Full-waveform inversion from compressively recovered updates

Xiang Li and Felix J. Herrmann
Motivation

Curse of dimensionality for $d>2$

- Exponentially increasing data volumes
- Helmholtz requires implicit solvers to address bandwidth
- Computational complexity grows linearly with # RHS’s
- Makes computation of the misfit functional & gradients prohibitively expensive
Wish list

An inversion technology that

• is based on a time-harmonic PDE solver, which is easily parallelizable, and scalable to 3D

• does not require multiple iterations with all data

• removes the linearly increasing costs of implicit solvers for increasing numbers of frequencies & RHS’s

• produces high-resolution inversion results
Key technologies

Simultaneous sources & phase encoding
• supershots [Krebs et.al., ’09, Operto et. al., ’09, Herrmann et.al., ’08-10’]

Stochastic optimization & machine learning [Bertsekas, ’96]
• stochastic gradient decent

Compressive sensing [Candès et.al, Donoho, ’06]
• sparse recovery & randomized subsampling
Imaging

Least-squares migration:

\[
\delta \tilde{m} = \arg \min_{\delta m} \frac{1}{2} \| \delta d - \nabla \mathcal{F}[m_0; Q] \delta m \|^2_2
\]

\[
\delta d = \text{Multi-source multi-frequency data residue}
\]

\[
\nabla \mathcal{F}[m_0; Q] = \text{Linearized Born-scattering operator}
\]

\[
m_0 = \text{Background velocity model}
\]

\[
Q = \text{Sources}
\]

\[
\delta \tilde{m} = \text{image}
\]
Phase encoding

Simultaneous source

Randomized amplitudes along the shot line

Create *supershot* via *superposition*

Simultaneous shot at 5 Hz

Sequential-source wavefield

Simultaneous-source wavefield
Image at 5 Hz

Sequential-source image

Simultaneous-source image

[Morton, '98, Romero, '00]
Collection of \( K \) simultaneous-source experiments with batch size \( K \ll n_f \times n_s \)

\[
Q = \text{RMQ}
\]
Phase encoding

Least-squares migration:

\[ \delta \tilde{m} = \arg \min_{\delta m} \frac{1}{2} \| \delta d - \nabla F[m_0; Q] \delta m \|_2^2 \]

\( \delta d \) = Simultaneous-source data residue

\( Q \) = Simultaneous sources
Sparse recovery

Least-squares migration with sparsity promotion

$$\delta \tilde{m} = S^* \arg \min_{\delta x} \frac{1}{2} \| \delta x \|_{\ell_1} \quad \text{subject to} \quad \| \delta d - \nabla F[m_0; Q]S^* \delta x \|_2 \leq \sigma$$

$$\delta x = \text{Sparse curvelet-coefficient vector}$$

$$S^* = \text{Curvelet synthesis}$$

leads to significant speedup as long as

$$n_{PDE}^{\ell_1} \times K \ll n_{PDE}^{\ell_2} \times n_f \times n_s$$
Experiment

Linearized *sparsity promoting* least-squares migration

- Marmousi model (128x256) with grid size 15 m
- use different
  - # of simultaneous shots (50, 20, 10)
  - # of frequencies (10, 10, 5)
Initial model
Linearized sparse inversion

30 simultaneous shots 10 random frequencies

true reflectivity

sparse recovery with wavelets

Speed up: X 86
Linearized sparse inversion

20 simultaneous shots 10 random frequencies

true reflectivity

sparse recovery with wavelets

Speed up: $\times 129$
Linearized sparse inversion

10 simultaneous shots 5 random frequencies

true reflectivity

sparse recovery with wavelets

Speed up: \( \times 517 \)
## Linearized sparse inversion

<table>
<thead>
<tr>
<th>Subsample ratio</th>
<th>0.015</th>
<th>0.006</th>
<th>0.002</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n'_f/n'_s )</td>
<td>recovery error (dB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>17.44 (1.32)</td>
<td>11.66 (0.78)</td>
<td>6.83 (-0.14)</td>
</tr>
<tr>
<td>1</td>
<td>17.53 (1.59)</td>
<td>11.89 (1.05)</td>
<td>7.19 (0.15)</td>
</tr>
<tr>
<td>0.2</td>
<td>18.22 (1.68)</td>
<td>12.11 (1.32)</td>
<td>7.46 (0.27)</td>
</tr>
<tr>
<td>Speed up (×)</td>
<td>66</td>
<td>166</td>
<td>500</td>
</tr>
</tbody>
</table>

