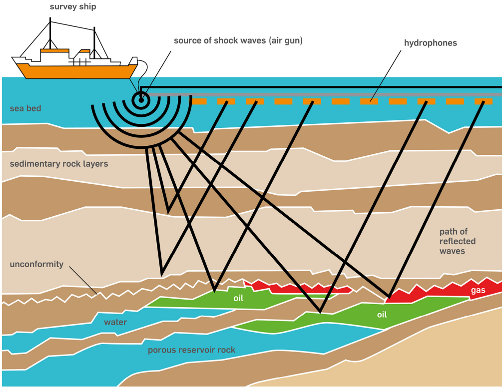
Nonlinear seismic imaging via reduced order model backprojection

Alexander V. Mamonov[1],
Vladimir Druskin[2] and Mikhail Zaslavsky[2]

[1]University of Houston,
[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

Acoustic wave equation

- Consider an acoustic wave equation in the **time domain**

$$u_{tt} = \mathbf{A}u \quad \text{in } \Omega, \quad t \in [0, T]$$

  with initial conditions

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

- The spatial operator $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a fine grid discretization of

$$A(c) = c^2 \Delta$$

  with the appropriate boundary conditions
- The solution is

$$u(t) = \cos(t\sqrt{-\mathbf{A}})u_0$$

Source model

- We stack all p sources in a single tall skinny matrix $\mathbf{S} \in \mathbb{R}^{N \times p}$ and introduce them in the initial condition

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

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

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

- We assume the form of the source matrix

$$\mathbf{S} = q^2(\mathbf{A})\mathbf{C}\mathbf{E},$$

  where $\mathbf{E}$ are p point sources supported on the surface, $q^2(\omega)$ is the Fourier transform of the source wavelet and $\mathbf{C} = \text{diag}(\mathbf{c})$
- Here we take $q^2(\omega) = e^{\sigma\omega}$ with small σ so that $\mathbf{S}$ is localized near $\mathbf{E}$, only assumes the knowledge of $\mathbf{c}$ and thus $\mathbf{A}$ 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**

$$\mathbf{F}(t; \mathbf{c}) = \mathbf{R}^T \cos(t\sqrt{-\mathbf{A}(\mathbf{c})})\mathbf{S},$$

  a $p \times p$ matrix function of time
- The data model can be fully symmetrized

$$\mathbf{F}(t) = \widehat{\mathbf{B}}^T \cos\left(t\sqrt{-\widehat{\mathbf{A}}}\right)\widehat{\mathbf{B}},$$

  with $\widehat{\mathbf{A}} = \mathbf{C}\mathbf{\Delta}\mathbf{C}$ and $\widehat{\mathbf{B}} = q(\widehat{\mathbf{A}})\mathbf{E}$

Seismic inversion and imaging

1. **Seismic inversion**: determine $\mathbf{c}$ from the knowledge of measured data $\mathbf{F}^\star(t)$ (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)
   $$\underset{\mathbf{c}}{\text{minimize}}\,\|\mathbf{F}^\star - \mathbf{F}(\,\cdot\,;\mathbf{c})\|_2^2$$
   - 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 τ is very important, optimally we want τ around Nyquist rate
- The discrete data samples are

$$\mathbf{F}_k = \mathbf{F}(k\tau) = \widehat{\mathbf{B}}^T \cos\left(k\tau\sqrt{-\widehat{\mathbf{A}}}\right)\widehat{\mathbf{B}} =$$

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

  where T_k is Chebyshev polynomial and the **propagator** is

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

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

$$\mathbf{F}_k = \widehat{\mathbf{B}}^T T_k(\widehat{\mathbf{P}})\widehat{\mathbf{B}} = \widetilde{\mathbf{B}}^T T_k(\widetilde{\mathbf{P}})\widetilde{\mathbf{B}}, \quad k = 0, \ldots, 2n-1$$

Projection ROMs

- Projection ROMs are obtained from

$$\widetilde{\mathbf{P}} = \mathbf{V}^T\widehat{\mathbf{P}}\mathbf{V}, \quad \widetilde{\mathbf{B}} = \mathbf{V}^T\widehat{\mathbf{B}},$$

  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**

$$\mathbf{U} = [\widehat{\mathbf{u}}_0, \widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_{n-1}] \in \mathbb{R}^{N\times np}, \quad \widehat{\mathbf{u}}_k = T_k(\widehat{\mathbf{P}})\widehat{\mathbf{B}}$$

Theorem (ROM data interpolation)

If $span(\mathbf{V}) = span(\mathbf{U})$ *and* $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ *then*

$$\mathbf{F}_k = \widehat{\mathbf{B}}^T T_k(\widehat{\mathbf{P}})\widehat{\mathbf{B}} = \widetilde{\mathbf{B}}^T T_k(\widetilde{\mathbf{P}})\widetilde{\mathbf{B}}, \quad k = 1, \ldots, 2n-1,$$

where $\widetilde{\mathbf{P}} = \mathbf{V}^T\widehat{\mathbf{P}}\mathbf{V} \in \mathbb{R}^{np\times np}$ *and* $\widetilde{\mathbf{B}} = \mathbf{V}^T\widehat{\mathbf{B}} \in \mathbb{R}^{np\times p}$.

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 $\mathbf{F}_k$?
- 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

$$(\mathbf{U}^T\mathbf{U})_{i,j} = \mathbf{u}_i^T\mathbf{u}_j = \frac{1}{2}(\mathbf{F}_{i+j} + \mathbf{F}_{i-j}),$$
$$(\mathbf{U}^T\widehat{\mathbf{P}}\mathbf{U})_{i,j} = \mathbf{u}_i^T\widehat{\mathbf{P}}\mathbf{u}_j = \frac{1}{4}(\mathbf{F}_{j+i+1} + \mathbf{F}_{j-i+1} + \mathbf{F}_{j+i-1} + \mathbf{F}_{j-i-1})$$

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

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

where $\mathbf{U}^T\widehat{\mathbf{P}}\mathbf{U}$ 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 $\widetilde{\mathbf{P}} = \mathbf{V}^T\widehat{\mathbf{P}}\mathbf{V}$, $\widetilde{\mathbf{B}} = \mathbf{V}^T\widehat{\mathbf{B}}$ how do we estimate $\mathbf{c}$ from it?

