Compressive sampling meets seismic imaging

Felix J. Herrmann
fherrmann@eos.ubc.ca
http://slim.eos.ubc.ca

joint work with Tim Lin and Yogi Erlangga

Seismic Laboratory for Imaging & Modeling
Department of Earth & Ocean Sciences
The University of British Columbia

SIAM Conference on Imaging Science
San Diego, July 2008
Motivation

Seismic data processing, modeling & imaging
- firmly rooted in Nyquist’s paradigm
  - sampling (e.g. of wavefields)
  - sampling of solutions (e.g. of PDEs)
- acquisition, modeling & inversion costs are proportional to the size of data and model

New paradigm of compressive sensing (CS)
- Nyquist is too pessimistic for signals with structure
  - existence of some sparsifying transform (e.g. wavelets)
  - existence of some low-dimensional structure (smooth manifolds)
- allows for recovery from sample rates $\approx$ computational cost proportional to the complexity of data and model
Main ingredients

New **preconditioner** for the *Helmholtz* operator
[Erlanga & Nabben, ‘06–’08, Elangga, Lin, F.J.H., ‘08]

Current advent of **simultaneous & continuous**
source acquisition and modeling
[Romero et. al., ‘00; Neelamani & C.E. Krohn, ‘08]

Sparsity-promoting **recovery** using results from **CS**
[Donoho, ‘06; Candes et al., ‘06; Candes and Tao, ‘06]
Consider the following (severely) underdetermined system of linear equations

\[ \text{data (simulations/observations)} \rightarrow \begin{bmatrix} y \\ A \end{bmatrix} = \begin{bmatrix} x_0 \\ \text{unknown} \end{bmatrix} \]

Is it possible to recover \( x_0 \) accurately from \( y \)?
CS

perfect recovery

conditions:
- $A$ obeys the *uniform uncertainty principle*
- $x_0$ is *sufficiently sparse*

procedure:

$$\min_x \|x\|_1 \quad \text{s.t.} \quad Ax = y$$

subject to sparsity

perfect reconstruction

performance:
- $S$-sparse vectors recovered from roughly on the order of $S$ measurements
  (to within constant and $\log$ factors)
Adjoint state method

Unconstrained nonlinear LS problem

\[
\min_{m \in M} \frac{1}{2} \| b - F[m] \|_2^2
\]

with

\[
F[m] = DA^{-1}[m]f
\]

and the gradient = \textit{migrated image},

\[
[\nabla J(m)]_i = -\Re \left( \sum_\omega \sum_s \langle \left( \frac{\partial A}{\partial m} e_i \right) u_s, v_s \rangle \right)
\]

involves for each monochromatic shot the solution of

\[
A[m]u = f \quad \text{and} \quad A^H[m]v = r
\]

with

\[
r = D^H(b - F[m])
\]
Forward modeling

**Current paradigm:** *time-domain finite differences*

**Pro:** relatively simple, implicit and fast

**Con:**
- discretization criteria for numerical stability
- storage requirements for
  - model (domain decompositions)
  - imaging conditions (check pointing)

**New ‘paradigm’:** *implicit preconditioned Helmholtz solvers*

**Pro:**
- matrix free, favorable criteria for numerical stability
- embarrassing parallelization over angular frequency

**Con:**
- slow or no convergence of indirect Krylov methods

**Solution:** preconditioner
Forward modeling

Discretize frequency-domain acoustic wave equation

$$\mathcal{H} u(\omega, x_s; x) := -\left(\nabla \cdot \nabla - \frac{\omega^2}{c(x)^2}\right) u(\omega, x_s; x) = b$$

Monochromatic linear system

$$A_\omega [c] u^s = b^s$$

Preconditioned system

$$A M^{-1} \hat{u} = b, \quad u = M^{-1} \hat{u},$$

derived from shifted Laplacian

$$\mathcal{M} := -\nabla \cdot \nabla - \frac{\omega^2}{c(x)^2}(1 - \beta i) \quad \text{with} \quad i = \sqrt{-1}, \beta > 0$$
Forward modeling cont’d

**Preconditioning** [Erlangga & Nabben, ‘06–’08]:
- moves eigenvalues to circle in complex plane
- is inverted using multigrid (no longer elliptic)

**Additional multi-level Krylov projection:**

\[ AM^{-1}Q\hat{u} = b, \quad u = M^{-1}Q\hat{u}, \]

- moves eigenvalues to real axis near 1
- improves condition number
Forward modeling cont’d

- implicit solvers converge
- number of iterations flat in grid size & frequency
- opens perspective to large-scale parallel solver for 3-D models
Despite significant improvement by Helmholtz preconditioner
- redundancy $\leftrightarrow$ extreme large size seismic data volumes
- multiple frequencies & multiple right-hand sides
- expensive modeling, imaging & inversion costs

Leverage new paradigm of CS ...
Relation to existing work

Simultaneous & continuous acquisition:
- *Simultaneous Sourcing without Compromise* by R. Neelamani & C.E. Krohn, ’08.

Simultaneous simulations & migration:
- *Phase encoding of shot records in prestack migration* by Romero et. al., ’00.

Imaging:
- *How to choose a subset of frequencies in frequency-domain finite-difference migration* by Mulder & Plessix, ’04.

Wavefield extrapolation:
- *Compressed wavefield extrapolation* by T. Lin and F.J.H, ’07
- *Compressive wave computations* by L. Demanet in MS79
Individual shots
Simultaneous & continuous shots
Simultaneous modeling & acquisition

Current paradigm:
- *separate* single-source experiments in the field
- *separate* single-shot simulations in the computer
- **Con:** expensive

New paradigm:
- *simultaneous* & *continuous* source experiments in the field
- *simultaneous* (continuous) simulations in the computer
- *continuous* simultaneous simulations are equivalent to *multiple* simultaneous experiments
- **Con:** *postprocessing necessary to separate into individual shots*

Key observation: this is *really* CS ...
Forward modeling
multishot

Helmholtz equation solved for:
- individual angular frequencies, i.e.,
  \[ \omega_i = 2\pi i \cdot \Delta f, \quad i = 1 \cdots n_f, \]
  - \( n_f \) number of frequencies
- \( \Delta f \) the sample interval in frequency
- individual shots, i.e., \( b_i = e_i \) for \( i = 1, \cdots, n_s \)
Forward modeling
multishot

Rewrite into

\[
\begin{bmatrix}
A_{\omega_1} & 0 \\
0 & A_{\omega_2} \\
& \ddots & \ddots & \ddots & 0 \\
0 & & & A_{\omega_{nf}} \\
\end{bmatrix}
\begin{bmatrix}
U_{\omega_1} \\
U_{nf} \\
\end{bmatrix}
=
\begin{bmatrix}
B_{\omega_1} \\
B_{nf} \\
\end{bmatrix}
\]

or

\[
LU = B.
\]

Modeling involves the inversion of the matrix

\[
L \in \mathbb{C}^{n_d \times n_d} \text{ with } n_d = 2n_f n_s n_r
\]
Equivalence

Show equivalence between
- CS sampling of **full** solution for separate single-source (sweep) experiments
- Solution of **reduced** system after CS sampling the collective single-shot source wavefield => simultaneous source experiments

\[
B = D^* \\
L^\omega L^{-1} B R^* \Sigma = U \Leftrightarrow U = \left(R_\omega L R^* \Sigma\right)^{-1} B
\]

Show that \( y = \underline{y} \) for which it is sufficient to show that
- full system
- reduced system
Equivalence cont’d

Fourier restriction:

\( R_\Omega : n'_f \times n_f \) block matrix, \( n'_f = \#\{\Omega\} , \Omega \subset \{\omega_i\}, i = 1, \ldots, n_f, n_f \gg n'_f \)

\[
[R_\Omega]_{J,I} = \begin{cases} 
I_{n_x \times n_z} , & I \in \mathcal{I} \\
0_{n_x \times n_z} , & I \notin \mathcal{I},
\end{cases}
\]

with \( \mathcal{I} \) the index set of \( \Omega \), and \( J = 1, \ldots, n'_f \).

Identity: \( R_\Omega L = L R_\Omega \), where

\[
L = \text{diag}(A_{\omega_I}) , \quad I \in \mathcal{I}.
\]

This implies: \( R_\Omega L^{-1} = L^{-1} R_\Omega \).
Equivalence cont’d

Shot restriction:

\( R_\Sigma : n'_s \times n_s \) rectangular matrix, \( n'_s = \# \{ \mathcal{N}'_s \} \), \( \mathcal{N}'_s \subset \mathcal{N}_s \), with \( \mathcal{N}_s \) the index set of \( b_i \).

\[
[R_\Sigma]_{j,i} = \begin{cases} 
1, & i \in \mathcal{N}'_s; \\
0, & i \notin \mathcal{N}'_s;
\end{cases}
\]

for \( j = 1, \ldots, n'_s \) with \( n'_s \ll n_s \).

So we have,

\[
R_\Omega L^{-1} = L^{-1} R_\Omega \Rightarrow \underbrace{R_\Omega L^{-1} B R^*_\Sigma}_{\text{U}} = \underbrace{L^{-1} R_\Omega B R^*_\Sigma}_{\text{B}} = \underbrace{U}_{\text{U}}.
\]

implying

\[
y = y
\]
Experiment

\[ y = RMDU \]

\[ \underline{y} = DU \]
Current paradigm: *Nyquist sampling*

**Pro:**
- linear
- signal independent (aside from Nyquist frequency)

**Con:**
- cost dependent on the Nyquist frequency and model size
- overly pessimistic for signals with *structure*

New paradigm: *Compressive sensing*

**Pro:**
- cost dependent on signal’s *complexity*

**Con:**
- solve a nonlinear recovery problem

Can lead to reduced cost when recovery cost < reduced simulation costs ...
CS

\[ \begin{align*}
\mathbf{y} &= \mathbf{RMf} \\
\tilde{\mathbf{x}} &= \arg \min_{\mathbf{x}} \| \mathbf{x} \|_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{y} \\
\mathbf{A} &= \mathbf{RMS}^* \\
\tilde{\mathbf{f}} &= \mathbf{S}^* \tilde{\mathbf{x}}
\end{align*} \]

P1:

CS provides conditions under which P1 recovers f:
- selection of CS-matrix (Measurement & Restriction matrices)
- selection of sparsifying transform
- large-to-extremely large problem size
- projected gradient with root finding method (SPG\(_1\), Friedlander & van den Berg, ‘07–’08)
- CS matrix has to lead to \textit{physically realizable} source wavefield for modeling & acquisition
CS

Selection of the CS-matrix

- natural restriction in Fourier ($F$) with *importance* sampling in the temporal direction
- CS with Gaussian ($N$) matrix along shots => simultaneous sources
- assures *incoherence* with sparsifying transform

For each *simultaneous* shot, define different restrictions

\[
RM = \begin{bmatrix}
R_1^{\Sigma} \otimes R_1^{\Omega} \\
\vdots \\
R_{n_s'}^{\Sigma} \otimes R_{n_s'}^{\Omega}
\end{bmatrix} \otimes (N \otimes F')
\]

yielding the reduced simulated data

\[
y = y = RMd, \; y \in \mathbb{C}^{n'_d}
\]

with \( n'_d = n'_f n'_s n_r \ll n_d = 2n_f n_s n_r \)
Selection of the sparsifying transform:

- wavelet transform is known to compresses seismic data [Donoho ‘99]
- successfully applied in MRI (reconstructions from incomplete Fourier data)[Lustig et. al. ‘07]

Define

$$S = W \otimes W \otimes W$$

**Bottom line:**
Computational gain of CS proportional to undersampling ratio

$$\frac{n'_d}{n_d} \text{ with } n'_d \approx 5 \times \# \{N_{\Omega} \circ N'_s\}$$

at the expense of solving a CS problem.
Complexity analysis

Assume discretization size in each dimension is $n$, and

$$n_s = n_t = n_f = \mathcal{O}(n)$$

Time-domain finite differences:
- $\mathcal{O}(n^4)$ in 2-D
- large constants

Preconditioned Helmholtz (Riyanti ‘06):
- $\mathcal{O}(n^5) = n_f n_s n_{it} \mathcal{O}(n^2)$ with $n_{it} = \mathcal{O}(n)$ asymptotically
- small constants

Multilevel-Krylov preconditioned (Erlangga and Nabben 08’)
- $\mathcal{O}(n^4) = n_f n_s n_{it} \mathcal{O}(n^2)$ with $n_{it} = \mathcal{O}(1)$
- small constants
Complexity analysis cont’d

Cost sparsity promoting optimization problem dominated by matrix-vector products

- 3-D wavelets are $O(n^3)$
- Gaussian projection $O(n^3)$ per frequency
- **Cost** $O(n^4)$, which does not lead to asymptotic improvement

Use fast transforms instead (e.g. Noiselets by Coifman ’01)

- fast projection in time & shot directions: $O(n \log n)$
- **Cost** $O(n^3 \log n)$ instead of $O(n^4)$

**Bottom line:** Computational cost for the $\ell_1$-solver is less ($O(n^3 \log n)$ vs. $O(n^4)$) than the cost for solving Helmholtz

- smaller memory imprint
- smaller data volume requirement
- cost reduction dependent on complexity
Freq sample 50%
Shot sample 50%
total sample 25%
Matched filter

**Simple model**

**Complex model**

![Graphs showing simple and complex models](image-url)
Recovered data

**simple model**

**complex model**

BP Solution ~3000 SPGL1 iteration
Green’s functions

simple model

complex model
Sample ratio SNR (dB)  
problem size $2^{21}$

Total computed data fraction

<table>
<thead>
<tr>
<th># Frequencies / # Shots</th>
<th>0.25</th>
<th>0.15</th>
<th>0.07</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>9.3</td>
<td>7.0</td>
<td>4.3</td>
</tr>
<tr>
<td>1</td>
<td>13.7</td>
<td>9.2</td>
<td>3.7</td>
</tr>
<tr>
<td>0.5</td>
<td>11.6</td>
<td>7.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

$$\text{SNR} = -20 \log \frac{\|d - \tilde{d}\|_2}{\|d\|_2}$$
Discussion & extensions

Compressive samplings are **cumulative**
- **more** simultaneous experiments **improve** recovery
- equivalent to longer simultaneous & continuous acquisition
and allow for design of beneficial insonifying waveforms.

Add sparsity-promoting prior to PDE constrained optimization problem:

$$\min_{\mathbf{u} \in \mathcal{U}, \mathbf{x} \in \mathcal{X}} \frac{1}{2} \| \mathbf{y} - \mathbf{D} \mathbf{U} \|_2^2 + \lambda \| \mathbf{x} \|_1 \quad \text{subject to} \quad \mathbf{L} [ \mathbf{S}^H \mathbf{x}] \mathbf{U} = \mathbf{B}$$

*Unconstrained* optimization problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \frac{1}{2} \| \mathbf{y} - \mathbf{F} [\mathbf{x}] \|_2^2 + \lambda \| \mathbf{x} \|_1 \quad \text{with} \quad \mathbf{F} [\mathbf{x}] = \mathbf{D} \mathbf{L}^{-1} [\mathbf{S}^H \mathbf{x}] \mathbf{B}$$

Requires extension of projected gradient $\ell_1$-solver to nonlinear forward map ...
Conclusions

Confluence of Compressive sensing, Simultaneous acquisition/modeling, and Helmholtz preconditioners leads to a formulation where cost to compute/acquire Green’s functions are

- no longer dependent on the problem size but on the complexity (=sparsity) of the wavefield
- computed/acquired with a gain in speed proportional to the compression rate of the wavefield & behavior CS matrix
- obtained with an overhead for the recovery problem that becomes negligible for large problem sizes.

Extends to other forward modeling operators.

Room for analyses.

Interesting
- link with simultaneous acquisition and source design
- outlook towards complexity-driven solutions to inversion problems.
Acknowledgments

E. van den Berg and M. P. Friedlander for \textit{SPGL1} (www.cs.ubc.ca/labs/scl/spgl1) & \textit{Sparco} (www.cs.ubc.ca/labs/scl/sparco)

Sergey Fomel for Madagascar (rsf.sf.net)

This work was carried out as part of the SINBAD project with financial support from the collaborative research & development (CRD) grant DNOISE (334810-05) funded by the Natural Science and Engineering Research Council (NSERC) and matching contributions from BG, BP, Chevron, ExxonMobil and Shell.

Thank you!