Part 1:
Inverse Problems

M D Sacchi
University of Alberta

LAPIS 2019 - UNLP - April 9, 2019

Course information

- Email: msacchi@ualberta.ca
- Web: https://sites.ualberta.ca/~msacchi/LAPIS/
- Approval: Assignment consisting of programming exercises due by April 30, 2019
- Delivery format: PDF by email

Inverse Problems

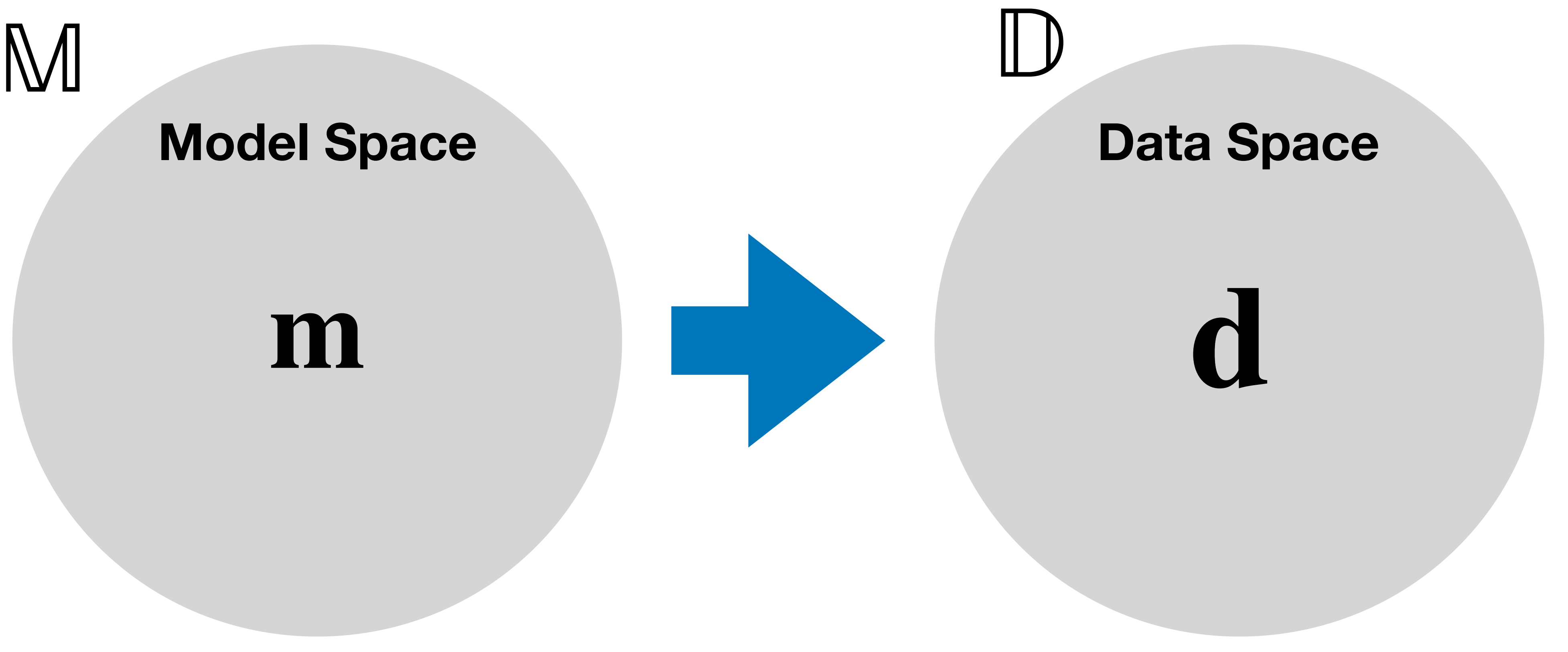
- An Inverse Problem is a mathematical problem where one attempts to estimate models that explain observations. We often name the observations **d** (data)

- Observations are generally measured on the surface of the earth and are discrete in time and space

- The subsurface is described by properties (density, velocity, reflectivity, resistivity, etc). These properties exist everywhere. We will refer to these properties as **m** (model parameters)

