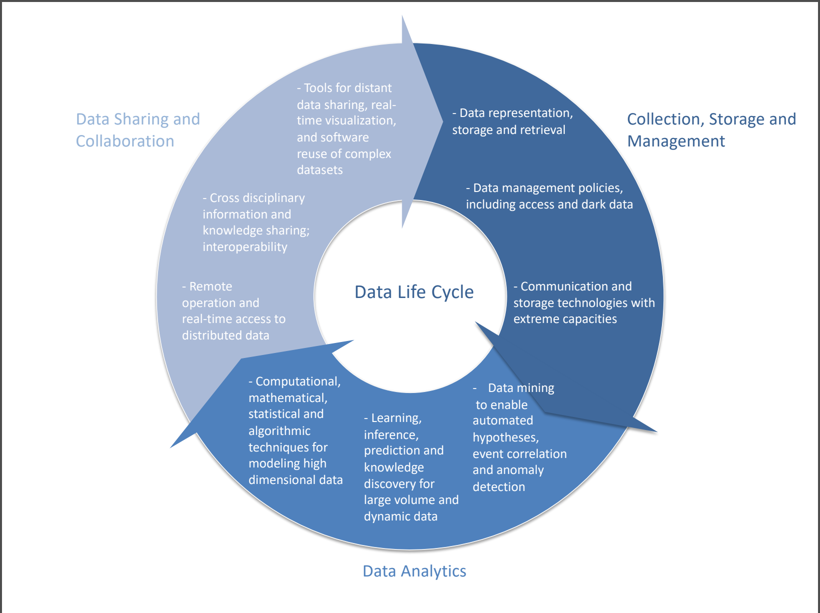
Overview of Data Analysis and Inverse Problems

Farzad Kamalabadi

University of Illinois Urbana-Champaign

Heliophysics Summer School, 2025 – Boulder

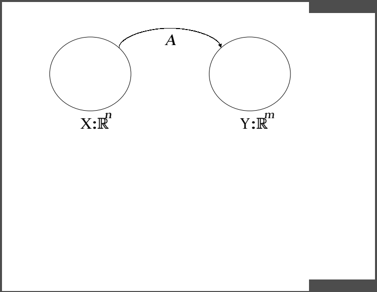
Scope


Data Sharing and Collaboration
- Tools for distant data sharing, real-time visualization, and software reuse of complex datasets
- Cross disciplinary information and knowledge sharing; interoperability
- Remote operation and real-time access to distributed data
Collection, Storage and Management
- Data representation, storage and retrieval
- Data management policies, including access and dark data
- Communication and storage technologies with extreme capacities
Data Life Cycle
Data Analytics
- Computational, mathematical, statistical and algorithmic techniques for modeling high dimensional data
- Learning, inference, prediction and knowledge discovery for large volume and dynamic data
- Data mining to enable automated hypotheses, event correlation and anomaly detection

Outline

1. State/Space Abstraction: Physical Parameters/Observations
2. Integral Equation/Differential Equation Model
3. Taxonomy of Inverse Problems/Solutions
4. Stationary Inverse Problem
   - Deterministic and variational methods
   - Statistical estimation methods
   - Iterative methods
5. Dynamic Inverse Problem
   - The Kalman filter
   - The ensemble Kalman filter (EnKF)
   - Transition to Learning

Examples of Observation/State Mappings



e.g., $y(t) = A(x)$

$\mathbf{x}$	$\mathbf{y}$
Temperature, Density, Composition, Electric Field,	Doppler spectrum
N_e	TEC
Volume Emission rate	Photometric Brightness

State/Observation abstraction

Observation:

- goal: given a set of observations and a forward model relating the state to the observables, determine the state parameters;
- challenges: observability, invertibility

Simulation:

- goal: given initial conditions and forcing parameters, generate simulated state parameters
- challenges: drivers often need to be (inferred) estimated from observation

Learning:

- goal: given a set of observations and the corresponding state parameters, learn the system (forward model)--system identification
- challenges: need sufficient and reliable observations

$$\{x_i(t)\} \rightarrow \mathrm{L}\{\cdot\} \rightarrow \{y_j(t)\}$$

state → observation

$$\{x_i(t)\} \rightarrow \mathrm{L}\{\cdot\} \rightarrow \{y_j(t)\}$$

initial conditions + forcing → simulated state

$$\{x_i(t)\} \rightarrow \mathrm{L}\{\cdot\} \rightarrow \{y_j(t)\}$$

?

- **Task of inference**: given 2 entities in this triplet, estimate (infer statistically) the third. Techniques for accomplishing this task have been in development for a century and continue to gain sophistication.

Integral Equation Model of Inverse Problems

$x(t)$: unknown quantity of interest
$y(t)$: observed (measured) quantity

- General case (nonlinear): $y(t) = h(t, x(\tau))$
- Nonlinear but additive: $y(t) = \int h(t, \tau, x(\tau)) d\tau$
- Linear observations:

$$y(t) = \int h(t; \tau) x(\tau) d\tau \tag{1}$$

FIEFK (when the integral has finite limits)
$h(t; \tau)$: kernel or response function of the system
in general not translation-invariiant

Linear Integral Equations: Examples

• Inverse source problems: Determine source distribution x from measured emitted radiation y:

$$\nabla^2 y + ky = -4\pi x, \quad k = \frac{2\pi}{\lambda} \text{ wave number of emitted radiation} \tag{2}$$

Partial differential equation; integral equation form can be written as:

$$y(r) = \int h(r - r')x(r')dr' \text{ where } h(r) = \frac{e^{\jmath kr}}{r} \tag{3}$$

