:

$$\tilde{\boldsymbol{x}} = (\boldsymbol{E}^T\boldsymbol{R}^{-1}\boldsymbol{E})^{-1}\boldsymbol{E}^T\boldsymbol{R}^{-1}\boldsymbol{y}$$

The estimate $\tilde{x}$

The estimate

$$\tilde{\boldsymbol{x}} = (\boldsymbol{E}^T \boldsymbol{R}^{-1} \boldsymbol{E})^{-1} \boldsymbol{E}^T \boldsymbol{R}^{-1} \boldsymbol{y}$$

we found is sometimes referred to as the BLUE

The estimate $\tilde{\boldsymbol{x}}$

The estimate

$$\tilde{\boldsymbol{x}} = (\boldsymbol{E}^T\boldsymbol{R}^{-1}\boldsymbol{E})^{-1}\boldsymbol{E}^T\boldsymbol{R}^{-1}\boldsymbol{y}$$

we found is sometimes referred to as the BLUE

This stands for the best linear unbiased estimate of $\boldsymbol{x}$ from $\boldsymbol{y}$

The estimate $\tilde{x}$

The estimate

$$\tilde{\boldsymbol{x}} = (\boldsymbol{E}^T\boldsymbol{R}^{-1}\boldsymbol{E})^{-1}\boldsymbol{E}^T\boldsymbol{R}^{-1}\boldsymbol{y}$$

we found is sometimes referred to as the BLUE

This stands for the best linear unbiased estimate of $\boldsymbol{x}$ from $\boldsymbol{y}$

The corresponding uncertainty is

$$\boldsymbol{P} = E\big((\tilde{\boldsymbol{x}} - \boldsymbol{x})(\tilde{\boldsymbol{x}} - \boldsymbol{x})^T\big)$$

The estimate $\tilde{x}$

The estimate

$$\tilde{\boldsymbol{x}} = (\boldsymbol{E}^T \boldsymbol{R}^{-1} \boldsymbol{E})^{-1} \boldsymbol{E}^T \boldsymbol{R}^{-1} \boldsymbol{y}$$

we found is sometimes referred to as the BLUE

This stands for the best linear unbiased estimate of $\boldsymbol{x}$ from $\boldsymbol{y}$

The corresponding uncertainty is

$$\boldsymbol{P} = E\left((\tilde{\boldsymbol{x}} - \boldsymbol{x})(\tilde{\boldsymbol{x}} - \boldsymbol{x})^T\right) = (\boldsymbol{E}^T \boldsymbol{R}^{-1} \boldsymbol{E})^{-1}$$

Sequential methods: Kalman filtering

Note: moving to time-dependent notation

We'll let $t \in [0, t_f]$, Δt a timestep. We have:

- $\boldsymbol{x}(t + \Delta t) = \boldsymbol{A}\boldsymbol{x}(t)$

Note: moving to time-dependent notation

We'll let $t \in [0, t_f]$, Δt a timestep. We have:

- $\boldsymbol{x}(t + \Delta t) = \boldsymbol{A}\boldsymbol{x}(t)$
- Observations are now time-dependent $\boldsymbol{y}(t)$

Note: moving to time-dependent notation

We'll let $t \in [0, t_f]$, Δt a timestep. We have:

- $\boldsymbol{x}(t + \Delta t) = \boldsymbol{A}\boldsymbol{x}(t)$
- Observations are now time-dependent $\boldsymbol{y}(t)$
- In general we'll just add a (t)

Kalman filtering

Suppose we have,

- A background estimate $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) + \boldsymbol{\varepsilon}(t)$ with error $\varepsilon(t)$,

Kalman filtering

Suppose we have,

- A background estimate $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) + \boldsymbol{\varepsilon}(t)$ with error $\varepsilon(t)$,
- corresponding covariance matrix $\boldsymbol{P}(t) = E(\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(t)^T)$ and,

Kalman filtering

Suppose we have,

- A background estimate $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) + \boldsymbol{\varepsilon}(t)$ with error $\varepsilon(t)$,
- corresponding covariance matrix $\boldsymbol{P}(t) = E(\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(t)^T)$ and,
- data/observation $\boldsymbol{y}(t + \Delta t) = \boldsymbol{E}\boldsymbol{x}(t + \Delta t) + \boldsymbol{n}(t + \Delta t)$

Kalman filtering

Suppose we have,

- A background estimate $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) + \boldsymbol{\varepsilon}(t)$ with error $\varepsilon(t)$,
- corresponding covariance matrix $\boldsymbol{P}(t) = E(\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(t)^T)$ and,
- data/observation $\boldsymbol{y}(t + \Delta t) = \boldsymbol{E}\boldsymbol{x}(t + \Delta t) + \boldsymbol{n}(t + \Delta t)$

Kalman filtering aims to produce the BLUE $\tilde{\boldsymbol{x}}(t + \Delta t)$ given both $\tilde{\boldsymbol{x}}(t)$ and $\boldsymbol{y}(t + \Delta t)$

Kalman filtering

Step 1: an intermediate prediction,

$$\tilde{\boldsymbol{x}}(t+\Delta t,-) = \boldsymbol{A}\tilde{\boldsymbol{x}}(t)$$

where "$-$" indicates no data has been incorporated at time t

Kalman filtering

Step 1: an intermediate prediction,

$$\tilde{\boldsymbol{x}}(t+\Delta t,-)=\boldsymbol{A}\tilde{\boldsymbol{x}}(t)$$

where "$-$" indicates no data has been incorporated at time t
Step 2: Computing the uncertainty $\boldsymbol{P}(t+\Delta t,-)$:

$$\boldsymbol{P}(t+\Delta t,-)=E\left((\boldsymbol{A}\tilde{x}(t)-\boldsymbol{A}\boldsymbol{x}(t))(\boldsymbol{A}\tilde{x}(t)-\boldsymbol{A}\boldsymbol{x}(t))^T\right)$$