- The ROM for the operator $\mathbf{A}$ itself is

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

  from truncated Taylor's expansion

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

- **Imaging**: Using a smooth kinematic model $\mathbf{c}_0$ **backproject** $\widetilde{\mathbf{A}}$ to get the coefficient $\mathbf{c}$ directly

Seismic inversion: optimization preconditioning

- Recall the conventional FWI (OLS)

$$\underset{\mathbf{c}}{\text{minimize}} \, \|\mathbf{F}^\star - \mathbf{F}(\,\cdot\,;\mathbf{c})\|_2^2$$

- Replace the objective with a **"nonlinearly preconditioned"** functional

$$\underset{\mathbf{c}}{\text{minimize}} \, \|\widetilde{\mathbf{A}}^\star - \widetilde{\mathbf{A}}(\mathbf{c})\|_F^2,$$

where $\widetilde{\mathbf{A}}^\star$ is computed from the data $\mathbf{F}^\star$ and $\widetilde{\mathbf{A}}(\mathbf{c})$ is a (highly) nonlinear mapping

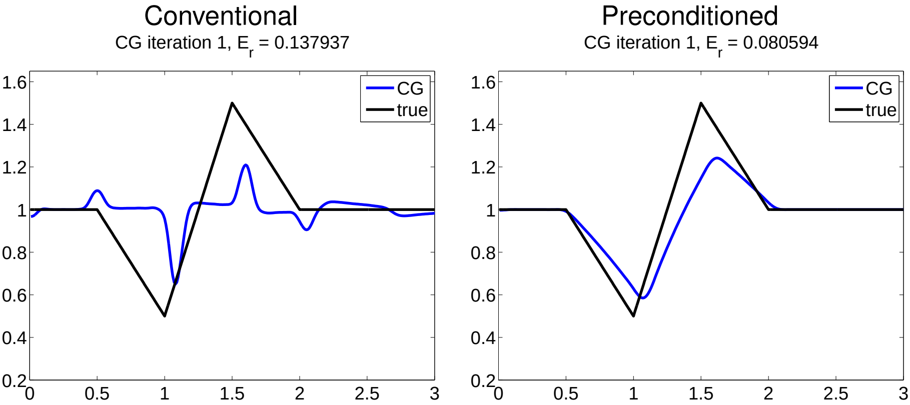
$$\widetilde{\mathbf{A}} : \mathbf{c} \to \mathbf{A}(\mathbf{c}) \to \mathbf{U} \to \mathbf{V} \to \widetilde{\mathbf{P}} \to \widetilde{\mathbf{A}}$$

- Why does this have a preconditioning effect?

Advantages of ROM-preconditioned optimization

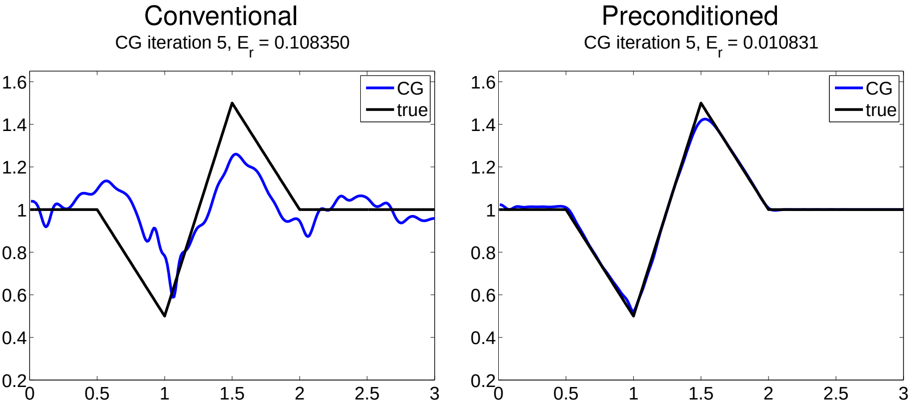
- The biggest issue of conventional OLS FWI is the abundance of **local minima** (cycle skipping)
- The dependency of $\mathbf{A}(\mathbf{c}) = \mathbf{c}^2\boldsymbol{\Delta}$ on $\mathbf{c}^2$ is linear
- In a certain parametrization the dependency of $\widetilde{\mathbf{A}}$ on $\mathbf{c}^2$ should 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 $\mathbf{V} = \mathbf{U}\mathbf{L}^{-T}$ removes the multiple reflections

Conventional vs. preconditioned in 1D



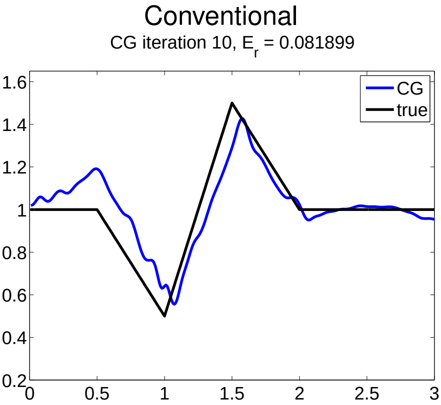
Faster convergence.

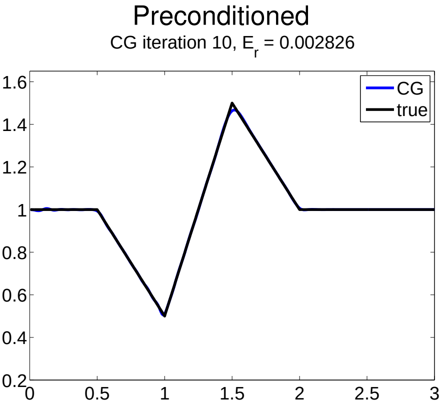
Conventional vs. preconditioned in 1D



Faster convergence.