Linear Integral Equations: Examples

• Heat Equation: Determine $y(s,0) = x(s)$, i.e., initial state, from $y(s,t)$

$$\frac{\partial y(s,t)}{\partial t} = \nabla^2 y(s,t), \quad t > 0 \tag{4}$$

Corresponding integral equation form can be written as:

$$y(s,t) = \int h(s-u,t)x(u))du \text{ where } h(s,t) = \frac{1}{4\pi t}e^{-|s|^2/4t} \tag{5}$$

Inverse problem: deblurring– deconvolution with a Gaussian blur kernel

• Atmospheric turbulence: $h(r,s) = e^{-\pi\alpha^2(r^2+s^2)}$

Linear Integral Equations: Examples (Continued)

- Linear system (signal processing) perspective

$$\sum_k (a_k \frac{d^k}{dt^k}) y(t) = \sum_j (b_j \frac{d^j}{dt^j}) x(t) \tag{6}$$

$$\Rightarrow \frac{Y(s)}{X(s)} = H(s) = \frac{\sum b_k s^k}{\sum a_k s^k} \tag{7}$$

$$y(t) = \int h(t-\tau) x(\tau) d\tau \tag{8}$$

Linear Integral Equations: Examples (Continued)

- Image reconstruction from projections:

$$y_\theta(u) = \int_{-\infty}^{\infty} x(t,s)\delta(t\cos\theta + s\sin\theta - u)dtds. \tag{9}$$

$$h(u,\theta;t,s) = \delta(t\cos\theta + s\sin\theta - u) \tag{10}$$

$$Y(\boldsymbol{r}) = \int_\Omega h(\boldsymbol{r};\boldsymbol{r}')\ X(\boldsymbol{r}')\ d\boldsymbol{r}' \tag{11}$$

For a two-dimensional observation geometry $\boldsymbol{r} = (r,s)$ with r and s denoting the two spatial variables and $\Omega \subset \mathbb{R}^2$ is the region of support. For a three-dimensional observation geometry $\boldsymbol{r} = (r,s,t)$ with t denoting the third spatial variable and $\Omega \subset \mathbb{R}^3$.

Discretized integral equation

For a nonanalytical solution, the unknown field $X(\boldsymbol{r})$ must be discretized. Assuming that the unknown field can be adequately represented by a weighted sum of N basis functions $\{\phi_j(\boldsymbol{r})\}_{j=1}^N$ as follows:

$$X(\boldsymbol{r}) = \sum_{j=1}^{N} x_j \; \phi_j(\boldsymbol{r}) \tag{12}$$

For instance, $\{\phi_j(\boldsymbol{r})\}_{j=1}^N$ are often chosen to be the set of unit height boxes corresponding to a 2-D or 3-D array of pixels. In that case, if a square $g \times f$ pixel array is used, for example, then $N = g \cdot f$ and the discretized field is completely described by the set of coefficients $\{x_j\}_{j=1}^N$, corresponding to the pixel values.

Algebraic form

Collecting all the observations into a vector $\boldsymbol{y}$ of length M, and the unknown image coefficients into a vector $\boldsymbol{x}$ of length N:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} \tag{13}$$

where $\boldsymbol{H} \in \mathbb{R}^{M \times N}$ is the linear operator relating the unknown field to the observations comprised of inner products of the basis functions with the corresponding observation kernel:

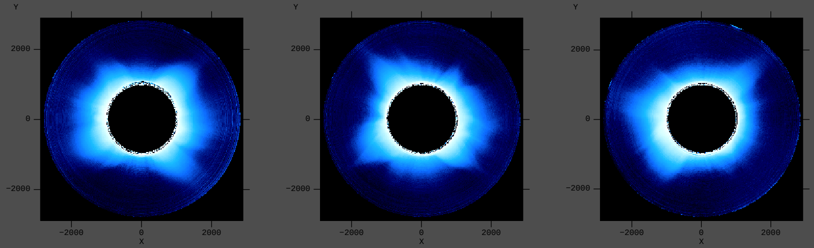
$$(\boldsymbol{H})_{ij} = \int_{\Omega} h_i(\boldsymbol{r}')\phi_j(\boldsymbol{r}')\, d\boldsymbol{r}',\ 1 \leq i \leq M,\ 1 \leq j \leq N \tag{14}$$

where $h_i(\boldsymbol{r}') = h(\boldsymbol{r}_i; \boldsymbol{r}')$ denotes the kernel function corresponding to the i-th observation.

A more complete model including measurement uncertainty component denoted by $\boldsymbol{w}$, i.e.,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w} \tag{15}$$

Example: pB Coronagraph Images

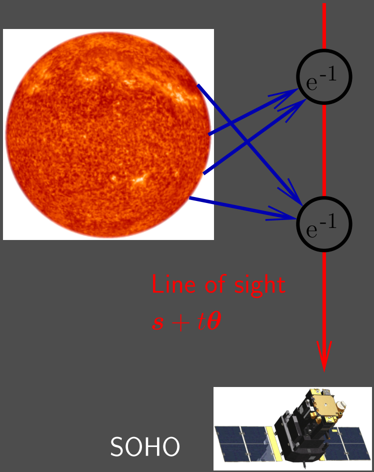


10/16/03   11/01/03   11/20/03

Data source:	Mk4 coronagraph.
Sampling rate:	1 time averaged image/day.
Image resolution:	960×960 pixels (65% are data).
Field of view:	1.12 – 2.79 $R_\odot$.
Pixel size:	4,350 km (at solar surface).

The Relationship Between pB and N_e

Each pixel of a white-light, polarized brightness (pB) coronagraph image is proportional to N_e integrated along the pixel's line of sight.



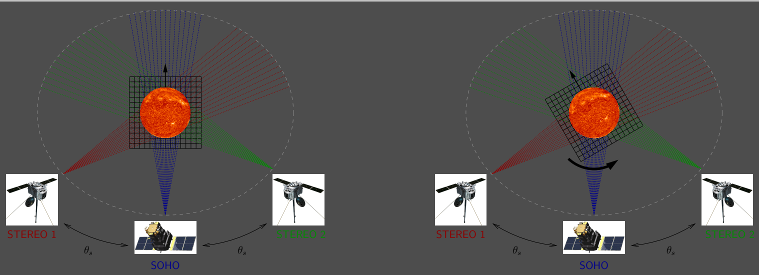
$$\text{pB} = C \int_{-\infty}^{\infty} \mathcal{A}(\boldsymbol{s} + t\boldsymbol{\theta}) N_e(\boldsymbol{s} + t\boldsymbol{\theta}) dt$$

$$\approx C \sum_{j=1}^{N} \mathcal{A}(\boldsymbol{s} + t_j\boldsymbol{\theta}) \boldsymbol{x}_j \Delta t_j$$

$$\Longrightarrow \boldsymbol{y}_i \approx \boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{v}_i$$

Exploitation of Solar Rotation

Solar rotation provides a unique set of line integral measurements of N_e over a 14 day period (corresponding to 180° of rotation).



(a) $\boldsymbol{y}_1 = \boldsymbol{A}_1\boldsymbol{x} + \boldsymbol{v}_1$ (b) $\boldsymbol{y}_2 = \boldsymbol{A}_2\boldsymbol{x} + \boldsymbol{v}_2$

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{14} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_{14} \end{pmatrix} \boldsymbol{x} + \begin{pmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_{14} \end{pmatrix} \implies \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}$$

Mathematical Statement of Inverse Problems

Given $\mathbf{y} \in Y$ and a linear operator $\mathbf{H} : X \to Y$ find $\mathbf{x} \in X$ such that

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{16}$$

• X: Object space
Space where you choose to look for solution
Choice of X encodes prior knowledge

• Y: Data space
Space where observations live
In general $Y \supset \mathbf{H}X$

System Invertibility

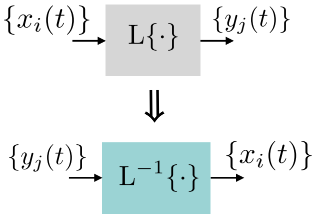
Long history... Hadamard 1915!

The inverse problem y = Lx

is well posed when

- Existence: ∀y, ∃x s.t. y = Lx.
- Uniqueness: Lx1 = Lx2 ⇒ x1 = x2
- Stability: L–1 is continuous.

▵ otherwise... ill posed!



Finite dimensional, linear case

▵ Existence: Least squares

▵ Uniqueness: Minimum norm solution

▵ Stability: Condition number?

Existence, Uniqueness, and Stability

- Case 1: Exact Solution

$N(\mathbf{H}) = \{0\}$; Mapping is injective; full col rank; uniqueness satisfied
$R(\mathbf{H}) = Y$ (Range = Co-domain); Mapping is surjective; full row rank; existence satisfied
$\mathbf{H}$ is square and full rank

- Case 2: Non-existence

$R(\mathbf{H}) \subset Y$; mapping not surjective; overdetermined case; n degrees of freedom and $m > n$ constraints

- Case 3: Non-uniqueness

$N(\mathbf{H}) \neq \{0\}$; mapping not injective; underdetermined case

Least Squares

Notation: l-norm: $||\mathbf{x}||_\mathbf{l} = \sqrt[l]{\sum_i |x_i|^l}$
$||\mathbf{x}||_\mathbf{2} = \sqrt{\sum_\mathbf{i} |\mathbf{x_i}|^\mathbf{2}}$: Usual measure of length
• Idea: Find $\hat{\mathbf{x}}_{\mathrm{LS}}$ that minimizes the length of the error vector $\mathbf{e} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_{\mathrm{LS}}$

$$\arg\min_{\mathbf{x}} ||\mathbf{e}||_2^2 = \arg\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\} \tag{17}$$

$$= \arg\min_{\mathbf{x}} \{(\mathbf{y} - \mathbf{H}\mathbf{x})^{\mathrm{T}}(\mathbf{y} - \mathbf{H}\mathbf{x})\} \tag{18}$$

Solving the minimization problem by setting $\partial/\partial\mathbf{x} = 0$ we arrive at the LS solution:

$$\mathbf{H}^{\mathrm{T}}\mathbf{H}\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{H}^{\mathrm{T}}\mathbf{y} \quad \text{or} \quad \hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y} \tag{19}$$

Weighted Least Squares

• Idea: If the measurements are not equally reliable, attach weights to the errors and minimize $||\mathbf{We}||_2^2 = ||\mathbf{W}(\mathbf{y} - \mathbf{Hx})||_2^2$.
In other words, find the least squares solution to $\mathbf{WHx} = \mathbf{Wy}$.
Solve

$$(\mathbf{WH})^{\mathrm{T}}\mathbf{WH}\hat{\mathbf{x}}_{\mathrm{WLS}} = (\mathbf{WH})^{\mathrm{T}}\mathbf{Wy} \tag{20}$$

$$\mathbf{H}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{WH}\hat{\mathbf{x}}_{\mathrm{WLS}} = \mathbf{H}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{Wy} \tag{21}$$

• Question: What is a rational way of determining an optimal $\mathbf{W}$?
• Approach: Use the knowledge of the average size (or expected value) of e_i, e_i^2, $e_i e_j$

Generalized inverse

A pseudoinverse for the system of equations in (13) may be calculated, for example, via the singular value decomposition (SVD). In terms of the SVD this solution is given by

$$\hat{\boldsymbol{x}}_{\mathsf{SVD}} = \sum_{i=1}^{n} \frac{\boldsymbol{v}_i < \boldsymbol{u}_i^T, \boldsymbol{y} >}{\sigma_i} \tag{22}$$

where n is the rank of $\boldsymbol{H}$; $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$ are the column vectors of the unitary matrices $\boldsymbol{V}$ and $\boldsymbol{U}$, respectively; and σ_i are the singular values or the diagonal elements of Σ in the singular value decomposition of $\boldsymbol{H}$: $\boldsymbol{H} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$. This inversion approach produces the miminum norm, least squares estimate of $\boldsymbol{x}$ and in the absence of noise can produce reasonable reconstructions.

Noise and the generalized inverse

An approach for coping with the ill-conditioning of the system matrix is based on truncating the sum in the SVD reconstruction. This idea is based on the observation that applying the SVD decomposition in the presence of noise to the system of equations in (15) can be shown to yield

$$\hat{\boldsymbol{x}}_{\mathsf{SVD}} = \boldsymbol{x} + \sum_{i=1}^{n} \frac{\boldsymbol{v}_i < \boldsymbol{u}_i^T, \boldsymbol{w} >}{\sigma_i} \tag{23}$$

Since the noise generally has power along all directions and σ_i gets small as i increases, we can see that this solution becomes dominated by noise in the second term. This is a reflection of the ill-conditioning of the problem.

Iterative solutions

The most common approach to reconstructing $\boldsymbol{x}$ is based on iterative inversion of $\boldsymbol{H}$. The reconstruction method consists of making an initial guess, denoted by $\boldsymbol{x}^{(0)}$, followed by cyclically projecting the initial guess onto the hyperplanes defined by the observation equations. The process is repeated until convergence is achieved. More specifically, the following recursion relation is used for the calculation of the object:

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} - \alpha^{(i)} \boldsymbol{H}^T \left(\boldsymbol{H} \boldsymbol{x}^{(i)} - \boldsymbol{y} \right) \tag{24}$$

Starting with an initially smooth solution, the iterations restore the modes that have the highest singular values iteratively.

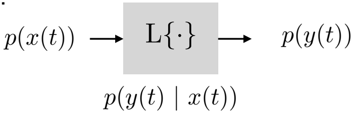
Stochastic System Model

- Ingredients

A stochastic system comprises the following:

- an input random process $\{x_i(t)\}$ characterized by a probability density function;
- an output random process $\{y_j(t)\}$;
- a description of the mapping $L\{.\}$ characterized by a conditional density.

stochastic model:



Stochastic System Model of Observation

A system model of observation:

- input: stochastic characterization of state parameters, e.g., temperature, density, ion composition, electric field, wind;
- output; observed Doppler power spectrum (stochastic quantity);
- a forward model based on conditional density

Estimation rule:

MMSE: $\mathbf{E}(x(t)|y(t))$

MAP: argmax $p(x(t)|y(t))$

$$p(x(t)) \longrightarrow \boxed{\mathrm{L}\{\cdot\}} \longrightarrow p(y(t))$$

state $\quad p(y(t) \mid x(t)) \quad$ observation

$$\Downarrow$$

$$p(y(t)) \longrightarrow \boxed{\mathrm{L}^{-1}\{\cdot\}} \longrightarrow p(x(t))$$

observation $\quad p(x(t) \mid y(t)) \quad$ state

Weighted Least Squares (Statistical Interpretation: ML)

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\max_{\mathbf{x}} \{p(\mathbf{y}|\mathbf{x})\} \tag{25}$$

If noise can be modeled as Gaussian, the conditional probability $p(\mathbf{y}|\mathbf{x})$ is also a Gaussian, with the following mean and covariance,

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &\sim \mathcal{N}(\mathbf{H}\mathbf{x}, \mathbf{R_e}) \\ &= e^{-\frac{1}{2}(\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R_e^{-1}}(\mathbf{y}-\mathbf{H}\mathbf{x})} \end{aligned} \tag{26}$$

the ML estimate takes the following optimization form:

$$\begin{aligned} \hat{\mathbf{x}}_{\mathrm{ML}} &= \arg\max_{\mathbf{x}} \{\ln p(\mathbf{y}|\mathbf{x})\} \\ &= \arg\max_{\mathbf{x}} \{-\frac{1}{2}(\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R_e^{-1}}(\mathbf{y}-\mathbf{H}\mathbf{x})\} \\ &= \arg\min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2_{\mathbf{R_e^{-1}}}\} \end{aligned} \tag{27}$$

Maximum Likelihood

Solving the minimization problem by setting $\partial/\partial\mathbf{x} = 0$ we arrive at the ML solution:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = (\mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{y} \tag{32}$$

The estimation error is defined as:

$$\mathbf{e}_{\mathrm{ML}} = \mathbf{x} - \hat{\mathbf{x}}_{\mathrm{ML}} \tag{33}$$

which can be shown, using substitution and simple algebra, to equal:

$$\mathbf{e}_{\mathrm{ML}} = -(\mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}w \tag{34}$$

Finally, the ML estimation error covariance is given by:

$$\begin{aligned}\mathbf{R}_{\mathrm{ML}} &= E\{\mathbf{e}\mathbf{e}^{\mathrm{T}}\} \\ &= (\mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{H})^{-1}\end{aligned} \tag{35}$$

MAP Estimation and Error Covariance

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R_x}) \tag{36}$$

$$\hat{\mathbf{x}}_{\text{map}} = \underset{\mathbf{x}}{\text{argmax}} \{p(\mathbf{x}|\mathbf{y})\} = \underset{\mathbf{x}}{\text{argmax}} \{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\} \tag{37}$$

$p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$ can be expanded respectively as:

$$p(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{Hx}, \mathbf{R_e}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\mathbf{R_e}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{y} - \mathbf{Hx})^T \mathbf{R_e}^{-1}(\mathbf{y} - \mathbf{Hx}) \tag{38}$$

$$p(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{R_x}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\mathbf{R_x}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}\mathbf{x}^T \mathbf{R_x}^{-1}\mathbf{x}\} \tag{39}$$

MAP Estimation and Error Covariance (continued)

$$\hat{\mathbf{x}}_{\text{map}} = \underset{\mathbf{x}}{\text{argmax}} \; \{-\tfrac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}_{\mathbf{e}}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) - \tfrac{1}{2}\mathbf{x}^T \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}\} \tag{40}$$

$$\hat{\mathbf{x}}_{\text{map}} = \underset{\mathbf{x}}{\text{argmin}} \; \{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2_{\mathbf{R}_{\mathbf{e}}^{-1}} + \|\mathbf{x}\|^2_{\mathbf{R}_{\mathbf{x}}^{-1}}\} \tag{41}$$

It can be shown that the solution to this minimization problem, $\hat{\mathbf{x}}_{\text{map}}$, is given by:

$$\hat{\mathbf{x}}_{\text{map}} = \mathbf{R}_{\hat{\mathbf{x}}_{\text{map}}}(\mathbf{H}^T \mathbf{R}_{\mathbf{e}}^{-1}\mathbf{y}) \tag{42}$$

where $\mathbf{R}_{\hat{\mathbf{x}}_{\text{map}}}$ is the estimation error covariance given by:

$$\mathbf{R}_{\hat{\mathbf{x}}_{\text{map}}} = (\mathbf{R}_{\mathbf{x}}^{-1} + \mathbf{H}^T \mathbf{R}_{\mathbf{e}}^{-1}\mathbf{H})^{-1} \tag{43}$$

Regularization

A general formulation for the cost function (the objective function) can be expressed as:

$$\Phi(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2_{\boldsymbol{\mathcal{W}}} + \sum_i \gamma_i C_i(\boldsymbol{x}) \tag{44}$$

where $\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2_{\boldsymbol{\mathcal{W}}}$ denotes the weighted residual norm, i.e., $(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})^T \boldsymbol{\mathcal{W}} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})$, C_i and γ_i are the i-th regularization functional and regularization parameters respectively, and $\boldsymbol{\mathcal{W}}$ is an appropriate weight, all to be chosen according to the specifics of the problem.

Tikhonov (quadratic) regularization

The most common technique used for regularization and is equivalent to maximum *a posteriori* (MAP) estimation assuming Gaussian statistics for both the unknown image and noise. Assuming $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$ and $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{\Sigma}_{\boldsymbol{x}})$, where $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ represents the normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, the MAP estimate is:

$$\begin{aligned} \hat{\boldsymbol{x}}_{\mathrm{MAP}} &= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \left[-\log p(\boldsymbol{y}|\boldsymbol{x}) - \log p(\boldsymbol{x})\right] \\ &= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \left[\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}} + \|\boldsymbol{x} - \boldsymbol{x}_0\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}}\right] \\ &= \boldsymbol{x}_0 + (\boldsymbol{H}^T\boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}\boldsymbol{H} + \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1})^{-1}\boldsymbol{H}^T\boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}_0) \end{aligned} \tag{45}$$

The connection between this MAP formulation of Tikhonov regularization and the variational form just discussed becomes apparent by assuming independent identically distributed (IID) Gaussian noise and taking $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \frac{1}{\gamma^2}(\boldsymbol{L}^T\boldsymbol{L})^{-1}$, hence arriving at the well known Tikhonov regularization functional:

$$\begin{aligned}
\hat{\boldsymbol{x}}_{\text{Tik}} &= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \left[\frac{1}{\sigma_{\boldsymbol{w}}^2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2 + \gamma^2\|\boldsymbol{L}(\boldsymbol{x} - \boldsymbol{x}_0)\|_2^2\right] \\
&= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \left[\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{L}(\boldsymbol{x} - \boldsymbol{x}_0)\|_2^2\right] \\
&= \boldsymbol{x}_0 + (\frac{1}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{H}^T\boldsymbol{H} + \gamma^2\boldsymbol{L}^T\boldsymbol{L})^{-1}\frac{1}{\sigma_{\boldsymbol{w}}^2}\boldsymbol{H}^T(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}_0) \qquad (46)
\end{aligned}$$

where $\boldsymbol{L}$ is the positive definite regularization matrix and $\lambda = (\gamma\sigma_w)^2$ where σ_w^2 is the variance of the noise samples. A special case is when $\boldsymbol{L} = \boldsymbol{I}$, which results in λ being inverse of the signal-to-noise ratio.

Regularization and its Stochastic Interpretation

$$\hat{\mathbf{x}}_{\mathrm{Tik}} = \underset{\mathbf{x}}{\text{argmin}} \quad \underbrace{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2}_{\text{Data Fidelity}} + \underbrace{\gamma^2 \|\mathbf{L}\mathbf{x}\|_2^2}_{\text{Prior Info}}$$

- Idea: Include prior information into solution
- Interpretations:
  - Add additional constraint: Penalize large values of $\mathbf{L}\mathbf{x}$

(e.g. $\mathbf{L} = \nabla$)
  - Improves conditioning: $(\mathbf{H}^{\mathbf{T}}\mathbf{H} + \gamma^2 \mathbf{L}^{\mathbf{T}}\mathbf{L})\mathbf{x} = \mathbf{H}^{\mathbf{T}}\mathbf{y}$
  - Equivalent to MAP estimate with prior: $p_{\mathbf{X}}(\mathbf{x}) \propto e^{-\gamma^2 \mathbf{x}^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}\mathbf{x}}$
- γ controls tradeoff between data and prior information
- Truncates "$\mathbf{H}^{-1}$" at high frequency

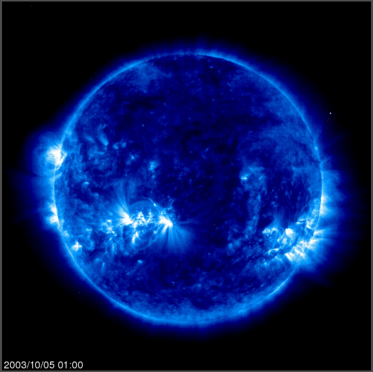
Dynamic Inverse problem: Example

Inputs

Measurements of the Sun at 171Å

2003/10/05

1024×1024 pixels
($M = 1024^2$)



$\boldsymbol{y}_1$

Height/width
$\approx 2 \times 10^6$ km

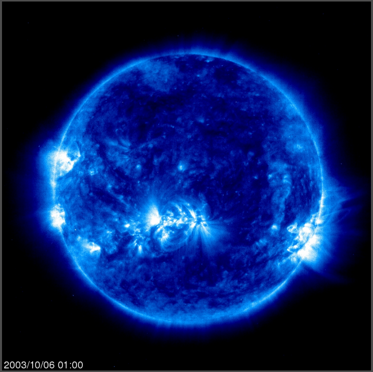
Solar Tomography

Inputs

Measurements of the Sun at 171Å

2003/10/06

1024×1024 pixels
$(M = 1024^2)$



Height/width
$\approx 2 \times 10^6$ km

$\boldsymbol{y}_2$

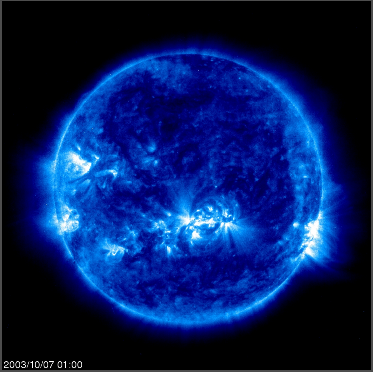
Solar Tomography

Inputs

Measurements of the Sun at 171Å

2003/10/07

1024×1024 pixels
$(M = 1024^2)$



Height/width
$\approx 2 \times 10^6$ km

$\boldsymbol{y}_3$

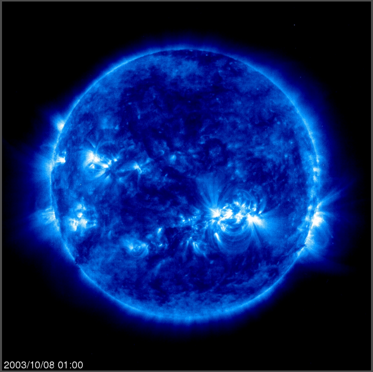
Solar Tomography

Inputs

Measurements of the Sun at 171Å

2003/10/08

1024×1024 pixels
($M = 1024^2$)



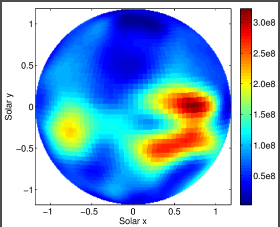
Height/width
$\approx 2 \times 10^6$ km

$\boldsymbol{y}_4$

Solar Tomography

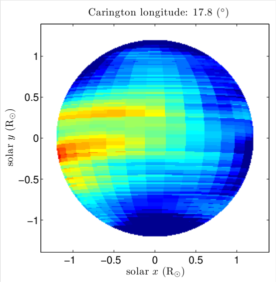
Outputs

(a) "Good" static reconstruction



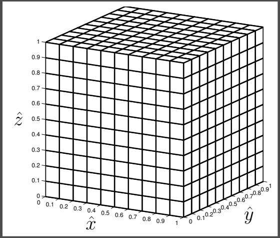
$\widehat{x}_{\text{good}}$

(b) Static reconstruction with "smearing" artifacts



$\widehat{x}_{\text{smeared}}$

(c) A $100 \times 100 \times 100$ voxel grid has one million unknowns ($N = 10^6$)!



3D voxel grid

General State-Space Signal Model

The general hidden Markov model (HMM):

$$\text{Initial prior:} \qquad p_{\boldsymbol{x}_1}(\boldsymbol{x}_1) \tag{47}$$

$$\text{Measurement/forward model:} \qquad h_i(\boldsymbol{y}_i|\boldsymbol{x}_i) \tag{48}$$

$$\text{State-transition model:} \qquad f_i(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i) \tag{49}$$

$$\dim(\boldsymbol{x}_i) = N \qquad \dim(\boldsymbol{y}_i) = M$$

Goal: Compute minimum mean square error (MMSE) estimates of the unknown state $\boldsymbol{x}_i$ given the measurements $\boldsymbol{y}_{1:j} \triangleq \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_j\}$.

$$\widehat{\boldsymbol{x}}_{i|j} \triangleq \mathbb{E}[\boldsymbol{x}_i|\boldsymbol{y}_{1:j}] = \int \boldsymbol{x}_i \, p(\boldsymbol{x}_i|\boldsymbol{y}_{1:j}) \, d\boldsymbol{x}_i \tag{50}$$

Filtering, Smoothing, and Prediction

Estimate ($\widehat{\boldsymbol{x}}_{i\|j}$)	Data ($\boldsymbol{y}_{1:j}$)	Purpose
Filtered	$j = i$	Online processing – estimates are based on currently available data.
Predicted	$j < i$	For forecasting the future evolution of the dynamic process.
Smoothed	$j > i$	Offline processing – estimates based on all available information.

- Smoothed and predicted estimates are typically computed by further processing of the filtered estimates $\widehat{\boldsymbol{x}}_{i|i}$.
- Filtered (posterior) estimates $\widehat{\boldsymbol{x}}_{i|i}$ may be recursively computed using the previous one-step prediction (prior) $\widehat{\boldsymbol{x}}_{i|i-1} \Longrightarrow j \in \{i,\, i-1\}$ for this talk.

Difficulties With The General Model

$$\widehat{\boldsymbol{x}}_{i|i} \triangleq \mathbb{E}[\boldsymbol{x}_i|\boldsymbol{y}_{1:i}] = \int \boldsymbol{x}_i \, p(\boldsymbol{x}_i|\boldsymbol{y}_{1:i}) \, d\boldsymbol{x}_i$$

Problem #1: The posterior PDF $p(\boldsymbol{x}_i|\boldsymbol{y}_{1:i})$ cannot generally be found in closed form. Special cases:

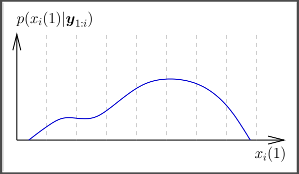
- The linear Gaussian model (discussed next).
- Exponential families with conjugate priors.

(The above list of special cases is nearly exhaustive)

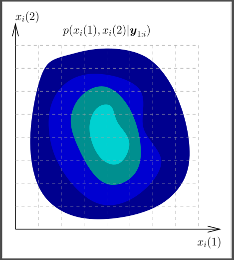
Difficulties With The General Model

$$\widehat{\boldsymbol{x}}_{i|i} \triangleq \mathbb{E}[\boldsymbol{x}_i|\boldsymbol{y}_{1:i}] = \int \boldsymbol{x}_i \, p(\boldsymbol{x}_i|\boldsymbol{y}_{1:i}) \, d\boldsymbol{x}_i$$

Problem #2: Numerical approximation of the posterior and evaluation of the conditional mean requires N dimensional quadrature.



1D – 8 function evaluations



2D – 8^2 function evaluations

Quadrature requires an exponential increase in computational effort with the dimension of the state (i.e., the curse of dimensionality)!

Linear Additive-Noise State-Space Signal Model (Linear Gaussian Model)

$$\text{Initial prior:} \quad \mathbb{E}[\boldsymbol{x}_1] = \boldsymbol{\mu}_1,\ \text{Cov}(\boldsymbol{x}_1) = \boldsymbol{\Pi}_1 \tag{51}$$

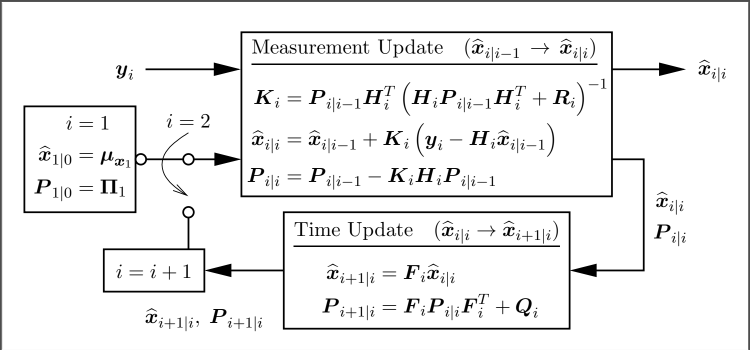
$$\text{Measurement/forward model:} \quad \boldsymbol{y}_i = \boldsymbol{H}_i\, \boldsymbol{x}_i + \boldsymbol{v}_i \tag{52}$$

$$\text{State-transition model:} \quad \boldsymbol{x}_{i+1} = \boldsymbol{F}_i\, \boldsymbol{x}_i + \boldsymbol{u}_i \tag{53}$$

- The second order statistics of the zero mean state ($\boldsymbol{u}_i$) and measurement ($\boldsymbol{v}_i$) noise are given: $\text{Cov}(\boldsymbol{u}_i) = \boldsymbol{Q}_i$ and $\text{Cov}(\boldsymbol{v}_i) = \boldsymbol{R}_i$.

Goal: Compute linear minimum mean square error (LMMSE) estimates of the unknown state $\boldsymbol{x}_i$ given the measurements $\boldsymbol{y}_{1:j}$.

The Kalman Filter



The $N \times N$ matrix $P_{i|j}$ is the estimator error covariance:

$$P_{i|j} \triangleq \text{Cov}(x_i - \widehat{x}_{i|j}) \qquad (54)$$

Almost **2 TB** of computer memory is required to store $P_{i|j}$ when the state dimension $N = 10^6$ and all operations involving $P_{i|j}$ become prohibitively computationally costly!

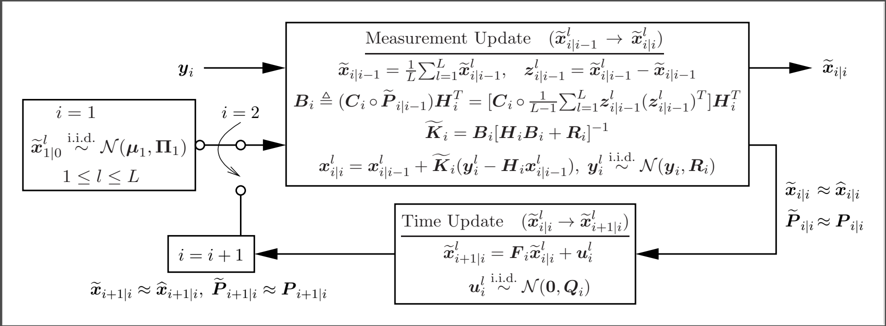
Approximate Kalman Filters

All large-scale methods make some form of dimension reducing approximation to reduce computational complexity.

4DVAR:	Deterministic state-transition model and a low-dimensional parametric model for the prior error covariance $\mathbf{\Pi}_i$. No mechanism to systematically adjust estimates when true system dynamics deviate from the deterministic model.
Banded KF:	Eliminate all bands of the error covariance beyond some distance from the diagonal. Will result in numerical instability.

Ensemble Kalman Filter (EnKF)

Idea: Update an ensemble of samples $\{\widetilde{\boldsymbol{x}}^1_{i|j}, \ldots, \widetilde{\boldsymbol{x}}^L_{i|j}\}$ such that the sample mean and covariance approximate the KF estimate $\widehat{\boldsymbol{x}}_{i|j}$ and estimator error covariance $\boldsymbol{P}_{i|j}$.



The ensemble size L is a trade-off between the estimate quality and computational effort.

Example: Electron Density Reconstruction

- The data source is the Mk4 coronagraph at the Mauna Loa Solar Observatory.

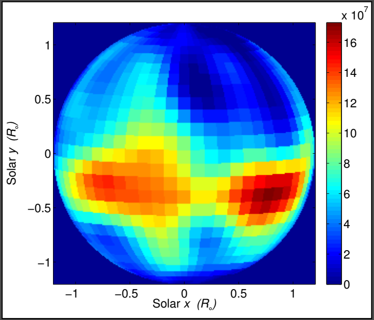
- We show reconstructions of electron density from Oct. 15 – Nov. 24, 2003 (which includes the intense "Halloween" solar storm).

- The ith 3-D static estimate is based on data from day i and 13 days prior to day i.

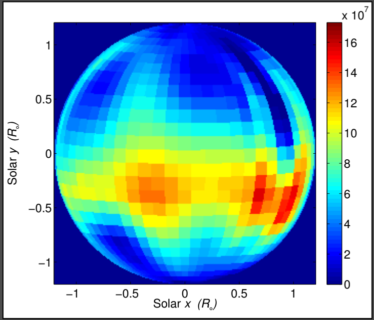
- We use the random walk state dynamic model $\boldsymbol{F}_i = \boldsymbol{I}$.

- Each 3-D static estimate required ≈ 10 minutes of processing and the complete 4-D EnKF reconstruction required ≈ 17 minutes.

4-D Reconstruction - 11/2



(a) Static reconstruction



(b) EnKF reconstruction

Learning Model Parameters Of Dynamic Systems

- In most geophysical applications, first principles physics drives the choice for the model of the dynamic system.

- In any case, the dynamic model will only be partially and imperfectly known and may depend on unobservable parameters.

- In many practical scenarios it is necessary to estimate the state of the unknown object jointly with the unknown parameters of the dynamic model.

Linear Conditional State-Space Model

$$
\begin{aligned}
\text{Initial prior:} \quad & \mathbb{E}[\boldsymbol{x}_1] = \boldsymbol{\mu}_1, \ \text{Cov}(\boldsymbol{x}_1) = \boldsymbol{\Pi}_1 & (55)\\
\text{Measurement/forward model:} \quad & \boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x}_i + \boldsymbol{v}_i & (56)\\
\text{State-transition model:} \quad & \boldsymbol{\theta}_i \sim p(\cdot|\boldsymbol{\theta}_{i-1}) & (57)\\
& \boldsymbol{x}_{i+1} = \boldsymbol{F}(\boldsymbol{\theta}_i)\, \boldsymbol{x}_i + \boldsymbol{G}(\boldsymbol{\theta}_i) \boldsymbol{u}_i & (58)
\end{aligned}
$$

- The conditional matrix $\boldsymbol{G}(\cdot)$ allows for mixing of the state noise $\boldsymbol{u}_i$.
- Both the state transition matrix $\boldsymbol{F}(\cdot)$ and the state noise mixing matrix $\boldsymbol{G}(\cdot)$ are dependent on the hidden Markov process $\boldsymbol{\theta}_i$.
- Simple example: binary parameter $\boldsymbol{\theta}_i$ which indicates if the Sun is in a low or high state of dynamic activity.