Jet Propulsion Laboratory
California Institute of Technology

State Estimation

Part 1: Basic Machinery

Ichiro Fukumori
Jet Propulsion Laboratory, Caltech

ECCO Summer School 2019

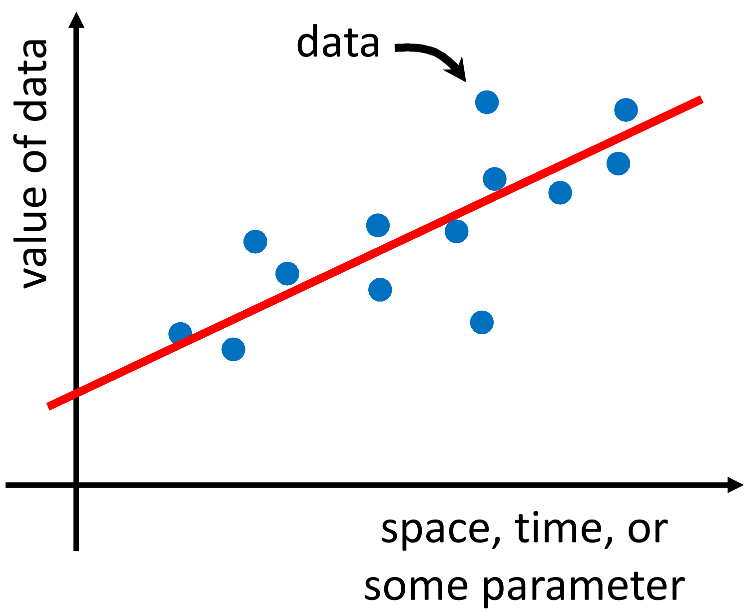
Scope of Lecture

State estimation (data assimilation) is about combining observations and models, but what is it actually doing?

- How is it done?
- What good is it?
- What use does it have?
- Are there caveats?
- What research issues are there?
- How best to use state estimation?
- Where to turn to to learn more?

What is State Estimation?

State estimation (data assimilation) is
a means to analyze observations using models,
equivalent to fitting a curve through data.

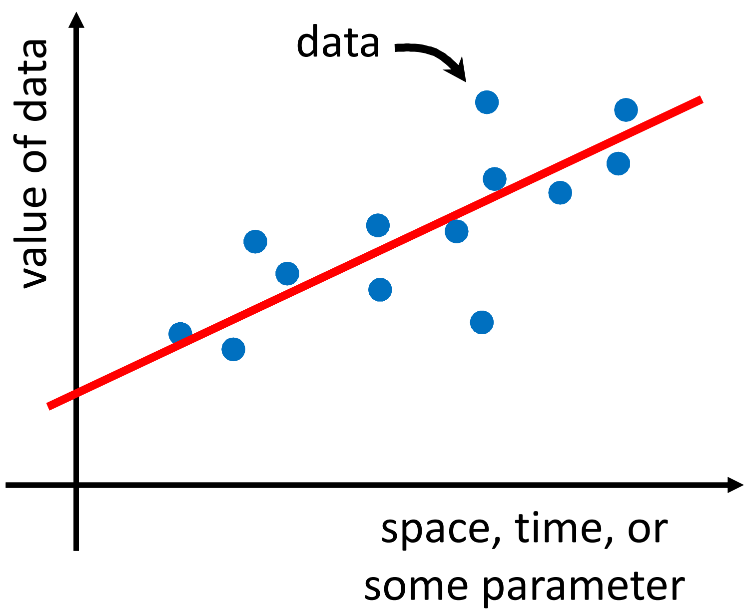


Purpose of curve fitting

- Filter out noise in the data to more accurately describe the system and to gain insight into underlying processes,
- Interpolate/extrapolate the data to aspects not directly measured,
- Test theories against observations.

What is State Estimation?

State estimation (data assimilation) is
a means to analyze observations using models,
equivalent to fitting a curve through data.



Purpose of ~~curve fitting~~ state estimation

- Filter out noise in the data to more accurately describe the system and to gain insight into underlying processes,
- Interpolate/extrapolate the data to aspects not directly measured,
- Test theories against observations.

Overview of the Lectures

1. State estimation is an inverse problem,
2. Estimation theory provides a framework to solving the problem,
3. Approximations and assumptions dictate what is being solved.

Lecture Outline

1. Basic Machinery (this lecture)

   The mathematical problem (inverse problem), Linear inverse methods, Singular value decomposition (SVD), Rank deficiency, Gauss-Markov theorem, Minimum variance estimate, Least-squares,

2. Methods of state estimation (tomorrow)

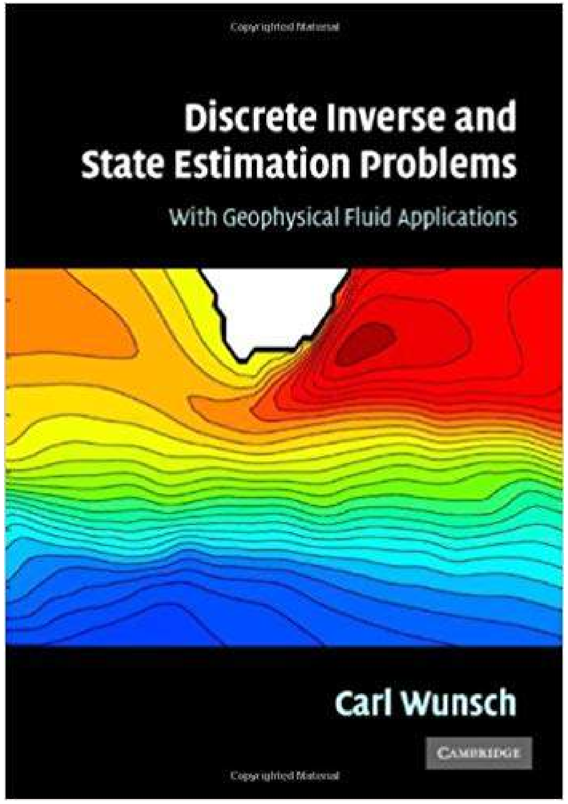
   Kalman filter, Rauch-Tung-Striebel smoother, Adjoint method,

3. Practical Matters (Saturday)

   Error estimation, representation error, covariance, approximate Kalman filters, other data assimilation methods (Optimal Interpolation, 3DVAR).

Reference

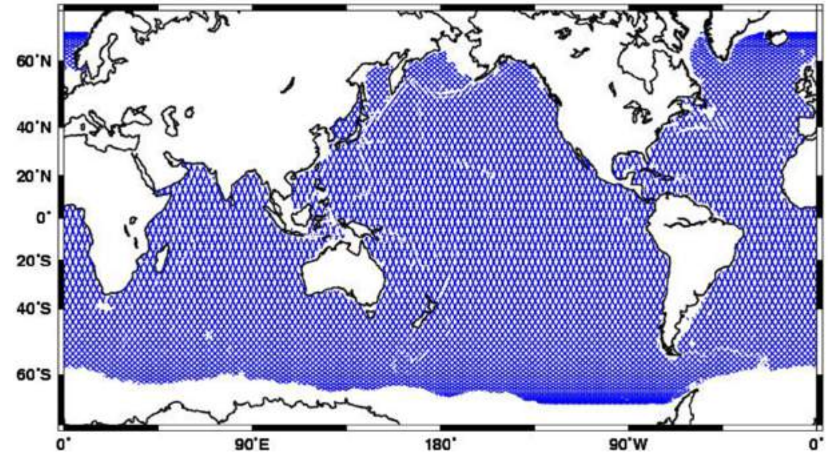
Wunsch, C., 2006: *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications.* Cambridge University Press, 371 pp.



