Linear Inverse Problems: some seismic data processing applications

UNLP

Inverse Problems

Mauricio D Sacchi

Problems in exploration seismology where regularization methods are important

- **Data Reconstruction**
- **Focusing with Radon Transforms**
- **Deconvolution**
- **Probing a reflector with seismic waves (AVO)**
- **Imaging (regularized LS imaging/Migration)**

These problems can all be tackled with regularization methods

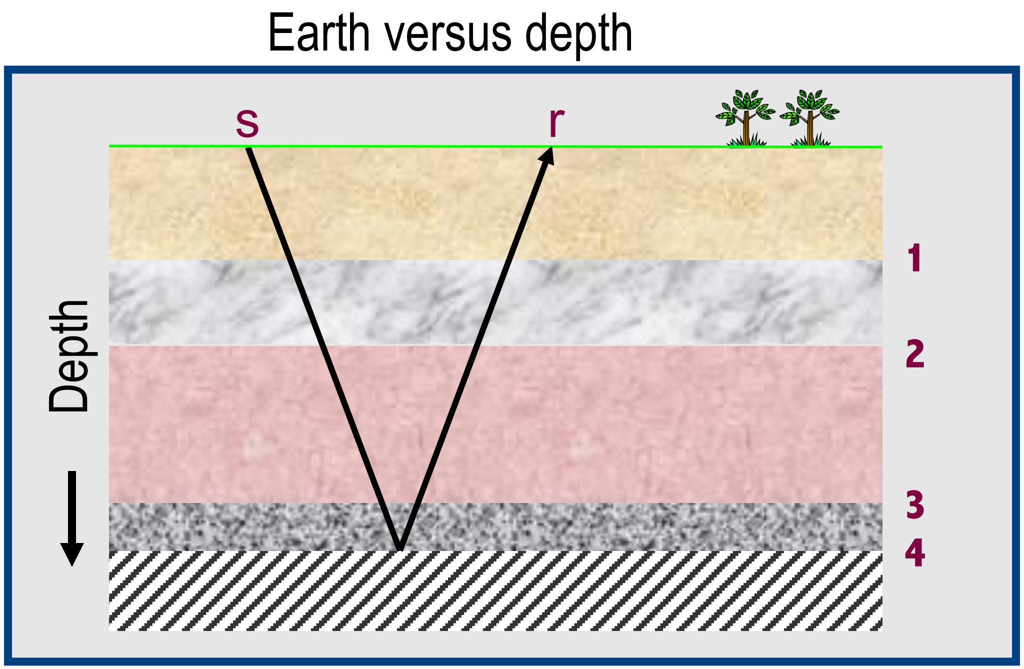
<u>Deconvolution</u>

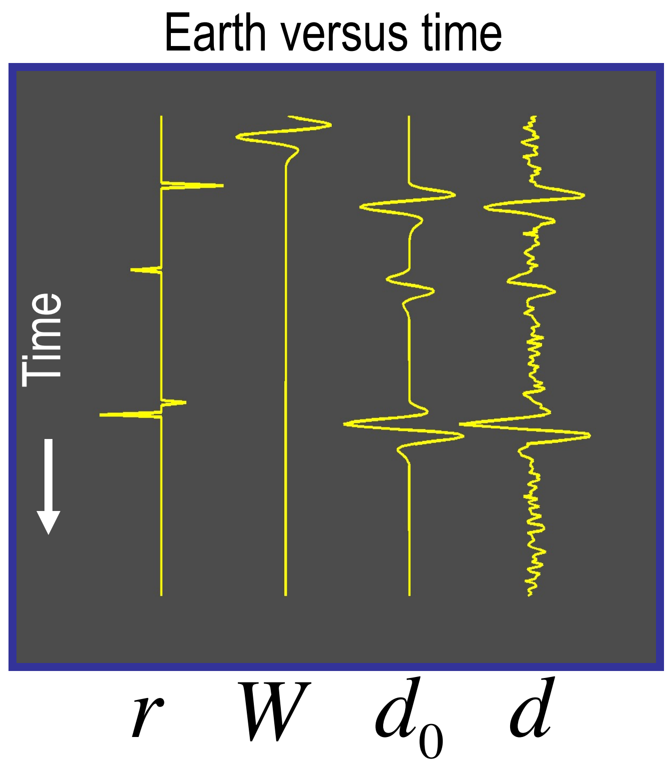
- debluring

- equalization

Increase resolution and our ability to “see” thin layers, stars, imperfections in materials, etc etc

Seismic Deconvolution





$$d(t) = \int w(\tau - t) r(\tau) d\tau + n(t)$$

$$d = d_0 + n$$

$$= Wr + n$$

$$d(t) = \int w(\tau - t) r(\tau) d\tau + n(t)$$

$$d = Wr + n$$

Ill-posed ?

This term can be extremely large when the source amplitude spectrum is small

$$D(\omega) = D_0(\omega) + N(\omega)$$

$$= W(\omega) R(\omega) + N(\omega)$$

$$\hat{R}(\omega) = \frac{D(\omega)}{W(\omega)} = \frac{D_0(\omega) + N(\omega)}{W(\omega)} = R(\omega) + \frac{N(\omega)}{W(\omega)}$$

Seismic Deconvolution

The <u>very Naïve</u> Solution (Just try to fit the data…)

$$d = W\, r + n$$

$$J = \| Wr - d \|_2^2$$

$$\nabla J = W^T W\, r - W' d = 0$$

$$\hat{r} = (W^T W)^{-1} W^T d$$

Small perturbations on d will produce large perturbations on r

Seismic Deconvolution

$$d = W\, r + n$$

Low Resolution Solution:

$$J = \| Wr - d \|_2^2$$

$$r = (W^T W)^{-1} W^T d$$

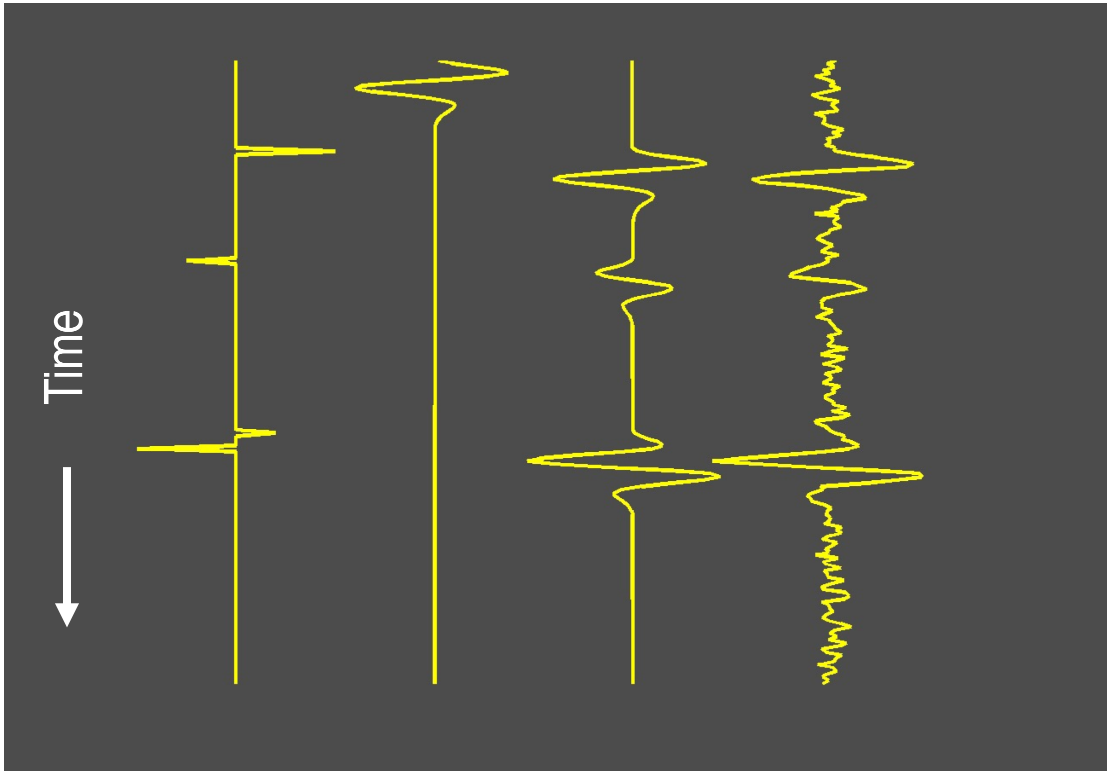
$$c\, I \approx W^T W$$

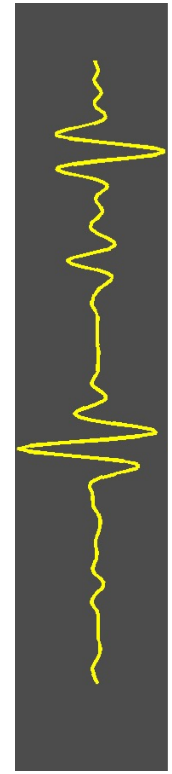
$$\tilde{r} \approx c^{-1} W^T d$$

Some sort of "backprojection"

Matching filter (Good to suppress noise)

Example





r W d_0 d $\tilde{r}$

Right phase but pulse has not been compressed

Seismic Deconvolution

Quadratic Regularization (add a quadratic penalty term)

$$d = W\, r + n$$

Quadratic penalty term

$$J = \| Wr - d \|_2^2 + \mu \| Lr \|_2^2$$

Misfit

$$\nabla J = W^T W\, r - W^T d + \mu L^T L\, r = 0$$

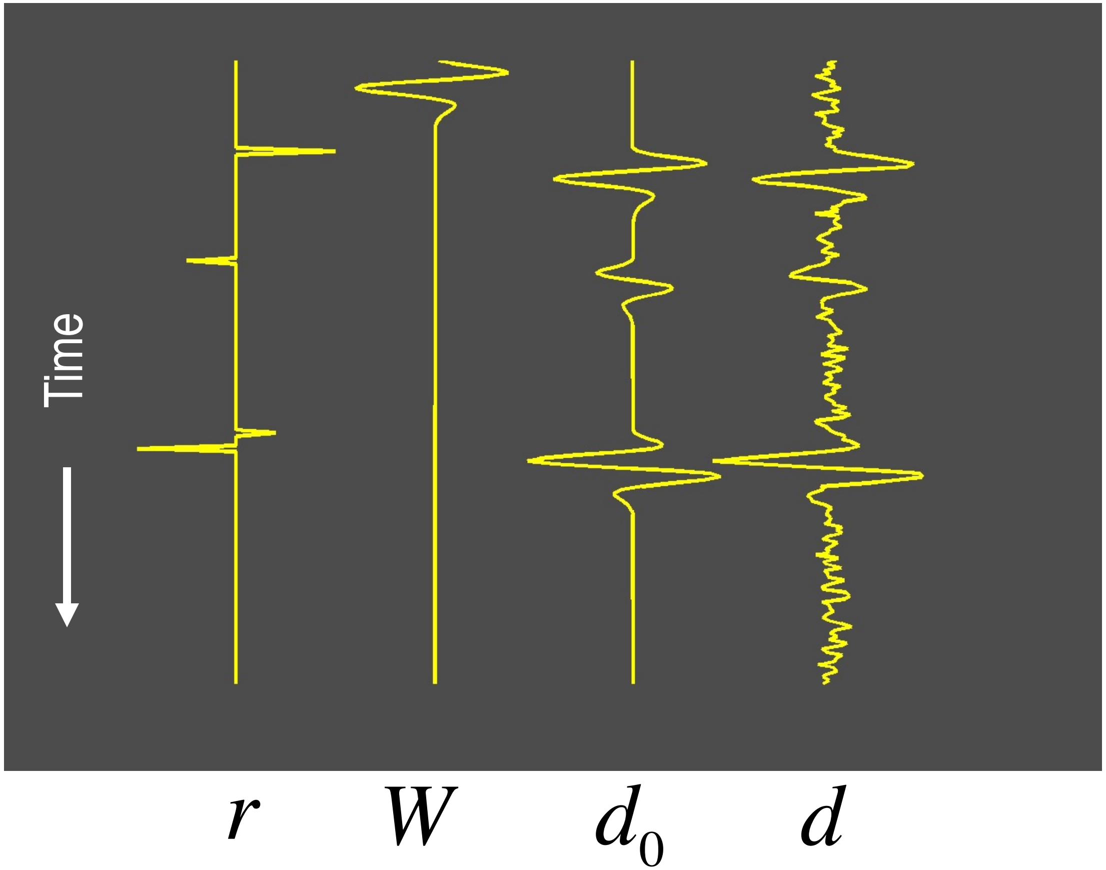
$$r = (W^T W + \mu L^T L)^{-1} W^T d$$

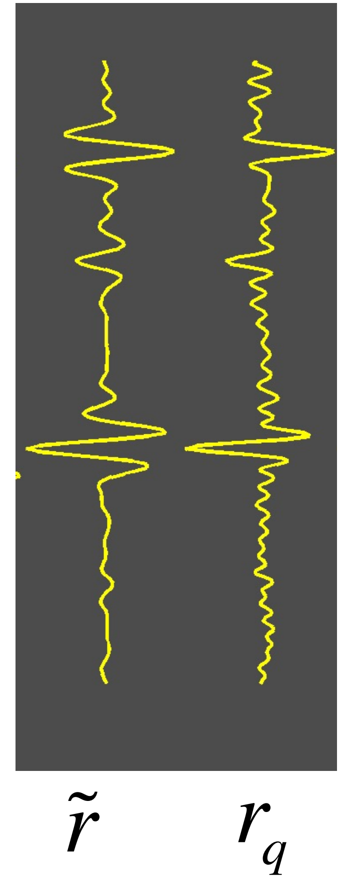
$$L = I \Rightarrow$$

Final solution: estimator of the reflectivity

$$r = (W'W + \mu I)^{-1} W' d$$

Example





<u>Right phase, pulse has been compressed</u>, ringing due to minimum norm requirement

Quadratic Regularization

Quadratic Regularization (general case)

Interpretation

Model Norm: Measure of "bad" features

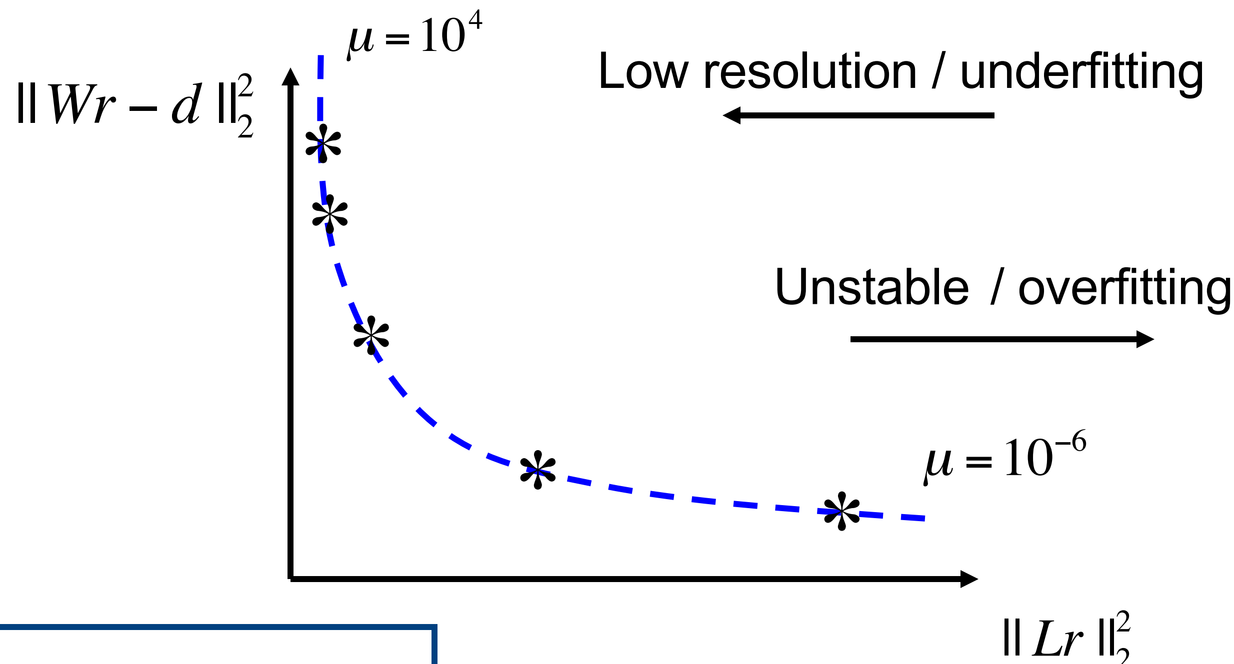
Minimize $$J = \| Wr - d \|_2^2 + \mu \| Lr \|_2^2$$

Misfit: Measure of Data Fidelity

Trade-off parameter

Quadratic Regularization

$$J = \| Wr - d \|_2^2 + \mu \| Lr \|_2^2$$



Trade-off curve / L-curve

Morozov's Discrepancy Principle for Tikhonov regularization

Quadratic Regularization

Quadratic Regularization

$$J = \| Wr - d \|_2^2 + \mu \| Lr \|_2^2$$

Typical regularizers

$L = I$ zero order quadratic regularization (smallest model)

$L = D_1$ First order quadratic regularization (flattest model)

$L = D_2$ Second order quadratic regularization (smoothest model)

D_1, D_2 Are high pass operators (First and Second order Derivatives)

Quadratic Regularization

Quadratic Regularization

$$L = I = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$L = D_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

$$L = D_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 \end{pmatrix}$$

Seismic Deconvolution:
Trying to increase *resolution/focusing* via mathematical tricks….

When in Doubt, Smooth

Sir Harold Jeffreys (Quoted by Moritz, 1980; Taken from Tarantola, 1987)

- **Or… define a regularization term that can switch smoothing on/off in order to focus/resolve signals/models**

- **Let's analyze high-resolution estimates obtained via <u>non-quadratic</u> regularization**

High-resolution via non-quadratic regularization
(GOAL: Suppress ringing & Increase BW)

$$J = \| Wr - d \|_2^2 + \mu R(r)$$

$R(r) = \sum_i | r_i |$ *L1 norm* → Non Quadratic Norm

$R(r) = \sum_i | r_i |^2$ *L2 norm* → Quadratic Norm

$R(r) = \sum_i ln(1 + \frac{r_i^2}{\sigma_c^2})$ *Cauchy norm* → Non Quadratic Norm

Let's study the Cauchy norm inversion

High-resolution via non-quadratic regularization

Non-quadratic norms are used to estimate SPARSE models

$$J = \| Wr - d \|_2^2 + \mu R(r)$$

$R(r) = \sum_i |r_i|$ — *L1 norm* → Non Quadratic Norm

$R(r) = \sum_i |r_i|^2$ — *L2 norm* → Quadratic Norm

$R(r) = \sum_i ln(1 + \frac{r_i^2}{\sigma_c^2})$ — *Cauchy norm* → Non Quadratic Norm

Let's study the Cauchy norm inversion

Seismic Deconvolution: Non-quadratic norms and sparse deconvolution

- Non-quadratic norms can be used to retrieve sparse models - solution with high frequency content
- Connected to the retrieval of non-Gaussian models/signals
- Sparse models are associated to broad-band solutions

  D. W. Oldenburg et al, 1982, Recovery of the acoustic impedance from reflection seismograms, Geophysics, vol. 48, No. 10, pp. 1318-1337

- New interest on these methods because of the importance for reconstruction & sparse coding algorithms:
  - Data reconstruction and spectral analysis from short records

    Sacchi, M.D., Ulrych, T.J, and Walker, C., 1998, Interpolation and extrapolation using a high resolution discrete Fourier transform: IEEE Trans. on Signal Processing, 46, No. 1, 31-38.

  - Modeling simple-cell receptive primary visual cortex in mammals

    Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381, 607-609

  - Compressed sensing and data recovery

    Baraniuk, Compressive Sensing, IEEE Signal Processing Magazine, July 2007)

Cauchy Norm Deconvolution

High-resolution via non-quadratic regularization

$$J = \| W\, r - d \|_2^2 + \mu S(r)$$

$$S(r) = \sum_i ln(1 + \frac{r_i^2}{\sigma^2})$$

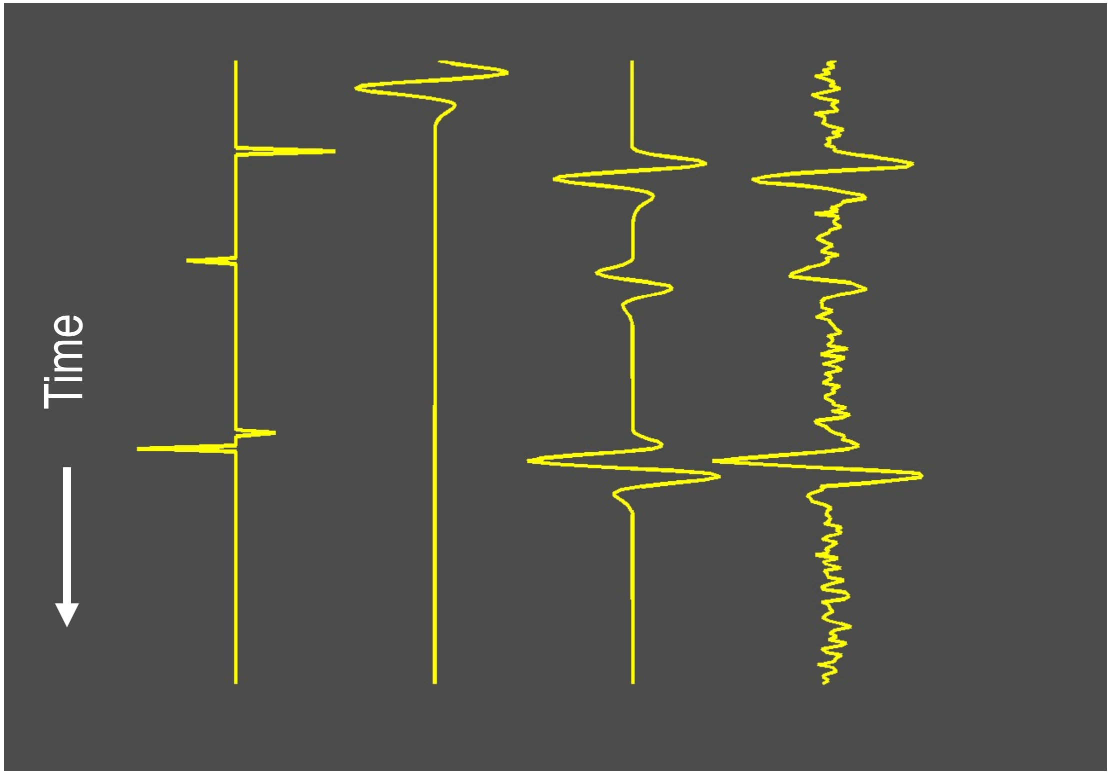
$$\nabla J = W' W\, r - W' d + \mu Q(r)\, r = 0$$