Kalman filtering

Step 1: an intermediate prediction,

$$\tilde{\boldsymbol{x}}(t + \Delta t, -) = \boldsymbol{A}\tilde{\boldsymbol{x}}(t)$$

where "$-$" indicates no data has been incorporated at time t
Step 2: Computing the uncertainty $\boldsymbol{P}(t + \Delta t, -)$:

$$\begin{aligned}\boldsymbol{P}(t + \Delta t, -) &= E\left((\boldsymbol{A}\tilde{x}(t) - \boldsymbol{A}\boldsymbol{x}(t))(\boldsymbol{A}\tilde{x}(t) - \boldsymbol{A}\boldsymbol{x}(t))^T\right) \\ &= E\left(\boldsymbol{A}\varepsilon(t)\varepsilon(t)^T\boldsymbol{A}^T\right)\end{aligned}$$

Kalman filtering

Step 1: an intermediate prediction,

$$\tilde{\boldsymbol{x}}(t+\Delta t,-) = \boldsymbol{A}\tilde{\boldsymbol{x}}(t)$$

where "$-$" indicates no data has been incorporated at time t

Step 2: Computing the uncertainty $\boldsymbol{P}(t+\Delta t,-)$:

$$\begin{aligned}
\boldsymbol{P}(t+\Delta t,-) &= E\left((\boldsymbol{A}\tilde{x}(t) - \boldsymbol{A}\boldsymbol{x}(t))(\boldsymbol{A}\tilde{x}(t) - \boldsymbol{A}\boldsymbol{x}(t))^T\right) \\
&= E\left(\boldsymbol{A}\varepsilon(t)\varepsilon(t)^T\boldsymbol{A}^T\right) \\
&= \boldsymbol{A}\boldsymbol{P}(t)\boldsymbol{A}^T
\end{aligned}$$

Kalman filtering

Step 1: an intermediate prediction,

$$\tilde{\boldsymbol{x}}(t+\Delta t, -) = \boldsymbol{A}\tilde{\boldsymbol{x}}(t)$$

where "$-$" indicates no data has been incorporated at time t
Step 2: Computing the uncertainty $\boldsymbol{P}(t+\Delta t, -)$:

$$\begin{aligned}\boldsymbol{P}(t+\Delta t, -) &= E\left((\boldsymbol{A}\tilde{x}(t) - \boldsymbol{A}\boldsymbol{x}(t))(\boldsymbol{A}\tilde{x}(t) - \boldsymbol{A}\boldsymbol{x}(t))^T\right) \\ &= E\left(\boldsymbol{A}\varepsilon(t)\varepsilon(t)^T\boldsymbol{A}^T\right) \\ &= \boldsymbol{A}\boldsymbol{P}(t)\boldsymbol{A}^T\end{aligned}$$

Which defines for us

$$\boldsymbol{P}(t+\Delta t, -) := \boldsymbol{A}\boldsymbol{P}(t)\boldsymbol{A}^T$$

Kalman filtering

We now have (1) an estimate $\tilde{\boldsymbol{x}}(t+\Delta t,-)$ and (2) its uncertainty $\boldsymbol{P}(t+\Delta t,-)$. How to add the data $\boldsymbol{y}(t+\Delta t)$?

Kalman filtering

We now have (1) an estimate $\tilde{\boldsymbol{x}}(t+\Delta t,-)$ and (2) its uncertainty $\boldsymbol{P}(t+\Delta t,-)$. How to add the data $\boldsymbol{y}(t+\Delta t)$?

As before, with the least-squares linear estimation!

Kalman filtering

We now have (1) an estimate $\tilde{\boldsymbol{x}}(t+\Delta t,-)$ and (2) its uncertainty $\boldsymbol{P}(t+\Delta t,-)$. How to add the data $\boldsymbol{y}(t+\Delta t)$?

As before, with the least-squares linear estimation!

Let's list the knowns:

- $\tilde{\boldsymbol{x}}(t+\Delta t,-)=\boldsymbol{I}\boldsymbol{x}(t+\Delta t)+\varepsilon$

Kalman filtering

We now have (1) an estimate $\tilde{\boldsymbol{x}}(t+\Delta t,-)$ and (2) its uncertainty $\boldsymbol{P}(t+\Delta t,-)$. How to add the data $\boldsymbol{y}(t+\Delta t)$?

As before, with the least-squares linear estimation!

Let's list the knowns:

- $\tilde{\boldsymbol{x}}(t+\Delta t,-) = \boldsymbol{I}\boldsymbol{x}(t+\Delta t) + \varepsilon$
- $\boldsymbol{y}(t+\Delta t) = \boldsymbol{E}\boldsymbol{x}(t+\Delta t) + \boldsymbol{n}(t+\Delta t)$

Kalman filtering

We now have (1) an estimate $\tilde{\boldsymbol{x}}(t+\Delta t,-)$ and (2) its uncertainty $\boldsymbol{P}(t+\Delta t,-)$. How to add the data $\boldsymbol{y}(t+\Delta t)$?

As before, with the least-squares linear estimation!

Let's list the knowns:

- $\tilde{\boldsymbol{x}}(t+\Delta t,-) = \boldsymbol{I}\boldsymbol{x}(t+\Delta t) + \varepsilon$
- $\boldsymbol{y}(t+\Delta t) = \boldsymbol{E}\boldsymbol{x}(t+\Delta t) + \boldsymbol{n}(t+\Delta t)$
- $E(\varepsilon\varepsilon^T) = \boldsymbol{P}(t+\Delta t,-)$

Kalman filtering

We now have (1) an estimate $\tilde{\boldsymbol{x}}(t + \Delta t, -)$ and (2) its uncertainty $\boldsymbol{P}(t + \Delta t, -)$. How to add the data $\boldsymbol{y}(t + \Delta t)$?

As before, with the least-squares linear estimation!

