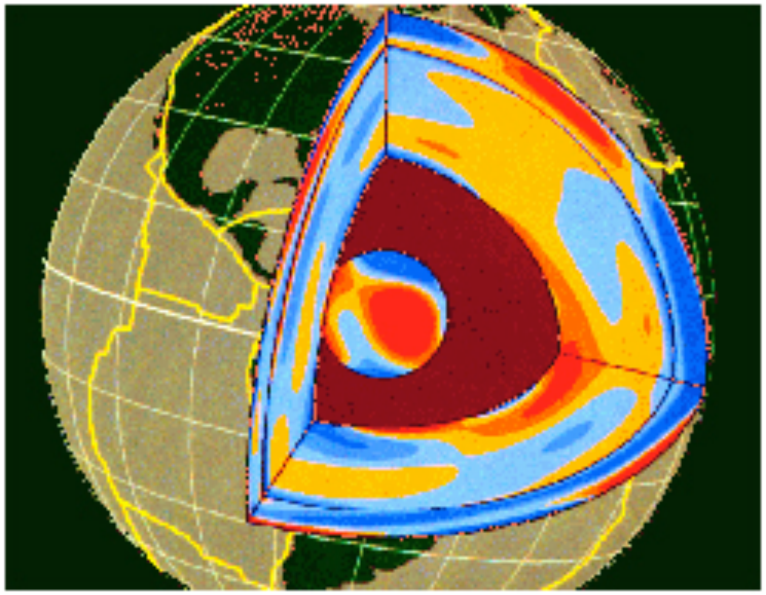


Woodhouse and Dziewonski, 1986

Inversion

(part I, history & formulation):

1. Earth is highly heterogeneous
2. The pattern can be roughly quantified by low degree (large-scale) anomalies.

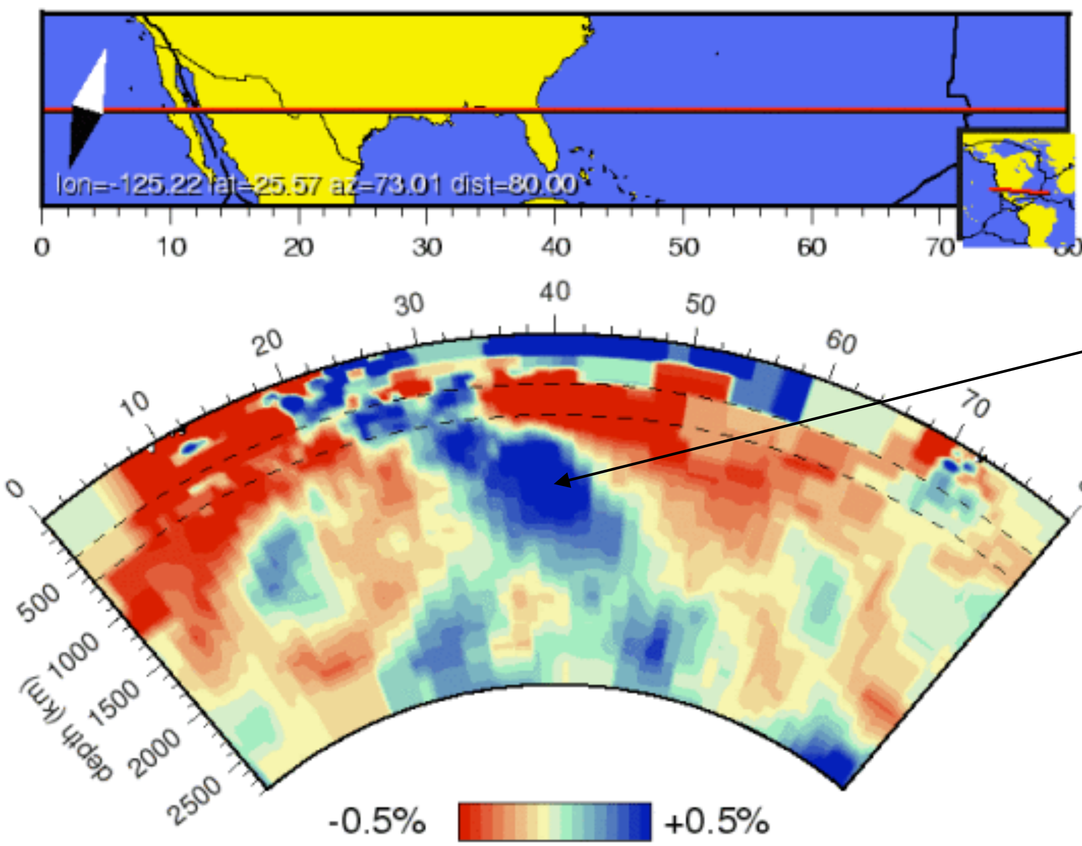
Limitations:

1. Limited observations make an inverse problem under-constrained

2. The "higher-degree" (small-scale) structures are inherently filtered out due to coarse parameterization, thereby emphasizing the long-wavelength patterns in the image.

Mantle tomography

- E.g., Bijwaard, Spakman, Engdahl, 1999.



Subducting Farallon slab

History of Seismic Tomography

Tomo— Greek for "tomos" (body), graphy --- study or subject

Where it all began: Radon transform: (Johan Radon, 1917): integral of function over a straight line segment.

