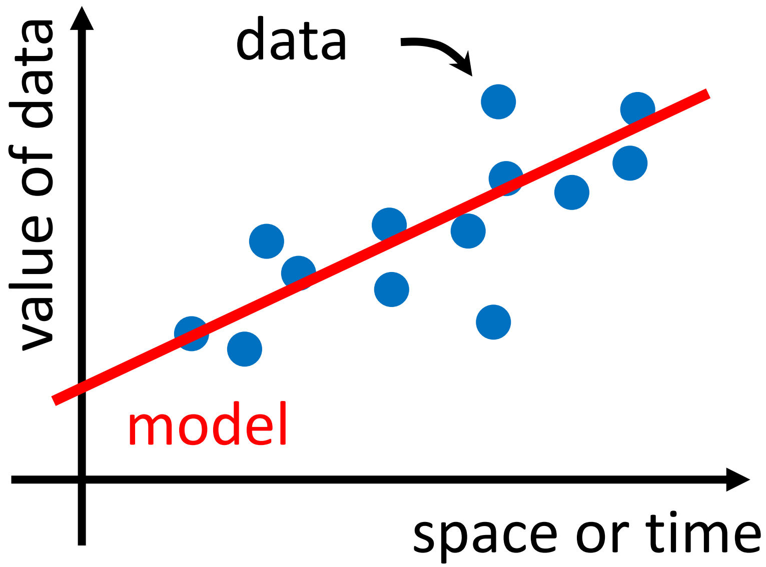
Jet Propulsion Laboratory
California Institute of Technology

An Overview of State Estimation

Ichiro Fukumori
Jet Propulsion Laboratory, Caltech

ECCO Summer School 2025

Scope of Lecture



"State Estimation" is
data analysis using a model.

- What is being solved?
- How is it being solved?

The Mathematical Problem (1/7)

Relation between model & data

model data

$$\mathbf{H}_t \mathbf{x}_t \approx \mathbf{y}_t \qquad t=1,2,\ldots,N$$

Observation matrix — state vector — data vector

model grid point

location of observation

$$\mathbf{x}_t \equiv \begin{pmatrix} \vdots \\ u_{ijk} \\ v_{ijk} \\ T_{ijk} \\ S_{ijk} \\ \eta_{ij} \\ \vdots \end{pmatrix}_t \qquad \mathbf{y}_t \equiv \begin{pmatrix} \vdots \\ \eta(x_n, y_n) \\ \vdots \\ T(x_n, y_n, z_n) \\ \vdots \end{pmatrix}_t$$

$$\left.\begin{array}{rl} u, v & \text{: zonal \& meridional velocity} \\ T, S & \text{: temperature \& salinity} \\ \eta & \text{: sea level} \end{array}\right\}$$

The Mathematical Problem (2/7)

Relation among model variables

$$\mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{u}_t \qquad t=1,2,\dots,N$$

State transition matrix

Forcing matrix

Control vector (forcing and other inhomogeneous terms, including errors of this relationship)

$$\mathbf{x}_t \equiv \begin{pmatrix} \vdots \\ u_{ijk} \\ v_{ijk} \\ T_{ijk} \\ S_{ijk} \\ \eta_{ij} \\ \vdots \end{pmatrix}_t \qquad \mathbf{u}_t \equiv \begin{pmatrix} \vdots \\ \tau^x_{ij} \\ \tau^y_{ij} \\ q_{ij} \\ e_{ij} \\ r_{ij} \\ \vdots \end{pmatrix}_t$$

model grid point

τ^x, τ^y : zonal & meridional wind stress

q : heat flux, e : evaporation

r : precipitation

The Mathematical Problem (3/7)

model state

data

model control

Observations $\quad \mathbf{H}_t \mathbf{x}_t \approx \mathbf{y}_t \qquad t=1,2,\dots,N$

Model $\quad \mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{u}_t \qquad t=1,2,\dots,N$

$$
\begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & \mathbf{H}_t & & & & & \\
 & & \mathbf{H}_{t+1} & & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & -\mathbf{A}_t & \mathbf{I} & & -\mathbf{G}_t & & \\
 & & -\mathbf{A}_{t+1} & \mathbf{I} & & -\mathbf{G}_{t+1} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
\vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots
\end{pmatrix}
\approx
\begin{pmatrix}
\vdots \\ \mathbf{y}_t \\ \mathbf{y}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots
\end{pmatrix}
$$

$$\mathbf{E}\,\mathbf{a} \approx \mathbf{b}$$

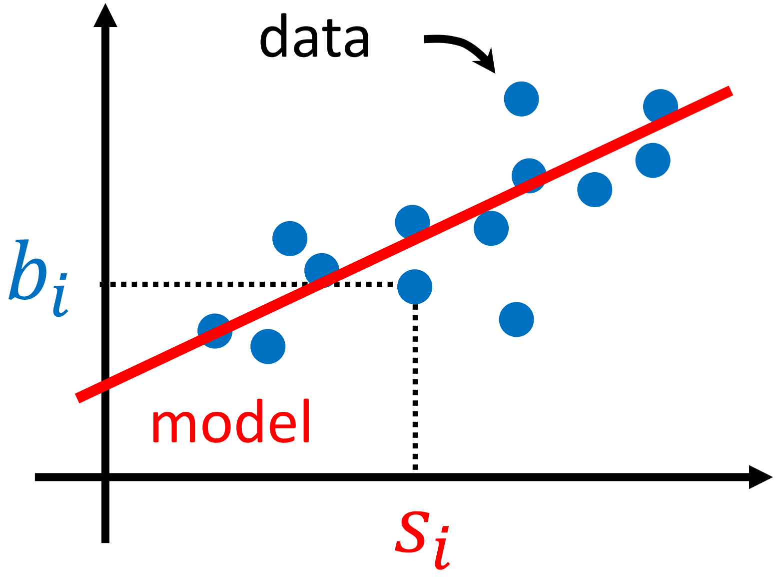
Combining model and data is an <u>inverse problem.</u>

The Mathematical Problem (4/7)

Linear Inverse Problem $\mathbf{E}\,\mathbf{a} \approx \mathbf{b}$

Given matrix $\mathbf{E}$ and vector $\mathbf{b}$ what is vector $\mathbf{a}$?

e.g., fitting a line through data $b = \alpha s + \beta$



$$\begin{pmatrix} s_1 & 1 \\ s_2 & 1 \\ \vdots & \vdots \\ s_n & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \approx \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

The Mathematical Problem (5/7)

$$
\begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & \mathbf{H}_t & & & & & \\
 & & \mathbf{H}_{t+1} & & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & -\mathbf{A}_t & \mathbf{I} & & -\mathbf{G}_t & & \\
 & & -\mathbf{A}_{t+1} & \mathbf{I} & & -\mathbf{G}_{t+1} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
\vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots
\end{pmatrix}
\approx
\begin{pmatrix}
\vdots \\ \mathbf{y}_t \\ \mathbf{y}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots
\end{pmatrix}
\Rightarrow
\mathbf{E}\,\mathbf{a} \approx \mathbf{b}
$$

$$N \times M$$

$$N \ll M$$

Challenges

1) The large size of the problem,

2) Ill-posedness with more unknowns than knowns.