Let's list the knowns:

- $\tilde{\boldsymbol{x}}(t + \Delta t, -) = \boldsymbol{I}\boldsymbol{x}(t + \Delta t) + \varepsilon$
- $\boldsymbol{y}(t+\Delta t) = \boldsymbol{E}\boldsymbol{x}(t+\Delta t)+\boldsymbol{n}(t+\Delta t)$
- $E(\varepsilon\varepsilon^T) = \boldsymbol{P}(t + \Delta t, -)$
- $E(\boldsymbol{n}\boldsymbol{n}^T) = \boldsymbol{R}(t + \Delta t)$

Kalman filtering

We now have (1) an estimate $\tilde{\boldsymbol{x}}(t+\Delta t, -)$ and (2) its uncertainty $\boldsymbol{P}(t+\Delta t, -)$. How to add the data $\boldsymbol{y}(t+\Delta t)$?

As before, with the least-squares linear estimation!

Let's list the knowns:

- $\tilde{\boldsymbol{x}}(t+\Delta t, -) = \boldsymbol{I}\boldsymbol{x}(t+\Delta t) + \varepsilon$
- $\boldsymbol{y}(t+\Delta t) = \boldsymbol{E}\boldsymbol{x}(t+\Delta t) + \boldsymbol{n}(t+\Delta t)$
- $E(\varepsilon\varepsilon^T) = \boldsymbol{P}(t+\Delta t, -)$
- $E(\boldsymbol{n}\boldsymbol{n}^T) = \boldsymbol{R}(t+\Delta t)$

Goal:

$$\tilde{\boldsymbol{x}}(t+\Delta t) =? \begin{bmatrix} \tilde{\boldsymbol{x}}(t+\Delta t, -) \\ \boldsymbol{y}(t+\Delta t) \end{bmatrix}$$

Kalman filtering

Luckily, we found what $?$ was for multi-dimensions:

$$? = (\underline{\boldsymbol{E}}^T \underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{E}})^{-1} \underline{\boldsymbol{E}}^T \underline{\boldsymbol{R}}^{-1}$$

Kalman filtering

Luckily, we found what $?$ was for multi-dimensions:

$$? = (\underline{\boldsymbol{E}}^T \underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{E}})^{-1} \underline{\boldsymbol{E}}^T \underline{\boldsymbol{R}}^{-1}$$

$$\underline{\boldsymbol{E}} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix},$$

Kalman filtering

Luckily, we found what $?$ was for multi-dimensions:

$$? = (\underline{\boldsymbol{E}}^T \underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{E}})^{-1} \underline{\boldsymbol{E}}^T \underline{\boldsymbol{R}}^{-1}$$

$$\underline{\boldsymbol{E}} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix}, \quad \underline{\boldsymbol{R}} = \begin{bmatrix} \boldsymbol{P}(-) & 0 \\ 0 & \boldsymbol{R} \end{bmatrix}$$

Kalman filtering: the Kalman gain matrix

Putting it all together, the BLUE from the Kalman filter is:

$$\tilde{\boldsymbol{x}} = \left\{ \begin{bmatrix} \boldsymbol{I} & \boldsymbol{E}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{P}^{-1}(-) & 0 \\ 0 & \boldsymbol{R}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{P}^{-1}(-) & 0 \\ 0 & \boldsymbol{R}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{x}}(-) \\ \boldsymbol{y} \end{bmatrix}$$

Kalman filtering: the Kalman gain matrix

Putting it all together, the BLUE from the Kalman filter is:

$$\tilde{\boldsymbol{x}} = \left\{ \begin{bmatrix} \boldsymbol{I} & \boldsymbol{E}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{P}^{-1}(-) & 0 \\ 0 & \boldsymbol{R}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{P}^{-1}(-) & 0 \\ 0 & \boldsymbol{R}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{x}}(-) \\ \boldsymbol{y} \end{bmatrix}$$

$$= \tilde{\boldsymbol{x}}(-) + \boldsymbol{P}(-)\boldsymbol{E}^T \left(\boldsymbol{E}\boldsymbol{P}(-)\boldsymbol{E}^T + \boldsymbol{R} \right)^{-1} \left(\boldsymbol{y} - \boldsymbol{E}\tilde{\boldsymbol{x}}(-) \right)$$

Kalman filtering: the Kalman gain matrix

Putting it all together, the BLUE from the Kalman filter is:

$$\tilde{\boldsymbol{x}} = \left\{ \begin{bmatrix} \boldsymbol{I} & \boldsymbol{E}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{P}^{-1}(-) & 0 \\ 0 & \boldsymbol{R}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{E} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{P}^{-1}(-) & 0 \\ 0 & \boldsymbol{R}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{x}}(-) \\ \boldsymbol{y} \end{bmatrix}$$

$$= \tilde{\boldsymbol{x}}(-) + \boldsymbol{P}(-)\boldsymbol{E}^T \left(\boldsymbol{E}\boldsymbol{P}(-)\boldsymbol{E}^T + \boldsymbol{R} \right)^{-1} \left(\boldsymbol{y} - \boldsymbol{E}\tilde{\boldsymbol{x}}(-) \right)$$

This is the **Kalman gain** matrix

Kalman filtering: improved uncertainty

Similarly, we can use prior results to update the uncertainty $\boldsymbol{P}(t + \Delta t, -)$:

Kalman filtering: improved uncertainty

Similarly, we can use prior results to update the uncertainty $\boldsymbol{P}(t+\Delta t,-)$:

$$\boldsymbol{P}(t+\Delta t) = (\underline{\boldsymbol{E}}^T \underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{E}})^{-1}$$