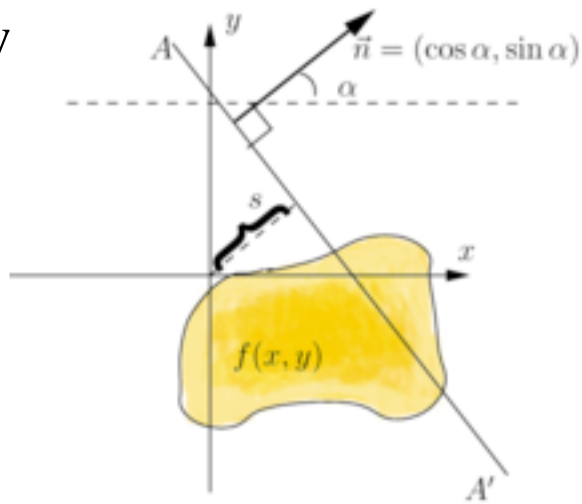
Radon transform

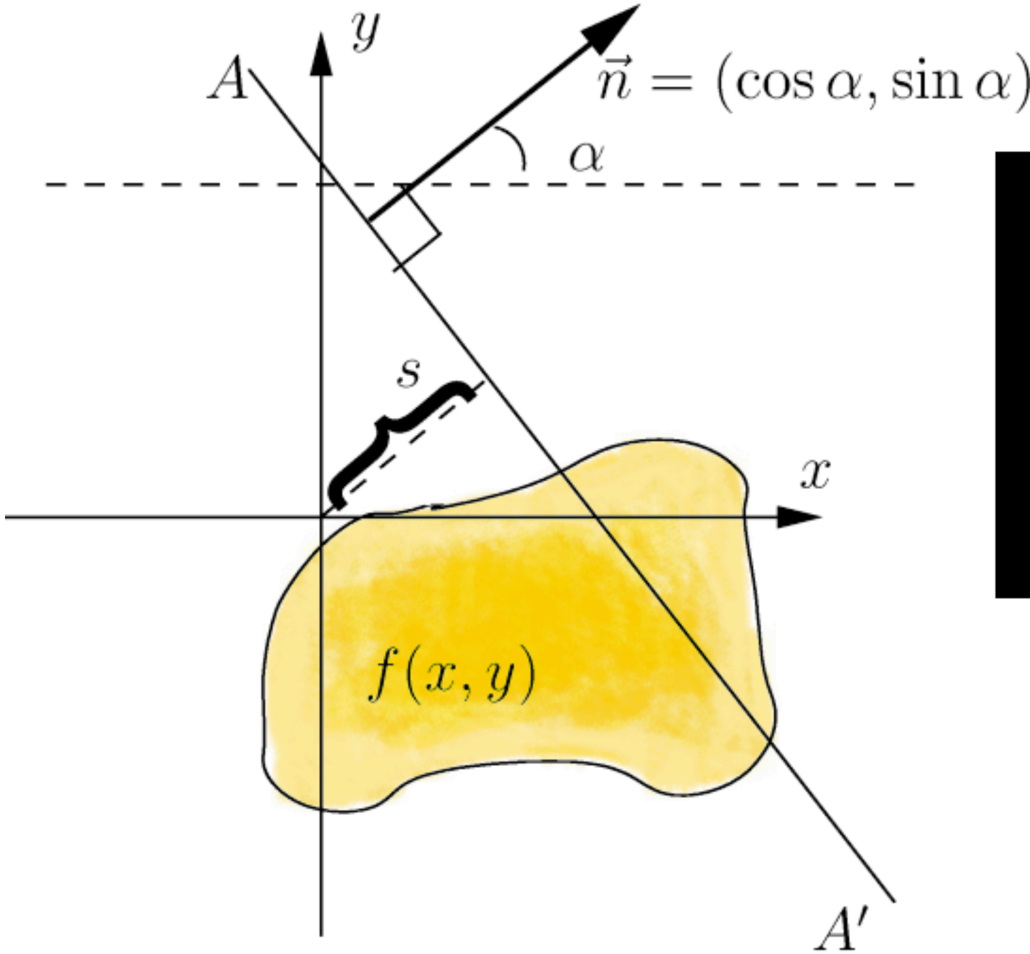
$$p(s,\alpha) = \int f(x,y)\delta(x\cos\alpha + y\sin\alpha - s)dxdy$$

where p is the radon transform of f(x, y), and δ is a Dirac Delta Function (an infinite spike at 0 with an integral area of 1)

p is also called sinogram, and it is a sine wave when f(x, y) is a point

v





Shepp-Logan Phantom
(human cerebral)

Input Radon Projected

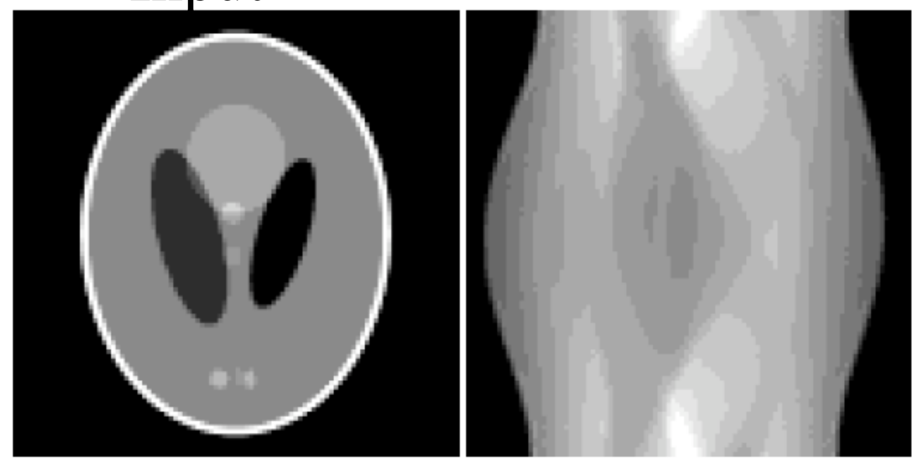
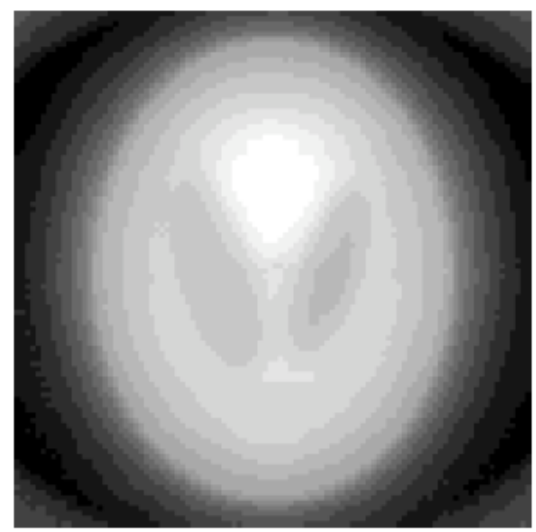


Figure 2. Shepp-Logan phantom and its Radon transform.

