Stable seismic data recovery

Felix J. Herrmann*
joint work with
Peyman Moghaddam*, Gilles Hennenfent* & Chris Stolk (Universiteit Twente)
*Seismic Laboratory for Imaging and Modeling
slim.eos.ubc.ca
AIP 2007, Vancouver, June 26
Combinations of **parsimonious** signal representations with nonlinear **sparsity** promoting programs hold the **key** to the next-generation of seismic inversion algorithms …

Since they allow for formulations that are **stable** w.r.t.

- noise
- incomplete data
- moderate phase rotations and amplitude errors

Finding a **sparse** representation for seismic data & images is complicated because of

- wavefronts & reflectors are multiscale & multi-directional
- the presence of caustics, faults and pinchouts
- the presence of operators (FIO’s & PsDO’s)
The seismic method
Seismic data acquisition
Exploration seismology

- create images of the subsurface
- need for higher resolution/deeper
- clutter and data incompleteness are problems
Exploration seismology

- create images of the subsurface
- need for higher resolution/deeper
- clutter and data incompleteness are problems
Forward problem

\[ F[c]u := \left( \frac{1}{c^2(x)} \cdot \frac{\partial^2}{\partial t^2} - \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} \right) u(x, t) = f(x, t) \]

- second order hyperbolic PDE
- interested in the singularities of

\[ m = c - \bar{c} \]
Inverse problem

Minimization:

\[ \tilde{m} = \arg \min_m \|d - F[m]\|_2^2 \]

After linearization (Born app.) forward model with noise:

\[ d(x_s, x_r, t) = (Km)(x_s, x_r, t) + n(x_s, x_r, t) \]

Conventional imaging:

\[ (K^T d)(x) = (K^T Km)(x) + (K^T n)(x) \]
\[ y(x) = (\Psi m)(x) + e(x) \]

\(\Psi\) is prohibitively expensive to invert
requires regular sampling ...
Sparsity promoting inversion
Formulate as inverse problem

$$\tilde{x} = \arg \min_{x} \|x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \epsilon$$

When a traveler reaches a fork in the road, the $l_1$-norm tells him to take either one way or the other, but the $l_2$-norm instructs him to head off into the bushes.

John F. Claerbout and Francis Muir, 1973

New field “compressive sampling”: D. Donoho, E. Candes et. al., M. Elad etc.

Preceded by others in geophysics: M. Sacchi & T. Ulrych and co-workers etc.
Sparsity promoting inversion

$x_0$ can be recovered by solving

$$\begin{align*}
P_\epsilon : & \quad \tilde{x} = \arg \min_{x} \| x \|_1 \quad \text{s.t.} \quad \| A x - y \|_2 \leq \epsilon \\
& \quad \tilde{f} = S^T \tilde{x}
\end{align*}$$

with

- $y = \text{(incomplete) data}$
- $A = \text{modeling matrix, e.g. } A = R S^T$
- $\tilde{x} = \text{recovered sparsity vector}$
- $\epsilon = \text{a number dependent on the noise level}$
- $S^T = \text{the synthesis matrix}$
- $\tilde{f} = \text{the recovered function } f$

**Crux lies in finding the sparse representation!**
Wish list

Transform that is parsimonious
- detects the wavefronts
- localized in space and frequency (phase space)
- some invariance under “wave propagation”

Events correspond to curved singularities with conflicting dips
- caustics
- faults & pinch outs

Need a transform that is
- multiscale
- multidirectional
- exactly reconstructs
Representations for seismic data

<table>
<thead>
<tr>
<th>Transform</th>
<th>Underlying assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>FK</td>
<td>plane waves</td>
</tr>
<tr>
<td>linear/parabolic Radon transform</td>
<td>linear/parabolic events</td>
</tr>
<tr>
<td>wavelet transform</td>
<td>point-like events (1D singularities)</td>
</tr>
<tr>
<td><strong>curvelet transform</strong></td>
<td><strong>curve-like events (2D singularities)</strong></td>
</tr>
</tbody>
</table>

Properties curvelet transform:

- **multiscale**: tiling of the FK domain into dyadic coronae
- **multi-directional**: coronae sub-partitioned into angular wedges, # of angle doubles every other scale
- **anisotropic**: parabolic scaling principle
- **Rapid decay space**
- **Strictly localized in Fourier**
- **Frame with moderate redundancy**
2-D curvelets

[Candès, Donoho, Demanet, Ying]

Oscillatory in one direction and smooth in the others!
Wavefront detection

curvelet coefficient is determined by the dot product of the curvelet function with the data
Compression

Interested in functions discontinuous along a piecewise smooth ($C^2$) interface, and otherwise smooth ($C^2$).

**Theorem** (Candès, Donoho). For such a model $f$, the best $m$-term curvelet expansion $f_m$ obeys

$$\|f - f_m\|^2 \leq C m^{-2} (\log m)^3.$$

Note: wavelets would give $O(m^{-1})$, so do ridgelets (Candès).

[From Demanet ‘05]
3-D curvelets

Curvelets live in wedges in the 3 D Fourier plane...
Nonlinear approximation

reconstructed data with p=5
Nonlinear approximation

reconstructed data with p=99
Curvelet-based seismic data recovery

joint work with Gilles Hennenfent
Sparsity-promoting inversion*

Reformulation of the problem

\[ \text{signal} \rightarrow y = RC^H x_0 + n \quad \text{noise} \]

Curvelet Reconstruction with Sparsity-promoting Inversion (CRSI)

- look for the **sparsest/most compressible, physical** solution

\[ P_\epsilon : \begin{cases} \hat{x} = \arg \min_x \| Wx \|_1 \quad \text{s.t.} \quad \| Ax - y \|_2 \leq \epsilon \\ \hat{f} = C^T \hat{x} \end{cases} \]

* inspired by Stable Signal Recovery (SSR) theory by E. Candès, J. Romberg, T. Tao, Compressed sensing by D. Donoho & Fourier Reconstruction with Sparse Inversion (FRSI) by P. Zwartjes
85 % missing
Curvelet recovery

$$A := RC^T$$
Observations

Inverted a rectangular matrix

- worked because the curvelet transform is sparse
- exploits the higher dimensional geometry of seismic wavefields
- curvelets are incoherent with the Dirac measurement basis

Data is recovered for large percentages of traces missing

Is an example of an inverse problem with incomplete data

Can these ideas be extended to recover migration amplitudes?

- approximately invert a PsDO
- diagonalize zero-order PsDO’s
Stable seismic amplitude recovery

“Sparsity- and continuity-promoting seismic image recovery with curvelet frames”

by

F.H, P. Moghaddam & C. Stolk

to appear in special issue on imaging in ACHA
Migrated data

Amplitude-corrected & denoised migrated data
Existing scaling methods

Methods are based on a diagonal approximation of $\Psi$

- Illumination-based normalization (Rickett ‘02)
- Amplitude preserved migration (Plessix & Mulder ‘04)
- Amplitude corrections (Guitton ‘04)
- Amplitude scaling (Symes ‘07)

We are interested in an ‘Operator and image adaptive’ scaling method which

- estimates the action of $\Psi$ from a *reference* vector close to the actual image
- assumes a *smooth* symbol of $\Psi$ in space and angle
- does not require the reflectors to be conormal $\iff$ allows for conflicting dips
- stably inverts the diagonal
Our approach

“Forward” model:

\[ y = K^T K m + \varepsilon \]

\[ \approx A x_0 + \varepsilon \]

with

- \( y = \) migrated data
- \( A := C^T \Gamma \)
- \( \text{AA}^T r \approx K^T Kr \)
- \( K = \) the demigration operator
- \( \varepsilon = \) migrated noise.

- diagonal approximation of the demigration-migration operator
- costs one demigration-migration to estimate the diagonal weighting
Solution

Solve

\[ \text{P : } \begin{cases} \min_{x} J(x) \quad \text{subject to} \quad \|y - Ax\|_2 \leq \epsilon \\ \tilde{m} = (A^H)^\dagger \tilde{x} \end{cases} \]

with

\[ J(x) = \alpha \|x\|_1 + \beta \|\Lambda^{1/2} (A^H)^\dagger x\|_p. \]

- need sparsity on the model
- invariance under the normal operator
Nonlinear approximation

Migrated mobil data set

reconstructed data with p=99
Nonlinear approximation

Recovery from largest 3 %

reconstructed data with $p=3$
Nonlinear approximation

Difference

Lateral (km)

Depth (km)

residue
Diagonal approximation of the Hessian
Normal/Gramm operator


In high-frequency limit $\Psi$ is a PsDO

$$(\Psi f)(x) = \int_{\mathbb{R}^d} e^{-ix \cdot \xi} a(x, \xi) \hat{f}(\xi) d\xi$$

- pseudolocal
- singularities are preserved

Inversion corrects for the ‘Hessian’
Invariance under Gramm matrix

- curvelets remain invariant
- approximation improves for higher frequencies
Approximation

So let $\Psi = \Psi(x, D)$ be a pseudodifferential operator of order 0, with homogeneous principal symbol $a(x, \xi)$.

$$K \mapsto K (-\Delta)^{-1/2} \quad \text{or} \quad K \mapsto \partial_t^{-1/2} K$$

$$m \mapsto (-\Delta)^{1/2} m \quad \text{with} \quad ((-\Delta)^{\alpha} f)^{\wedge}(\xi) = |\xi|^{2\alpha} \cdot \hat{f}(\xi).$$

**Lemma 1.** With $C'$ some constant, the following holds

$$\| (\Psi(x, D) - a(x_{\nu}, \xi_{\nu})) \varphi_{\nu} \|_{L^2(\mathbb{R}^n)} \leq C' 2^{-|\nu|/2}. \quad (14)$$

To approximate $\Psi$, we define the sequence $u := (u_\mu)_{\mu \in \mathcal{M}} = a(x_\mu, \xi_\mu)$. Let $D_\Psi$ be the diagonal matrix with entries given by $u$. Next we state our result on the approximation of $\Psi$ by $C^T D_\Psi C$. 
Approximation

**Theorem 1.** The following estimate for the error holds

\[ \|(\Psi(x, D) - C^T D\Psi C)\varphi_\mu\|_{L^2(\mathbb{R}^n)} \leq C'' 2^{-|\mu|/2}, \]

where \(C''\) is a constant depending on \(\Psi\).

Allows for the decomposition

\[
\left(\Psi \varphi_\mu\right)(x) \simeq \left(C^T D\Psi C \varphi_\mu\right)(x) = \left(A A^T \varphi_\mu\right)(x)
\]

with \(A := \sqrt{D\Psi} C\) and \(A^T := C^T \sqrt{D\Psi}\).
Approximation

\[ y(x) = (\Psi m)(x) + e(x) \]
\[ \simeq (AA^T m)(x) + e(x) \]
\[ = Ax_0 + e, \]

Wavelet-vaguelette like
Amenable to nonlinear recovery
Estimation of the diagonal scaling
Diagonal estimation

Define a reference vector (say conventional image).

Calculate ‘data’

\[ b = \Psi r \]

Define the matrix

\[ P := C^T \text{diag}(v) \quad \text{with} \quad v = Cr \]

Invert

\[ \tilde{u} = \arg\min_u \frac{1}{2} \|b - Pu\|_2^2 + \eta^2 \|Lu\|_2^2 \]
Diagonal estimation

Impose smoothness in phase space

\[ L = \begin{bmatrix} D_1 & D_2 & D_\theta \end{bmatrix} \]

Calculate: \( b = \Psi r \) and \( v = Cr \).

Set: \( \eta = \eta_{\text{min}} \);

\textbf{while} \( \exists (\hat{u}_\mu)_{\mu \in M} < 0 \) \textbf{do}

Solve

\[ \tilde{u} = \arg \min_u \frac{1}{2} \| b - Pu \|_2^2 + \eta^2 \| Lu \|_2^2 \]

Increase the Lagrange multiplier

\[ \lambda = \eta + \Delta \eta \]

\textbf{end while}
Figure 5: Estimates for the diagonal $\tilde{u}$ are plotted in (a-d) for increasing $\eta = \{0.01, 0.1, 1, 10\}$. The diagonal is estimated according the procedure outlined in Table 1 with the reference and 'data' vectors, and plotted in Fig. 4(b) and 4(c). As expected the diagonal becomes more positive for increasing $\eta$. 

Herrmann et.al. – 54
Seismic amplitude recovery
Recovery

Final form

\[ y = Ax_0 + \varepsilon \]

with \( x_0 = \Gamma Cm \) and \( \varepsilon = Ae. \)

Solve

\[
P : \begin{cases} 
\min_x J(x) & \text{subject to} & \|y - Ax\|_2 \leq \epsilon \\
\tilde{m} = (A^H)\dagger \tilde{x} 
\end{cases}
\]

with

\[
J(x) = \alpha \|x\|_1 + \beta \|\Lambda^{1/2} (A^H)^\dagger x\|_p.
\]
Image recovery
anisotropic diffusion

[Black et. al '98, Fehmers et. al. '03 and Shertzer '03]

Define

\[ J_c(m) = \| \Lambda^{1/2} \nabla m \|_p \]

with \( p=2 \)

\[ \Lambda[r] = \frac{1}{\| \nabla r \|^2 + 2\nu} \left\{ \begin{pmatrix} +D_2r \\ -D_1r \end{pmatrix} \begin{pmatrix} +D_2r & -D_1r \end{pmatrix} + \nu \text{Id} \right\} \]
Gradient of the reference vector

lateral (m)  depth (m)

2000  4000  6000  8000  10000  12000  14000

500  1000  1500  2000  2500  3000  3500
Recovery

Step 1: Update of the Jacobian of $\frac{1}{2}\|y - Ax\|_2^2$:

$$x \leftarrow x + A^T (y - Ax);$$

Step 2: projection onto the $\ell_1$ ball $S = \{||x||_1 \leq ||x_0||_1\}$ by soft thresholding

$$x \leftarrow T_{\lambda w}(x);$$

Step 3: projection onto the anisotropic diffusion ball $C = \{x : J(x) \leq J(x_0)\}$ by

$$x \leftarrow x - \kappa \nabla x J_c(x)$$
Initialize:

\[ m = 0; \]
\[ x^0 = 0; \]
\[ y = K^T d; \]

Choose:

\[ M \text{ and } L \]

\[ \|A^T y\|_\infty > \lambda_1 > \lambda_2 > \cdots \]

while \[ \|y - A\tilde{x}\|_2 > \epsilon \] do

\[ m = m + 1; \]
\[ x^m = x^{m-1}; \]

for \( l = 1 \) to \( L \) do

\[ x^m = T_{\lambda_m}(x^m + A^T (y - x^m)) \{ \text{Iterative thresholding} \} \]

end for

Anisotropic descent update;

\[ x^m = x^m - \beta \nabla x^m J_c(x^m); \]

end while

\[ \tilde{x} = x_m; \tilde{m} = (A^T)^\dagger \tilde{x}. \]

Table 2: Sparsity-and continuity-enhancing recovery of seismic amplitudes.
Application to the SEG AA’ model
Example

SEGAA’ data:
- “broad-band” half-integrated wavelet [5-60 Hz]
- 324 shots, 176 receivers, shot at 48 m
- 5 s of data

Modeling operator
- Reverse-time migration with optimal check pointing (Symes ‘07)
- 8000 time steps
- **linearized** modeling 64, and migration 294 minutes on 68 CPU’s

Scaling required 1 extra migration-demigration
Migrated data

Amplitude-corrected & denoised migrated data
Noise-free data

Noisy data (3 dB)

Data from migrated image

Data from amplitude-corrected & denoised migrated image
Example

SEGAA’ data:
- “broad-band” half-integrated wavelet [5-60 Hz]
- 324 shots, 176 receivers, shot at 48 m
- 5 s of data

Modeling operator
- Reverse-time migration with optimal check pointing (Symes ‘07)
- 8000 time steps
- **full** modeling

Scaling required 1 extra migration-demigration
Conclusions

Curvelet-domain scaling
- handles conflicting dips (conormality assumption)
- exploits invariance under the PsDO
- robust w.r.t. noise

Diagonal approximation
- exploits smoothness of the symbol
- uses “neighbor” structure of the curvelet transform

Results on the SEG AA’ show
- recovery of amplitudes beneath the Salt
- successful recovery of clutter
- improvement of the continuity
Acknowledgments

The authors of CurveLab (Demanet, Ying, Candes, Donoho)

Dr. W.W. Symes for his reverse-time migration code

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE (334810-05) of F.J.H. This research was carried out as part of the SINBAD project with support, secured through ITF (the Industry Technology Facilitator), from the following organizations: BG Group, BP, Chevron, ExxonMobil and Shell.