Nonlinear seismic imaging via reduced order model backprojection

Alexander V. Mamonov\textsuperscript{1},
Vladimir Druskin\textsuperscript{2} and Mikhail Zaslavsky\textsuperscript{2}

\textsuperscript{1}University of Houston,
\textsuperscript{2}Schlumberger-Doll Research Center
Motivation: seismic oil and gas exploration

- **Seismic exploration**
- Seismic waves in the subsurface induced by sources (shots)
- Measurements of seismic signals on the surface or in a well bore
- Determine the acoustic or elastic parameters of the subsurface
Consider an acoustic wave equation in the **time domain**

\[ u_{tt} = Au \quad \text{in } \Omega, \quad t \in [0, T] \]

with initial conditions

\[ u|_{t=0} = u_0, \quad u_t|_{t=0} = 0 \]

The spatial operator \( A \in \mathbb{R}^{N \times N} \) is a fine grid discretization of

\[ A(c) = c^2 \Delta \]

with the appropriate boundary conditions

The solution is

\[ u(t) = \cos(t\sqrt{-A})u_0 \]
We stack all $p$ sources in a single tall skinny matrix $S \in \mathbb{R}^{N \times p}$ and introduce them in the initial condition

$$u|_{t=0} = S, \quad u_t|_{t=0} = 0$$

The solution matrix $u(t) \in \mathbb{R}^{N \times p}$ is

$$u(t) = \cos(t\sqrt{-A})S$$

We assume the form of the source matrix

$$S = q^2(A)CE,$$

where $E$ are $p$ point sources supported on the surface, $q^2(\omega)$ is the Fourier transform of the source wavelet and $C = \text{diag}(c)$

Here we take $q^2(\omega) = e^{\sigma\omega}$ with small $\sigma$ so that $S$ is localized near $E$, only assumes the knowledge of $c$ and thus $A$ near the surface
**Receiver and data model**

- For simplicity assume that the sources and receivers are collocated.
- Then the receiver matrix \( \mathbf{R} \in \mathbb{R}^{N \times p} \) is
  \[
  \mathbf{R} = \mathbf{C}^{-1} \mathbf{E}
  \]
- Combining the source and receiver we get the **data model**
  \[
  \mathbf{F}(t; \mathbf{c}) = \mathbf{R}^T \cos(t\sqrt{-\hat{\mathbf{A}}(\mathbf{c})})\mathbf{S},
  \]
  a \( p \times p \) matrix function of time.
- The data model can be fully symmetrized
  \[
  \mathbf{F}(t) = \hat{\mathbf{B}}^T \cos \left( t\sqrt{-\hat{\mathbf{A}}} \right) \hat{\mathbf{B}},
  \]
  with \( \hat{\mathbf{A}} = \mathbf{C} \Delta \mathbf{C} \) and \( \hat{\mathbf{B}} = q(\hat{\mathbf{A}})\mathbf{E} \)
Seismic inversion and imaging

1. **Seismic inversion**: determine $\mathbf{c}$ from the knowledge of measured data $\mathbf{F}^*(t)$ (full waveform inversion, FWI); highly nonlinear since $\mathbf{F}(\cdot;\mathbf{c})$ is nonlinear in $\mathbf{c}$
   - Conventional approach: non-linear least squares (output least squares, OLS)
     \[
     \minimize_{\mathbf{c}} \| \mathbf{F}^* - \mathbf{F}(\cdot;\mathbf{c}) \|^2_2
     \]
   - Abundant local minima
   - Slow convergence
   - Low frequency data needed

2. **Seismic imaging**: estimate $\mathbf{c}$ or its discontinuities given $\mathbf{F}(t)$ and also a smooth kinematic model $\mathbf{c}_0$
   - Conventional approach: linear migration (Kirchhoff, reverse time migration - RTM)
   - Major difficulty: multiple reflections
Reduced order models