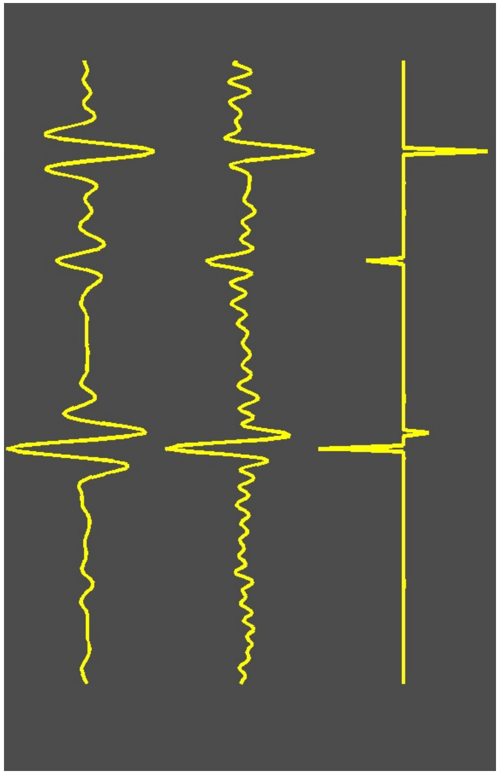
$$\boxed{r = (W' W + \mu Q(r))^{-1} W' d}$$

Non-Linear system of equations - can be solved with IRLS

Iterative Reweighted least squares

Example - High Res. Result with Cauchy regularization





r W d_0 d $\tilde{r}$ r_q r_{nq}

$\tilde{r}$: Matched filtering (cross-correlation of wavelet with trace)

r_q : Quadratic regularization (l2 norm on reflectivity)

r_{nq} : Non-quadratic regularization (Cauchy norm on the reflectivity)

Cauchy Norm Deconvolution

High-resolution via non-quadratic regularization

Assignment:

Derive $\boldsymbol{Q}$ for the Cauchy norm, ell-1 norm, and ell1-ell2 norm and the ell2 norm

Cauchy Norm Deconvolution

Super-resolution via non-quadratic regularization

$$J = \| W\, r - d \|_2^2 + \mu S(r)$$

$$S(r) = \sum_i ln(1 + \frac{r_i^2}{\sigma^2})$$

$$\nabla J = W^T W\, r - W^T d + \mu Q(r)\, r = 0$$

$$\boxed{r = (W^T W + \mu Q(r))^{-1} W^T d}$$

Non-Linear system of equations can be solved with IRLS (Iterative Reweighted Least Squares)

Cauchy Norm Deconvolution

$$r = (W^T W + \mu Q(r))^{-1} W^T d$$

$W'W$: Autocorrelation matrix of the wavelet (Toeplitz)

$$\mu Q(r) = \begin{pmatrix} \mu Q(r_1) & & & & \\ & \mu Q(r_1) & & & \\ & & \ddots & & \\ & & & \mu Q(r_{M-1}) & \\ & & & & \mu Q(r_M) \end{pmatrix}$$

$$r = (W^T W + \mu Q(r))^{-1} W^T d$$

$$Q_{ii} = Q(r_i) = \frac{1}{\sigma^2 (1 + r_i^2 / \sigma_c{}^2)}, \text{ we now have an adaptive damping}$$

$r_i \rightarrow$ small, $\quad Q(r_i) \rightarrow 1/\sigma^2 \rightarrow$ large if σ is small

$r_i \rightarrow$ large, $\quad Q(r_i) \rightarrow$ small $\rightarrow$ if $\quad |r_i| >> \sigma$

Cauchy Norm Deconvolution

The problem is non-linear, the reflectivity depends on the reflectivity…

$$r = (W^T W + \mu Q(r))^{-1} W^T d$$

$$Q_{ii} = \frac{1}{\sigma^2 (1 + r_i^2 / \sigma^2)}$$

IRLS (Iterative Re-weighted Least Squares)

$$iter = 0$$

$$r^{iter} = 0$$

For $iter = 1 : \text{iter\_max}$

$$Q^{iter}{}_{ii} = \frac{1}{\sigma_c{}^2 (1 + (r_i^{iter})^2 / \sigma^2)}$$

$$r^{iter+1} = (W^T W + \mu Q^{iter})^{-1} W^T d$$

End

I used IRLS…

But but . there are 10^{10} methods algorithms to solve the *l1-l2* problem..

IRLS

ISTA

FISTA = Fast Iterative Soft Threshold Algorithm

Bregmann Iterations

Proximity Methods

ADMM

Cauchy Norm Deconvolution

For large inverse problems, IRLS requires the following modification: the inversion stage is done with a semi-iterative solver :

- *Conjugate Gradients (CG)*
- *Gauss-Seidel*
- *Pre-conditioned CG*
- *LSQR*

Iterative solvers can compute an approximation to the solution of the linear system

- HR Radon uses the above trick

IRLS (Iterative Re-weighted Least Squares)

$$iter = 1$$

$$r^{iter} = 0$$

For $iter = 1 : \text{iter\_max}$

$$Q^{iter}{}_{ii} = \frac{1}{\sigma^2 (1 + (r_i^{iter})^2 / \sigma^2)}$$

$$r^{iter+1} = (W'W + \mu Q^{iter})^{-1} W' d$$

End

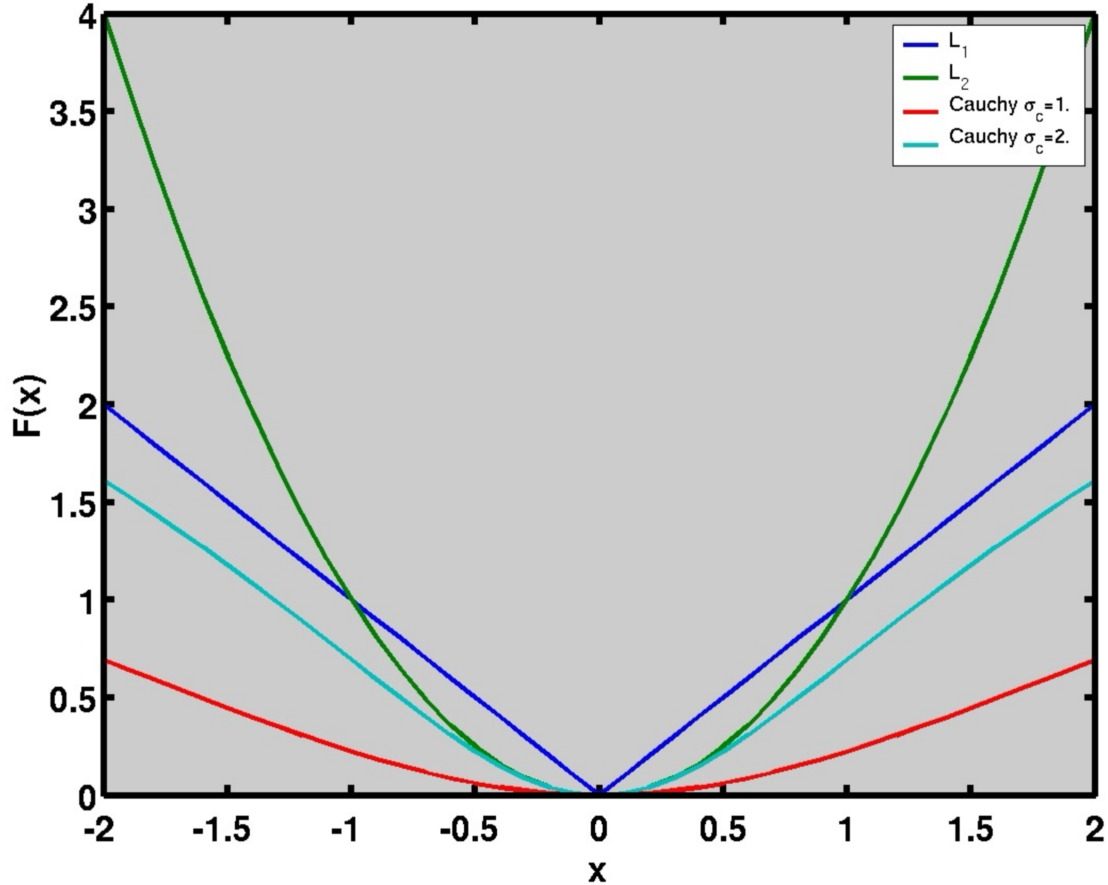
$$R(r) = \sum_i |r_i|$$

$$R(r) = \sum_i |r_i|^2$$

$$R(r) = \sum_i ln(1 + \frac{r_i^2}{\sigma^2})$$

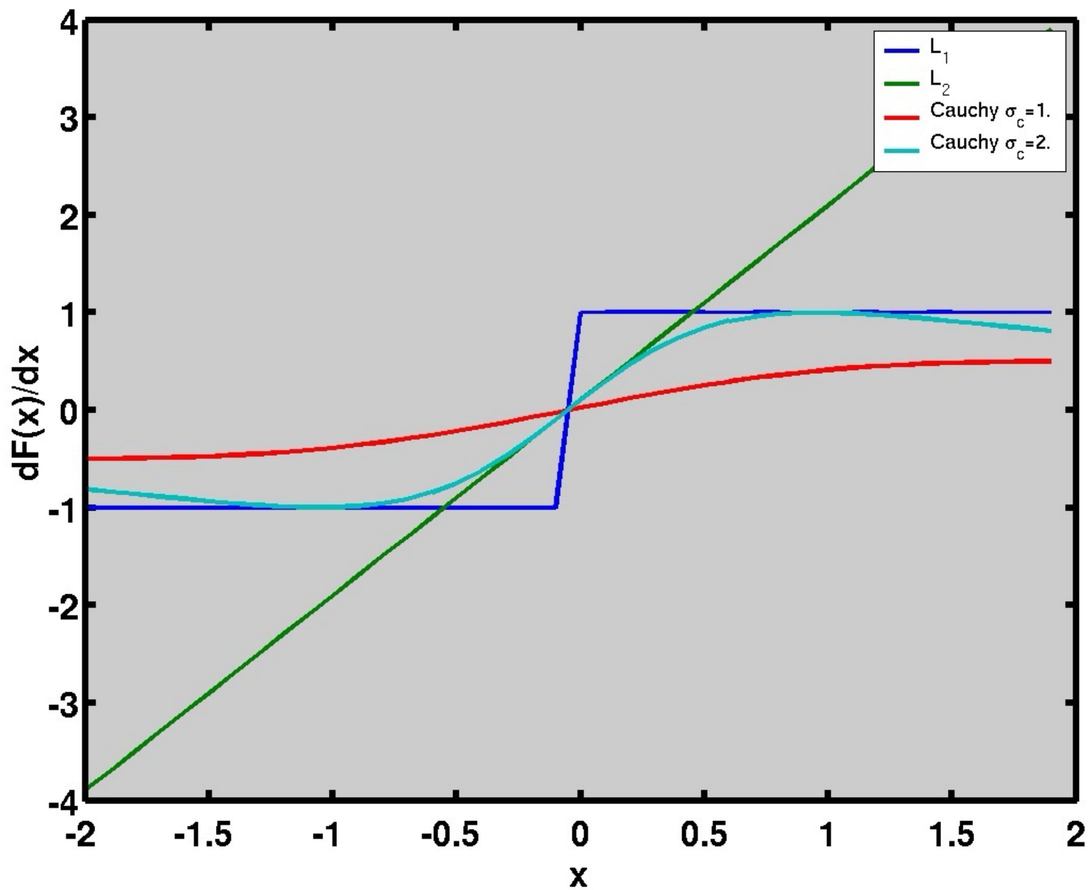
$$\boxed{R(r) = \sum_i F(r_i)}$$

General Form of the norm



$$R(r) = \sum_i F(r_i)$$

$$\frac{dF(x)}{dx}: \text{ Influence Function}$$



- L2 norm: all samples are weighted the same
- Cauchy norm: the weight is model dependent. This is what allows the solution to become sparse

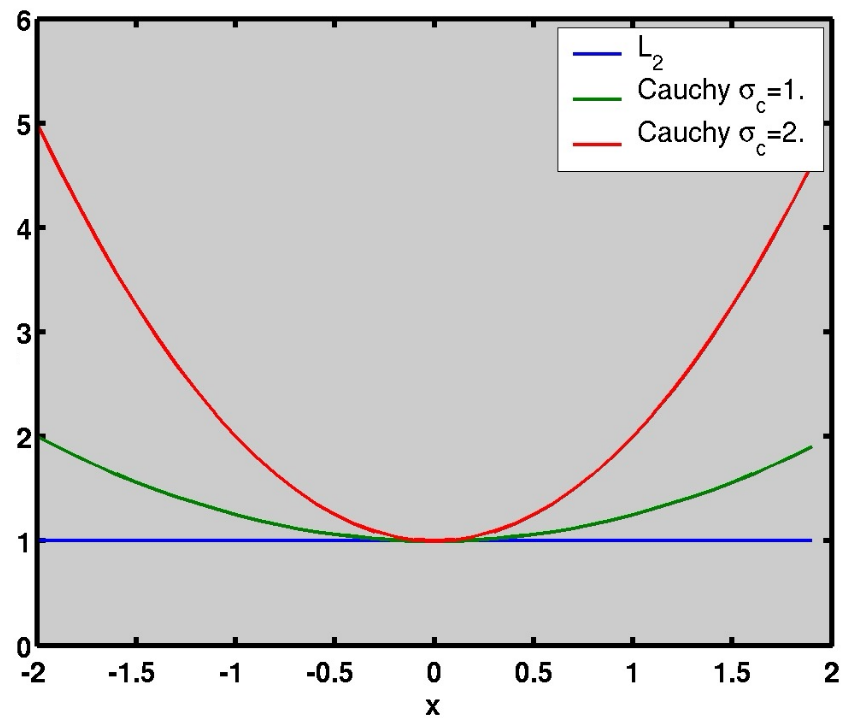
$$r = (W'W + \mu Q(r))^{-1} W' d$$

$$Q_{ii} = \frac{1}{\sigma^2 (1 + r_i^2 / \sigma^2)}$$

μQ_{ii} : Model Dependent
Pre - whitening

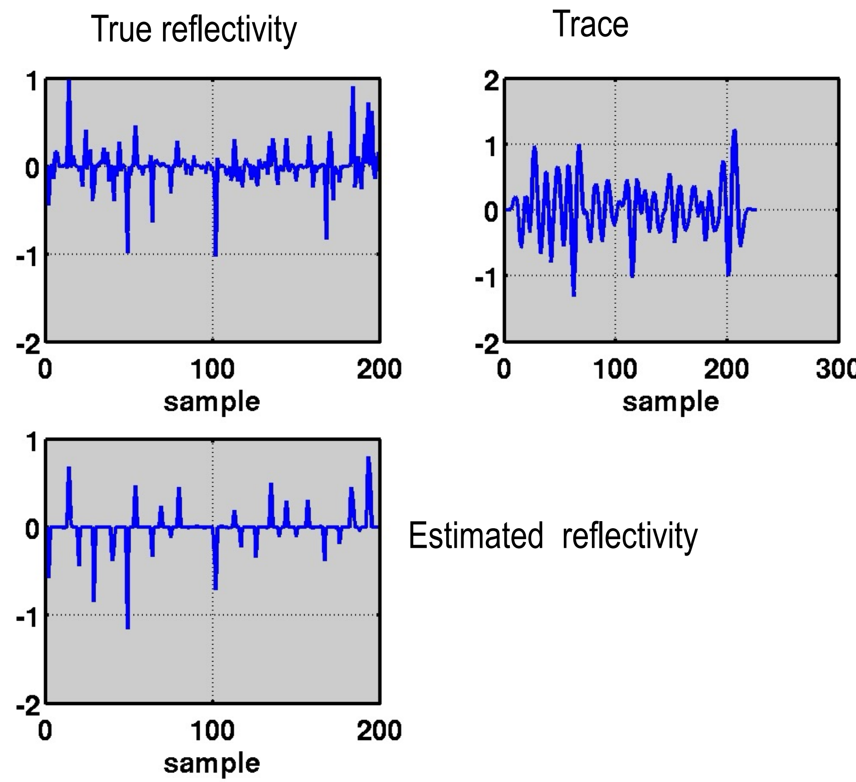
Q_{ii}^{-1} : Can be intepreted as
model - dependent variance

Q_{ii}^{-1}



Cauchy Norm Deconvolution: Algorithm [IRLS]

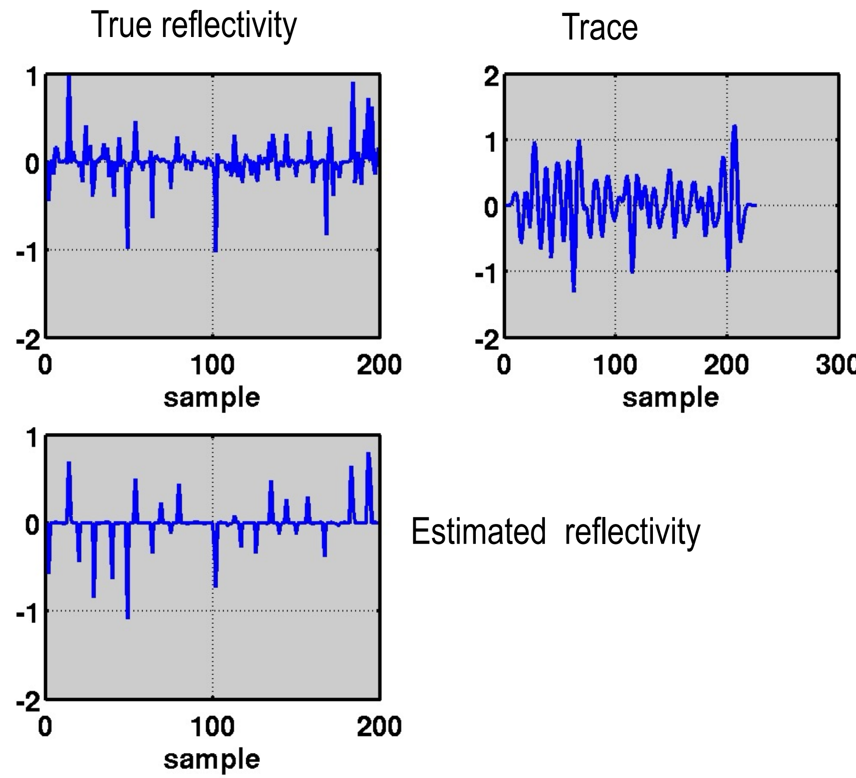
```
r = zeros(N,1);
sc=0.01;
mu =0.01
iter_max = 10;

R = W'*W;
iter_max;

for k=1:iter_max;

 Q = diag(1./(sc^2+r.^2));
 Matrix = R + mu*Q;
 r = inv(Matrix) * W'*s;

end
```



Cauchy Norm Deconvolution: Algorithm [IRLS]

```
r = zeros(N,1);
sc=0.01;
mu = .01
iter_max = 10;
R = W'*W;

for k=1:iter_max;

 Q =diag(1./(sc^2+r.^2));
 Matrix = R + mu*Q;
 r = inv(Matrix) * W'*s;

 sc = 0.01*max(abs(r));

end;
```

Adaptive version



Cauchy Norm Deconvolution:
Algorithm for real data … (HFRestoration or Cauchy Norm Inversion)

Similar to MATLAB prototype but in f77

- Pre-conditioning is used to avoid specifying the trade-off parameter (mu)
- The number of iteration of the external loop is used as trade-off
- Linear inversion is done in the flight with CG (the internal loop)
- Pre-conditioning makes it robust to parameter selection (you do not have to play to much with trade-off parameters!)
- We will briefly discuss the pre-conditioning problem when dealing with de-migration/migration operators

High frequency imaging methods....

You can use any norm that drives the algorithm to sparse solutions (Ell-1, Ell-p with p close to 1, etc)

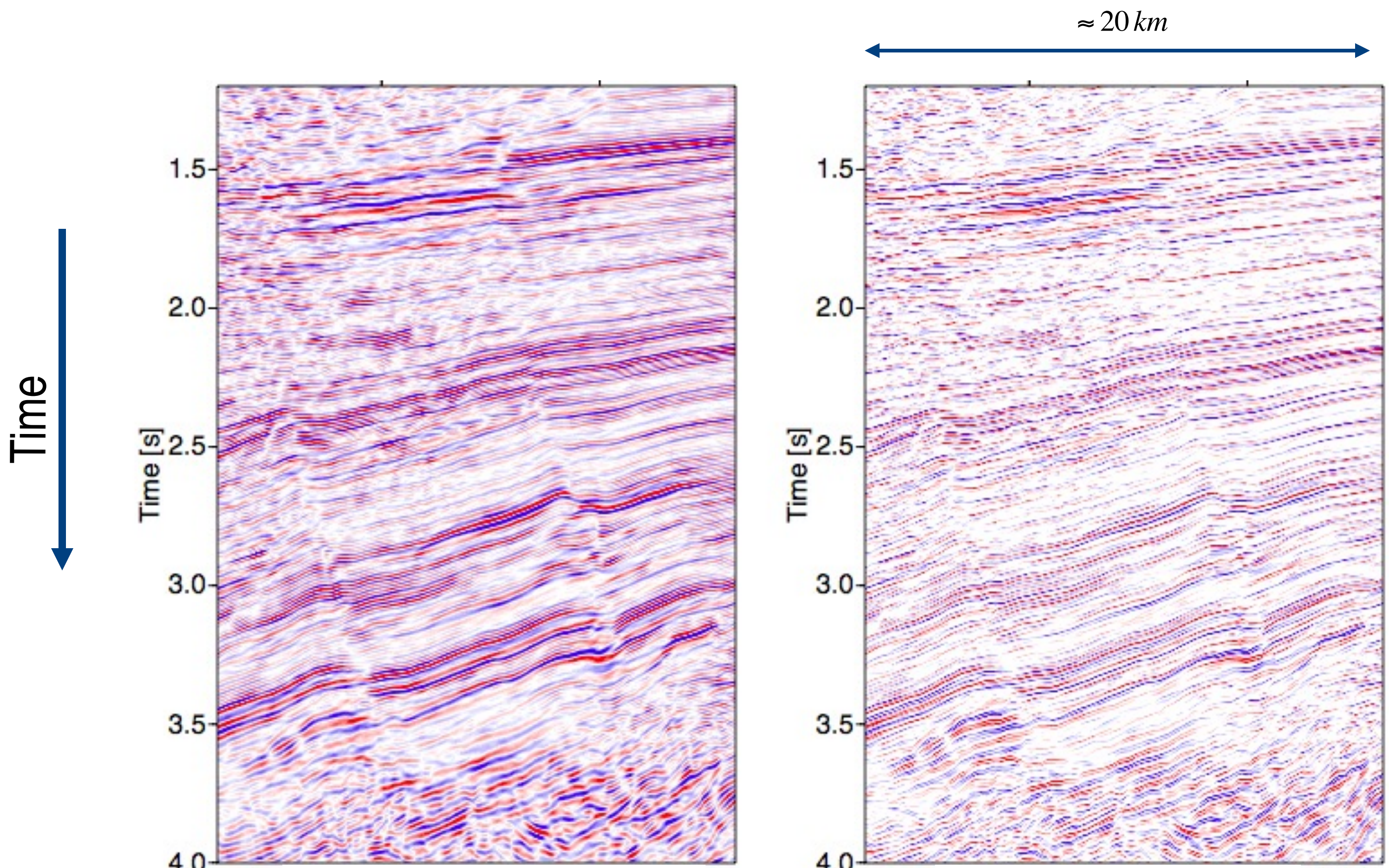
DEBEYE, H. W. J. & RIEL, P. (1990) Lp-NORM DECONVOLUTION. Geophysical Prospecting 38 (4), 381-403.

Other methods exits - they all attempt to retrieve a sparse reflectivity sequence:
Atomic Decompostion/Matching Pursuit / Basic Pursuit (like an L1)

Chopra, S., J. Castagna, and O. Portniaguine, 2006, Seismic Resolution and Thin-Bed Reflectivity Inversion: Recorder, 31, 19-25. (www.cseg.ca)

Portniaguine, O., and J. P. Castagna, 2004, Inverse spectral decomposition: 74th Annual International Meeting, SEG, Expanded Abstracts, 1786-1789.