Recovered
(output)

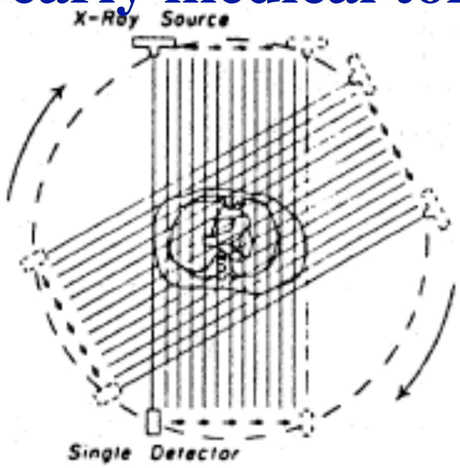


Back projection of the function is a way to solve f() from p() ("Inversion"):

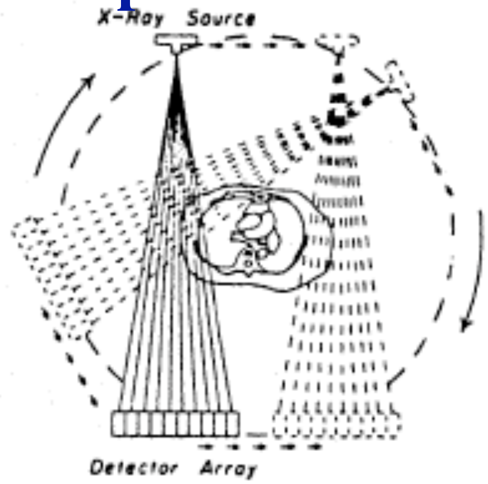
$$f(x,y) = \int_0^\pi p(x\cos\alpha + y\sin\alpha),\alpha)d\alpha$$

A few of the early medical tomo setups

Parallel beam



1st Generation CT Scanner
(Parallel Beam, Translate-Rotate)

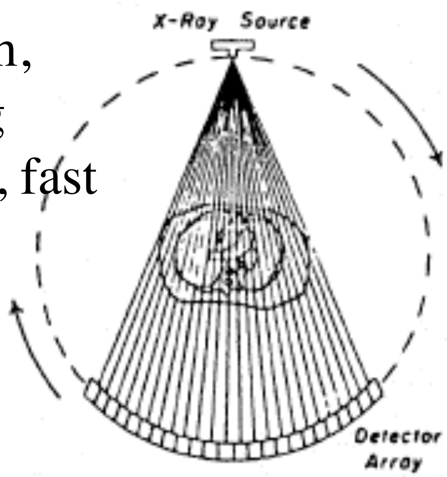


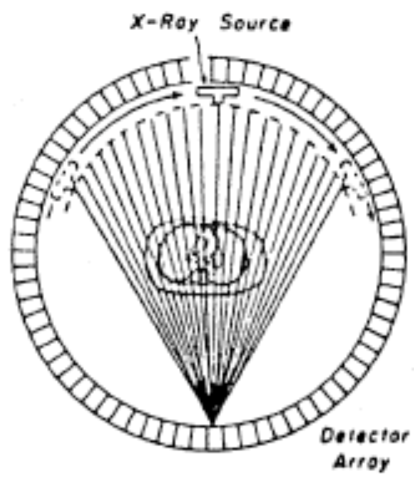
2nd Generation CT Scanner
(Fan Beam, Translate-Rotate)

Fan beam,
Multi-receiver,
Moves in big steps

Cunningham & Jurdy, 2000

Broader fan beam,
Coupled, moving
source receivers, fast
moving





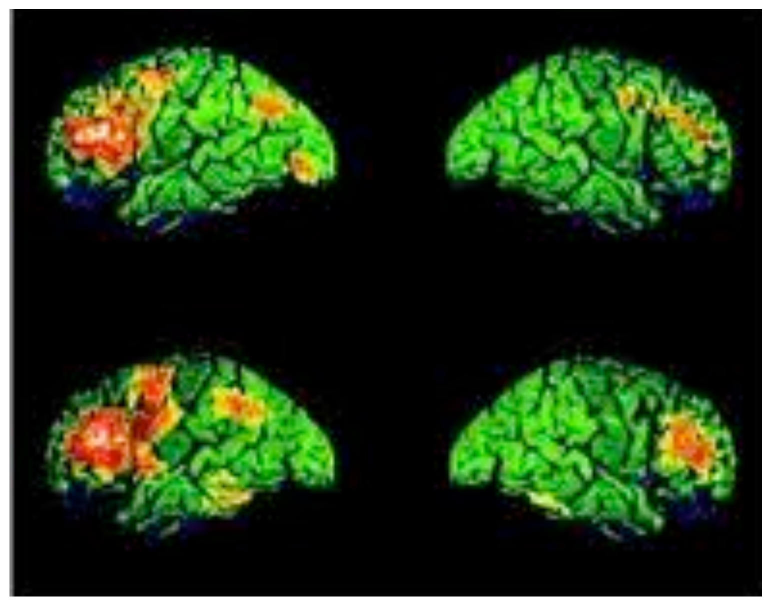
Broader fan beam,
Moving source,
fixed receivers, fast
moving
(1976)

Different "generations" of X-Ray Computed Tomography (angled beams are used to increase resolution). Moral: good coverage & cross-crossing rays a must in tomography (regardless of the kind)