Two spaces

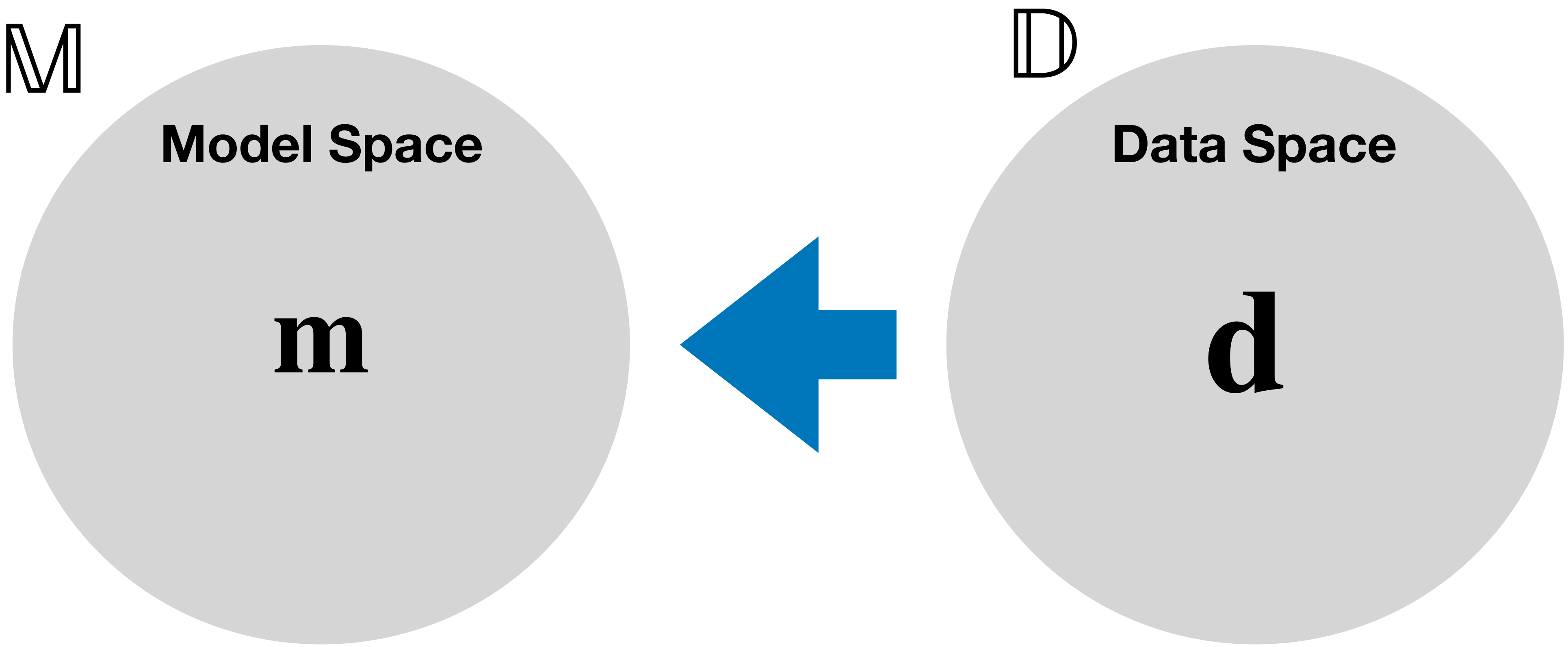
- Forward Problem



$$F[\mathbf{m}] = \mathbf{d}$$

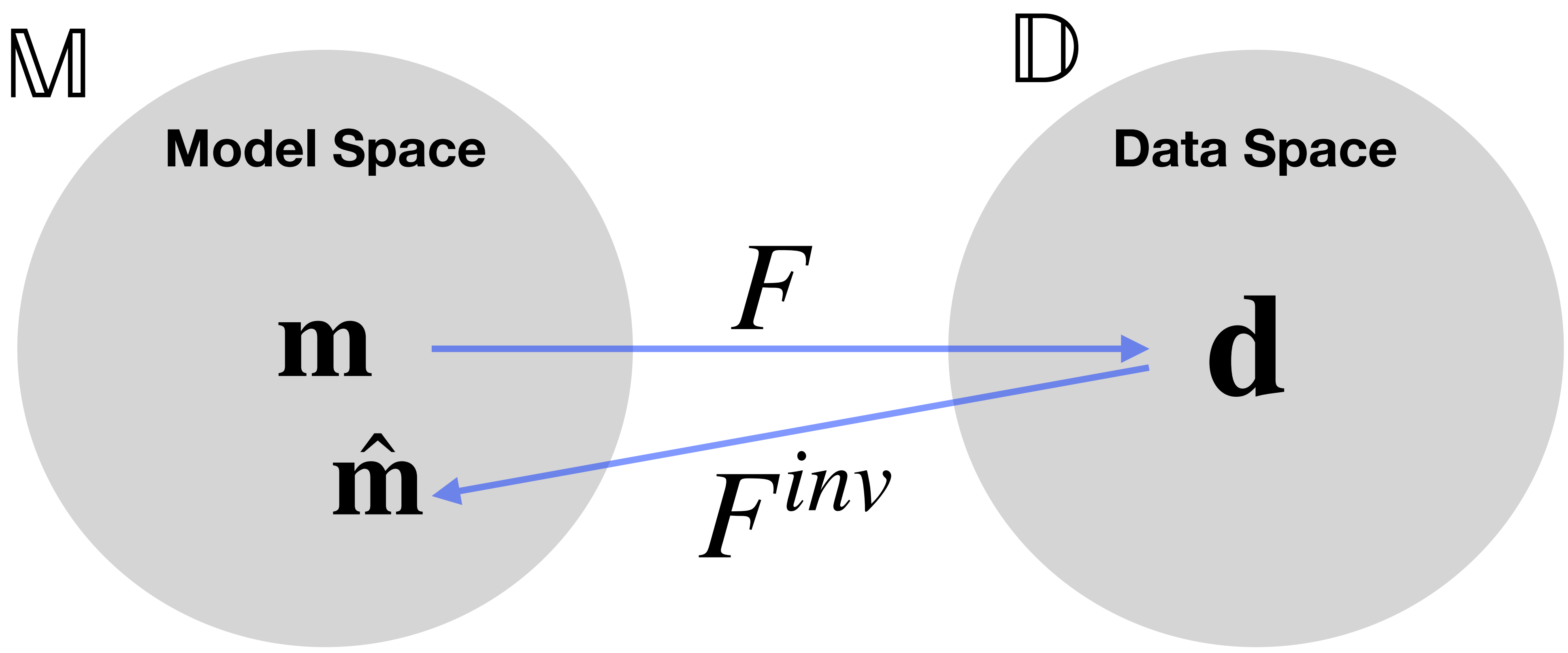
Two spaces

- Inverse Problem



$$F^{inv}[\mathbf{d}] = \mathbf{m}$$

It is more complicated…



$\hat{\mathbf{m}}$: **Solution**

Geophysics and IPs

Data (What you can measure)	Model (What you would like to know)	Method (Course/Materia)
Gravity anomalies	Density	Potential Field Methods
Electrical potential	Resistivity	Potential Field Methods
Electrical and Magnetic Field	Electrical conductivity	EM/MT methods
Magnetic Fields	Susceptibility	Potential Field Methods
Seismic Waves	Velocities	Seismic Methods

ILL-posed problems

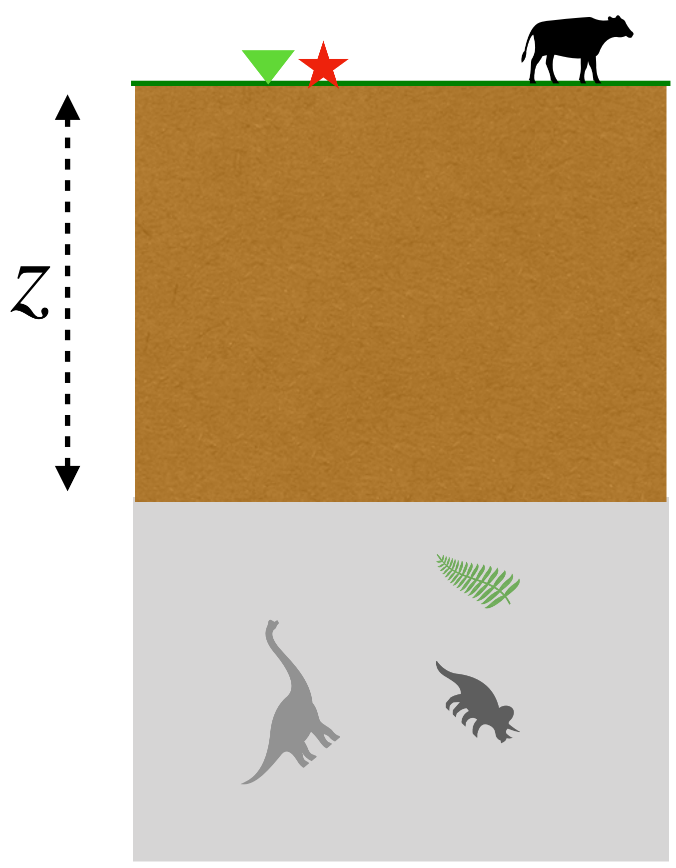
- Well-posed problem. A problem is said to be well-posed when
  - There is a solution
  - The solution is unique
  - The solution is stable
- If one of the above is not true, the problem is called an ill-posed problem
- Typical geophysical inverse problems are ill-posed problems (2 and 3 are not true)

ILL-posed problems

- There is a solution: YES
  - There is a solution otherwise we wouldn't be here
  - e.g. Rocks have density causing gravity anomalies