3D Seismic Volume (Venezuela)



Conventional Deconv

High Res. Deconv

$t = 2z/v$

$t \propto z$

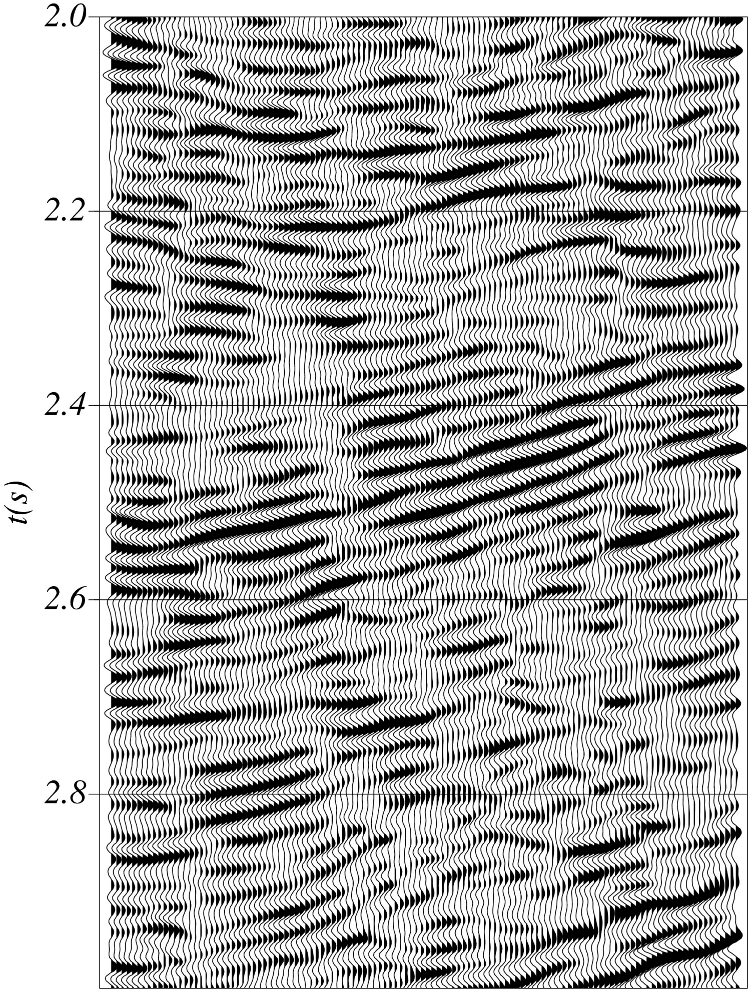
$t = 3s$

$<v> = 2000 m/s$

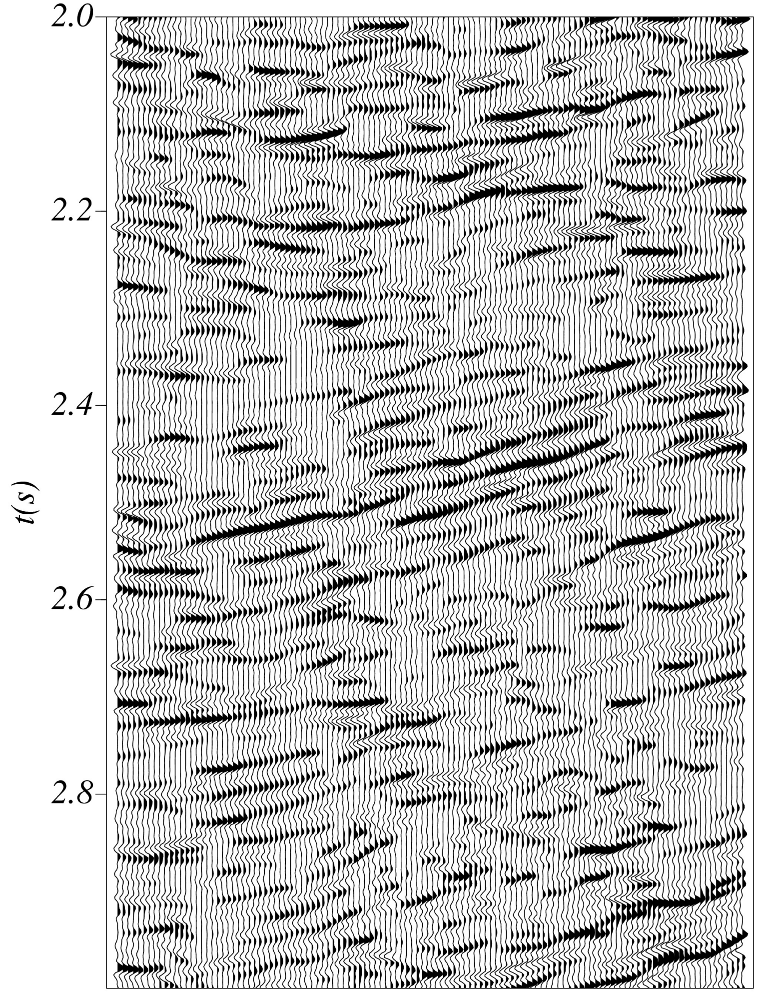
$z = 3 km$

Time: Time that a wave takes to go to and come back from a layer

Red: Positive polarity Blue: negative polarity - The image portrays some sort of reflectivity

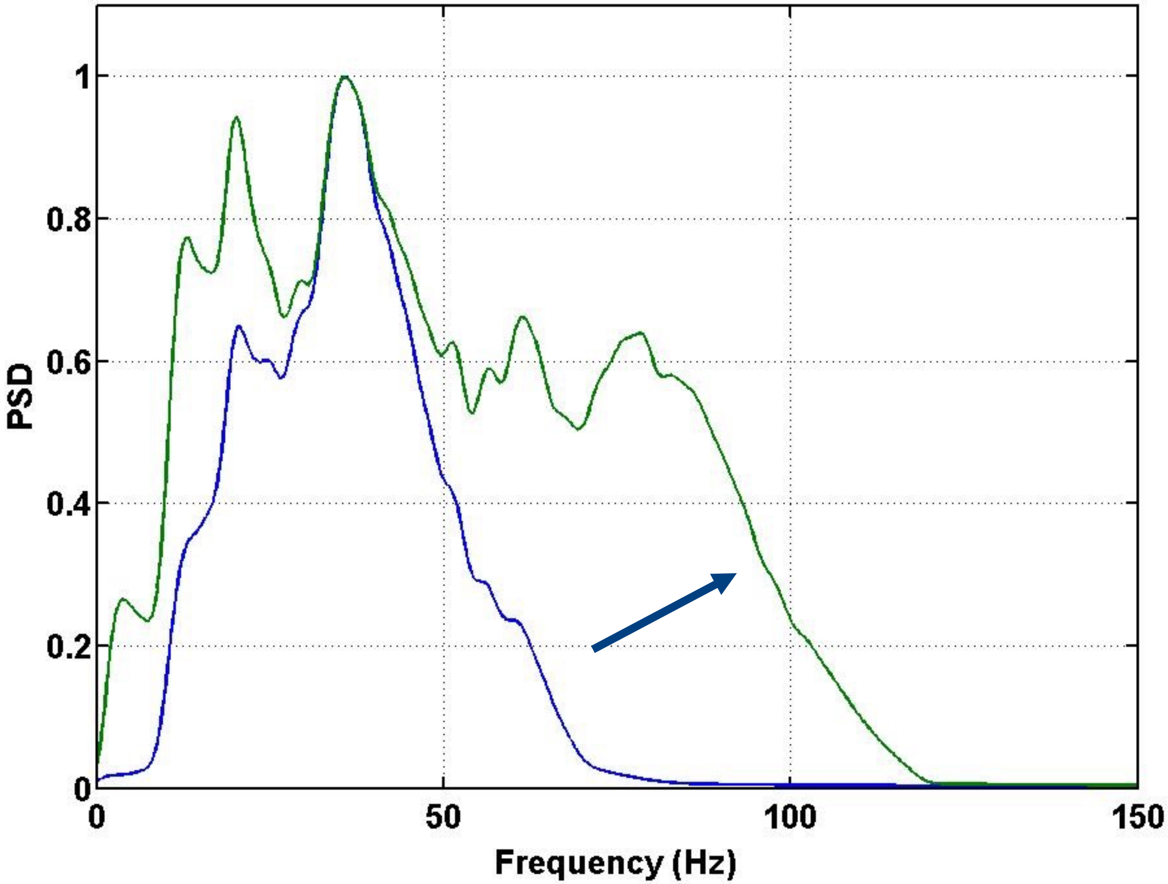


Conventional Deconv



High Res. Deconv

Amplitude spectrum



Can you trust the new bandwidth? I am not saying this is the true answer but a possible answer if you believe in the sparse reflectivity assumption.

Cauchy Norm Deconvolution: with impedance constraints

$$r_k = \frac{I_{k+1} - I_k}{I_{k+1} + I_k}$$

Reflectivity as a function of P-impedance

$$\xi_k = \frac{1}{2}\log(I_k / I_0) \approx \sum_{j=0}^{k} r_j$$

Log Approximation (Logarithm impedance and reflectivity are related by a linear operator)

$$\boxed{J = \| W\, r - d \|_2^2 + \mu S(r) + \beta \| Cr - \xi \|_2^2}$$

Fit the data

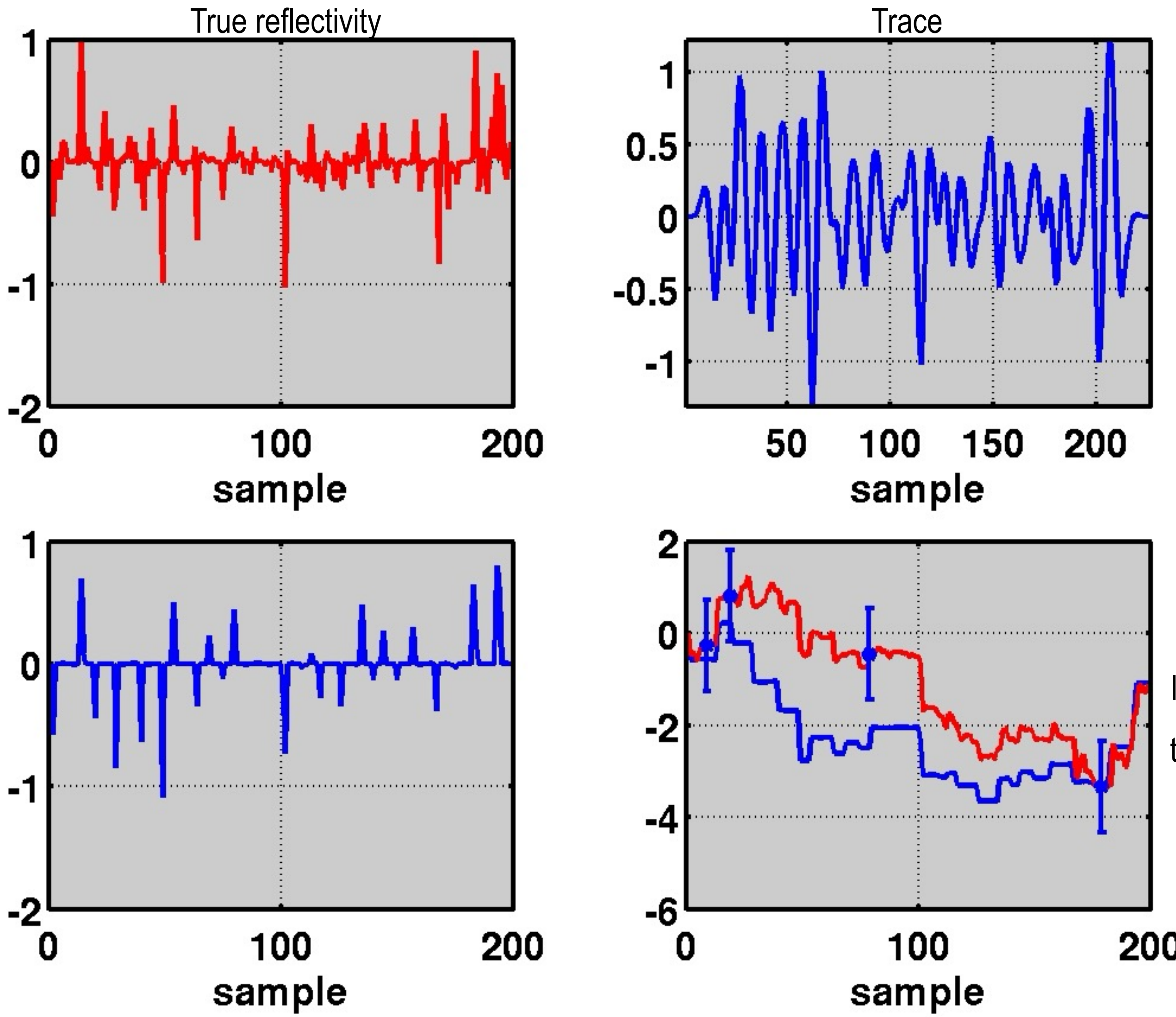
Fit impedance constraints

Solution must be sparse (High freq)

Cauchy Norm Deconvolution: with impedance constraints

$\beta = 0$

Constraints are not honored

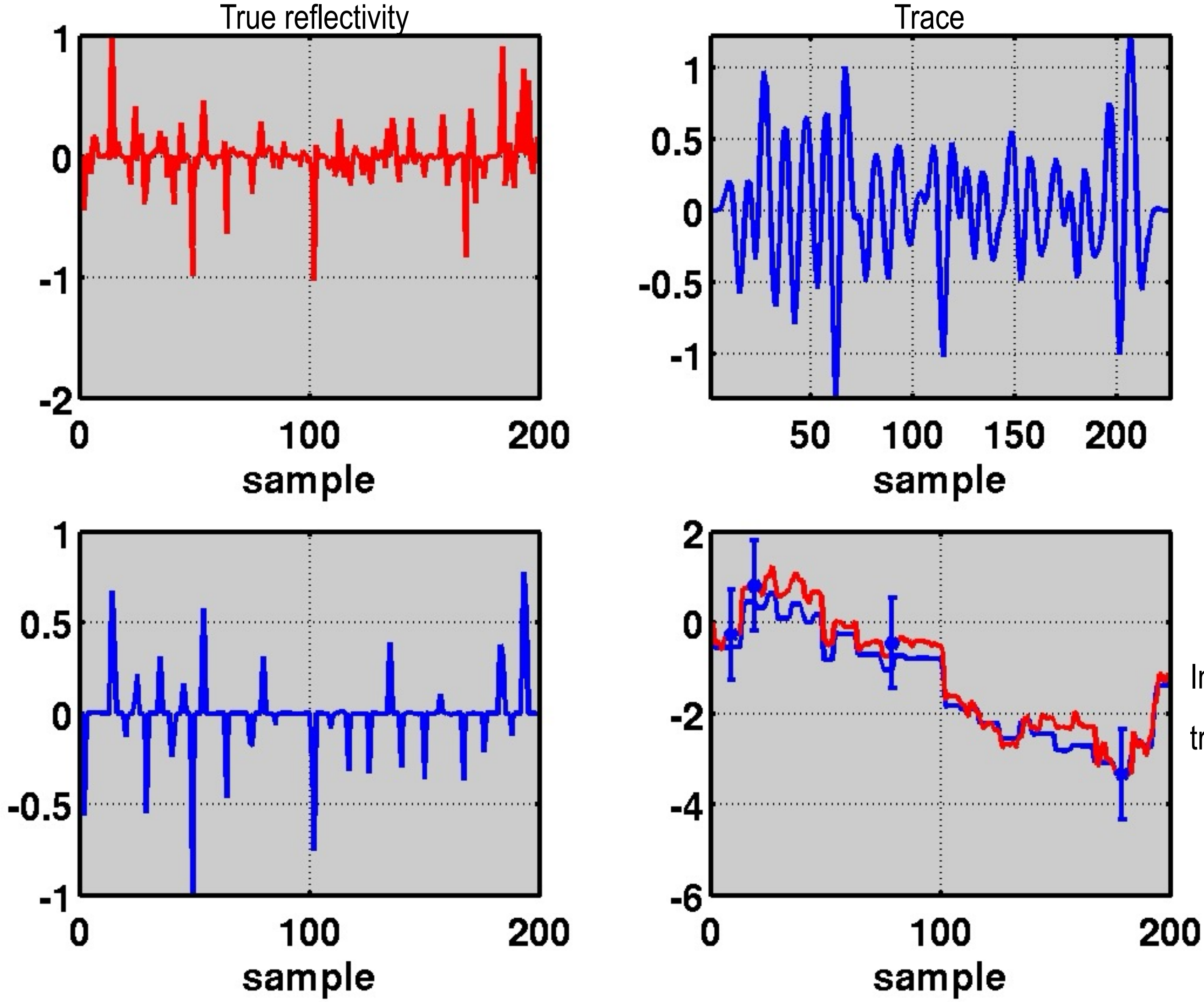


Inverted impedance & true impedance (red)

Cauchy Norm Deconvolution: with impedance constraints

$\beta = 0.25$

Constraints are honored



Cauchy Norm Deconvolution: with impedance constraints - Algorithm [IRLS + soft bounds]

```
beta=0.25
iter_max=10
R = W'*W+beta*C'*C;
r = zeros(N,1);

for k=1:iter_max;
  Q = diag(1./(sc^2+r.^2));
  Matrix = R + mu*Q;
  r = inv(Matrix) * (W'*s+beta*C'*psi)
  sc = 0.01*max(abs(r));
 end;
```

Angle dependent reflectivity

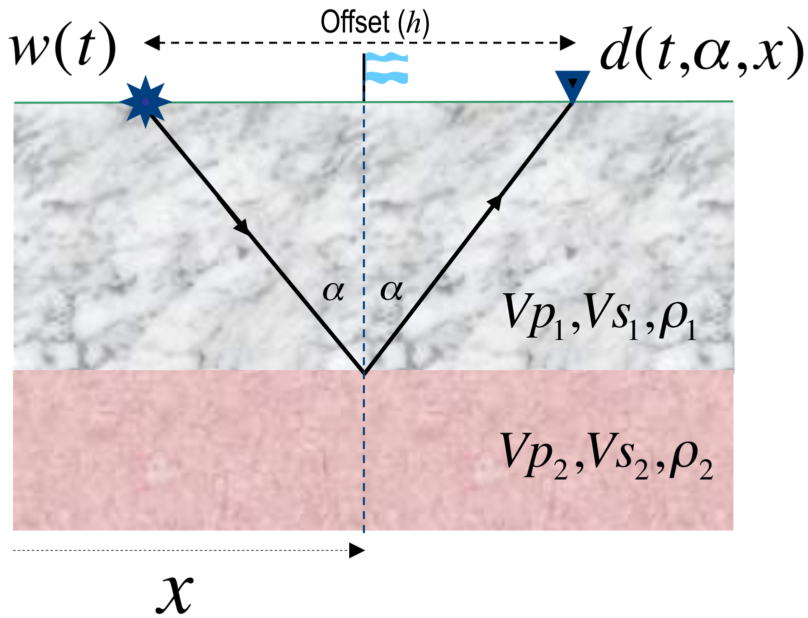
AVO: Amplitude versus offset

AVA: Amplitude versus angle

AVP: Amplitude versus Ray Parameter

Estimation of rock properties & HCI

AVO as a Multi-channel Seismic Deconvolution Problem



$$d(t,\alpha,x) = \int r(\tau,\alpha,x)\, w(\tau - t)\, d\tau$$

$$r(t,\alpha,x) = f(Vp_1, Vs_1, \rho_1, Vp_2, Vs_2, \rho_2) = A(t,x) + B(t,x) \times F(\alpha)$$

Shuey's approximation

AVO as a Multi-channel Seismic Deconvolution Problem

$$d(t,\alpha,x) = \int r(\tau,\alpha,x)\, w(\tau - t) d\tau$$

$$r(t,\alpha,x) = f(Vp_1, Vs_1, \rho_1, Vp_2, Vs_2, \rho_2) = A(t,x) + B(t,x) \times F(\alpha)$$

$$\mathbf{d} \approx \mathbf{Lm}, \qquad \mathbf{m} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}$$

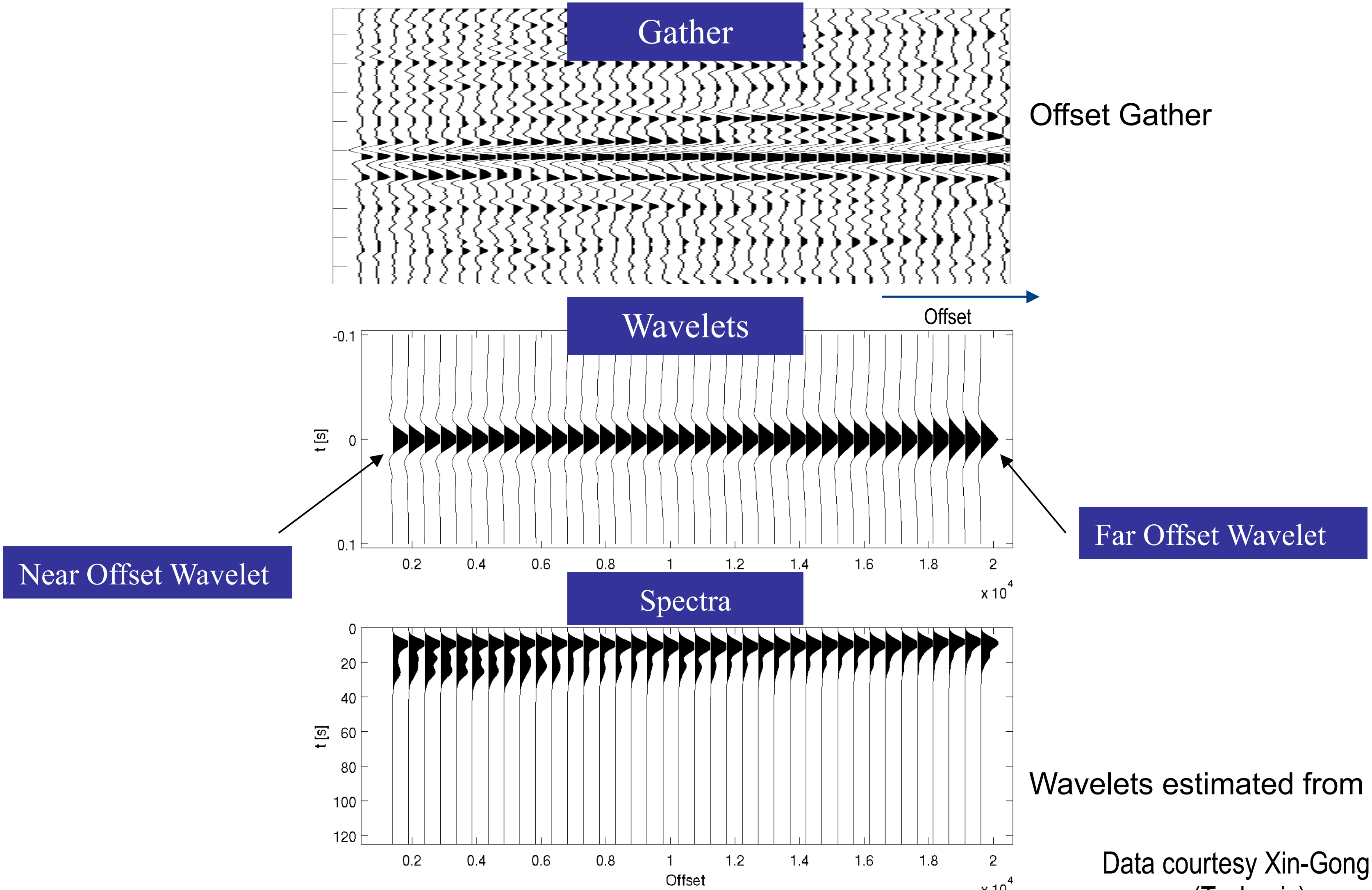
$$J = \| \mathbf{d} - \mathbf{Lm} \|_2^2 + \mu R(\mathbf{m})$$