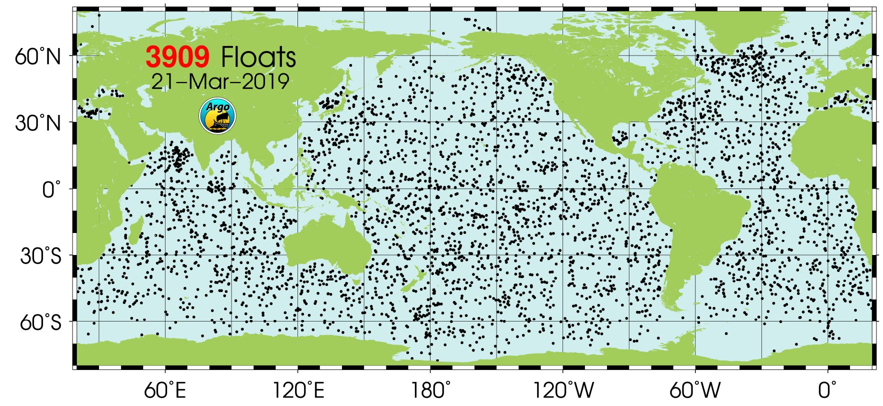
Ocean Observations

Observations are sparse, intermittent, irregular, noisy and limited in what can be measured.

Jason: Global sea level measurements every 10-days with a 300km cross-track distance at the Equator



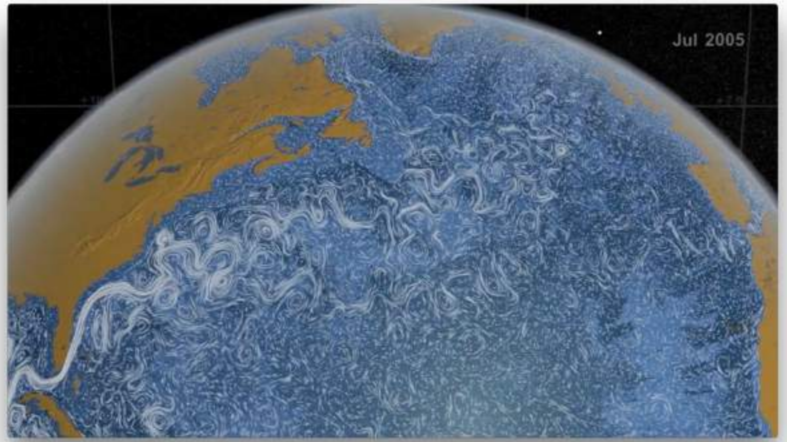
Argo: TS profile down to 2000m depth, once every 10-days & every 300 km

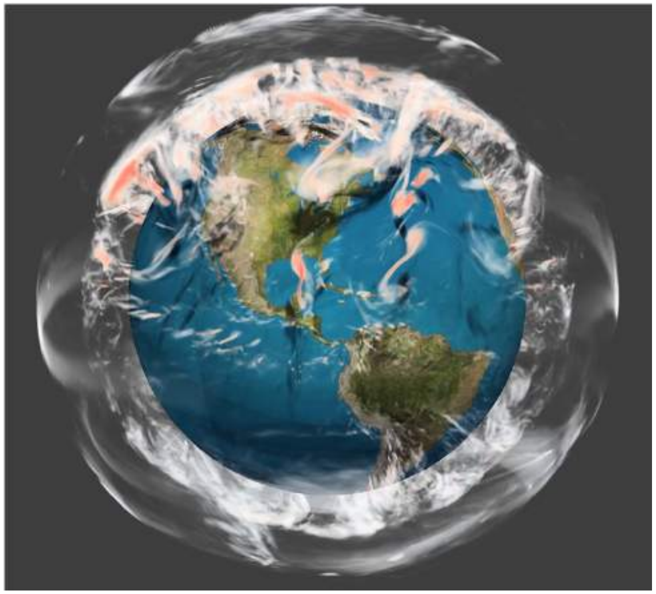


Models

General circulation models provide complete descriptions of the ocean, motivating their use as a "curve" to fit the observations.

"Perpetual Ocean"
ECCO2 model simulation of
surface current (drifter tracks)





Atmospheric Reanalyses: Combines observations with weather forecasting models to yield the most complete description of the global atmosphere. e.g., ERA-5 relative vorticity (FZ Juelich)

State Estimation

State estimation (data assimilation) is about combining observations with models so as to

a) Reconcile diverse measurements into complete and coherent descriptions of the entire ocean,

b) Improve the accuracy of the model.

Mathematically, the problem is an *inverse problem* and is most commonly solved by *least-squares*.

The Mathematical Problem

It is instructive to describe the problem mathematically to gain insight into what combining model and data is about.

Relating observations to model

model ... data

$$\mathbf{H}_t \mathbf{x}_t \approx \hat{\mathbf{y}}_t \qquad t = 1, M$$

$\mathbf{H}_t$: Observation matrix; $\mathbf{x}_t$: state vector; $\hat{\mathbf{y}}_t$: data vector

state & data vectors:

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

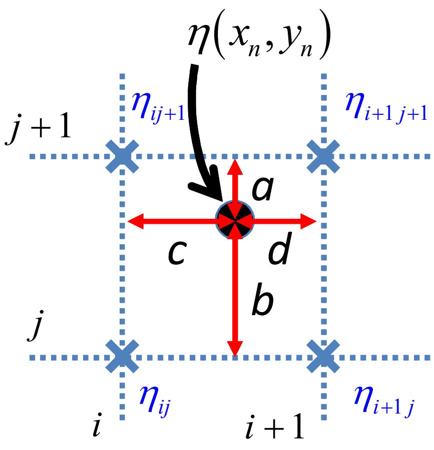
u_{ijk}: model grid point

(x_n, y_n): location of observation

u, v: zonal & meridional velocity
T, S: temperature & salinity
η: sea level

The Mathematical Problem

e.g., bilinear interpolation



$$\eta(x_n, y_n) \equiv w_1\eta_{ij} + w_2\eta_{i+1\,j} + w_3\eta_{ij+1} + w_4\eta_{i+1\,j+1}$$

where $w_1 = \frac{a}{a+b}\frac{d}{c+d}$ $\quad w_2 = \frac{a}{a+b}\frac{c}{c+d}$ etc

Then, corresponding row of

$\mathbf{H}_t\mathbf{x}_t \approx \hat{\mathbf{y}}_t$ is

$$\begin{pmatrix} & & & & \vdots & & & & \\ \cdots & w_1 & \cdots & w_2 & \cdots & w_3 & \cdots & w_4 & \cdots \\ & & & & \vdots & & & & \end{pmatrix} \begin{pmatrix} \vdots \\ \eta_{ij} \\ \vdots \\ \eta_{i+1\,j} \\ \vdots \\ \eta_{ij+1} \\ \vdots \\ \eta_{i+1\,j+1} \\ \vdots \end{pmatrix}_t \approx \begin{pmatrix} \vdots \\ \hat{\eta}(x_n, y_n) \\ \vdots \end{pmatrix}_t$$

The Mathematical Problem

Model $\quad \mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t \qquad t = 0, M-1$

state transition matrix ($\mathbf{A}$)

forcing matrix ($\mathbf{G}$)

control vector ($\mathbf{u}_t$)
(external forcing and other inhomogeneous terms including errors of the model, i.e., errors of this relationship)

state & control vectors:

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

- ${}^x\tau$, ${}^y\tau$: zonal & meridional wind stress
- q: heat flux, e: evaporation,
- r: precipitation