- The data is always discretely sampled, say uniformly at $t_k = k\tau$
- The choice of $\tau$ is very important, optimally we want $\tau$ around Nyquist rate
- The discrete data samples are

$$F_k = F(k\tau) = \hat{B}^T \cos \left( k\tau \sqrt{\hat{-A}} \right) \hat{B} =$$

$$= \hat{B}^T \cos \left( k \arccos \left( \cos \tau \sqrt{\hat{-A}} \right) \right) \hat{B} = \hat{B}^T T_k(\hat{P}) \hat{B},$$

where $T_k$ is Chebyshev polynomial and the propagator is

$$\hat{P} = \cos \left( \tau \sqrt{\hat{-A}} \right)$$

- We want a reduced order model (ROM) $\tilde{P}, \tilde{B}$ that fits the measured data

$$F_k = \tilde{B}^T T_k(\tilde{P}) \tilde{B} = \tilde{B}^T T_k(\tilde{P}) \tilde{B}, \quad k = 0, \ldots, 2n - 1$$
Projection ROMs

- Projection ROMs are obtained from
  \[ \tilde{P} = V^T \hat{P} V, \quad \tilde{B} = V^T \hat{B}, \]

  where \( V \) is an orthonormal basis for some subspace

- How do we get a ROM that fits the data?
- Consider a matrix of solution snapshots
  \[ U = [\hat{u}_0, \hat{u}_1, \ldots, \hat{u}_{n-1}] \in \mathbb{R}^{N \times np}, \quad \hat{u}_k = T_k(\hat{P})\hat{B} \]

Theorem (ROM data interpolation)

If \( \text{span}(V) = \text{span}(U) \) and \( V^T V = I \) then

\[ F_k = \hat{B}^T T_k(\hat{P})\hat{B} = \tilde{B}^T T_k(\tilde{P})\tilde{B}, \quad k = 1, \ldots, 2n - 1, \]

where \( \tilde{P} = V^T \hat{P} V \in \mathbb{R}^{np \times np} \) and \( \tilde{B} = V^T \hat{B} \in \mathbb{R}^{np \times p} \).
Obtaining the ROM from the data

- We do not know the solutions in the whole domain $U$ and thus $V$ is unknown.
- How do we obtain the ROM from just the data $F_k$?
- The data does not give us $U$, but it gives us the inner products!
- A basic property of Chebyshev polynomials is

$$T_i(x)T_j(x) = \frac{1}{2}(T_{i+j}(x) + T_{|i-j|}(x))$$

- Then we can obtain

$$(U^TU)_{i,j} = u_i^Tu_j = \frac{1}{2}(F_{i+j} + F_{i-j}),$$

$$(U^T\hat{P}U)_{i,j} = u_i^T\hat{P}u_j = \frac{1}{4}(F_{j+i+1} + F_{j-i+1} + F_{j+i-1} + F_{j-i-1})$$
Obtaining the ROM from the data

Suppose $\mathbf{U}$ is orthogonalized by a block QR procedure

$$\mathbf{U} = \mathbf{V} \mathbf{L}^T,$$

so $\mathbf{V} = \mathbf{U} \mathbf{L}^{-T}$, where $\mathbf{L}$ is a block Cholesky factor of the Gramian $\mathbf{U}^T \mathbf{U}$ known from the data

$$\mathbf{U}^T \mathbf{U} = \mathbf{L} \mathbf{L}^T$$

The projection is given by

$$\tilde{\mathbf{P}} = \mathbf{V}^T \hat{\mathbf{P}} \mathbf{V} = \mathbf{L}^{-1} \left( \mathbf{U}^T \hat{\mathbf{P}} \mathbf{U} \right) \mathbf{L}^{-T},$$

where $\mathbf{U}^T \hat{\mathbf{P}} \mathbf{U}$ is also known from the data

The use of Cholesky for orthogonalization is essential, (block) lower triangular structure is the linear algebraic equivalent of causality.
Use of ROMs

- Once we have the ROM $\tilde{P} = V^T \tilde{P} V$, $\tilde{B} = V^T \tilde{B}$ how do we estimate $c$ from it?

- The ROM for the operator $A$ itself is

$$\tilde{A} = \frac{2}{\tau^2} (\tilde{P} - I)$$

from truncated Taylor’s expansion.

- **Inversion**: transform $\tilde{A}$ to a **block finite difference** (bFD) scheme, use the bFD coefficients in optimization.