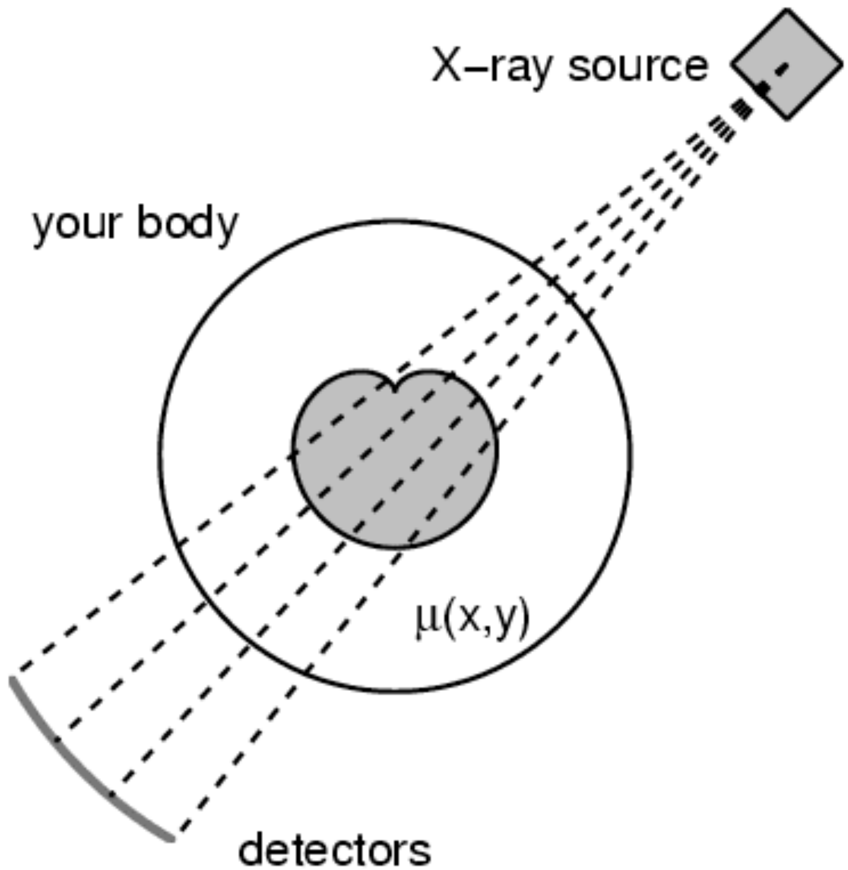
Present Generation of models: Dense receiver sets, all rotating, great coverage and crossing rays.

Brain Scanning Cool Fact:

According to an earlier report, the best valentine's gift to your love ones is a freshly taken brainogram. The spots of red shows your love, not your words!



What is $f(x, y)$? Medical applications.



X-ray absorption & scattering

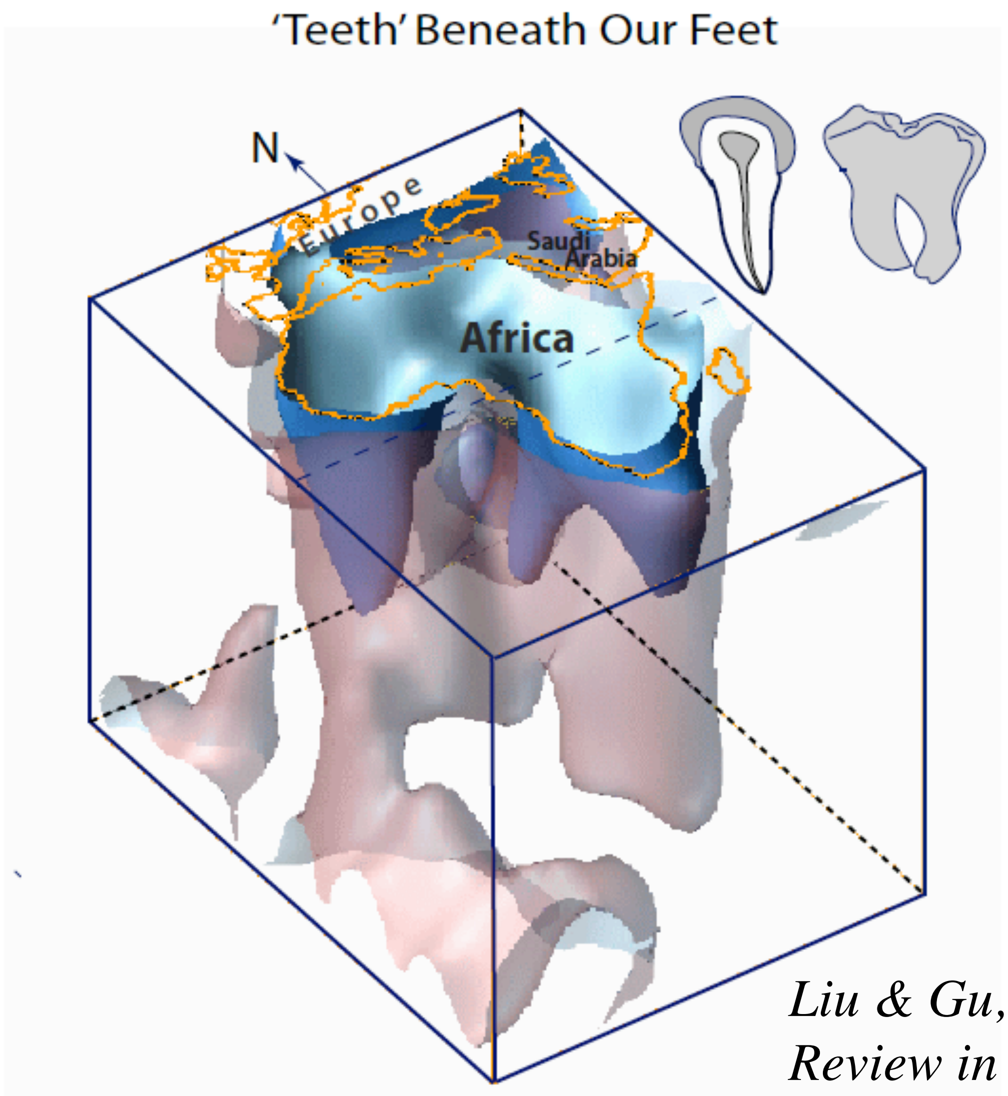
Tissues and bones have $\neq$ absorption and scattering coefficients $\mu(x, y)$.

Recorded intensity goes as

$$I = I_0 \exp\left[\int_{\text{ray}} -\mu(x, y)\, ds\right]. \qquad (2)$$

Sources and detectors rotate to achieve perfect "coverage".

Attractive images like this are why the term "seismic tomography" got hot.



Liu & Gu, 2012, Review in Tectonophysics.