Kalman filtering: improved uncertainty

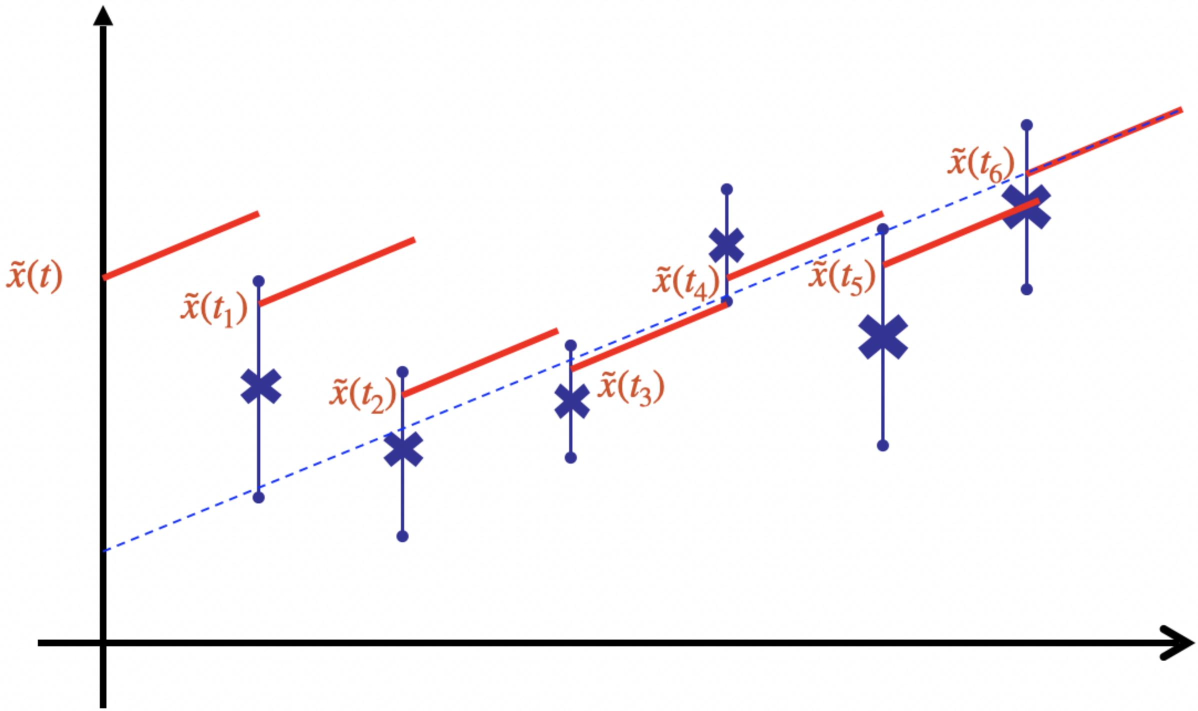
Similarly, we can use prior results to update the uncertainty $\boldsymbol{P}(t+\Delta t,-)$:

$$\begin{aligned}\boldsymbol{P}(t+\Delta t) &= (\underline{\boldsymbol{E}}^T\underline{\boldsymbol{R}}^{-1}\underline{\boldsymbol{E}})^{-1} \\ &= \boldsymbol{P}(t+\Delta t,-) - \boldsymbol{K}(t+\Delta)\boldsymbol{E}\boldsymbol{P}(t+\Delta t,-)\end{aligned}$$

Kalman filter equations

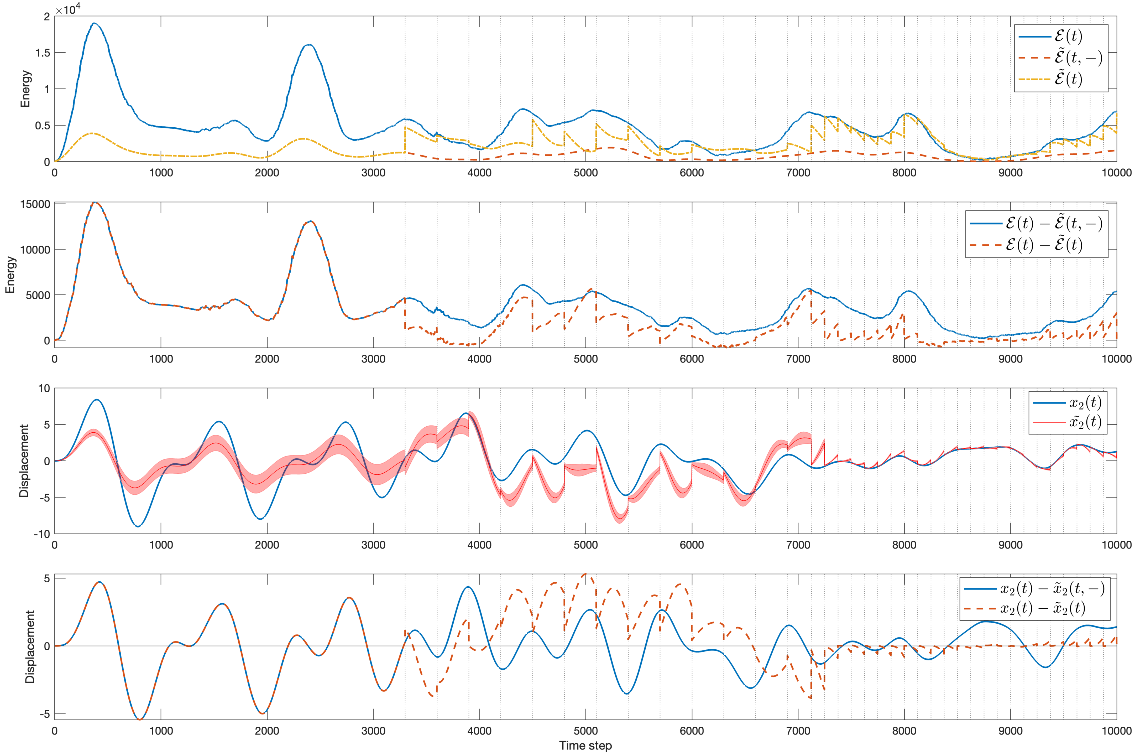
$$
\begin{aligned}
\tilde{\boldsymbol{x}}(t+\Delta t,-) &= \boldsymbol{A}\tilde{\boldsymbol{x}}(t)\\
\boldsymbol{P}(t+\Delta t,-) &= \boldsymbol{A}\boldsymbol{P}(t)\boldsymbol{A}^T\\
\boldsymbol{K}(t+\Delta t) &= \boldsymbol{P}(t+\Delta t,-)\boldsymbol{E}^T\left(\boldsymbol{E}\boldsymbol{P}(t+\Delta t,-)\boldsymbol{E}^T+\boldsymbol{R}(t+\Delta t)\right)^{-1}\\
\tilde{\boldsymbol{x}}(t+\Delta t) &= \tilde{\boldsymbol{x}}(t+\Delta t,-)+\boldsymbol{K}(t+\Delta t)\left(\boldsymbol{y}(t+\Delta t)-\boldsymbol{E}\tilde{\boldsymbol{x}}(t+\Delta t,-)\right)\\
\boldsymbol{P}(t+\Delta t) &= \boldsymbol{P}(t,-)-\boldsymbol{K}(t+\Delta t)\boldsymbol{E}\boldsymbol{P}(t,-)
\end{aligned}
$$

Sequential estimation



Kalman filter discontinuities

From “Potential artifacts in conservation laws and invariants inferred from sequential state estimation” by Wunsch, **Williamson**, Heimbach



Variational methods: smoothing problem

Variational methods

Consider the model-data misfit

$$J(\boldsymbol{x}) = \sum_{t=t_0}^{t_f} (\boldsymbol{E}\boldsymbol{x}(t) - \boldsymbol{y}(t))^T \boldsymbol{R}^{-1} (\boldsymbol{E}\boldsymbol{x}(t) - \boldsymbol{y}(t))$$

Variational methods

Consider the model-data misfit

$$J(\boldsymbol{x}) = \sum_{t=t_0}^{t_f} (\boldsymbol{E}\boldsymbol{x}(t) - \boldsymbol{y}(t))^T \boldsymbol{R}^{-1} (\boldsymbol{E}\boldsymbol{x}(t) - \boldsymbol{y}(t))$$

- Our goal is still the BLUE $\tilde{\boldsymbol{x}}(t)$ given available data $\boldsymbol{y}$

Variational methods

Consider the model-data misfit

$$J(\boldsymbol{x}) = \sum_{t=t_0}^{t_f} (\boldsymbol{Ex}(t) - \boldsymbol{y}(t))^T \boldsymbol{R}^{-1} (\boldsymbol{Ex}(t) - \boldsymbol{y}(t))$$

- Our goal is still the BLUE $\tilde{\boldsymbol{x}}(t)$ given available data $\boldsymbol{y}$
- In variational methods, we'll minimize J under the constraint $\boldsymbol{x}(t + \Delta t) = \boldsymbol{Ax}(t)$

Variational methods

Consider the model-data misfit

$$J(\boldsymbol{x}) = \sum_{t=t_0}^{t_f} (\boldsymbol{Ex}(t) - \boldsymbol{y}(t))^T \boldsymbol{R}^{-1} (\boldsymbol{Ex}(t) - \boldsymbol{y}(t))$$

- Our goal is still the BLUE $\tilde{\boldsymbol{x}}(t)$ given available data $\boldsymbol{y}$
- In variational methods, we'll minimize J under the constraint $\boldsymbol{x}(t + \Delta t) = \boldsymbol{Ax}(t)$

Variational methods combine available observations with the numerical model by globally adjusting the model to observations

Variational methods

- The method we'll use to minimize J is the Lagrange multiplier method

Variational methods

- The method we'll use to minimize J is the Lagrange multiplier method
- Step zero in a variational scheme is to decide the control variables

Variational methods

- The method we'll use to minimize J is the Lagrange multiplier method
- Step zero in a variational scheme is to decide the control variables
- Our control variable will be the initial condition

Variational methods

- The method we'll use to minimize J is the Lagrange multiplier method
- Step zero in a variational scheme is to decide the control variables
- Our control variable will be the initial condition

Goal

Utilize Lagrange multipliers to compute

$$\frac{\partial J}{\partial \mathrm{x}(t_0)}$$

Minimizing J

- Let $\mu(t)$ denote the Lagrange multipliers

Minimizing J

- Let $\mu(t)$ denote the Lagrange multipliers
- Our Lagrangian then is:

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \left(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t - \Delta t)\right)$$

Minimizing J

- Let $\mu(t)$ denote the Lagrange multipliers
- Our Lagrangian then is:

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \left(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t - \Delta t)\right)$$

- Stationary points of $\mathcal{L}$ correspond to the minima of the constrained optimization problem!

Finding stationary points of $\mathcal{L}$: the normal equations

We now take derivatives of

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \big(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t - \Delta t)\big)$$

Finding stationary points of $\mathcal{L}$: the normal equations

We now take derivatives of

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \big(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t-\Delta t)\big)$$

$$\frac{\partial \mathcal{L}}{\partial \mu(t)}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_f)}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t)}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_0)}$$

Finding stationary points of $\mathcal{L}$: the normal equations

We now take derivatives of

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \big(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t - \Delta t)\big)$$

$$\frac{\partial \mathcal{L}}{\partial \mu(t)} = \boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t + \Delta t) = 0 \qquad t_0 \leq t \leq t_f$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_f)}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t)}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_0)}$$

Finding stationary points of $\mathcal{L}$: the normal equations

We now take derivatives of

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \big(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t-\Delta t)\big)$$

$$\frac{\partial \mathcal{L}}{\partial \mu(t)} = \boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t+\Delta t) = 0 \qquad t_0 \le t \le t_f$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_f)} = \frac{\partial J}{\partial \boldsymbol{x}(t_f)} + \mu(t_f) = 0 \qquad t = t_f$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t)}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_0)}$$

Finding stationary points of $\mathcal{L}$: the normal equations

We now take derivatives of

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \left(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t-\Delta t)\right)$$

$$\frac{\partial \mathcal{L}}{\partial \mu(t)} = \boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t+\Delta t) = 0 \qquad t_0 \le t \le t_f$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_f)} = \frac{\partial J}{\partial \boldsymbol{x}(t_f)} + \mu(t_f) = 0 \qquad t = t_f$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t)} = \frac{\partial J}{\partial \boldsymbol{x}(t)} + \mu(t) - \left(\frac{\partial \boldsymbol{A}(\boldsymbol{x}(t))}{\partial \boldsymbol{x}(t)}\right)^T \mu(t+\Delta t) = 0 \qquad t_0 < t \le t_f - \Delta t$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_0)}$$