The Mathematical Problem (6/7)

Curve-fitting is also fundamentally ill-posed, as no solution exactly satisfies the equations.

$$\begin{pmatrix} s_1 & 1 \\ s_2 & 1 \\ \vdots & \vdots \\ s_n & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \approx \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Explicitly writing the misfit of each equation $\gamma_1, \gamma_2, \cdots, \gamma_n$ where $\gamma_i = b_i - (\alpha s_i + \beta)$ the problem is mathematically,

$$\begin{pmatrix} s_1 & 1 & 1 & & & \\ s_2 & 1 & & 1 & & \\ \vdots & \vdots & & & \ddots & \\ s_n & 1 & & & & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

The Mathematical Problem (7/7)

The classic oceanographic inverse problem is that of determining reference level velocities in geostrophic calculations.

$$v(z) = v\left(z_{ref}\right) + \frac{g}{f\rho_0}\int_{z_{ref}}^{z} \frac{\partial\rho}{\partial x}\,dz$$

Wunsch (1977, Science)

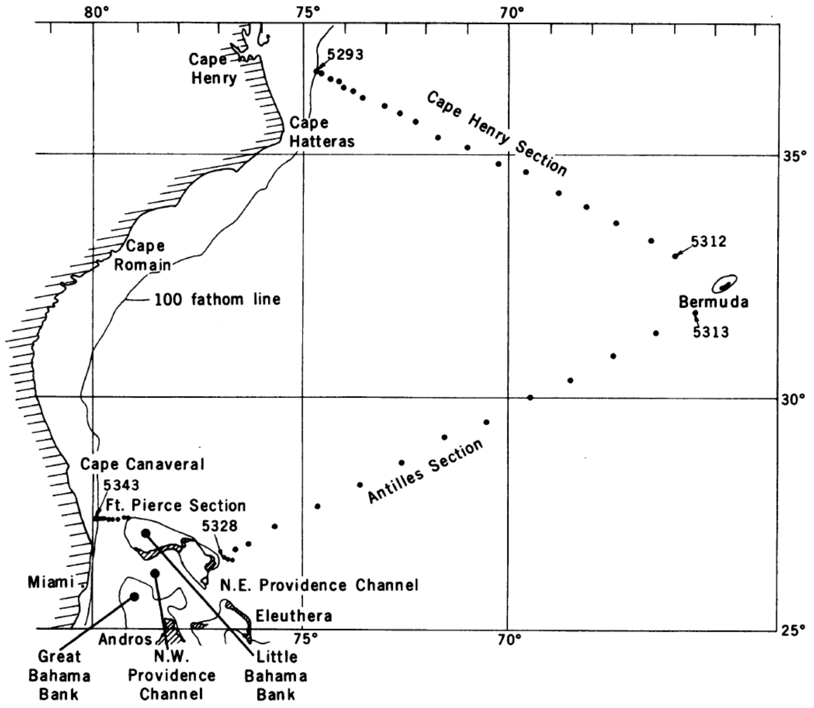


Fig. 1. Locations of *Atlantis* 215 stations used here.

range; then the statement that property C_k is conserved may be written

$$\sum_{j=1}^{M} (\bar{v}_{kj} + b_j)\,\Delta p_{kj}\Delta x_j = 0 \qquad (1)$$

and let there be $k = 1, \ldots, N$ such properties. Then we can combine Eq. 1 into matrix form

$$\mathbf{Ab} = -\Gamma \qquad (2)$$

where A is the $N \times M$ matrix of elements

Singular Value Decomposition (SVD)

The inverse problem $\mathbf{E}\,\mathbf{a} \approx \mathbf{b}$ can be solved using

Singular Value Decomposition $\mathbf{E} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i\, \boldsymbol{v}_i{}^T = \mathbf{U}\, \Xi\, \mathbf{V}^T$

where $r \leq min(N, M)$ and $\boldsymbol{u}_i{}^T \boldsymbol{u}_j = \boldsymbol{v}_i{}^T \boldsymbol{v}_j = \delta_{ij}$

Given this decomposition, we can write

$\mathbf{a} = \sum_{i=1}^{M} \gamma_i \boldsymbol{v}_i$ where $\gamma_i = \dfrac{\boldsymbol{u}_i{}^T \mathbf{b}}{\sigma_i}$ for $i = 1,2, \ldots r$

By "Ockham's Razor" we set $\gamma_i = 0$ for $i = r + 1, \ldots M$

Since $||\mathbf{a}|| = \sqrt{\sum_{i=1}^{M} \gamma_i{}^2}$ this is a Minimum Length Solution.

Other Inverse Methods

$$\mathbf{H}\,\mathbf{x} \approx \hat{\mathbf{y}}$$

Solve $\mathbf{E}\,\mathbf{a} \approx \mathbf{b}$ incorporating prior statistical information

- Minimum Variance Estimate

  aka Gauss-Markov inverse, basis of objective mapping.
  Closely related to the Kalman filter and related smoothers in state estimation.

- Least-Squares

  Closely related to the Adjoint Method (4dVAR) in state estimation.

… which turn out to be the same.

Minimum Variance Estimate

Solve $\mathbf{H}\mathbf{x} \approx \hat{\mathbf{y}}$ by incorporating prior statistical information

$$\mathbf{H}\mathbf{x} + \mathbf{n} = \hat{\mathbf{y}} \qquad \langle \mathbf{x}\mathbf{x}^T \rangle \equiv \mathbf{P} \qquad \langle \mathbf{n}\mathbf{n}^T \rangle \equiv \mathbf{R}$$

$$\langle \mathbf{x} \rangle = \mathbf{0} \qquad \langle \mathbf{n} \rangle = \mathbf{0}$$

A linear solution $\hat{\mathbf{x}} = \mathbf{B}\hat{\mathbf{y}}$ that minimizes error variance

$$\begin{aligned} \widehat{\mathbf{P}} &\equiv \langle (\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T \rangle \\ &= \langle (\mathbf{B}\hat{\mathbf{y}} - \mathbf{x})(\mathbf{B}\hat{\mathbf{y}} - \mathbf{x})^T \rangle \\ &= \mathbf{B} \langle \hat{\mathbf{y}}\hat{\mathbf{y}}^T \rangle \mathbf{B}^T - \langle \mathbf{x}\hat{\mathbf{y}}^T \rangle \mathbf{B}^T - \mathbf{B} \langle \hat{\mathbf{y}}\mathbf{x}^T \rangle + \langle \mathbf{x}\mathbf{x}^T \rangle \end{aligned}$$

is given by

$$\boxed{\hat{\mathbf{x}} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}\hat{\mathbf{y}}}$$

with error

$$\widehat{\mathbf{P}} = \mathbf{P} - \mathbf{P}\mathbf{H}^T (\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H}\mathbf{P}$$

Least-Squares Estimate (1/4)

Solve $\mathbf{Hx} \approx \hat{\mathbf{y}}$ that minimizes

$$\boxed{J = (\hat{\mathbf{y}} - \mathbf{Hx})^T \mathbf{R}^{-1} (\hat{\mathbf{y}} - \mathbf{Hx}) + \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}}$$

where $\mathbf{R} = \left\langle (\hat{\mathbf{y}} - \mathbf{Hx})(\hat{\mathbf{y}} - \mathbf{Hx})^T \right\rangle = \langle \mathbf{nn}^T \rangle$ and $\mathbf{P} = \langle \mathbf{xx}^T \rangle$

Why scale by inverse covariance?

Using Cholesky decomposition $\mathbf{R} = \mathbf{R}^{T/2}\, \mathbf{R}^{1/2}$

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ 0 & r_{22} & \cdots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{NN} \end{pmatrix}$$

define $\mathbf{n}' \equiv \mathbf{R}^{-T/2} (\hat{\mathbf{y}} - \mathbf{Hx})$

then $(\hat{\mathbf{y}} - \mathbf{Hx})^T \mathbf{R}^{-1} (\hat{\mathbf{y}} - \mathbf{Hx}) = \mathbf{n}'^T \mathbf{n}'$ Notably $\langle \mathbf{n}' \mathbf{n}'^T \rangle = \mathrm{I}$

Least-Squares Estimate (2/4)

Example of de-correlating variables

$$\mathbf{R} = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.99 & 0.14 \end{pmatrix} \begin{pmatrix} 1 & 0.99 \\ 0 & 1 \end{pmatrix} = \mathbf{R}^{T/2}\, \mathbf{R}^{1/2}$$

$$\mathbf{R}^{-T/2} = \begin{pmatrix} 1 & 0 \\ 0.99 & 0.14 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -7.0 & 7.1 \end{pmatrix}$$

$$\mathbf{n}' = \mathbf{R}^{-T/2}\mathbf{n} = \begin{pmatrix} 1 & 0 \\ -7.0 & 7.1 \end{pmatrix} \begin{pmatrix} n_1 \\ n_1 \end{pmatrix}$$

Instead of having two of the same, the scaled version has just one of them and the scaled difference between them.