Real data example - comparison of quadratic vs non-quadratic regularization

- Quadratic - damped least-squares
- Non-quadratic - Cauchy regularization + Internal iteration with Conjugate Gradients

- To stress: The two solutions that I will show fit the data at the same level

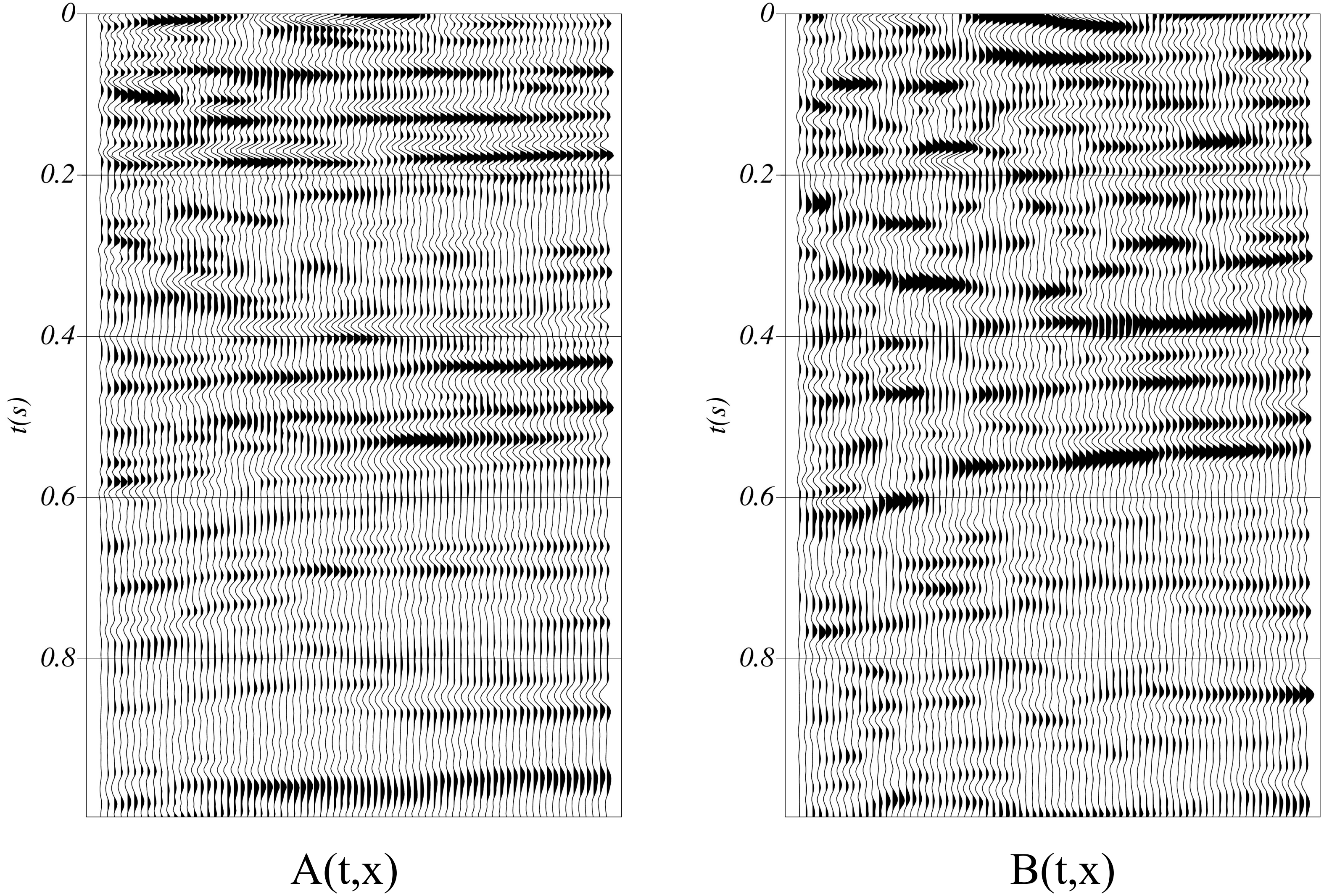
AVO as a Multi-channel Seismic Deconvolution



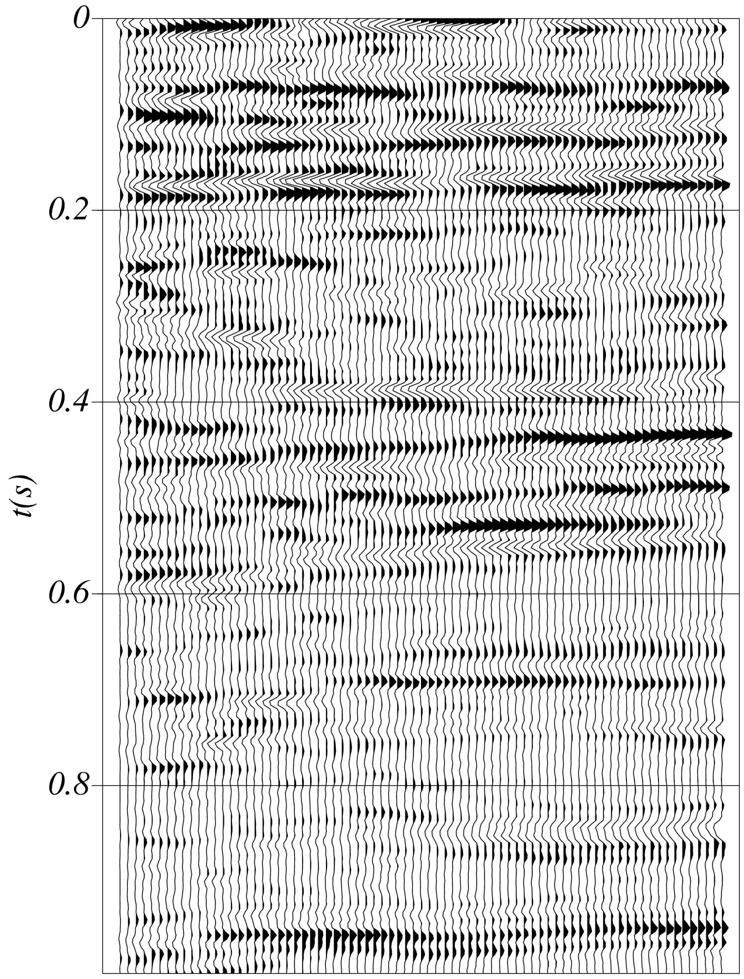
Offset Gather

Wavelets estimated from first gather

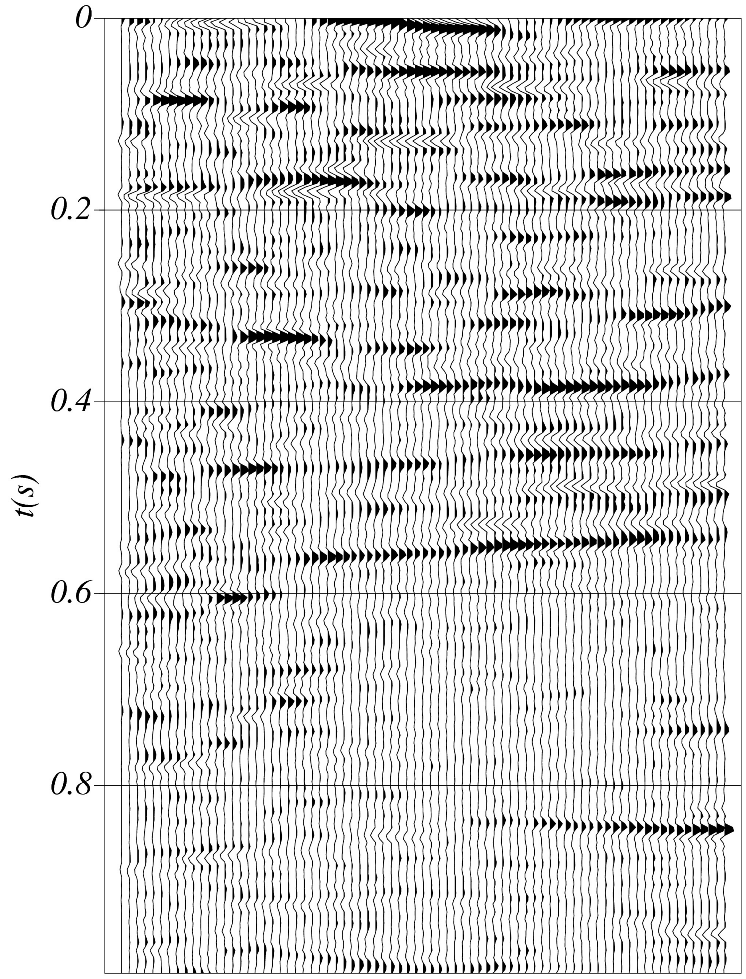
Data courtesy Xin-Gong Li (Techseis)



Quadratic regularization solution



A(t,x)

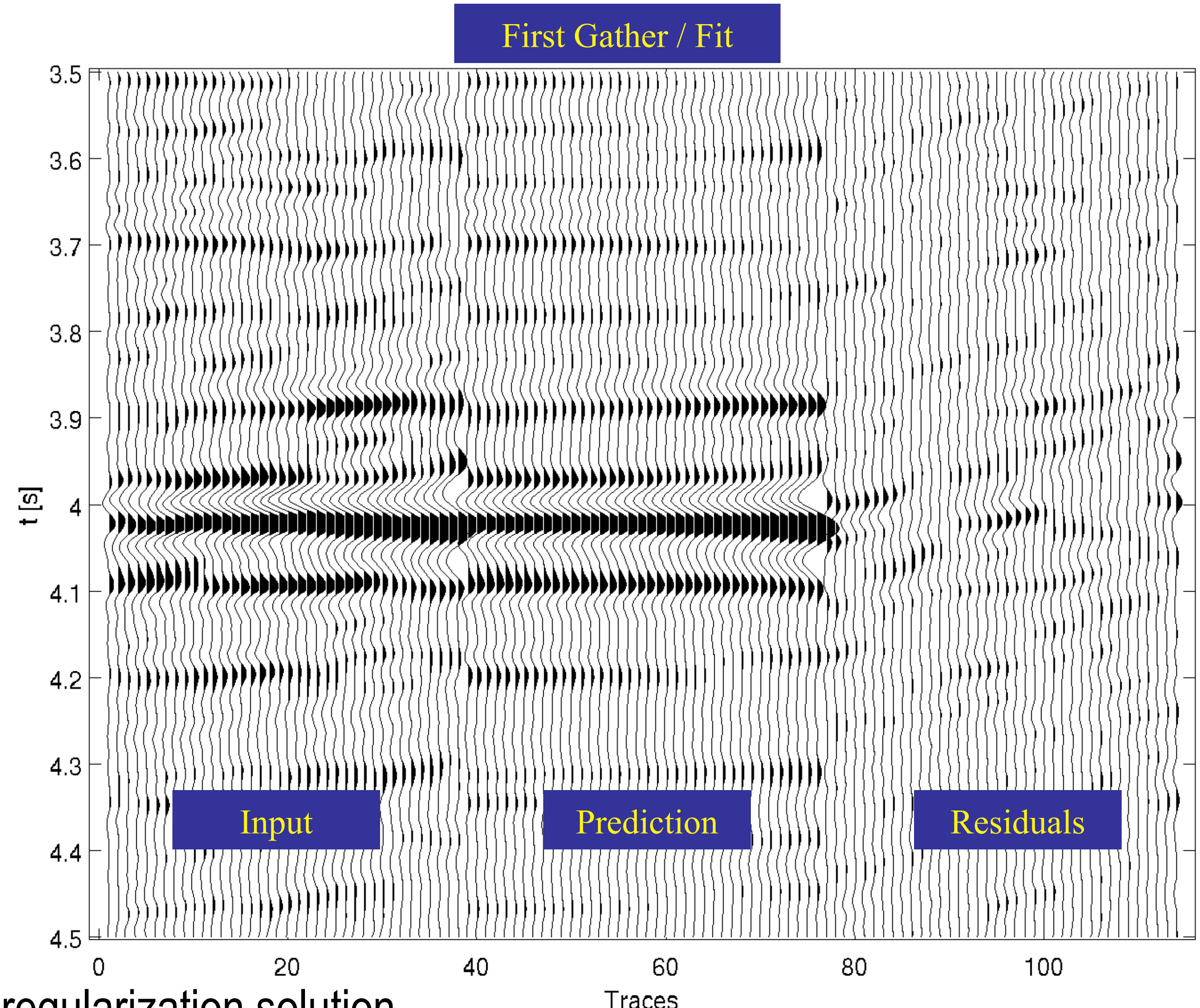


B(t,x)

Nonquadratic regularization solution

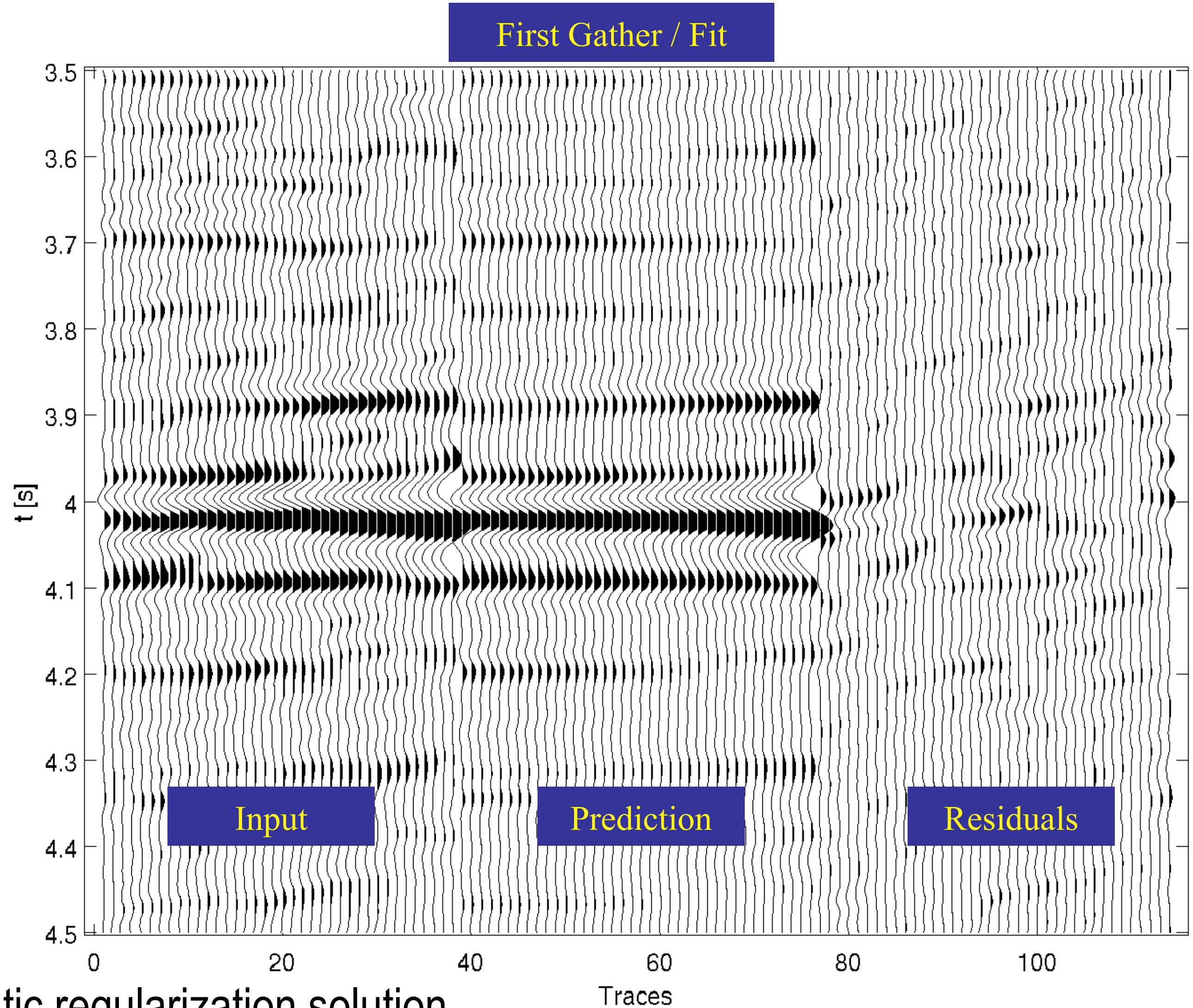
- More details: *Dey, A.K., M.D. Sacchi, and A. Gisolf, 2006, High-resolution reservoir rock properties via joint prestack seismic amplitude inversion: 76th Ann. Internat. Mtg.: SEG, Expanded Abstracts, 2156-2160.*

Data fitting for first gather



Quadratic regularization solution

Data fitting for first gather



Nonquadratic regularization solution

Large Scale Problems

Inverse problems -
Angle dependent reflectivity and
LS migration

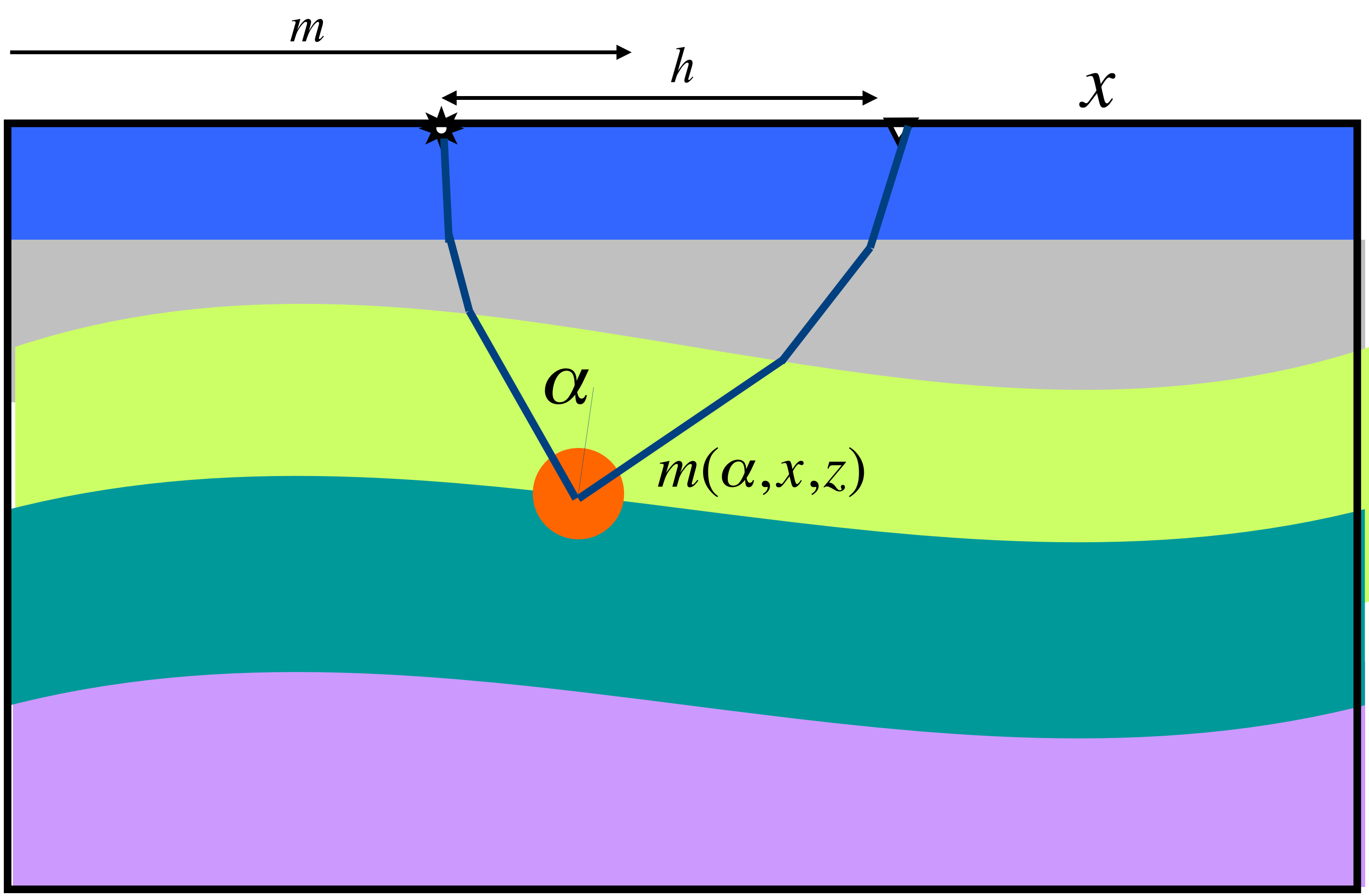
Large inverse problems: Pre-stack depth Imaging with EXTENDED DOMAIN IMAGES

Imaging Angle Dependent Reflectivity

Wang J., Kuehl H. and Sacchi M.D., 2005, High-resolution wave-equation AVA imaging: Algorithm and tests with a data set from the Western Canadian Sedimentary Basin: Geophysics, 70, 891-899

Wang J. and Sacchi M.D., 2007, High-resolution wave equation AVP imaging with sparseness constraints, Geophysics, 72 (1), S11-S18.

Imaging - Large scale problems



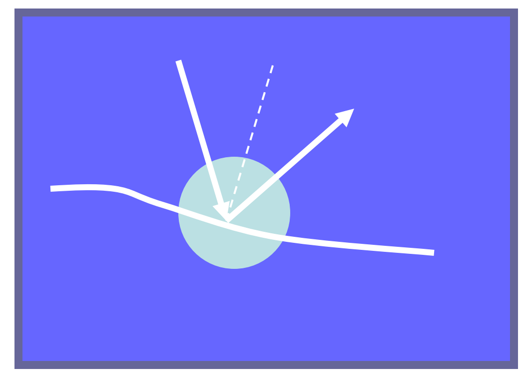
m: midpoint
h: offset

Wang, Kuehl and Sacchi, Geophysics 2005

Imaging in Operator Form..

$$\mathbf{L\,m} \approx \mathbf{d}$$

$$m = m(x,z,p), \qquad p = \frac{\sin(\alpha)}{\bar{v}(x,z)}$$



$$d = d(g,s,t)$$

d: is now pre-processed seismic data

Migration/Inversion Hierarchy..

$$\mathbf{Lm} = \mathbf{d}$$ Modeling (De-migration operator)

$$\mathbf{m}_{\mathbf{mig}} = \mathbf{L}^T\mathbf{d}$$ Migration

$$\mathbf{m}_{inv} = \mathbf{L}^{-1}\mathbf{d}$$ Inversion

Regularized Imaging

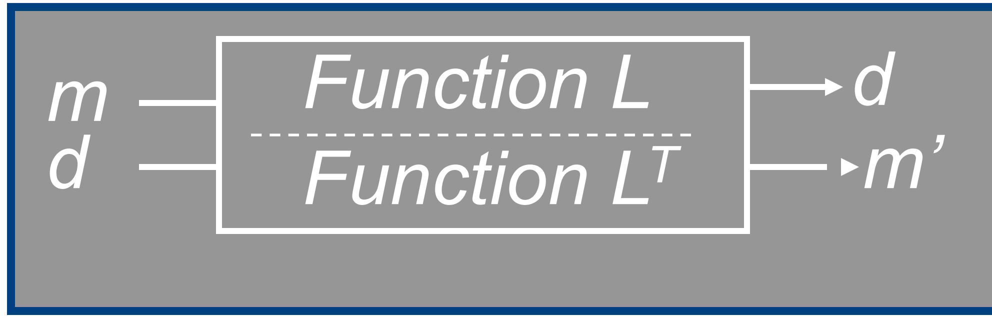
$$J = \| \mathbf{W}(\mathbf{L}\mathbf{m} - \mathbf{d}) \|_2^2 + \mu R(\mathbf{m})$$

Remarks:

- We don’t have $\mathbf{L}$, we only have a code that knows how to apply $\mathbf{L}$ to $\mathbf{m}$ and, another, that knows how to apply $\mathbf{L}^T$ to $\mathbf{d}$
- $\mathbf{L}$ and $\mathbf{L}'$ are built to pass the dot product test
- *Importance of sampling Matrix* $\mathbf{W}$ *for reducing footprints*

Kuehl H. and Sacchi M.D., 2003, Least-squares wave-equation migration for AVP/AVA inversion: Geophysics, 68, 262-273

Regularized Imaging



Dot Product Test

(only for programmers..)

1) $m_1, \; d_1 = L m_1$

2) $d_2, \; m_2 = L^T d_2$

3) *if* $d_1^T d_2 = m_1^T m_2,$

then, L^T is the adjoint of L

Goal: Guarantee ***adjointness*** of *L* and L^T (numerically)

Regularized Imaging with Conjugate Gradients

$$J = \| \mathbf{W}(\mathbf{Lm} - \mathbf{d}) \|_2^2 + \mu F(\mathbf{m})$$

$$m = m(x, z, p)$$

F : Smoothness in x, y, p ,

Sparseness in z

Regularized Imaging with Conjugate Gradients + *Preconditioning*

Reduce computational cost

$$J = \| W(Lm - d) \|_2^2 + \mu \| Rm \|_2^2$$

$$J' = \| W(L\,P\nu - d) \|_2^2 + \mu \| \nu \|_2^2$$

$$m = P\nu, \; P \approx R^{-1}$$

Regularized Imaging with Conjugate Gradients + Preconditioning + Regularization by Iteration

$$J' = \| W(L\,Pv - d) \|_2^2 \cancel{+ \mu \| v \|_2^2}$$

k : CG Iteration

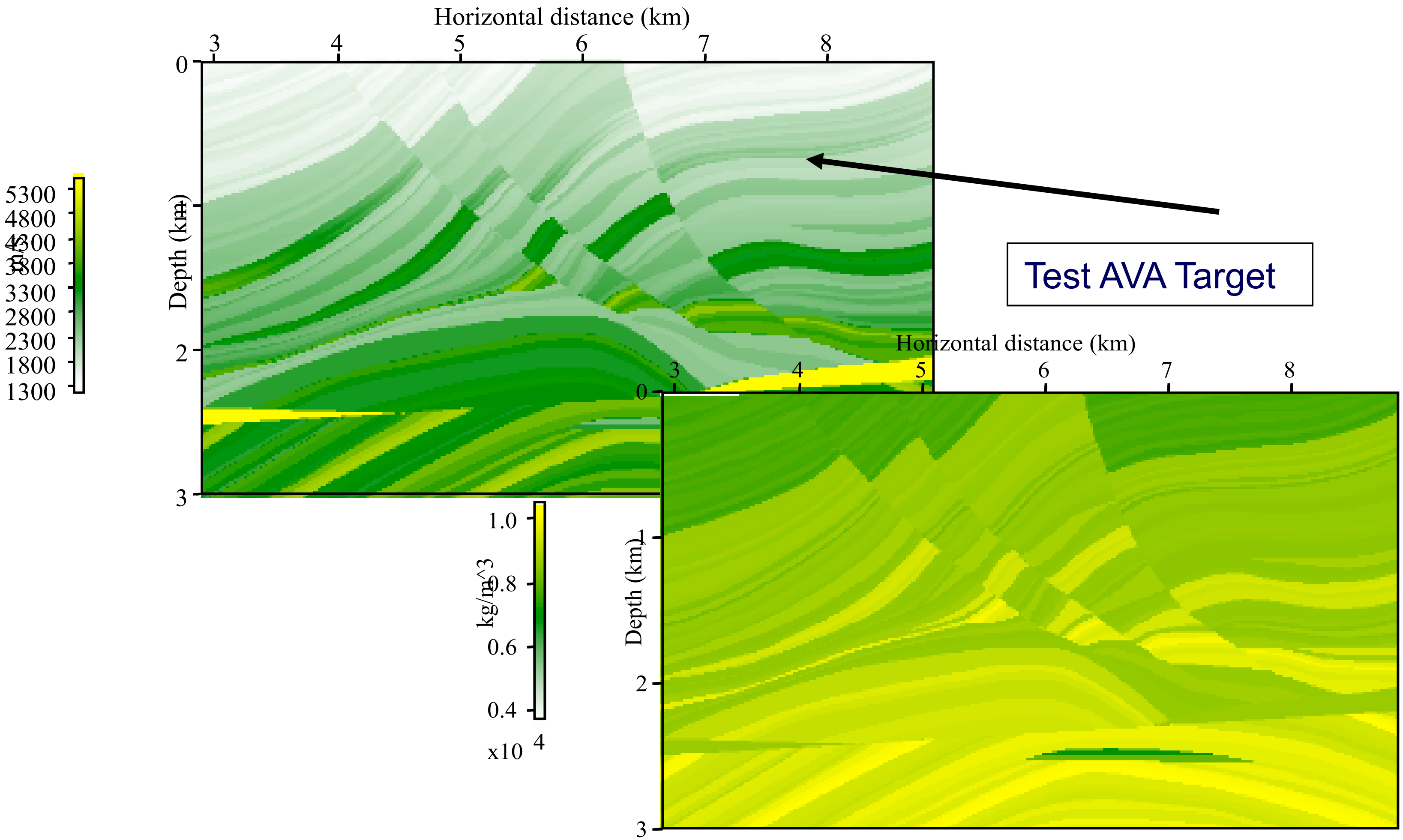
$$J'' = \| W(LPv_k - d) \|_2^2 < tol$$

Hansen, 1987, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion (SIAM Monographs on Mathematical Modeling and Computation)

Examples:

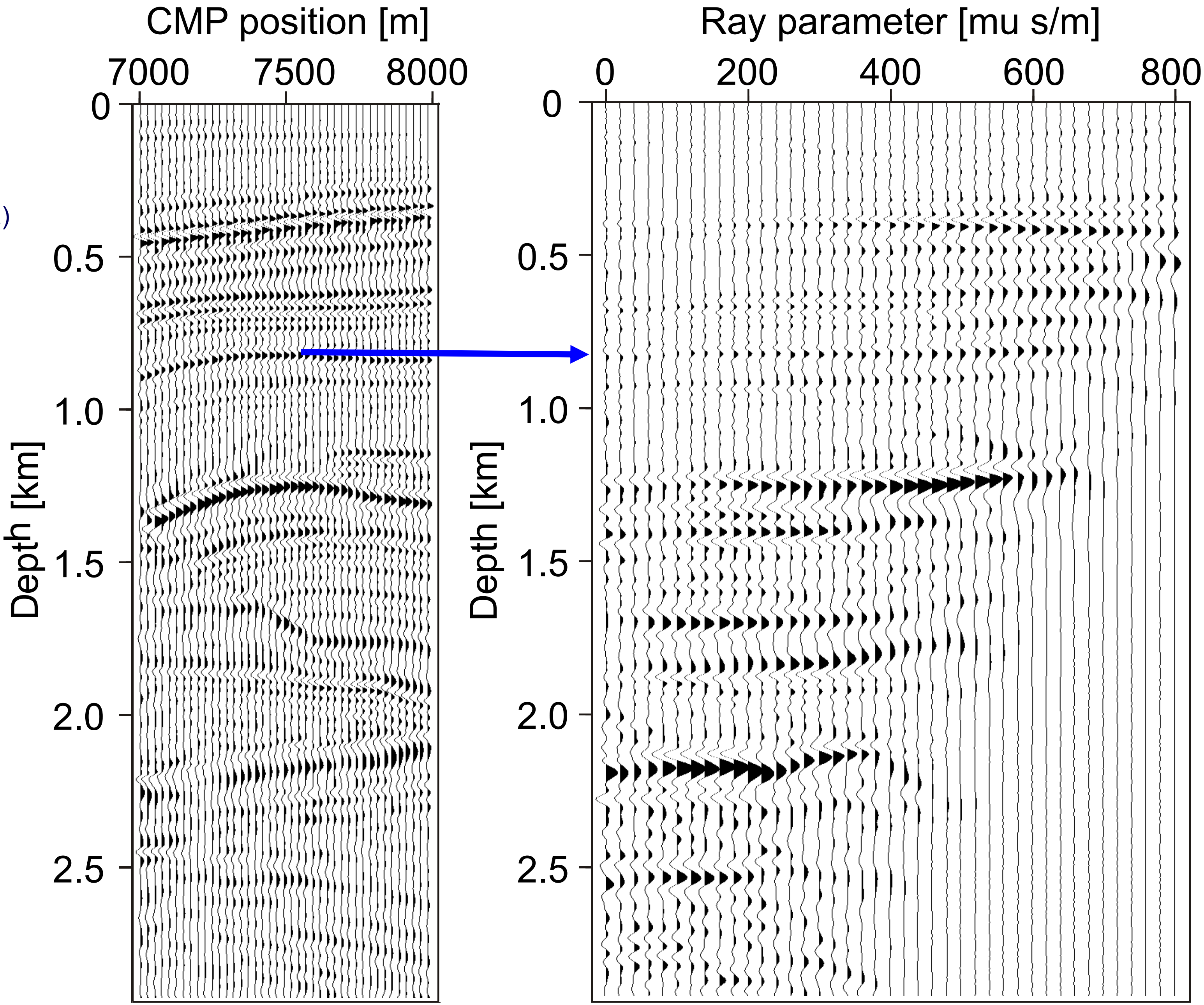
- **Regularized Imaging to attenuate the influence of missing data**
- **Increase resolution of seismic images**

Marmousi model (P-wave velocity and density)



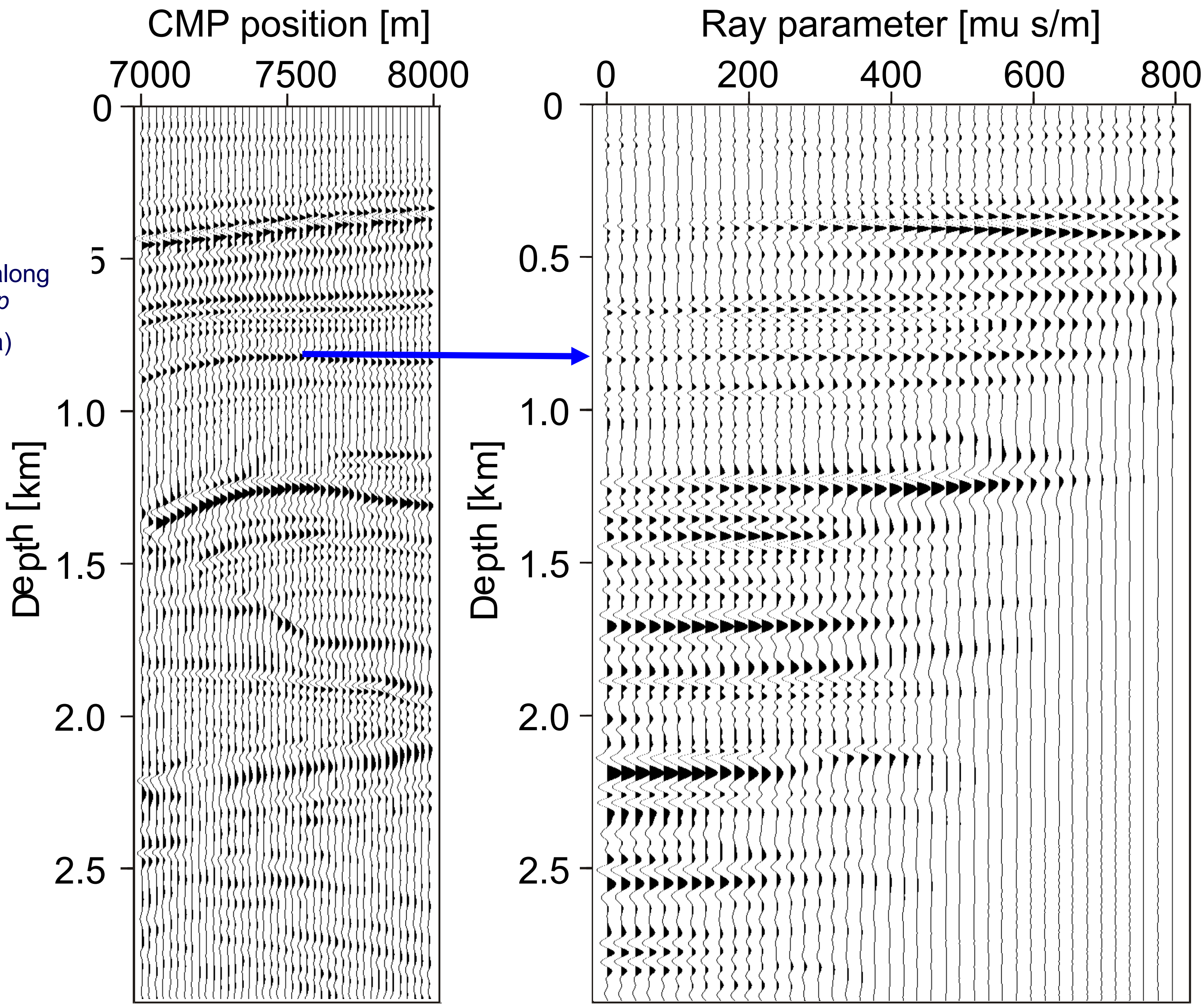
Migration

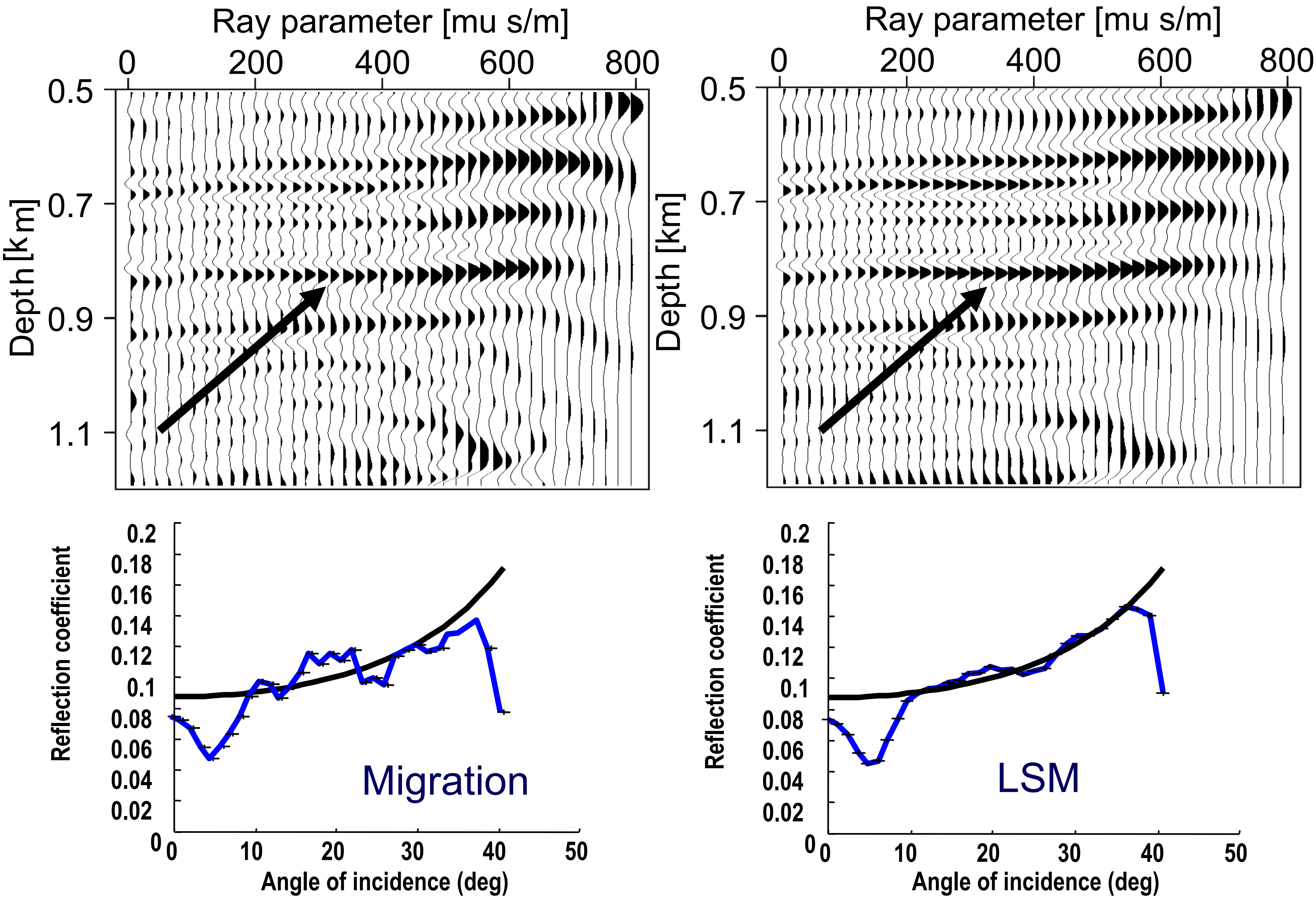
(Complete data)

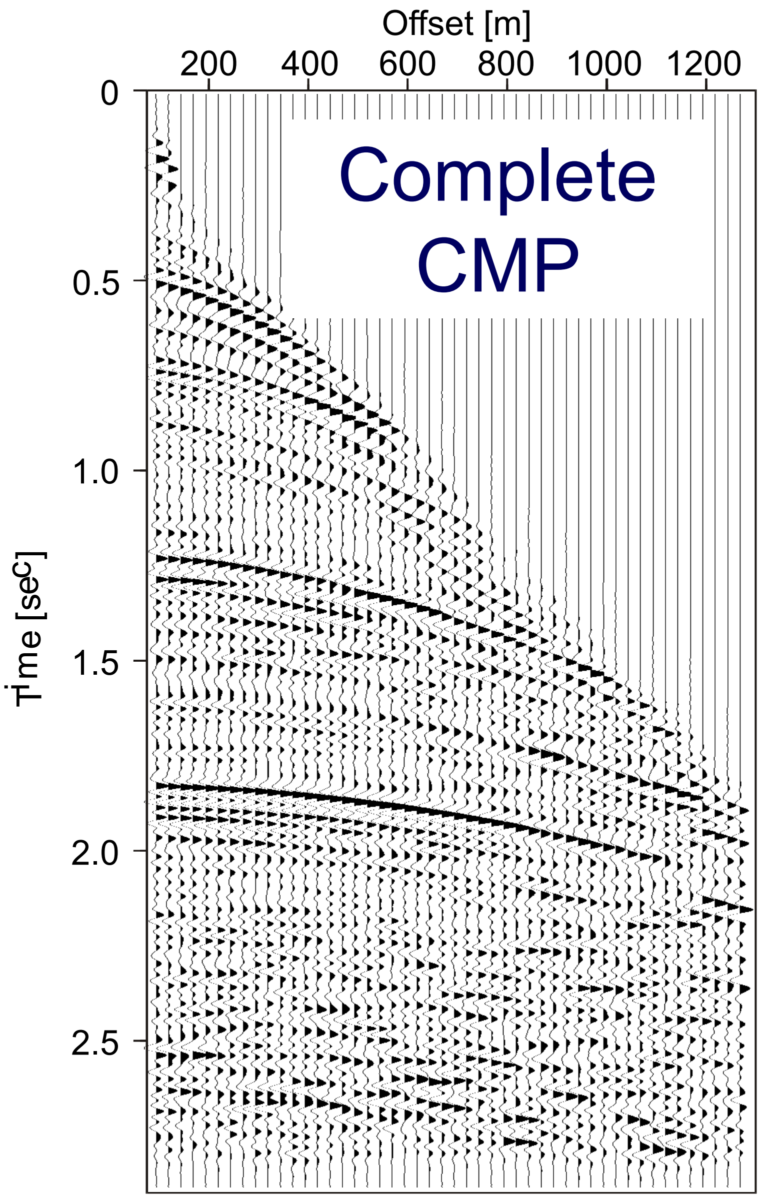


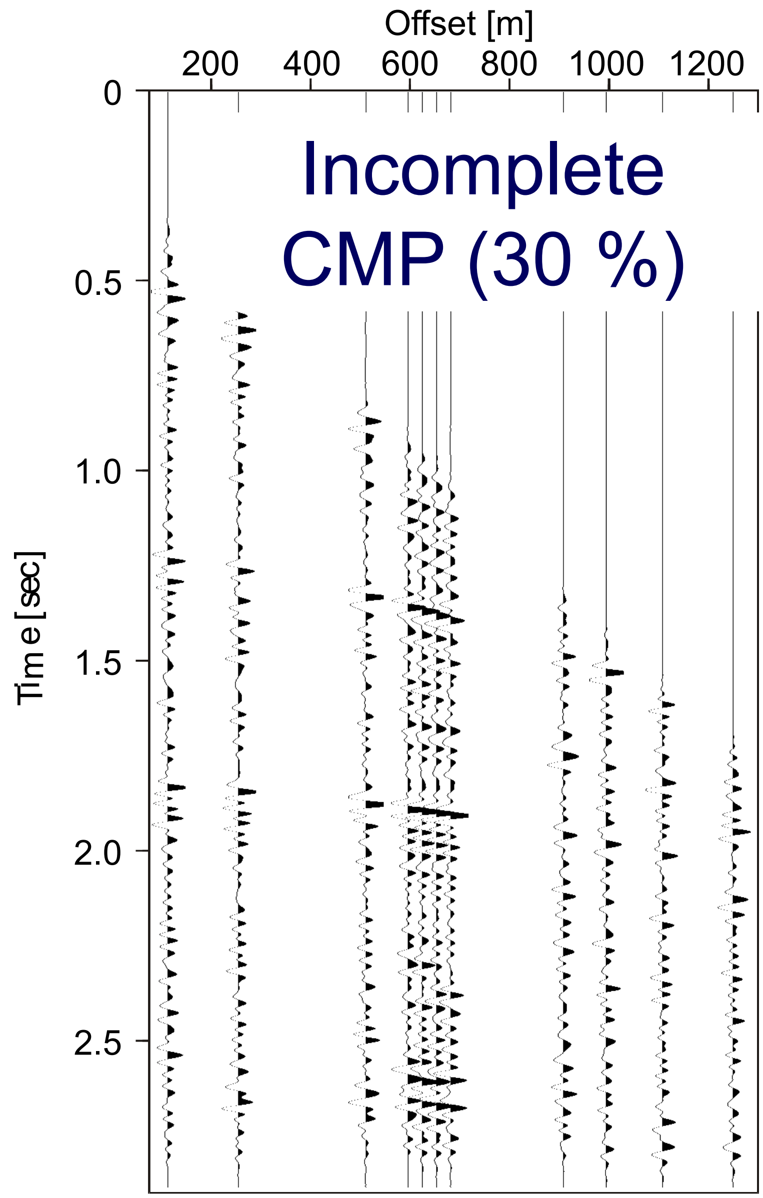
Least Squares Migration with Smoothing regularization along ray parameter p

(Complete data)

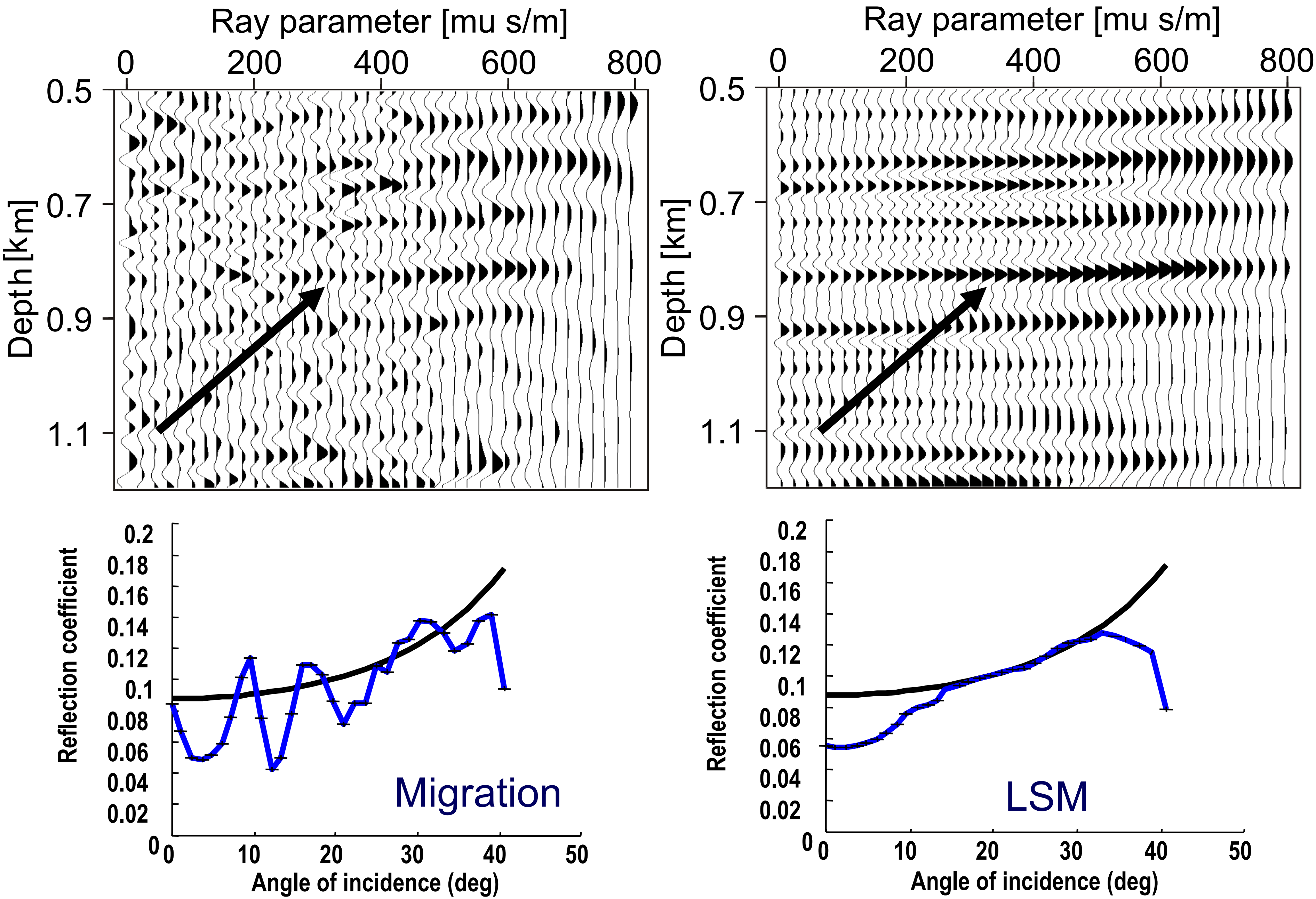



Ray parameter [mu s/m]
0
200
400
600
800
0.5
0.7
0.9
1.1
Depth [km]
Reflection coefficient
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
10
20
30
40
50
Angle of incidence (deg)
Migration
LSM





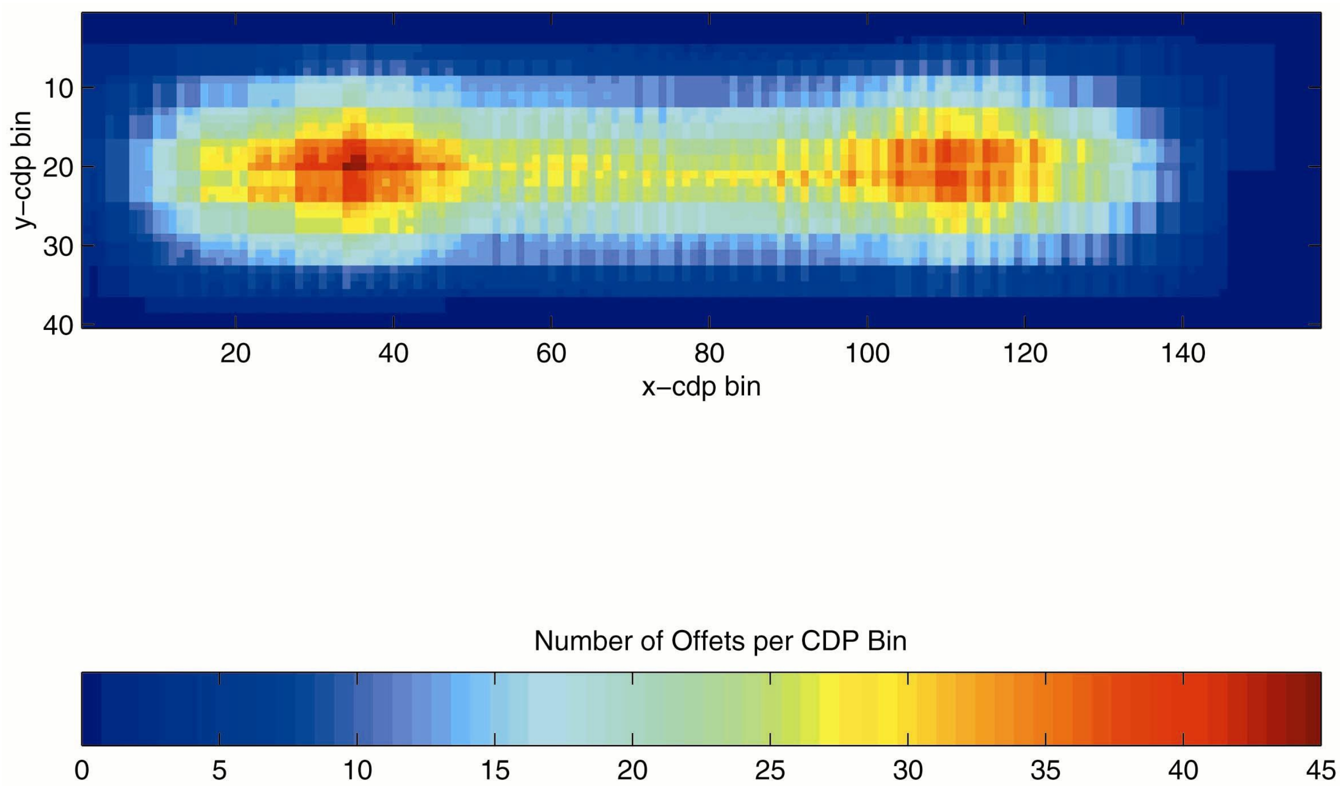
All the data (240 shots) were decimated



AVP from Incomplete data

ERSKINE 3D Data / WCSB /Alberta

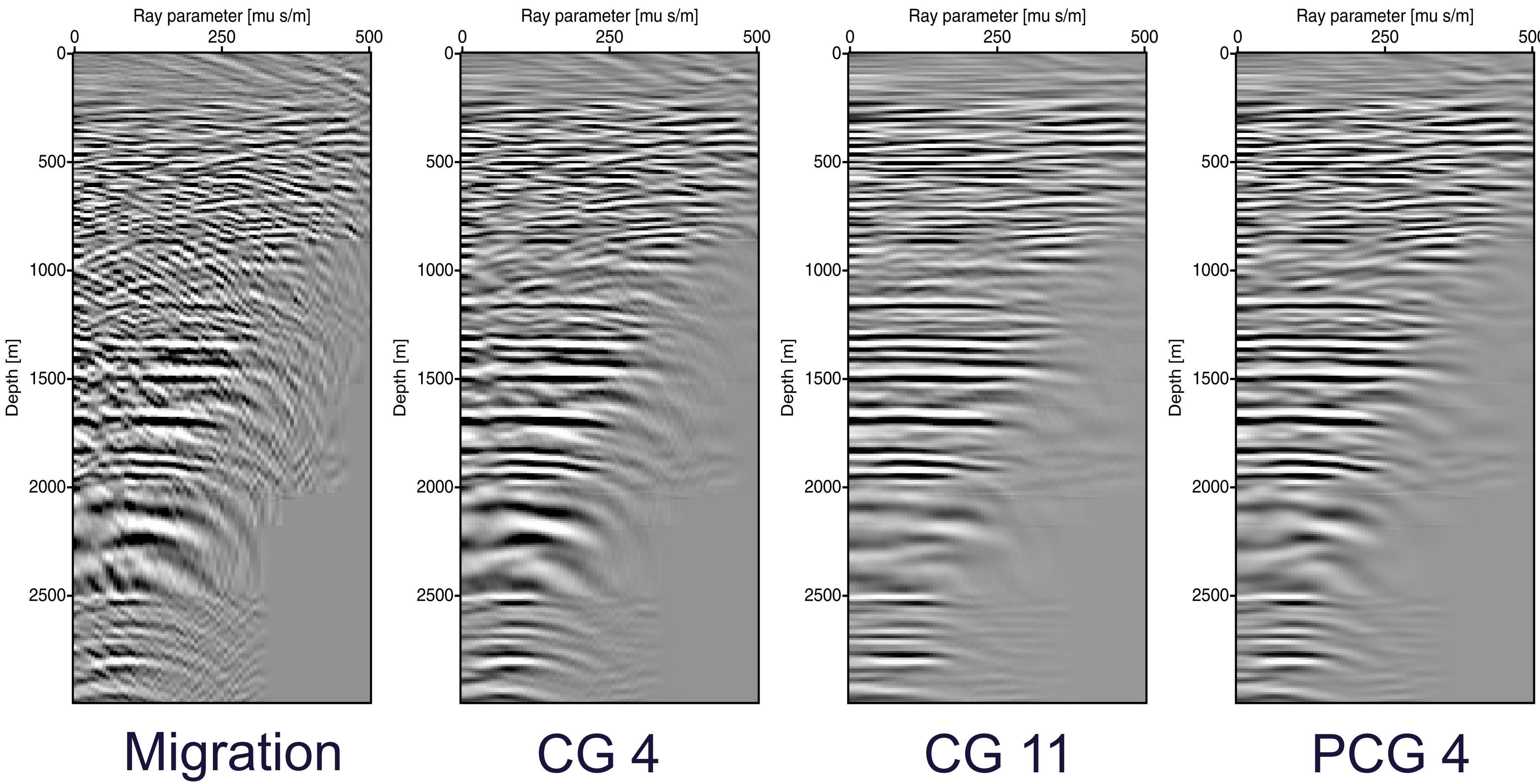
ERSKINE orthogonal 3-D sparse land data set



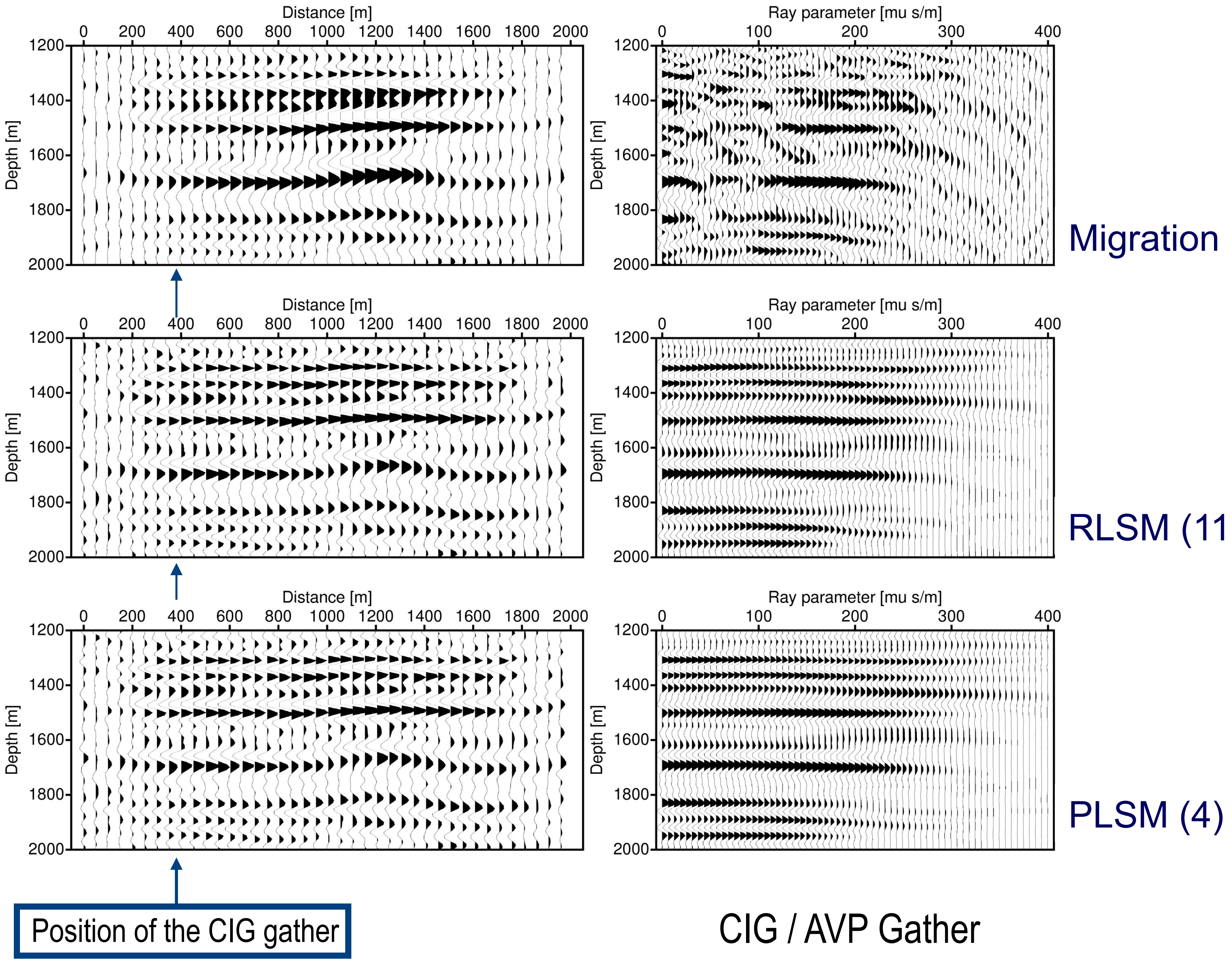
dx=50.29m

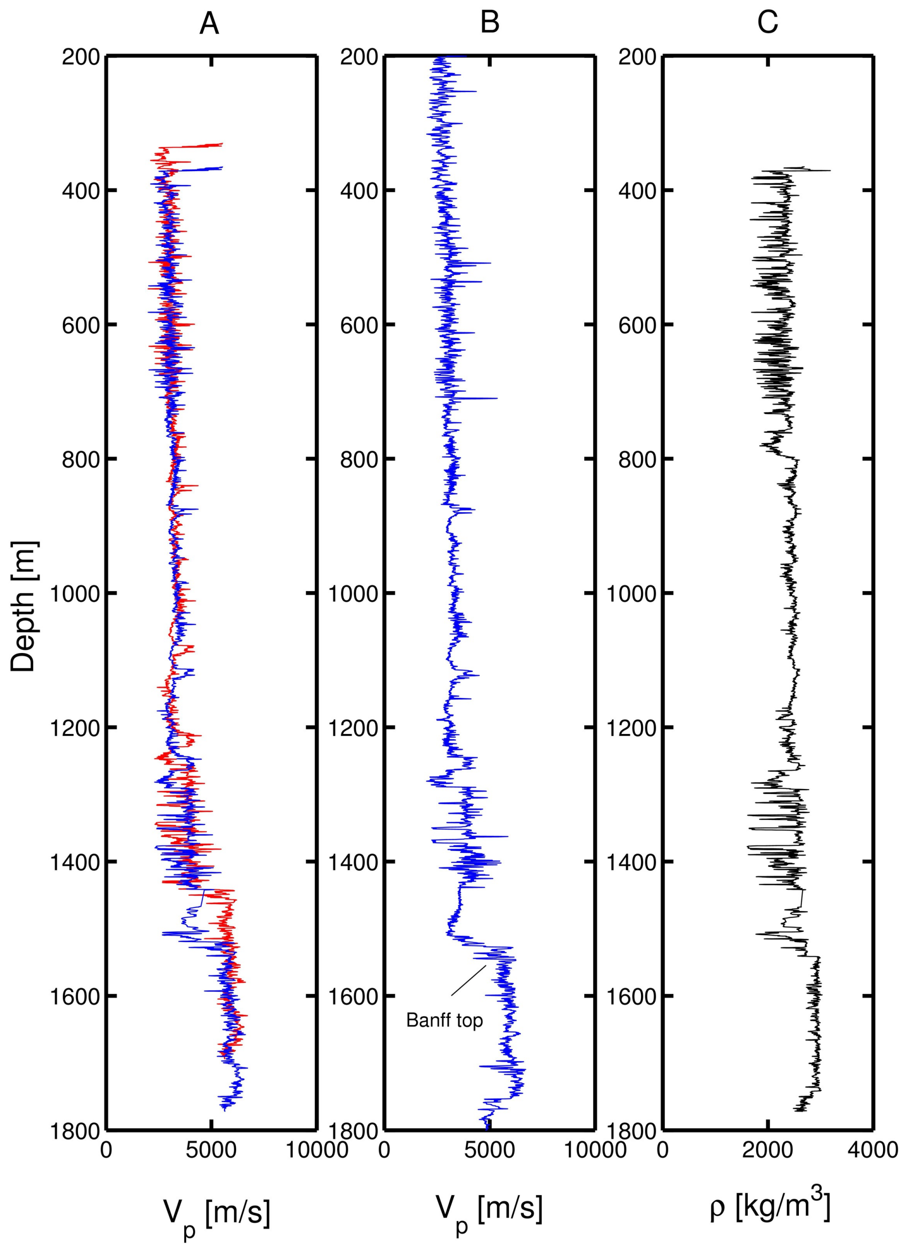
dy=33.5m

CIG at x-line #10, in-line #71



Comparison of LSM solution with regularization in the ray parameter direction. CG is the solution using conjugate gradients, PCG is the solution with Pre-conditioned CG.


Distance [m]
0 200 400 600 800 1000 1200 1400 1600 1800 2000
Depth [m]
1200 1400 1600 1800 2000
Ray parameter [mu s/m]
0 100 200 300 400
Migration
RLSM (11)
PLSM (4)
Position of the CIG gather
CIG / AVP Gather

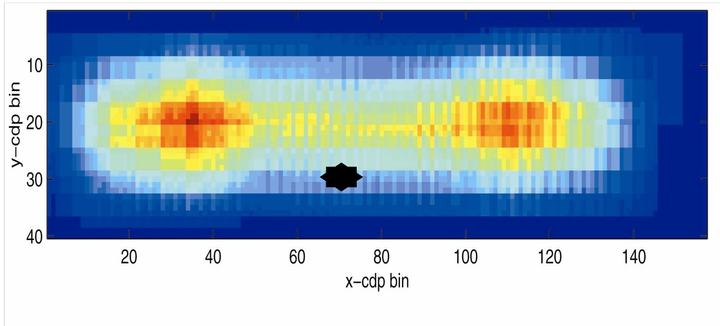


Wang J., Kuehl H. and Sacchi M.D., 2005, High-resolution wave-equation AVA imaging: Algorithm and tests with a data set from the Western Canadian Sedimentary Basin: Geophysics, 70, 891-899.

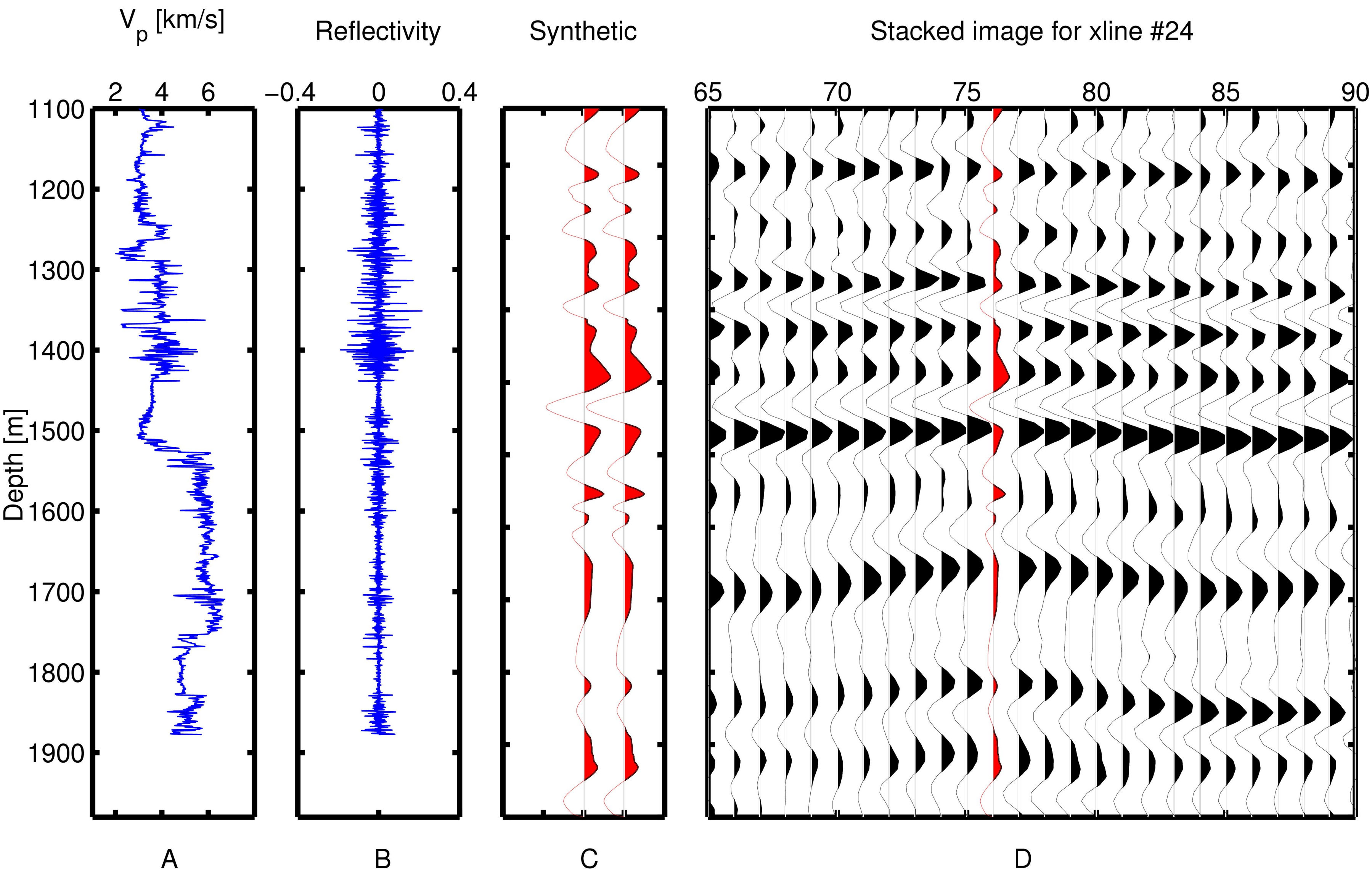
$V_P = (V_S - 1360)/1.16\ m/s$ (Castagna, 1985)

The analysis is restricted to the Ellerslie (sandstone) and Banff (shales) formations.