Finding stationary points of $\mathcal{L}$: the normal equations

We now take derivatives of

$$\mathcal{L}(\boldsymbol{x}, \mu) = J(\boldsymbol{x}) + \sum_{t=t_0+\Delta t}^{t_f} \mu(t)^T \left(\boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t - \Delta t)\right)$$

$$\frac{\partial \mathcal{L}}{\partial \mu(t)} = \boldsymbol{x}(t) - \boldsymbol{A}\boldsymbol{x}(t + \Delta t) = 0 \qquad t_0 \le t \le t_f$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_f)} = \frac{\partial J}{\partial \boldsymbol{x}(t_f)} + \mu(t_f) = 0 \qquad t = t_f$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t)} = \frac{\partial J}{\partial \boldsymbol{x}(t)} + \mu(t) - \left(\frac{\partial \boldsymbol{A}(\boldsymbol{x}(t))}{\partial \boldsymbol{x}(t)}\right)^T \mu(t + \Delta t) = 0 \qquad t_0 < t \le t_f - \Delta t$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}(t_0)} = \frac{\partial J}{\partial \boldsymbol{x}(t_0)} + \left(\frac{\partial \boldsymbol{A}(\boldsymbol{x}(t_0))}{\partial \boldsymbol{x}(t_0)}\right)^T \mu(t_0 + \Delta t) \qquad t = t_0$$

Variational framework (for adjoint method)

- Step 0: Define the control variables

Variational framework (for adjoint method)

- Step 0: Define the control variables
- Step 1: Locate Lagrangian stationary point (i.e. solve the normal equations)

Variational framework (for adjoint method)

- Step 0: Define the control variables
- Step 1: Locate Lagrangian stationary point (i.e. solve the normal equations)
- Step 2: Use computed derivative $\partial\mathcal{L}/\partial\boldsymbol{x}(t_0)$ in gradient based optimization

One sticky point in adjoint method

Computing model derivative

$$\frac{\partial \boldsymbol{A}(\boldsymbol{x}(t))}{\partial \boldsymbol{x}(t)}^{T}$$

is difficult

One sticky point in adjoint method

Computing model derivative

$$\frac{\partial \boldsymbol{A}(\boldsymbol{x}(t))}{\partial \boldsymbol{x}(t)}^{T}$$

is difficult

- One can write this out by hand, but it would be tedious

One sticky point in adjoint method

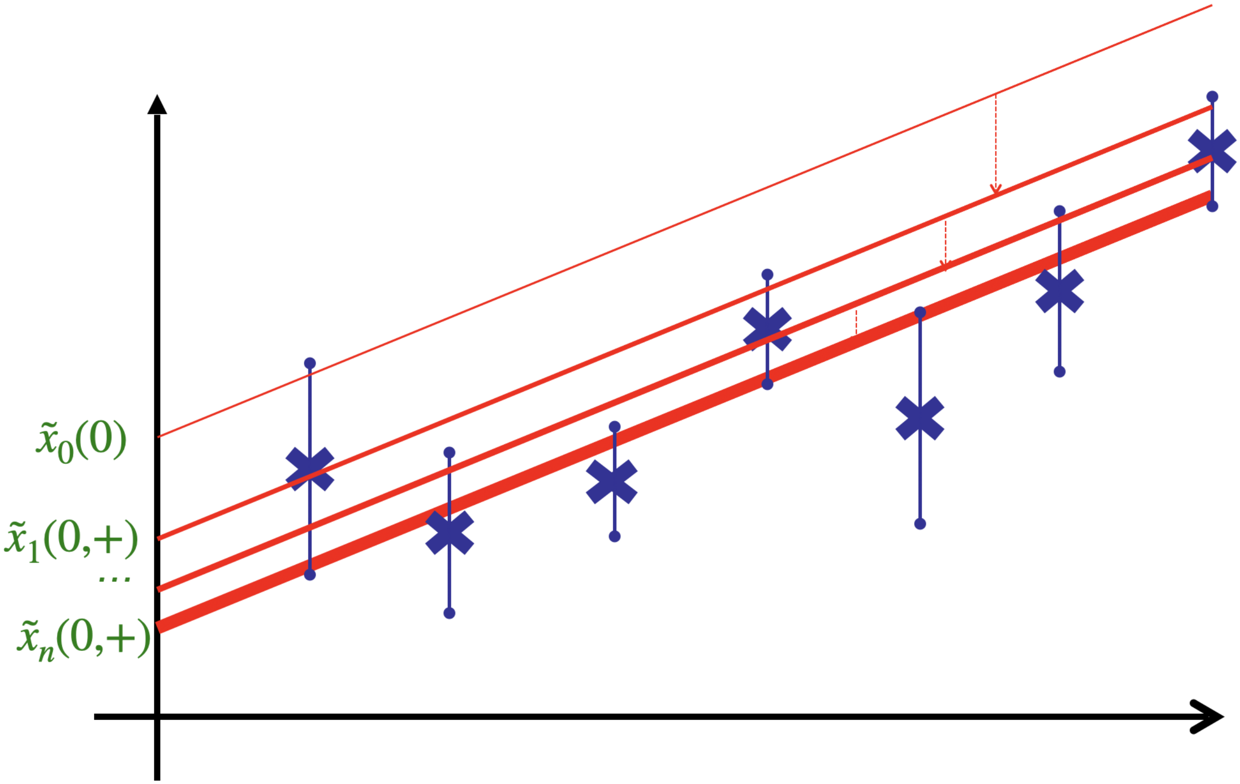
Computing model derivative

$$\frac{\partial \boldsymbol{A}(\boldsymbol{x}(t))}{\partial \boldsymbol{x}(t)}^T$$

is difficult

- One can write this out by hand, but it would be tedious
- Making use of automatic differentiation is ideal

Sketch of the variational method: improving initial condition



What about other control variables?