Least-Squares Estimate (3/4)

To minimizes $$J = (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x}) + \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}$$

Solve $$\frac{\partial J}{\partial \mathbf{x}} = \mathbf{0}$$

Basic notation of vector differentiation

$$\frac{\partial s}{\partial \mathbf{x}} \equiv \begin{pmatrix} \frac{\partial s}{\partial x_1} & \cdots & \frac{\partial s}{\partial x_N} \end{pmatrix}^T \quad \Rightarrow \quad \frac{\partial (\boldsymbol{q}^T \mathbf{x})}{\partial \mathbf{x}} = \frac{\partial (\mathbf{x}^T \boldsymbol{q})}{\partial \mathbf{x}} = \boldsymbol{q}$$

$$\frac{\partial \boldsymbol{q}}{\partial \mathbf{x}} \equiv \begin{pmatrix} \frac{\partial q_1}{\partial x_1} & \cdots & \frac{\partial q_M}{\partial x_1} \\ \vdots & \cdots & \vdots \\ \frac{\partial q_1}{\partial x_N} & \cdots & \frac{\partial q_M}{\partial x_N} \end{pmatrix} \quad \Rightarrow \quad \frac{\partial (\mathbf{B}\mathbf{x})}{\partial \mathbf{x}} = \mathbf{B}^T \qquad \frac{\partial (\mathbf{x}^T \mathbf{B})}{\partial \mathbf{x}} = \mathbf{B}$$

Least-Squares Estimate (4/4)

To minimizes $$\boxed{J = (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x}) + \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}}$$

Solve $$\frac{\partial J}{\partial \mathbf{x}} = \mathbf{0}$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}} = \frac{1}{2}\frac{\partial (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x})}{\partial \mathbf{x}} \frac{\partial}{\partial (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x})} (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x}) + \frac{1}{2}\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}$$

$$= -\mathbf{H}^T \mathbf{R}^{-1} (\hat{\mathbf{y}} - \mathbf{H}\mathbf{x}) + \mathbf{P}^{-1} \mathbf{x}$$

$$= (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1}) \mathbf{x} - \mathbf{H}^T \mathbf{R}^{-1} \hat{\mathbf{y}}$$

Therefore, $$\boxed{\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \hat{\mathbf{y}}}$$

Comparing Solutions

Solve $\mathbf{Hx} \approx \hat{\mathbf{y}}$ by incorporating prior statistical information

Least-Squares $\boxed{\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1}\ \mathbf{H}^T\mathbf{R}^{-1}\,\hat{\mathbf{y}}}$

Minimum Variance $\boxed{\hat{\mathbf{x}} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}\,\hat{\mathbf{y}}}$

The two are identical based on the “Matrix Inversion Lemma”

$$\mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-\mathbf{1}}\ \mathbf{H}^T\mathbf{R}^{-1}$$

Recap

1. State estimation is an inverse problem,

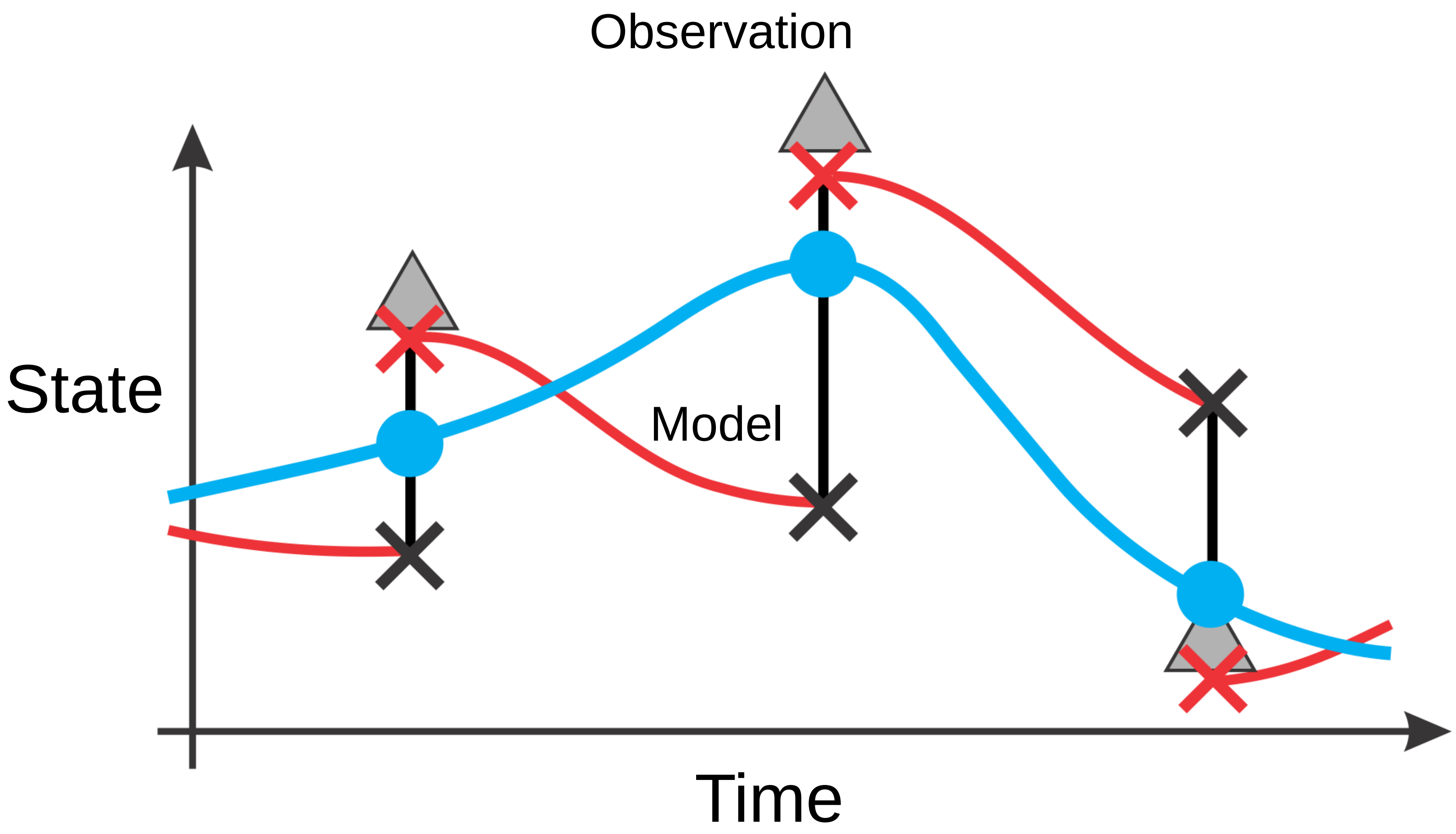
2. Various inverse methods are available;
    - Minimum length
    - Minimum Variance
    - Least-Squares
    - Maximum likelihood
    - Bayesian

Motivation for State Estimation (3/3)

From Monday

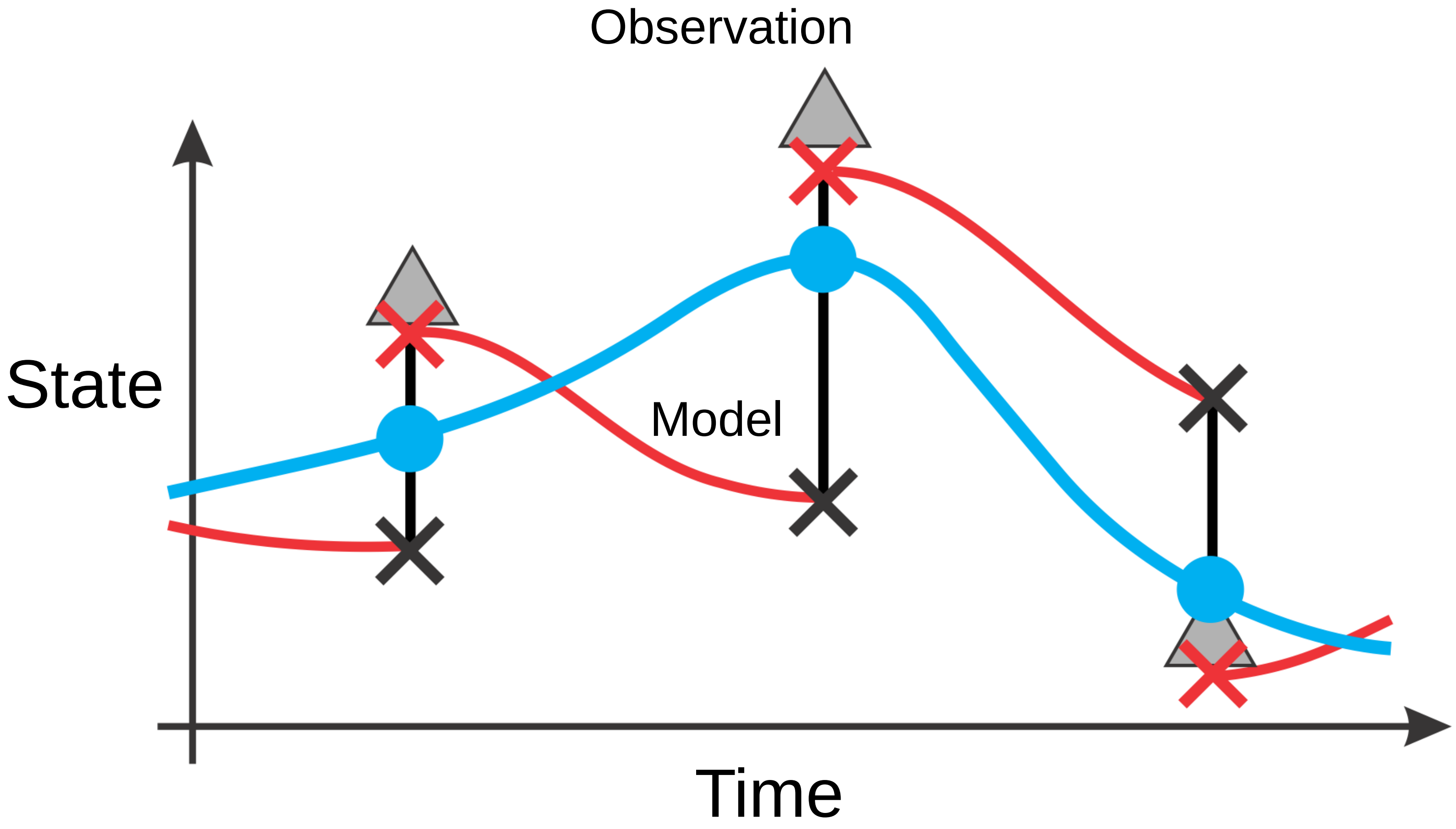
Data Assimilation vs State Estimation

From Monday



Kalman Filter (1/2)

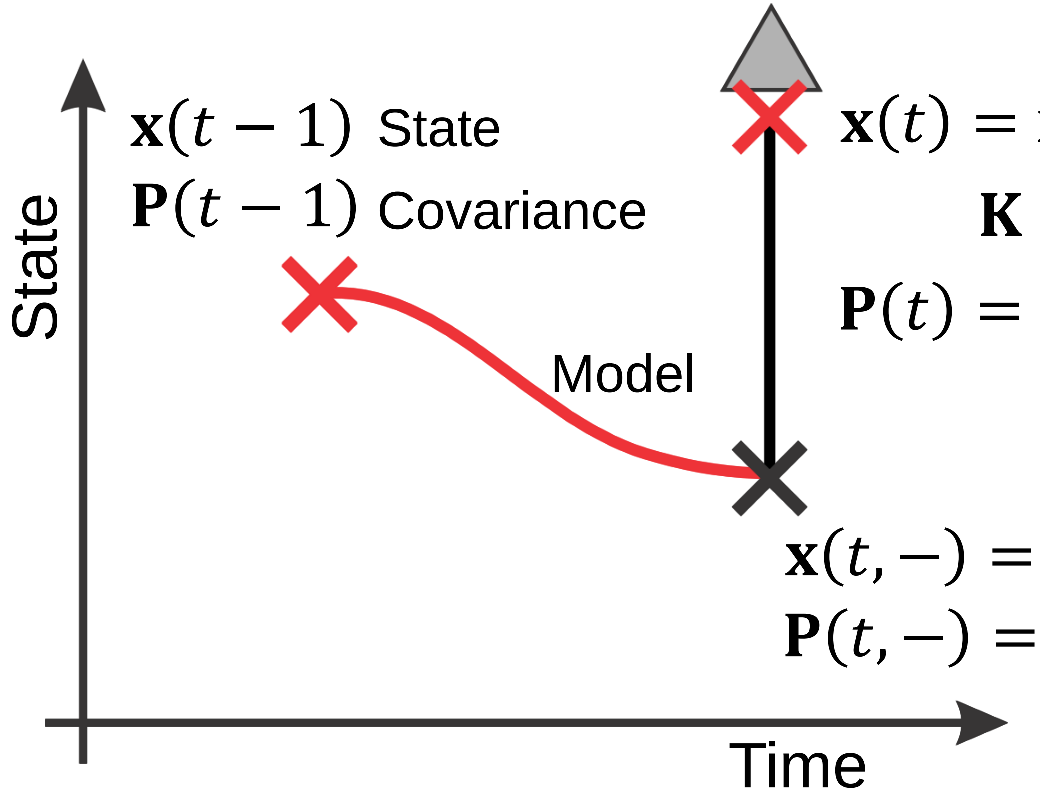
A recursive Minimum Variance Estimate



Kalman Filter (2/2)

A recursive Minimum Variance Estimate

Observation $\mathbf{y}(t)$ $\quad$ $\mathbf{Hx} \approx \mathbf{y}$ $\quad \Rightarrow \quad$ $\mathbf{H}\big(\mathbf{x} - \mathbf{x}(-)\big) \approx \mathbf{y} - \mathbf{Hx}(-)$



$$\mathbf{x}(t) = \mathbf{x}(t,-) + \mathbf{K}\big(\mathbf{y}(t) - \mathbf{Hx}(t,-)\big)$$

$$\mathbf{K} \equiv \mathbf{P}(t,-)\mathbf{H}^T\left(\mathbf{HP}(t,-)\mathbf{H}^T + \mathbf{R}\right)^{-1} \quad \text{Kalman Gain}$$

$$\mathbf{P}(t) = \mathbf{P}(t,-) - \mathbf{P}(t,-)\mathbf{H}^T\left(\mathbf{HP}(t,-)\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{HP}(t,-)$$

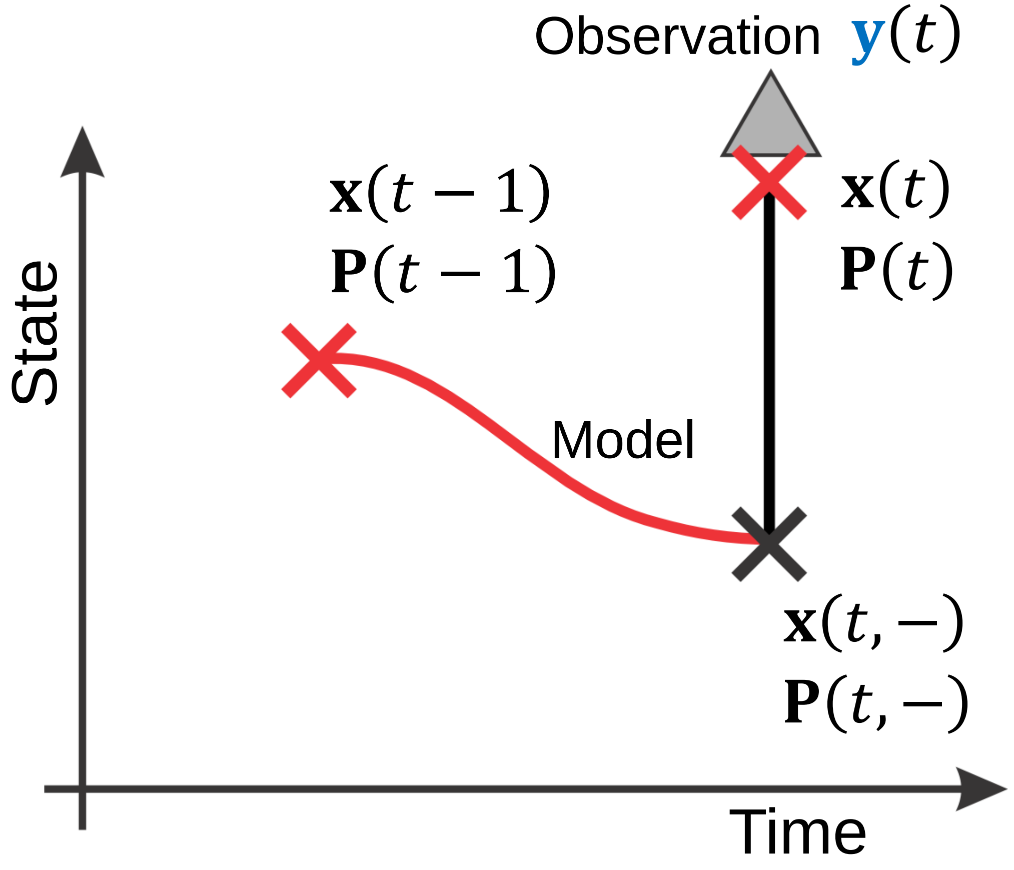
$$\mathbf{x}(t,-) = \mathbf{Ax}(t-1) + \mathbf{Gu}(t-1)$$

$$\mathbf{P}(t,-) = \mathbf{AP}(t-1)\mathbf{A}^T + \mathbf{GQG}^T$$

$$\mathbf{Q} \equiv \langle \mathbf{uu}^T \rangle$$

Kalman Filter (2/2)

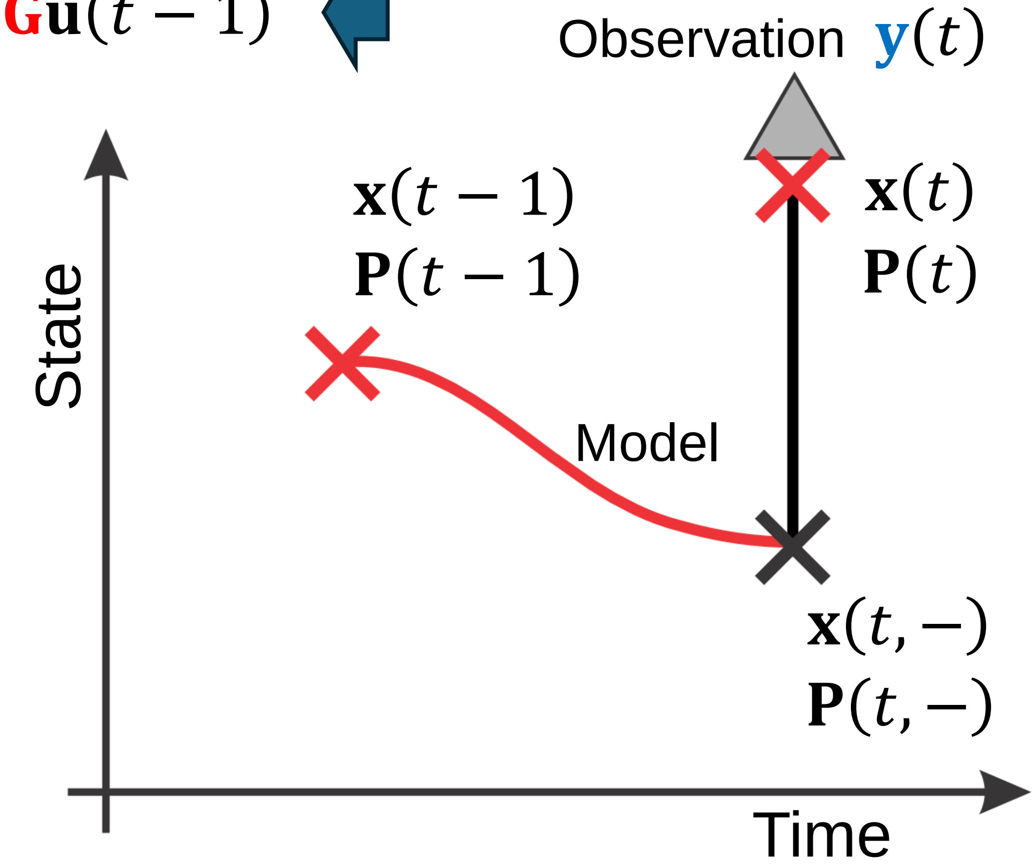
A recursive Minimum Variance Estimate



Rauch-Tung-Striebel Smoother (1/2)

A recursive Minimum Variance Estimate

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$$



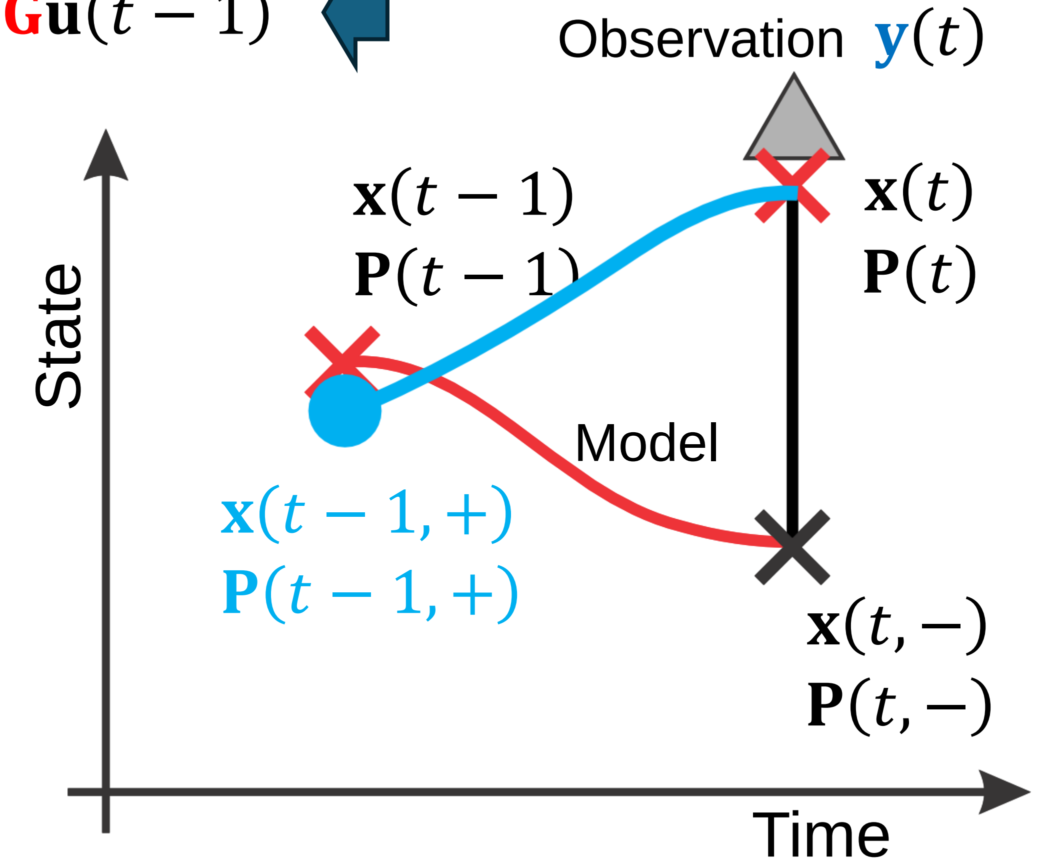
Rauch-Tung-Striebel Smoother (1/2)

A recursive Minimum Variance Estimate

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$$

$$\left.\begin{aligned} \mathbf{x}(t-1,+) &= \mathbf{x}(t-1) + \mathbf{S}[\mathbf{x}(t) - \mathbf{x}(t,-)] \\ \mathbf{S} &\equiv \mathbf{P}(t-1)\,\mathbf{A}^T\,\mathbf{P}(t,-)^{-1} \\ \mathbf{P}(t-1,+) &= \mathbf{P}(t-1) + \mathbf{S}[\mathbf{P}(t) - \mathbf{P}(t,-)]\mathbf{S}^T \end{aligned}\right\}$$

$$\left.\begin{aligned} \mathbf{u}(t-1,+) &= \mathbf{u}(t-1) + \mathbf{M}[\mathbf{x}(t) - \mathbf{x}(t,-)] \\ \mathbf{M} &\equiv \mathbf{Q}(t-1)\,\mathbf{G}^T\,\mathbf{P}(t,-)^{-1} \\ \mathbf{Q}(t-1,+) &= \mathbf{Q}(t-1) + \mathbf{M}[\mathbf{P}(t) - \mathbf{P}(t,-)]\mathbf{M}^T \end{aligned}\right\}$$



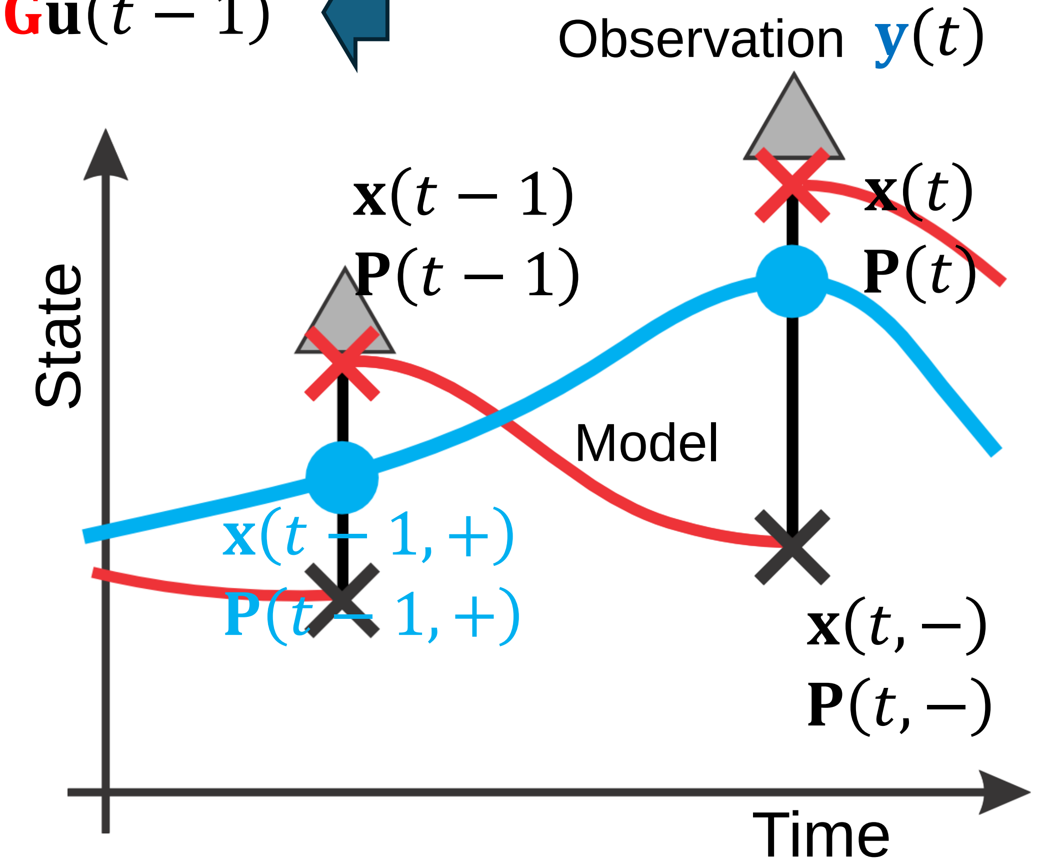
Rauch-Tung-Striebel Smoother (2/2)

A recursive Minimum Variance Estimate

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{G}\mathbf{u}(t-1)$$

$$\left.\begin{aligned} \mathbf{x}(t-1,+) &= \mathbf{x}(t-1) + \mathbf{S}[\mathbf{x}(t,+) - \mathbf{x}(t,-)] \\ \mathbf{S} &\equiv \mathbf{P}(t-1)\,\mathbf{A}^T\,\mathbf{P}(t,-)^{-1} \\ \mathbf{P}(t-1,+) &= \mathbf{P}(t-1) + \mathbf{S}[\mathbf{P}(t,+) - \mathbf{P}(t,-)]\mathbf{S}^T \end{aligned}\right\}$$

$$\left.\begin{aligned} \mathbf{u}(t-1,+) &= \mathbf{u}(t-1) + \mathbf{M}[\mathbf{x}(t,+) - \mathbf{x}(t,-)] \\ \mathbf{M} &\equiv \mathbf{Q}(t-1)\,\mathbf{G}^T\,\mathbf{P}(t,-)^{-1} \\ \mathbf{Q}(t-1,+) &= \mathbf{Q}(t-1) + \mathbf{M}[\mathbf{P}(t,+) - \mathbf{P}(t,-)]\mathbf{M}^T \end{aligned}\right\}$$



Adjoint Method (4D-var) (1/9)

A least-squares method that solves the problem iteratively by descent optimization.

Solve $\mathbf{Hx} \approx \mathbf{y}$ that minimizes

$$\boxed{J = (\mathbf{y} - \mathbf{Hx})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{Hx}) + \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}}$$

analytically

$$\boxed{\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1})^{-1} \; \mathbf{H}^T \mathbf{R}^{-1} \, \mathbf{y}}$$

Adjoint Method (4D-var) (2/9)

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & \mathbf{H}_t & & & & & \\ & & \mathbf{H}_{t+1} & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & & \\ & & -\mathbf{A} & \mathbf{I} & & -\mathbf{G} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ \mathbf{x}_{t+2} \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \mathbf{y}_t \\ \mathbf{y}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

In terms of least-squares, the problem is to minimize

$$J \equiv \sum_{t=1}^{M} (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t)^T \mathbf{R}_t^{-1} (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t)$$

$$+ (\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T \mathbf{P}_0^{-1} (\mathbf{x}_0 - \hat{\mathbf{x}}_0)$$

$$+ \sum_{t=0}^{M-1} (\mathbf{u}_t - \hat{\mathbf{u}}_t)^T \mathbf{Q}_t^{-1} (\mathbf{u}_t - \hat{\mathbf{u}}_t)$$

$$- 2\sum_{t=1}^{M} \boldsymbol{\lambda}_t^T [\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1}]$$

Lagrange multipliers

Adjoint Method (4D-var) (3/9)

$$J \equiv \sum_{t=1}^{M}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)^T\mathbf{R}_t{}^{-1}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)$$

$$+ (\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T\mathbf{P}_0{}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)$$

$$+ \sum_{t=0}^{M-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t)^T\mathbf{Q}_t{}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t)$$

$$- 2\sum_{t=1}^{M}\boldsymbol{\lambda}_t^T[\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1}]$$

Adjoint Method (4D-var) (4/9)

$$J \equiv \sum_{t=1}^{M}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)^T \mathbf{R}_t{}^{-1}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)$$
$$+ (\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T \mathbf{P}_0{}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)$$
$$+ \sum_{t=0}^{M-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t)^T \mathbf{Q}_t{}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t)$$
$$- 2\sum_{t=1}^{M} \boldsymbol{\lambda}_t^T[\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1}]$$

$$\left\{ \begin{array}{ll}
\frac{1}{2}\frac{\partial J}{\partial \mathbf{u}_t} = \mathbf{Q}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t) + \mathbf{G}^T\boldsymbol{\lambda}_{t+1} = \mathbf{0} & t = 0, 1, \ldots, M\text{-}1 \\
\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{P}_0{}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) + \mathbf{A}^T\boldsymbol{\lambda}_1 = \mathbf{0} & \\
\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_t} = -\mathbf{H}_t\mathbf{R}_t{}^{-1}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t) - \boldsymbol{\lambda}_t + \mathbf{A}^T\boldsymbol{\lambda}_{t+1} = \mathbf{0} & t = 1, 2, \ldots, M\text{-}1 \\
\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_M} = -\mathbf{H}_M\mathbf{R}_M{}^{-1}(\mathbf{y}_M - \mathbf{H}_M\mathbf{x}_M) - \boldsymbol{\lambda}_M = \mathbf{0} & \\
\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\lambda}_t} = \mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1} = \mathbf{0} & t = 1, 2, \ldots, M
\end{array} \right.$$