ILL-posed problems

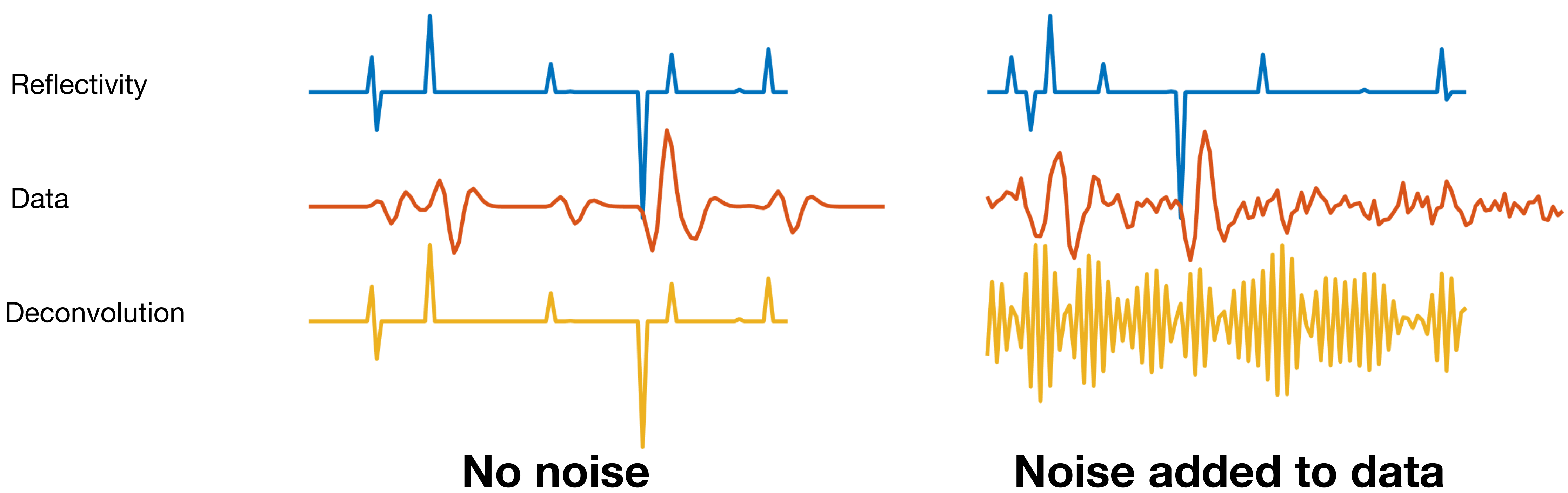
- The solution is unique: Generally NO
  - e.g., Depth-Velocity ambiguity $t = \frac{2z}{v}$



$$t = \frac{2 \times 1000m}{2000m/s} = \frac{2 \times 1500m}{3000m/s}$$

ILL-posed problems

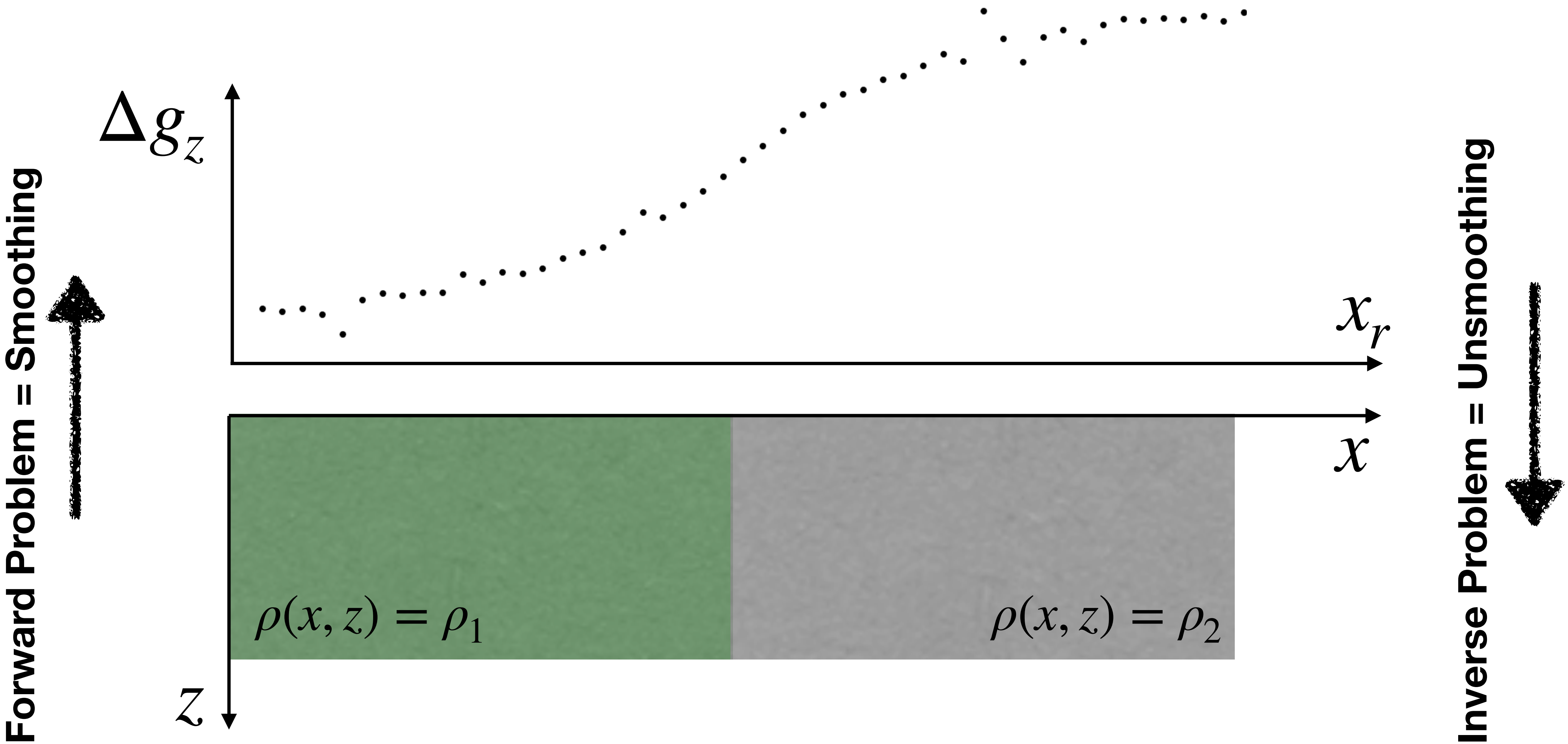
- The solution is stable: Not true
  - A small perturbation in the data causes a large perturbation in the solution
  - e.g. Deconvolution



Why geophysical problems are ill-posed?

- Not enough data (insufficient spatial data or insufficient bandwidth)
- Presence of noise leads to unstable solutions
- The forward problem itself leads to ill-posed problems. Forward problem smoothes material properties. Seismic waves, gravity anomalies, electrical potential, etc are nice smooth functions of space and time. Subsurface properties, on the other hand, might or might not be smooth!
- Next slide illustrated the aforementioned concept.