Official Credit in Seismic Tomo: K. Aki and coauthors (1976)

First tomographic study of california

VOL. 81, NO. 23 JOURNAL OF GEOPHYSICAL RESEARCH AUGUST 10, 1976

DETERMINATION OF THREE-DIMENSIONAL VELOCITY ANOMALIES UNDER A SEISMIC ARRAY USING FIRST P ARRIVAL TIMES FROM LOCAL EARTHQUAKES
1. A HOMOGENEOUS INITIAL MODEL

Keiiti Aki

Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

W. H. K. Lee

first P arrival time as

$$T_{ij}^{obs} = T_{ij}^{cal} + \left(\frac{\partial T}{\partial X}\right)_{ij} \Delta X_j + \left(\frac{\partial T}{\partial Y}\right)_{ij} \Delta Y_j + \left(\frac{\partial T}{\partial Z}\right)_{ij} \Delta Z_j + \Delta T_j + \sum_k T_{ij}^{(k)} F_k + E_{ij} \quad (1)$$

Structure Term
F_k = *fraction of slowness change*
T_{ji}=*time spend in a 'cell'*

where T_{ij}^{cal} is the calculated first P arrival time based on the homogeneous initial model:

$$T_{ij}^{cal} = T_j^o + (X_i - X_j^o)^2 + (Y_i - Y_j^o)^2 + (Z_i - Z_j^o)^2 \; {}^{1/2}/V_o \quad (2)$$

Eventually, something familiar & simple

$$\tau = G\chi + \varepsilon \quad (4)$$

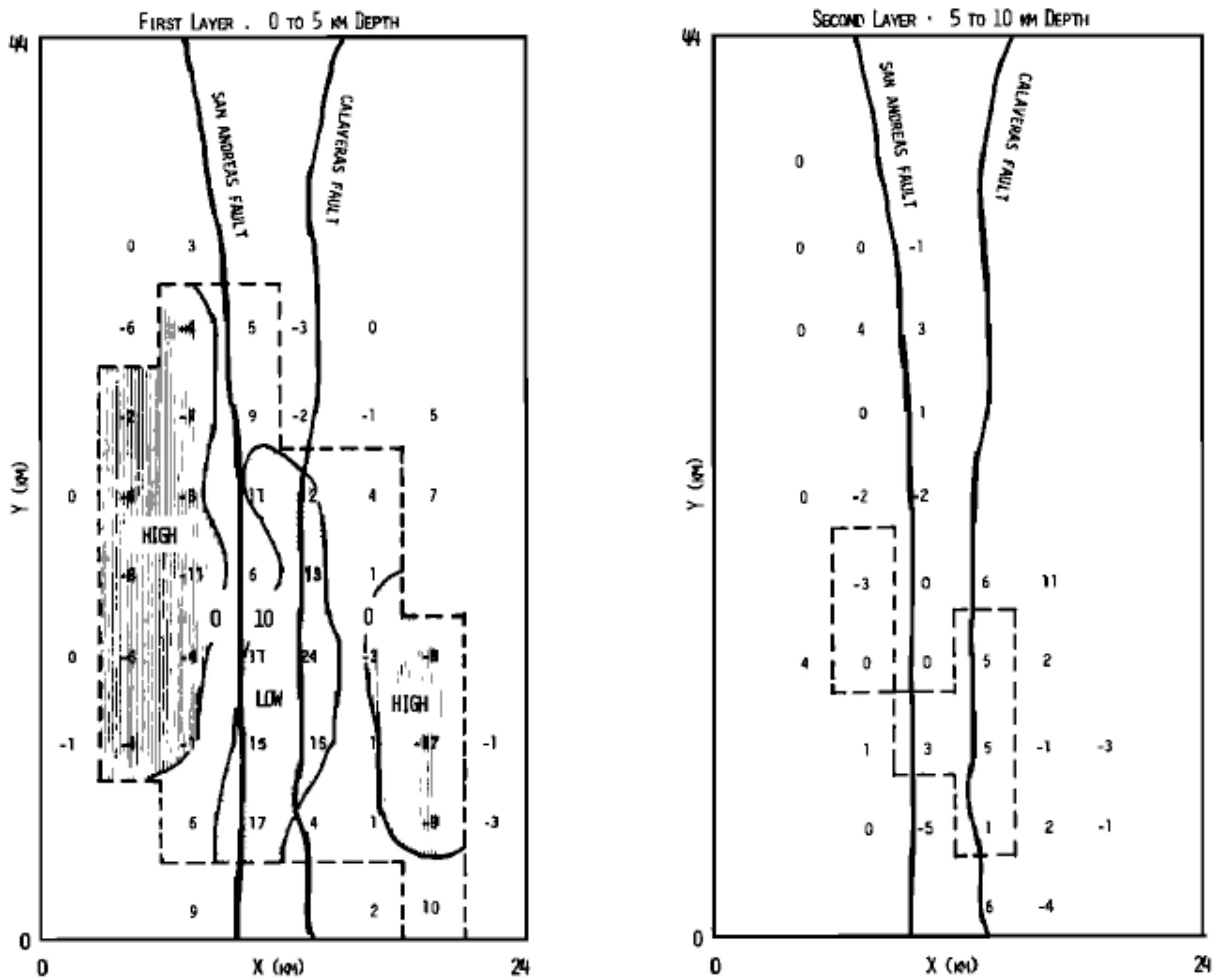


Fig. 11. Map showing contours of slowness in percent (upper number) and its standard error (lower number) for case 2.

Real Credit: John Backus & Free Gilbert (1968)

First established the idea of Differential Kernels inside integral as a way to express dependency of changes in a given seismic quantity (e.g., time) to velocities/densities (**use of Perturbation Theory**).

Freeman Gilbert

Adam Dziewonski

First "Global Seismic Tomographic Model" (1984)

Don Anderson

PREM 1D model (1981, Preliminary Reference Earth Model) ($500 K Crawford/Nobel Price)

John Woodhouse

Moral: Established the proper reference to express perturbations

Except:

- **X-ray**: exponential of a line integral
S-ray: raypath itself is a function of velocity } **non-linear** functions!
- Earth **coverage** is non-continuous
- "Experiment" is done by nature and **not repeatable**
- Earthquake **source parameters** (location, time) is uncertain

Remedy:

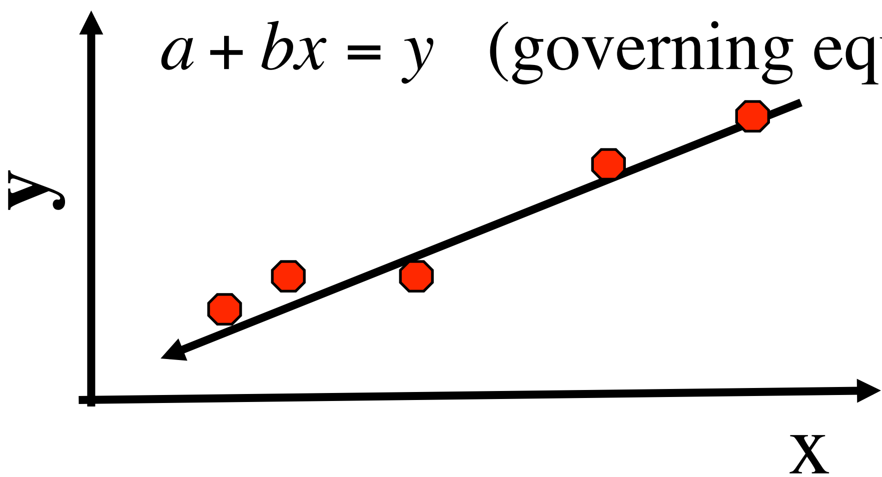
- Linearization
- Discretization
- Regularization (*a priori* information)

Recipe Step 1: Linearize

At the end of the day, write your data as a sum of some unknown coefficients multiply by the independent variable.

<u>Simple inverse problem</u>: find intercept and slope of a linear trend.

$a + bx = y$ (governing equation)



$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \dots & \dots \\ 1 & x_n \end{bmatrix} \bullet \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{bmatrix}$$

<u>Slightly more difficult inverse problem:</u>

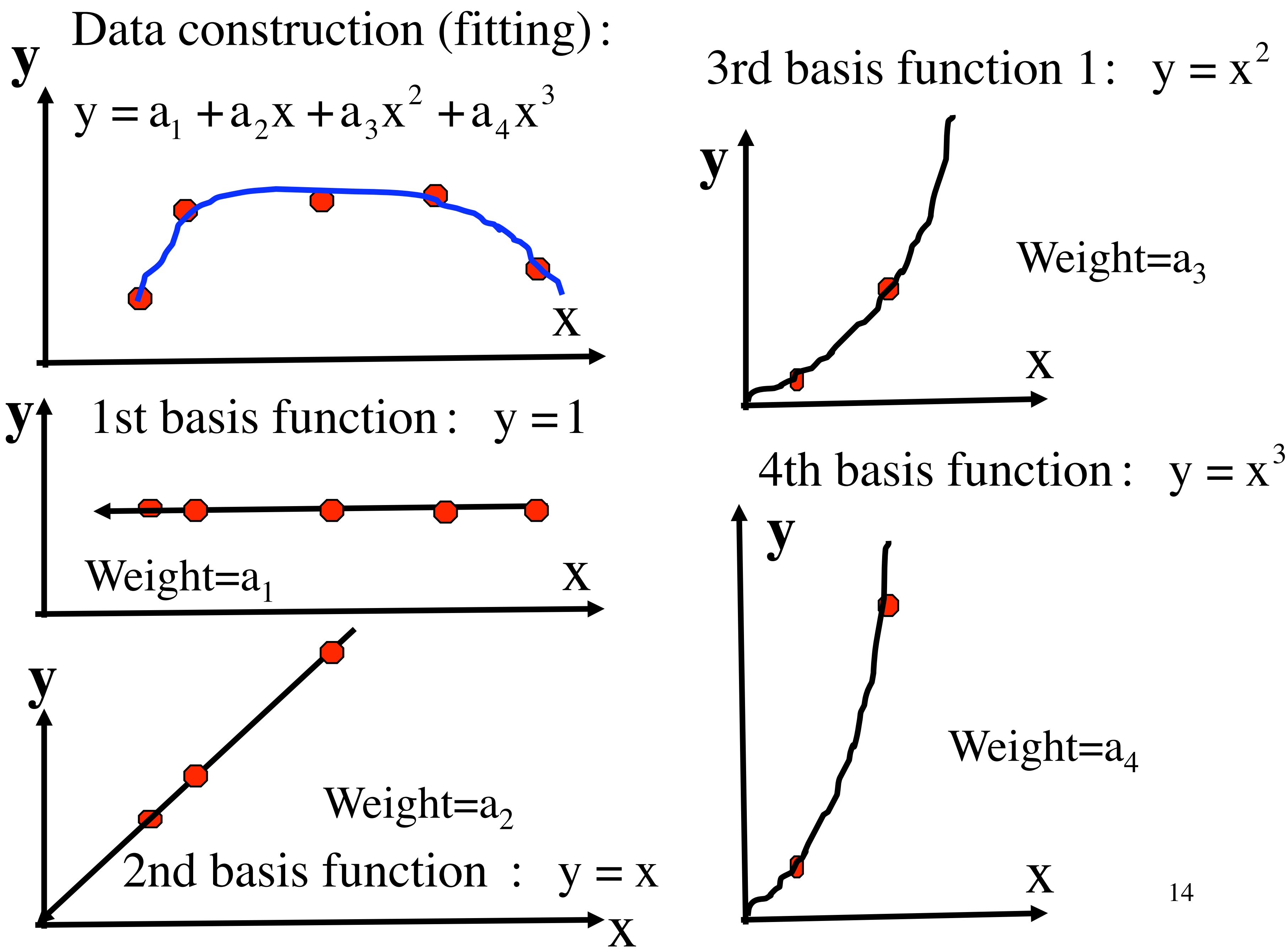
$$\mathbf{A} \cdot \mathbf{X} = \mathbf{B}$$

Cubic Polynomial Interpolation (governing equation)

$$a_1 + a_2 x_i + a_3 x_i^{\ 2} + a_4 x_i^{\ 3} = b_i$$

$$\begin{bmatrix} x_1^{\ 0} & x_1^{\ 1} & x_1^{\ 2} & x_1^{\ 3} \\ x_2^{\ 0} & x_2^{\ 1} & x_2^{\ 2} & x_2^{\ 3} \\ \dots & \dots & \dots & \dots \\ x_n^{\ 0} & x_n^{\ 1} & x_n^{\ 2} & x_n^{\ 3} \end{bmatrix} \bullet \begin{bmatrix} a1 \\ a2 \\ a3 \\ a4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ \dots \\ b_n \end{bmatrix}$$

Basis Functions and Construction of Data



Travel time (or slowness) inversions:

$$t = \int_0^{\Delta} \frac{1}{v} ds \quad \Rightarrow \delta t = \int_0^{\Delta} -\frac{1}{v^2}\delta v ds = \int_0^{\Delta} -\frac{1}{v}\left(\frac{\delta v}{v}\right) ds$$

$\dfrac{\delta v}{v}$ ---> use *basis* functions to represent

$$\delta t = \int_0^{\Delta} -\frac{1}{v} \sum_r \sum_\theta \sum_\phi x_{ijk}\, {}_r f_\theta{}^\phi ds$$

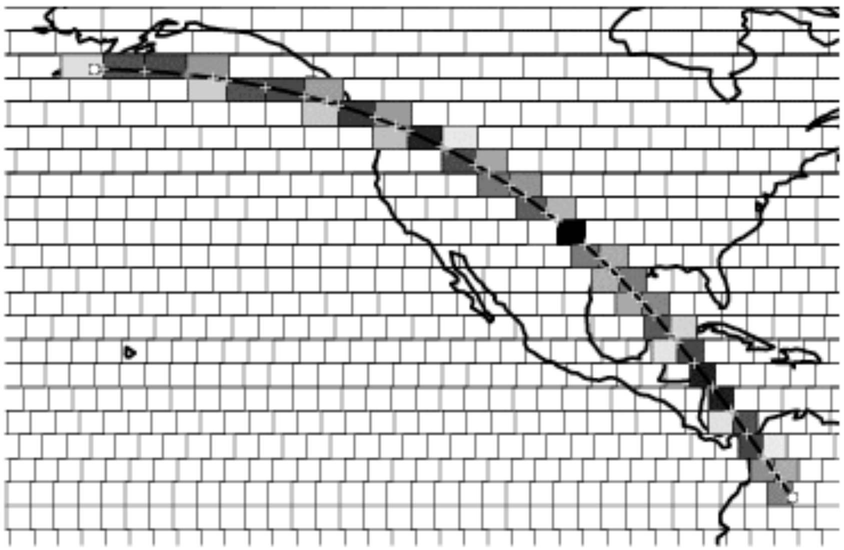
$$\delta t = \sum_r \sum_\theta \sum_\phi x_{ijk} \boxed{\int_0^{\Delta} -\frac{1}{v}\, {}_r f_\theta{}^\phi ds}$$

"sensitivity kernel", Elements form an "A" matrix

How to go from here?

Solve for x_{ijk} (which are weights to the original basis functions). They corresponds to a unique representation of of seismic wave speed Perturbation. This process is often referred to as "Travel-time tomography".

Recipe, Step 2: **Discretize!**



For a set of seismic rays $i = 1 \rightarrow M$, calculate the length spent in each of $j = 1 \rightarrow N$ grid boxes, in each of which it accumulates a proportional fraction of the total travel-time anomaly δt.

$$\delta t_i = L_{ij}\delta s_j \quad \text{or} \quad \boldsymbol{\delta}\mathbf{t} = \mathbf{L} \cdot \boldsymbol{\delta}\mathbf{s} \tag{5}$$

$$\text{M travel-time anomalies} \begin{bmatrix} \vdots \\ \delta t_i \\ \vdots \end{bmatrix} = \begin{bmatrix} & \vdots & \\ \ldots & L_{ij} & \ldots \\ & \vdots & \end{bmatrix} \times \begin{bmatrix} \vdots \\ \delta s_j \\ \vdots \end{bmatrix} \text{N slowness perturbations} \tag{6}$$

M×N **sensitivity matrix**

Least-Squares Solutions

Suppose we have a simple set of linear equations

$$\mathbf{A\,X = d}$$

We can define a simple scalar quantity E

$$E = \varepsilon^T \varepsilon = (\mathbf{AX} - \mathbf{d})^T (\mathbf{AX} - \mathbf{d}) = \left\| (\mathbf{AX} - \mathbf{d}) \right\|^2$$

Mean square error (or total error)

Error function

We want to minimize the total error, to do so, find first derivative of function ***E*** and set to 0.

So, do $\frac{\partial \mathbf{E}}{\partial \mathbf{X}} = 0$, we should have

$$2\left\| (\mathbf{AX} - \mathbf{d}) \right\| \mathbf{A} = 0 \longrightarrow \mathbf{A}^T\mathbf{AX} = \mathbf{A}^T\mathbf{d}$$

This is known as the system of normal equations.

$$\longrightarrow \mathbf{X} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{d}$$

So this involves the inversion of the term $\mathbf{A}^T\mathbf{A}$**,** This matrix is often called the ***inner-product matrix, or Toeplitz matrix*****. The solution is called the least -squares solution,** while $\mathbf{X}=\mathbf{A}^{-1}\,\mathbf{d}$ is **not** a least squares solution**.**

Pre-conditioning for ill-conditioned inverse problem (damping, smoothing, regularization. *Purpose: Stabilize, enhance smoothness/simplicity*)

Lets use the same definition $E = \varepsilon^T \varepsilon = \|\mathbf{AX} - \mathbf{d}\|^2$

Define: An *objective function* **J** where $\mathbf{J} = \mathbf{E} + \mu\|\mathbf{X}\|^2$

where μ is the damping or regularization parameter.

$$\frac{\partial \mathbf{J}}{\partial \mathbf{X}} = \frac{\partial \mathbf{E}}{\partial \mathbf{X}} + \frac{\partial(\mu\|\mathbf{X}\|^2)}{\partial \mathbf{X}} = 2\|(\mathbf{AX} - \mathbf{d})\|\mathbf{A} + 2\mu\|\mathbf{X}\|$$

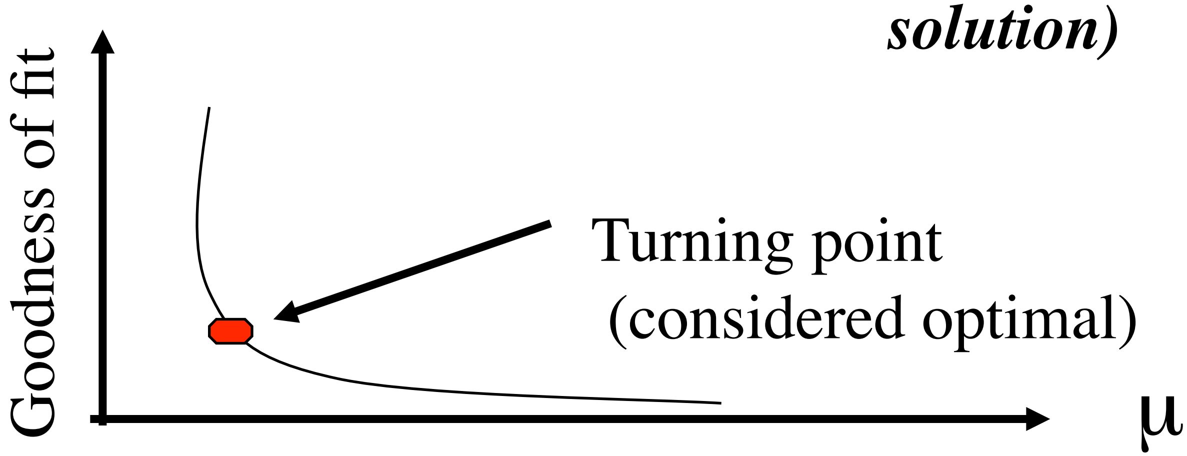
Minimize the above by $\frac{\partial \mathbf{J}}{\partial \mathbf{X}} = 0$

$$\longrightarrow (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})\mathbf{X} = \mathbf{A}^T\mathbf{d}$$

Left multiply by $(\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}$ **I** is identity matrix

$$\longrightarrow \mathbf{X} = (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T\mathbf{d}$$

(Damped Least Squares solution)



Main Reasons for Damping,

Purpose 1: Damping stabilizes the inversion process in case of a singular (or near singular matrix). Singular means the determinant = 0. Furthermore, keep in mind there is a slight difference between a singular A matrix and a singular A^TA. A singular A matrix don't always get a singular A^TA. A^TA inversion is more stable.

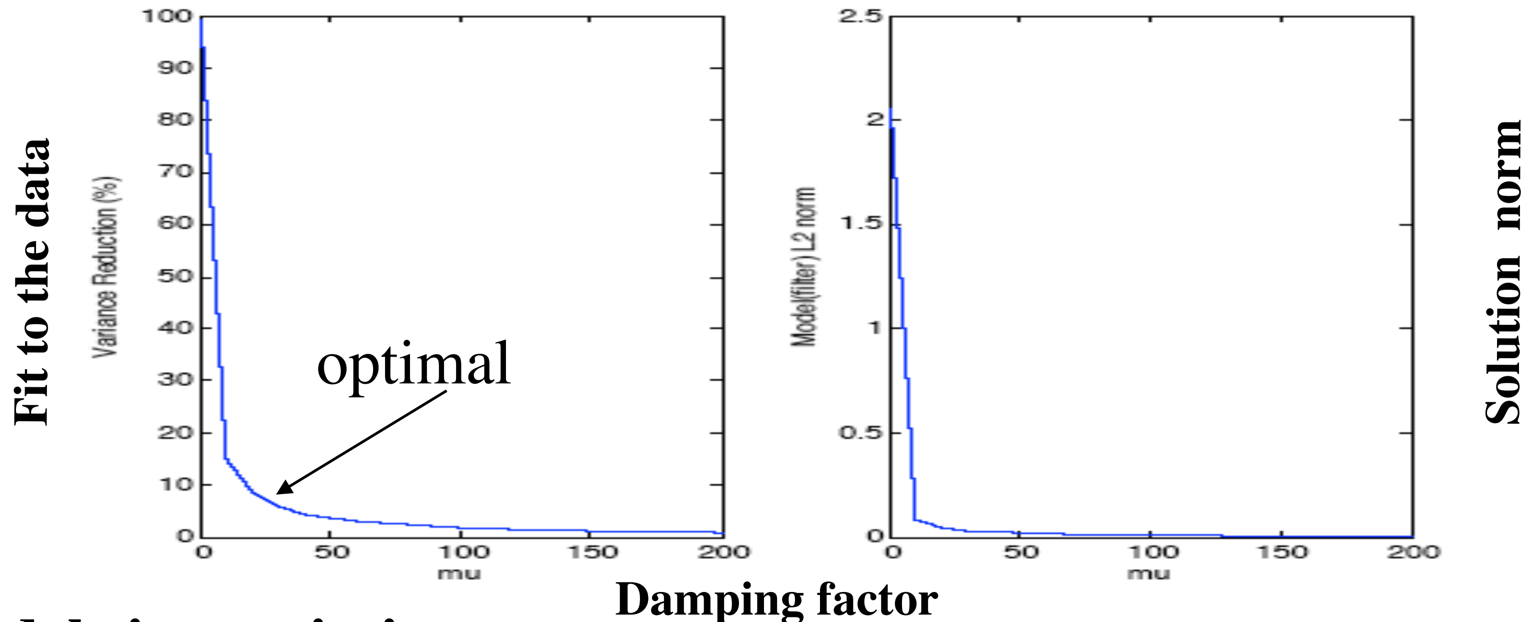
In a mathematical sense, why more stable? Related to Pivoting -→ one can show that if the diagonal elements are too small, error is large (the reason for partial pivoting). Adding a factor to diagonal will help keep the problem stable.

When adding to the diagonal of a given A^TA matrix, the matrix condition is modified. As a result, **A * X = d** problem is also modified. So we no longer solve the original problem exactly, but a modified one depending on the size of μ. We are sacrificing some accuracy for stability and for some **desired properties** in solution vector.

Purpose 2: **obtain some desired properties in the solution vector X. The most important property = smoothness.**

Damped Least Squares: $(\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})\mathbf{X} = \mathbf{A}^T\mathbf{D}$

Tradeoff curves



Model size variation

The sum of the squared values of elements of **X** (***norm*** of **X**) goes to 0 since when we increase μ, $\mathbf{A}^T\mathbf{A}$ matrix effectively becomes diagonal (with a very large number on the diagonal), naturally, ***X ----> 0 as a_{ii} ----> infinity.***