The Mathematical Problem

Model $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t$

e.g., temperature equation $\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x}\cdots + \kappa\frac{\partial^2 T}{\partial z^2}\cdots = q$

linearizing around background state $\frac{\partial T}{\partial t} + \bar{u}\frac{\partial T}{\partial x} + u\frac{\partial \bar{T}}{\partial x}\cdots + \kappa\frac{\partial^2 T}{\partial z^2}\cdots = q$

$$\frac{\left(T_{ijk}\right)_{t+1} - \left(T_{ijk}\right)_t}{\Delta t} + \left(\bar{u}_{ijk}\right)_t \frac{\left(T_{i+1\,jk}\right)_t - \left(T_{i-1\,jk}\right)_t}{2\Delta x} + \left(u_{ijk}\right)_t \left(\frac{\partial \bar{T}}{\partial x}\right)_t \cdots$$

$$+\kappa\frac{\left(T_{ijk+1}\right)_t - 2\left(T_{ijk}\right)_t + \left(T_{ijk-1}\right)_t}{\Delta z^2}\cdots = \left(q_{ijk}\right)_t$$

$$\left(T_{ijk}\right)_{t+1} = \left(1 + 2\kappa\frac{\Delta t}{\Delta z^2}\right)\left(T_{ijk}\right)_t + \left(\bar{u}_{ijk}\right)_t \frac{\Delta t}{2\Delta x}\left(T_{i-1\,jk}\right)_t - \left(\bar{u}_{ijk}\right)_t \frac{\Delta t}{2\Delta x}\left(T_{i+1\,jk}\right)_t$$

$$-\kappa\frac{\Delta t}{\Delta z^2}\left(T_{ijk+1}\right)_t - \kappa\frac{\Delta t}{\Delta z^2}\left(T_{ijk-1}\right)_t + \Delta t\left(\frac{\partial \bar{T}}{\partial x}_{ijk}\right)_t \left(u_{ijk}\right)_t + \Delta t\left(q_{ijk}\right)_t + \cdots$$

The Mathematical Problem

Model $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t$

$$\left(T_{ijk}\right)_{t+1} = \left(1 + 2\kappa\frac{\Delta t}{\Delta z^2}\right)\left(T_{ijk}\right)_t + \left(\bar{u}_{ijk}\right)_t \frac{\Delta t}{2\Delta x}\left(T_{i-1\,jk}\right)_t - \left(\bar{u}_{ijk}\right)_t \frac{\Delta t}{2\Delta x}\left(T_{i+1\,jk}\right)_t$$

$$- \kappa\frac{\Delta t}{\Delta z^2}\left(T_{ijk+1}\right)_t - \kappa\frac{\Delta t}{\Delta z^2}\left(T_{ijk-1}\right)_t + \Delta t\left(\frac{\partial \bar{T}}{\partial x}_{ijk}\right)_t\left(u_{ijk}\right)_t + \Delta t\left(q_{ijk}\right)_t + \cdots$$

$$\begin{pmatrix} \vdots \\ T_{ijk} \\ \vdots \end{pmatrix}_{t+1} = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & -\kappa\frac{\Delta t}{\Delta z^2} & \left(u_{ijk}\right)_t\frac{\Delta t}{2\Delta x} & 1+2\kappa\frac{\Delta t}{\Delta z^2} & -\left(u_{ijk}\right)_t\frac{\Delta t}{2\Delta x} & -\kappa\frac{\Delta t}{\Delta z^2} & \Delta t\left(\frac{\partial \bar{T}}{\partial x}_{ijk}\right)_t & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ T_{ijk-1} \\ T_{i-1\,jk} \\ T_{ijk} \\ T_{i+1\,jk} \\ T_{ijk+1} \\ u_{ijk} \\ \vdots \end{pmatrix}_t$$

$$+ \begin{pmatrix} \vdots & \vdots & \vdots \\ \cdots & \Delta t & \cdots \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ q_{ijk} \\ \vdots \end{pmatrix}_t$$

The Mathematical Problem

Model $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t$

second-order time-stepping (e.g., Adams-Bashforth)

$$\mathbf{x}_{t+1} = {}^{1}\mathbf{A}\mathbf{x}_t + {}^{2}\mathbf{A}\mathbf{x}_{t-1} + {}^{1}\mathbf{G}\mathbf{u}_t + {}^{2}\mathbf{G}\mathbf{u}_{t-1}$$

$$\begin{pmatrix} \mathbf{x}_{t+1} \\ \mathbf{x}_t \end{pmatrix} = \begin{pmatrix} {}^{1}\mathbf{A} & {}^{2}\mathbf{A} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \end{pmatrix} + \begin{pmatrix} {}^{1}\mathbf{G} & {}^{2}\mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{u}_t \\ \mathbf{u}_{t-1} \end{pmatrix}$$

multiple time-steps

$$\mathbf{x}_{t+\Delta t} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t = \mathbf{A}\left(\mathbf{A}\mathbf{x}_{t-\Delta t} + \mathbf{G}\mathbf{u}_{t-\Delta t}\right) + \mathbf{G}\mathbf{u}_t$$

$$= \mathbf{A}^2\mathbf{x}_{t-\Delta t} + \begin{pmatrix} \mathbf{A} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{u}_{t-\Delta t} \\ \mathbf{u}_t \end{pmatrix}$$

$$= \mathbf{A}^{n+1}\mathbf{x}_{t-n\Delta t} + \begin{pmatrix} \mathbf{A}^{\mathbf{n}} & \mathbf{A}^{\mathbf{n-1}} & \cdots & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{G} & & & \\ & \mathbf{G} & & \\ & & \ddots & \\ & & & \mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{u}_{t-n\Delta t} \\ \mathbf{u}_{t-(n-1)\Delta t} \\ \vdots \\ \mathbf{u}_t \end{pmatrix}$$

The Mathematical Problem

Given observations $\mathbf{y}$ what is the ocean state $\mathbf{x}$?

Observations $\mathbf{H}_t \mathbf{x}_t \approx \hat{\mathbf{y}}_t \qquad t = 1, M$

Model $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t \qquad t = 0, M-1$

data ↘

$$\hat{\mathbf{y}}_t \equiv \begin{pmatrix} \vdots \\ \hat{\eta}(x_n, y_n) \\ \vdots \\ \hat{T}(x_m, y_m, z_m) \\ \vdots \end{pmatrix}_t$$

state ↘

$$\mathbf{x}_t \equiv \begin{pmatrix} \vdots \\ u_{ijk} \\ v_{ijk} \\ T_{ijk} \\ S_{ijk} \\ \eta_{ij} \\ \vdots \end{pmatrix}_t$$

control ↘

$$\mathbf{u}_t = \begin{pmatrix} \vdots \\ {}^x\tau_{ij} \\ {}^y\tau_{ij} \\ q_{ij} \\ e_{ij} \\ r_{ij} \\ \vdots \end{pmatrix}_t$$

The Mathematical Problem

$$\begin{pmatrix} \vdots \\ \mathbf{Hx}_t \\ \mathbf{Hx}_{t+1} \\ \vdots \\ \mathbf{x}_{t+1}-\mathbf{Ax}_t-\mathbf{Gu}_t \\ \mathbf{x}_{t+2}-\mathbf{Ax}_{t+1}-\mathbf{Gu}_{t+1} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \hat{\mathbf{y}}_t \\ \hat{\mathbf{y}}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

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

Combining models and data
is mathematically
an inverse problem

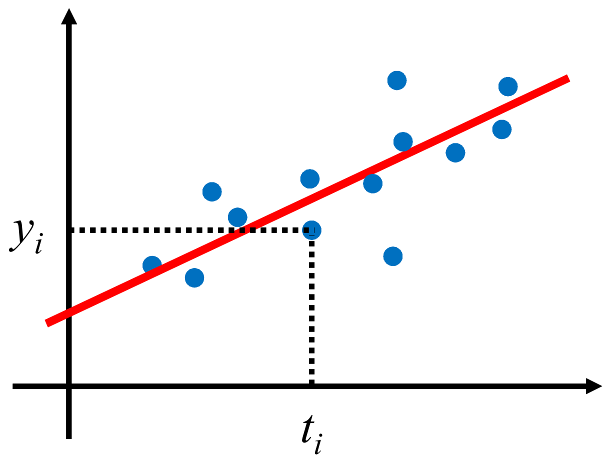
$$\mathbf{Ex} \approx \hat{\mathbf{y}}$$

The Mathematical Problem

Linear Inverse Problem $\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}}$

Given matrix $\mathbf{E}$ and vector $\hat{\mathbf{y}}$ what is vector $\mathbf{x}$?

e.g., fitting a line through data $y = at + b$



$$\begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \approx \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

The Mathematical Problem

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

$$\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}}$$

$$M \times N$$

$$M \ll N$$

There are always more unknowns (number of elements) than knowns (number of data), rendering inverse problems (state estimation) mathematically ill-posed; i.e., there is no unique solution. One needs to change what it means to solve a problem, recognizing what is resolved and what is not.