Example: Inversion of gravity anomalies



$$\Delta g_z(x_r, z_r = 0) = \int_{x,y} G(x, y \,|\, x_r, z_r = 0)\, \Delta\rho(x, y)\, dx\, dy$$

Key points

- Previous example shows an interesting feature of the Forward Problem:
  - Observations (data) are a weighted average of density. The averaging kernel smoothes the density anomaly
  - Inverse Problem: Recovering the density from the data entails the opposite of smoothing (un-smoothing); an unstable operation
  - Smoothing = Low pass operator = Stable
  - Un-smoothing = High pass operator = Unstable
  - Solving an inverse problem in many cases entails controlling instability

Linear inverse problems

- Many inverse problems in seismology entail solving integral equations. Others entail solving PDEs. We will start with simple problems that can be written as integral equations that after discretization lead to discrete system of equations.
    - Integral equations
    - Linear discrete system of equations
    - Regularization methods
    - Connection to exploration seismology:
        - Deconvolution
        - AVO inversion
        - Linearized seismic imaging (Forward and Adjoint Operators)
            - Migration & Least-Squares migration

Notation

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{pmatrix}$$

- Real vector

$$\mathbf{u} : N \times 1 \longrightarrow \|\mathbf{u}\|_2^2 = \mathbf{u}^T\mathbf{u} = \sum_{i=1}^{N} u_i^2$$

- Complex vector

$$\mathbf{u} : N \times 1 \longrightarrow \|\mathbf{u}\|_2^2 = \mathbf{u}^H\mathbf{u} = \sum_{i=1}^{N} u_i u_i^* = \sum_{i=1}^{N} |u_i|^2$$

- The ell-2 norm

$$l_2 = \|\mathbf{u}\|_2$$

Linear inverse problems

- Fredholm integral equation of 1st kind

$$d(r_j) = \int_X G(r_j, x)\, m(x)\, dx \qquad j = 1 \ldots N$$

- We can discretize the model

$$x_k = x_0 + (k-1)\Delta x \qquad k = 1 \ldots M$$

$$d(r_j) = \sum_{k=1}^{M} G(r_j, x_k)\, m(x_k)\, \Delta x \qquad j = 1 \ldots N$$

- Which leads to a system of equations

$$\mathbf{d} = \mathbf{G}\mathbf{m}, \quad \mathbf{d} \in R^N, \quad \mathbf{m} \in R^M$$

Examples

$$d(r_j) = \int_X G(r_j, x)\, m(x)\, dx \qquad j = 1 \ldots N$$

$$s(t_i) = \int w(t_i - \tau)\, r(\tau)\, d\tau$$

Convolution

$$T(r_j) = \int_0^{r_j} s(l)\, dl$$

Travel-time tomography

$$d(\omega, s_i, r_j) = \int_x \int_z B(\omega, x, y \,|\, s_i, r_j) m(x, z) dx\, dz$$

Born imaging

Forward and Adjoint Operators

- **Two special operations (or operators?)**

Forward : $\mathbf{d} = \mathbf{L}\mathbf{m}$

Conjugate transpose or adjoint : $\tilde{\mathbf{m}} = \mathbf{L}'\mathbf{d}$

- $\mathbf{L}$ is a linear operator (Forward modelling operator).
- Why are them special?
  - Iterative solvers for large inverse problems only need to know how to evaluate $\mathbf{L}[\,\cdot\,]$ and $\mathbf{L}'[\,\cdot\,]$
  - I usually interpret operators as matrices. In reality, operators are codes applied *on the flight* (al voleo)
  - *We will see these operators everywhere today*

Forward and Adjoint Operator

- **Two special operations (or operators?)**

$$\textbf{Forward} : \mathbf{d} = \mathbf{L}\mathbf{m} \quad \textbf{(1)}$$

$$\textbf{Conjugate transpose or adjoint} : \tilde{\mathbf{m}} = \mathbf{L}'\mathbf{d} \quad \textbf{(2)}$$

- Replace (1) into (2)

$$\tilde{\mathbf{m}} = \mathbf{L}'\mathbf{L}\mathbf{m}$$

- Questions:
  - When can I say the adjoint model is a goo representation of the true mode?
  - Can I remove $\mathbf{L'L}$ from the mode obtained via the adjoint ?
  - Migration vs. Least-squares Migration: *Least-squares migration of incomplete reflection data. Nemeth et al. GEOPHYSICS(1999),64(1)*

Forward and Adjoint Operators

- Two special operations (or operators?)

Forward : $\mathbf{d} = \mathbf{L}\mathbf{m}$ (1)

Conjugate transpose or adjoint : $\tilde{\mathbf{m}} = \mathbf{L}'\mathbf{d}$ (2)

$$\tilde{\mathbf{m}} = \mathbf{L}'\mathbf{L}\mathbf{m}$$

$\tilde{\mathbf{m}}$ **is a distorted version of** $\mathbf{m}$

Linear Inverse Problems

- Including noise into the problem

$$\mathbf{d} = \mathbf{L}\mathbf{m} + \mathbf{e}$$

$$\mathbf{d} : N \times 1 \qquad \mathbf{m} : M \times 1 \qquad \mathbf{L} : N \times M$$

- $\mathbf{e}$ represents “nice” Gaussian additive noise

- Given the vector of observed data $\mathbf{d}$, one wishes to estimate the vector of model parameters $\mathbf{m}$

Linear Inverse Problems

Cost function for least-squares problems

$$J = \|\mathbf{e}\|_2^2 = \|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2$$

The principle is simple, find $\mathbf{m}$ that minimize the sum of the squares of the errors

$$\nabla J = 0 \rightarrow (\mathbf{L}'\mathbf{L})\mathbf{m} = \mathbf{L}'\mathbf{d}$$

To compute the solution we now have to invert a matrix. Assume the matrix is invertible then

$$\mathbf{m}_{sol} = (\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\mathbf{d}$$

Linear Inverse Problems

The latter is a naive solution because inverse problems are ill-posed and regularization is needed. In other words, one cannot safely compute:

$$(\mathbf{L}'\mathbf{L})^{-1}$$

- Either is non invertible
- Or the Matrix has a large condition number (It will amplify noise)
- Condition Number $$\kappa = \frac{\lambda_{max}(\mathbf{L}'\mathbf{L})}{\lambda_{min}(\mathbf{L}'\mathbf{L})}$$

Matrix computations
GH Golub, CF Van Loan, 2013 (4th edition)

Regularization

- Main idea of regularization methods
    - Take an ill-posed problem and turn into a well-posed problem by introducing constraints that lead to a stable solution. The solution often depends on a trade-off parameter. Therefore, regularization methods create a family of solutions with different properties (e.g. smoothness). One can alter the solution by modifying the tradeoff parameter
    - *Often call Tikhonov regularization, after Andrey Nikolayevich Tikhonov 1906-1993.*

Regularization

- Replace cost function to minimize

$$J = \|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2$$

- by

$$J = \|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2 + \mu\|\mathbf{W}\,\mathbf{m}\|_2^2$$

- μ Is the infamous trade-off parameter or regularization parameter. Why infamous?

- $\mathbf{W}$ is a matrix/operator of weights

Regularization

- Evaluate the solution, as usual, by minimizing the cost function

$$\begin{aligned}\mathbf{m}_{sol} &= \underset{\mathbf{m}}{\text{argmin}}\ J \\ &= \underset{\mathbf{m}}{\text{argmin}}\ \{ \|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2 + \mu \|\mathbf{W}\,\mathbf{m}\|_2^2 \}\end{aligned}$$

- Taking derivatives and equating them to zero

$$\nabla J = 0 \rightarrow \mathbf{m}_{sol} = (\mathbf{L}'\,\mathbf{L} + \mu \mathbf{R})^{-1} \mathbf{L}'\,\mathbf{d}$$

$$\mathbf{R} = \mathbf{W}'\,\mathbf{W}$$

Regularization

- Anatomy of the cost function

$$J = \underbrace{\|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{W}\,\mathbf{m}\|_2^2}_{Model\ Norm}$$

$$J = \underbrace{\|\mathbf{e}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{u}\|_2^2}_{Model\ Norm}$$

- To minimize J is equivalent to simultaneously minimize $\mathbf{e}$ and $\mathbf{u}$

Regularization

- To minimize J is equivalent to simultaneously minimize $\mathbf{e}$ and $\mathbf{u}$

$$\mathbf{e} = \mathbf{d} - \mathbf{Lm} \approx \mathbf{0}$$ *Make residuals small*

$$\mathbf{u} = \mathbf{Wm} \approx \mathbf{0}$$ *Make bad features of* $\mathbf{m}$ *small*

- $\mathbf{W}$ is a high-pass operator therefore it penalizes roughness. $\mathbf{u}$ are amplified bad features of $\mathbf{m}$
- Examples of $\mathbf{W}$ are first and second order derivatives

Regularization with first order derivative smoothing

$$\mathbf{W} = \mathbf{D}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

$$\frac{\partial f(x)}{\partial x} \leftrightarrow ikF(k)$$

Derivative is high pass

Regularization with second order derivative smoothing

$$\mathbf{W} = \mathbf{D}_2 = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

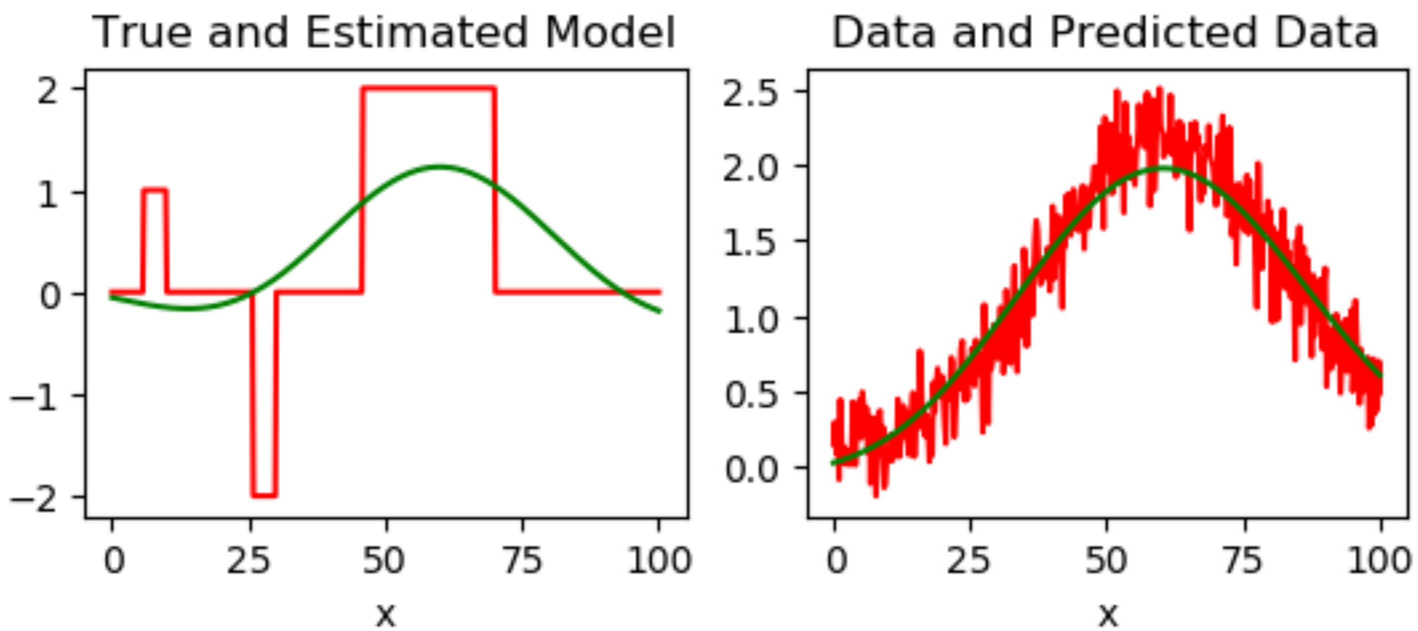
$$\frac{\partial^2 f(x)}{\partial x^2} \leftrightarrow - k^2 F(k)$$ *Derivative is high pass*

Example: Dampled LS solution

`Demo_1_Lapis_2019.ipynb`

$$\mathbf{m}_{sol} = \underset{\mathbf{m}}{\text{argmin}} \{ \|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2 + \mu \, \|\mathbf{m}\|_2^2 \}$$

$$\nabla J = 0 \rightarrow \mathbf{m}_{sol} = (\mathbf{L}'\mathbf{L} + \mu \, \mathbf{I})^{-1}\mathbf{L}'\mathbf{d}$$



Red: True Model | Observed data
Green: Estimated Model | Predicted data

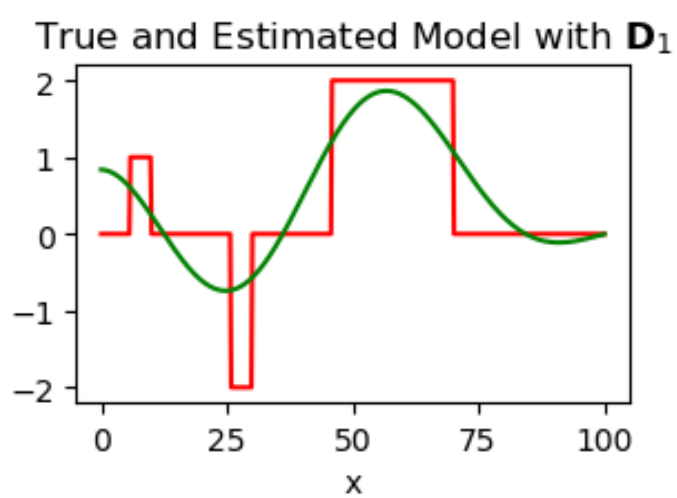
Example: Solution with smoothing

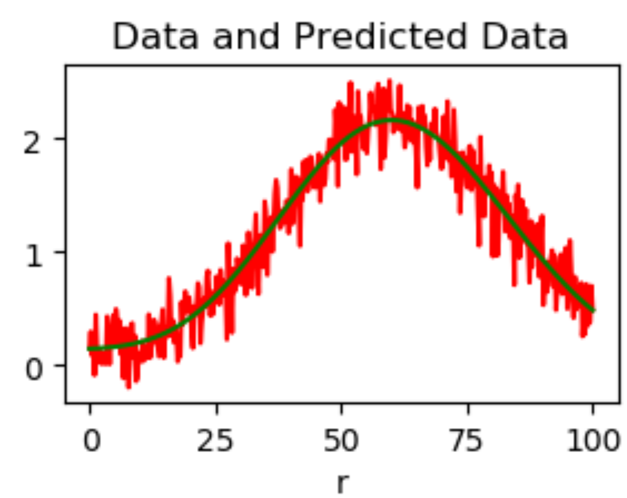
`Demo_1_Lapis_2019.ipynb`

$$\mathbf{m}_{sol} = \underset{\mathbf{m}}{\text{argmin}} \{ \|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2 + \mu \, \|\mathbf{W}\,\mathbf{m}\|_2^2 \}$$

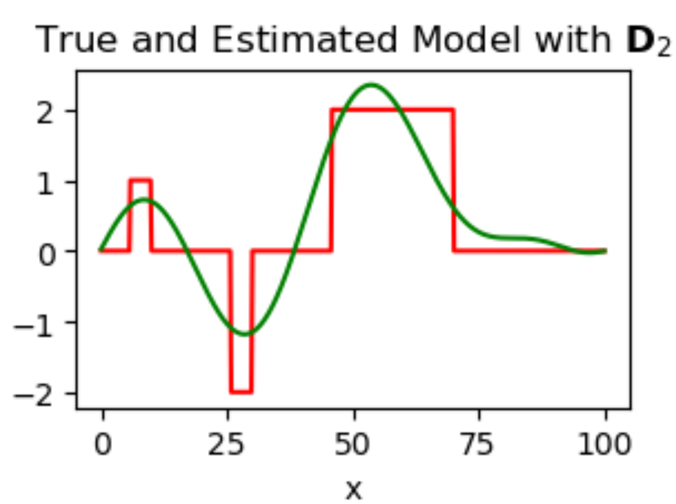
$$\nabla J = 0 \rightarrow \mathbf{m}_{sol} = (\mathbf{L}'\mathbf{L} + \mu \, \mathbf{W}'\mathbf{W})^{-1}\mathbf{L}'\mathbf{d}$$

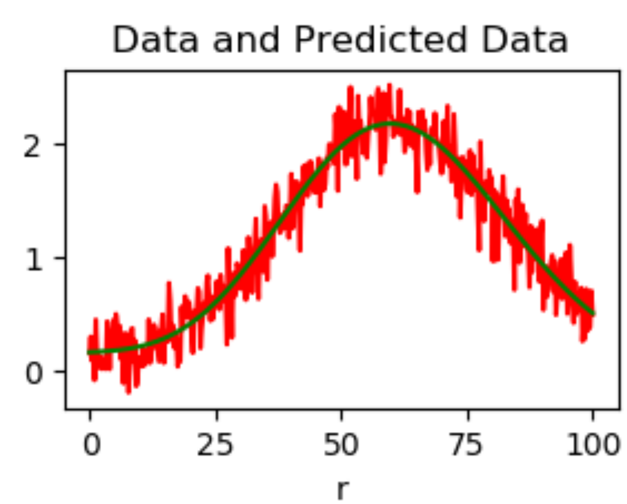
$\mathbf{W} = \mathbf{D}_1$





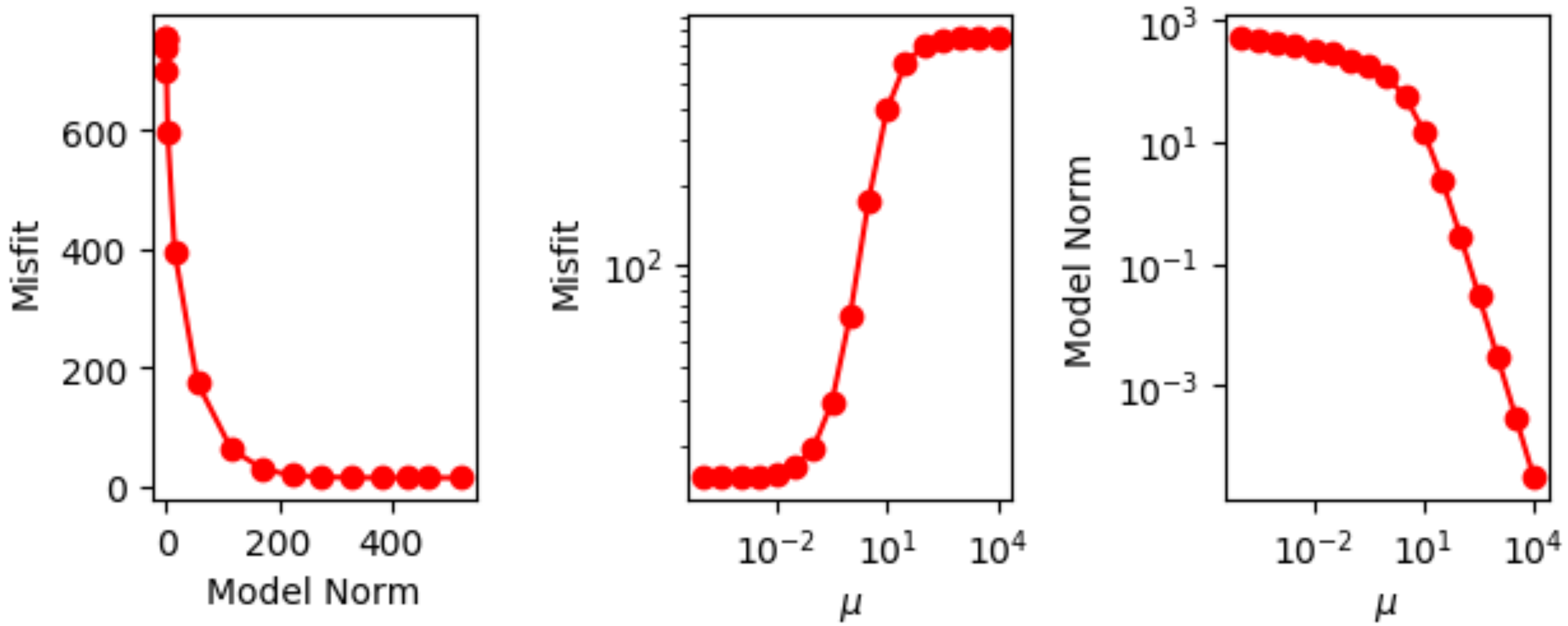
$\mathbf{W} = \mathbf{D}_2$





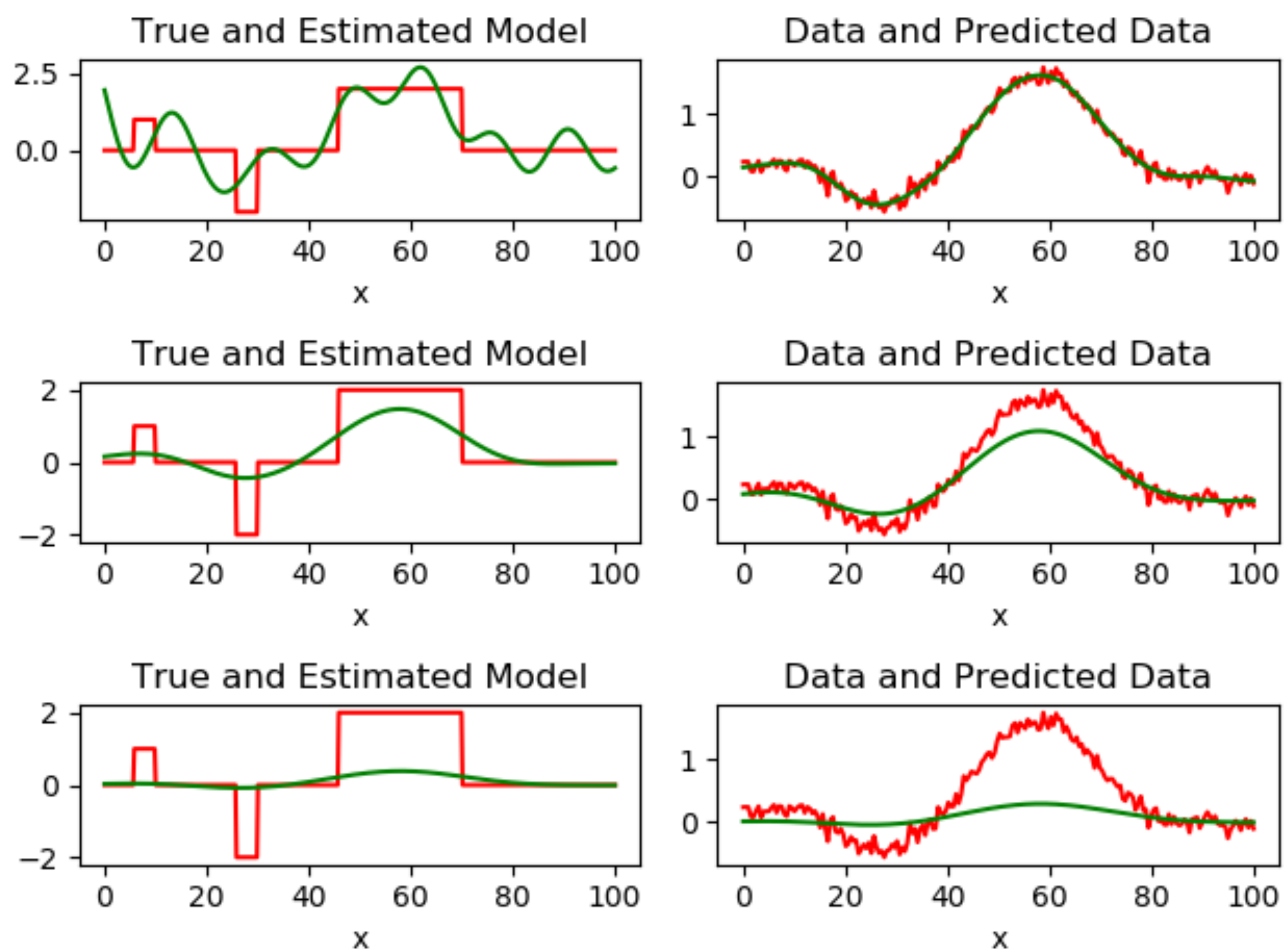
Trade-off curves (for Damped least-squares)

`Demo_1_Lapis_2019.ipynb`



Trade-off curves (for Damped least-squares)

`Demo_1_Lapis_2019.ipynb`



$\mu = 0.0001$

$\mu = 0.1$

$\mu = 1$

Edge preserving regularization (EPR)

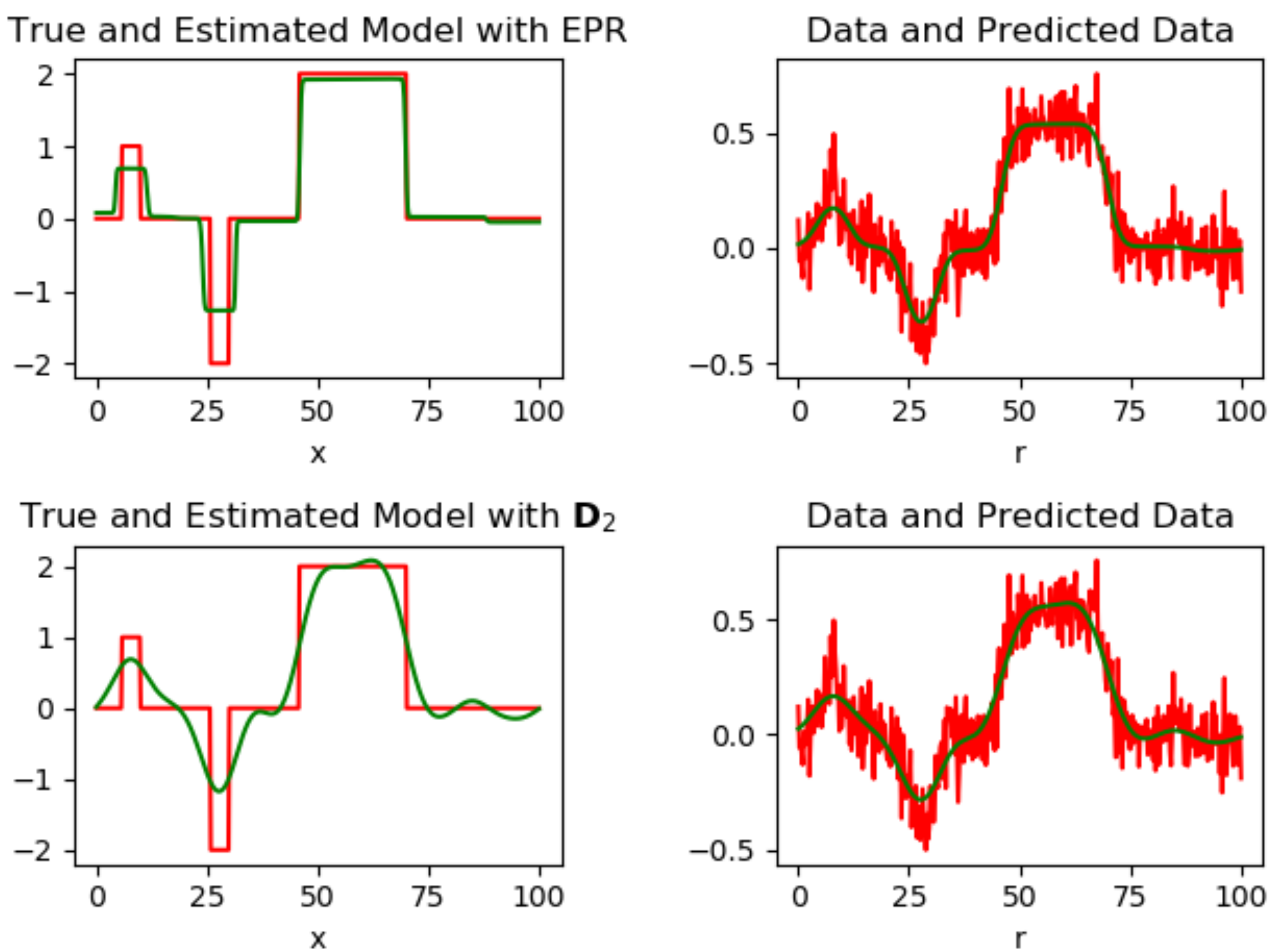
- Avoid smoothing to preserve edges
- We adopt the ell-1 norm of the first order derivative of model parameters
- Make the derivative of the model sparse

$$J = \underbrace{\|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{D}_1\,\mathbf{m}\|_1}_{Model\ Norm}$$

We ask $\mathbf{u} = \mathbf{D}_1\mathbf{m}$ **to be sparse**

Edge preserving regularization (EPR)

`Demo_2_Lapis_2019.ipynb`



EPR
(non-quadratic)

Second Order Derivative Reg.
(quadratic)

Edge preserving regularization (EPR)

- Quadratic leads to close form solution (linear system of equations)

$$J = \underbrace{\|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{W}\,\mathbf{m}\|_2^2}_{Model\ Norm}$$

- EPR: Non-quadratic regularization leads to non-linear solution that tries to recover edges

$$J = \underbrace{\|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{D}_1\,\mathbf{m}\|_1}_{Model\ Norm}$$

Edge preserving regularization (EPR)

- Non-quadratic regularization leads to non-linear solution

$$J = \underbrace{\|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{D}_1\,\mathbf{m}\|_1}_{Model\ Norm}$$

- Solution

$$\nabla J = 0 \rightarrow (\mathbf{L}'\,\mathbf{L} + \mu \mathbf{D}_1'\,\mathbf{Q}\,\mathbf{D}_1)\mathbf{m} = \mathbf{L}'\mathbf{d}$$

- Where

$$\mathbf{v} = \mathbf{D}_1\,\mathbf{m} \qquad Q_{ii} = \frac{1}{|\,v_i\,|}$$

- To avoid division by zero

$$Q_{ii} = \frac{1}{\epsilon + |\,v_i\,|}$$

Edge preserving regularization (EPR)

- Iterative re-weighted least-squares

$$\mathbf{m}^1 = \mathbf{m}_{initial}$$

For $k=1$ **until convergence**

$$\mathbf{v} = \mathbf{D}_1\,\mathbf{m}$$

$$Q_{ii}^k = \frac{1}{\epsilon + |v_i^k|}$$

$$\mathbf{m}^{k+1} = (\mathbf{L}'\,\mathbf{L} + \mu\mathbf{D}_1'\,\mathbf{Q}^k\,\mathbf{D}_1)^{-1}\mathbf{L}'\mathbf{d}$$

End

Sacchi, 1997, Re-weighting strategies for seismic deconvolution, GJI

Connection to sparsity

- This cost function generates a sparse solution

$$J = \underbrace{\|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{m}\|_1}_{Model\ Norm}$$

- Often used for Deconvolution
- Pre-stack data Reconstruction *(Liu and Sacchi, 2004, Geophysics; Hermann, 2010, Geophysics)*
- Radon Transfroms, etc etc etc *(Sacchi & Ulrych, 1995, Geophysics)*
- AVO Inversion (*Alemie and Sacchi, 2011, Geophysics*)

Connection to sparsity

- Make $\mathbf{m}$ sparse:

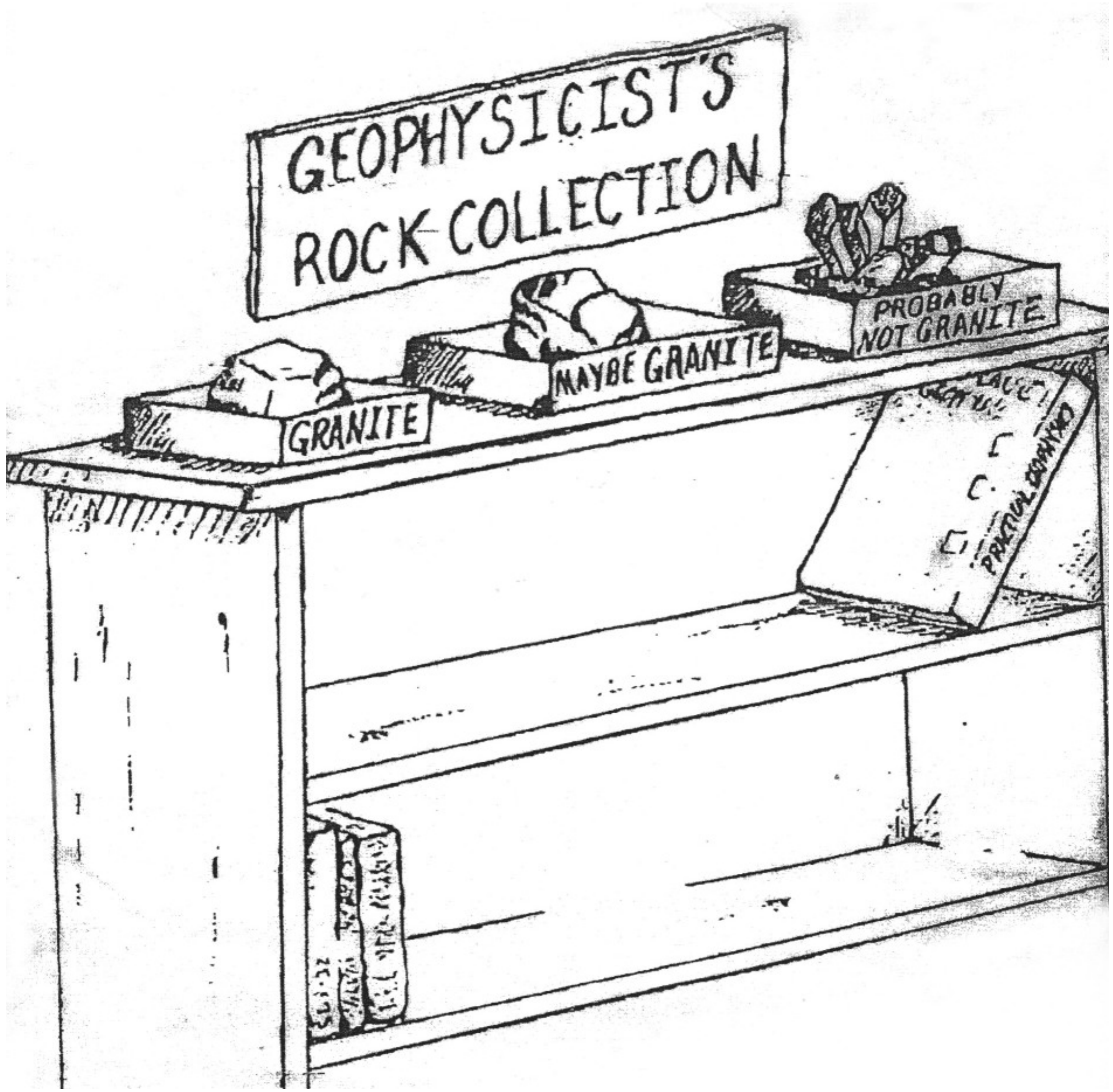
$$J = \underbrace{\|\mathbf{d} - \mathbf{Lm}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{m}\|_1}_{Model\ Norm}$$

- Make the derivative of $\mathbf{m}$ sparse = Make $\mathbf{m}$ blocky

$$J = \underbrace{\|\mathbf{d} - \mathbf{Lm}\|_2^2}_{Error\ Norm} + \mu \underbrace{\|\mathbf{D}_1\,\mathbf{m}\|_1}_{Model\ Norm}$$

Sparsity

- **IRLS is the simplest solver one can imagine**
  - Many new solvers in recent year
    - ISTA, FISTA, SALSA, SPG-L1, ADMM, L1-Magic,etc etc etc
    - I usually use IRLS or FISTA
    - ISTA/FISTA: Only need to know how to apply $\mathbf{L}$ and $\mathbf{L}$’ *(on-the-flight)*


GEOPHYSICIST'S
ROCK COLLECTION
GRANITE
MAYBE GRANITE
PROBABLY
NOT GRANITE
PRACTICAL GEOPHYSICS