**SNRs for “migration” in parentheses**
Observations

Reconstruct model updates

- from *randomized* subsamplings
- with correct amplitudes (like Gauss-Newton updates)

Recovery quality depends on *degree of subsampling*

*Significant* speedups attainable...
FWI formulation

Multiexperiment unconstrained optimization problem:

$$\min_{m \in \mathcal{M}} \frac{1}{2} \| D - \mathcal{F}[m; Q] \|_{2,2}^2 \quad \text{with} \quad \mathcal{F}[m; Q] := PH^{-1}Q$$

- requires large number of PDE solves
- linear in the sources
- apply randomized dimensionality reduction

[Tarantola, 84; Pratt, ’98; Plessix, 06] [Haber, Chung, and Herrmann, ’10]
Algorithm 1: Gauss Newton

Result: Output estimate for the model $m$

$m \leftarrow m_0; \quad k \leftarrow 0; \quad \text{// initial model}$

while not converged do

$p^k \leftarrow \arg \min_p \frac{1}{2} \| \delta d - \nabla \mathcal{F}[m^k; Q] p \|_2^2 + \lambda^k \| p \|_2^2; \quad \text{// search dir.}$

$m^{k+1} \leftarrow m^k + \gamma^k p^k; \quad \text{// update with linesearch}$

$k \leftarrow k + 1; \quad \text{// update iteration count}$

end
FWI with phase encoding

**Multiexperiment** unconstrained optimization problem:

\[
\min_{m \in \mathcal{M}} \frac{1}{2} \|D - \mathcal{F}[m; Q]\|_2^2, \quad \text{with} \quad \mathcal{F}[m; Q] := PH^{-1}Q
\]

- requires **smaller** number of PDE solves
- exploits **linearity** in the sources & **block-diagonal** structure of the **Helmholtz system**
- uses **randomized frequency selection** and **phase encoding**

[Krebs et.al., ’09, Operto et. al., ’09 ; Herrmann et. al. ’08–’10]
Renewals

Use *different* simultaneous shots for each *subproblem*, i.e.,

\[ Q \rightarrow Q^k \]

Requires *fewer* PDE solves for each GN *subproblem*...

- motivated by *stochastic approximation* [Nemirovski, ’09]
- related to Kaczmarz (’37) method applied by Natterer, ‘01
- *supersedes ad hoc* approach by Krebs *et.al.*, 2009
Phase encoding

Algorithm 1: Gauss Newton with renewed phase encodings

Result: Output estimate for the model $m$

$m \leftarrow m_0; \ k \leftarrow 0$; \hspace{1cm} // initial model

while not converged do

\[ p^k \leftarrow \text{arg min}_p \frac{1}{2} \| \delta d^k - \nabla \mathcal{F}[m^k; Q^k] p \|^2_2 + \lambda^k \| p \|^2_2 \]; \hspace{1cm} // search dir.

\[ m^{k+1} \leftarrow m^k + \gamma^k p^k \]; \hspace{1cm} // update with linesearch

$\ k \leftarrow k + 1$;

end
Observations

Stochastic optimization

- introduces noisy search directions
- interferences go down \textit{slowly} as batch size \textit{increases}
- requires \textit{averaging} over \textit{previous} model \textit{updates}

Formulation does not exploit \textit{sparsity} on the \textit{model}

[Bertsekas, '96]
[Krebs et.al, '09]
Sparse Linearized inversion

Suggests that sparsity promotion recovers search directions accurately from randomized source encoding.
Our approach

Leverage findings from *sparse recovery & compressive sensing*

- consider each *phase-encoded* Gauss-Newton update as separate *compressive-sensing* experiment
- remove *interferences* by *curvelet-domain sparsity* promotion
- exploit properties of the Pareto curve

[Candes et al., ’06; Donoho, ’06]
[Demanet et. al. ’07; Herrmann & Li, ’08–’09]
Compressive updates

Algorithm 1: Gauss Newton with sparse updates

Result: Output estimate for the model $m$

$m \leftarrow m_0; \quad k \leftarrow 0$; \hspace{1cm} // initial model

while not converged do

\[ p^k \leftarrow S^* \arg \min_x \frac{1}{2} \| \delta d^k - \nabla F[m^k; Q^k] S^* x \|_2^2 \] s.t. $\| x \|_1 \leq \tau^k$ \hspace{1cm} // update with linesearch

$m^{k+1} \leftarrow m^k + \gamma^k p^k$;

$k \leftarrow k + 1$;

end

[van den Berg & Friedlander, ’08]
Example

Marmousi model:

- 128x384 with a mesh size of 24 meters
- 384 co-located shots and receivers with offset = 3 X depth
- 2.4s recording time

Explicit Time-harmonic Helmholtz solver

- 9-point finite difference
- Absorbing boundary condition
Example

FWI specs:

• Committed inversion crime
• Frequency continuation over 10 bands
• 15 simultaneous shots with 10 frequencies each

\[ K = 10 \times 15 \ll 100 \times 384 \]
True model
Initial model
Inverted model
True model

Lateral (x 24 meters)

Depth (x 24 meters)
Initial model
Inverted model
True model

Lateral (× 24 meters)

Depth (× 24 meters)

50 100 150 200 250 300 350

20 40 60 80 100 120 140 160 180 200

2000 2500 3000 3500 4000 4500 5000 5500
Difference

Tuesday, October 19, 2010
Performance

Remember per subproblem

\[ n_{PDE}^{\ell_1} \times K \ll n_{PDE}^{\ell_2} \times n_f \times n_s \]

\[ n_{PDE}^{\ell_1} \approx 200 \quad \text{versus} \quad n_{PDE}^{\ell_2} \approx 10 \]

\[ K = 150 \quad \text{versus} \quad K = 38400 \]

SPEEDUP of 13 X
Conclusions

Because Compressive Sensing does not rely on averaging but on sparsity, our approach is a viable alternative to the stochastic approximation.

Sparse recoveries offset random interferences due to source encoding.

High-quality & high-resolution inversions have been achieved with significant accelerations.

No need for additional migration step.

Improvements come from sparsity promotion & curvelets.

Indications that the curse of dimensionality can be removed...
Future plans

Investigate

・ *Noise sensitivity*

・ *continuation* with batch size (ref latest paper with Haber)

・ explore multiscale structure of curvelets

・ incomplete data

・ extension to 3D
Acknowledgments

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 (375142-08).

We also would like to thank the authors of CurveLab.

This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.
Thank you

slim.eos.ubc.ca
Further reading

Compressive sensing

- Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information by Candes, 06.
- Compressed Sensing by D. Donoho, '06
- Curvelets and Wave Atoms for Mirror-Extended Images by L. Demanet, L. Ying, 07.

Simultaneous acquisition

- A new look at simultaneous sources by Beasley et. al., '98.
- Changing the mindset in seismic data acquisition by Berkhout '08.

Simultaneous simulations, imaging, and full-wave inversion:

- Faster shot-record depth migrations using phase encoding by Morton & Ober; '98.
- Phase encoding of shot records in prestack migration by Romero et. al., '00.
- Efficient Seismic Forward Modeling using Simultaneous Random Sources and Sparsity by N. Neelamani et. al., '08.
- Compressive simultaneous full-waveform simulation by FJH et. al., '09.
- Fast full-wavefield seismic inversion using encoded sources by Krebs et. al., '09
- Randomized dimensionality reduction for full-waveform inversion by FJH & X. Li, '10

Stochastic optimization and machine learning:

- A Stochastic Approximation Method by Robbins and Monro, 1951
- Neuro-Dynamic Programming by Bertsekas, '96
- Robust stochastic approximation approach to stochastic programming by Nemirovski et. al., '09
- Stochastic Approximation and Recursive Algorithms and Applications by Kushner and Lin
- Stochastic Approximation approach to Stochastic Programming by Nemirovski
- An effective method for parameter estimation with PDE constraints with multiple right hand sides. by Eldad Haber, Matthias Chung, and Felix J. Herrmann. '10

Tuesday, October 19, 2010