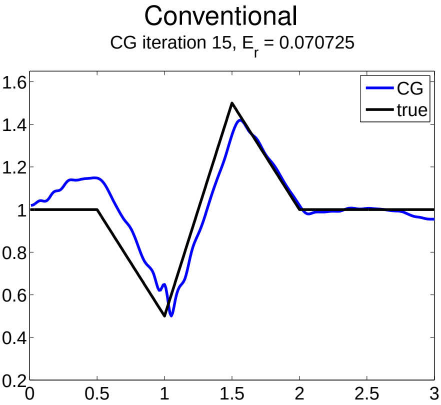
Conventional vs. preconditioned in 1D

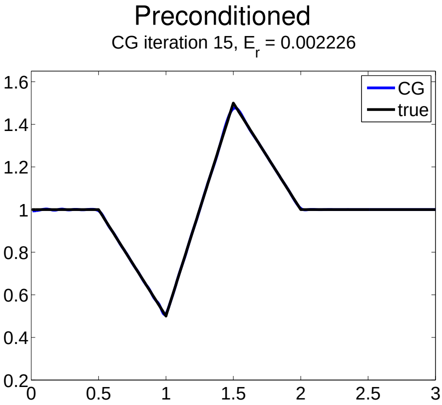




Faster convergence.

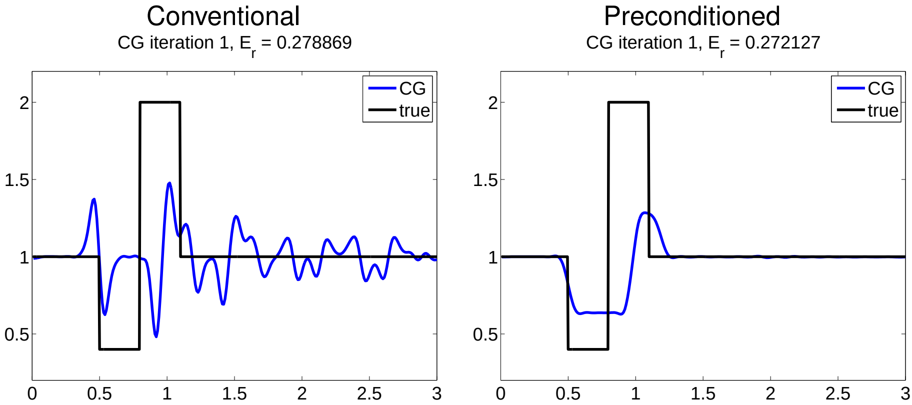
Conventional vs. preconditioned in 1D





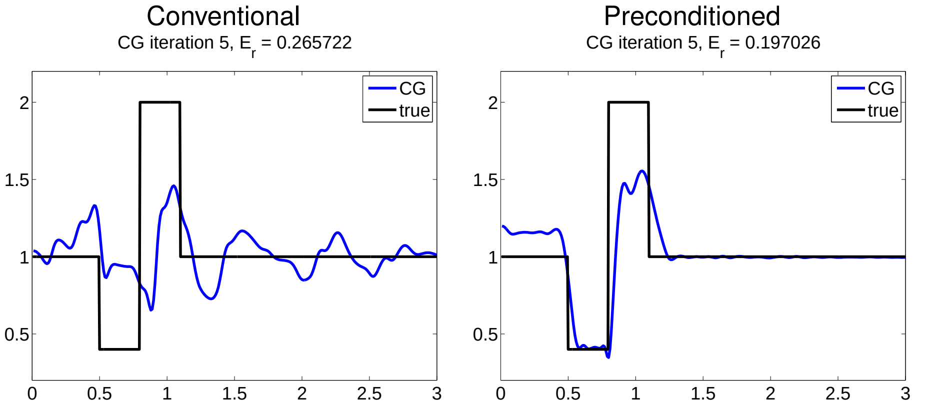
Faster convergence.

Conventional vs. preconditioned in 1D



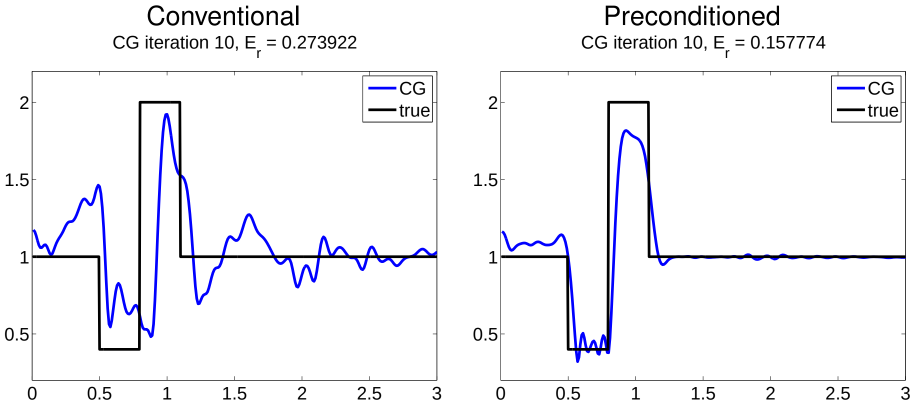
Automatic removal of multiple reflections.

Conventional vs. preconditioned in 1D



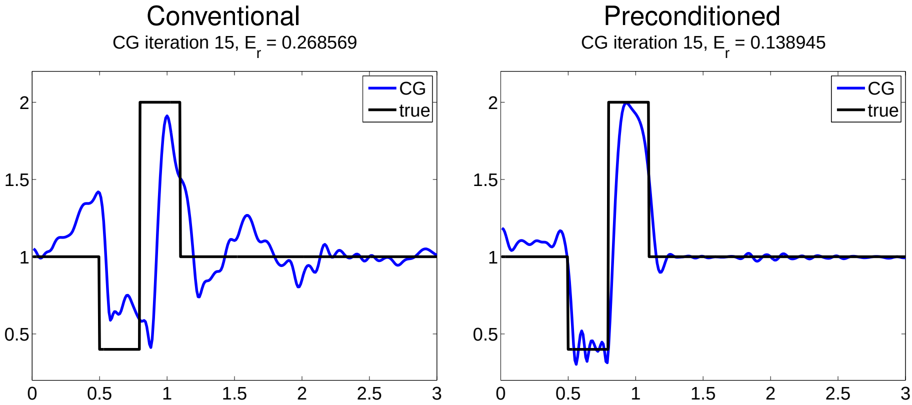
Automatic removal of multiple reflections.

Conventional vs. preconditioned in 1D



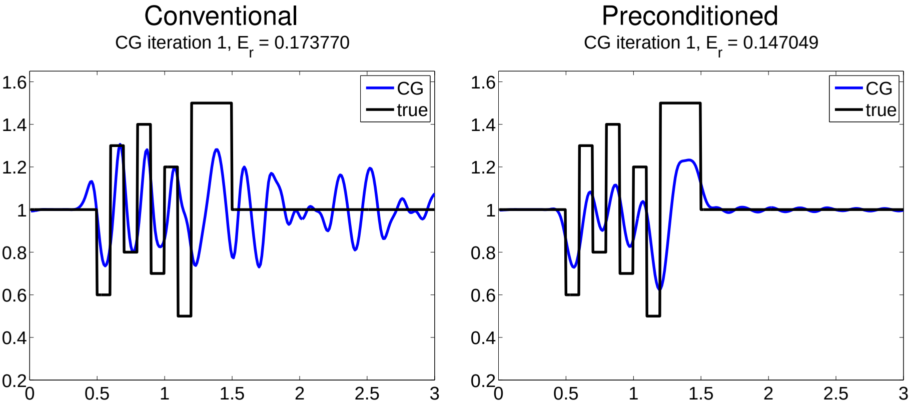
Automatic removal of multiple reflections.

Conventional vs. preconditioned in 1D



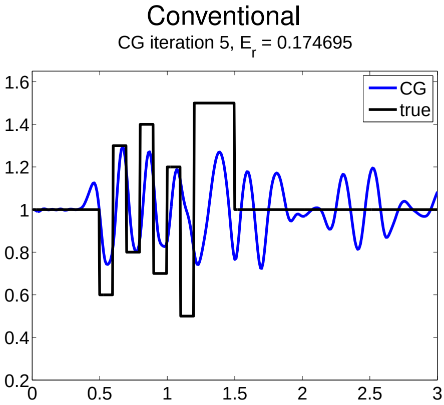
Automatic removal of multiple reflections.

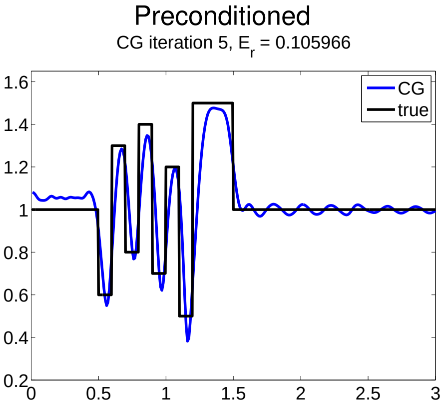
Conventional vs. preconditioned in 1D



Avoiding the cycle skipping.

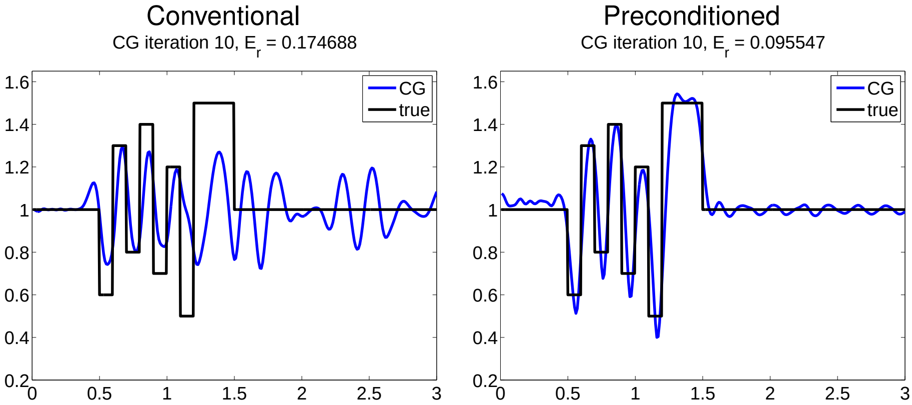
Conventional vs. preconditioned in 1D





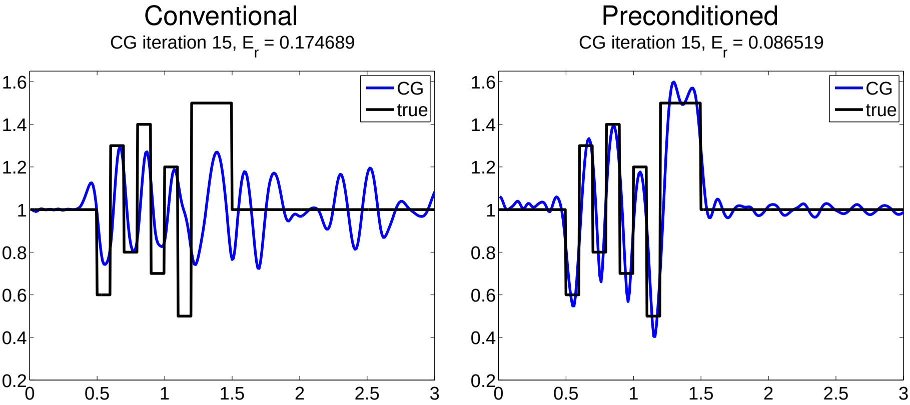
Avoiding the cycle skipping.

Conventional vs. preconditioned in 1D



Avoiding the cycle skipping.

Conventional vs. preconditioned in 1D



Avoiding the cycle skipping.

Imaging: backprojection

- The ROM for $\widetilde{\mathbf{A}}$ approximately satisfies
$$\widetilde{\mathbf{A}} \approx \mathbf{V}^T \widehat{\mathbf{A}} \mathbf{V}$$
- If the subspace spanned by $\mathbf{V}$ is sufficiently rich, then
$$\mathbf{V}\mathbf{V}^T \approx \mathbf{I},$$
so we can **backproject** the ROM to the fine grid space
$$\widehat{\mathbf{A}} \approx \mathbf{V}\widetilde{\mathbf{A}}\mathbf{V}^T \approx \mathbf{V}\mathbf{V}^T \widehat{\mathbf{A}} \mathbf{V}\mathbf{V}^T$$
- **Problem**: we do not know $\mathbf{V}$, since the snapshots $\mathbf{U}$ are unknown to us in the whole domain
- Known smooth **kinematic model** $\mathbf{c}_0$ is needed
- From c_0 we can explicitly compute everything: $\widehat{\mathbf{A}}_0$, $\widetilde{\mathbf{A}}$, $\mathbf{U}_0$ and, most important, $\mathbf{V}_0$
- Replace the unknown true $\mathbf{V}$ by known $\mathbf{V}_0$
$$\widehat{\mathbf{A}} \approx \mathbf{V}_0 \widetilde{\mathbf{A}} \mathbf{V}_0^T$$

Backprojection: extracting the PDE coefficient

- We do not need the whole operator $\mathbf{A}$ or $\widehat{\mathbf{A}}$, just the fine grid coefficient $\mathbf{c}^2$
- Recall that $\widehat{\mathbf{A}} = \mathbf{C}\boldsymbol{\Delta}\mathbf{C}$, thus

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

- Similarly for the difference we have

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

- Approximate $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{A}}_0$ by their backprojections to obtain an **imaging relation**

$$\delta\mathbf{c}^2 \propto \text{diag}\left(\mathbf{V}_0(\widetilde{\mathbf{A}} - \widetilde{\mathbf{A}}_0)\mathbf{V}_0^T\right)$$

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

$$\mathbf{c}^\star = \mathbf{c}_0\sqrt{1 + \alpha\delta\mathbf{c}^2}$$

Backprojection imaging: features

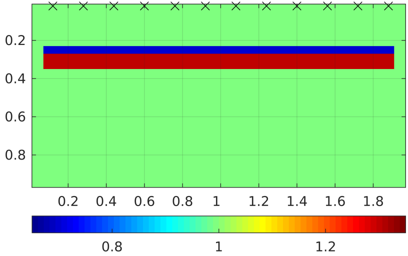
- Conventional imaging techniques (Kirchhoff, RTM) are **linear** in the data
- Our approach is **non-linear** because of **implicit orthogonalization**

$$\widetilde{\mathbf{P}} = \mathbf{L}^{-1} \left(\mathbf{U}^T \widehat{\mathbf{P}} \mathbf{U} \right) \mathbf{L}^{-T}, \quad \mathbf{U}^T \mathbf{U} = \mathbf{L} \mathbf{L}^T$$

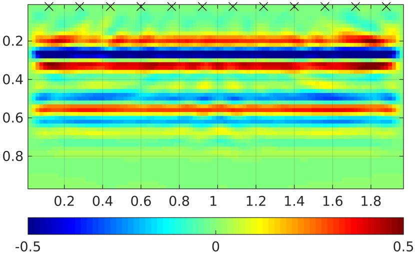
- Block Cholesky: **causal orthogonalization**, removes the “tail”, only the wavefront survives
- Thus, **multiple reflection** artifacts are removed
- We image correctly not only the locations of reflectors, but also their strength: **true amplitude imaging**
- Computationally cheap: we need a forward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

Removal of multiple reflection artifacts

True sound speed **c**

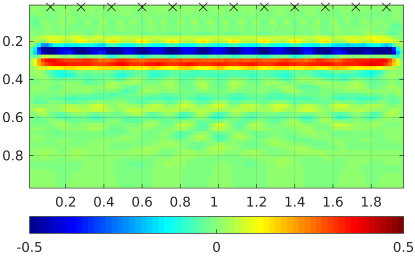


RTM image



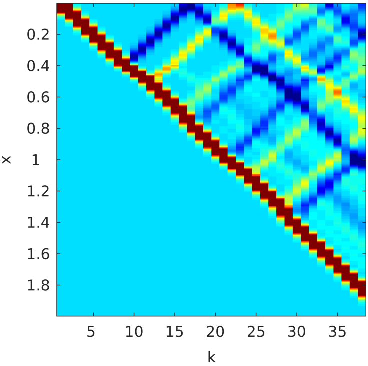
- A simple layered model, $p = 12$ sources/receivers (black $\times$)
- Multiple reflections from waves bouncing between layers and surface
- Each multiple creates an RTM artifact below actual layers

Backprojection image

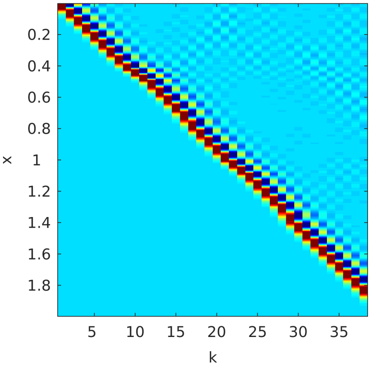


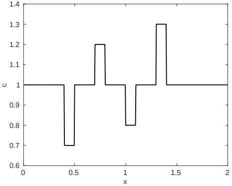
Solution snapshot orthogonalization

Solution snapshots **U**



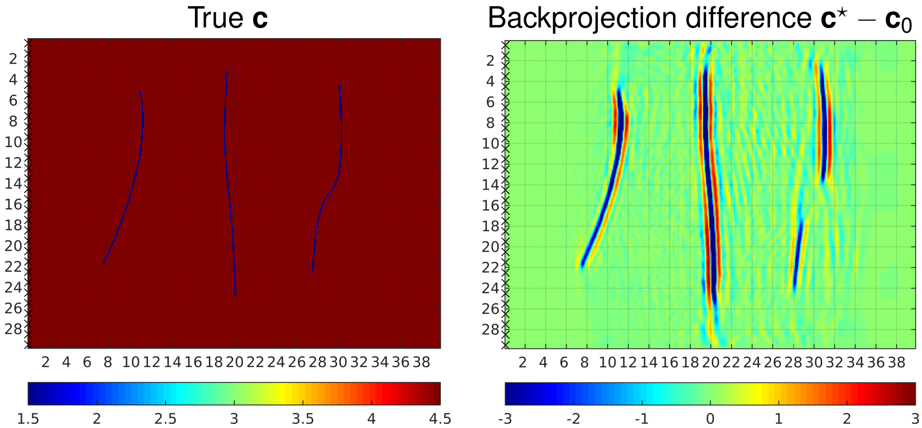
Orthogonalized basis **V**





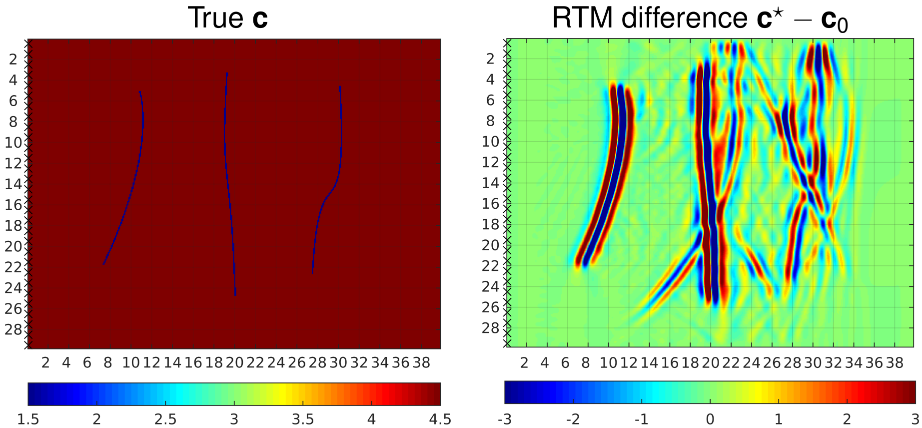
- A 1D analogue of the previous example
- Strong primaries/multiples in **U**, almost none in **V**
- The operator $\widehat{\mathbf{A}}$ is probed with **V** that is mostly a single propagating wavefront

High contrast imaging: hydraulic fractures



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

High contrast imaging: hydraulic fractures

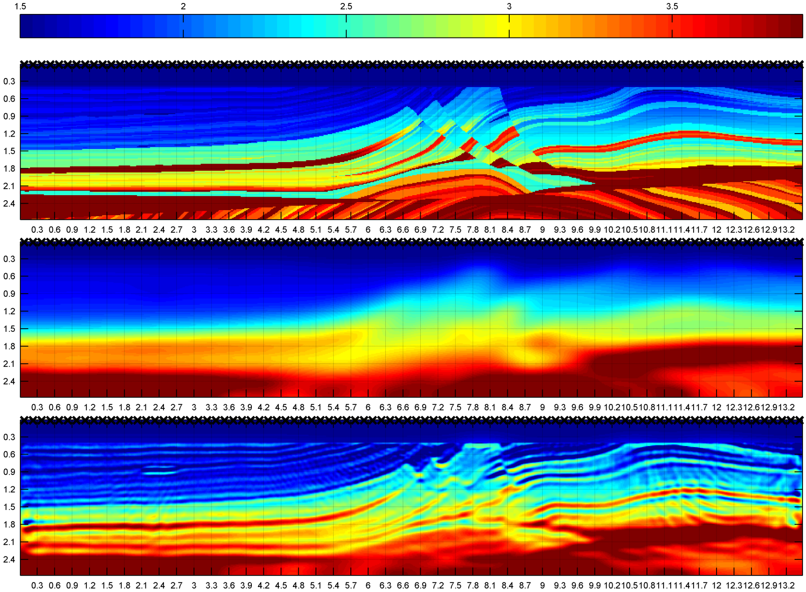


- Important application: seismic monitoring of hydraulic fracturing
- Multiple thin fractures (down to $1cm$ in width, here $10cm$)
- Very high contrasts: $c = 4500m/s$ in the surrounding rock, $c = 1500m/s$ in the fluid inside fractures