The Mathematical Problem

Line-fitting is also fundamentally an ill-posed problem, as typically no solution exactly satisfies the problem when using observations.

$$\begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \approx \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

Explicitly writing misfits $r_1, r_2 \cdots r_n$ where $r_i = \hat{y}_i - (at_i + b)$
the problem is mathematically

$$\begin{pmatrix} t_1 & 1 & 1 & & & \\ t_1 & 1 & & 1 & & \\ \vdots & \vdots & & & \ddots & \\ t_1 & 1 & & & & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

The Mathematical Problem

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

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

Wunsch (1977, Science)

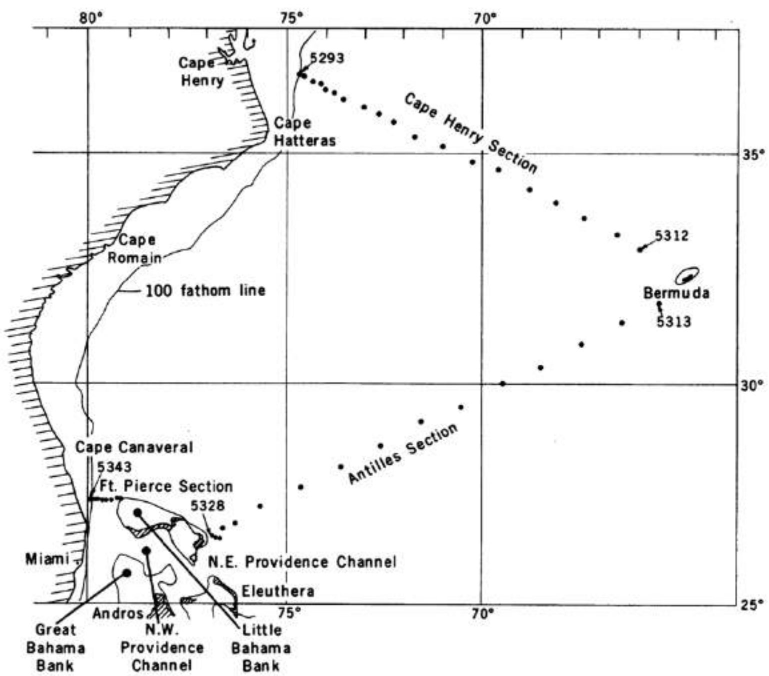


Fig. 1. Locations of *Atlantis* 215 stations used here.

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

$$A_{ij} = \Delta p_{ij}\, \Delta x_j \qquad (3)$$

b is the $M \times 1$ column vector of barotropic velocities, and Γ is the $N \times 1$ column vector

$$\Gamma_i = \sum_{j=1}^{M} \bar{v}_{ij}\, \Delta p_{ij}\, \Delta x_j \qquad (4)$$

representing the imbalance of properties

Singular Value Decomposition (SVD)

Inverse Problem $\quad \mathbf{Ex} \approx \hat{\mathbf{y}}$

The singular value decomposition of matrix $\mathbf{E}$ is useful in gaining insight into the problem and its solution.

Singular Value Decomposition (SVD)

$M \times N$ matrix $\quad \mathbf{E} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^T$

where

$\mathbf{u}_i^T \mathbf{u}_j = \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$ *singular vectors*

M-vector $\quad N$-vector

$\lambda_i \geq \lambda_j > 0$ for $i < j$ *singular values*

$r \leq \min(M, N)$ *rank*

$$\mathbf{E}\mathbf{v}_i = \lambda_i \mathbf{u}_i$$

$$\mathbf{u}_i^T \mathbf{E} = \lambda_i \mathbf{v}_i^T$$

EOFs are Singular Vectors

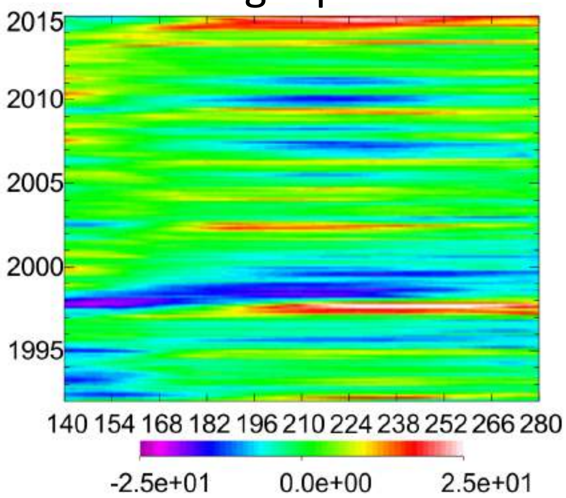
SVD $$\mathbf{E} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^T$$

Empirical orthogonal functions and principal components. are singular vectors of data. Geometrically $\mathbf{u}_k$ ($\mathbf{v}_k$) can be interpreted as the most common structure among the columns (rows) of $\mathbf{E}$ after $\mathbf{u}_i$ ($\mathbf{v}_i$) $i=1,k-1$.

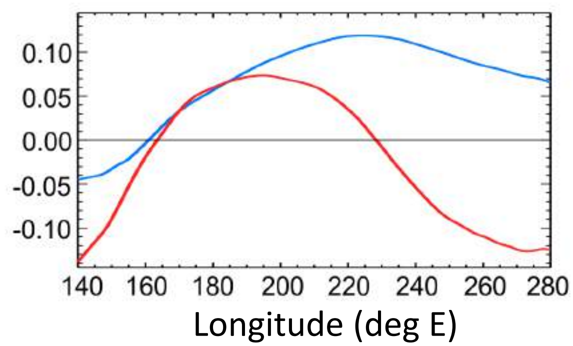
$$\mathbf{E}\mathbf{E}^T\mathbf{u}_i = \lambda_i^2\mathbf{u}_i$$

$$\mathbf{E}^T\mathbf{E}\mathbf{v}_i = \lambda_i^2\mathbf{v}_i$$

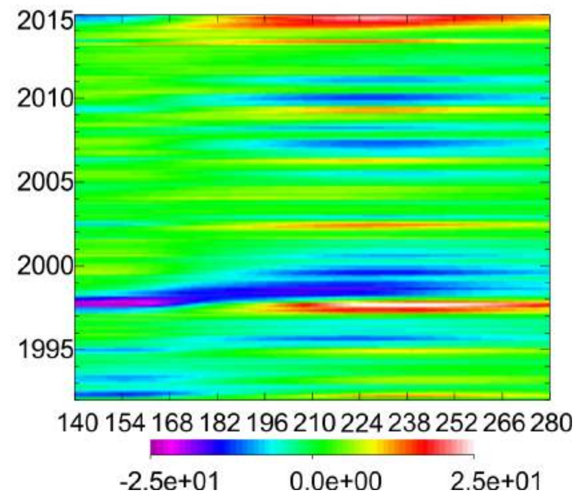
e.g., sea level anomaly along Equator



First two EOFs



Reconstruction with first two EOFs



SVD Inversion

Inverse Problem $\mathbf{Ex} \approx \hat{\mathbf{y}}$ where. $\mathbf{E}$ is a $M \times N$ matrix with

Singular Value Decomposition $\mathbf{E} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^T$ $\quad \mathbf{u}_i^T \mathbf{u}_j = \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$

$r \le \min(M, N)$

Let $\mathbf{x} = \sum_{i=1}^{N} a_i \mathbf{v}_i$ and solve for $a_i \quad i = 1, N$

By substitution $$\mathbf{Ex} = \left(\sum_i^r \lambda_i \mathbf{u}_i \mathbf{v}_i^T \right) \left(\sum_i^N a_i \mathbf{v}_i \right) = \sum_i^r a_i \lambda_i \mathbf{u}_i \approx \hat{\mathbf{y}}$$

Left multiply by $\mathbf{u}_k^T$ with $k = 1, r$ yields $\sum_i^r a_i \lambda_i \mathbf{u}_k^T \mathbf{u}_i = a_k \lambda_k \approx \mathbf{u}_k^T \hat{\mathbf{y}}$

Therefore, $$a_k = \frac{\mathbf{u}_k^T \hat{\mathbf{y}}}{\lambda_k} \qquad \text{for} \quad k = 1, r$$

a_i for $i = r+1, N$ remain undetermined, but they have no bearing on the inverse problem and, therefore, could be chosen arbitrarily;

i.e., there is an infinite number of possible solutions.

SVD Inversion

Inverse Problem $\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}}$ where. $\mathbf{E}$ is a $M \times N$ matrix with

Singular Value Decomposition $\mathbf{E} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^T$ $\quad \mathbf{u}_i^T \mathbf{u}_j = \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$

$r \le \min(M, N)$

We have $\mathbf{x} = \sum_{i=1}^{N} a_i \mathbf{v}_i$ where $a_k = \dfrac{\mathbf{u}_k^T \hat{\mathbf{y}}}{\lambda_k}$ for $k = 1, r$

How can we choose a_k for $k = r+1, N$?

One approach is to seek the "simplest" solution; "Ockham's razor"

Set $a_k = 0$ for $k = r+1, N$ and choose

$$\hat{\mathbf{x}} = \sum_{i=1}^{r} a_i \mathbf{v}_i = \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \hat{\mathbf{y}}}{\lambda_i} \mathbf{v}_i$$

(*SVD solution*)

As $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{N} a_i^2}$ the SVD solution is also the *Minimum Length Solution*.

Properties of SVD Inversion

1) The SVD solution $\hat{\mathbf{x}} = \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \hat{\mathbf{y}}}{\lambda_i} \mathbf{v}_i$

can also be written as $\hat{\mathbf{x}} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T\hat{\mathbf{y}}$ where

$$\underset{M\times r}{\mathbf{U}} = \left(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r\right) \quad \underset{N\times r}{\mathbf{V}} = \left(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\right) \quad \underset{r\times r}{\mathbf{\Lambda}} = diag\left(\boldsymbol{\lambda}\right) = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_r \end{pmatrix}$$

$$\mathbf{E} = \sum_{i=1}^{r} \lambda \mathbf{u}_i \mathbf{v}_i^T = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \qquad \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \underset{r\times r}{\mathbf{I}} \qquad \mathbf{U}\mathbf{U}^T \neq \mathbf{I}_{M\times M} \quad r < M$$

$$\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}} \qquad \mathbf{V}\mathbf{V}^T \neq \mathbf{I}_{N\times N} \quad r < N$$

a) $\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T$ is the SVD inverse of $\mathbf{E}$.

b) The SVD inverse is equivalent to Moore-Penrose inverse, pseudo-inverse, right-inverse, left-inverse.

Properties of SVD Inversion

2) The SVD solution is identical to ordinary least-squares solution (when the latter exists).

Seek solution that minimizes residual norm of the inverse problem;

$$J = (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})^T (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})$$

By setting $\dfrac{\partial J}{\partial \mathbf{x}} = 0 \quad \Rightarrow \quad \hat{\mathbf{x}} = \left(\mathbf{E}^T\mathbf{E}\right)^{-1} \mathbf{E}^T \hat{\mathbf{y}}$

a) $\left(\mathbf{E}^T\mathbf{E}\right)^{-1} \mathbf{E}^T$ is the left-inverse of $\mathbf{E}$; $\left(\mathbf{E}^T\mathbf{E}\right)^{-1} \mathbf{E}^T \; \mathbf{E} = \mathbf{I}$

b) Equivalence to SVD can be shown by substitution.

$$\mathbf{E} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T$$

$$\left(\mathbf{E}^T\mathbf{E}\right)^{-1} \mathbf{E}^T = \left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T \; \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T\right)^{-1} \mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T = \left(\mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^T\right)^{-1} \mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T$$

$$= \mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}^T \; \mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T$$

Properties of SVD Inversion

3) Error estimate of $\hat{\mathbf{x}} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\hat{\mathbf{y}}$

SVD estimate with error-free data $\bar{\mathbf{x}} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\bar{\mathbf{y}}$

Estimation error due to data error $\hat{\mathbf{x}} - \bar{\mathbf{x}} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\left(\hat{\mathbf{y}} - \bar{\mathbf{y}}\right)$

Estimation error covariance matrix

statistical expectation → $\left\langle (\hat{\mathbf{x}} - \bar{\mathbf{x}})(\hat{\mathbf{x}} - \bar{\mathbf{x}})^T \right\rangle = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T \left\langle (\hat{\mathbf{y}} - \bar{\mathbf{y}})(\hat{\mathbf{y}} - \bar{\mathbf{y}})^T \right\rangle \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T$

Defining $\mathbf{R}_{yy} = \left\langle (\hat{\mathbf{y}} - \bar{\mathbf{y}})(\hat{\mathbf{y}} - \bar{\mathbf{y}})^T \right\rangle$ $= \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{R}_{yy}\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T$

If $\mathbf{R}_{yy} = \sigma_{yy}\mathbf{I}$ $= \sigma_{yy}\mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}^T$

The smaller the singular values, the larger the estimation error;
i.e., there is a trade-off between accuracy & resolution.

4) Row and column weighting changes SVD;

$$\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}} \quad \Rightarrow \quad \mathbf{W}\mathbf{E}\mathbf{S}\left(\mathbf{S}^{-1}\mathbf{x}\right) \approx \mathbf{W}\hat{\mathbf{y}}$$

Summary of Inverse Problem and SVD

a) State estimation (data assimilation) is an inverse problem,

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

$$M \times N$$

$$M \ll N$$

b) Most (all) oceanographic inverse problems are rank deficient (mathematically ill-posed). Choices are made to obtain particular (optimal, objective) solutions; e.g., SVD solution

$$\hat{\mathbf{x}} = \sum_{i=1}^{N} a_i \mathbf{v}_i \qquad \begin{cases} a_k \approx \dfrac{\mathbf{u}_k{}^T \hat{\mathbf{y}}}{\lambda_k} & \text{for} \quad k = 1,\ r \\ a_k = 0 & \text{for} \quad k = r+1,\ N \end{cases}$$

Other Inverse Methods

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

- Minimum Variance Estimate

  aka Gauss-Markov theorem, 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.

Gauss-Markov Theorem

Suppose we estimate $\mathbf{x}$ from $\hat{\mathbf{y}}$ using prior statistical knowledge;

statistical expected value

$$\langle \mathbf{x} \rangle = 0 \qquad \langle \mathbf{x}\mathbf{x}^T \rangle = \mathbf{R}_{xx}$$

$$\langle \mathbf{y} \rangle = 0 \qquad \langle \mathbf{y}\mathbf{y}^T \rangle = \mathbf{R}_{yy} \qquad \langle \mathbf{x}\mathbf{y}^T \rangle = \mathbf{R}_{xy}$$

Seek a linear solution of form $\hat{\mathbf{x}} = \mathbf{B}\hat{\mathbf{y}}$ that would have the least posterior error for *each* of its elements.

Error covariance of $\hat{\mathbf{x}}$

$$\begin{aligned}\mathbf{P}_{\mathbf{xx}} &\equiv \left\langle (\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T \right\rangle = \left\langle (\mathbf{B}\mathbf{y} - \mathbf{x})(\mathbf{B}\mathbf{y} - \mathbf{x})^T \right\rangle \\ &= \mathbf{B}\langle \mathbf{y}\mathbf{y}^T \rangle \mathbf{B}^T - \langle \mathbf{x}\mathbf{y}^T \rangle \mathbf{B}^T - \mathbf{B}\langle \mathbf{y}\mathbf{x}^T \rangle + \langle \mathbf{x}\mathbf{x}^T \rangle \\ &= \mathbf{B}\mathbf{R}_{\mathbf{yy}}\mathbf{B}^T - \mathbf{R}_{\mathbf{xy}}\mathbf{B}^T - \mathbf{B}\mathbf{R}_{\mathbf{xy}}{}^T + \mathbf{R}_{\mathbf{xx}}\end{aligned}$$

Gauss-Markov Theorem

$$\mathbf{P_{xx}} = \mathbf{B}\mathbf{R_{yy}}\mathbf{B}^T - \mathbf{R_{xy}}\mathbf{B}^T - \mathbf{B}\mathbf{R_{xy}}^T + \mathbf{R_{xx}}$$

$$= \left(\mathbf{B} - \mathbf{R_{xy}}\mathbf{R_{yy}}^{-1}\right)\mathbf{R_{yy}}\left(\mathbf{B} - \mathbf{R_{xy}}\mathbf{R_{yy}}^{-1}\right)^T - \mathbf{R_{xy}}\mathbf{R_{yy}}^{-1}\mathbf{R_{xy}}^T + \mathbf{R_{xx}}$$

re-written by "completing the square" $ax^2 + bx = a\left[x + b/2a\right]^2 - b^2/4a$

$$\mathbf{ACA}^T - \mathbf{BA}^T - \mathbf{AB}^T = \left(\mathbf{A} - \mathbf{BC}^{-1}\right)\mathbf{C}\left(\mathbf{A} - \mathbf{BC}^{-1}\right)^T - \mathbf{BC}^{-1}\mathbf{B}^T$$

Thus, choosing $\mathbf{B} = \mathbf{R_{xy}}\mathbf{R_{yy}}^{-1}$ minimizes *all* diagonal elements of $\mathbf{P_{xx}}$ leading to

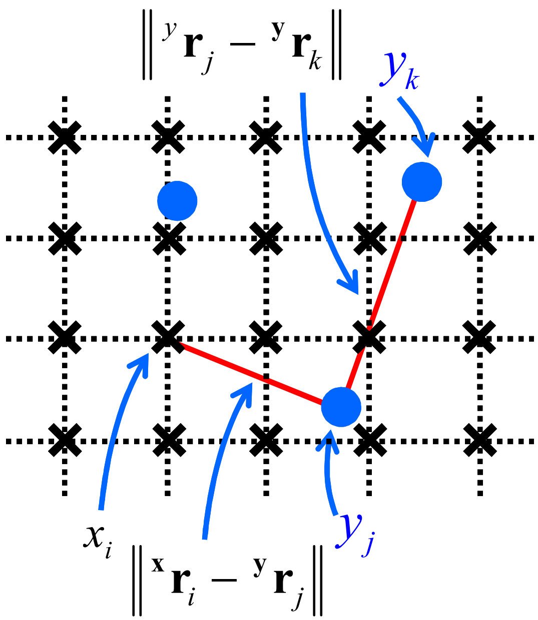
$$\hat{\mathbf{x}} = \mathbf{R_{xy}}\mathbf{R_{yy}}^{-1}\hat{\mathbf{y}}$$

$$\mathbf{P_{xx}} = \mathbf{R_{xx}} - \mathbf{R_{xy}}\mathbf{R_{yy}}^{-1}\mathbf{R_{xy}}^T$$

1) The estimate (GM Estimate) is a Best Linear Unbiased Estimate (BLUE),
2) Errors are reduced from prior estimates by information from $\mathbf{y}$ (2nd term in $\mathbf{P_{xx}}$),
3) Estimate is the basis of objective mapping.

Objective Mapping is a GM Estimate

Objective mapping



Map irregularly sampled observations $\hat{\mathbf{y}}$ to values on a regular grid $\hat{\mathbf{x}}$.

Assuming that the field has a spatially uniform Gaussian covariance function with standard deviation σ and correlation distance λ , and that the observations y have a random white noise of variance n^2 ,

$$\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xy}} \mathbf{R}_{\mathbf{yy}}^{-1} \hat{\mathbf{y}}$$

where

$$\left(\mathbf{R}_{\mathbf{xy}}\right)_{ij} = \sigma^2 \exp\left(-\frac{\left\| {}^{\mathbf{x}}\mathbf{r}_i - {}^{\mathbf{y}}\mathbf{r}_j \right\|^2}{\lambda^2} \right)$$

$$\left(\mathbf{R}_{\mathbf{yy}}\right)_{jk} = \sigma^2 \exp\left(-\frac{\left\| {}^{\mathbf{y}}\mathbf{r}_j - {}^{\mathbf{y}}\mathbf{r}_k \right\|^2}{\lambda^2} \right) + n^2 \delta_{jk}$$

[Bretherton et al., 1976]

Minimum Variance Estimate

Use Gauss-Markov theorem

$$\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{yy}}^{-1}\hat{\mathbf{y}}$$

$$\mathbf{P}_{\mathbf{xx}} = \mathbf{R}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{xy}}^{T}$$

to solve $\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}}$ ⇒ $\mathbf{E}\mathbf{x} + \mathbf{n} = \hat{\mathbf{y}}$ $\quad \langle \mathbf{x}\mathbf{n}^T \rangle = \mathbf{0}$

$$\mathbf{R}_{\mathbf{xy}} = \left\langle \mathbf{x}\left(\mathbf{E}\mathbf{x}+\mathbf{n}\right)^T \right\rangle = \left\langle \mathbf{x}\mathbf{x}^T\mathbf{E}^T \right\rangle = \mathbf{R}_{\mathbf{xx}}\mathbf{E}^T$$

$$\mathbf{R}_{\mathbf{yy}} = \left\langle \left(\mathbf{E}\mathbf{x}+\mathbf{n}\right)\left(\mathbf{E}\mathbf{x}+\mathbf{n}\right)^T \right\rangle = \mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}$$

where

$$\mathbf{R}_{\mathbf{xx}} \equiv \left\langle \mathbf{x}\mathbf{x}^T \right\rangle$$

$$\mathbf{R}_{\mathbf{nn}} \equiv \left\langle \mathbf{n}\mathbf{n}^T \right\rangle$$

Then,

$$\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xx}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}\right)^{-1}\hat{\mathbf{y}}$$

$$\mathbf{P}_{\mathbf{xx}} = \mathbf{R}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}\right)^{-1}\mathbf{E}\mathbf{R}_{\mathbf{xx}}$$

Properties of Minimum Variance Estimate

Minimum Variance Solution of $\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}}$

given prior error $\mathbf{R}_{\mathbf{xx}} = \left\langle \mathbf{x}\mathbf{x}^T \right\rangle \quad \mathbf{R}_{\mathbf{nn}} = \left\langle (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})^T \right\rangle$

is $\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xx}}\mathbf{E}^T \left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}} \right)^{-1} \hat{\mathbf{y}}$

with posterior error $\mathbf{P}_{\mathbf{xx}} = \mathbf{R}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}}\mathbf{E}^T \left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}} \right)^{-1} \mathbf{E}\mathbf{R}_{\mathbf{xx}}$

1) The product $\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T \left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}} \right)^{-1}$ can be regarded as an inversion of $\mathbf{E}$ incorporating prior statistical knowledge,

2) Assumptions about $\mathbf{R}_{\mathbf{xx}}$ $\mathbf{R}_{\mathbf{nn}}$ are not arbitrary. Solution $\hat{\mathbf{x}}$ and residual $\hat{\mathbf{n}} = \hat{\mathbf{y}} - \mathbf{E}\hat{\mathbf{x}}$ must be consistent with these assumptions, otherwise the assumptions (and solution) must be rejected.

3) $\mathbf{n}$ is not simply data error (i.e., error of $\hat{\mathbf{y}}$) but the residual of the inverse problem.

Least-Squares

Find solution to $\mathbf{Ex} \approx \hat{\mathbf{y}}$

symmetric & positive definite weights

that minimizes $J = (\hat{\mathbf{y}} - \mathbf{Ex})^T \mathbf{W}^{-1} (\hat{\mathbf{y}} - \mathbf{Ex}) + \mathbf{x}^T \mathbf{S}^{-1} \mathbf{x}$

- ordinary least-squares $\quad \mathbf{W} = \mathbf{I} \quad \mathbf{S}^{-1} = \mathbf{0}$
- weighted least-squares $\quad \mathbf{W} = diag(\mathbf{w}) \quad \mathbf{S}^{-1} = \mathbf{0}$
- tapered least-squares $\quad \mathbf{S} = diag(\boldsymbol{\gamma})$
- generalized least-squares $\quad (\mathbf{W})_{ij} \neq 0$
- regularized least-squares $\quad (\mathbf{S})_{ij} \neq 0$

Typically, one chooses

$$\mathbf{W} = \mathbf{R}_{\mathbf{nn}} \equiv \langle \mathbf{nn}^T \rangle \qquad \mathbf{S} = \mathbf{R}_{\mathbf{xx}} \equiv \langle \mathbf{xx}^T \rangle$$

$$\mathbf{n} = \hat{\mathbf{y}} - \mathbf{Ex}$$

Why choose inverse error covariance as weights?

By choosing $\mathbf{W} = \mathbf{R}_{\mathbf{nn}} \equiv \left\langle (\mathbf{y} - \mathbf{E}\mathbf{x})(\mathbf{y} - \mathbf{E}\mathbf{x})^T \right\rangle$ $\quad \mathbf{S} = \mathbf{R}_{\mathbf{xx}} \equiv \left\langle \mathbf{x}\mathbf{x}^T \right\rangle$
elements of the scaled least-square problem become normalized (i.e., uncorrelated and equal variance, so elements are on equal footing).

Write Cholesky decomposition $\quad \mathbf{W} = \mathbf{W}^{T/2}\mathbf{W}^{1/2} \quad \mathbf{S} = \mathbf{S}^{T/2}\mathbf{S}^{1/2}$

$$\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1M} \\ & w_{22} & \cdots & w_{2M} \\ & & \ddots & \vdots \\ 0 & & & w_{MM} \end{pmatrix}$$

non-singular
upper triangle matrix

In terms of scaled variables $\quad \mathbf{n}' \equiv \mathbf{W}^{-T/2}(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}) \quad \mathbf{x}' \equiv \mathbf{S}^{-T/2}\mathbf{x}$

elements are uncorrelated and are normalized (unit variance)

$$\left\langle \mathbf{n}'\mathbf{n}'^T \right\rangle = \mathbf{W}^{-T/2} \left\langle (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})^T \right\rangle \mathbf{W}^{-1/2} = \mathbf{W}^{-T/2}\mathbf{W}^{T/2}\mathbf{W}^{1/2}\mathbf{W}^{-1/2} = \mathbf{I}$$

and J becomes

$$J = (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})^T \mathbf{W}^{-1} (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}) + \mathbf{x}^T\mathbf{S}^{-1}\mathbf{x} = \mathbf{n}'^T\mathbf{n}' + \mathbf{x}'^T\mathbf{x}'$$

$$= \sum_i^M n_i'^2 + \sum_i^N x_i'^2$$

Why choose inverse error covariance as weights?

Example of de-correlating variables

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

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

$$\mathbf{n}' = \mathbf{W}^{-T/2}\mathbf{n} = \begin{pmatrix} 1 & 0 \\ -7.0 & 7.1 \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}$$

Instead of having two of the same in original form, the scaled version has just one of them as its variable and the scaled difference between them as another.

Least-Squares

Minimize $$J = \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{x}^T\mathbf{S}^{-1}\mathbf{x}$$

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

Basic notation of vector differentiation

when $\mathbf{A}$ is symmetric

$$\frac{\partial s}{\partial \mathbf{x}} \equiv \begin{pmatrix} \frac{\partial s}{\partial x_1} & \cdots & \frac{\partial s}{\partial x_N} \end{pmatrix}^T \Rightarrow \frac{\partial\left(\mathbf{q}^T\mathbf{r}\right)}{\partial \mathbf{q}} = \frac{\partial\left(\mathbf{r}^T\mathbf{q}\right)}{\partial \mathbf{q}} = \mathbf{r} \qquad \frac{\partial\left(\mathbf{q}^T\mathbf{A}\mathbf{q}\right)}{\partial \mathbf{q}} = \left(\mathbf{A} + \mathbf{A}^T\right)\mathbf{q} = 2\mathbf{A}\mathbf{q}$$

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

Least-Squares

Minimize $$J = \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{x}^T\mathbf{S}^{-1}\mathbf{x}$$

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

$$\frac{\partial\left(\mathbf{q}^T\mathbf{r}\right)}{\partial \mathbf{q}} = \frac{\partial\left(\mathbf{r}^T\mathbf{q}\right)}{\partial \mathbf{q}} = \mathbf{r} \qquad \frac{\partial\left(\mathbf{q}^T\mathbf{A}\mathbf{q}\right)}{\partial \mathbf{q}} = \left(\mathbf{A} + \mathbf{A}^T\right)\mathbf{q} = 2\mathbf{A}\mathbf{q} \qquad \frac{\partial}{\partial \mathbf{q}}\left(\mathbf{B}\mathbf{q}\right) = \mathbf{B}^T \qquad \frac{\partial}{\partial \mathbf{q}}\left(\mathbf{q}^T\mathbf{B}\right) = \mathbf{B}$$

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}} = \frac{1}{2}\frac{\partial\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)}{\partial \mathbf{x}}\frac{\partial}{\partial\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \frac{1}{2}\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{S}^{-1}\mathbf{x}$$

$$= -\mathbf{E}^T\mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{S}^{-1}\mathbf{x}$$

$$= \left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{E} + \mathbf{S}^{-1}\right)\mathbf{x} - \mathbf{E}^T\mathbf{W}^{-1}\hat{\mathbf{y}}$$

Therefore, $$\hat{\mathbf{x}} = \left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{E} + \mathbf{S}^{-1}\right)^{-1}\mathbf{E}^T\mathbf{W}^{-1}\hat{\mathbf{y}}$$

Property of Least-Squares Solution

$$J = \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{x}^T\mathbf{S}^{-1}\mathbf{x}$$

is minimized by $\hat{\mathbf{x}} = \left(\mathbf{E}^T\mathbf{W}^{-1}\mathbf{E} + \mathbf{S}^{-1}\right)^{-1}\mathbf{E}^T\mathbf{W}^{-1}\hat{\mathbf{y}}$

1. When $\mathbf{S}^{-1} = \mathbf{0}$ $\mathbf{W} = \mathbf{I}$ (ordinary least-squares),

$$\hat{\mathbf{x}} = \left(\mathbf{E}^T\mathbf{E}\right)^{-1}\mathbf{E}^T\hat{\mathbf{y}}$$

which reduces to familiar forms in particular examples; e.g.,

If $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} x = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_M \end{pmatrix}$ $\mathbf{E} = \left(1 \quad \cdots \quad 1\right)^T$

then $x = \dfrac{1}{M}\sum_{i=1}^{M}\hat{y}_i$

Property of Least-Squares Solution

$$J = (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x})^T \mathbf{W}^{-1} (\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}) + \mathbf{x}^T \mathbf{S}^{-1} \mathbf{x}$$

is minimized by $\hat{\mathbf{x}} = \left(\mathbf{E}^T \mathbf{W}^{-1} \mathbf{E} + \mathbf{S}^{-1}\right)^{-1} \mathbf{E}^T \mathbf{W}^{-1} \hat{\mathbf{y}}$

2. This solution can also be written as $\hat{\mathbf{x}} = \mathbf{S}\mathbf{E}^T \left(\mathbf{E}\mathbf{S}\mathbf{E}^T + \mathbf{W}\right)^{-1} \hat{\mathbf{y}}$
   using a variant of the "matrix inversion lemma"

$$\mathbf{A}\mathbf{B}^T \left(\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C}\right)^{-1} = \left(\mathbf{B}^T \mathbf{C}^{-1} \mathbf{B} + \mathbf{A}^{-1}\right)^{-1} \mathbf{B}^T \mathbf{C}^{-1}$$

Remarkably, the Least-Squares solution is the same as the Minimum Variance Estimate

$$\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xx}} \mathbf{E}^T \left(\mathbf{E}\mathbf{R}_{\mathbf{xx}} \mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}\right)^{-1} \hat{\mathbf{y}}$$

when $\mathbf{S} = \mathbf{R}_{xx}$ and $\mathbf{W} = \mathbf{R}_{\mathbf{nn}}$ as is usually done.

Property of Least-Squares Solution

$$J = \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{R}_{\mathbf{nn}}^{-1} \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{x}^T \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{x}$$

is minimized by $\hat{\mathbf{x}} = \left(\mathbf{E}^T \mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{E} + \mathbf{R}_{\mathbf{xx}}^{-1}\right)^{-1} \mathbf{E}^T \mathbf{R}_{\mathbf{nn}}^{-1} \hat{\mathbf{y}}$

3. The formal error of the canonical least-squares estimate is therefore,

$$\mathbf{P}_{\mathbf{xx}} = \mathbf{R}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}} \mathbf{E}^T \left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}\right)^{-1} \mathbf{E}\mathbf{R}_{\mathbf{xx}}$$

This can also be written as

$$\mathbf{P}_{\mathbf{xx}} = \left(\mathbf{R}_{\mathbf{xx}}^{-1} + \mathbf{E}^T \mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{E}\right)^{-1}$$

using the “matrix inversion lemma”

$$\left(\mathbf{C}^{-1} + \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B}\right)^{-1} = \mathbf{C} - \mathbf{C}\mathbf{B}^T \left(\mathbf{B}\mathbf{C}\mathbf{B}^T + \mathbf{A}\right)^{-1} \mathbf{B}\mathbf{C}$$

This latter expression of error is the inverse of the Hessian of J ;

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}} = \left(\mathbf{E}^T \mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{E} + \mathbf{R}_{\mathbf{xx}}^{-1}\right)\mathbf{x} - \mathbf{E}^T \mathbf{R}_{\mathbf{nn}}^{-1} \hat{\mathbf{y}}$$

$$\therefore \frac{1}{2} H = \frac{1}{2}\frac{\partial}{\partial \mathbf{x}}\frac{\partial J}{\partial \mathbf{x}} = \left(\mathbf{E}^T \mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{E} + \mathbf{R}_{\mathbf{xx}}^{-1}\right)$$

Summary of GM Inverse and Least-Squares

Solving $\mathbf{E}\mathbf{x} \approx \hat{\mathbf{y}}$ given

$$\mathbf{R}_{\mathbf{xx}} = \left\langle \mathbf{x}\mathbf{x}^T \right\rangle \qquad \mathbf{R}_{\mathbf{nn}} = \left\langle \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \right\rangle$$

a) The *minimum variance solution* (Gauss-Markov inversion) is

$$\hat{\mathbf{x}} = \mathbf{R}_{\mathbf{xx}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R}_{\mathbf{xx}}\mathbf{E}^T + \mathbf{R}_{\mathbf{nn}}\right)^{-1}\hat{\mathbf{y}}$$

b) The least-squares solution minimizing the sum of residual and solution norms weighted by their respective error covariance

$$J = \left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{R}_{\mathbf{nn}}^{-1}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{x}^T\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{x}$$

is *the same* as the *minimum variance solution*.

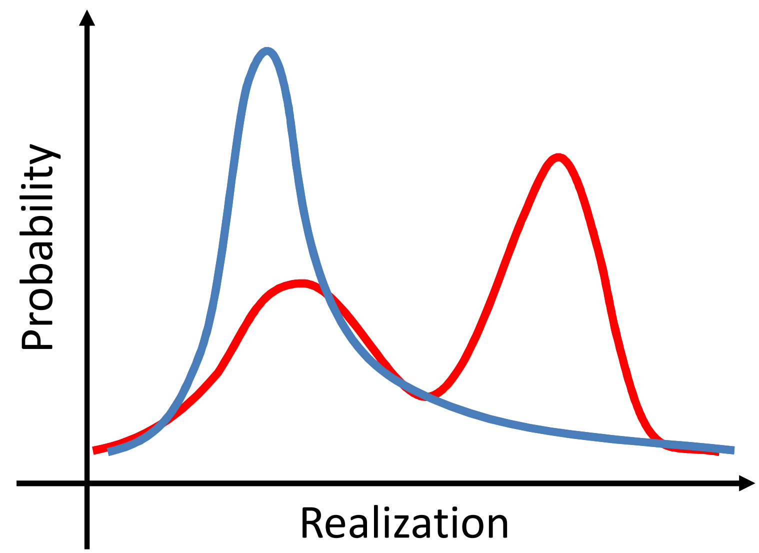
Summary of GM Inverse and Least-Squares

Minimum Variance Estimate $\hat{\mathbf{x}} = \mathbf{R_{xx}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R_{xx}}\mathbf{E}^T + \mathbf{R_{nn}}\right)^{-1}\hat{\mathbf{y}}$

Least-Squares Estimate $\min J \equiv \min\left[\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right)^T \mathbf{R_{nn}}^{-1}\left(\hat{\mathbf{y}} - \mathbf{E}\mathbf{x}\right) + \mathbf{x}^T\mathbf{R_{xx}}^{-1}\mathbf{x}\right]$

c) Neither solution assumes Gaussian probability distribution. The methods above only assumed covariances and should not be confused with Maximum Likelihood Solutions and/or related Bayesian methods that are based on probability distributions.

The solutions are the same when the probability distribution is Gaussian, but are generally different otherwise.



Concluding Remarks (Lecture 1)

1) Combining data and model is mathematically an inverse problem,

2) Inverse problems with data are invariably ill-posed and do not have unique solutions in the strict mathematical sense,

3) Inverse methods provide objective means to obtaining optimal solutions,

    a) Minimum Length (Singular Value Decomposition),
    b) Minimum Variance,
    c) Least-Squares,
    d) Maximum Likelihood,

4) Minimum error variance estimate and least-squares estimate are equivalent.

Next Topic

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

$$M \times N$$

$$M \ll N$$

Typical dimensions of $\mathbf{E}$ in state estimation are $O(10^6 \sim 10^9)$, making direct application of basic inverse methods impractical.

However, the problem can be re-formulated into a series of smaller ones, taking advantage of the problem's structure, and solving them using these basic methods.

- Kalman filter and related smoothers
- Adjoint method

References

Bretherton, F. P., R. E. Davis, and C. B. Fandry, 1976: Technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Research*, **23,** 559-582, doi:10.1016/0011-7471(76)90001-2.

Wunsch, C., 1977: Determining the General Circulation of the Oceans: A Preliminary Discussion. *Science*, **196,** 871-875, doi:10.1126/science.196.4292.871.

Wunsch, C., 2006: *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications.* Cambridge University Press, 371 pp.

Jet Propulsion Laboratory
California Institute of Technology

© 2019 California Institute of Technology.
Government sponsorship acknowledged