Variational methods are a little flexible

- We can opt to define (tune) a different control variable

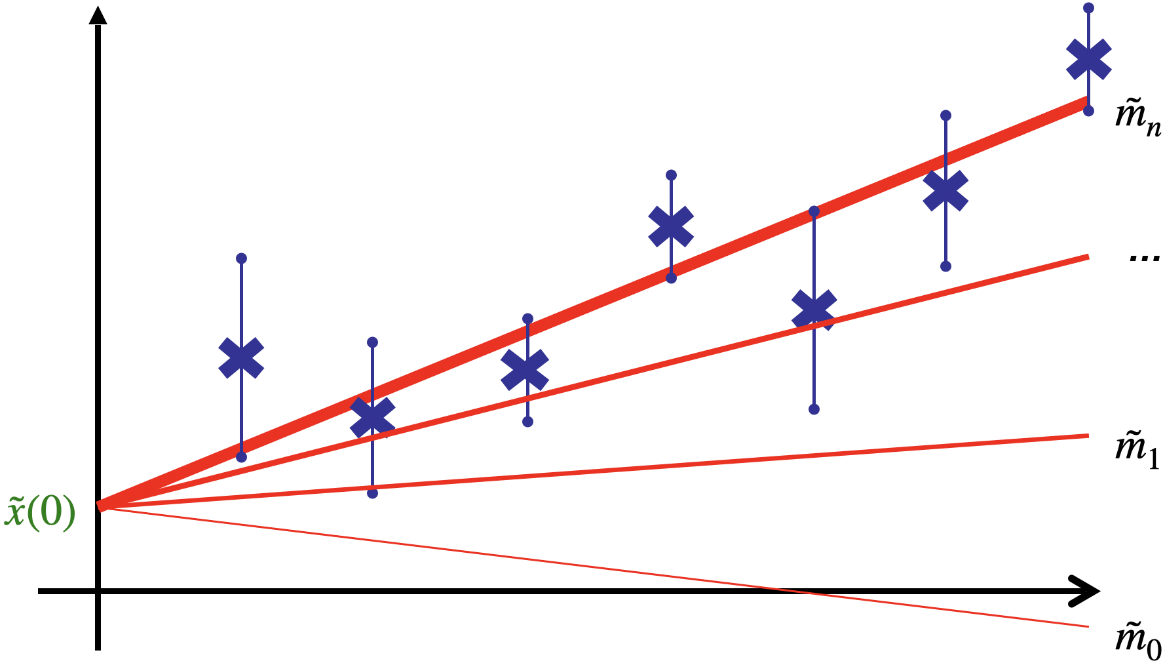
Variational methods are a little flexible

- We can opt to define (tune) a different control variable
- In these examples we've had a line $x(t) = mt + b$

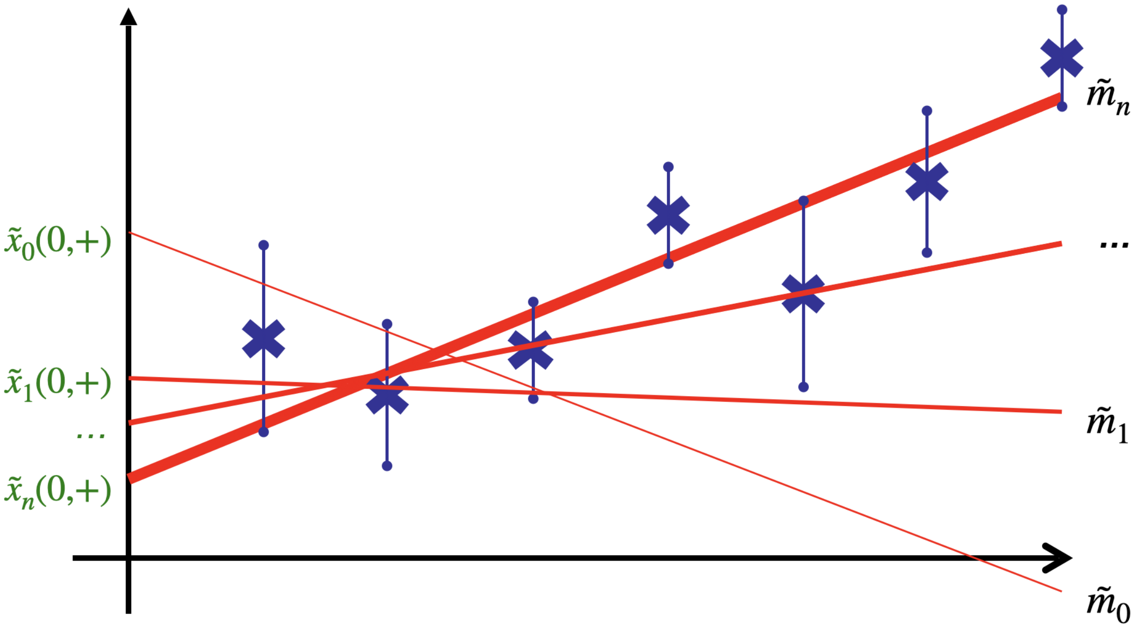
Variational methods are a little flexible

- We can opt to define (tune) a different control variable
- In these examples we've had a line $x(t) = mt + b$
- What about tuning other parameters?