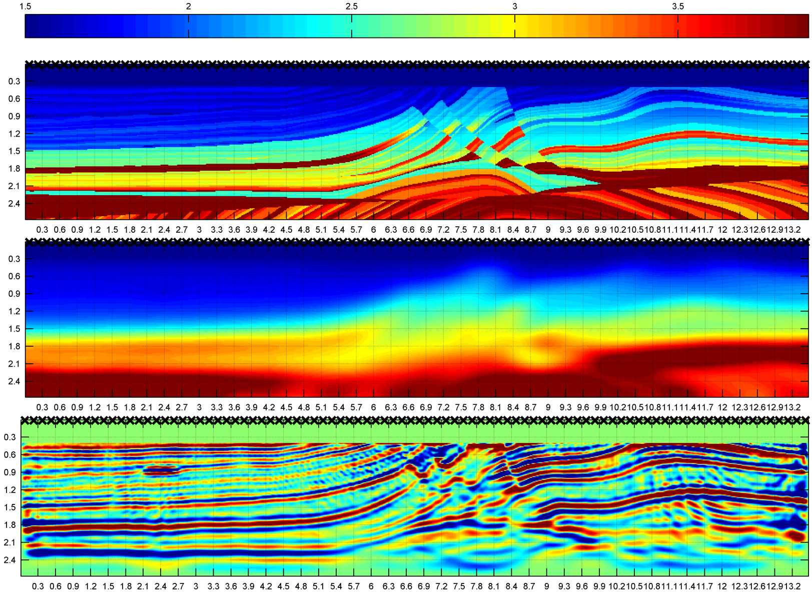
Numerical example: Marmousi model

- Classical Marmousi model, $13.5km \times 2.7km$
- Forward problem is discretized on a $15m$ grid with $N = 900 \times 180 = 162,000$ nodes
- Kinematic model $\mathbf{c}_0$: smoothed out true $\mathbf{c}$ ($465m$ horizontally, $315m$ vertically)
- Time domain data sample rate $\tau = 33.5ms$, source frequency about $15Hz$, $n = 35$ data samples measured
- Number of sources/receivers $p = 90$ uniformly distributed with spacing $150m$
- Data is split into 17 overlapping windows of 10 sources/receivers each ($1.5km$ max offset)
- Reflecting boundary conditions
- No data filtering, everything used as is (surface wave, reflections from the boundaries, multiples)

