Statistical Inversion Techniques: Indirect Measurements and Aeronomy

Farzad Kamalabadi

University of Illinois at Urbana-Champaign, Urbana, IL

June 19, 2002

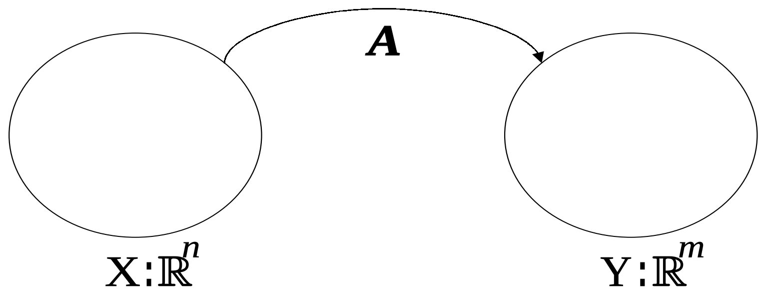
Acknowledgements

Néstor Aponte	NAIC
Sixto Gonzalez	NAIC
Craig Heinselman	SRI
Bob Kerr	SSI
Erhan Kudeki	UIUC
John Noto	SSI
Joshua Semeter	SRI
Michael Sulzer	NAIC

Overview

- Nomenclature and context: Common themes in observational aeronomy
- Linear inverse problems and linear estimation
- Taxonomy of solutions:
- Least squares, weighted least squares, and maximum likelihood estimation
- Regularization and Bayesian techniques
- Recursive estimation and optimal filtering

Examples of Observation/State Mappings



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

x	**y**
Temperature, Density, Ion Composition, Electric Field, Wind	Doppler spectrum
N_e	TEC
Volume Emission rate	Photometric Brightness

Integral Equation Model of Inverse Problems

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

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

$x(t)$: unknown quantity of interest

$y(t)$: observed (measured) quantity

$h(t; \tau)$: kernel or response function of the system

Linear Integral Equations: Examples

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

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

Partial differential equation whose solution can be written as:

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

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

Linear Integral Equations: Examples (Continued)

- Linear system (signal processing) perspective

$$\sum (a_k \frac{d^k}{dt^k}) y(t) = \sum (b_k \frac{d^k}{dt^k}) x(t) \tag{4}$$

$$\Rightarrow \frac{Y(s)}{X(s)} = H(s) = \frac{\sum b_k s^k}{\sum a_k s^k} \tag{5}$$

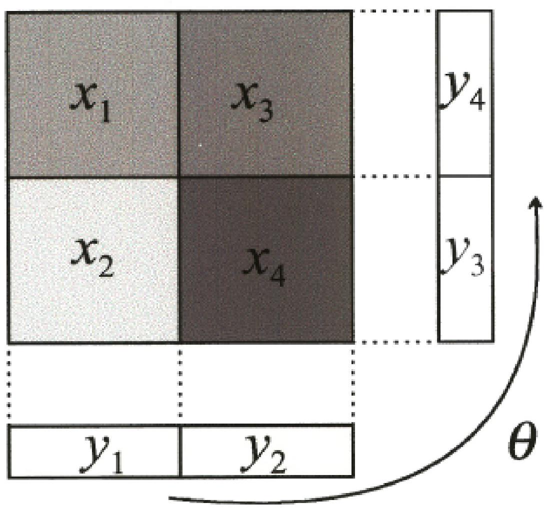
$$y(t) = \int h(t-s) x(s) ds \tag{6}$$

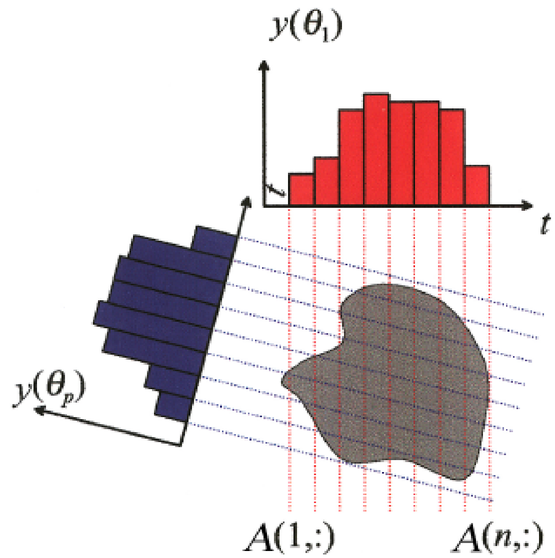
- Image reconstruction from projections:

$$y_\theta(u) = \int_{-\infty}^{\infty} x(t,s) \delta(t \cos\theta + s \sin\theta - u) dt ds. \tag{7}$$