- **Imaging**: Using a smooth kinematic model $c_0$ **backproject** $\tilde{A}$ to get the coefficient $c$ directly.
Recall the conventional FWI (OLS)

\[ \min_c \| F^* - F(\cdot; c) \|_2^2 \]

Replace the objective with a “nonlinearly preconditioned” functional

\[ \min_c \| \tilde{A}^* - \tilde{A}(c) \|_F^2, \]

where \( \tilde{A}^* \) is computed from the data \( F^* \) and \( \tilde{A}(c) \) is a (highly) nonlinear mapping

\[ \tilde{A} : c \rightarrow A(c) \rightarrow U \rightarrow V \rightarrow P \rightarrow \tilde{A} \]

Why does this have a preconditioning effect?
Advantages of ROM-preconditioned optimization

The biggest issue of conventional OLS FWI is the abundance of **local minima** (cycle skipping)

The dependency of $A(c) = c^2 \Delta$ on $c^2$ is linear

In a certain parametrization the dependency of $\tilde{A}$ on $c^2$ should be close to linear

The preconditioned objective functional is close to quadratic, thus close to **convex**

Approximate convexity leads to faster, more robust convergence

Implicit orthogonalization of solution snapshots $V = UL^{-T}$ removes the multiple reflections
Conventional vs. preconditioned in 1D

Conventional
CG iteration 1, $E_r = 0.137937$

Preconditioned
CG iteration 1, $E_r = 0.080594$

Faster convergence.
Conventional vs. preconditioned in 1D

Conventional
CG iteration 5, $E_r = 0.108350$

Preconditioned
CG iteration 5, $E_r = 0.010831$

Faster convergence.
Conventional vs. preconditioned in 1D

Faster convergence.
Conventional vs. preconditioned in 1D

**Conventional**

CG iteration 15, $E_r = 0.070725$

**Preconditioned**

CG iteration 15, $E_r = 0.002226$

Faster convergence.
Conventional vs. preconditioned in 1D

**Conventional**

CG iteration 1, $E_r = 0.278869$

**Preconditioned**

CG iteration 1, $E_r = 0.272127$

Automatic removal of multiple reflections.
Conventional vs. preconditioned in 1D

Conventional
CG iteration 5, $E_r = 0.265722$

Preconditioned
CG iteration 5, $E_r = 0.197026$

Automatic removal of multiple reflections.
Conventional vs. preconditioned in 1D

Conventional
CG iteration 10, $E_r = 0.273922$

Preconditioned
CG iteration 10, $E_r = 0.157774$

Automatic removal of multiple reflections.
Conventional vs. preconditioned in 1D

Conventional
CG iteration 15, $E_r = 0.268569$

Preconditioned
CG iteration 15, $E_r = 0.138945$

Automatic removal of multiple reflections.
Conventional vs. preconditioned in 1D

Avoiding the cycle skipping.

Conventional
CG iteration 1, $E_r = 0.173770$

Preconditioned
CG iteration 1, $E_r = 0.147049$
Conventional vs. preconditioned in 1D

Conventional
CG iteration 5, $E_r = 0.174695$

Preconditioned
CG iteration 5, $E_r = 0.105966$

Avoiding the cycle skipping.
Conventional vs. preconditioned in 1D

Avoiding the cycle skipping.
Conventional vs. preconditioned in 1D

Conventional
CG iteration 15, $E_r = 0.174689$

Preconditioned
CG iteration 15, $E_r = 0.086519$

Avoiding the cycle skipping.
The ROM for \( \tilde{A} \) approximately satisfies\[ \tilde{A} \approx V^T \tilde{A} V \]

If the subspace spanned by \( V \) is sufficiently rich, then\[ VV^T \approx I, \]
so we can **backproject** the ROM to the fine grid space\[ \hat{A} \approx V \tilde{A} V^T \approx VV^T \hat{A} VV^T \]

**Problem**: we do not know \( V \), since the snapshots \( U \) are unknown to us in the whole domain

Known smooth **kinematic model** \( c_0 \) is needed

From \( c_0 \) we can explicitly compute everything: \( \hat{A}_0, \tilde{A}, U_0 \) and, most important, \( V_0 \)

Replace the unknown true \( V \) by known \( V_0 \)\[ \hat{A} \approx V_0 \tilde{A} V_0^T \]
We do not need the whole operator $A$ or $\hat{A}$, just the fine grid coefficient $c^2$.