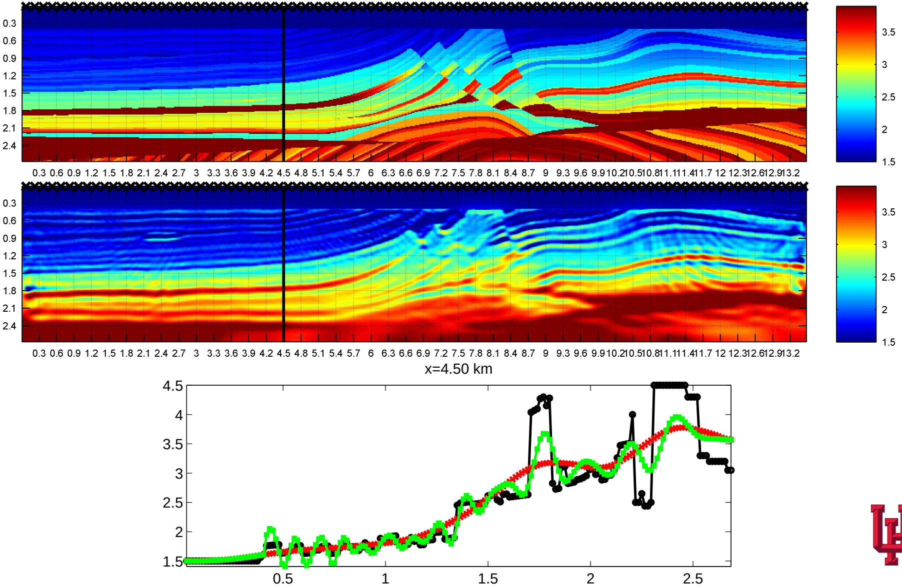
Backprojection imaging: Marmousi model



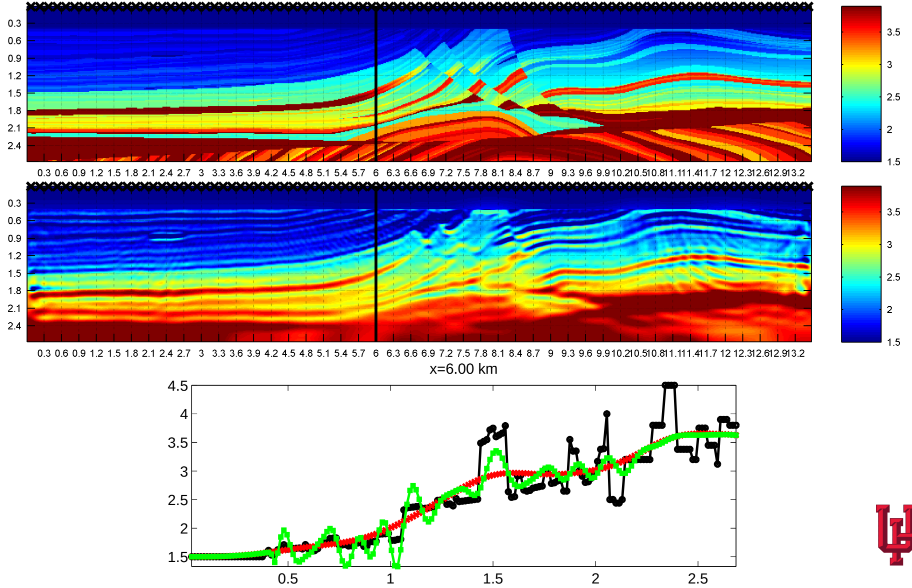
Backprojection imaging: Marmousi model



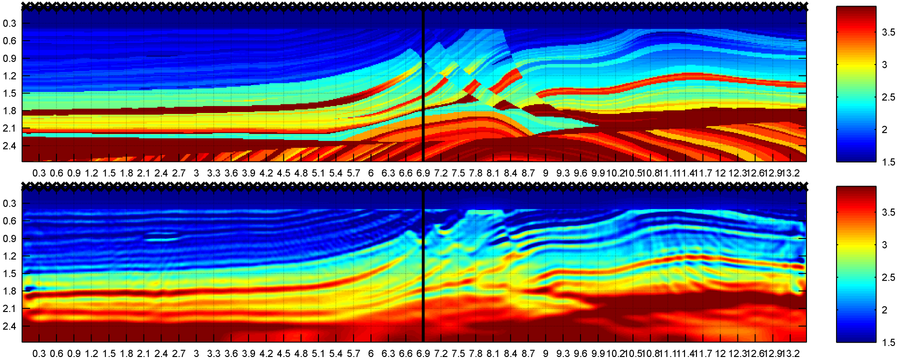
Marmousi backprojection image: well log

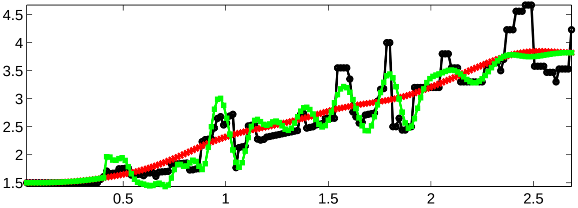


Marmousi backprojection image: well log

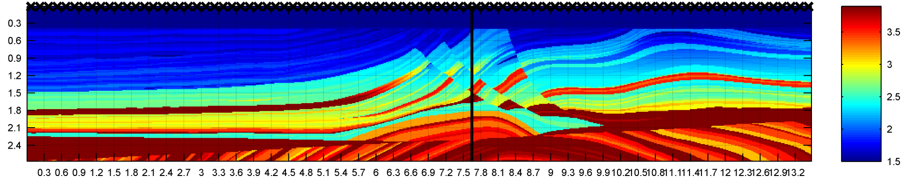


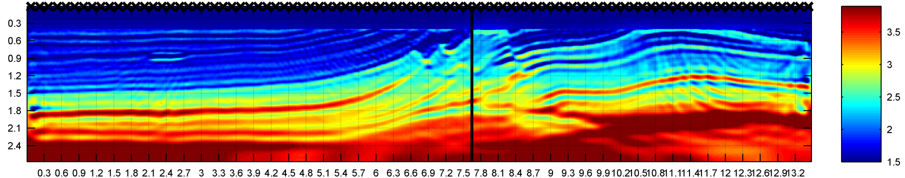
Marmousi backprojection image: well log

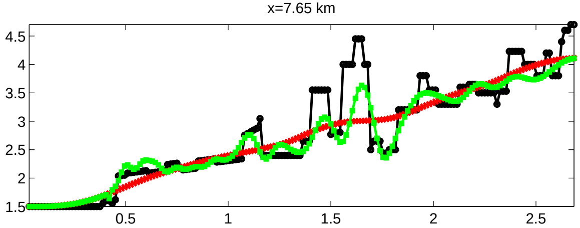




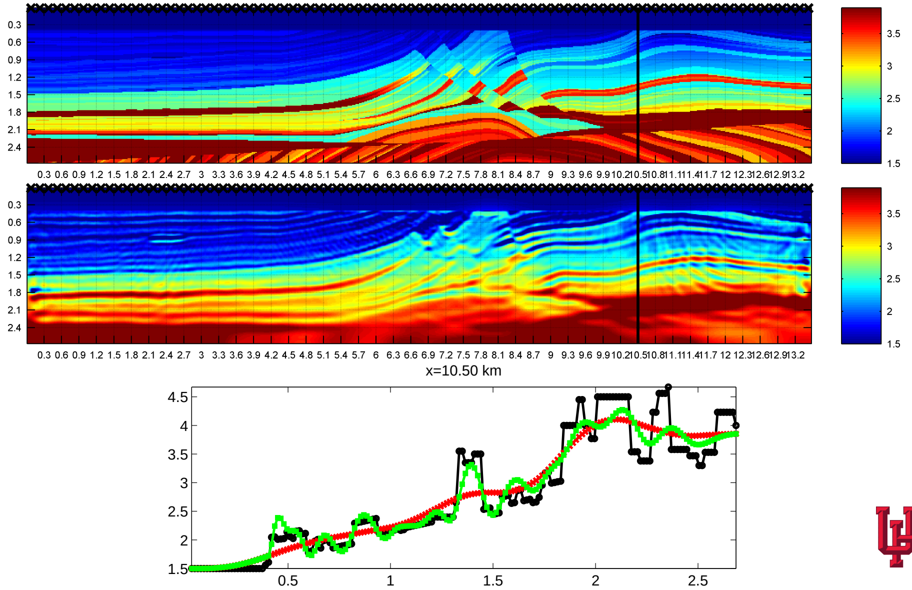
Marmousi backprojection image: well log



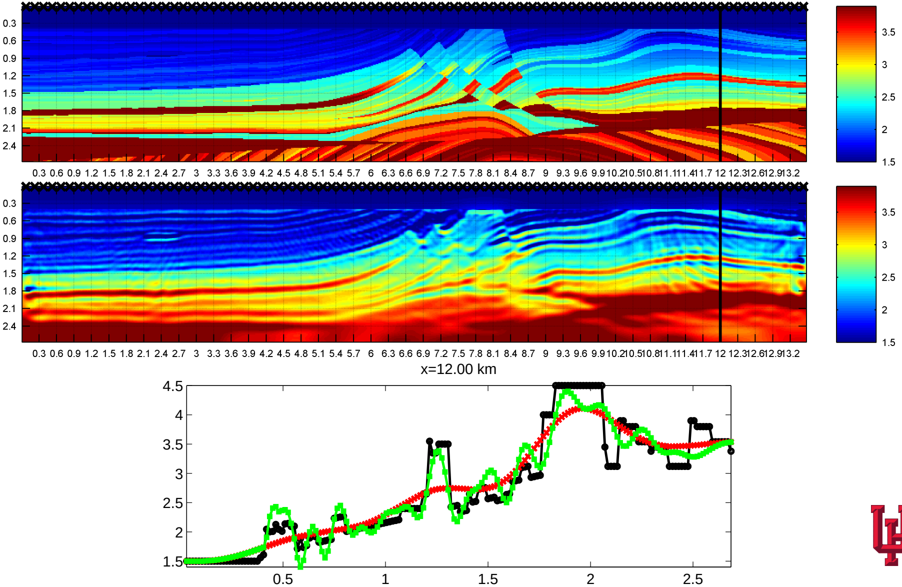




Marmousi backprojection image: well log



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:

[1] A.V. Mamonov, V. Druskin, M. Zaslavsky, *Nonlinear seismic imaging via reduced order model backprojection*, SEG Technical Program Expanded Abstracts 2015: pp. 4375–4379.

[2] V. Druskin, A. Mamonov, A.E. Thaler and M. Zaslavsky, *Direct, nonlinear inversion algorithm for hyperbolic problems via projection-based model reduction*. arXiv:1509.06603 [math.NA], 2015.