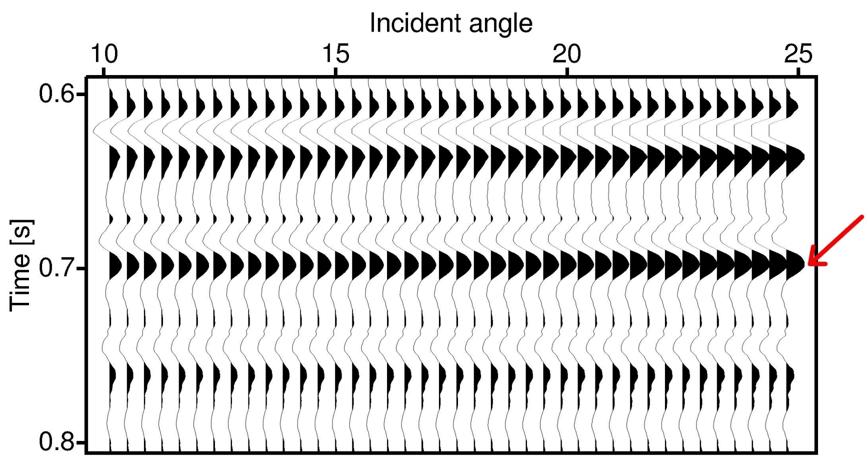
The economical target, the Leduc Reef, is excluded from the AVA analysis due the difficulty of estimating carbonate shear velocities.



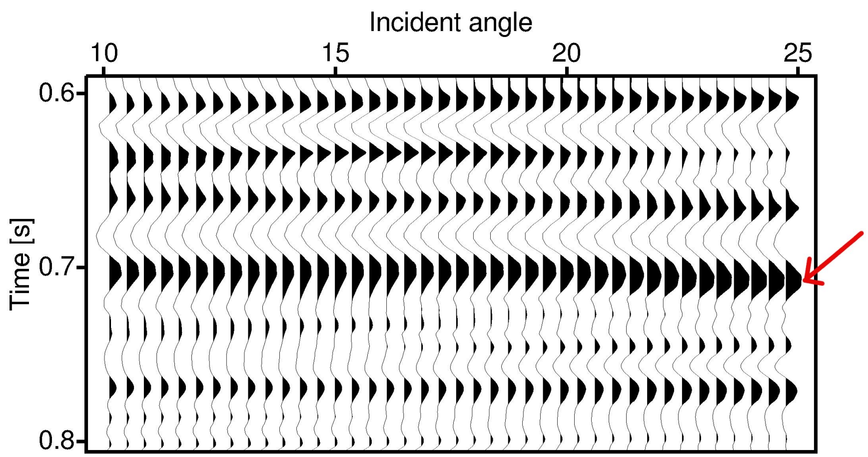
Synthetic trace



CIG (synthetic vs. inverted)

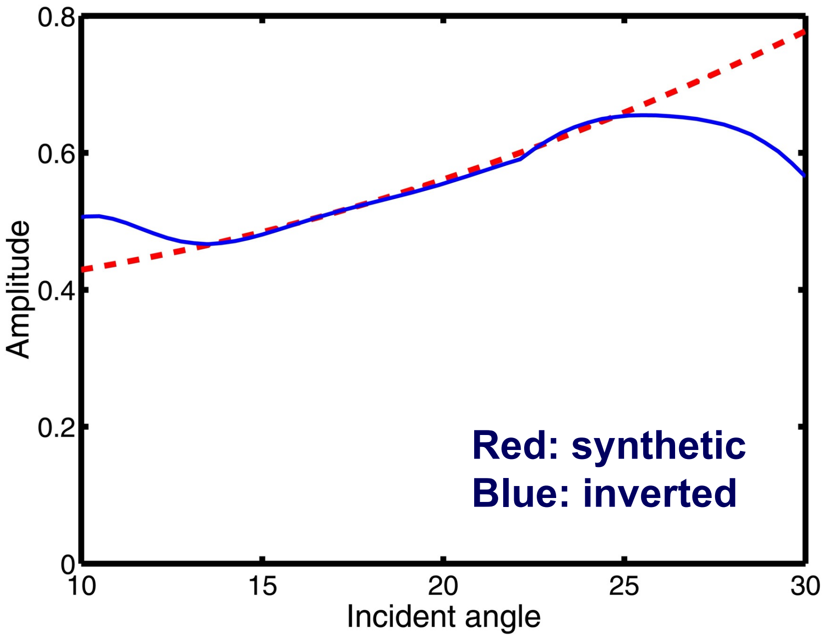


A



B

A: synthetic, B: inverted



Synthetic CIG is computed using Aki & Richards' approximation

Large scale Imaging - Including Resolution Enhancement (using Cauchy norm R)

$$J(\mathbf{m}) = \| \mathbf{W}(\mathbf{Lm} - \mathbf{d}) \|_2^2 + \mu \; R(\mathbf{SH}(\mathbf{m}))$$

$\mathbf{H}$: High Pass operator along p - gathers

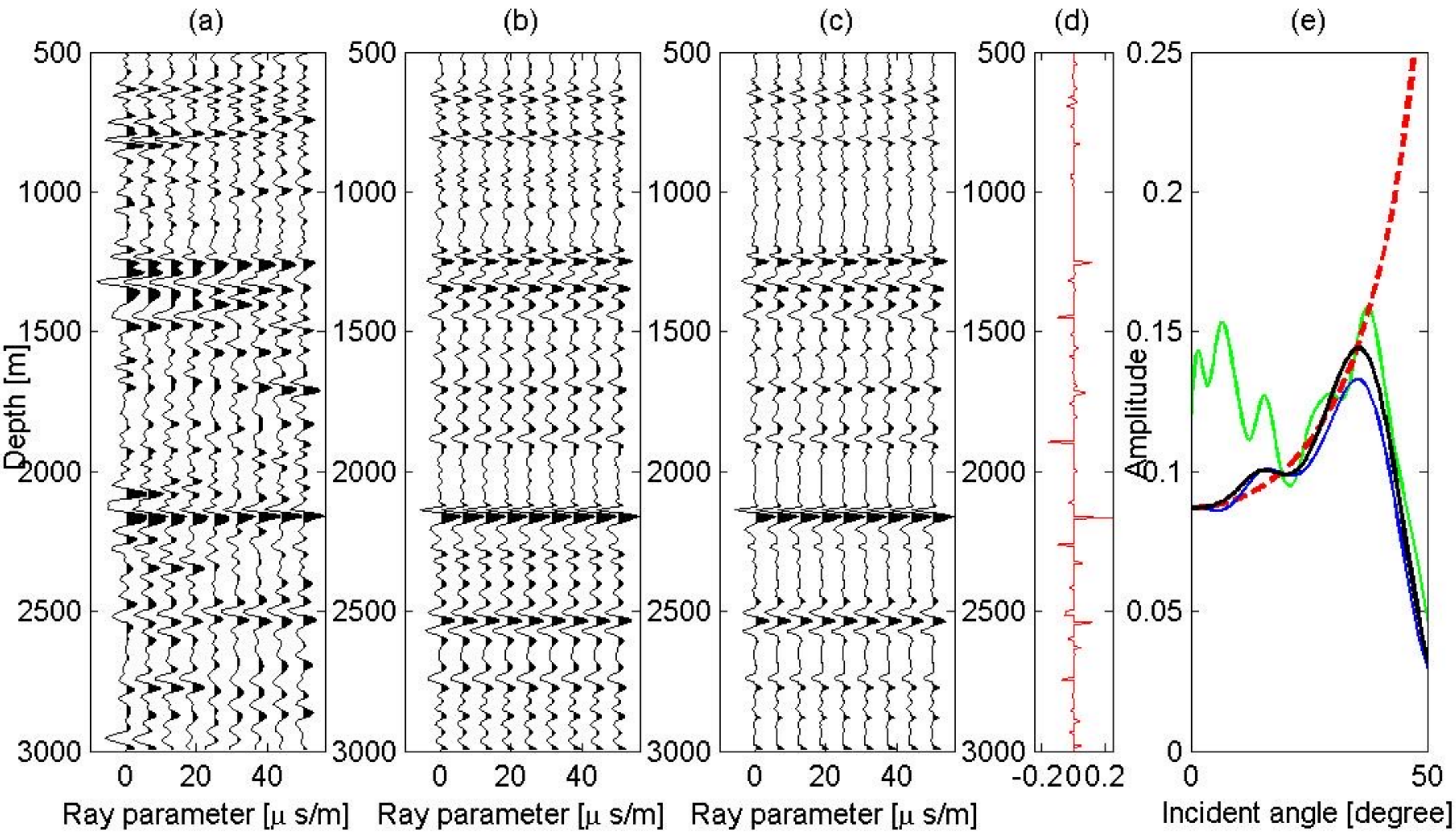
$\mathbf{S}$: Stacking on p – gathers

$$\mathbf{m} = \mathbf{Pz}$$

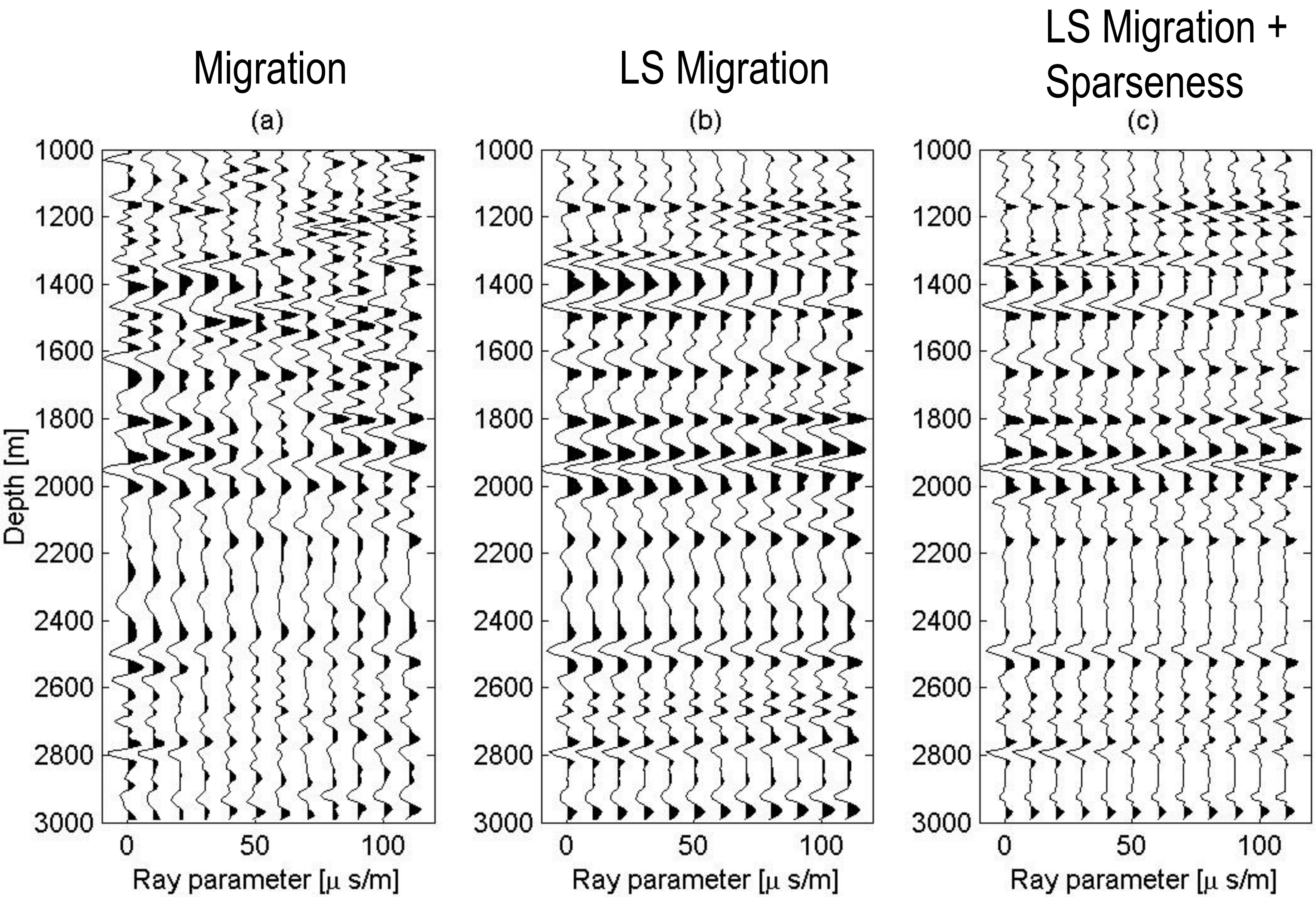
$$J'(\mathbf{m}) = \| \mathbf{W}(\mathbf{LPz} - \mathbf{d}) \|_2^2 + \mu \; R(\mathbf{Sz})$$

$\mathbf{P}$ is chosen to behave like $\mathbf{H}^{-1}$

Marmousi data / Common image gathers



Erskine data /Common image gathers



Addendum

The Bayesian Framework

A bit about Bayes

Thomas Bayes was born in London in 1702 into a religious atmosphere. His, father, the Rev. Joshua Bayes, was one of the first six Nonconformist ministers to be ordained in England. Like his father, Thomas was ordained a Nonconformist minister and assisted his father until the late 1720s when he become a Presbyterian minister.

Bayes theory appeared posthumously in “Essay Towards Solving a Problem in the Doctrine of Chances” published in the *Philosophical Transactions of the Royal Society of London* in 1764. Thomas Bayes died in England in 1761.

From: A Bayes tour of Inversion: A tutorial. Ulrych, Sacchi and Woodbury, *Geophysics*, 2001, 66-1, p.55-64.

Cost functions derived from Bayes rule

- Bayes rule
- MAP solution
- Cost functions induced by a Posterior
- Examples

Goal of this section

- Make the connection between probabilities and regularization
- Laplace distribution → l1 norm
- Gauss distribution → l2 norm
- Cauchy distribution → Cauchy criterion

Bayes theorem

$$p(A \mid B)\,p(B) = P(A)\,p(B \mid A)$$

$$p(A \mid B) = \frac{p(A)\,p(B \mid A)}{p(B)}$$

$p(A)$: Prior

$p(B \mid A)$: Probability of B given A

$p(A \mid B)$: Posterior

Update rule for

probabilities = state of information

Relation to Inverse problems

- Consider a linear problem $\mathbf{d} = \mathbf{L}\mathbf{m} + \mathbf{n}$

- Apply Bayes rule $p(A \mid B) = \dfrac{p(A)\, p(B \mid A)}{p(B)}$

$$p(A) = \mathrm{p(m)} \quad \text{Prior}$$

$$p(B \mid A) = p(d \mid m) \ \text{Likelihood}$$

$$p(A \mid B) = p(m \mid d) \ \text{Posterior}$$

$$p(\mathbf{m} \mid \mathbf{d}) = \frac{p(\mathbf{m})\, p(\mathbf{d} \mid \mathbf{m})}{p(\mathbf{d})}$$

Relation to Inverse Problems

$$\int p(\mathbf{m} \mid \mathbf{d}) d\mathbf{m} = 1 \quad \Rightarrow p(\mathbf{d}) = \int p(\mathbf{m}) p(\mathbf{d} \mid \mathbf{m}) d\mathbf{m}$$

$$p(\mathbf{m} \mid \mathbf{d}) = \frac{p(\mathbf{m}) p(\mathbf{d} \mid \mathbf{m})}{p(\mathbf{d})}$$

$$p(\mathbf{m} \mid \mathbf{d}) \propto p(\mathbf{m}) p(\mathbf{d} \mid \mathbf{m})$$

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

Example: Gaussian Noise and "Gaussian Model"

$$p(\mathbf{d}\,|\,\mathbf{m}) \propto \exp[-\frac{1}{2}(\mathbf{d}-\mathbf{Lm})^T \mathbf{C}_d{}^{-1}(\mathbf{d}-\mathbf{Lm})]$$

$$p(\mathbf{m}) \propto \exp[-\frac{1}{2}(\mathbf{m}-\mathbf{m_0})^T \mathbf{C}_m{}^{-1}(\mathbf{m}-\mathbf{m_0})]$$

$$\mathbf{m_0} = 0, \quad \mathbf{C}_m = \sigma_m^2 \mathbf{I}, \quad \mathbf{C}_d = \sigma_d^2 \mathbf{I}$$

$$p(\mathbf{m}\,|\,\mathbf{d}) \propto p(\mathbf{m})\, p(\mathbf{d}\,|\,\mathbf{m})$$

$$p(\mathbf{m}\,|\,\mathbf{d}) \propto \exp[-(\frac{1}{2\sigma_m^2}\mathbf{m}^T\mathbf{m} + \frac{1}{2\sigma_d^2}(\mathbf{d}-\mathbf{Lm})^T(\mathbf{d}-\mathbf{Lm}))]$$

Example: Gaussian Noise and "Gaussian Model"

Maximze

$$p(\mathbf{m}\,|\,\mathbf{d}) \propto \exp[-\frac{1}{2\sigma_m^2}\mathbf{m}^T\mathbf{m} - \frac{1}{2\sigma_d^2}(\mathbf{d}-\mathbf{L}\mathbf{m})^T(\mathbf{d}-\mathbf{L}\mathbf{m})] = \exp[-J]$$

Minimize

$$J = \frac{1}{2\sigma_m^2}\mathbf{m}^T\mathbf{m} + \frac{1}{2\sigma_d^2}(\mathbf{d}-\mathbf{L}\mathbf{m})^T(\mathbf{d}-\mathbf{L}\mathbf{m})$$

or

$$\boxed{\begin{aligned}\text{Cost} &= \frac{\sigma_d^2}{\sigma_m^2}\mathbf{m}^T\mathbf{m} + (\mathbf{d}-\mathbf{L}\mathbf{m})^T(\mathbf{d}-\mathbf{L}\mathbf{m}) \\ &= \mu\mathbf{m}^T\mathbf{m} + (\mathbf{d}-\mathbf{L}\mathbf{m})^T(\mathbf{d}-\mathbf{L}\mathbf{m})\end{aligned}}$$

Statistical Interpretation of the trade-off parameter

$$\boxed{\mu = \frac{\sigma_d^2}{\sigma_m^2}}$$

Noise $\uparrow$ then $\mu \uparrow$

Noise $\downarrow$ then $\mu \downarrow$

MAP Solution (Maximum a Posteriori)

Minimize

$$\text{Cost} = \frac{\sigma_d^2}{\sigma_m^2}\mathbf{m}^T\mathbf{m} + (\mathbf{d} - \mathbf{L}\mathbf{m})^T(\mathbf{d} - \mathbf{L}\mathbf{m})$$

$$= \mu\mathbf{m}^T\mathbf{m} + (\mathbf{d} - \mathbf{L}\mathbf{m})^T(\mathbf{d} - \mathbf{L}\mathbf{m})$$

$$\Rightarrow$$

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}$$

Remarks

- Today I’ve used the idea of regularization methods to construct a cost function
- Now I will derive the cost function for quadratic and non-quadratic regularization and quadratic misfit using Bayes theorem:

    - Probability of data given model is the probability of the noise and we assume Gaussian noise. The assuption will lead to a Quadratic Misfit

    - If we model our state of information about the model is encoded in a Gaussian pdf then the regularization term is Quadratic

    - Other priors are possible, they all lead to non-quadratic regularization methods

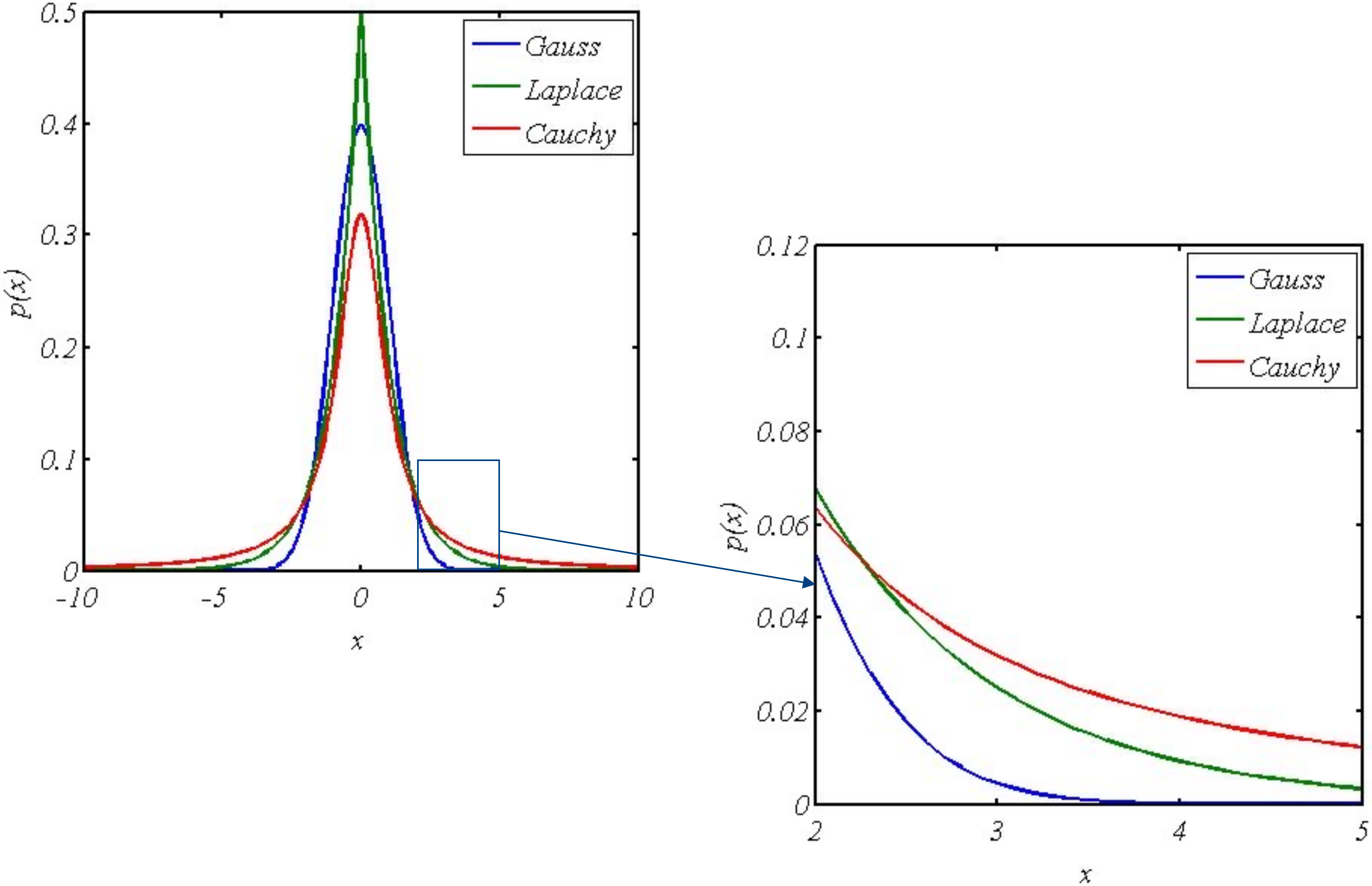
More on Bayes

Assume that the parameters to estimate (model) are non-Gaussian (say the follow a Laplace or a Cauchy distribution)

$$p(x) \propto \exp[-\lambda |x|^p]$$

$$p(x) \propto \exp[-\lambda |x|]$$ Laplace *p=1*

$$p(x) \propto \frac{1}{1+\frac{x^2}{\sigma^2}}$$ Cauchy


Gauss
Laplace
Cauchy
p(x)
x
0.5
0.4
0.3
0.2
0.1
0
-10
-5
0
5
10
Gauss
Laplace
Cauchy
p(x)
x
0.12
0.1
0.08
0.06
0.04
0.02
0
2
3
4
5