$$h(u, \theta; t, s) = \delta(t \cos\theta + s \sin\theta - u) \tag{8}$$

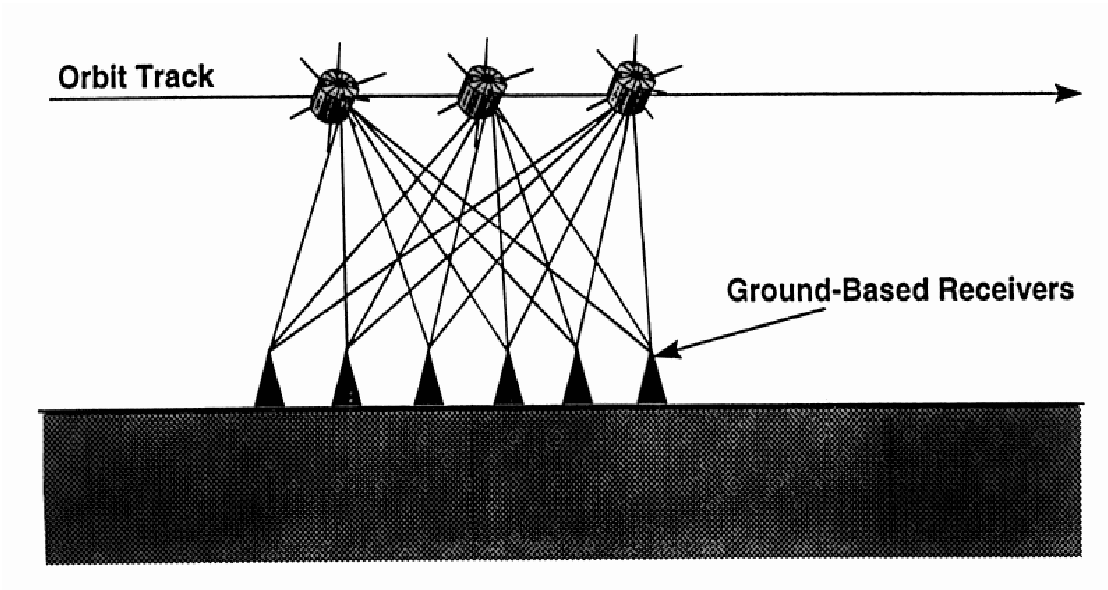
General Mathematical Model



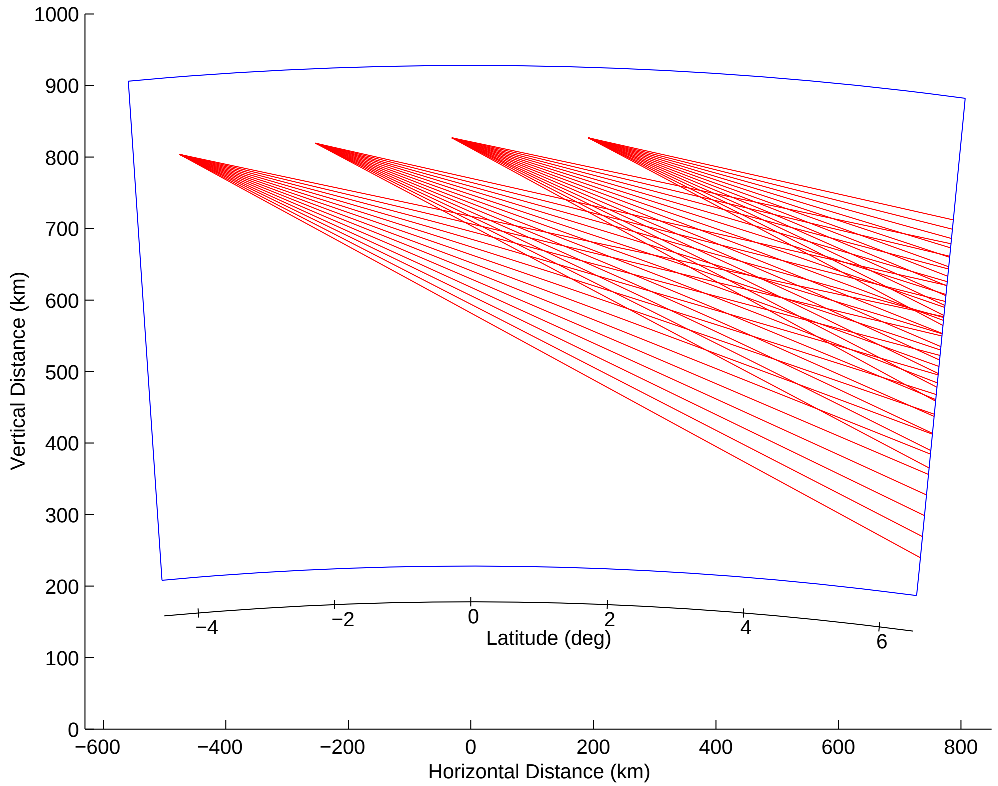


$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \Longleftrightarrow \quad \boxed{\mathbf{y} = \mathbf{A}\mathbf{x}}$$

Ionospheric Radio Tomography



Space-based Limb Scanning Observation Geometry



Discrete Representation of Integral Equations

Assuming there are m observations,

$$y_i = y(t_i) = \int h_i(\tau)x(\tau)d\tau, \quad 1 \leq i \leq m, \tag{9}$$

where $h_i = h(t_i;\tau)$ denotes the kernel corresponding to the ith observation.

The unknown quantity $x(t)$ itself can be described in terms of a discrete and finite set of parameters by a weighted sum of n basis functions $\phi_j(t)$, $j = 1, \ldots, n$ as follows:

$$x(t) = \sum_{j=1}^{n} x_j \phi_j(t). \tag{10}$$

Mathematical Statement of Inverse Problems

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

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{11}$$

- 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 $Y \supset \mathbf{A}X$

Existence, Uniqueness, and Posedness

- Case 1: Exact Solution

$N(\mathbf{A}) = 0$; Mapping is one-to-one

$R(\mathbf{A}) = Y$; Mapping is onto

$\mathbf{A}$ is square and full rank

- Case 2: Non-existence

$R(\mathbf{A}) \subset Y$; Overdetermined case

- Case 2: Non-uniqueness

$N(\mathbf{A}) \neq 0$; Underdetermined case

Least Squares

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 $\hat{\mathbf{x}}_{\text{LS}}$ that minimizes the length of the error vector $\mathbf{e} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{\text{LS}}$

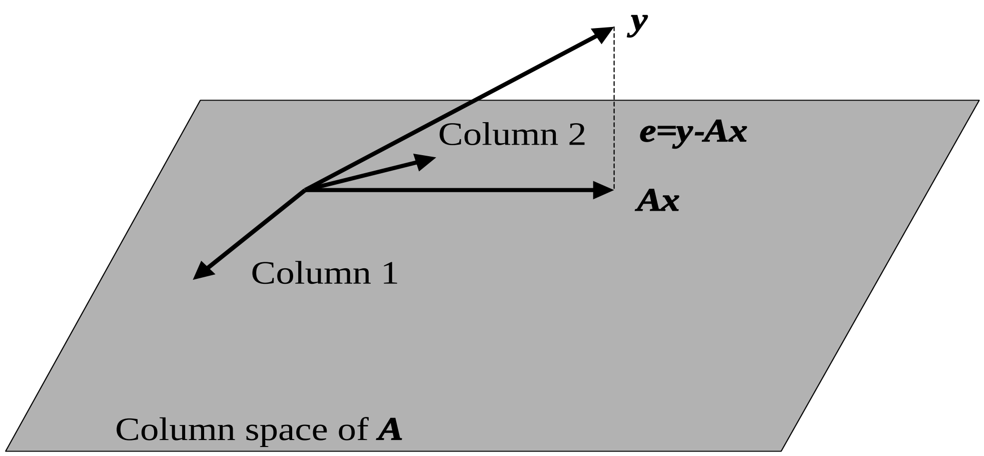
$$\arg\min_{\mathbf{x}} ||\mathbf{e}||_2^2 = \arg\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2\} \tag{12}$$

$$= \arg\min_{\mathbf{x}} \{(\mathbf{y} - \mathbf{A}\mathbf{x})^{\mathrm{T}}(\mathbf{y} - \mathbf{A}\mathbf{x})\} \tag{13}$$

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

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\hat{\mathbf{x}}_{\text{LS}} = \mathbf{A}^{\mathrm{T}}\mathbf{y} \quad \text{or} \quad \hat{\mathbf{x}}_{\text{LS}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y} \tag{14}$$

Geometrical Interpretation of Least Squares



$\mathbf{A}\hat{\mathbf{x}}_{\mathrm{LS}}$: projection of $\mathbf{y}$; the closest vector to $\mathbf{y}$ among all possible vectors $\mathbf{Ax}$

Weighted Least Squares

• Idea: If the measurements are not equally reliable, attach weights to the errors and minimize $||\mathbf{We}||_2^2 = ||\mathbf{W}(\mathbf{b} - \mathbf{Ax})||_2^2$.

In other words, find the least squares solution to $\mathbf{WAx} = \mathbf{Wy}$.

Solve

$$(\mathbf{WA})^{\mathrm{T}}\mathbf{WA}\hat{\mathbf{x}}_{\mathrm{WLS}} = (\mathbf{WA})^{\mathrm{T}}\mathbf{Wy} \quad (15)$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{WA}\hat{\mathbf{x}}_{\mathrm{WLS}} = \mathbf{A}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{Wy} \quad (16)$$

• 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_ie_j

Weighted Least Squares

$$\text{mean} = E[e] = \int xp(x)dx \text{ and variance} = E[e^2] = \int x^2 p(x)dx \quad (17)$$

$$covariance = E[e_i e_j] = \int\int (e_i)(e_j)(joint\ probability\ of\ e_i\ and\ e_j) \quad (18)$$

$$\mathbf{R}_e = E[\mathbf{e}\mathbf{e}^T] \quad (19)$$

Assumptions:
Unbiased errors: $E[\mathbf{e}] = 0$;
The estimation rule $\hat{\mathbf{x}} = \mathbf{L}\mathbf{y}$ is *linear* and *unbiased*

$$E[\mathbf{x} - \hat{\mathbf{x}}] = E[\mathbf{x} - \mathbf{L}\mathbf{y}] = E[\mathbf{x} - \mathbf{L}\mathbf{A}\mathbf{x} - \mathbf{L}\mathbf{e}] = E[(\mathbf{I} - \mathbf{L}\mathbf{A})\mathbf{x}] = 0 \quad (20)$$

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{R}_e^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{R}_e^{-1}\mathbf{y} \quad (21)$$

Statistical Estimation Models

Nature:
(i) $p(\mathbf{x})$
(ii) $\mathbf{x}$ unknown

Observation Model: $p(\mathbf{y}|\mathbf{x})$
(i) Conditional pdf
(ii) Parameterized density

Estimation Rule:
(i) Bayes:
MMSE $\rightarrow \hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{y}]$
MAP $\rightarrow \arg\max p(\mathbf{x}|\mathbf{y})$
(ii) Fisher:
ML $\rightarrow \arg\max p(\mathbf{y}|\mathbf{x})$

Weighted Least Squares (Statistical Interpretation: ML)

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\max_{\mathbf{x}}\{p(\mathbf{y}|\mathbf{x})\} \tag{22}$$

The conditional probability $p(\mathbf{y}|\mathbf{x})$ is also a Gaussian, with the following mean and covariance,

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &\sim \mathcal{N}(\mathbf{A}\mathbf{x}, \mathbf{R_e}) \\ &= e^{-\frac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^{\mathrm{T}}\mathbf{R_e^{-1}}(\mathbf{y}-\mathbf{A}\mathbf{x})} \end{aligned} \tag{23}$$

the ML estimate takes the following optimization form:

$$\begin{aligned} \hat{\mathbf{x}}_{\mathrm{ML}} &= \arg\max_{\mathbf{x}}\{\ln p(\mathbf{y}|\mathbf{x})\} \\ &= \arg\max_{\mathbf{x}}\{-\frac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^{\mathrm{T}}\mathbf{R_e^{-1}}(\mathbf{y}-\mathbf{A}\mathbf{x})\} \\ &= \arg\min_{\mathbf{x}}\{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2_{\mathbf{R_e^{-1}}}\} \end{aligned} \tag{24}$$

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

$$\hat{\mathbf{x}}_{\mathrm{ML}} = (\mathbf{A}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{y} \tag{25}$$

The estimation error is defined as:

$$\mathbf{e}_{\mathrm{ML}} = \mathbf{x} - \hat{\mathbf{x}}_{\mathrm{ML}} \tag{26}$$

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

$$\mathbf{e}_{\mathrm{ML}} = -(\mathbf{A}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1} \tag{27}$$

Finally, the ML estimation error covariance is given by:

$$\begin{aligned}\mathbf{R}_{\mathrm{ML}} &= E\{\mathbf{e}\mathbf{e}^{\mathrm{T}}\} \\ &= (\mathbf{A}^{\mathrm{T}}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{A})^{-1}\end{aligned} \tag{28}$$

Stability and Conditioning

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq c(\mathbf{A}) \frac{\|\delta \mathbf{y}\|}{\|\mathbf{y}\|} \tag{29}$$

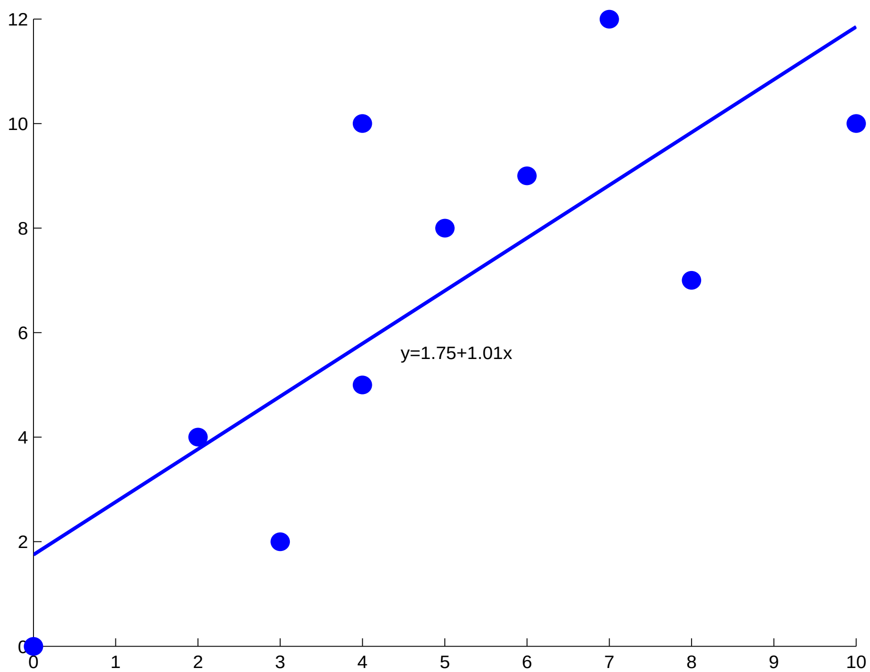
where $c(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$ is defined as the *condition number* of $\mathbf{A}$ and it can be interpreted as a measure of the singularity of the system.

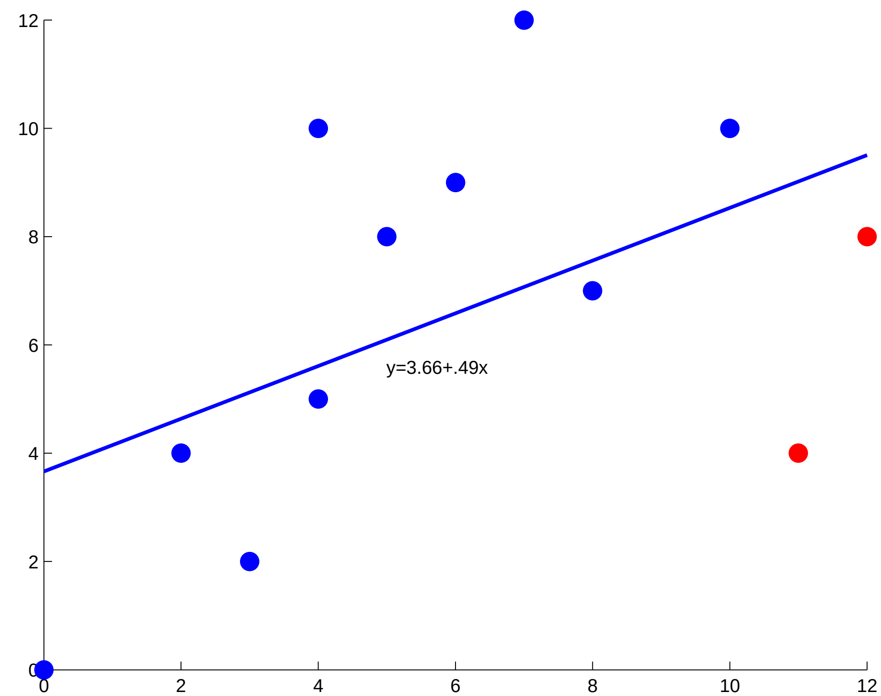
$$\mathbf{y}_0 = \mathbf{A}\mathbf{x}_0$$

$$\mathbf{y}_0 = \begin{bmatrix} 0.26 \\ 0.28 \\ 3.31 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{bmatrix} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{30}$$

$$\text{Suppose } \mathbf{y} = \mathbf{y}_0 + \delta\mathbf{y} = \begin{bmatrix} 0.27 \\ 0.25 \\ 3.33 \end{bmatrix} \quad \delta\mathbf{y} = \begin{bmatrix} 0.01 \\ -0.03 \\ 0.02 \end{bmatrix} \quad 1\ \%\ \text{change} \tag{31}$$

$$\hat{\mathbf{x}}_{LS} = \begin{bmatrix} 7.01 \\ -8.40 \end{bmatrix}$$


y=1.75+1.01x
0
1
2
3
4
5
6
7
8
9
10
2
4
6
8
10
12


y=3.66+.49x
0
2
4
6
8
10
12

MAP Estimation and Error Covariance

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R_x}) \tag{32}$$

$$\hat{\mathbf{x}}_{\text{map}} = \underset{\mathbf{x}}{\text{argmax}} \quad \{p(\mathbf{x}|\mathbf{y})\} = \underset{\mathbf{x}}{\text{argmax}} \quad \{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\} \tag{33}$$

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

$$p(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{Ax}, \mathbf{R_e}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\mathbf{R_e}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{y} - \mathbf{Ax})^T \mathbf{R_e}^{-1}(\mathbf{y} - \mathbf{Ax})\} \tag{34}$$

$$p(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{R_x}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\mathbf{R_x}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}\mathbf{x}^T \mathbf{R_x}^{-1}\mathbf{x}\} \tag{35}$$

and

$$\hat{\mathbf{x}}_{\text{map}} = \underset{\mathbf{x}}{\text{argmax}} \quad \{\exp\{-\tfrac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^T\mathbf{R}_{\mathbf{e}}^{-1}(\mathbf{y}-\mathbf{A}\mathbf{x})\}\exp\{-\tfrac{1}{2}\mathbf{x}^T\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}\}\} \tag{36}$$

$$\hat{\mathbf{x}}_{\text{map}} = \underset{\mathbf{x}}{\text{argmax}} \quad \{-\tfrac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^T\mathbf{R}_{\mathbf{e}}^{-1}(\mathbf{y}-\mathbf{A}\mathbf{x}) - \tfrac{1}{2}\mathbf{x}^T\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}\} \tag{37}$$

$$\hat{\mathbf{x}}_{\text{map}} = \underset{\mathbf{x}}{\text{argmin}} \quad \{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2_{\mathbf{R}_{\mathbf{e}}^{-1}} + \|\mathbf{x}\|^2_{\mathbf{R}_{\mathbf{x}}^{-1}}\} \tag{38}$$

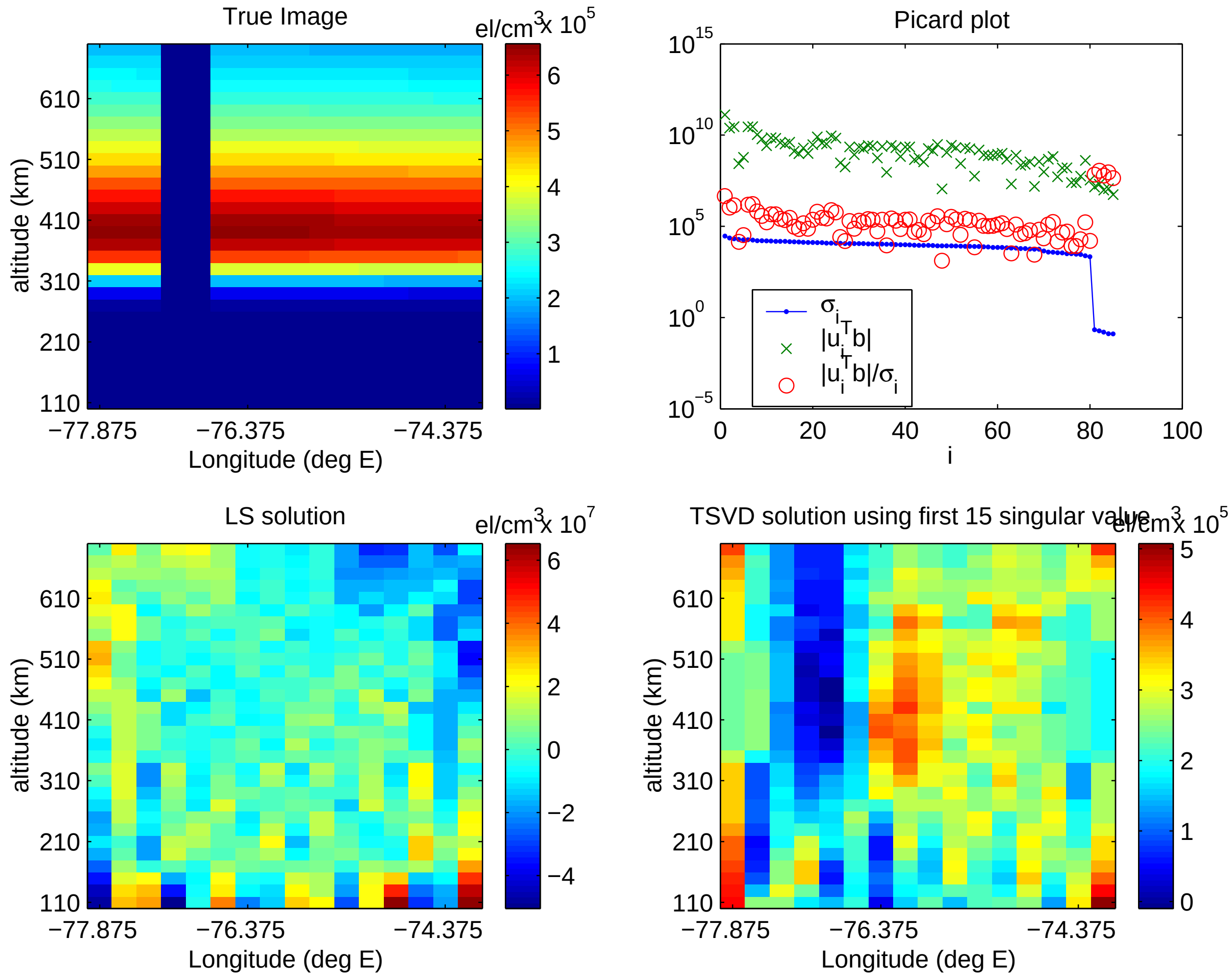
It can be shown that the solution to this minimization problem, $\hat{\mathbf{x}}_{\text{map}}$, is given by:

$$\hat{\mathbf{x}}_{\text{map}} = \mathbf{R}_{\hat{\mathbf{x}}_{\text{map}}}(\mathbf{A}^T\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{y}) \tag{39}$$

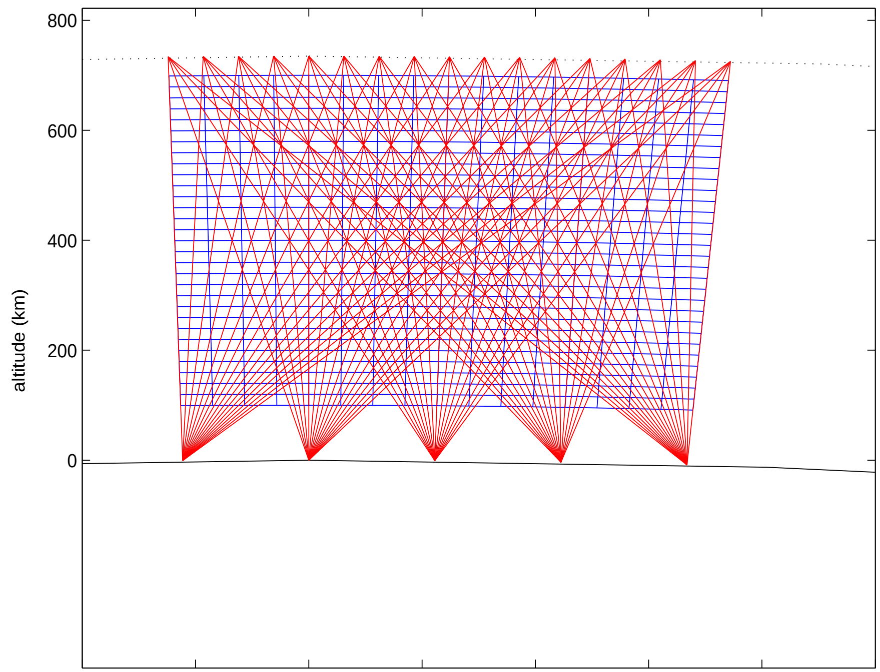
where $\mathbf{R}_{\hat{\mathbf{x}}_{\text{map}}}$ is the estimation error covariance given by:

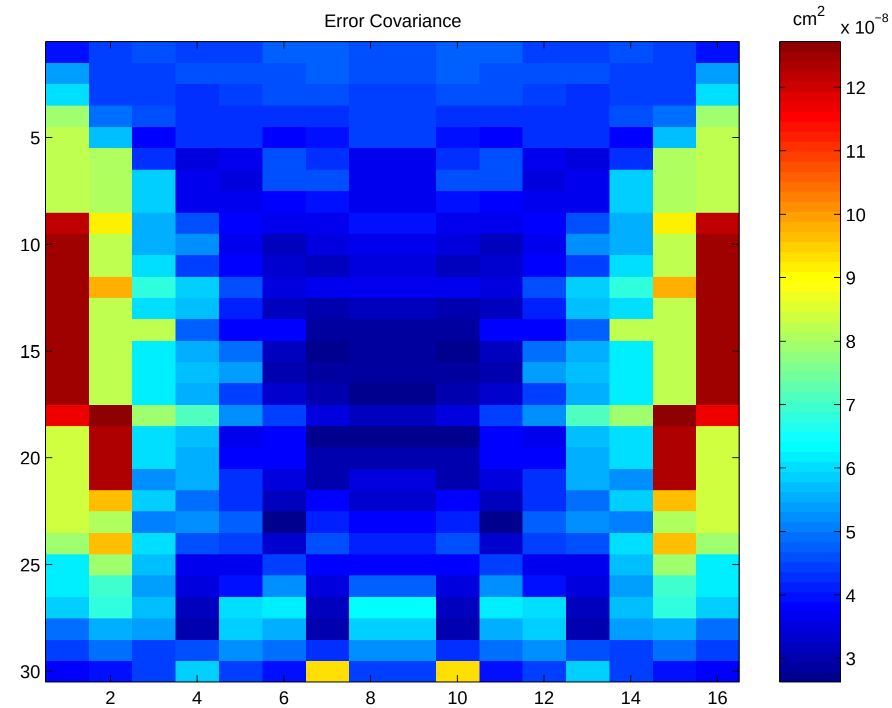
$$\mathbf{R}_{\hat{\mathbf{x}}_{\text{map}}} = (\mathbf{R}_{\mathbf{x}}^{-1} + \mathbf{A}^T\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{A})^{-1} \tag{40}$$

Radio Tomography Example: Reconstruction



Radio Tomography Example: Error Covariance





Regularization and its Stochastic Interpretation

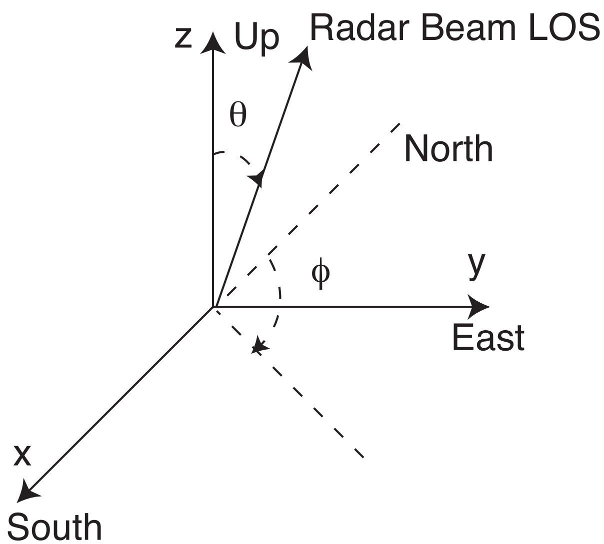
$$\hat{\mathbf{x}}_{\text{Tik}} = \underset{\mathbf{x}}{\operatorname{argmin}} \quad \underbrace{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}_{\text{Data Fidelity}} + \underbrace{\gamma^2 \|\mathbf{L}\mathbf{x}\|_2^2}_{\text{Prior Info}}$$

- Idea: Include prior information into solution
- Interpretations:
  - Add additional constraint: Penalize large values of $\mathbf{Lx}$ (e.g. $\mathbf{L} = \nabla$)
  - Improves conditioning: $(\mathbf{A}^\mathbf{T}\mathbf{A} + \gamma^2\mathbf{L}^\mathbf{T}\mathbf{L})\mathbf{x} = \mathbf{A}^\mathbf{T}\mathbf{y}$
  - 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 "$\mathbf{A}^{-1}$" at high frequency

Example: Ion Velocity Field Estimation

• Objective: Measurement of the three dimensional ion velocity field

• Problem: There are three unknown components for each line of sight velocity.



$$\begin{bmatrix} V_{LOS}^1 \\ V_{LOS}^2 \end{bmatrix} = \begin{bmatrix} -cos\phi sin\theta^1 & sin\phi sin\theta^1 & cos\theta^1 \\ -cos\phi sin\theta^2 & sin\phi sin\theta^2 & cos\theta^2 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$$

Example: Ion Velocity Field Estimation (Continued)

- Least squares approach [*Hagfors and Behnke*, 1974]: Assume that the unknowns do not change in the time of one rotation, and use many samples.

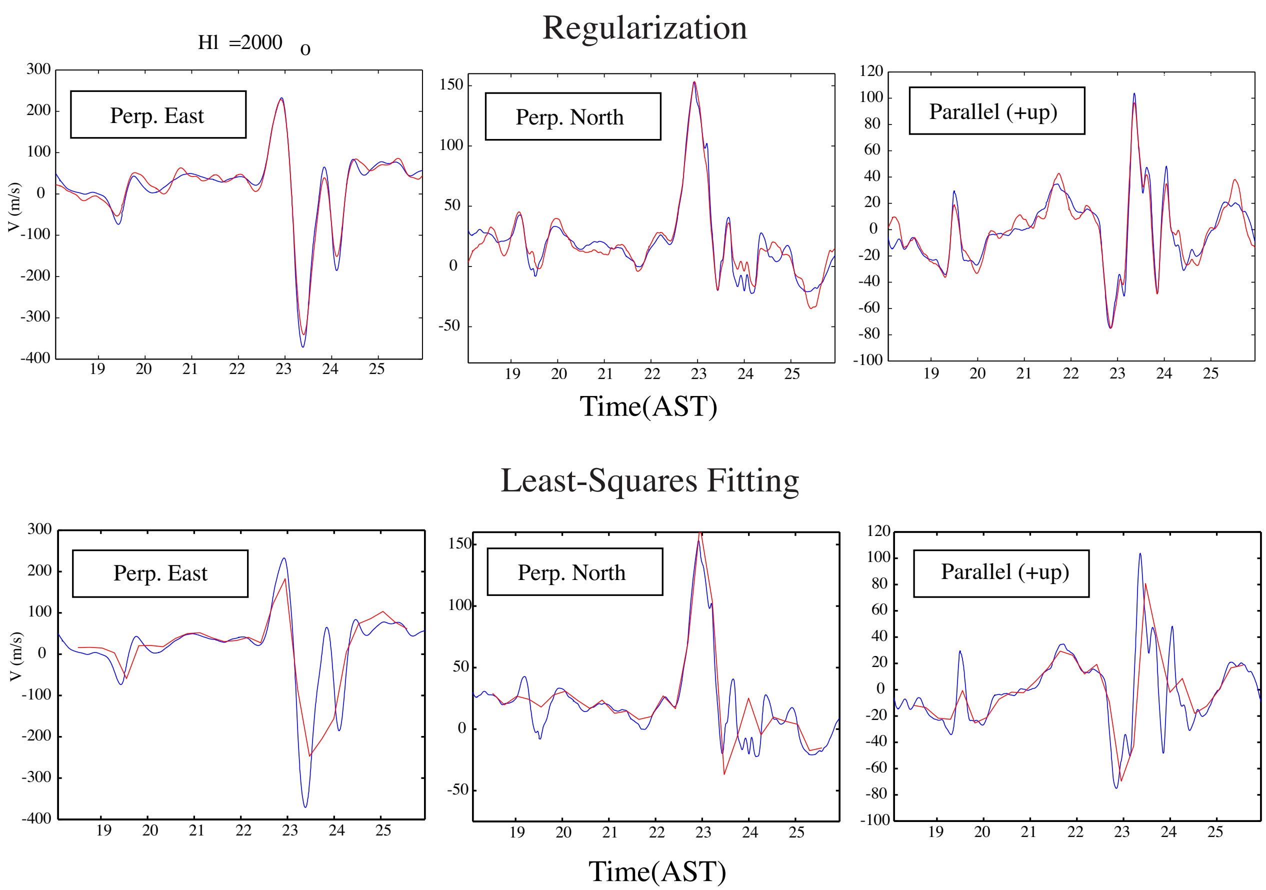
$$\begin{bmatrix} V_{LOS}(1) \\ \vdots \\ V_{LOS}(n) \end{bmatrix} = \begin{bmatrix} -cos\phi_1 sin\theta & sin\phi_1 sin\theta & cos\theta \\ & \vdots & \\ -cos\phi_n sin\theta & sin\phi_n sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \tag{41}$$