Adjoint Method (4D-var) (5/9)

$$\begin{cases} \frac{1}{2}\frac{\partial J}{\partial \mathbf{u}_t} = \mathbf{Q}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t) + \mathbf{G}^T \boldsymbol{\lambda}_{t+1} = \mathbf{0} & t = 0, 1, \ldots, M\text{-}1 \\ \frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{P}_0{}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) + \mathbf{A}^T \boldsymbol{\lambda}_1 = \mathbf{0} & \\ \frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_t} = -\mathbf{H}_t \mathbf{R}_t{}^{-1}(\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t) - \boldsymbol{\lambda}_t + \mathbf{A}^T \boldsymbol{\lambda}_{t+1} = \mathbf{0} & t = 1, 2, \ldots, M\text{-}1 \\ \frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_M} = -\mathbf{H}_M \mathbf{R}_M{}^{-1}(\mathbf{y}_M - \mathbf{H}_M \mathbf{x}_M) - \boldsymbol{\lambda}_M = \mathbf{0} & \\ \frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\lambda}_t} = \mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1} = \mathbf{0} & t = 1, 2, \ldots, M \end{cases}$$

1) Integrate model from first guess controls.

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{G}\mathbf{u}_{t-1} \qquad t = 1, 2, \ldots, M$$

with $\mathbf{x}_0 = \hat{\mathbf{x}}_0$ $\quad \mathbf{u}_t = \hat{\mathbf{u}}_t$

Adjoint Method (4D-var) (6/9)

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{u}_t} = \mathbf{Q}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t) + \mathbf{G}^T \boldsymbol{\lambda}_{t+1} \qquad t = 0, 1, \ldots, M\text{-}1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{P}_0{}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) + \mathbf{A}^T \boldsymbol{\lambda}_1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_t} = -\mathbf{H}_t \mathbf{R}_t{}^{-1}(\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t) - \boldsymbol{\lambda}_t + \mathbf{A}^T \boldsymbol{\lambda}_{t+1} \qquad t = 1, 2, \ldots, M\text{-}1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_M} = -\mathbf{H}_M \mathbf{R}_M{}^{-1}(\mathbf{y}_M - \mathbf{H}_M \mathbf{x}_M) - \boldsymbol{\lambda}_M = \mathbf{0}$$

$$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\lambda}_t} = \mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1} = \mathbf{0} \qquad t = 1, 2, \ldots, M$$

2) Determine terminal Lagrange multiplier.

$$\boldsymbol{\lambda}_M = -\mathbf{H}_M \mathbf{R}_M{}^{-1}(\mathbf{y}_M - \mathbf{H}_M \mathbf{x}_M)$$

Adjoint Method (4D-var) (7/9)

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{u}_t} = \mathbf{Q}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t) + \mathbf{G}^T \boldsymbol{\lambda}_{t+1} \qquad t = 0, 1, \ldots, M\text{-}1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{P}_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) + \mathbf{A}^T \boldsymbol{\lambda}_1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_t} = -\mathbf{H}_t \mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t) - \boldsymbol{\lambda}_t + \mathbf{A}^T \boldsymbol{\lambda}_{t+1} = \mathbf{0} \qquad t = 1, 2, \ldots, M\text{-}1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_M} = -\mathbf{H}_M \mathbf{R}_M^{-1}(\mathbf{y}_M - \mathbf{H}_M \mathbf{x}_M) - \boldsymbol{\lambda}_M = \mathbf{0}$$

$$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\lambda}_t} = \mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1} = \mathbf{0} \qquad t = 1, 2, \ldots, M$$

3) Determine all other Lagrange multipliers by the adjoint model.

$$\boldsymbol{\lambda}_t = \mathbf{A}^T \boldsymbol{\lambda}_{t+1} - \mathbf{H}_M \mathbf{R}_M^{-1}(\mathbf{y}_M - \mathbf{H}_M \mathbf{x}_M) \qquad t = M-1, M-2, \ldots, 1$$

with $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_M$ at $t = M$

Adjoint Method (4D-var) (8/9)

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{u}_t} = \mathbf{Q}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t) + \mathbf{G}^T\boldsymbol{\lambda}_{t+1} \qquad t = 0, 1, \ldots, M\text{-}1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{P}_0{}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) + \mathbf{A}^T\boldsymbol{\lambda}_1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_t} = -\mathbf{H}_t\mathbf{R}_t{}^{-1}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t) - \boldsymbol{\lambda}_t + \mathbf{A}^T\boldsymbol{\lambda}_{t+1} = \mathbf{0} \qquad t = 1, 2, \ldots, M\text{-}1$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_M} = -\mathbf{H}_M\mathbf{R}_M{}^{-1}(\mathbf{y}_M - \mathbf{H}_M\mathbf{x}_M) - \boldsymbol{\lambda}_M = \mathbf{0}$$

$$\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\lambda}_t} = \mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1} = \mathbf{0} \qquad t = 1, 2, \ldots, M$$

4) Calculate cost function gradient for descent algorithm;

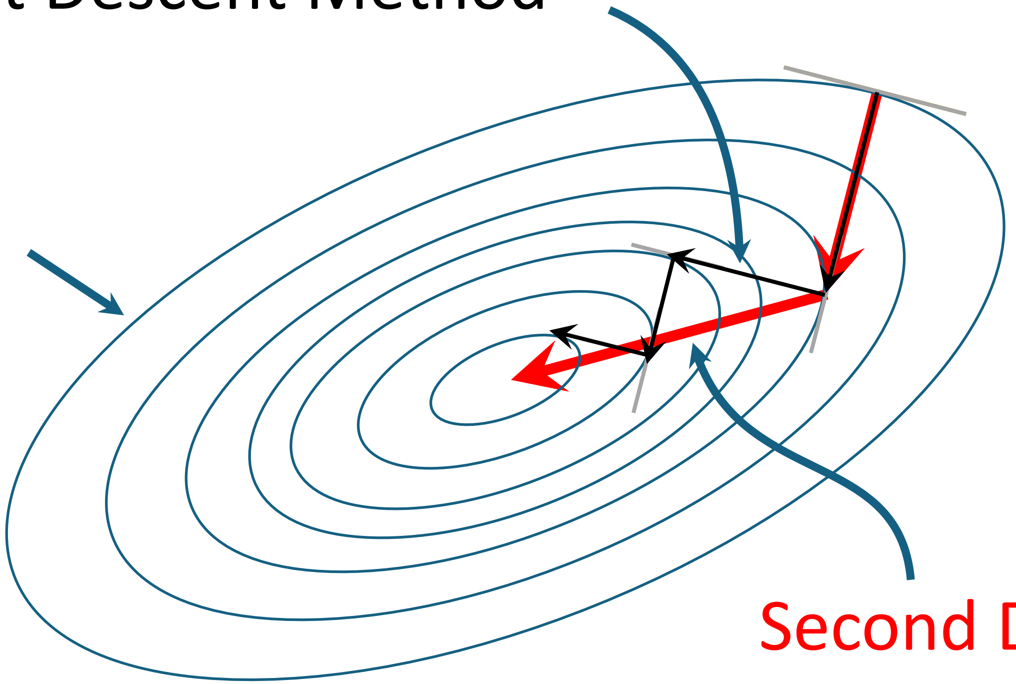
$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{P}_0{}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) + \mathbf{A}^T\boldsymbol{\lambda}_1 \qquad \frac{1}{2}\frac{\partial J}{\partial \mathbf{u}_t} = \mathbf{Q}_t{}^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t) + \mathbf{G}^T\boldsymbol{\lambda}_{t+1} \quad t = 0, 1, \ldots, M-1$$

Adjoint Method (4D-var) (9/9)

Descent Optimization

Steepest Descent Method

isopleths of J



Second Derivative Methods

e.g., Conjugate-Gradient Method
Quasi-Newton Method

What are "data error" R & "state error" P?

$$J \equiv \sum_{t=1}^{M}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)^T \mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)$$
$$+ (\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T \mathbf{P}_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)$$
$$+ \sum_{t=0}^{M-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t)^T \mathbf{Q}_t^{-1}(\mathbf{u}_t - \hat{\mathbf{u}}_t)$$
$$- 2\sum_{t=1}^{M} \boldsymbol{\lambda}_t^T [\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{G}\mathbf{u}_{t-1}]$$

$$\mathbf{P}(-) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T$$
$$\mathbf{P}(t) = \mathbf{P}(-) - \mathbf{P}(-)\mathbf{H}^T(\mathbf{H}\mathbf{P}(-)\mathbf{H}^T + \mathbf{R})^{-1}\,\mathbf{H}\mathbf{P}(-)$$

Data   True state of model

Data error $\hat{\mathbf{y}} - \mathbf{H}\mathbf{x}$

State Estimate

State error $\hat{\mathbf{x}} - \mathbf{x}$

Actual ocean

Representation error $\mathbf{x} - \mathbf{w}$

because $\mathbf{x} = \boldsymbol{\Gamma}\mathbf{w}$

Modeling

Example estimating $\mathbf{R}$

$\mathbf{R}$ can be estimated from observations and model simulation

$$\mathbf{y} = \mathbf{s} + \mathbf{n}$$

$$\mathbf{m} = \mathbf{s} + \mathbf{p}$$

Assuming signal and errors are mutually uncorrelated, i.e.,

$$\langle \mathbf{s}\mathbf{n}^T \rangle = \mathbf{0} \qquad \langle \mathbf{s}\mathbf{p}^T \rangle = \mathbf{0} \qquad \langle \mathbf{n}\mathbf{p}^T \rangle = \mathbf{0}$$

covariances among data and model simulation are,

$$\langle \boldsymbol{y}\boldsymbol{y}^T \rangle = \langle \mathbf{s}\mathbf{s}^T \rangle + \langle \mathbf{n}\mathbf{n}^T \rangle$$

$$\langle \mathbf{m}\mathbf{m}^T \rangle = \langle \mathbf{s}\mathbf{s}^T \rangle + \langle \mathbf{p}\mathbf{p}^T \rangle$$

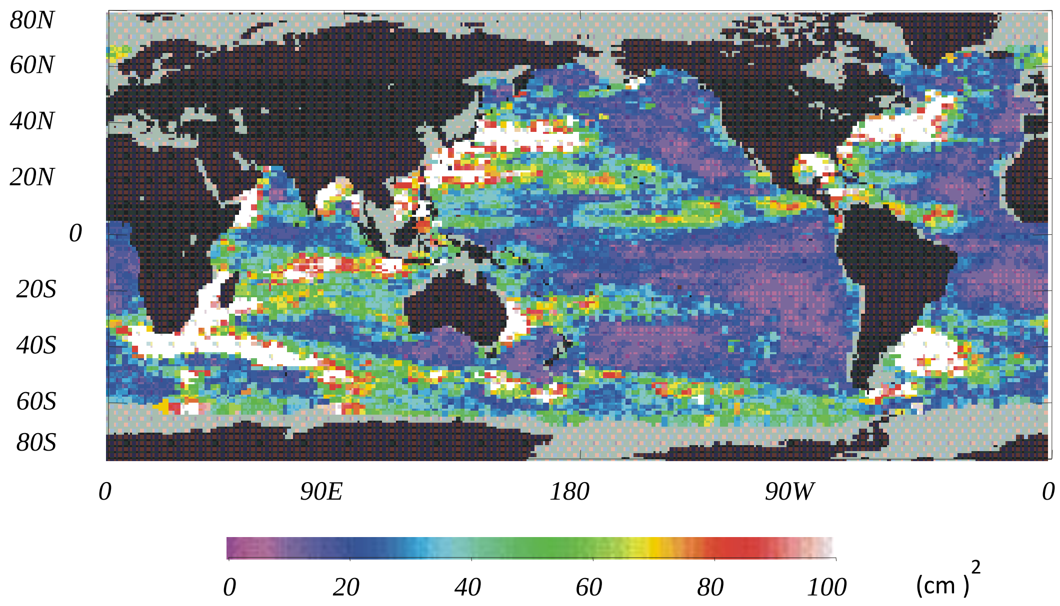
$$\langle \boldsymbol{y}\mathbf{m}^T \rangle = \langle \mathbf{s}\mathbf{s}^T \rangle$$

Which can be solved for $\langle \mathbf{n}\mathbf{n}^T \rangle \equiv \mathrm{R}$ and $\langle \mathbf{p}\mathbf{p}^T \rangle \equiv \mathbf{H}\mathbf{P}_{\boldsymbol{sim}}{}^T\ \mathbf{H}^T$

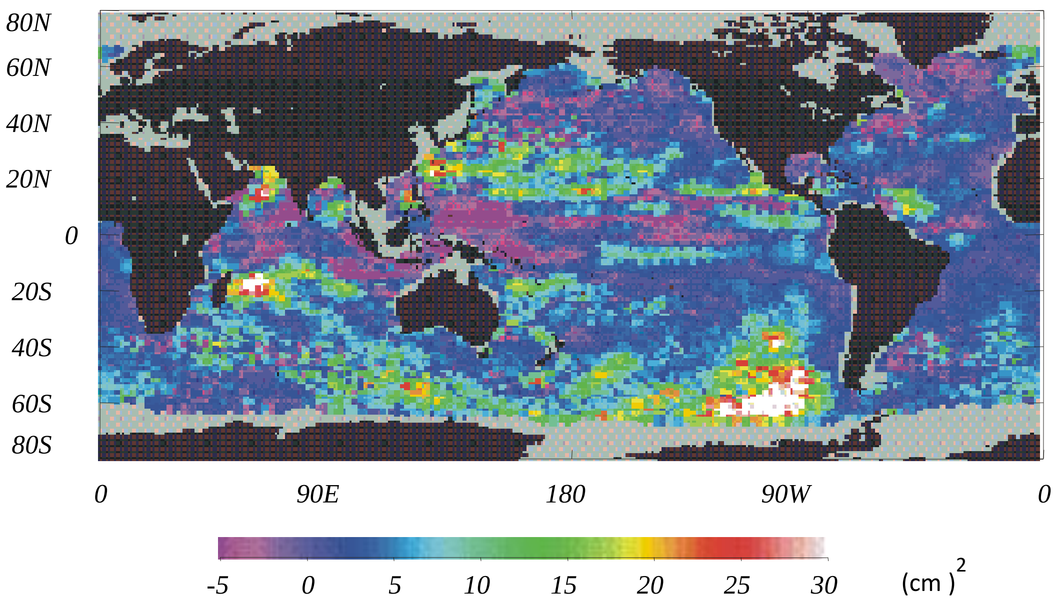
Example estimating $\mathbf{R}$

Error estimates comparing altimetry data with a coarse resolution OGCM.

$$\mathbf{R} \equiv \langle \mathbf{n}\mathbf{n}^T \rangle = \langle \boldsymbol{y}\boldsymbol{y}^T \rangle - \langle \boldsymbol{y}\mathbf{m}^T \rangle$$



$$\mathbf{H}\mathbf{P}_{sim}{}^{T}\,\mathbf{H}^T = \langle \mathbf{p}\mathbf{p}^T \rangle = \langle \mathbf{m}\mathbf{m}^T \rangle - \langle \boldsymbol{y}\mathbf{m}^T \rangle$$



[*Fukumori et al., 1999*]

Summary

1. State estimation is an inverse problem,
2. Various inverse methods are available,
   - Minimum Variance
   - Least-Squares
3. Kalman filter & RTS smoother are recursive minimum variance estimates,
4. Adjoint method is an iterative least-squares estimate,
5. Practical consideration of errors.

Jet Propulsion Laboratory
California Institute of Technology

© 2025 California Institute of Technology.
Government sponsorship acknowledged