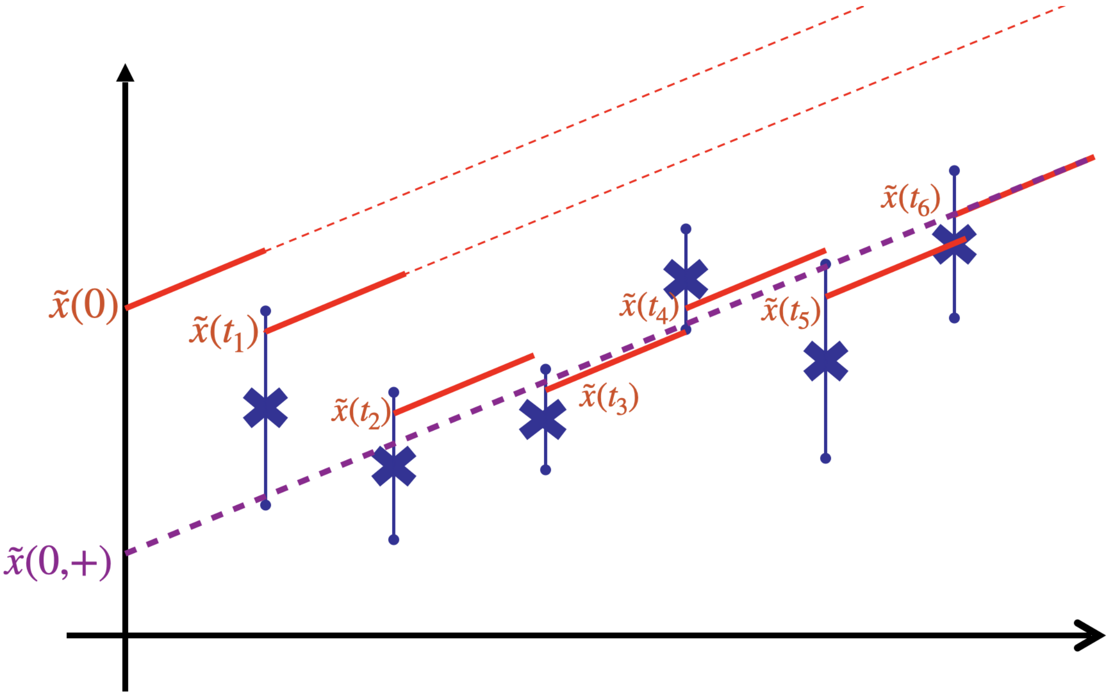
Variational method sketch: tuning m



Variational method sketch: tuning m and b

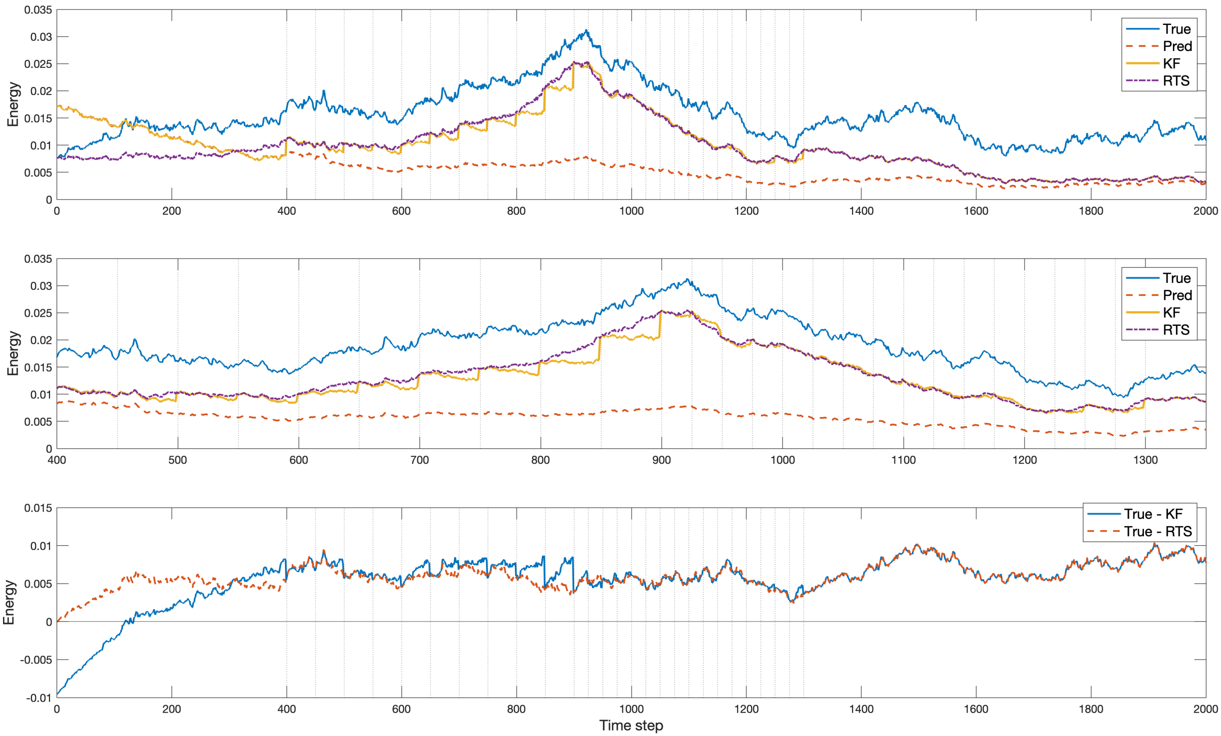


Sequential vs. variational methods



Sequential vs. variational methods

From "Potential artifacts in conservation laws and invariants inferred from sequential state estimation" by Wunsch, **Williamson**, Heimbach



Conclusions

- Data assimilation aims to combine observations, models, and their uncertainties

Conclusions

- Data assimilation aims to combine observations, models, and their uncertainties
- Depending on application, different methods warranted:
    - forecasting: sequential methods

Conclusions

- Data assimilation aims to combine observations, models, and their uncertainties
- Depending on application, different methods warranted:
  - forecasting: sequential methods
  - reconstruction: smoother methods

Conclusions

- Data assimilation aims to combine observations, models, and their uncertainties
- Depending on application, different methods warranted:
  - forecasting: sequential methods
  - reconstruction: smoother methods
- The solution to the DA problem (either filtering or smoothing) depends on the uncertainties that are provided for the background state and the observational error

Thank you for listening!