- Alternative approach using regularization:

$$\begin{bmatrix} V_{LOS}(1) & V_{LOS}(2) & ... & V_{LOS}(N) \end{bmatrix} \tag{42}$$

$$\begin{bmatrix} v_x(1) & v_y(1) & v_z(1) & \cdots & v_x(N) & v_y(N) & v_z(N) \end{bmatrix} \tag{43}$$

Example: Ion Velocity Field Estimation (Results)



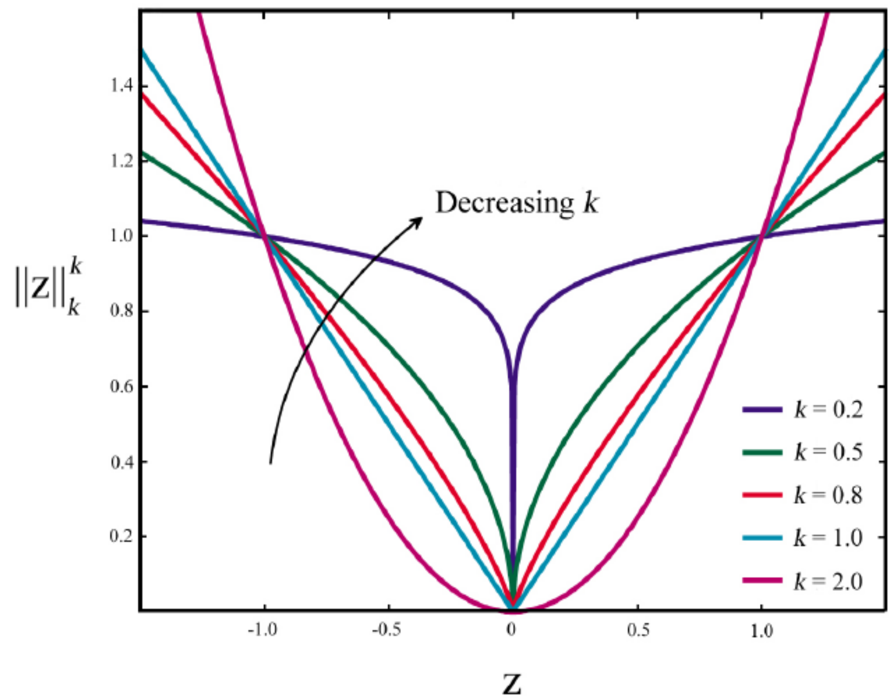
Interpretation of Cost as Statistical Model

$$\begin{aligned}
\arg\min_{\mathbf{x}} \quad & \|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2_{\mathbf{w_1}} + \|\mathbf{x}\|^2_{\mathbf{w_2}} \\
= \arg\max_{\mathbf{x}} \quad & -\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2_{\mathbf{w_1}} - \frac{1}{2}\|\mathbf{x}\|^2_{\mathbf{w_2}} \\
= \arg\max_{\mathbf{x}} \quad & \mathrm{e}^{-\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2_{\mathbf{w_1}} - \frac{1}{2}\|\mathbf{x}\|^2_{\mathbf{w_2}}} \\
= \arg\max_{\mathbf{x}} \quad & \underbrace{\mathrm{e}^{-\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2_{\mathbf{w_1}}}}_{\swarrow} \; \underbrace{\mathrm{e}^{-\frac{1}{2}\|\mathbf{x}\|^2_{\mathbf{w_2}}}}_{\searrow}
\end{aligned}$$

$$\mathrm{P}(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \mathbf{w}_1^{-1}) \qquad \mathrm{P}(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{w}_2^{-1})$$

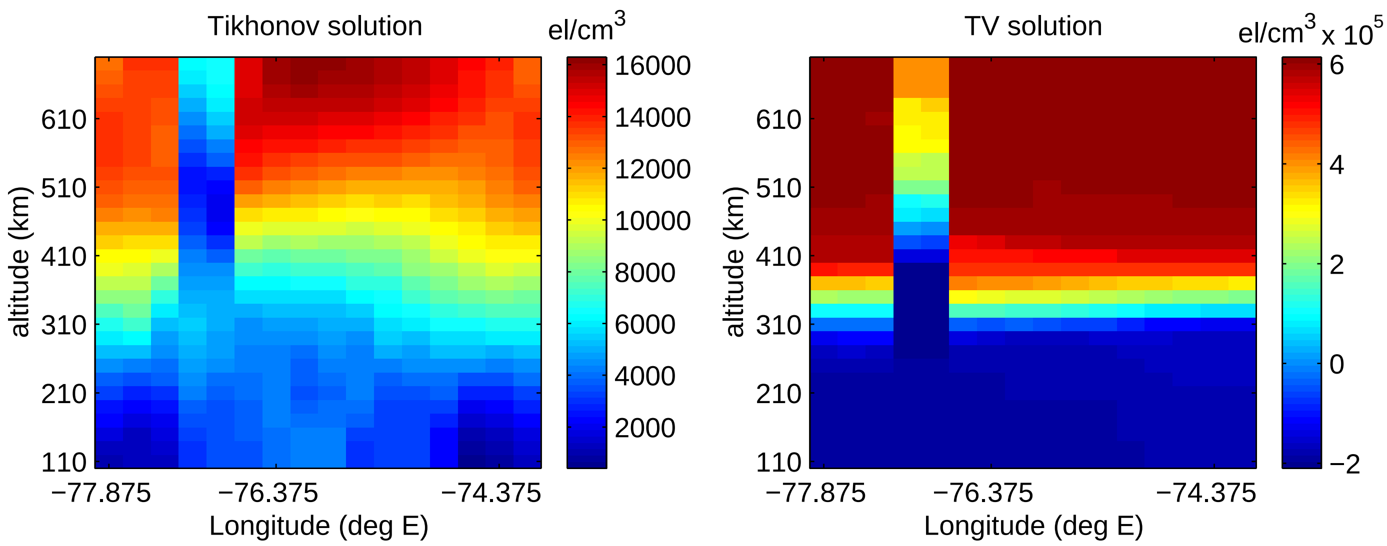
More Generally, One Can Imagine...

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \gamma^2 \|\mathbf{L}\mathbf{x}\|_k^k \quad \text{where} \quad \|\mathbf{z}\|_k^k = \sum_i |\mathbf{z}_i|^k$$



- Smaller values of k suppress small values of $\mathbf{Lx}$ in favor of large values of $\mathbf{Lx}$

Radio Tomography Example: Total Variation Reconstruction



Recursive Estimation

$$\mathbf{x}_o = (\mathbf{A}_o^T \mathbf{V}_o^{-1} \mathbf{A}_o)^{-1} \mathbf{A}_o^T \mathbf{V}_o^{-1} \mathbf{y}_o \tag{44}$$

$$\mathbf{P}_o = E[(\mathbf{x} - \mathbf{x}_o)(\mathbf{x} - \mathbf{x}_o)^T] = (\mathbf{A}_o^T \mathbf{V}_o^{-1} \mathbf{A}_o)^{-1} \tag{45}$$

- Question: If more data arrives, can the best estimate for the combined system be computed from $\mathbf{x}_0$ and $\mathbf{y}_1$ without restarting the calculation from $\mathbf{y}_0$?

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_o & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_1 \end{bmatrix} \text{ is the covariance matrix of the errors } \begin{bmatrix} \mathbf{e}_o \\ \mathbf{e}_1 \end{bmatrix} \tag{46}$$

$$\mathbf{P}_1^{-1} = \begin{bmatrix} \mathbf{A}_o \\ \mathbf{A}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{V}_o & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}_o \\ \mathbf{A}_1 \end{bmatrix} = \mathbf{A}_o^T \mathbf{V}_o^{-1} \mathbf{A}_o + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{A}_1 \tag{47}$$

$$\mathbf{x}_1 = \mathbf{P}_1 \begin{bmatrix} \mathbf{A}_o \\ \mathbf{A}_1 \end{bmatrix}^T \mathbf{V}^{-1} \begin{bmatrix} \mathbf{y}_o \\ \mathbf{y}_1 \end{bmatrix} = \mathbf{P}_1(\mathbf{A}_o^T \mathbf{V}_o^{-1} \mathbf{y}_o + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{y}_1) \tag{48}$$

$$\mathbf{P}_1^{-1} = \mathbf{P}_o^{-1} + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{A}_1 \tag{49}$$

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{P}_1(\mathbf{P}_o^{-1}\mathbf{x}_o + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{y}_1) \\ &= \mathbf{P}_1(\mathbf{P}_1^{-1}\mathbf{x}_o - \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{A}_1 \mathbf{x}_o + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{y}_1) \\ &= \mathbf{x}_o + \mathbf{K}_1(\mathbf{y}_1 - \mathbf{A}_1 \mathbf{x}_o) \end{aligned} \tag{50}$$

The Kalman Filter

$$\mathbf{y}_i = \mathbf{A}_i \mathbf{x}_i + \mathbf{e}_i \tag{51}$$

$$\mathbf{x}_{i+1} = \mathbf{F}_i \mathbf{x}_i + \mathbf{E}_i \tag{52}$$

Statistical Fusion of Multi-Sensor Data

• Definition: Data fusion is the process by which data from a multitude of sensors is used to yield an optimal estimate of a specified state vector pertaining to the observed system.

The measurement model for ionospheric tomography can be expressed as:

$$\mathbf{y} = f(\mathbf{x}) + \mathbf{e} = \mathbf{A}\mathbf{x} + \mathbf{e} \tag{53}$$

where $\mathbf{x}$ represents the state vector of the system of interest, and by $\mathbf{y}$ we denote the total set of measured quantities.

The statistical inference problem is then to estimate the value of $\mathbf{x}$ from the knowledge of $\mathbf{y}$ in accordance with some specified optimality criterion. In this case the outputs of more than one type of sensor can be represented as:

$$\mathbf{y} = \left[\frac{\mathbf{y}_1}{\mathbf{y}_2}\right] \tag{54}$$

Statistical Fusion of Multi-Sensor Data (continued)

The task of the fusion process is then to produce the optimal estimate $\hat{\mathbf{x}}(\mathbf{y_1}, \mathbf{y_2})$, based on the total set of measured data $\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$

- Question: How to decompose (modularize) the total process into parts that have stand-alone significance?

The measurement model then translates into the following multisensor formalism:

$$\mathbf{y_i} = f(\mathbf{x_i}) + \mathbf{e_i} \tag{55}$$

The modularization of the problem is based on the relation:

$$p(\mathbf{x}|\mathbf{y_1}, \mathbf{y_2}) = p(\mathbf{y_1}|\mathbf{x})p(\mathbf{y_2}|\mathbf{x})p(\mathbf{x})p(\mathbf{y_1}, \mathbf{y_2})^{-1} \tag{56}$$

Summary and Conclusions

- A statistical formulation of an inverse problem provides a rational framework for the inclusion of prior knowledge about the unknown.

- Challenges concerning the inversion can be addressed suitably by incorporating statistical models to ensure meaningful results.

- There often exist intimate links between a deterministic and statistical view of the same inverse problem.

- A statistical approach provides quantitative measures of estimation uncertainty; an important attribute for assimilation into models.