For instance use Cauchy

Maximize

$$p(\mathbf{m} \mid \mathbf{d}) \propto \prod_i \frac{1}{1 + m_i^2 / \sigma_m^2} \times \exp[-\frac{1}{2\sigma_d^2}(\mathbf{d} - \mathbf{Lm})^T (\mathbf{d} - \mathbf{Lm}))]$$

Minimize

$$Cost = \mu \sum_i \ln(1 + m_i^2 / \sigma_m^2) + (\mathbf{d} - \mathbf{Lm})^T (\mathbf{d} - \mathbf{Lm})$$

$$= \underbrace{\mu \sum_i \ln(1 + m_i^2 / \sigma_m^2)}_{\text{Cauchy “norm”}} + |\mathbf{d} - \mathbf{Lm}|_2^2$$

For instance, let's use the Laplace prior

Maximize

$$p(\mathbf{m} \mid \mathbf{d}) \propto \prod_i \exp(-\lambda \mid x_i \mid) \times \exp[-\frac{1}{2\sigma_d^2}(\mathbf{d}-\mathbf{Lm})^T(\mathbf{d}-\mathbf{Lm}))]$$

Minimize

$$Cost = \mu \sum_i \mid m_i \mid + (\mathbf{d}-\mathbf{Lm})^T(\mathbf{d}-\mathbf{Lm})$$

$$= \mu \underbrace{\mid \mathbf{m} \mid_1}_{\text{Ell-1 norm}} + \mid \mathbf{d}-\mathbf{Lm} \mid_2^2$$

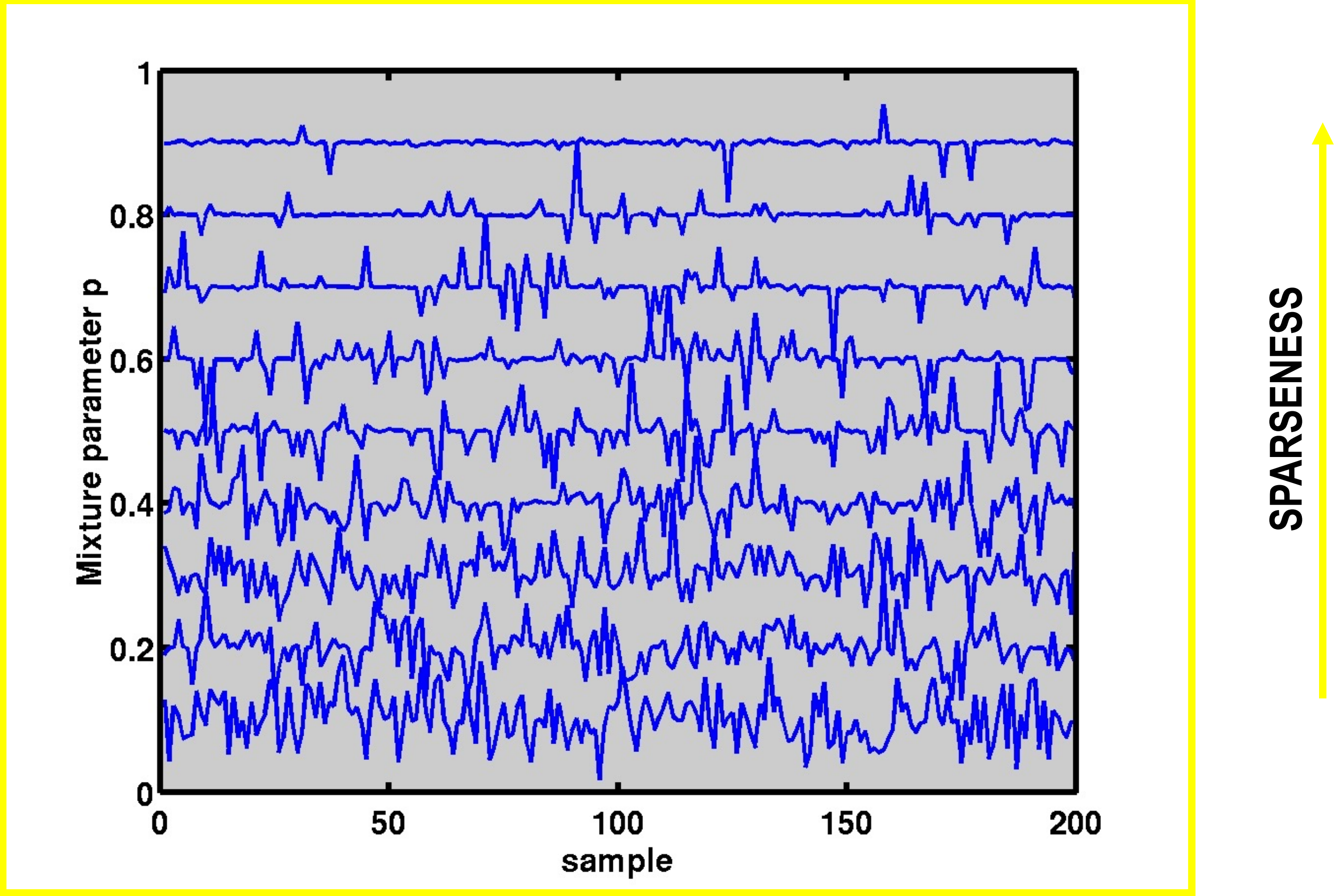
Addendum: What controls resolution? Sparsity and BW

- These examples are to show how critical is to understand that priors can create information not in the data

- One has to be careful when using HR solutions to problems such as decon and AVO inversion in the use of prior and, if possible, try to rely on extra information or physics/petrophysical constrains rather than sparse priors.

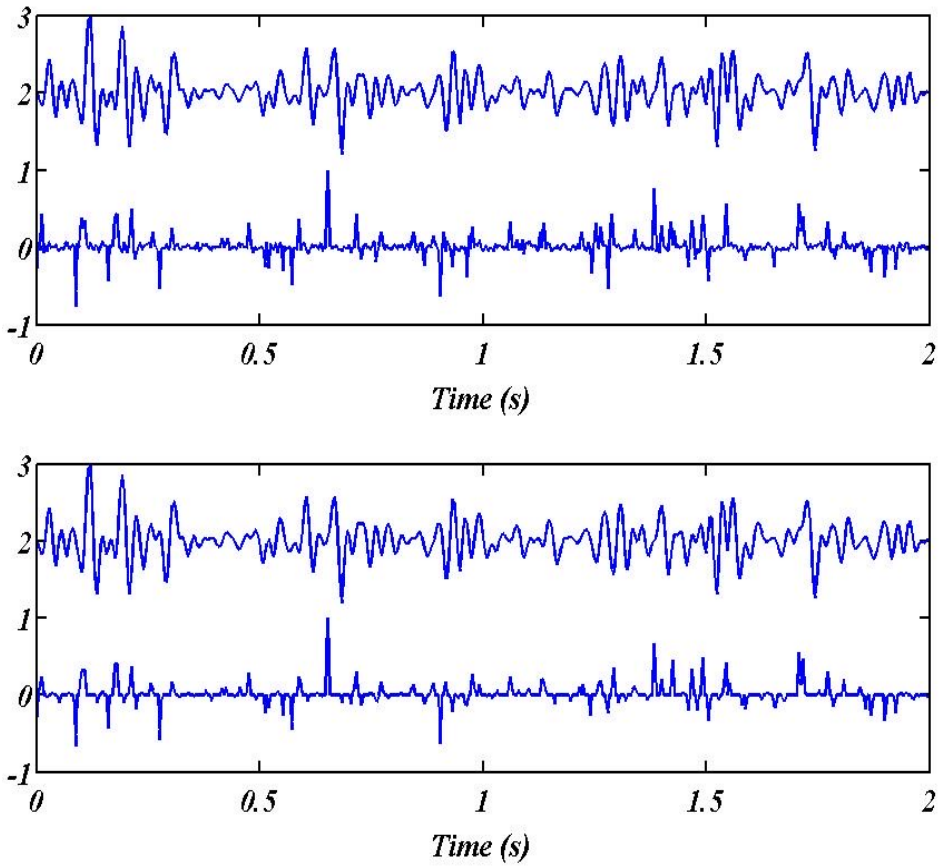
Example (Non-Gaussian reflectivity) retrieval with sparse inversion

- We assume a non-Gaussian reflectivity and known wavelet



Mixing Parameter *p=0.8 (very sparse)*

Cauchy inversion



Data

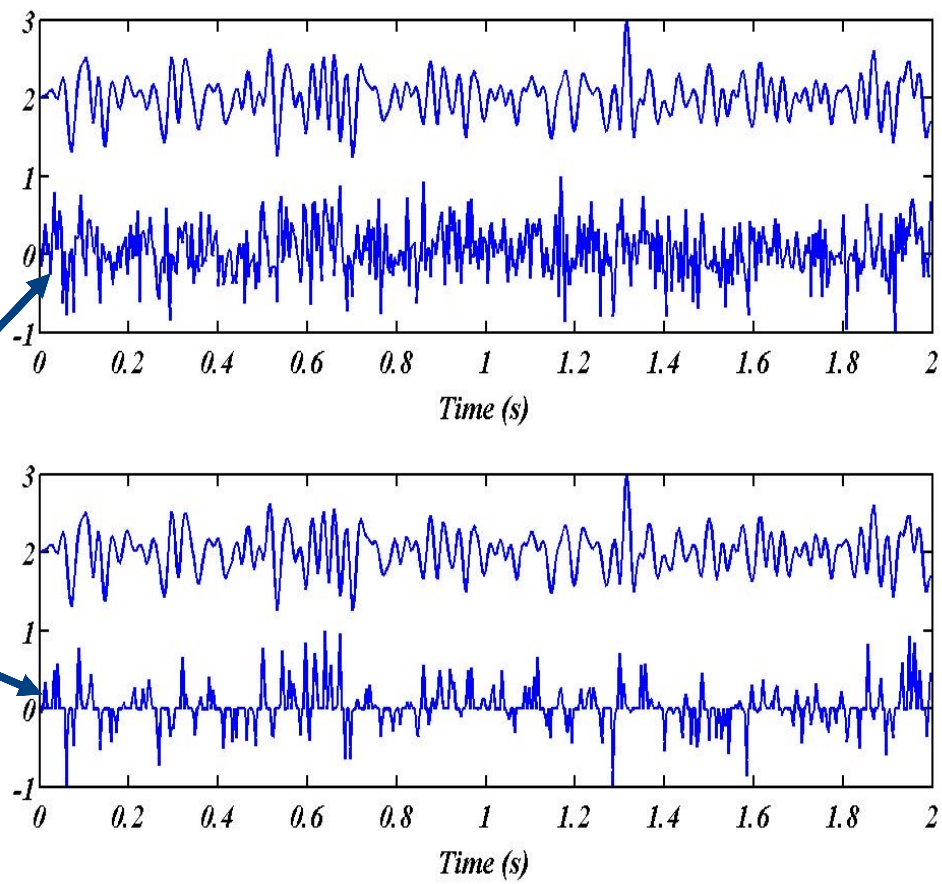
True Impulse Response

Predicted Data

Estimate Impulse Response

Mixing Parameter *p=0.2*

Cauchy inversion



The solution is too simple (sparse).

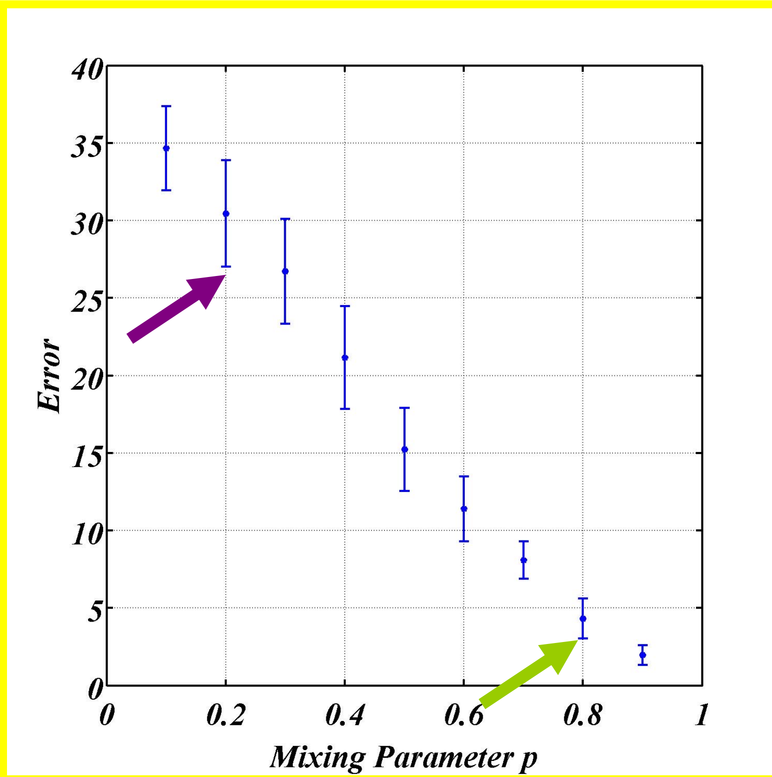
Data

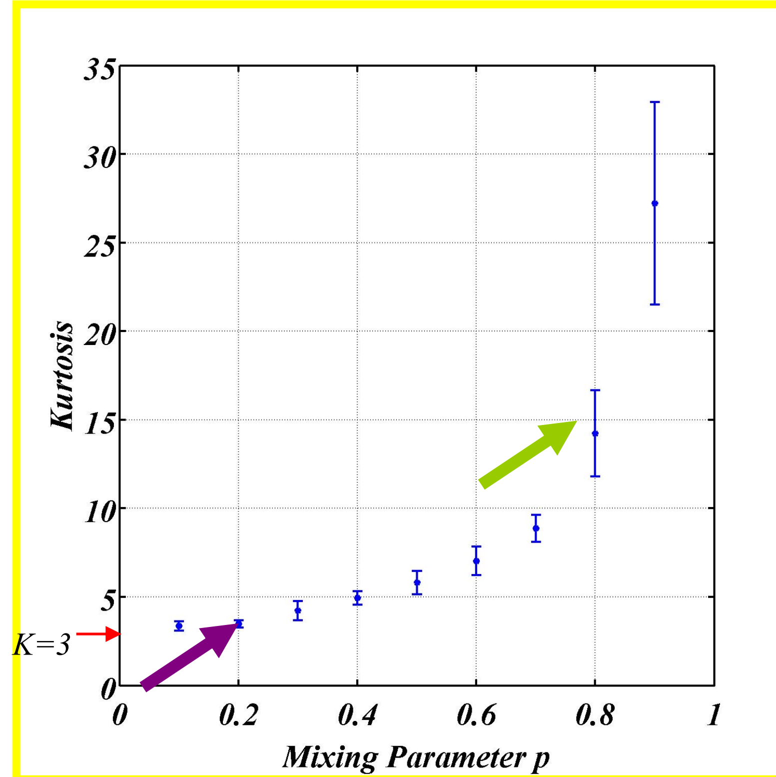
True Impulse Response

Predicted Data

Estimate Impulse Response

Kurtosis





Error = difference between true and estimated impulse response (reflectivity)

Results obtained after 20 realizations

Remark

- Key feature for proper recovery of the impulse response with sparse inversion..
  - Sparsity (of course)….. But something else is needed

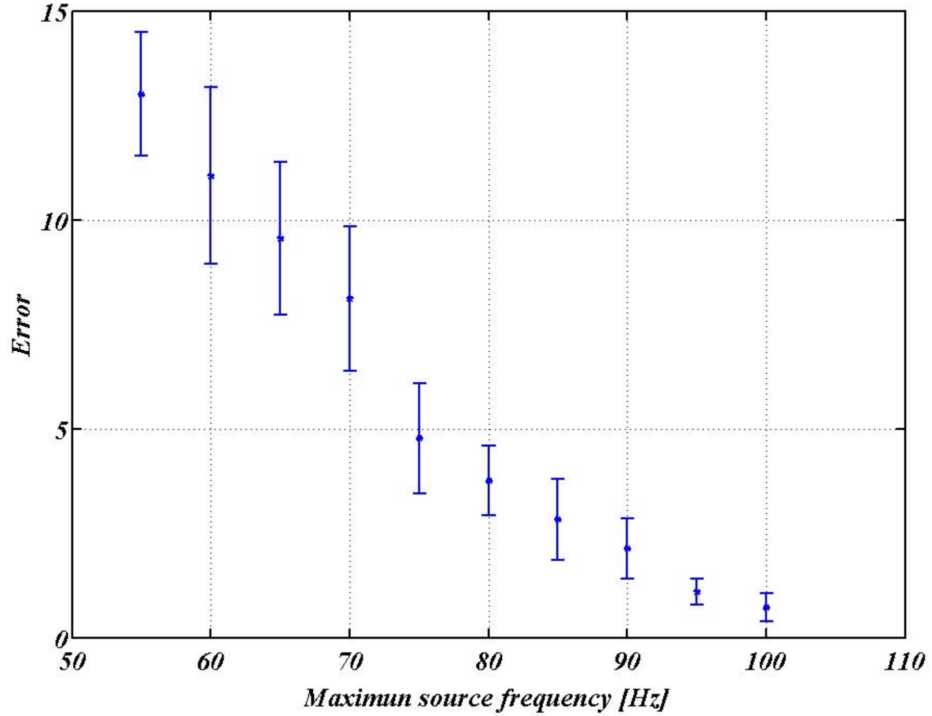
?

Remark

- Key feature for proper recovery of the impulse response with sparse inversion..
    - Sparsity (of course)….. But something else is needed

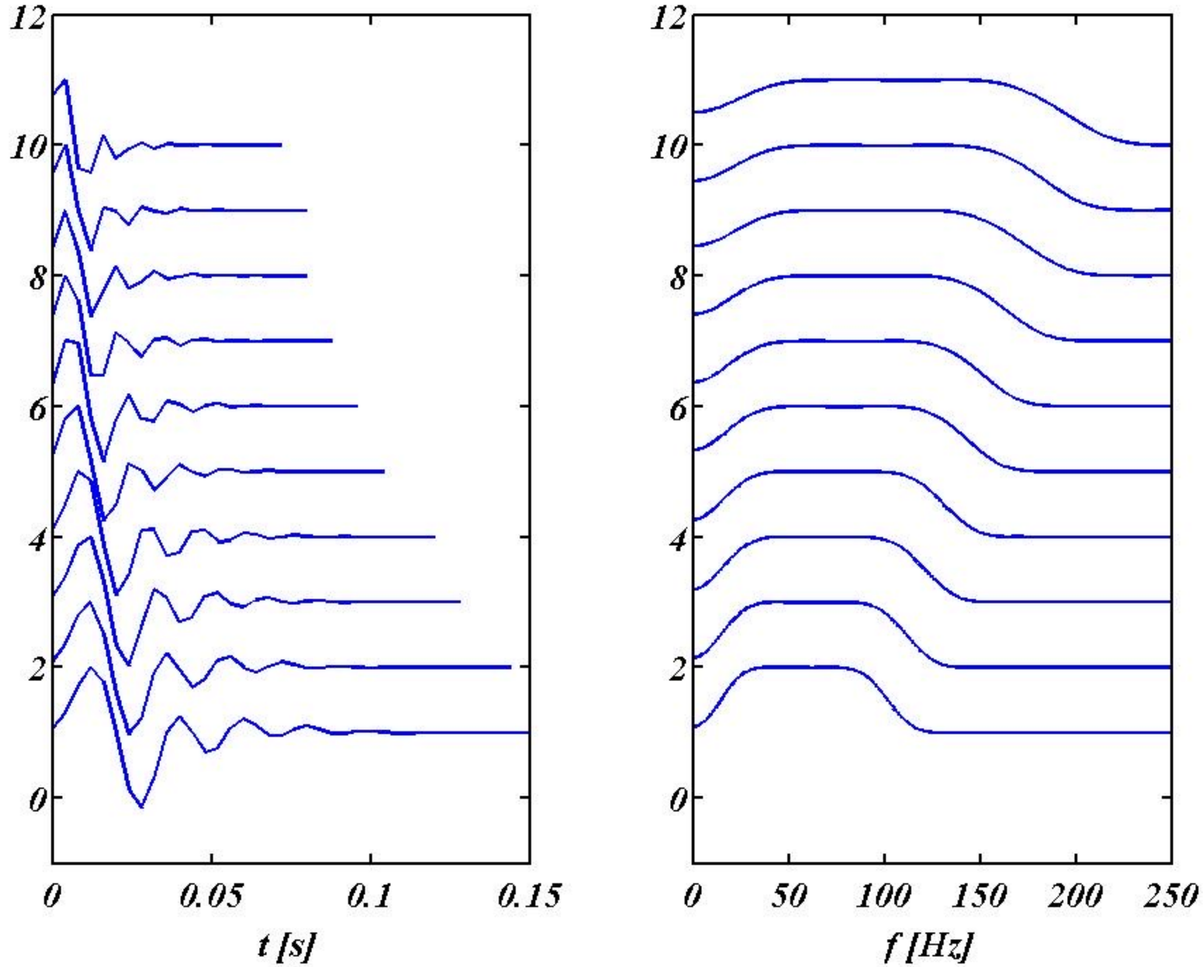
MEASURABLE INFORMATION…. Which is controlled by the source band-width

Source BW vs Estimation error for a sparse reflectivity (*p=0.5*)



Error = difference between true and estimated impulse response

Source functions used in the simulation



- BW is not only important for de-phasing (e.g. maximum Kurtosis de-phasing) but also for the inversion of the reflectivity
- *Nothing new*: more BW means that more information about the reflectivity is preserved in the seismic trace.

  – In the absence of sufficient BW, the prior will dominate the solution and dictate how the solution looks like. The geologist/interpreter (with access to borehole-derived reflectivity) will feel deeply discontented with the simplicity of your model.