Recall that $\hat{A} = C\Delta C$, thus

$$c^2 \propto \text{diag}(\hat{A})$$

Similarly for the difference we have

$$\delta c^2 = c^2 - c_0^2 \propto \text{diag}(\hat{A} - \hat{A}_0)$$

Approximate $\hat{A}$ and $\hat{A}_0$ by their backprojections to obtain an imaging relation

$$\delta c^2 \propto \text{diag} \left( V_0(\tilde{A} - \tilde{A}_0)V_0^T \right)$$

Choosing different proportionality factors leads to various imaging formulae, for example a multiplicative

$$c^* = c_0 \sqrt{1 + \alpha \delta c^2}$$
Backprojection imaging: features

- Conventional imaging techniques (Kirchhoff, RTM) are **linear** in the data.
- Our approach is **non-linear** because of implicit orthogonalization.

\[
\tilde{P} = L^{-1} \left( U^T \hat{P} U \right) L^{-T}, \quad U^T U = LL^T
\]

- Block Cholesky: **causal orthogonalization**, removes the “tail”, only the wavefront survives.
- Thus, **multiple reflection** artifacts are removed.
- We image correctly not only the locations of reflectors, but also their strength: **true amplitude imaging**.
- Computationally cheap: we need a forward solution (same as RTM) and an extra orthogonalization step.
Removal of multiple reflection artifacts

True sound speed $c$

- A simple layered model, $p = 12$ sources/receivers (black $\times$)
- Multiple reflections from waves bouncing between layers and surface
- Each multiple creates an RTM artifact below actual layers

RTM image

Backprojection image
Solution snapshot orthogonalization

Solution snapshots $\mathbf{U}$

Orthogonalized basis $\mathbf{V}$

- A 1D analogue of the previous example
- Strong primaries/multiples in $\mathbf{U}$, almost none in $\mathbf{V}$
- The operator $\hat{\mathbf{A}}$ is probed with $\mathbf{V}$ that is mostly a single propagating wavefront

Mamonov, Druskin, Zaslavsky
Backprojection imaging
High contrast imaging: hydraulic fractures

- Important application: seismic monitoring of hydraulic fracturing
- Multiple thin fractures (down to 1 cm in width, here 10 cm)
- Very high contrasts: $c = 4500 \text{ m/s}$ in the surrounding rock, $c = 1500 \text{ m/s}$ in the fluid inside fractures
High contrast imaging: hydraulic fractures

- Important application: seismic monitoring of hydraulic fracturing
- Multiple thin fractures (down to 1 cm in width, here 10 cm)
- Very high contrasts: $c = 4500 m/s$ in the surrounding rock, $c = 1500 m/s$ in the fluid inside fractures
Numerical example: Marmousi model

- Classical Marmousi model, $13.5\, km \times 2.7\, km$
- Forward problem is discretized on a $15\, m$ grid with $N = 900 \times 180 = 162,000$ nodes
- Kinematic model $c_0$: smoothed out true $c$ ($465\, m$ horizontally, $315\, m$ vertically)
- Time domain data sample rate $\tau = 33.5\, ms$, source frequency about $15\, Hz$, $n = 35$ data samples measured
- Number of sources/receivers $p = 90$ uniformly distributed with spacing $150\, m$
- Data is split into 17 overlapping windows of 10 sources/receivers each ($1.5\, km$ max offset)
- Reflecting boundary conditions
- No data filtering, everything used as is (surface wave, reflections from the boundaries, multiples)
Backprojection imaging: Marmousi model
Backprojection imaging: Marmousi model
Marmousi backprojection image: well log
Marmousi backprojection image: well log

Mamonov, Druskin, Zaslavsky
Marmousi backprojection image: well log

Mamonov, Druskin, Zaslavsky

Backprojection imaging
Marmousi backprojection image: well log
Marmousi backprojection image: well log
Conclusions and future work

- Novel approach to seismic imaging using reduced order models
- Time domain formulation is essential, makes use of causality (linear algebraic analogue - Cholesky decomposition)
- Nonlinear construction of ROM via implicit causal orthogonalization of solution snapshots
- Strong suppression of multiple reflection artifacts

Future work:

- Non-symmetric setting (non-collocated sources/receivers)
- Full waveform inversion in higher dimensions
- Better theoretical understanding

References: