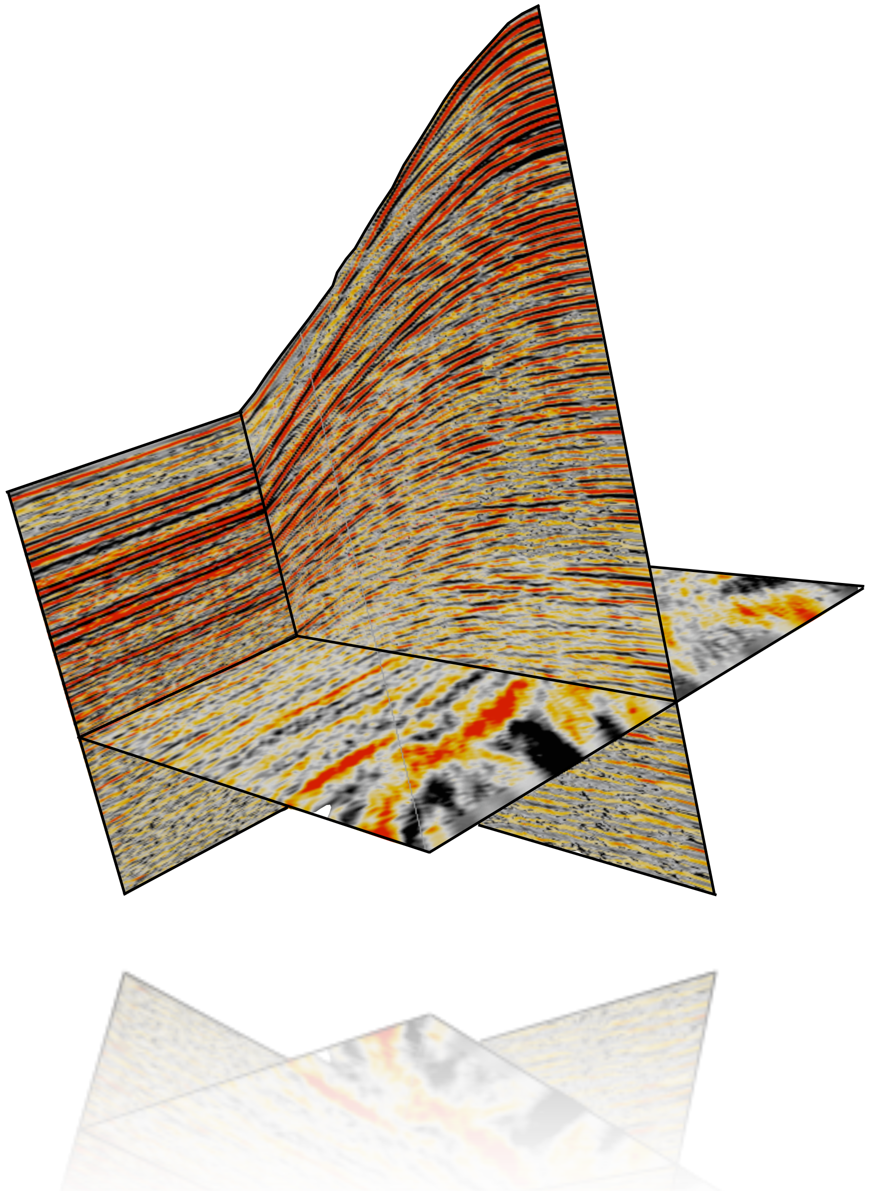
THE UNIVERSITY OF BRITISH COLUMBIA | VANCOUVER

Compressive modeling, imaging, and inversion



PIMS Summer School
Seattle, August 10-14, 2009

Outline

III. Compressive modeling, imaging, and inversion

- Inversion Helmholtz system by multi-level Krylov
- Linearized inversion by joint sparsity promotion
- Extensions & open problems
- **application:**
  - primary prediction from simultaneous data by curvelet-based wavefield inversion
  - compressive image volume recovery by *focused* curvelet-based wavefield inversion

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\boldsymbol{\mathcal{U}},\, \mathbf{m}\in\boldsymbol{\mathcal{M}}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$\mathbf{P}$ = Total multi-source and multi-frequency data volume

$\mathbf{D}$ = Detection operator

$\mathbf{U}$ = Solution of the Helmholtz equation

$\mathbf{H}$ = Discretized multi-frequency Helmholtz system

$\mathbf{Q}$ = Unknown seismic sources

$\mathbf{m}$ = Unknown model, e.g. $c^{-2}(x)$

Adjoint state methods [Plessix ‘06 & many others]

For each ***separate*** source **q** solve the **unconstrained problem:**

$$\min_{\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{p}-\mathcal{F}[\mathbf{m}]\|_2^2 \qquad \text{with} \qquad \mathcal{F}[\mathbf{m},\mathbf{q}] = \mathbf{D}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{q}$$

where ***model updates*** <=> ***migrated image***

$$\delta\mathbf{m} = \Re\left(\sum_{\omega}\omega^2\sum_{s}\bar{\mathbf{u}}\odot\mathbf{v}\right) = \mathbf{K}^*[\mathbf{m},\mathbf{Q}]\delta\mathbf{d}$$

$$\text{with } \boldsymbol{\delta}\mathbf{d} = \text{vec}(\mathbf{P}-\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{Q}])$$

involve single ***implicit*** solves of Helmholtz system

$$\mathbf{H}[\mathbf{m}]\mathbf{u} = \mathbf{q} \quad \text{and} \quad \mathbf{H}^*[\mathbf{m}]\mathbf{v} = \mathbf{r}$$

with

$$\mathbf{r} = \mathbf{D}^H(\mathbf{p}-\mathcal{F}[\mathbf{m}])$$

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\boldsymbol{\mathcal{U}},\,\mathbf{m}\in\boldsymbol{\mathcal{M}}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$$\begin{aligned}
\mathbf{P} &= \text{Total multi-source and multi-frequency data volume} \\
\mathbf{D} &= \text{Detection operator} \\
\mathbf{U} &= \text{Solution of the Helmholtz equation} \\
\mathbf{H} &= \text{Discretized multi-frequency Helmholtz system} \\
\mathbf{Q} &= \text{Unknown seismic sources} \\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x)
\end{aligned}$$

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\mathcal{U},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$\mathbf{P}$ = Total multi-source and multi-frequency data volume

$\mathbf{D}$ = Detection operator

$\mathbf{U}$ = Solution of the Helmholtz equation

$\mathbf{H}$ = Discretized multi-frequency Helmholtz system

$\mathbf{Q}$ = Unknown seismic sources

$\mathbf{m}$ = Unknown model, e.g. $c^{-2}(x)$

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\mathcal{U},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$$\begin{aligned}
\mathbf{P} &= \text{Total multi-source and multi-frequency data volume} \\
\mathbf{D} &= \text{Detection operator} \\
\mathbf{U} &= \text{Solution of the Helmholtz equation} \\
\mathbf{H} &= \text{Discretized multi-frequency Helmholtz system} \\
\mathbf{Q} &= \text{Unknown seismic sources} \\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x)
\end{aligned}$$

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\boldsymbol{\mathcal{U}},\,\mathbf{m}\in\boldsymbol{\mathcal{M}}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$$\begin{aligned}
\mathbf{P} &= \text{Total multi-source and multi-frequency data volume}\\
\mathbf{D} &= \text{Detection operator}\\
\mathbf{U} &= \text{Solution of the Helmholtz equation}\\
\mathbf{H} &= \text{Discretized multi-frequency Helmholtz system}\\
\mathbf{Q} &= \text{Unknown seismic sources}\\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x)
\end{aligned}$$

Challenges: there are many ...

Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, '09]

- data cannot be explained fully
- the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\mathcal{U},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$\mathbf{P}$ = Total multi-source and multi-frequency data volume

$\mathbf{D}$ = Detection operator

$\mathbf{U}$ = Solution of the Helmholtz equation

$\mathbf{H}$ = Discretized multi-frequency Helmholtz system

$\mathbf{Q}$ = Unknown seismic sources

$\mathbf{m}$ = Unknown model, e.g. $c^{-2}(x)$

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\mathcal{U},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$\mathbf{P}$ = Total multi-source and multi-frequency data volume

$\mathbf{D}$ = Detection operator

$\mathbf{U}$ = Solution of the Helmholtz equation

$\mathbf{H}$ = Discretized multi-frequency Helmholtz system

$\mathbf{Q}$ = Unknown seismic sources

$\mathbf{m}$ = Unknown model, e.g. $c^{-2}(x)$

Time domain vs. Frequency domain

	Time domain	Frequency domain
Solution of wave equation	explicit, easy	implicit, not easy
Imaging	time history, checkpointing, not trivial	all frequencies, freq. subsampling, easy
Computational algorithm	paralellizable via domain decomposition (DD)-type algorithm	embarrasingly parallel in frequency, no communication, DD-type can apply for very large problem (3D)
Boundary condition and damping layer	not trivial	trivial, use complex velocity
Modeling relaxation	not trivial	trivial, use freq. dep. complex velocity

Multiexperiment wavefield simulations

Based on discretization of the Helmholtz equation:

$$\mathcal{H}u = -\Delta u - \omega^2 m u = q$$

$$\begin{bmatrix} \boldsymbol{\mathcal{H}}_{\omega_1} & 0 & & \\ 0 & \boldsymbol{\mathcal{H}}_{\omega_2} & \ddots & \\ & \ddots & \ddots & 0 \\ & & 0 & \boldsymbol{\mathcal{H}}_{\omega_{n_f}} \end{bmatrix} \begin{bmatrix} \overbrace{[\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_{n_s}]_{\omega_1}}^{\mathbf{U}_{\omega_1}} \\ \vdots \\ \vdots \\ \underbrace{[\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_{n_s}]_{\omega_{n_f}}}_{\mathbf{U}_{n_f}} \end{bmatrix} = \begin{bmatrix} \overbrace{[\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_{n_s}]_{\omega_1}}^{\mathbf{Q}_{\omega_1}} \\ \vdots \\ \vdots \\ \underbrace{[\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_{n_s}]_{\omega_{n_f}}}_{\mathbf{Q}_{n_f}} \end{bmatrix}$$

$$\boldsymbol{\mathcal{H}}_{\omega_j} := \boldsymbol{\mathcal{H}}(\omega_j), \quad \omega_j = 2\pi j \Delta f, \quad j = 1, \ldots, n_f$$

Δf frequency sample interval

Computational complexity

Multiple-shots (right-hand sides), multiple frequencies

n_s : number of shots n_f : number of frequencies

2D	Direct methods	Iterative Methods
LU factors	$n_f \mathcal{O}(n^4)$	-
Solution	$n_s n_f \mathcal{O}(n^3)$	$n_s n_f n_{iter} \mathcal{O}(n^2)$

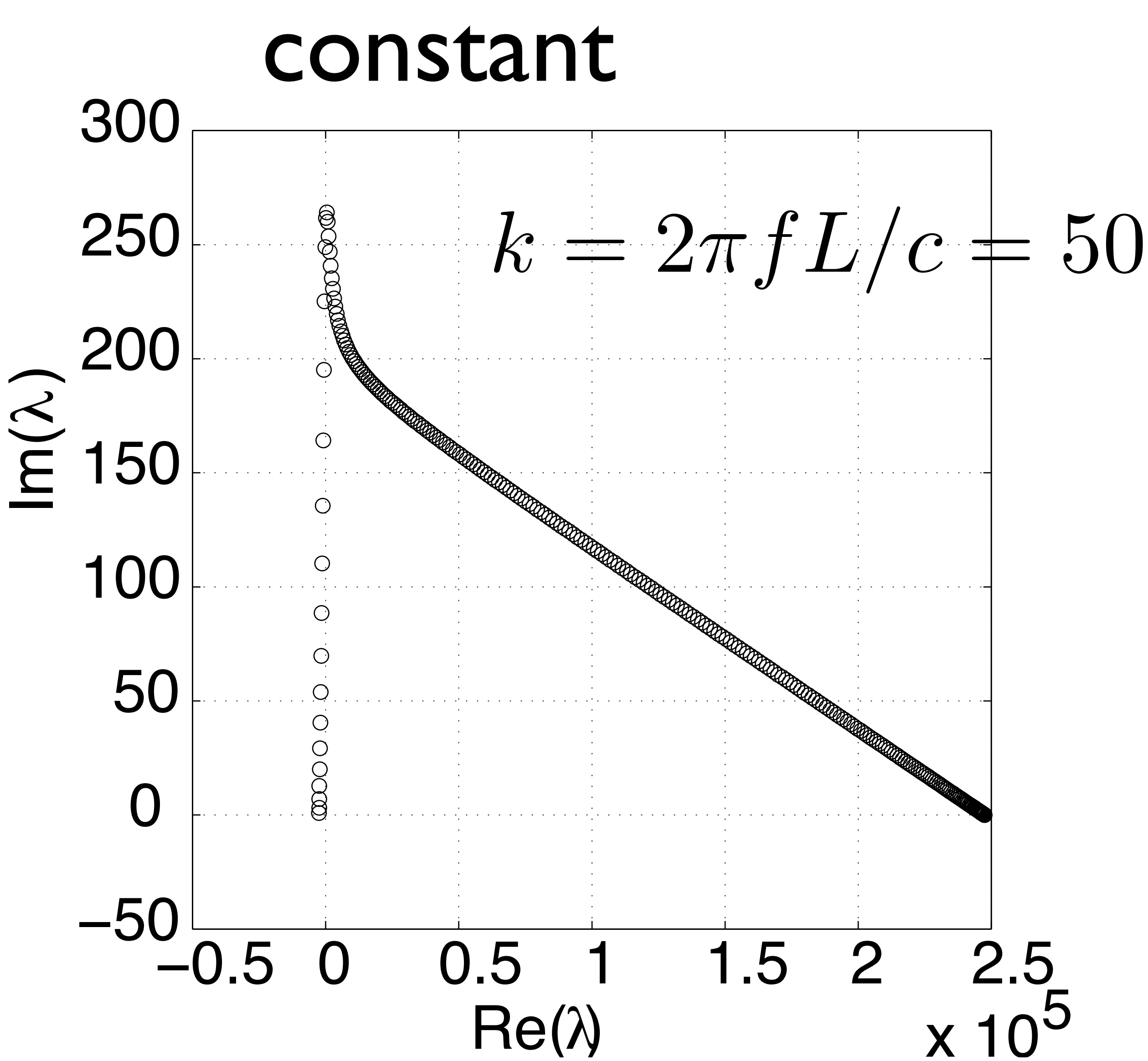
3D	Direct methods	Iterative Methods
LU factors	$n_f \mathcal{O}(n^9)$	-
Solution	$n_s n_f \mathcal{O}(n^5)$	$n_s n_f n_{iter} \mathcal{O}(n^3)$

IM can be competitive if $n_{iter} \ll n^d$ (with, e.g., preconditioner)

For similar analysis for MUMPS, see [Virieux, The Leading Edge, 2009]

Complications

One-*d* example: not of practical interest but tells the story



- Small eigenvalues close to zero, large eigenvalues unbounded: ill-conditioned
- Real parts of eigenvalues change signs: indefinite

Convergence is not guaranteed.

Indefiniteness the most difficult to handle. No iterative method for indefinite system

Tackle indefiniteness by Laplacian shift

Use as preconditioner the damped Helmholtz op.:

$$\mathbf{M} \triangleq -\nabla \cdot (\nabla) - (1 - \frac{1}{2}\hat{j}) \left(\frac{\omega}{c}\right)^2, \quad \hat{j} = \sqrt{-1}.$$

Then solve using iterative method the system

$$\boxed{\mathbf{H}\mathbf{M}^{-1}\mathbf{w} = \mathbf{f}, \quad \mathbf{u} = \mathbf{M}^{-1}\mathbf{w}}$$

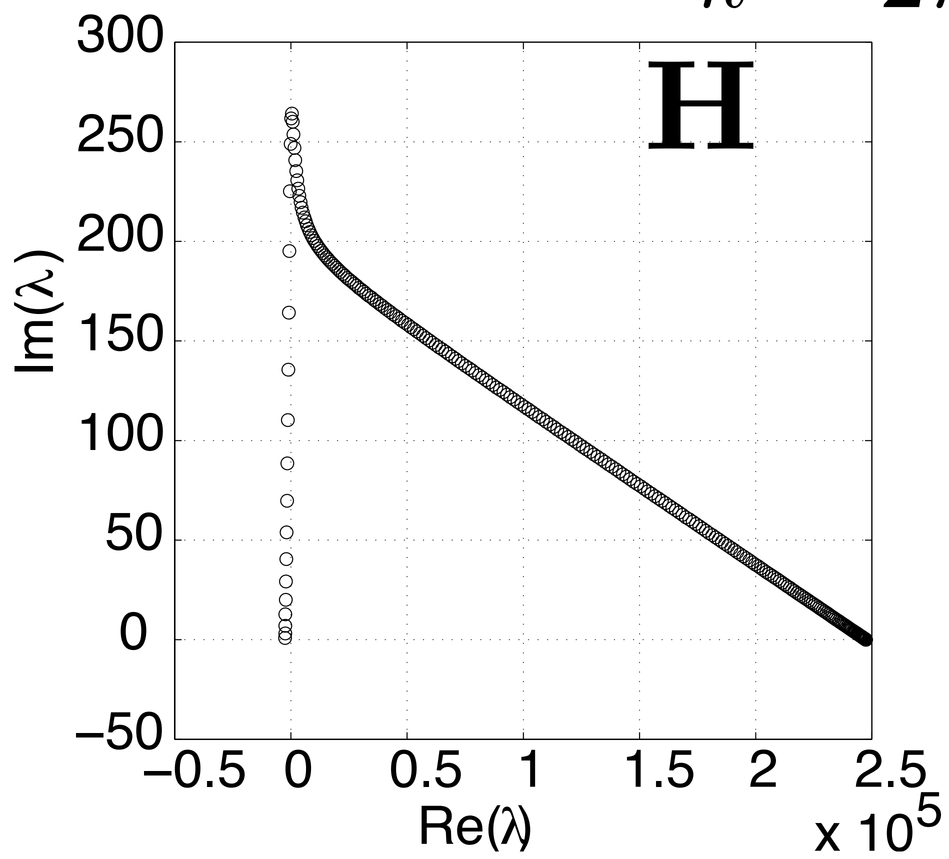
(And similarly for back-propagated wavefield)

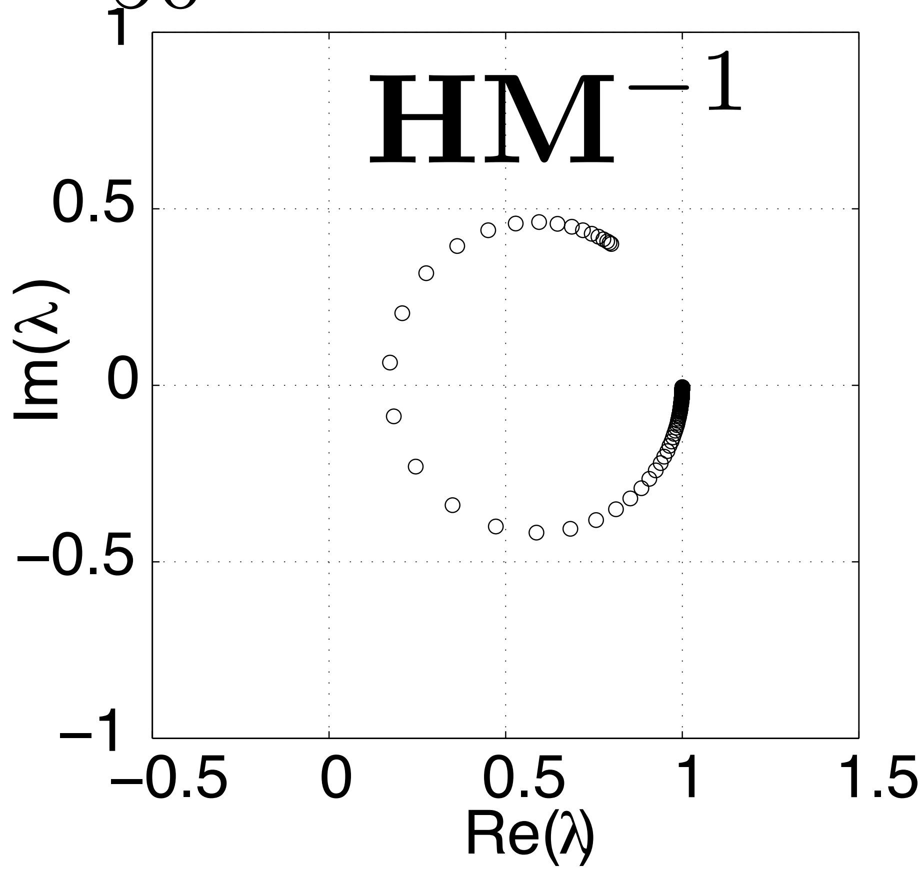
$$\mathbf{H}\mathbf{M}^{-1} =: \hat{\mathbf{H}}$$

[Erlangga, Oosterlee, Vuik, 2006]
[Riyanti et al., 2006]
[Plessix et al., 2007]

Indefiniteness removed

$$k = 2\pi f L / c = 50$$





- Real parts of eigenvalues have the same signs: definite! Iterative methods will converge easier $n_{iter} < n^d$
- To obtain $\mathcal{O}(n^d)$ method , $\mathbf{M}^{-1}$ computed by one multigrid iteration
- Large eigenvalue bounded by one, still some small eigenvalues ill-conditioned

Tackle ill conditioning

Multilevel/scale operator:

$$\mathbf{Q} = \overbrace{\mathbf{I} - \mathbf{Z}\widehat{\mathbf{H}}^{-1}\mathbf{Z}^T\mathbf{H}\mathbf{M}^{-1}}^{\text{shift small eigenvalues to 0}} + \overbrace{\mathbf{Z}\widehat{\mathbf{H}}^{-1}\mathbf{Z}^T}^{\text{shift zero eigenvalues to 1}},$$

with

$$\widehat{\mathbf{H}} = \mathbf{Z}^T\mathbf{H}\mathbf{M}^{-1}\mathbf{Z}, \quad dim\widehat{\mathbf{H}} \ll dim\mathbf{H}$$

$\mathbf{Z}$: interpolation/fining operator

Then, solve $\boxed{\mathbf{H}\mathbf{M}^{-1}\mathbf{Q}\mathbf{y} = \mathbf{f}, \quad \mathbf{u} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{y}}$

[Erlangga, Nabben, 2008]
[Erlangga, Herrmann, 2008]

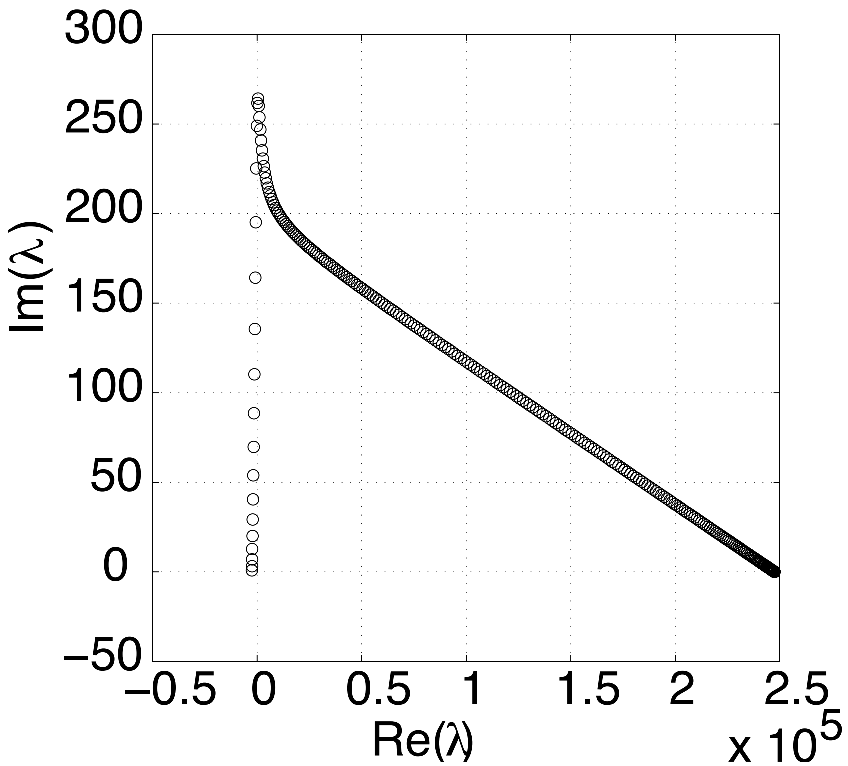
Tackle ill conditioning

$$\mathbf{Q} = \mathbf{I} - \mathbf{Z}\widehat{\mathbf{H}}^{-1}\mathbf{Z}^T\mathbf{H}\mathbf{M}^{-1} + \mathbf{Z}\widehat{\mathbf{H}}^{-1}\mathbf{Z}^T$$

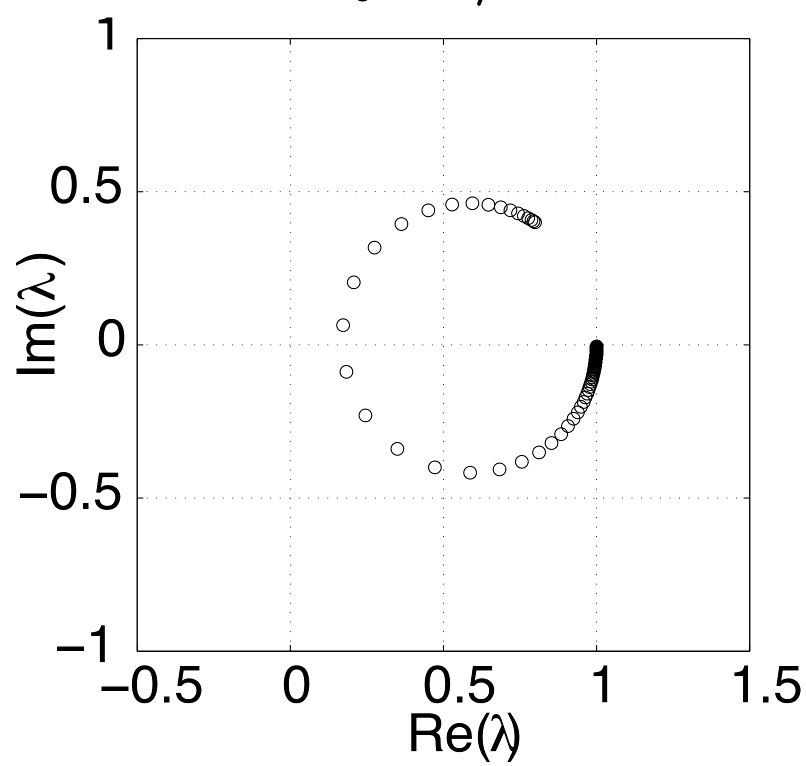
- The action of $\mathbf{Z}^T$ restricts components of errors, which are responsible for slow convergence, into the coarse grid (level)
- The action of $\widehat{\mathbf{H}}^{-1}$ reduces those components in coarse grid (level)
- The action of $\mathbf{Z}$ interpolates the reduction back into the fine level
- $\widehat{\mathbf{H}}^{-1}$ is computed recursively: Multilevel method

Ill conditioning removed

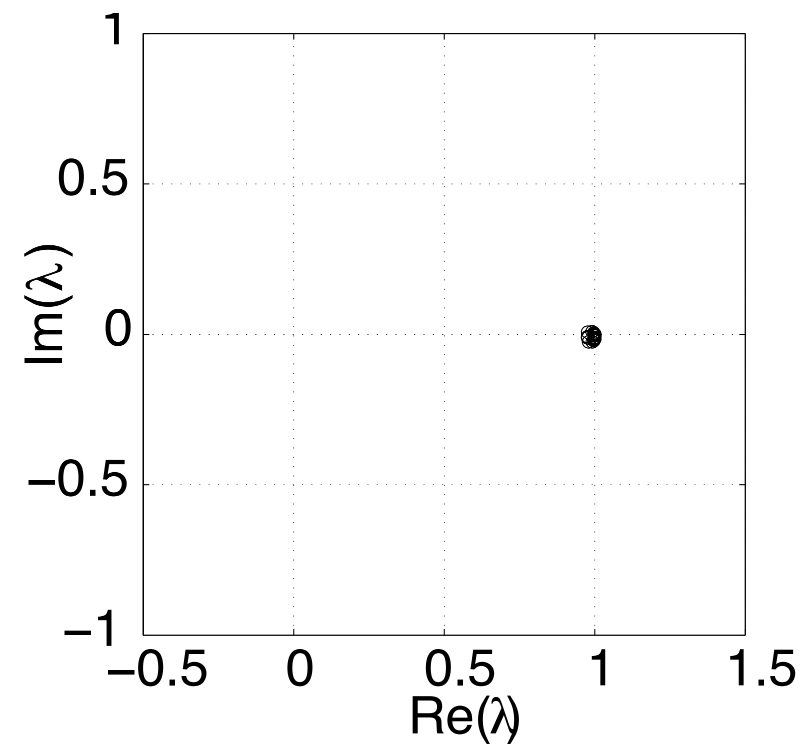
$$k = 2\pi f L / c = 50$$



$$\mathbf{H}$$



$$\mathbf{HM^{-1}}$$



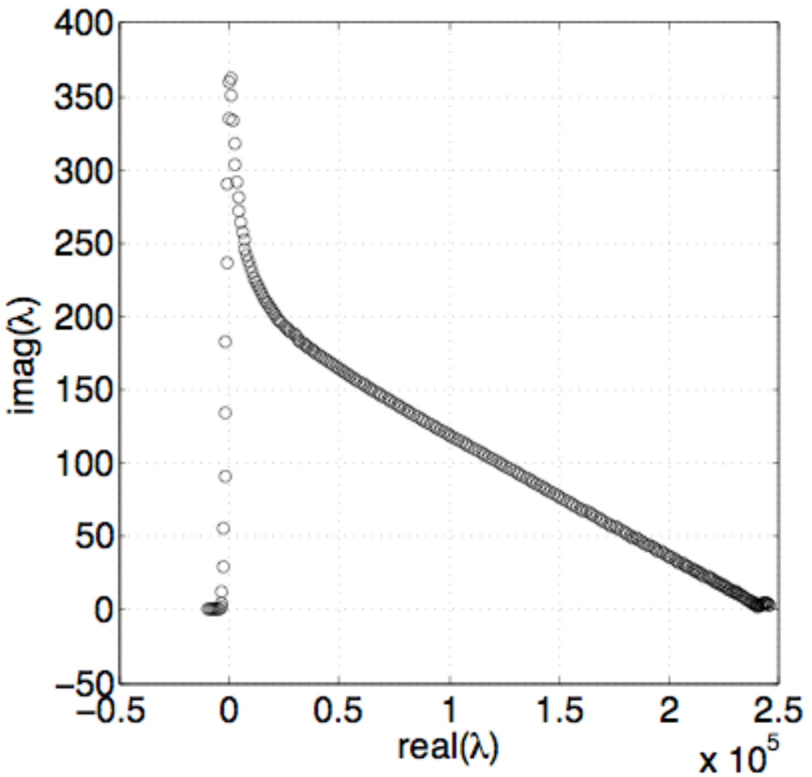
$$\mathbf{HM^{-1}Q}$$

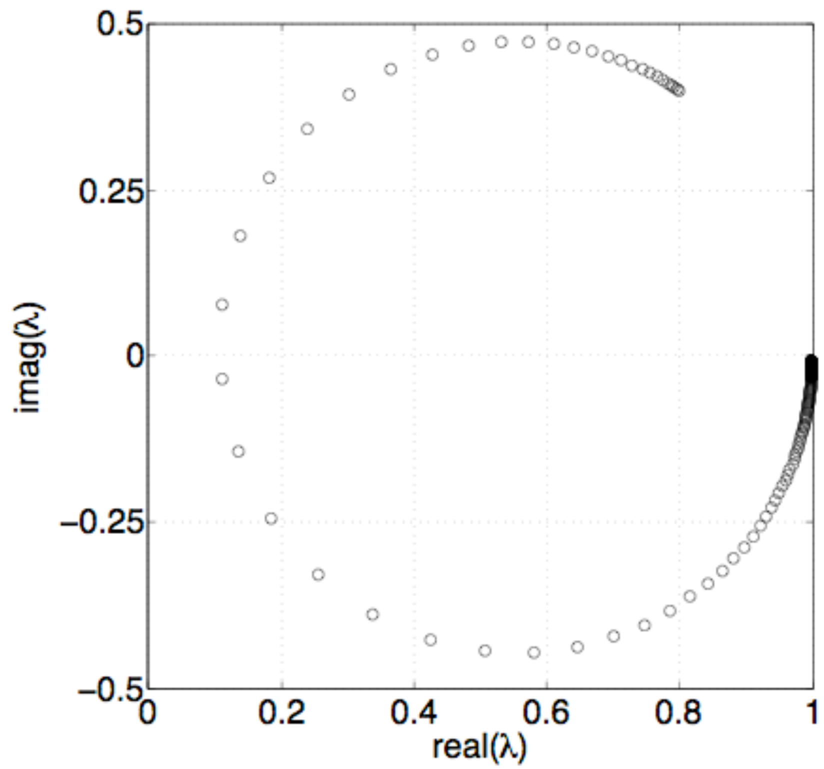
- Notice shift of eigenvalues towards one due to Q!
- The spectrum of $\mathbf{HM^{-1}Q}$ is favorable for iterative methods

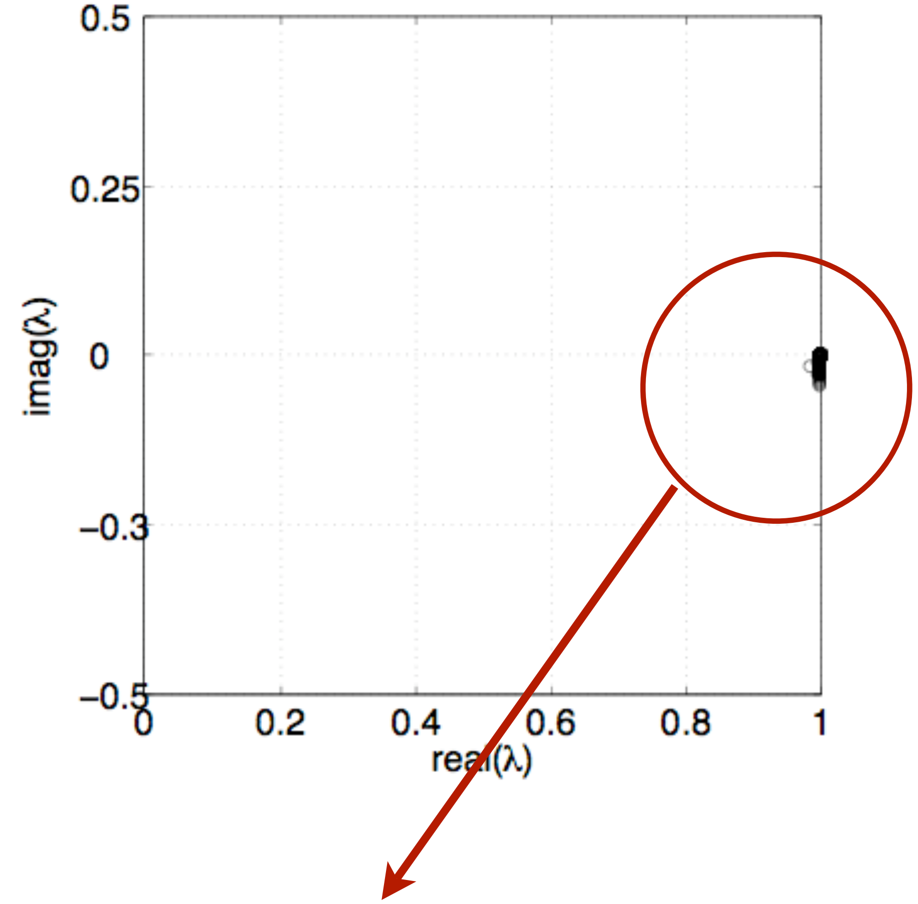
More on eigenvalues

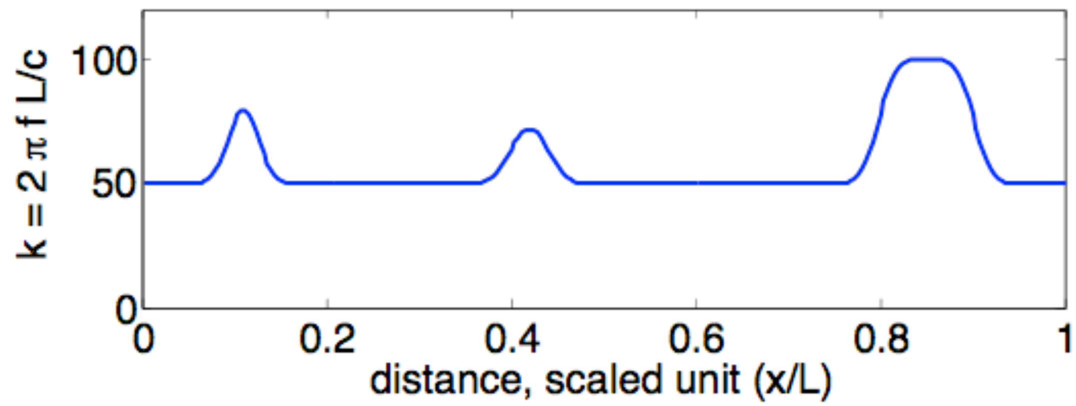
1D non-constant wavenumber k, smooth model $k = (50, 100)$

$\mathbf{H}$ $\mathbf{HM}^{-1}$ $\mathbf{HM}^{-1}\mathbf{Q}$







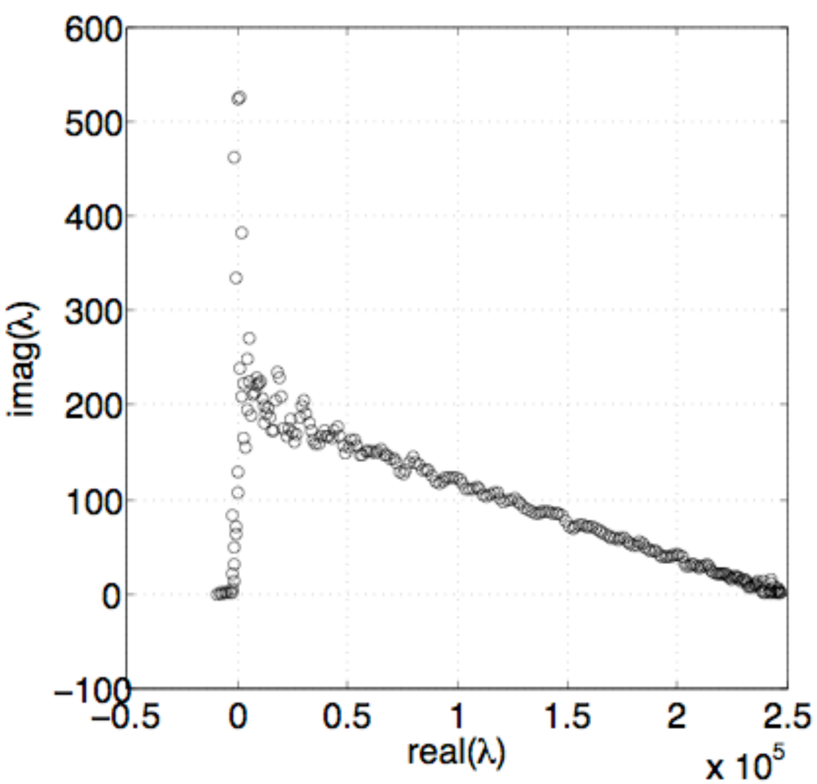


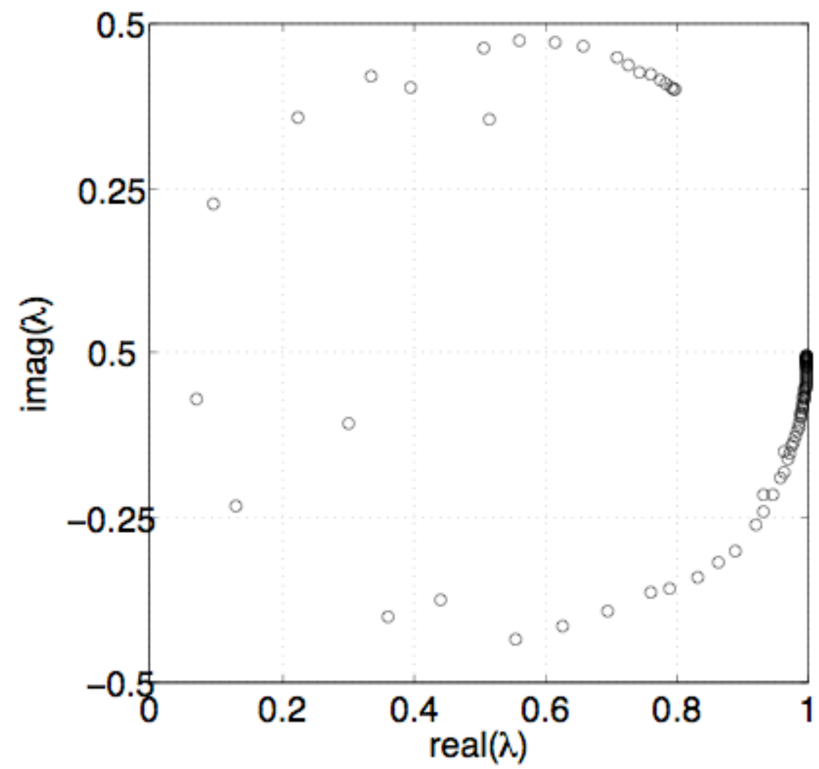
Clustering around one

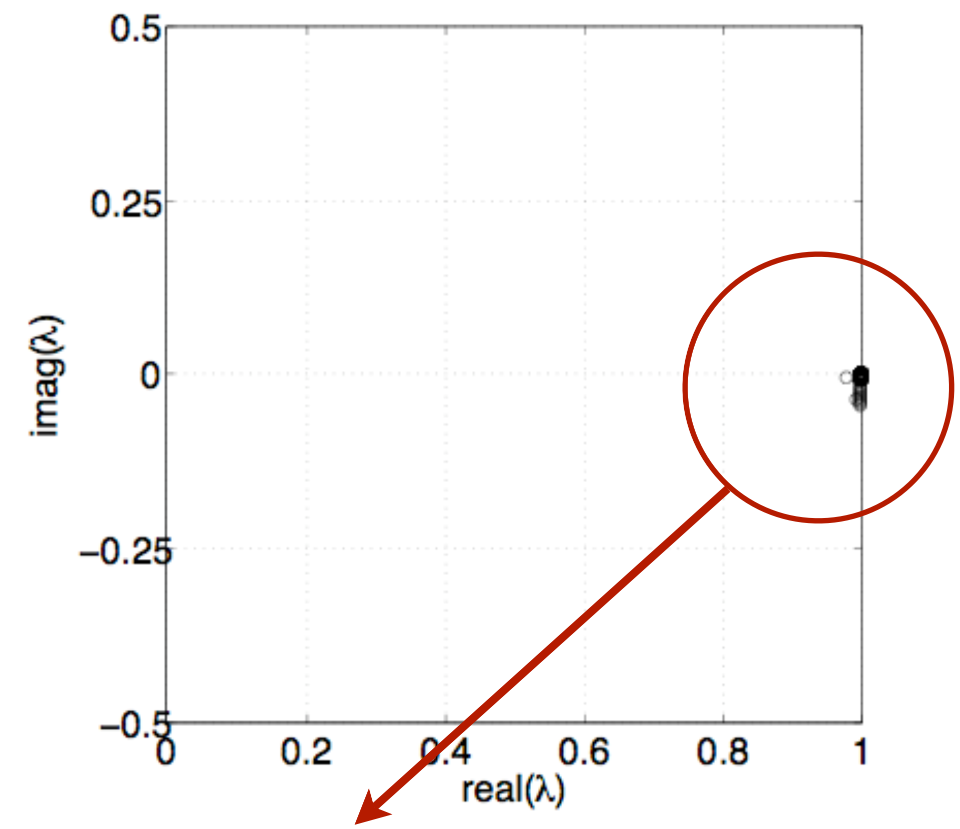
More on eigenvalues

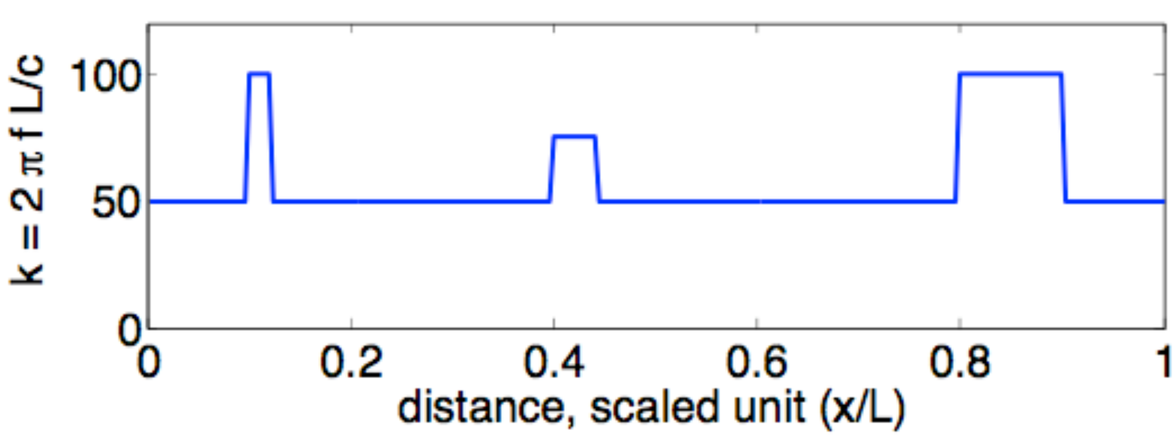
1D non-constant wavenumber k, hard model $k = (50, 100)$

$\mathbf{H}$ $\mathbf{HM}^{-1}$ $\mathbf{HM}^{-1}\mathbf{Q}$







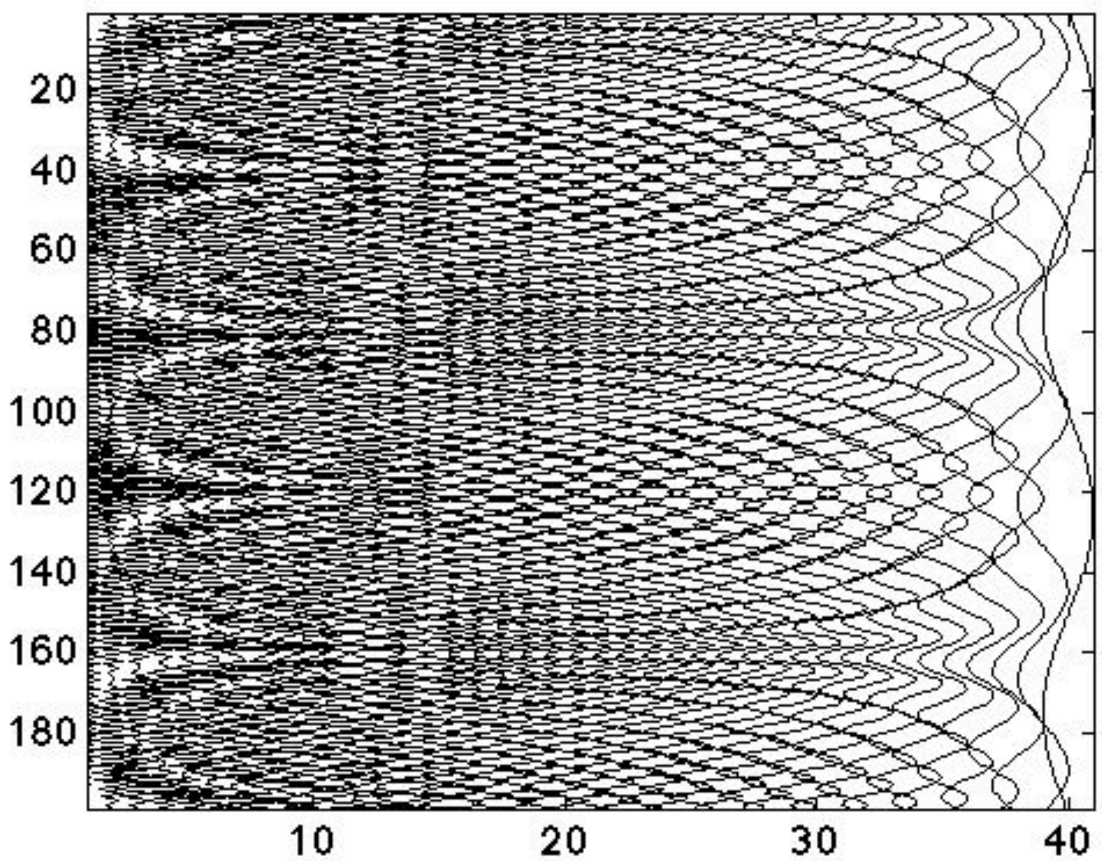


Clustering around one

For constant, smooth, or hard model, one can expect the same convergence rate

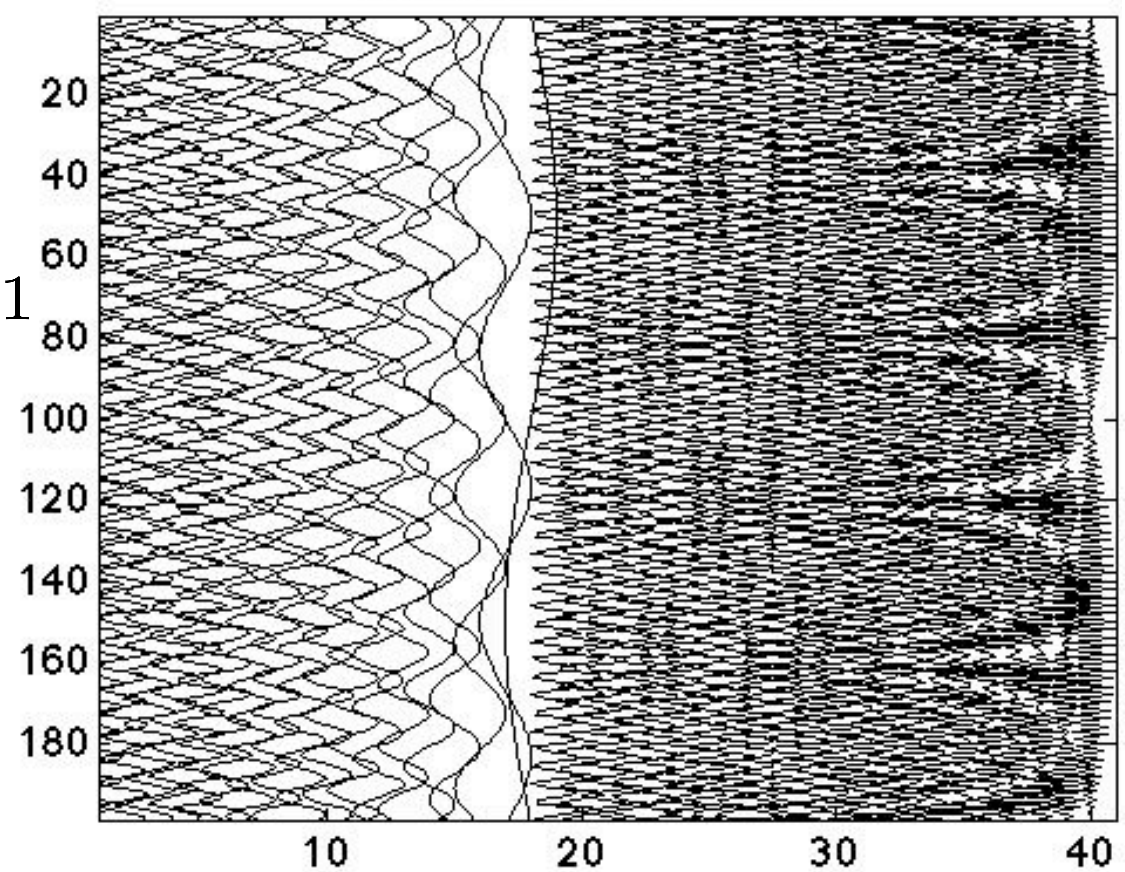
Eigenvectors: 1D constant velocity

$\mathbf{H}$



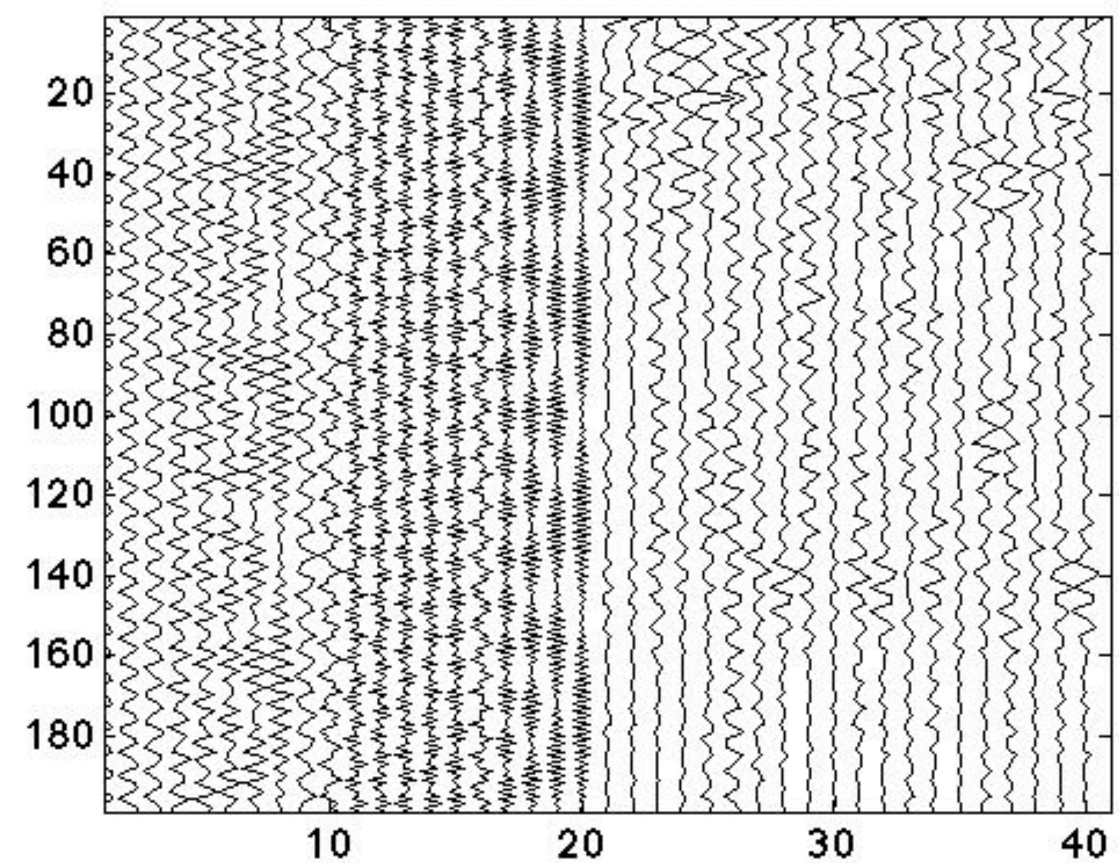
Real part of eigenvectors

$\mathbf{HM^{-1}}$



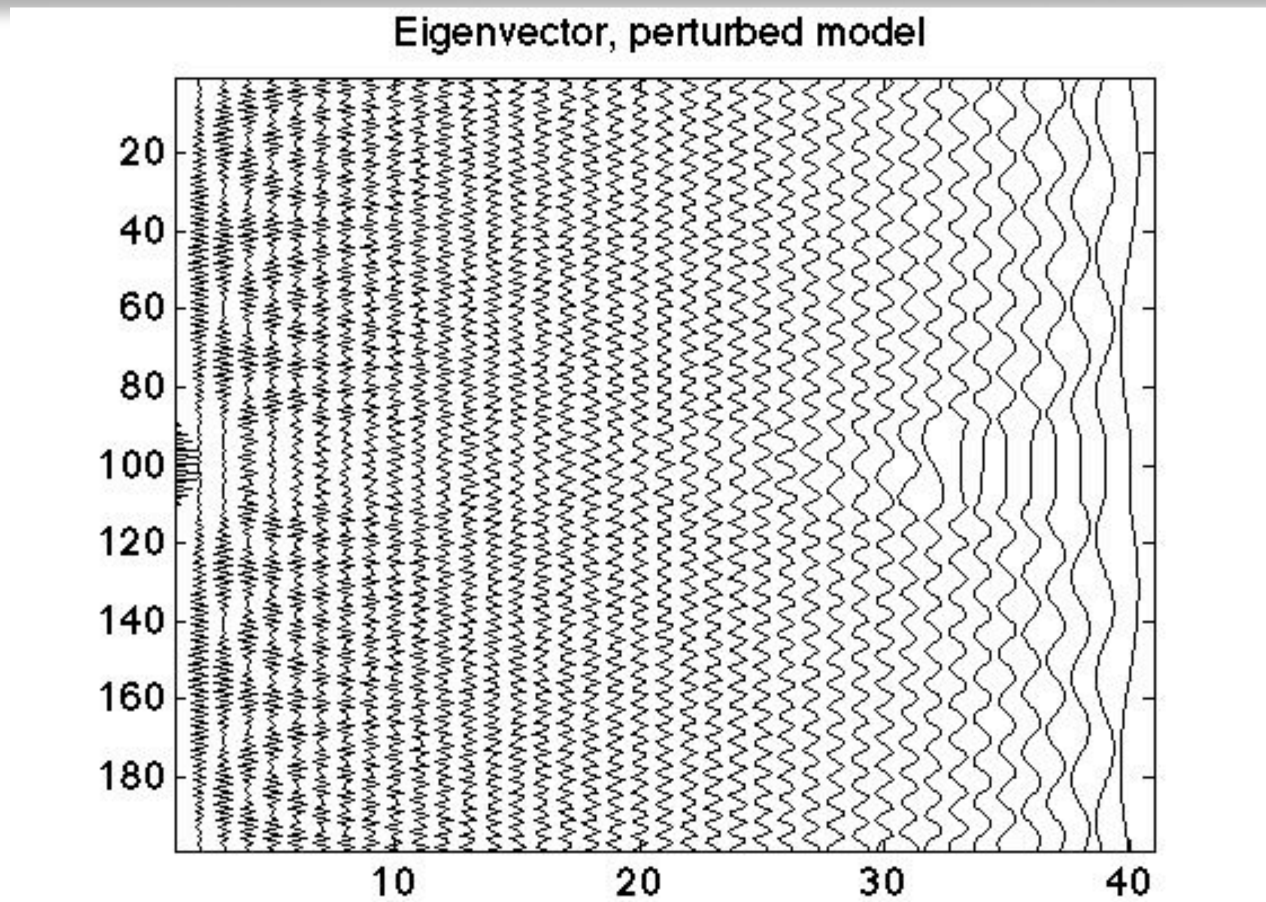
$\mathbf{HM^{-1}Q}$

Real part of eigenvectors

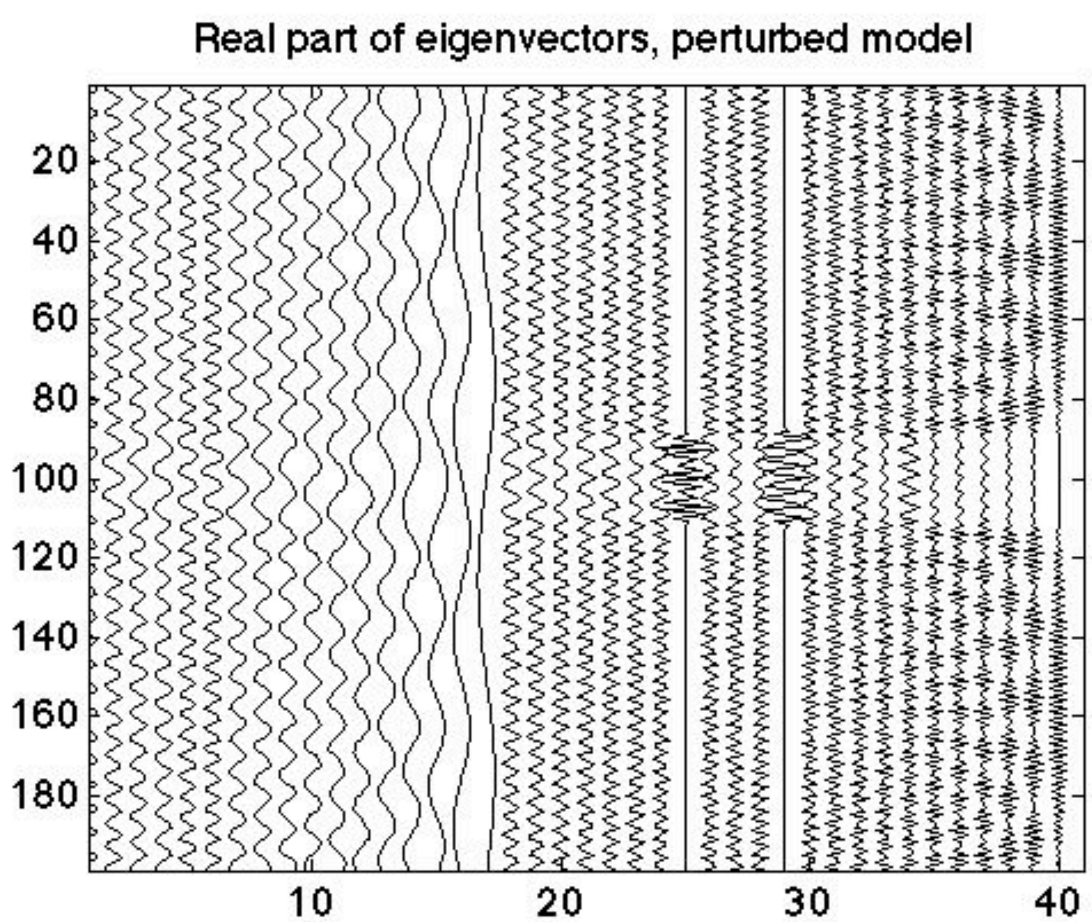


Eigenvectors: 1D with velocity jump

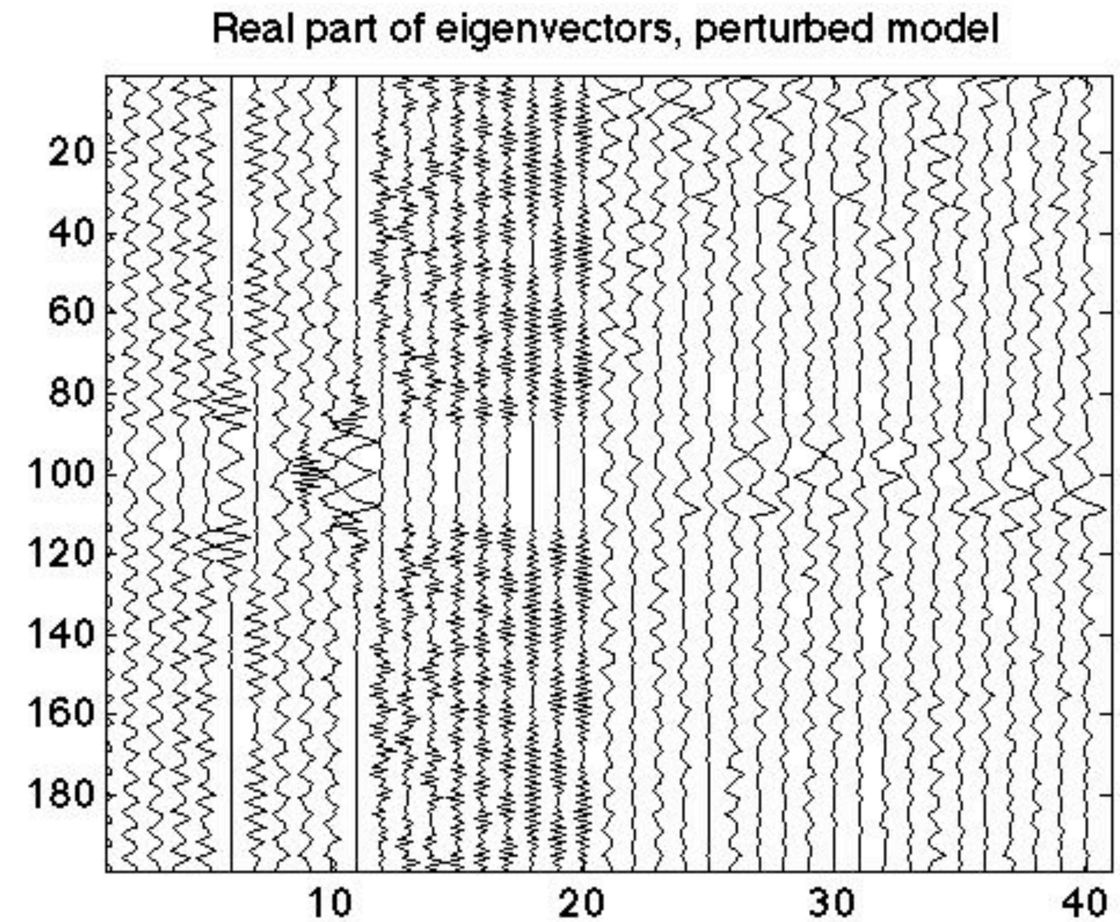
$\mathbf{H}$



$\mathbf{H}\mathbf{M}^{-1}$

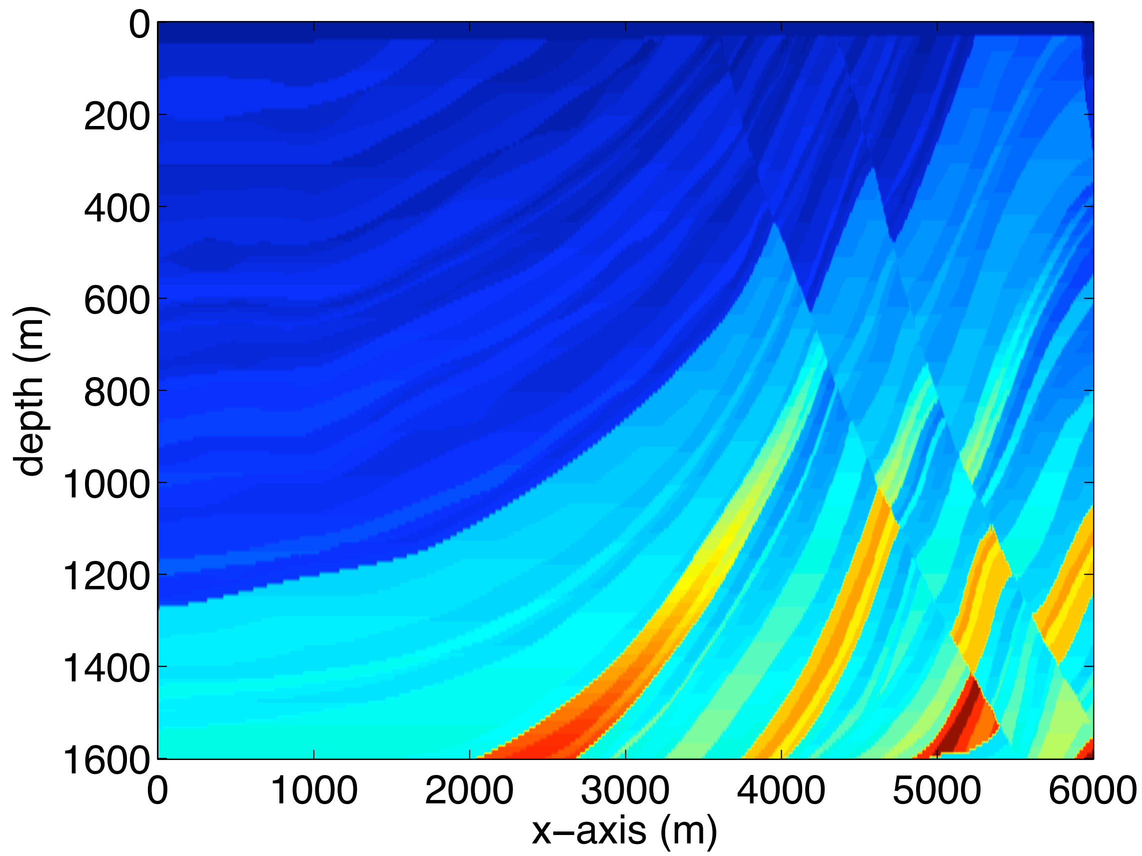


$\mathbf{H}\mathbf{M}^{-1}\mathbf{Q}$

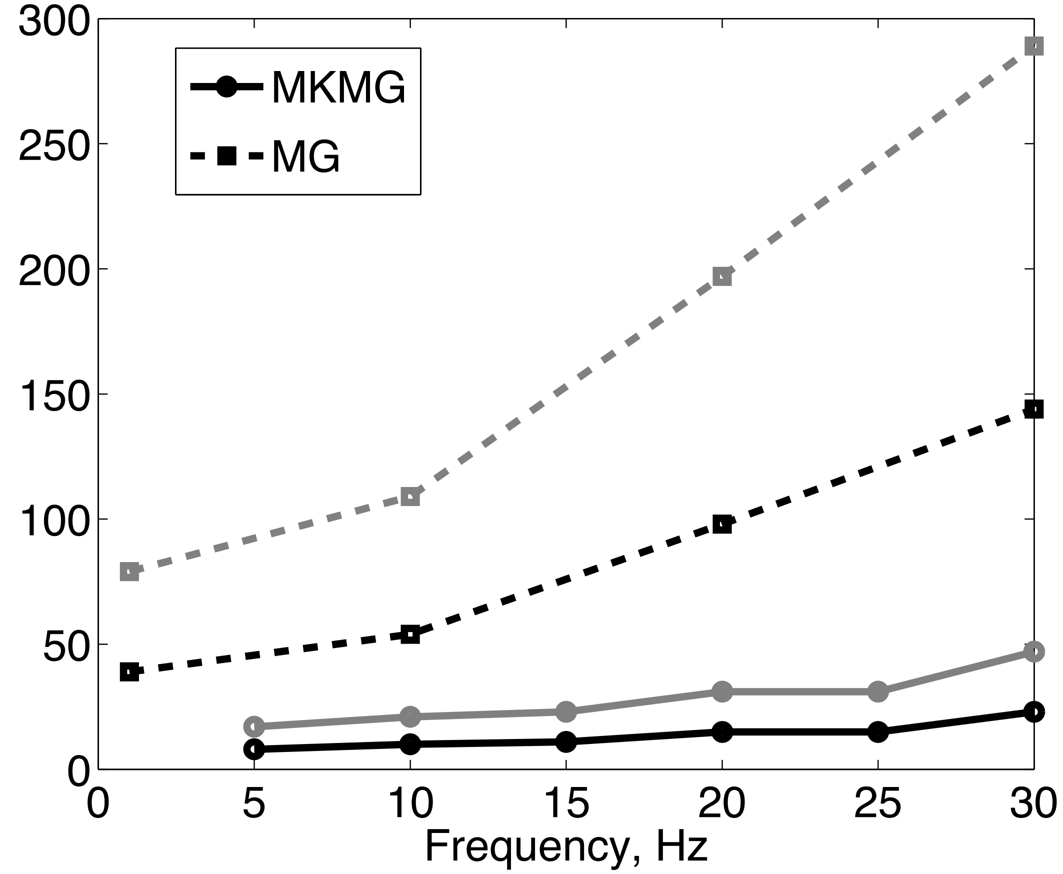


Example: forward modeling

Forward modeling, one shot position, hard model



- Velocity contrast: 1500 - 4000 m/s
- Convergence is less dependent of frequency



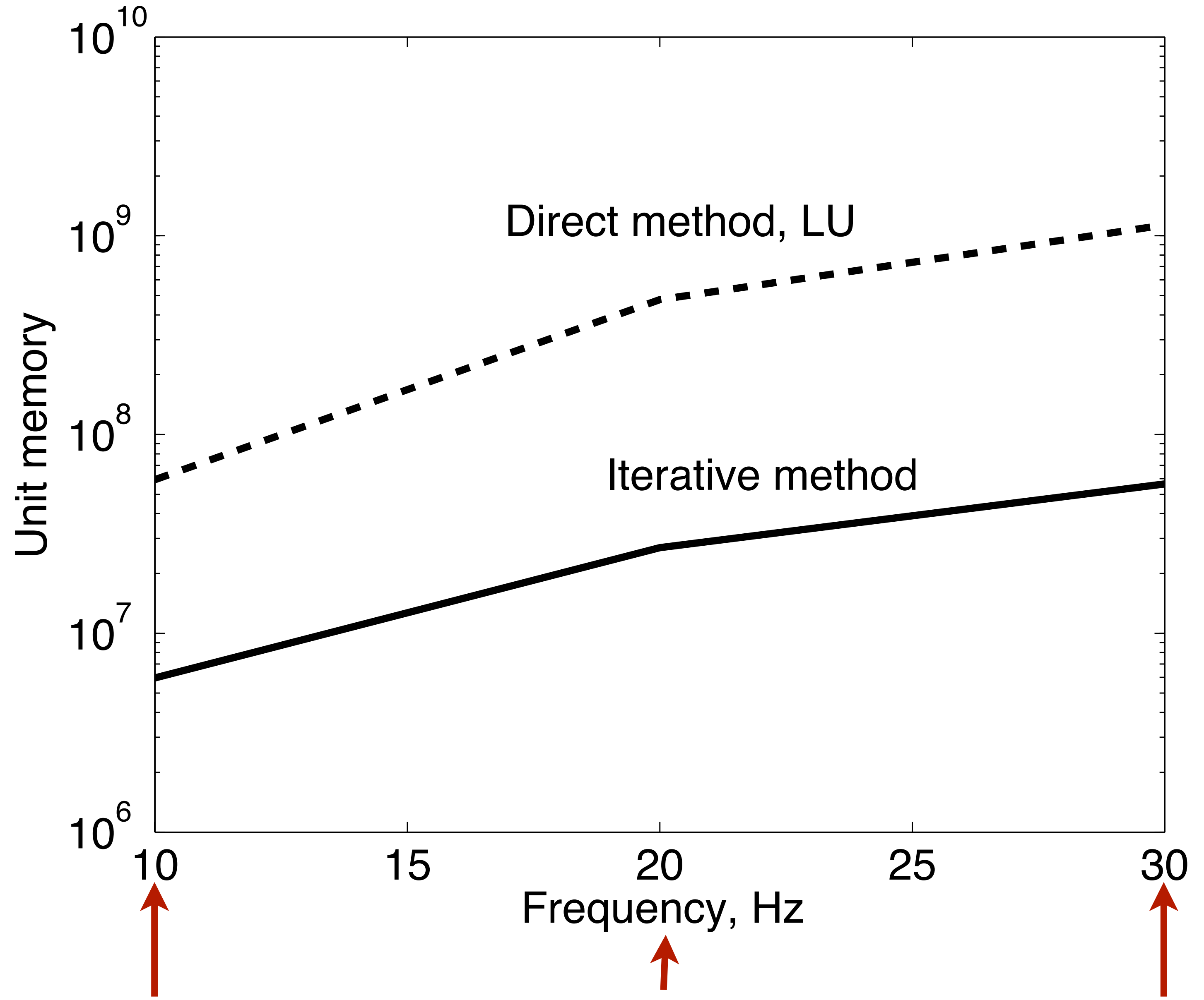
Number of iteration

Number of matrix-vec multiplies

MG: $\mathbf{HM}^{-1}$

MKMG: $\mathbf{HM}^{-1}\mathbf{Q}$

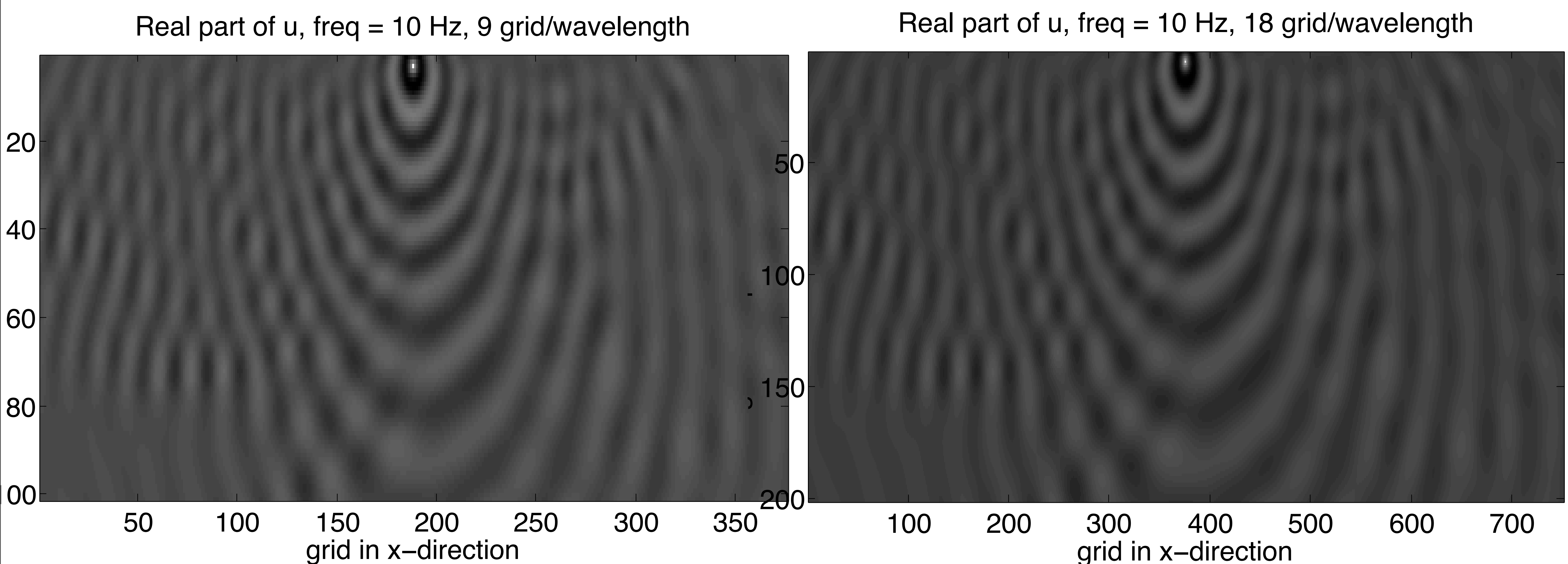
Example: Marmousi, cont'd



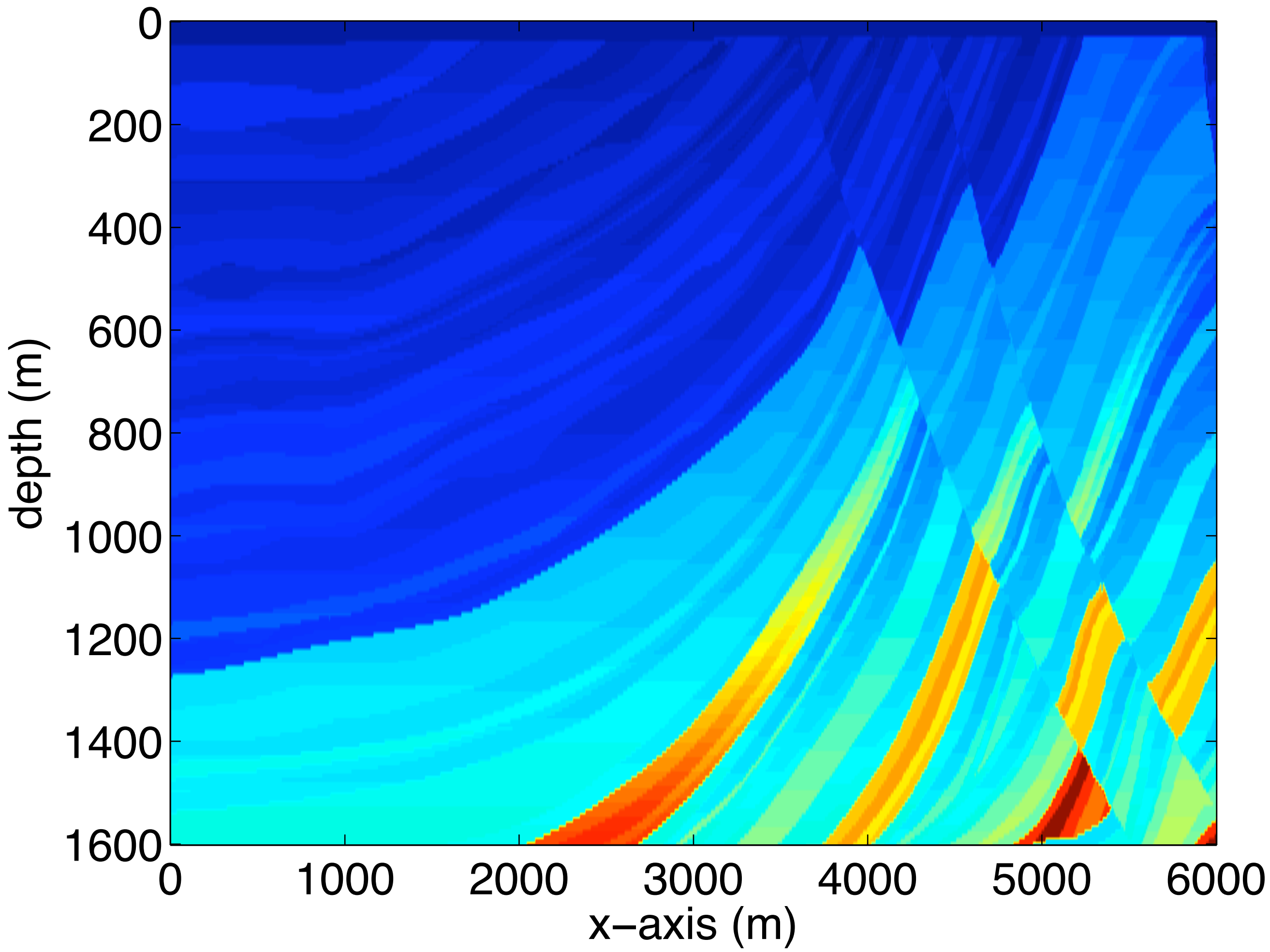
gridpoint : 751×201 1501×401 2001×534

Example: forward modeling

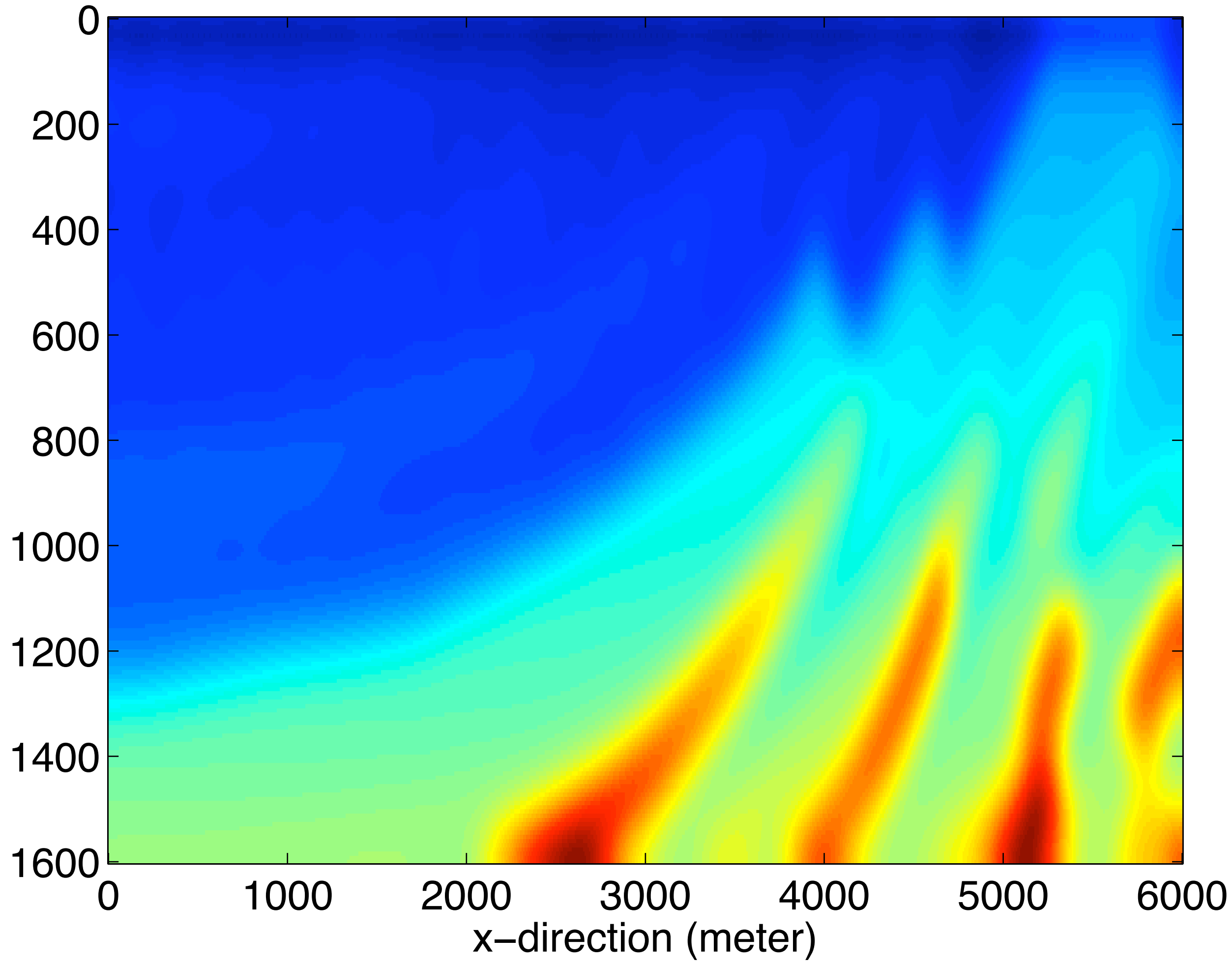
One shot position, hard model : wavefield



Example: imaging



Target model

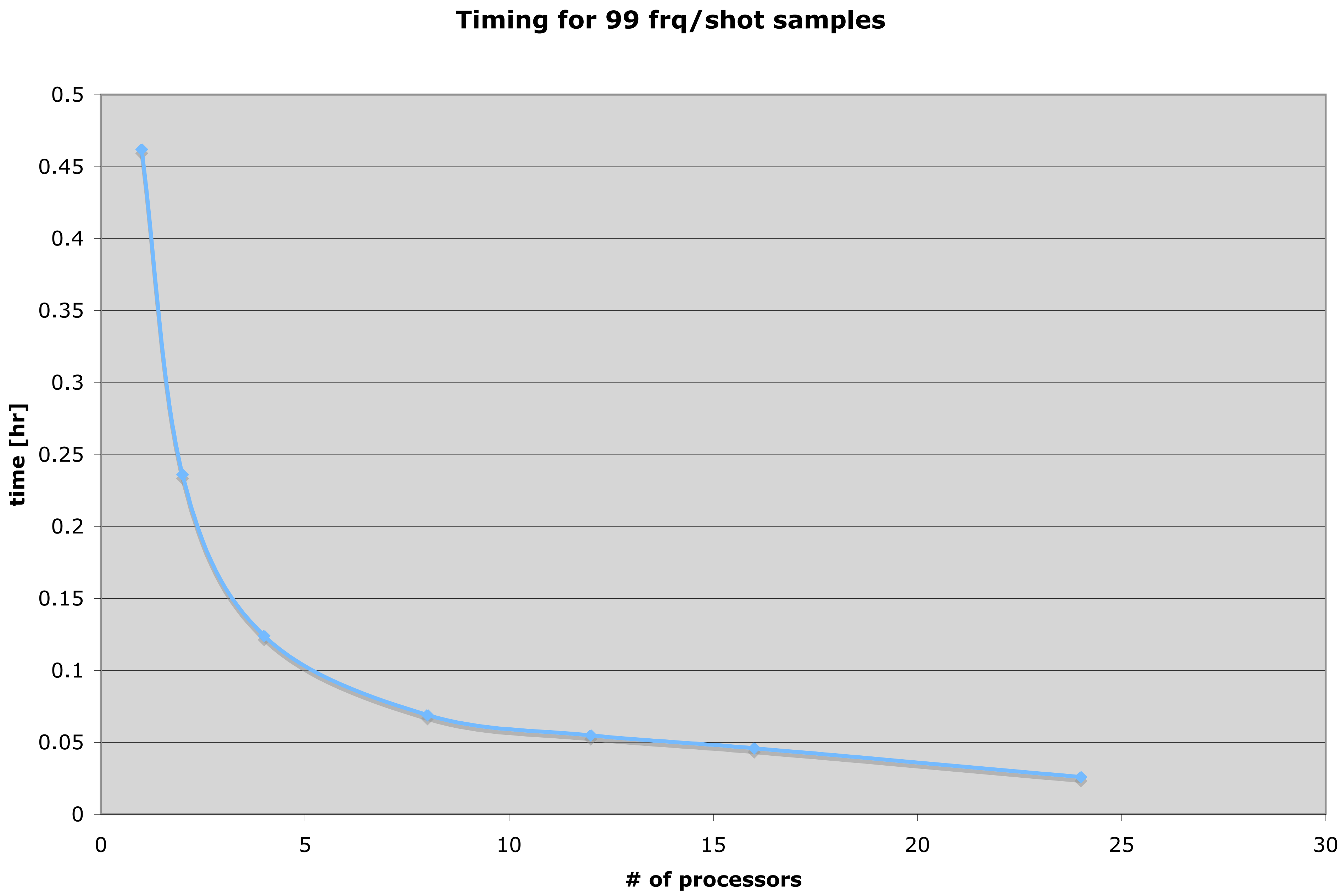


After first gradient-based update

$$\mathbf{m}_1 = \mathbf{m}_0 + \delta\mathbf{m}$$

$\delta\mathbf{m}$ (not shown) is computed using data from 188 shots and 11 frequencies (0.5-5.0 Hz)

Parallellism



CPU time, single processor: (28 min) vs. Symes's (7 min)

Challenges: there are many ...

✓ Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, '09]

- data cannot be explained fully
- the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

System-size reduction

- Apply CS to reduce cost of ***wavefield*** simulation with Helmholtz
  - use ***simultaneous sources*** instead of ***separated sources***
  - leverage transform-domain sparsity & randomized subsampling by **one-norm sparsity promotion**
  - reduce size Helmholtz system
    - sources (number of right-hand sides)
    - angular frequencies (number of blocks)

- Apply CS to reduce cost of computing ***image volumes*** by multi-dimensional correlations via ***explicit*** matrix-matrix multiplies
  - randomize and subsample wavefields in **model space**
  - leverage transform-domain sparsity and focusing in the ***model space*** by ***joint sparsity promotion with mixed (1,2) norms***
  - reduce costs of storage and explicit matrix-matrix multiplies
    - sources (right-hand sides), receivers, depth
    - angular frequencies (blocks)

Relation to existing work

- **Simultaneous & continuous acquisition:**
  - *Efficient Seismic Forward Modeling using Simultaneous Random Sources and Sparsity* by N. Neelamani and C. Krohn and J. Krebs and M. Deffenbaugh and J. Romberg, ‘08
- **Simultaneous simulations & migration:**
  - *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.
- **Imaging:**
  - *How to choose a subset of frequencies in frequency-domain finite-difference migration* by Mulder & Plessix, ’04.
  - *Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies* by Sirque and Pratt, ’04.
- **Full-waveform inversion:**
  - *3D prestack plane-wave, full-waveform inversion* by Vigh and Starr, ‘08
- **Wavefield extrapolation:**
  - *Compressed wavefield extrapolation* by T. Lin and F.J.H, ’07
  - *Compressive wave computations* by L. Demanet (SIA ’08 MS79 & Preprint)

Tools

Compressive sensing based on Johnson-Lindenstrauss embeddings

- *Compressive sensing* [Donoho, 06‘, Candes, Romberg, Tao, ‘06]

$$\mathbf{b} = \mathbf{RMx} \quad \textbf{[randomized subsampling]}$$

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{RMx} - \mathbf{b}\|_2 \leq \sigma$$

$$\tilde{\mathbf{x}} \approx \mathbf{x}$$

Fast matrix computations based on Johnson-Lindenstrauss embeddings

- Improved Approximation Algorithms for Large Matrices via Random Projections by Tamás Sarlós, ’08

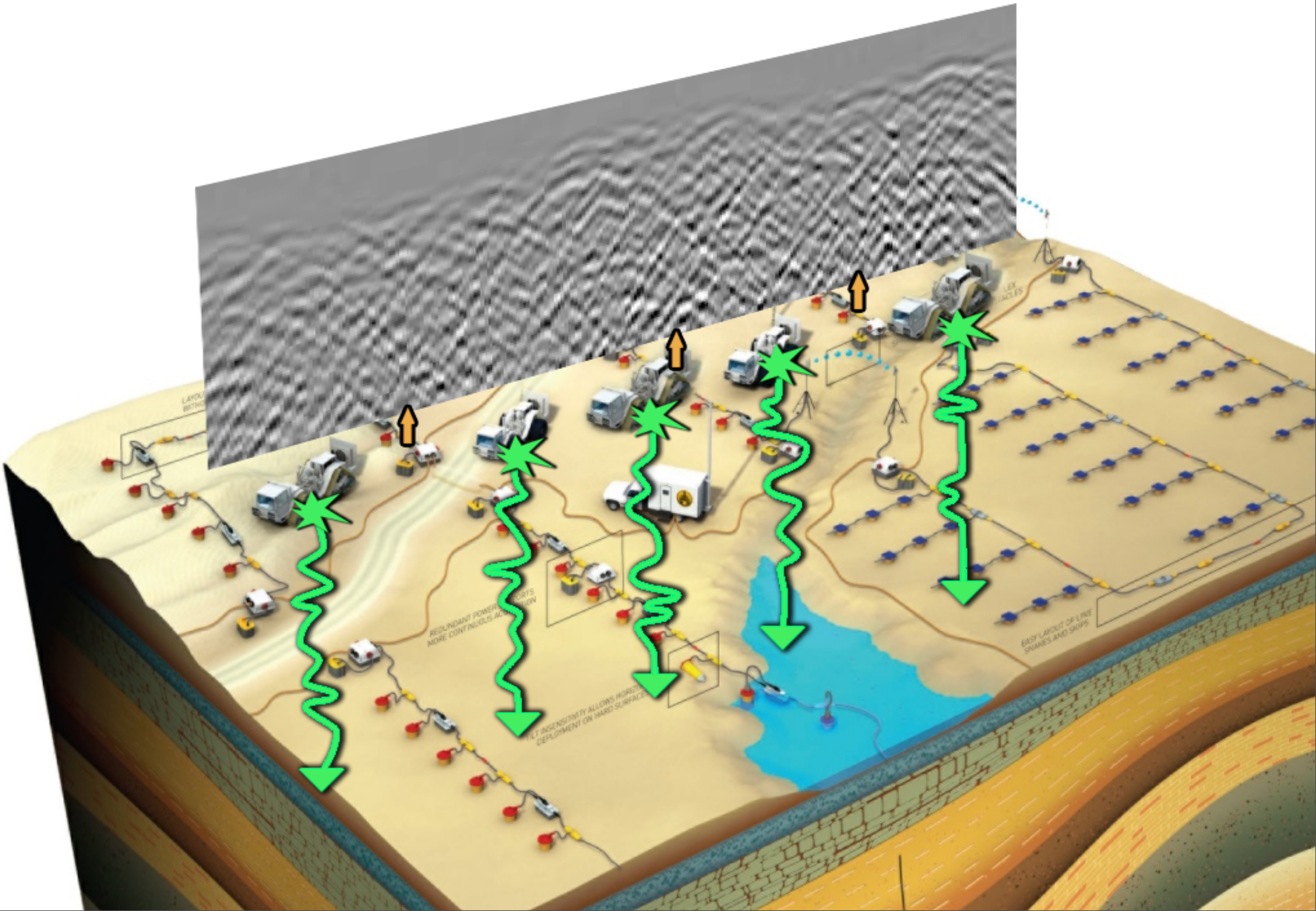
$$\mathbf{AB} \approx \mathbf{A}\left(\mathbf{RM}\right)^{*}\left(\mathbf{RM}\right)\mathbf{B}$$

Joint sparsity-promotion with mixed (1,2) norm minimization

- *Joint-sparse recovery from multiple measurements* by E. van den Berg and M. Friedlander, ‘09

$$\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{1,2} \quad \text{subject to} \quad \|\mathbf{AX} - \mathbf{B}\|_{2,2} \leq \sigma,$$

Simultaneous & continuous sources



Wavefield computations

$$\overbrace{\begin{bmatrix} \mathcal{H}_{\omega_1} & 0 & & \\ 0 & \mathcal{H}_{\omega_2} & \ddots & \\ & \ddots & \ddots & 0 \\ & & 0 & \mathcal{H}_{\omega_{n_f}} \end{bmatrix}}^{\mathbf{H}} \overbrace{\begin{bmatrix} \mathbf{U}_{\omega_1} \\ \mathbf{U}_{\omega_1} \\ \vdots \\ \mathbf{U}_{n_f} \end{bmatrix}}^{\mathbf{U}} = \overbrace{\begin{bmatrix} \mathbf{B}_{\omega_1} \\ \mathbf{B}_{\omega_2} \\ \vdots \\ \mathbf{B}_{n_f} \end{bmatrix}}^{\mathbf{B}} \quad \Rightarrow \quad \mathbf{HU} = \mathbf{B}$$

- Matrix-free preconditioned indirect solver based on multilevel Krylov with deflation [Erlanga, Nabben, '08, Erlanga and F.J.H, '08]
- Solution gives multidimensional wavefield $\mathbf{u}(x_s, x_r, t)$
- Block-diagonal structure **H** and multiple rhs are amenable to CS as long as CS sampling matrix ***commutes*** with **H**
- Corresponds to simultaneous acquisition
  - replaces *impulsive* individual sources by *simultaneous* randomized sources
  - reduces number *simultaneous* sources (rhs) & *angular* frequencies (blocks)

Sparse recovery

$$\mathbf{P_1}: \begin{cases} \mathbf{y} &= \mathbf{RMd} \\ \tilde{\mathbf{x}} &= \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{y} \\ \mathbf{A} &= \mathbf{RMS}^* \\ \tilde{\mathbf{d}} &= \mathbf{S}^* \tilde{\mathbf{x}} \end{cases}$$

Challenges:

- large to extreme large system size (number of unknowns is 2^{25} for a really small problem)
- find proper subsampling matrix that is physically realizable and numerically fast
- find proper sparsifying transforms that balances ***sparsity*** with **mutual coherence**

Solver:

- bring in as many entries per iteration as possible
- projected gradient with root finding method ($\mathrm{SPG}\ell_1$, Friedlander & van den Berg, '07-'08)
- few matrix-vector multiplies
- use matrix-free implementations where possible

CS sampling matrix

Subsample along source and frequency coordinates

Use ***fast*** transform-based sampling algorithms such as ***scrambled Fourier*** [Romberg, '08] or ***Hadamard*** ensembles [Gan et. al., '08]

$$\mathbf{RM} = \overbrace{\begin{bmatrix} \mathbf{R}_1^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{R}_1^{\Omega} \\ \vdots \\ \mathbf{R}_{n_{s'}}^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{R}_{n_{s'}}^{\Omega} \end{bmatrix}}^{\text{sub sampler}} \overbrace{\left(\mathbf{F}_2^* \operatorname{diag}\left(e^{\hat{i}\boldsymbol{\theta}}\right) \otimes \mathbf{I}\right)\mathbf{F}_3}^{\text{random phase encoder}},$$

$$\theta_w = \text{Uniform}([0, 2\pi])$$

- Different random restriction for each $n_s' \ll n_s$ simultaneous experiments
- Restriction reduces system size
- Different from implementations of sampling matrices based on Kronecker-products
- Numerical complexity CS sampling

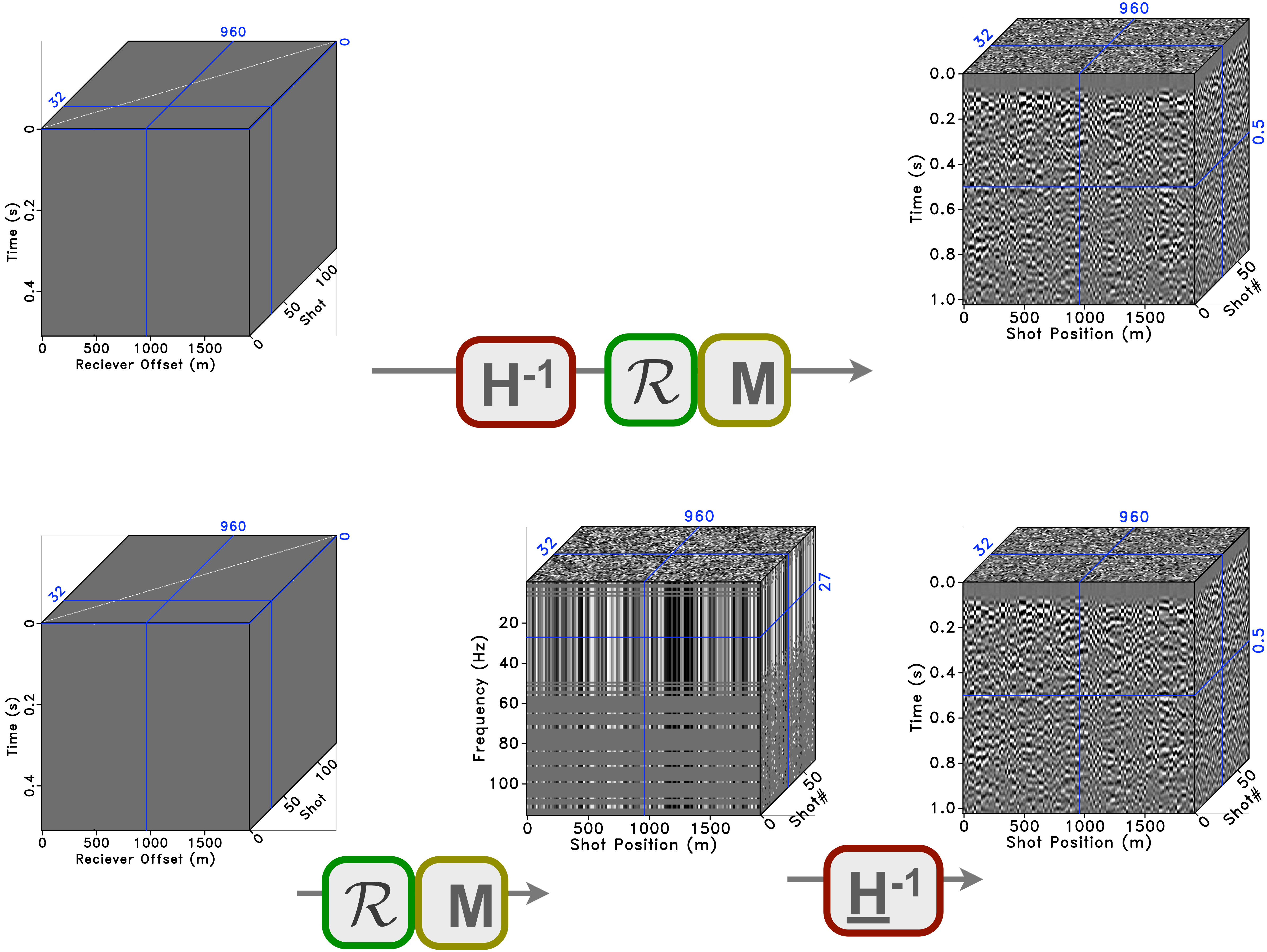
$$\boxed{\mathcal{O}(n^3 \log n)}$$

Source-solution sampling equivalence

$$\begin{cases} \mathbf{Q} = \mathbf{D}^* \underbrace{\mathbf{s}}_{\text{single shots}} \\ \mathbf{HU} = \mathbf{B} \\ \mathbf{y} = \mathbf{RMDU} \end{cases} \iff \begin{cases} \underline{\mathbf{Q}} = \underline{\mathbf{D}}^* \underbrace{\mathbf{RMs}}_{\text{simul. shots}} \\ \underline{\mathbf{HU}} = \underline{\mathbf{B}} \\ \underline{\mathbf{y}} = \underline{\mathbf{DU}} \end{cases}$$

Full data can be recovered via sparsity promotion, i.e.,

$$\mathbf{P}_1: \begin{cases} \tilde{\mathbf{x}} & = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{y} \\ \mathbf{A} & = \mathbf{RMS}^* \\ \tilde{\mathbf{d}} & = \mathbf{S}^* \tilde{\mathbf{x}} \end{cases}$$



960
32
0
Time (s)
0 0.2 0.4
0 500 1000 1500
Reciever Offset (m)
50 100
Shot
H-1
R
M
960
32
0.0 0.2 0.4 0.6 0.8 1.0
Time (s)
0 500 1000 1500
Shot Position (m)
0.5
0 50
Shot#
960
32
0
Time (s)
0 0.2 0.4
0 500 1000 1500
Reciever Offset (m)
50 100
Shot
R
M
960
32
27
20 40 60 80 100
Frequency (Hz)
0 500 1000 1500
Shot Position (m)
0 50
Shot#
H-1
960
32
0.0 0.2 0.4 0.6 0.8 1.0
Time (s)
0 500 1000 1500
Shot Position (m)
0.5
0 50
Shot#

Sparsifying transform

- Use fast discrete 2-D Curvelet transform based on wrapping [Demanet '06] along shot and receiver coordinates
  - compresses highly geometrical features of monochromatic wavefields
  - incoherent with compressive-sampling matrix that acts along the source coordinate

- Use fast discrete wavelet transform along the time coordinate
  - compresses front-like features arriving along the time direction
  - reasonable incoherent with sampling of angular frequencies

- Combine both transforms through a **Kronecker** product

$$\mathbf{S} = \mathbf{C}_{2d} \otimes \mathbf{W}$$

- Numerical complexity *sparsifying* transform

$$\mathcal{O}(n^3 \log n)$$

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

Multilevel-Krylov preconditioned [Erlangga, Nabben, FJH, '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 dominated by matrix-vector products

- Sparsity transform is $\mathcal{O}(n^3 \log n)$
- Gaussian projection $\mathcal{O}(n^3)$ per frequency
- **Cost** $\mathcal{O}(n^4)$, which does not lead to asymptotic improvement

Use fast transforms (e.g. Random Convolutions by Romberg '08)

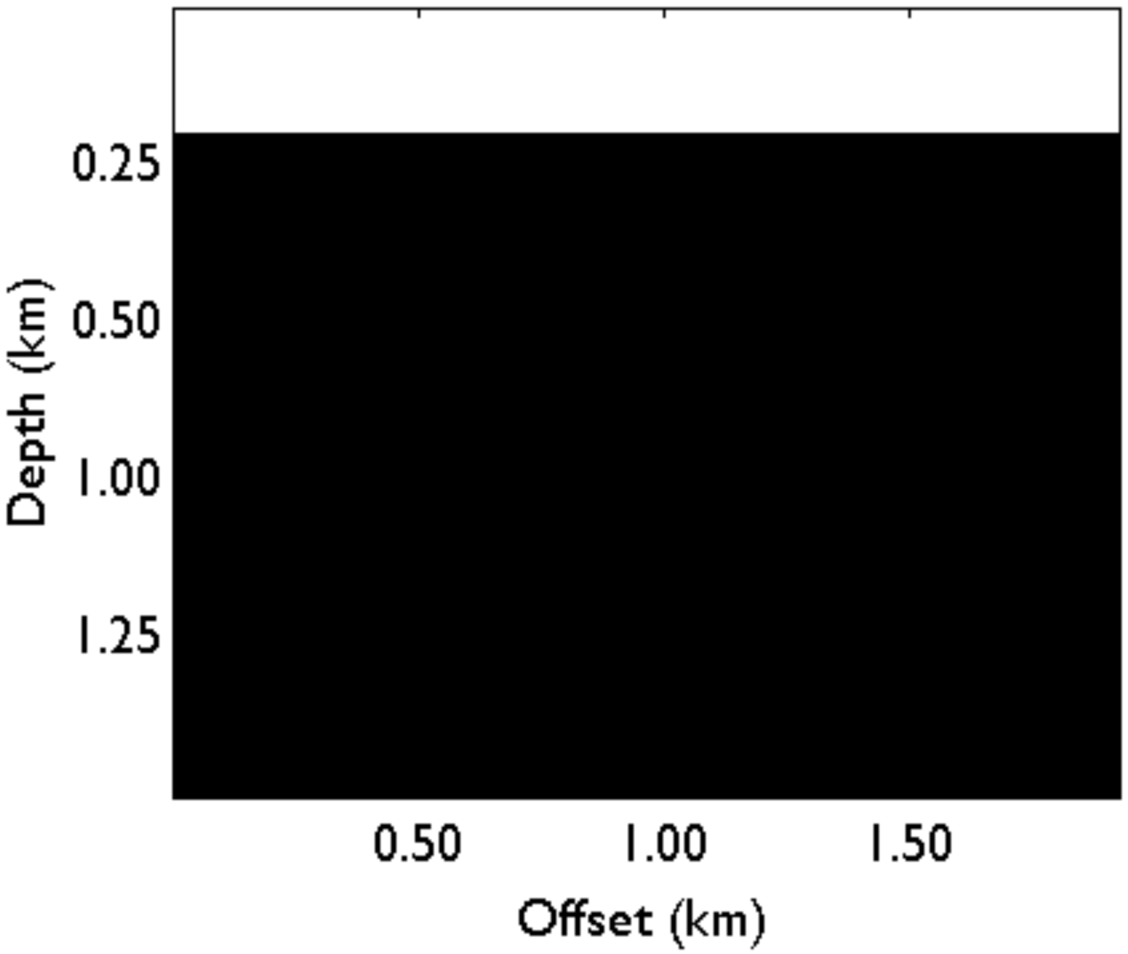
- fast projection in time & shot directions: $\mathcal{O}(n \log n)$
- **Cost** $\mathcal{O}(n^3 \log n)$ instead of $\mathcal{O}(n^4)$

Bottom line: Computational cost for the ℓ_1-solver is less $(\mathcal{O}(n^3 \log n)$ vs. $\mathcal{O}(n^4))$ than the cost of solving Helmholtz

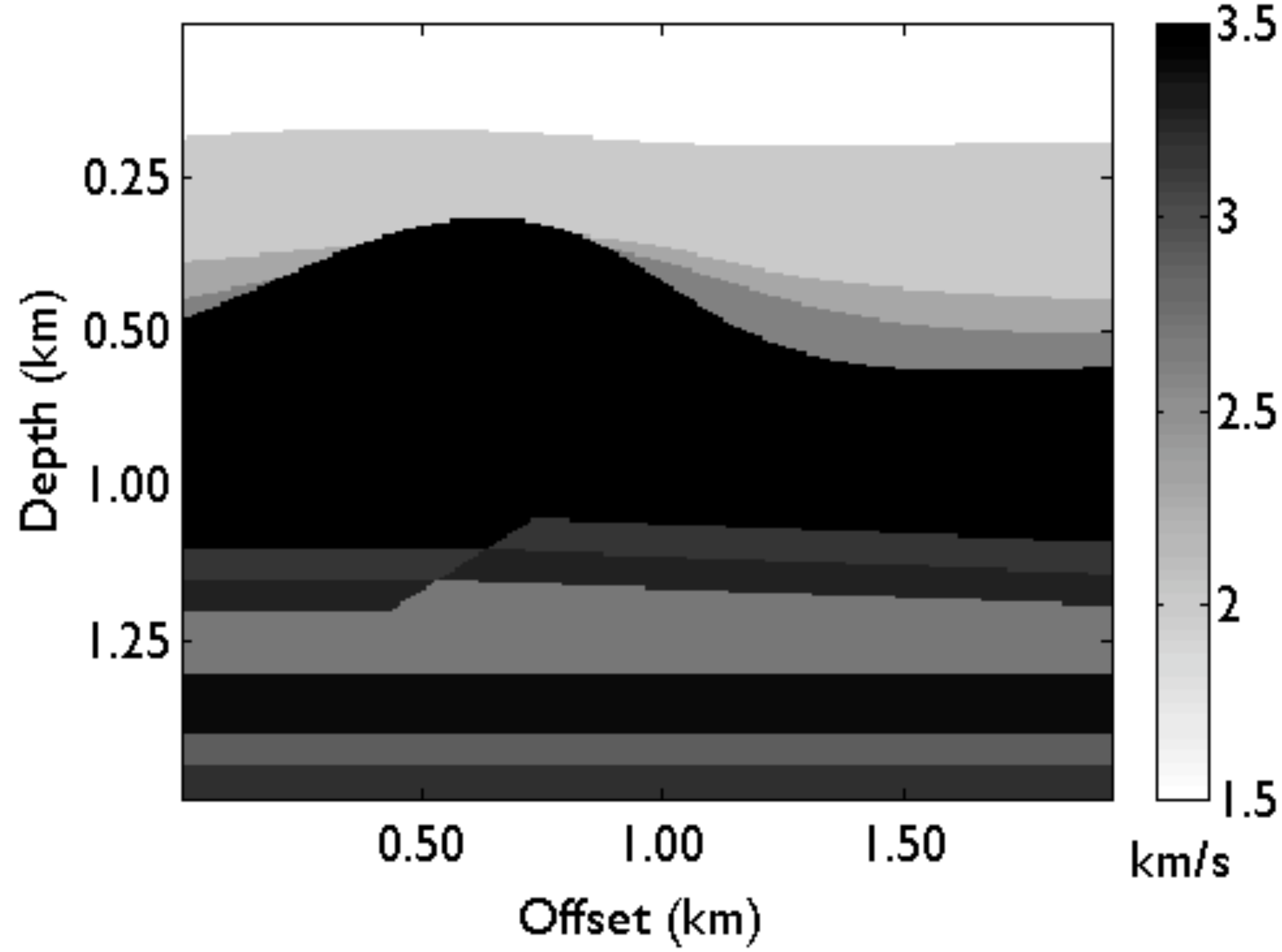
- smaller memory imprint
- cost reduction dependent on complexity = transform-domain sparsity of the solution

Velocity models

simple model

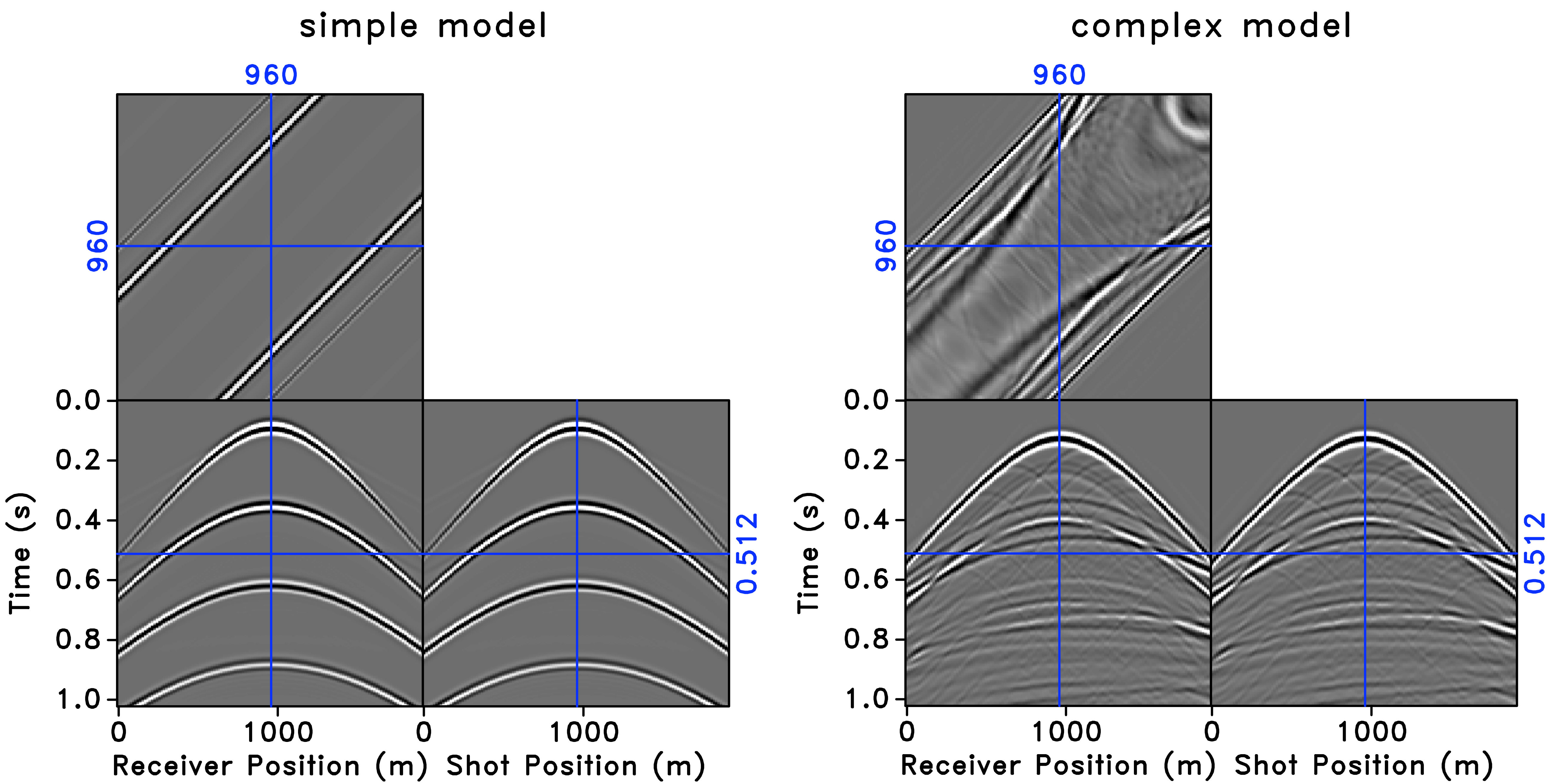


complex model



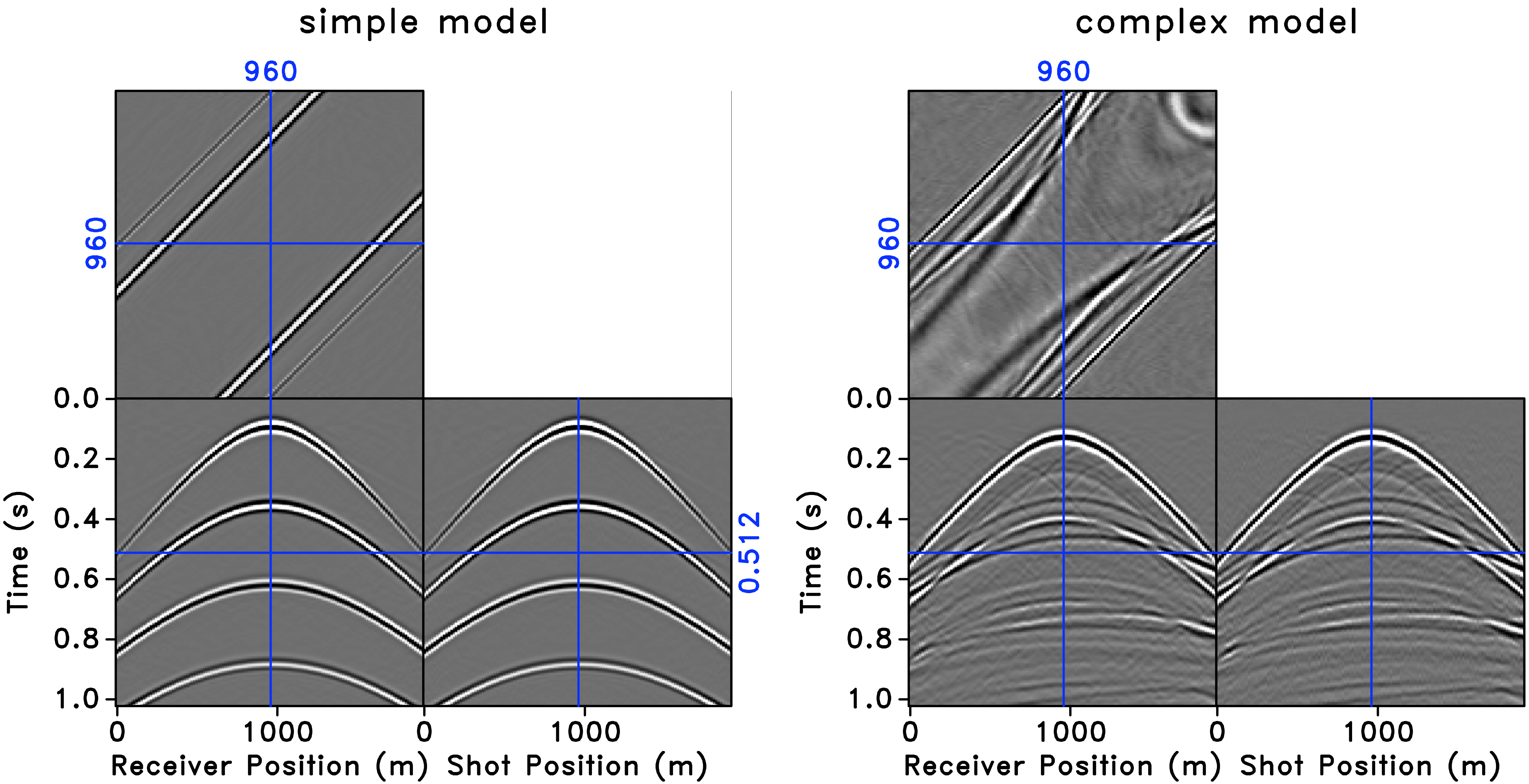
SLIM
Seismic Laboratory for Imaging and Modeling

Green’s functions



SLIM
Seismic Laboratory for Imaging and Modeling

Recovered data

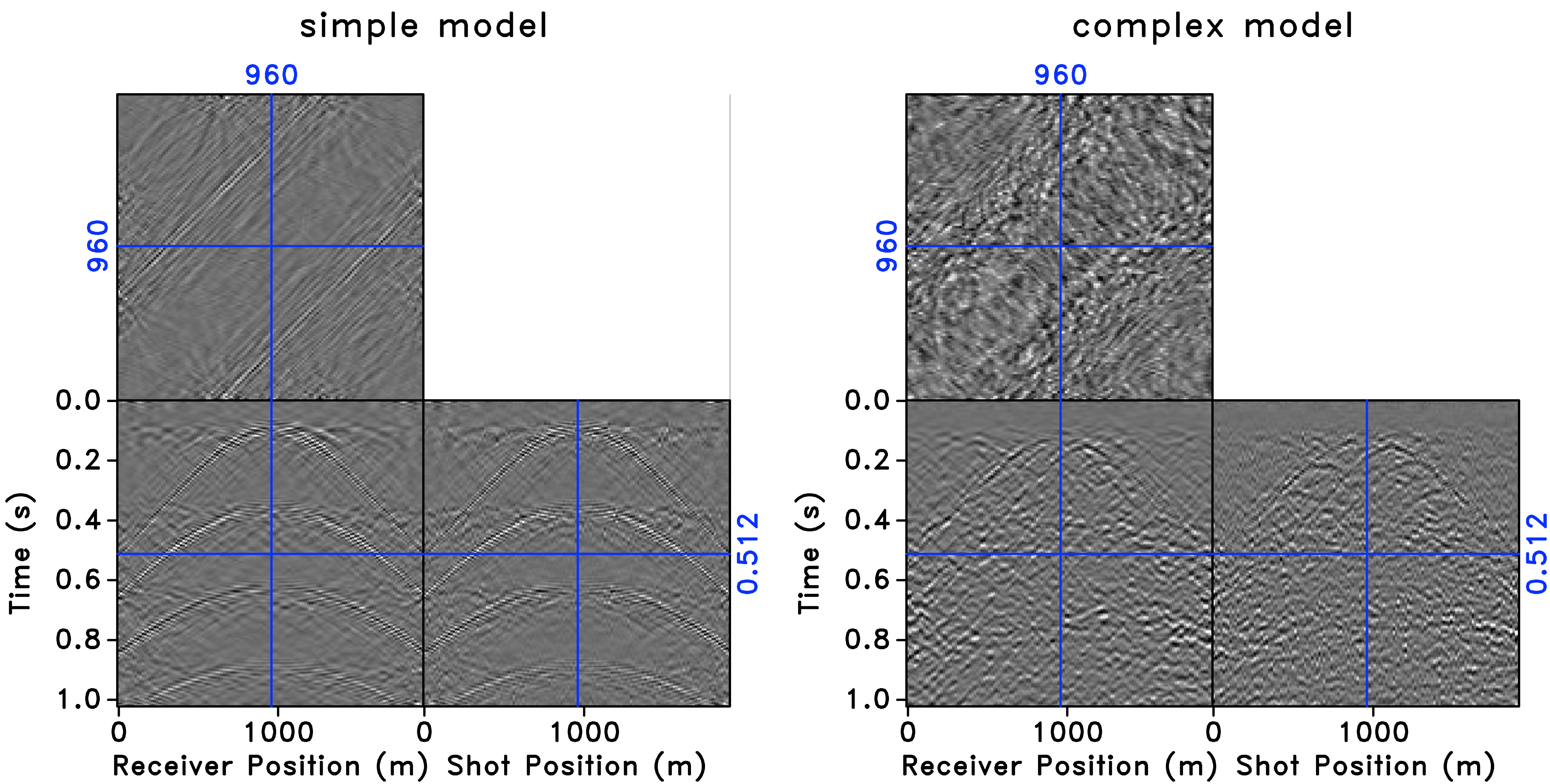


28.1dB

18.2dB

300 SPGL1 iteration

Difference



28.1dB

18.2dB

300 SPGL1 iteration

Sample ratio SNR (dB)

problem size 2^{22}

Total computed data fraction

# Frequencies / # Shots	0.25	0.15	0.07
2	14.3	12.1	8.6
1	18.2	14.5	10.2
0.5	22.2	16.5	10.7

$$\mathrm{SNR} = -20 \log \frac{\|\mathbf{d} - \tilde{\mathbf{d}}\|_2}{\|\mathbf{d}\|_2}$$

SLIM
Seismic Laboratory for Imaging and Modeling

Implications

CSed PDE *constrained* optimization problem

$$\min_{\mathbf{U}\in\mathcal{U},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{RM}\big(\mathbf{d}-\mathbf{DU}\big)\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

is equivalent to

$$\min_{\underline{\mathbf{U}}\in\underline{\mathcal{U}},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{y}-\underline{\mathbf{DU}}\|_2^2 \quad \text{subject to} \quad \underline{\mathbf{H}}[\mathbf{m}]\underline{\mathbf{U}}=\underline{\mathbf{Q}}$$

- ***reduced*** system of equations for forward modeling
- freedom to choose amount of undersampling and **M** ***commensurate*** complexity of the model
- Solutions requires **sparsity-promotion**
- CS on the model side

Implications

Add sparsity promoting ***prior***

$$\min_{\underline{\mathbf{U}}\in\underline{\mathcal{U}},\,\mathbf{x}\in\mathcal{X}} \frac{1}{2}\|\mathbf{y}-\underline{\mathbf{DU}}\|_2^2 \quad \text{subject to} \quad \underline{\mathbf{H}}[\mathbf{S}^H\mathbf{x}]\underline{\mathbf{U}}=\underline{\mathbf{Q}} \quad \wedge \quad \|\mathbf{x}\|_1\leq\tau$$

Recast into *unconstrained* optimization problem:

$$\min_{\mathbf{x}\in\mathcal{X}} \frac{1}{2}\|\mathbf{y}-\underline{\mathcal{F}}[\mathbf{x}]\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1\leq\tau$$

with

$$\underline{\mathcal{F}}[\mathbf{x}] = \underline{\mathbf{DH}}^{-1}[\mathbf{S}^H\mathbf{x}]\underline{\mathbf{Q}}$$

- requires extension of projected gradient ℓ_1 -solver to nonlinear forward map ...
- preconditioning for nonlinear operators

Challenges: there are many ...

✓ Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, ‘09]

- data cannot be explained fully
- the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

Challenges: there are many ...

✓ Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

±✓ Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, ‘09]

- data cannot be explained fully
- the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

Challenges: there are many ...

✓ Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

±✓ Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, '09]

- data cannot be explained fully
- ➡ the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\boldsymbol{\mathcal{U}},\,\mathbf{m}\in\boldsymbol{\mathcal{M}}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$\mathbf{P} \;=\;$ Total multi-source and multi-frequency data volume

$\mathbf{D} \;=\;$ Detection operator

$\mathbf{U} \;=\;$ Solution of the Helmholtz equation

$\mathbf{H} \;=\;$ Discretized multi-frequency Helmholtz system

$\mathbf{Q} \;=\;$ Unknown seismic sources

$\mathbf{m} \;=\;$ Unknown model, e.g. $c^{-2}(x)$

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\mathcal{U},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$$\begin{aligned}
\mathbf{P} &= \text{Total multi-source and multi-frequency data volume}\\
\mathbf{D} &= \text{Detection operator}\\
\mathbf{U} &= \text{Solution of the Helmholtz equation}\\
\mathbf{H} &= \text{Discretized multi-frequency Helmholtz system}\\
\mathbf{Q} &= \text{Unknown seismic sources}\\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x)
\end{aligned}$$

Seismic imaging & inversion

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{U}\in\boldsymbol{\mathcal{U}},\,\mathbf{m}\in\boldsymbol{\mathcal{M}}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

+ Free surface BC

$$\begin{aligned}
\mathbf{P} &= \text{Total multi-source and multi-frequency data volume} \\
\mathbf{D} &= \text{Detection operator} \\
\mathbf{U} &= \text{Solution of the Helmholtz equation} \\
\mathbf{H} &= \text{Discretized multi-frequency Helmholtz system} \\
\mathbf{Q} &= \text{Unknown seismic sources} \\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x)
\end{aligned}$$

Source-function & surface nonlinearity

- Estimation of source function & removal of free-surface nonlinearity are intrinsically related.

- Removal of these effects involves
  - inversion of Fredholm integral equation of the first kind => full matrices
  - blind deconvolution problem

- Based on the following monochromatic expression:

unknown source function

$$\overbrace{\widehat{\mathbf{P}}_i}^{\text{upgoing wavefield}} \approx \underbrace{\widehat{\mathbf{G}}_i}_{\text{surface-free impulse response}} \overbrace{\left[\widehat{\mathbf{Q}_i} - \widehat{\mathbf{P}_i}\right]}^{\text{downgoing wavefield}}$$

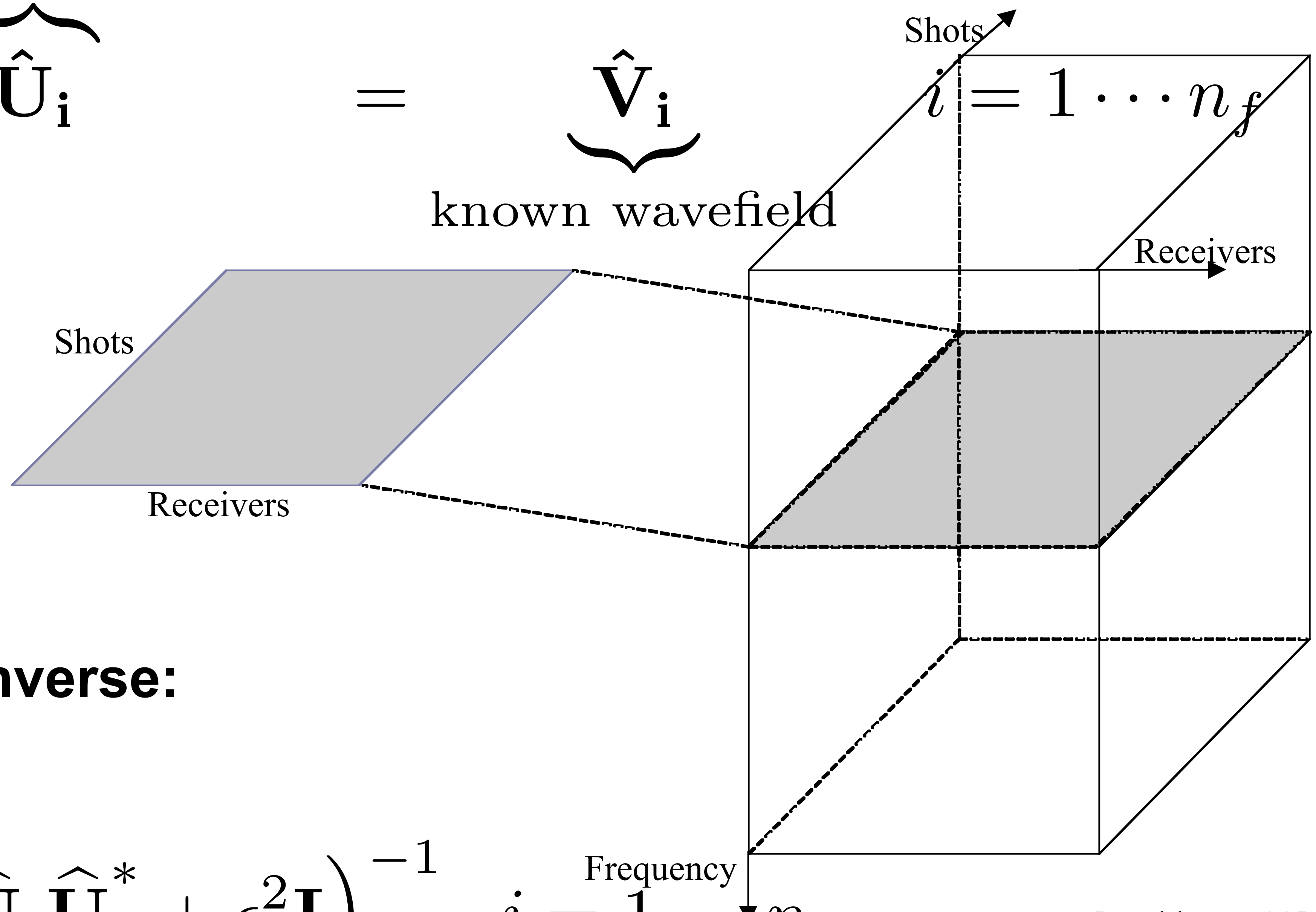
unknown source function

- "Informed" blind deconvolution problem

Common approach: damped least-squares

Monochromatic **forward model:**

$$\underbrace{\hat{\mathbf{G}}_\mathbf{i}}_{\text{unknown "image"}}\ \overbrace{\hat{\mathbf{U}}_\mathbf{i}}^{\text{to be inverted wavefield}} = \underbrace{\hat{\mathbf{V}}_\mathbf{i}}_{\text{known wavefield}} \qquad i = 1 \cdots n_f$$



Monochromatic **pseudo-inverse:**

$$\widetilde{\widehat{\mathbf{G}}}_i = \widehat{\mathbf{V}}_i \widehat{\mathbf{U}}_i^* \left(\widehat{\mathbf{U}}_i \widehat{\mathbf{U}}_i^* + \epsilon_i^2 \mathbf{I}\right)^{-1},\ i = 1 \cdots n_f,$$

[Berkhout ‘82]
[F.J.H ’07-’08]
[Wapenaar ‘08]

Curvelet-based wavefield inversion (CWI)

Cast into rigorous *linear-algebra* framework, i.e.

$$\widehat{\mathbf{G}}_i \widehat{\mathbf{U}}_i \approx \widehat{\mathbf{V}}_i,\; i = 1 \cdots n_f,$$

which with the Kronecker identity

$$\mathrm{vec}\left(\mathbf{AXB}\right) = \left(\mathbf{B}^H \otimes \mathbf{A}\right) \mathrm{vec}\left(\mathbf{X}\right)$$

becomes for each *frequency*

$$\left(\widehat{\mathbf{U}}_i^* \otimes \mathbf{I}\right) \mathrm{vec}\left(\widehat{\mathbf{G}}_i\right) \approx \mathrm{vec}\left(\widehat{\mathbf{V}}_i\right),\; i = 1 \cdots n_f,$$

Set up a system for ***all*** *frequencies* and incorporate the *temporal* ***Fourier*** transform ....

Curvelet-based wavefield inversion (CWI)

$$\overbrace{\begin{bmatrix} \left(\widehat{\mathbf{U}}_1^* \otimes \mathbf{I}\right) & & \\ & \ddots & \\ & & \left(\widehat{\mathbf{U}}_{n_f}^* \otimes \mathbf{I}\right) \end{bmatrix}}^{\mathbf{A}} \mathbf{F}_t \overbrace{\begin{bmatrix} \mathrm{vec}\left(\mathbf{G}_1\right) \\ \vdots \\ \mathrm{vec}\left(\mathbf{G}_{n_t}\right) \end{bmatrix}}^{\mathbf{x}} \approx \overbrace{\begin{bmatrix} \mathrm{vec}\left(\widehat{\mathbf{V}}_1\right) \\ \vdots \\ \mathrm{vec}\left(\widehat{\mathbf{V}}_{n_f}\right) \end{bmatrix}}^{\mathbf{b}}$$

with $\mathbf{F}_t = (\mathbf{I} \otimes \mathbf{I} \otimes \mathcal{F})$ (*temporal* Fourier transform)

Linear system is

- conducive to curvelet-based wavefield *inversion* with *sparsity* promotion
- versatile
- conducive to compressive subsampling (e.g. simultaneous acquisition)

Estimation of primaries by sparse inversion (EPSI)

- **Forward model:**

$$\overbrace{\widehat{\mathbf{P}}_i}^{\text{upgoing wavefield}} \approx \underbrace{\widehat{\mathbf{G}}_i}_{\text{surface-free impulse response}} \overbrace{\left[\widehat{\mathbf{Q}}_i - \widehat{\mathbf{P}}_i\right]}^{\text{downgoing wavefield}}$$

P total upgoing data

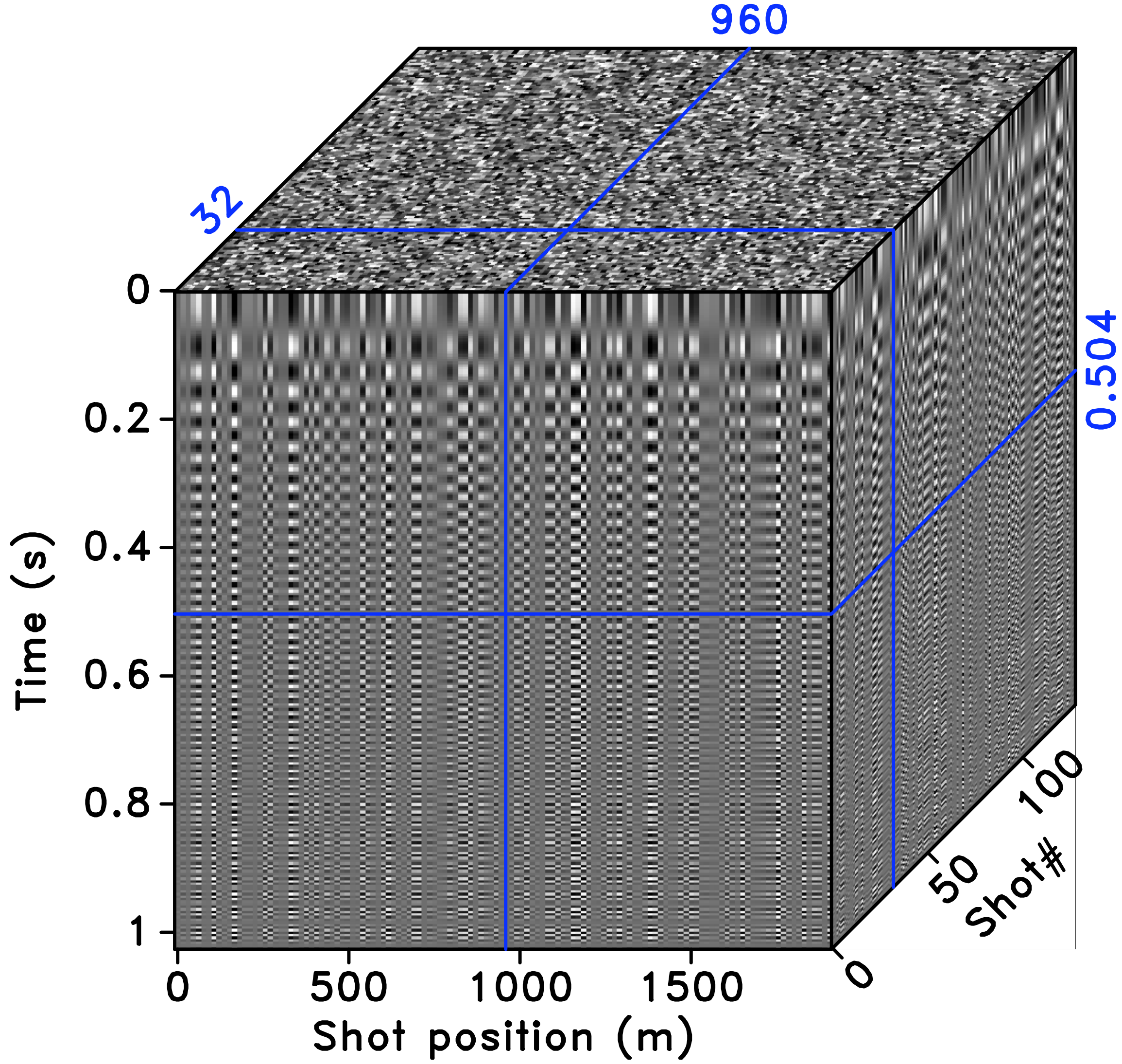
Q the source function

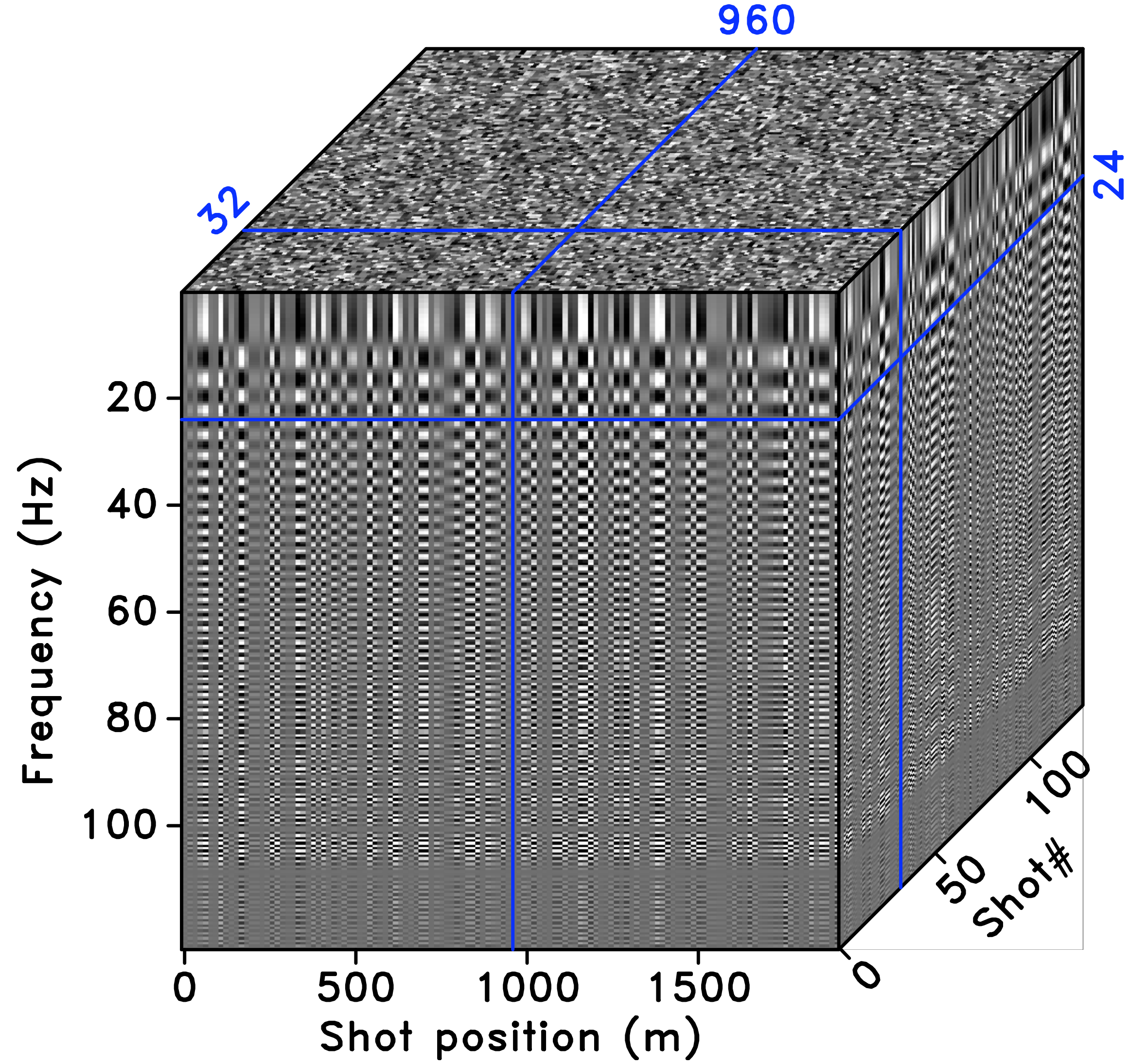
- ***Randomized*** **simultaneous acquisition:**

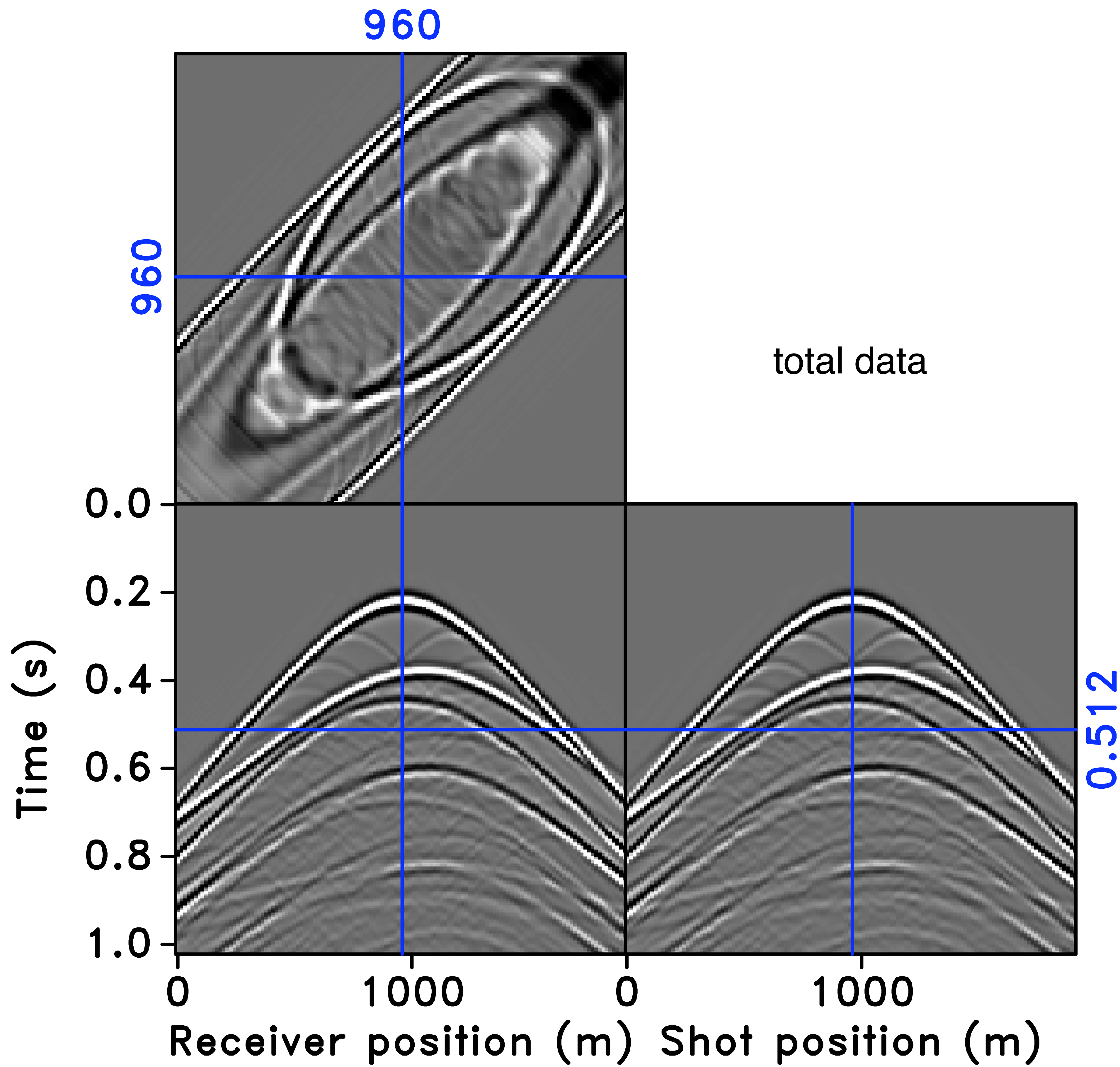
$$\mathbf{RM} = \overbrace{\left[\mathbf{R}^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{I}\right]}^{\text{sub sampler}} \overbrace{\left(\mathbf{F}_{\mathrm{s}}^{*}\mathrm{diag}\left(e^{\hat{i}\boldsymbol{\theta}}\right)\mathbf{F}_{\mathrm{s}} \otimes \mathbf{I} \otimes \mathbf{I}\right)}^{\text{random phase encoder}}$$

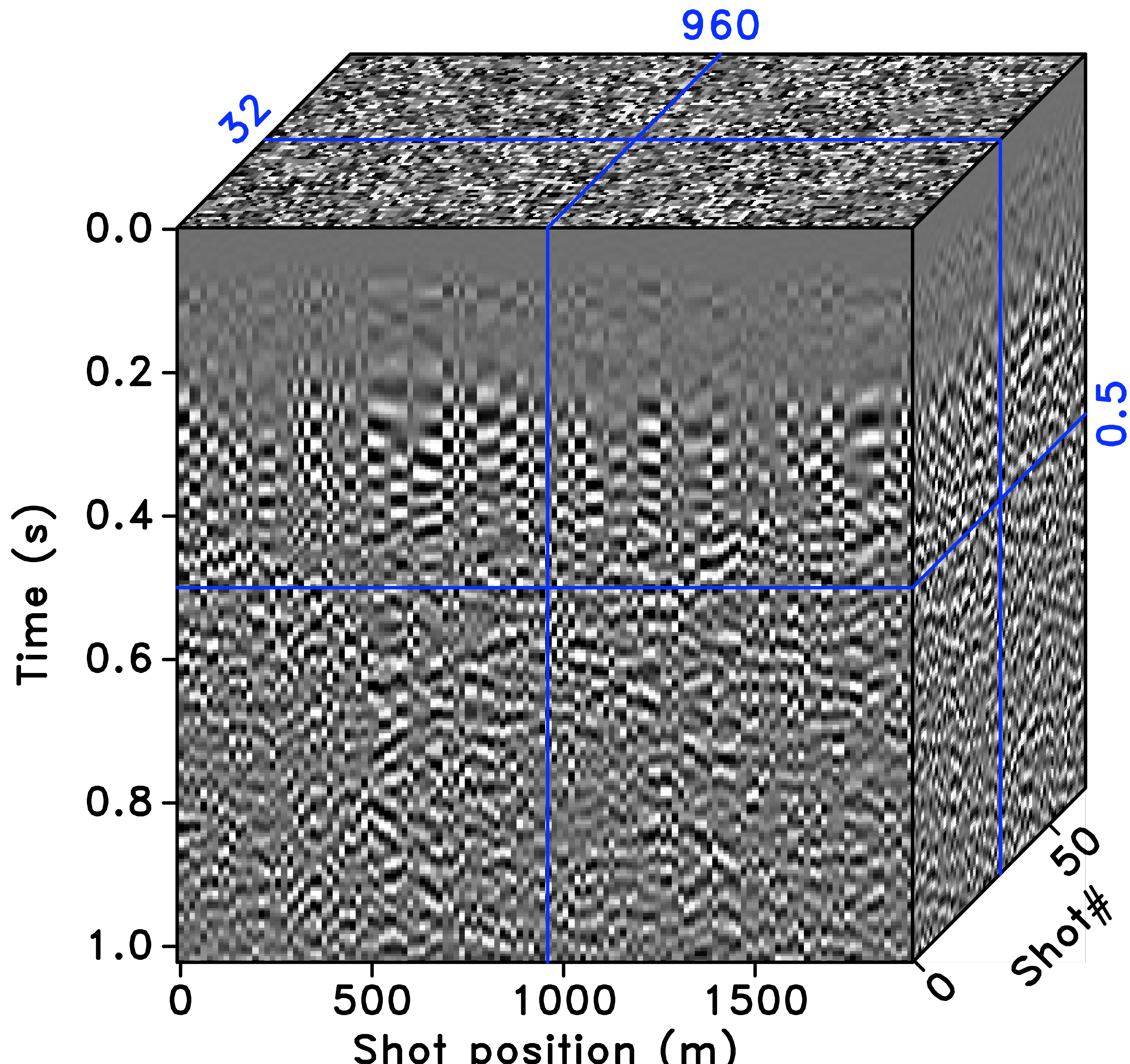
Randomized simultaneous sweep signals

- Linearly ramping seismic sweep, 5 to 110 Hz
- Simultaneous source at all positions, each randomly phase encoded

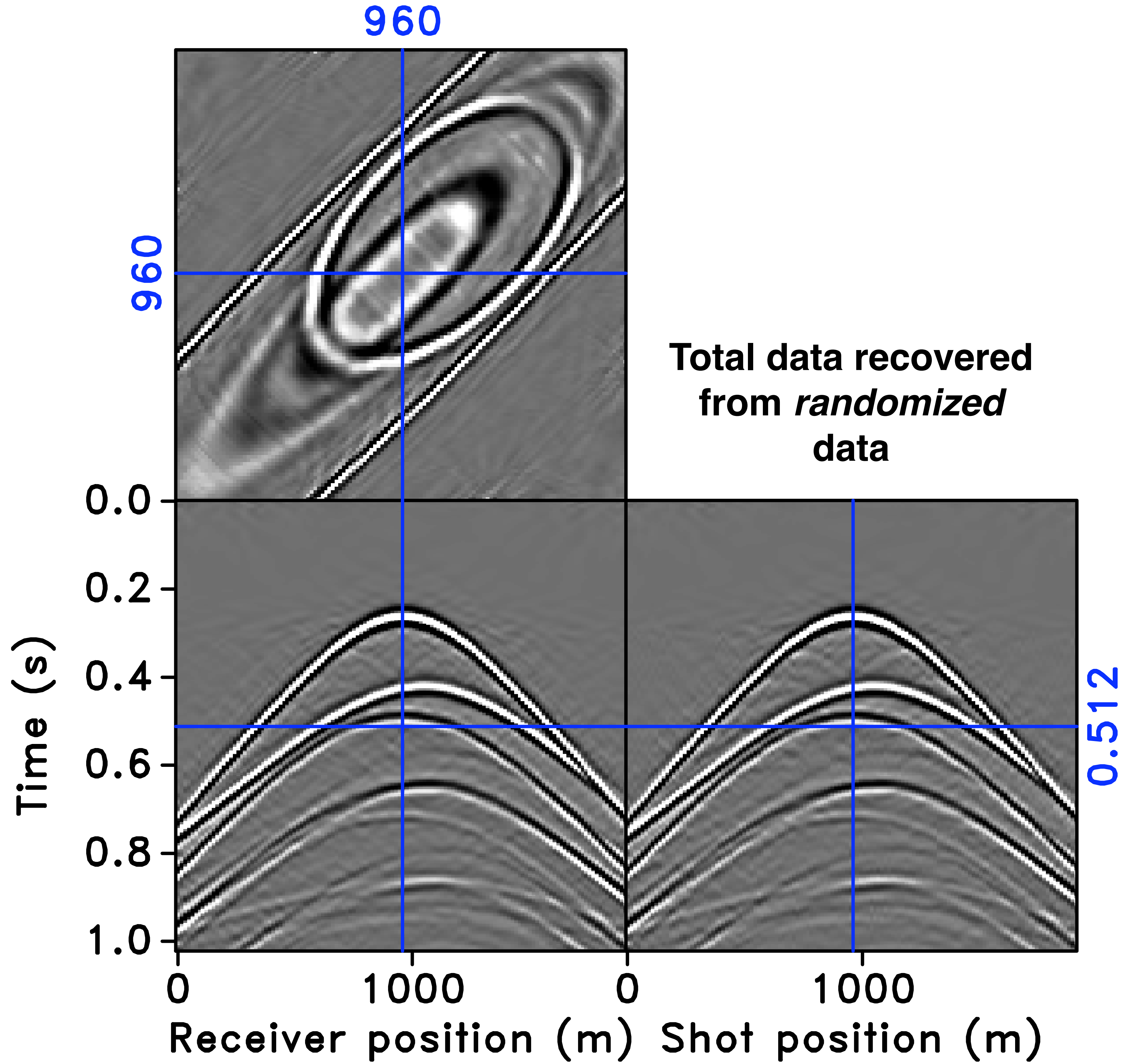




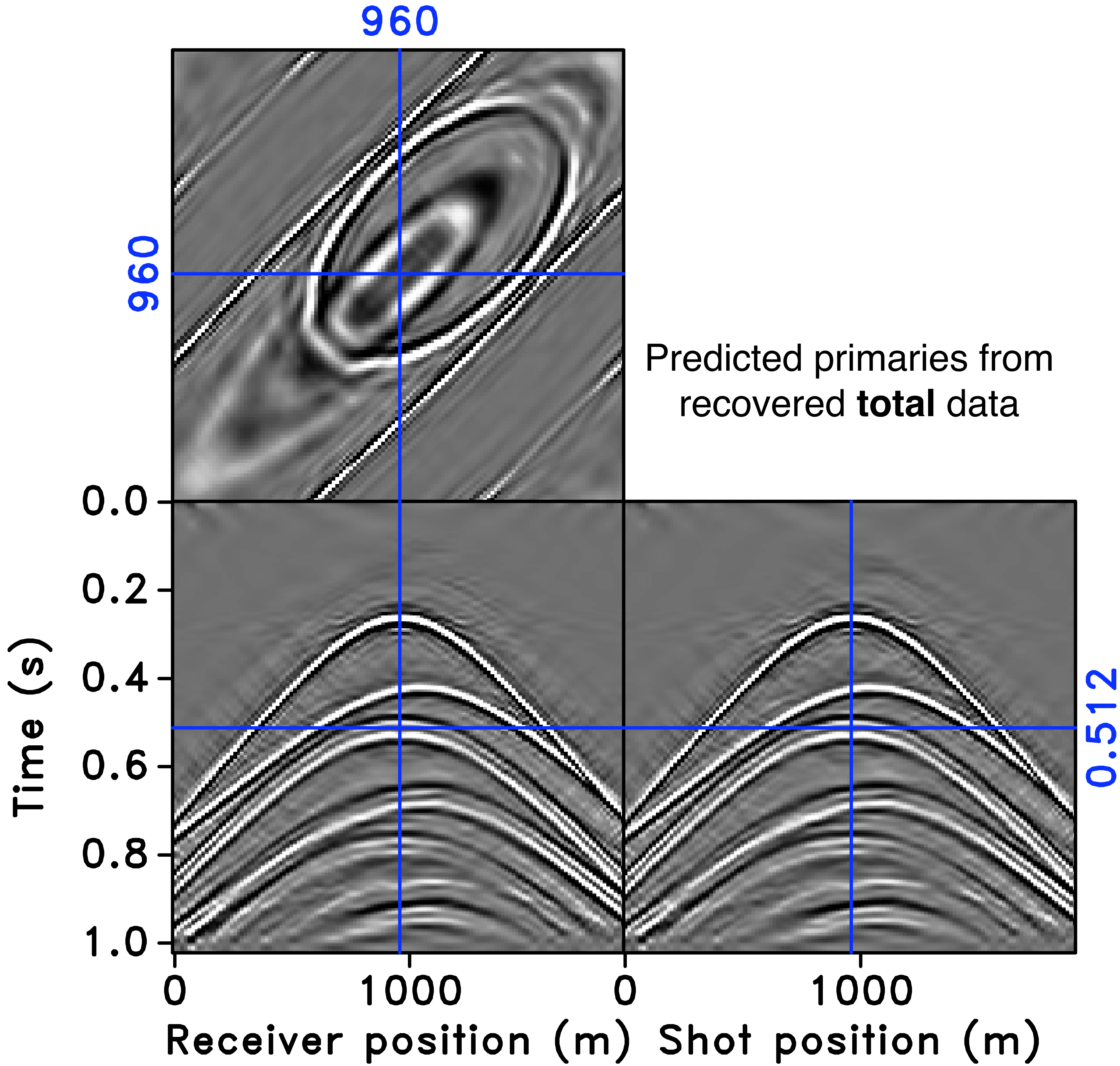

960
960
total data
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
0.512
0
1000
0
1000
Receiver position (m)
Shot position (m)



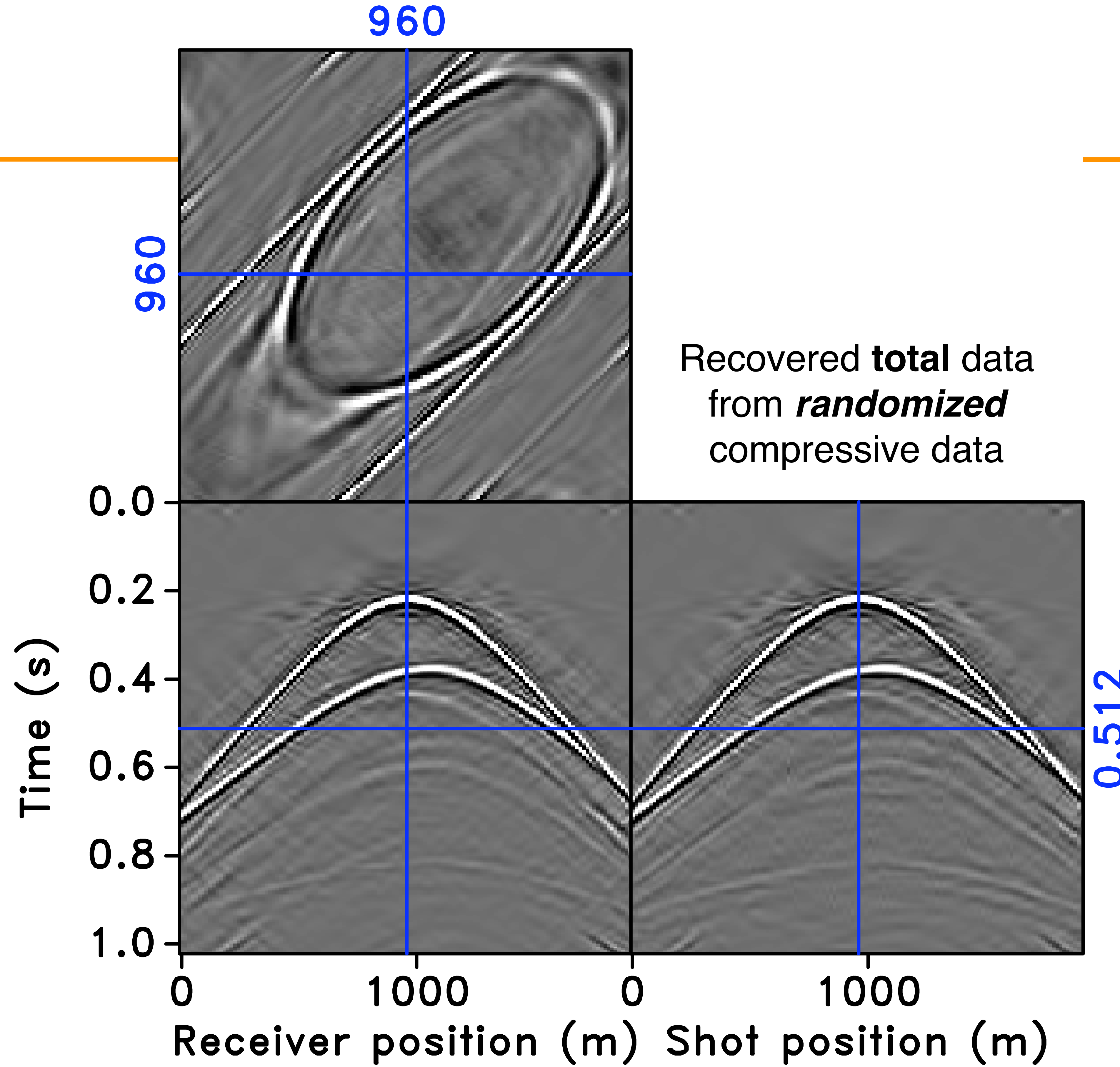
960
32
0.5
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
0
500
1000
1500
Shot position (m)
0
50
Shot#



Total data recovered from *randomized* data



Predicted primaries from recovered **total** data



Recovered **total** data from ***randomized*** compressive data

Implication

Multiexperiment **PDE-constrained optimization problem:**

$$\min_{\mathbf{G}\in\mathcal{G},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathrm{vec}\big(\mathbf{P}\big) - \mathbf{A}\mathrm{vec}\big(\mathbf{DG}\big)\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{G} = \mathbf{I}$$

$$\begin{aligned}
\mathbf{P} &= \text{Total multi-source and multi-frequency upgoing data volume} \\
\mathbf{A} &= \text{Matrix operator representation of downgoing wavefield} \\
\mathbf{D} &= \text{Detection operator} \\
\mathbf{G} &= \text{Solution of the surface-free Helmholtz equation} \\
\mathbf{H} &= \text{Discretized multi-frequency Helmholtz system} \\
\mathbf{I} &= \text{Delta Dirac} \\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x)
\end{aligned}$$

Challenges: there are many ...

✓ Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, '09]

- data cannot be explained fully
- the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

Challenges: there are many ...

✓ Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

±✓ Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, ‘09]

- data cannot be explained fully
- the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

Challenges: there are many ...

✓ Helmholtz system is ***indefinite*** & ***ill conditioned*** => lack of convergence ***indirect*** Krylov solvers

±✓ Multiexperiment setup with ***multiple*** *right-hand-sides* is *computationally* **prohibitive** as part of ***iterative*** Newton methods

Inversion problem can be both ***over-*** and ***underdetermined*** [Symes, ‘09]

- data cannot be explained fully
- ±✓ the source function is unknown & surface causes large nonlinearity
- there are local minima, many velocity models may explain data within some error

Proposed ideas to tackle *multimodality* by ***extensions & focusing*** make the situation worse by additional ***degrees of freedom***

System-size reduction

- Apply CS to reduce cost of ***wavefield*** simulation with Helmholtz
  - use ***simultaneous sources*** instead of ***separated sources***
  - leverage transform-domain sparsity & randomized subsampling by **one-norm sparsity promotion**
  - reduce size Helmholtz system
    - sources (number of right-hand sides)
    - angular frequencies (number of blocks)

- Apply CS to reduce cost of computing ***image volumes*** by multi-dimensional correlations via ***explicit*** matrix-matrix multiplies
  - randomize and subsample wavefields in **model space**
  - leverage transform-domain sparsity and focusing in the ***model space*** by ***joint sparsity promotion with mixed (1,2) norms***
  - reduce costs of storage and explicit matrix-matrix multiplies
    - sources (right-hand sides), receivers, depth
    - angular frequencies (blocks)

Tools

Compressive sensing based on Johnson-Lindenstrauss embeddings

- *Compressive sensing* [Donoho, 06‘, Candes, Romberg, Tao, ‘06]

$$\begin{aligned} \mathbf{b} &= \mathbf{RMx} \quad \textbf{[randomized subsampling]} \\ \tilde{\mathbf{x}} &= \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{RMx}-\mathbf{b}\|_2 \leq \sigma \\ \tilde{\mathbf{x}} &\approx \mathbf{x} \end{aligned}$$

Fast matrix computations based on Johnson-Lindenstrauss embeddings

- Improved Approximation Algorithms for Large Matrices via Random Projections by Tamás Sarlós, ’08

$$\mathbf{AB} \approx \mathbf{A}\left(\mathbf{RM}\right)^{*}\left(\mathbf{RM}\right)\mathbf{B}$$

Joint sparsity-promotion with mixed (1,2) norm minimization

- *Joint-sparse recovery from multiple measurements* by E. van den Berg and M. Friedlander, ‘09

$$\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{1,2} \quad \text{subject to} \quad \|\mathbf{AX}-\mathbf{B}\|_{2,2} \leq \sigma,$$

Differential semblance

- Invoke *physical* principle of ***focusing*** [Claerbout & many others] <=> *mathematical* principle of ***extensions*** [Symes '09]

- Motivated by Symes' differential semblance principle [Symes '09]: "*Amongst all possible quadratic forms in the data, parameterized by velocity, of the form*

$$\min_{\mathbf{m}} \|(\mathsf{P}_h \overbrace{\boldsymbol{\delta}\mathbf{I}(\cdot,\ \underset{\substack{\uparrow \\ \text{redundant coordinate}}}{h}; \mathbf{m}, \delta\mathbf{d})}^{\text{image volume}})\|_2 \quad \text{with} \ \overbrace{\mathsf{P}_h \cdot = \mathbf{h}\cdot}^{\text{annihilator}},$$

only differential semblance is smooth jointly as function of smooth perturbations in velocity and finite energy perturbations in data [Stolk & Symes, '03]"

- Forms the basis of ***nonlinear*** migration velocity analysis on ***linearized*** data [Symes, '09].

Image volume

Compute multi-D ***cross-correlations*** on ***multiexperiment*** solutions of the forward- and reverse-time Helmholtz systems--i.e,

$$\boldsymbol{\delta}\mathbf{I}(m,h,t) = \left(\bar{\mathbf{U}} * \mathbf{V}^T\right)$$

with

$$\mathbf{U}_f = \left[\mathbf{u}_1 \cdots \mathbf{u}_{n_f}\right] \text{ and } \mathbf{V}_f = \left[\mathbf{v}_1 \cdots \mathbf{v}_{n_f}\right]$$

and

$$\left(\bar{\mathbf{U}} * \mathbf{V}^T\right) := \mathbf{T}_{(x_s,x_r,\omega)\mapsto(m,h,t)} \begin{bmatrix} \bar{\mathbf{U}}_1 & & \\ & \ddots & \\ & & \bar{\mathbf{U}}_{n_f} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \vdots \\ \mathbf{V}_{n_f}^T \end{bmatrix}$$

where

$$m = \frac{1}{2}(x_s + x_r) \quad \text{and} \quad h = \frac{1}{2}(x_s - x_r)$$

High **dimensional** and highly ***redundant ...***

Imaging condition

Claerbout's imaging principle:

$$\begin{aligned}\delta\mathbf{m} &= \boldsymbol{\delta}\mathbf{I}(\cdot, h = 0, t = 0) \\ &= \mathbf{K}^*\boldsymbol{\delta}\mathbf{d}\end{aligned}$$

- implicit in adjoint state method
- Image volume
  - very large because of additional degree of freedom
  - expensive to store

System-size reduction by CS

For each angular frequency, subsample with CS matrix

$$\mathbf{RM} := \overbrace{\begin{bmatrix} \mathbf{R}_1^{\sigma} \otimes \mathbf{R}_1^{\rho} \otimes \mathbf{R}_1^{\zeta} \\ \vdots \\ \mathbf{R}_{n'_f}^{\sigma} \otimes \mathbf{R}_{n'_f}^{\rho} \otimes \mathbf{R}_{n'_f}^{\zeta} \end{bmatrix}}^{\text{sub sampler}} \overbrace{\left(\mathbf{F}_3^* \left(e^{\hat{i}\boldsymbol{\theta}}\right)\right) \mathbf{F}_3}^{\text{random phase encoder}} ,$$

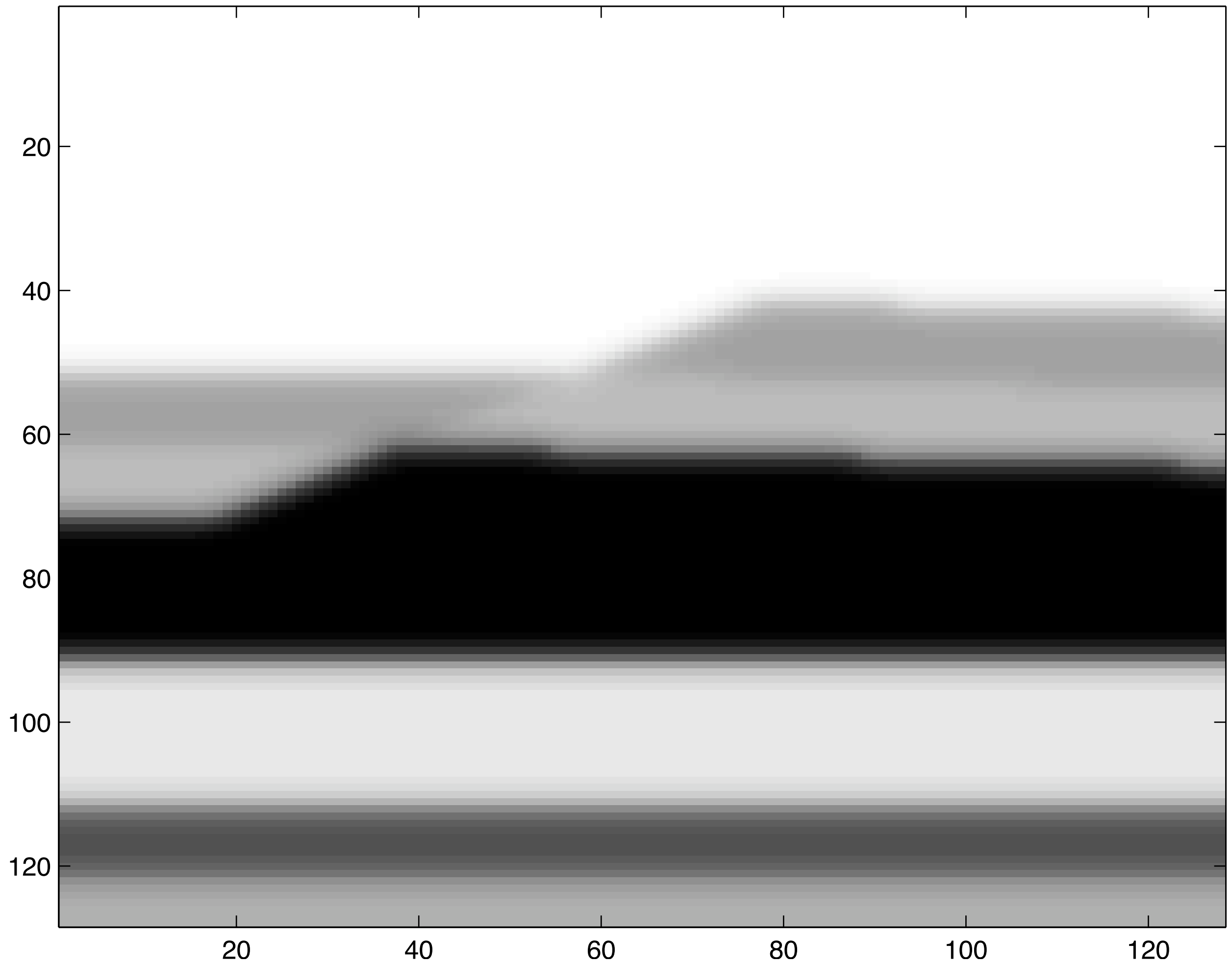
with

$$n'_f \times n'_\sigma \times n'_\rho \times n'_\zeta \ll n_f \times n_s \times n_r \times n_z$$

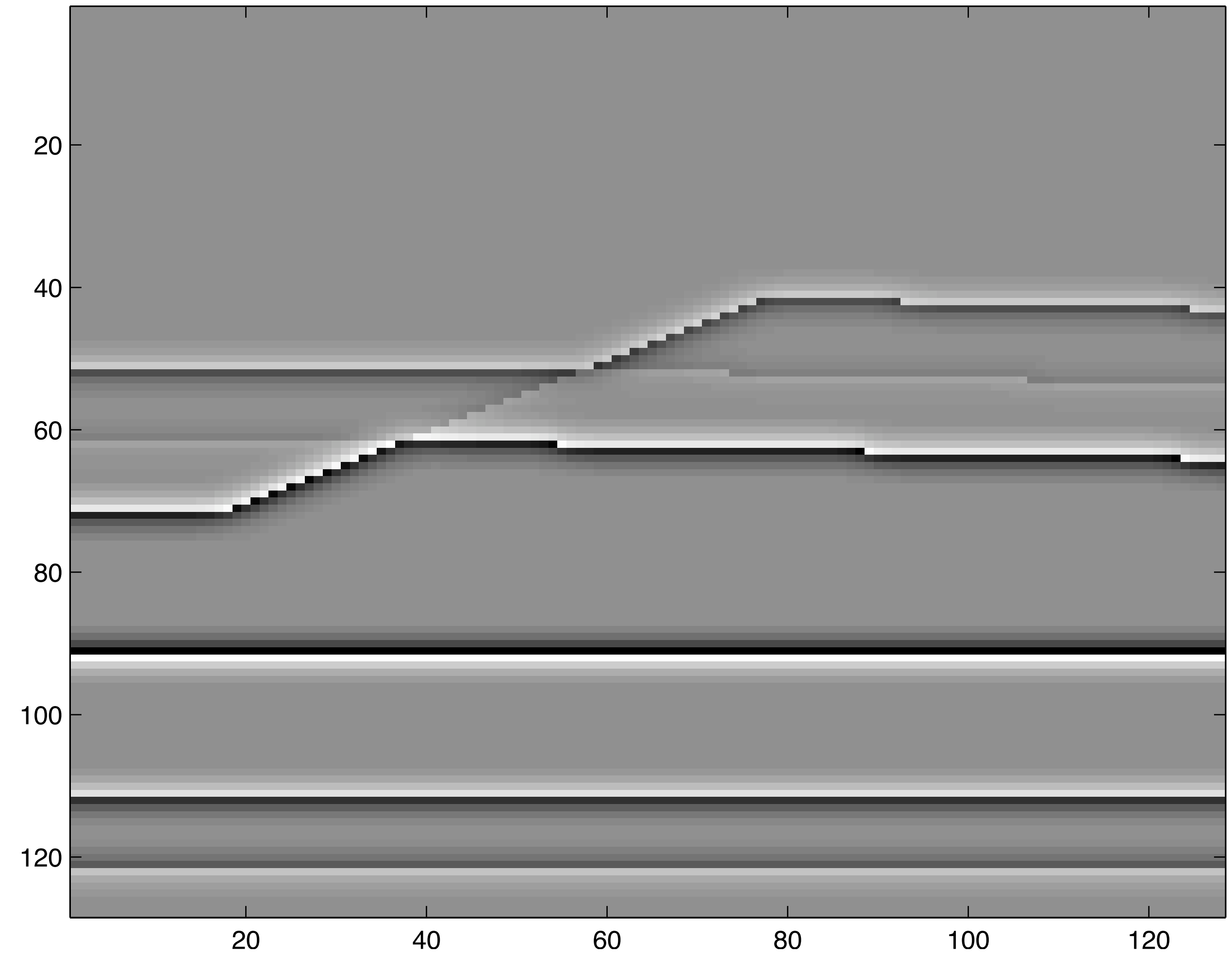
Model-space CS subsampling along subsurface source, receiver, and depth coordinates yielding an *approximate* **extended** image

$$\boldsymbol{\delta}\mathbf{I}(m,h,t) \approx \left(\bar{\mathbf{U}}(\mathbf{RM})^* * \mathbf{RMV}^T\right)$$

Example

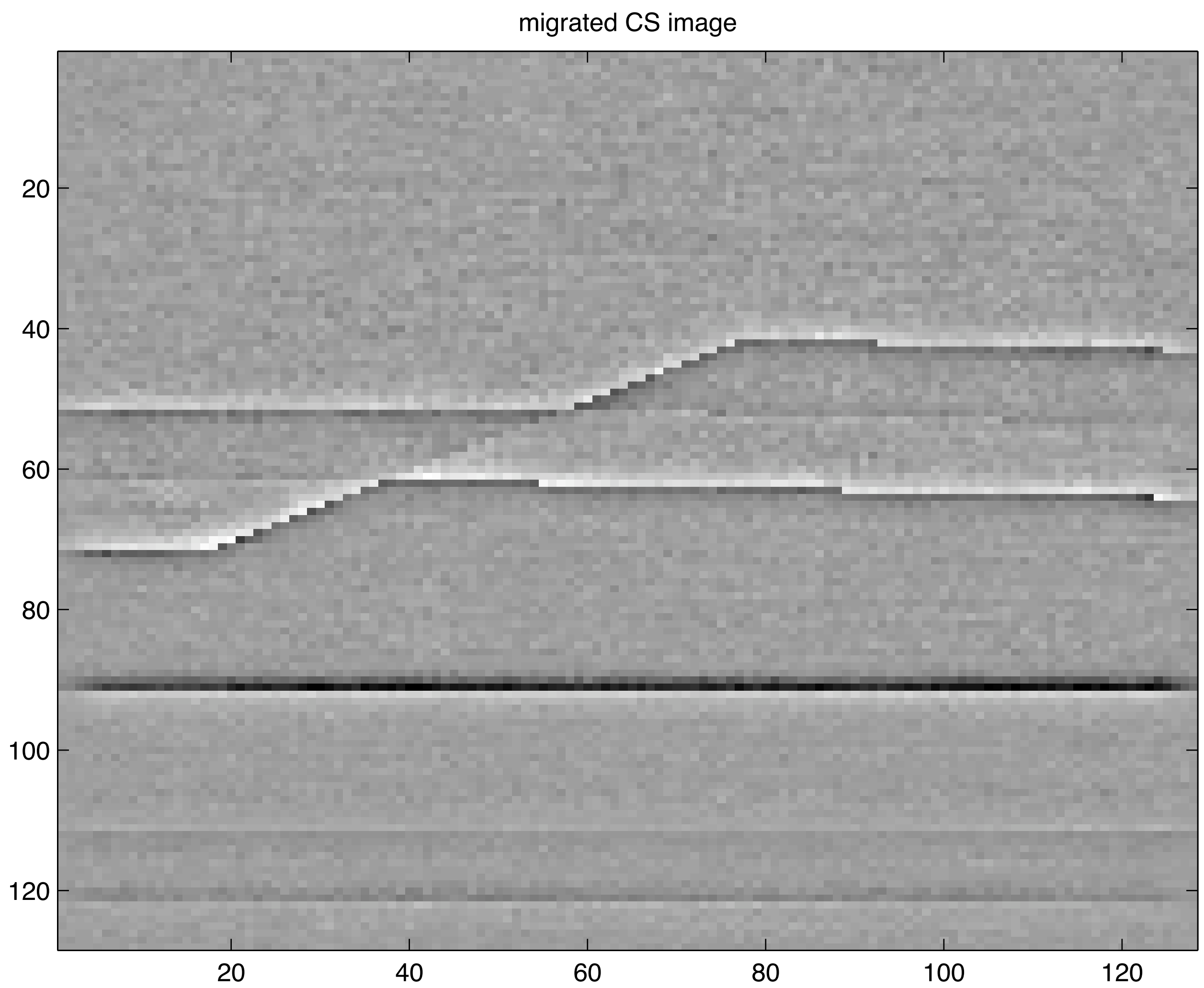


background velocity model

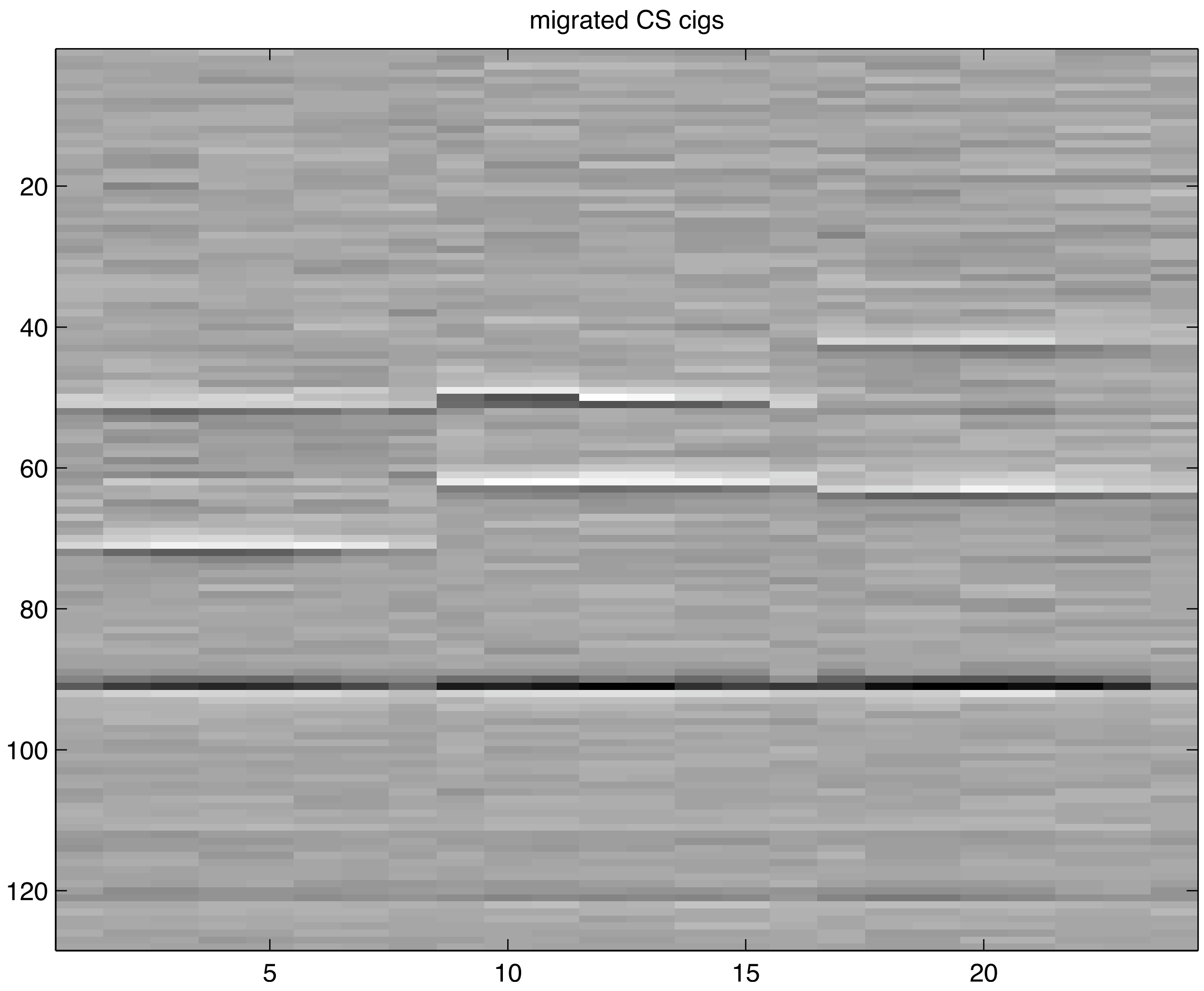


perturbation

Example: matched filter



$\boldsymbol{\delta}\mathbf{I}(\cdot, h = 0, t = 0)$



$\boldsymbol{\delta}\mathbf{I}([m_1,\ m_2,\ m_3], h, t = 0)$

Recovery from 64-fold subsampling ...

- **Noisy**
- **Not focused**

Tools

Compressive sensing based on Johnson-Lindenstrauss embeddings

- *Compressive sensing* [Donoho, 06‘, Candes, Romberg, Tao, ‘06]

$$
\begin{aligned}
\mathbf{b} &= \mathbf{RMx} \quad \textbf{[randomized subsampling]} \\
\tilde{\mathbf{x}} &= \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{RMx}-\mathbf{b}\|_2 \leq \sigma \\
\tilde{\mathbf{x}} &\approx \mathbf{x}
\end{aligned}
$$

Fast matrix computations based on Johnson-Lindenstrauss embeddings

- Improved Approximation Algorithms for Large Matrices via Random Projections by Tamás Sarlós, ’08

$$\mathbf{AB} \approx \mathbf{A}\left(\mathbf{RM}\right)^{*}\left(\mathbf{RM}\right)\mathbf{B}$$

Joint sparsity-promotion with mixed (1,2) norm minimization

- *Joint-sparse recovery from multiple measurements* by E. van den Berg and M. Friedlander, ‘09

$$\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{1,2} \quad \text{subject to} \quad \|\mathbf{AX}-\mathbf{B}\|_{2,2} \leq \sigma,$$

Extended Born & focusing

Define ***extended*** *linearized* forward model [Symes, '09]:

$$\bar{\mathbf{K}}[\mathbf{m},\mathbf{Q}]\boldsymbol{\delta}\mathbf{I} \approx \boldsymbol{\delta}\mathbf{D}$$

- multiexperiment form *amenable* for ***joint sparsity promotion***
- introduce penalty term that penalizes ***defocusing***

Form augmented system with **focusing**:

$$\begin{aligned} \bar{\mathbf{K}}\boldsymbol{\delta}\mathbf{I} &\approx \boldsymbol{\delta}\mathbf{D} && \text{data fit} \\ \lambda^2 \mathsf{P}_h \boldsymbol{\delta}\mathbf{I} &\approx \mathbf{0} && \text{focusing} \end{aligned}$$

with $\mathsf{P}_h\cdot = \mathbf{h}\cdot$ *annihilator* that increasingly *penalizes* non-zero offsets.

Solution involves multi-D "deconvolution" (adjoint of cross correlation):

$$\left(\mathbf{U}^* \star \boldsymbol{\delta}\mathbf{I}\right) \approx \mathbf{V}^T$$

Compressed linearized inversion

Compressively sample augmented system that includes sparsity synthesis operator--i.e,

$$\begin{aligned} \mathbf{RM}\left(\mathbf{U}^* \star \mathbf{S}^*\mathbf{X}\right) &\approx \mathbf{RMV}^T \\ \mathsf{P}_h\mathbf{X} &\approx \mathbf{0} \end{aligned} \quad \text{or} \quad \mathbf{AX} \approx \mathbf{B}$$

with the sparsifying transform **S** for each offset h given by the curvelet or wavelet transform

Recover focused solution by mixed (1,2)-norm minimization.

Promote sparsity amongst images though one-norm on columns

Penalize energy amongst rows => focusing

Joint-sparsity promotion [van den berg & Friedlander, '09]

Recover focused solution by mixed (1,2)-norm minimization:

$$\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{1,2} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{X} - \mathbf{B}\|_{2,2} \leq \sigma,$$

with

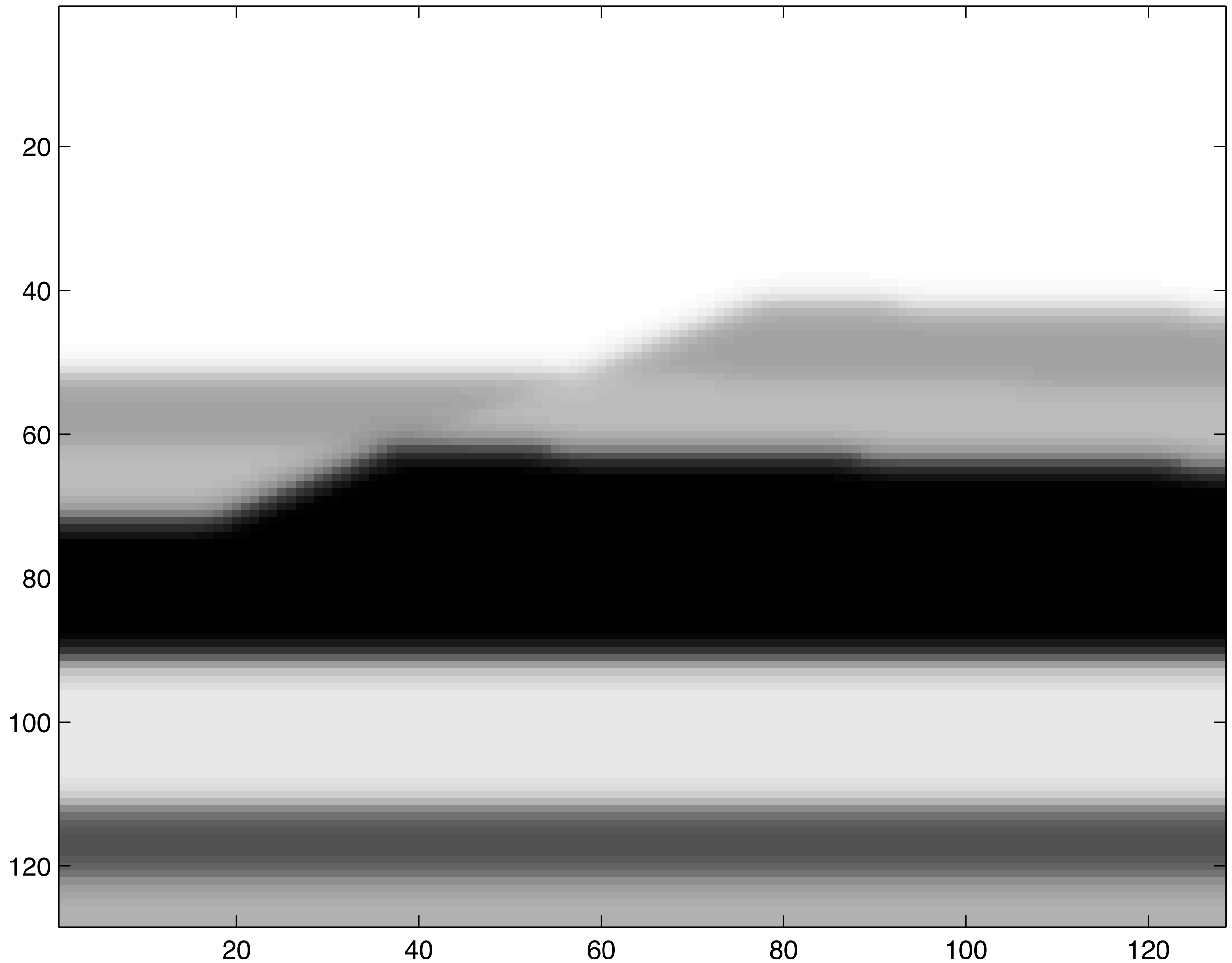
$$\|\mathbf{X}\|_{1,2} := \sum_{i \in \mathrm{rows}(\mathbf{X})} \|\mathrm{row}_i(\mathbf{X})^*\|_2$$

and

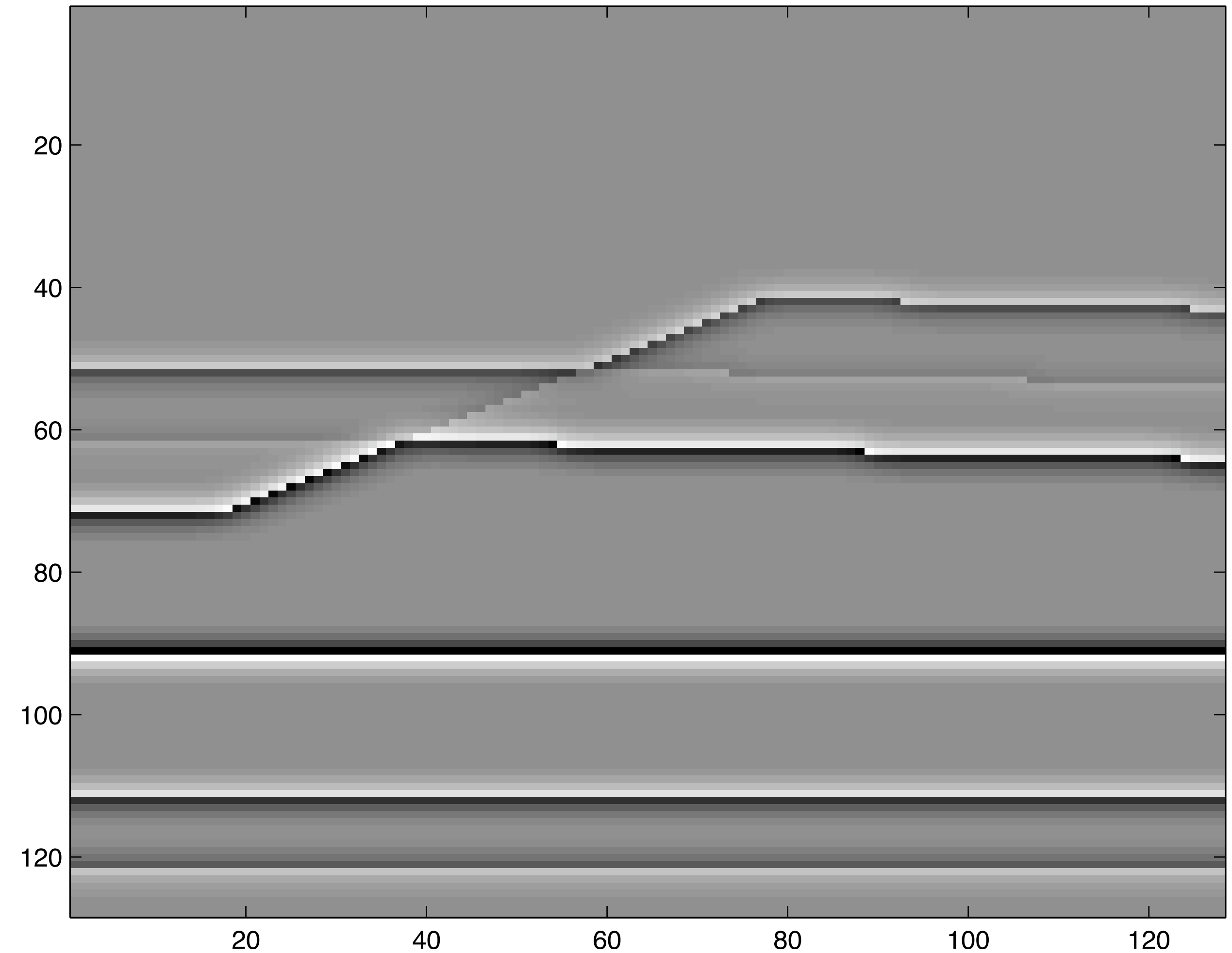
$$\|\mathbf{X}\|_{2,2} := \left(\sum_{i \in \mathrm{rows}(\mathbf{X})} \|\mathrm{row}_i(\mathbf{X})^*\|_2^2 \right)^{\frac{1}{2}} .$$

Solved with SPGL1.

Example

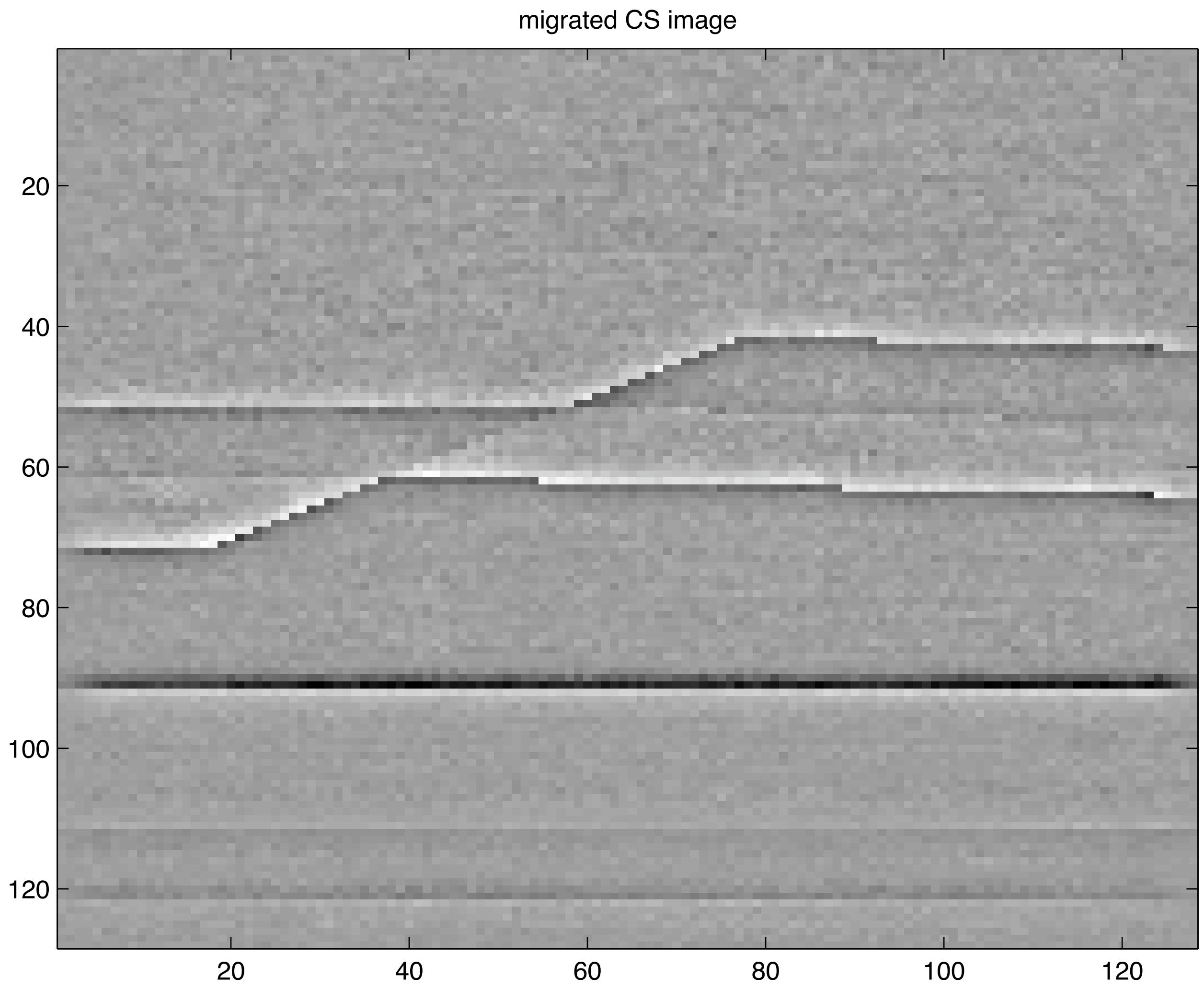


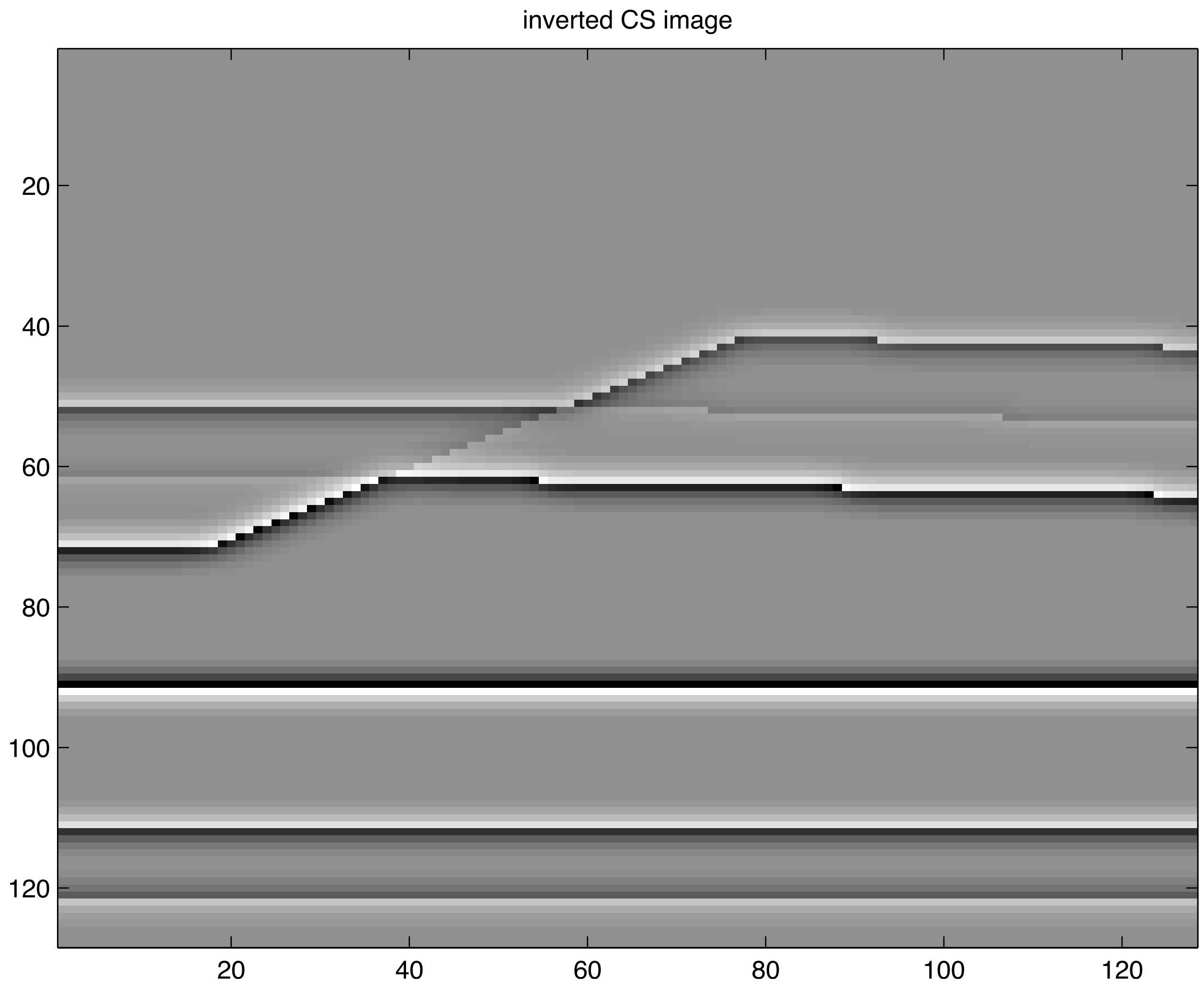
background velocity model



perturbation

Example

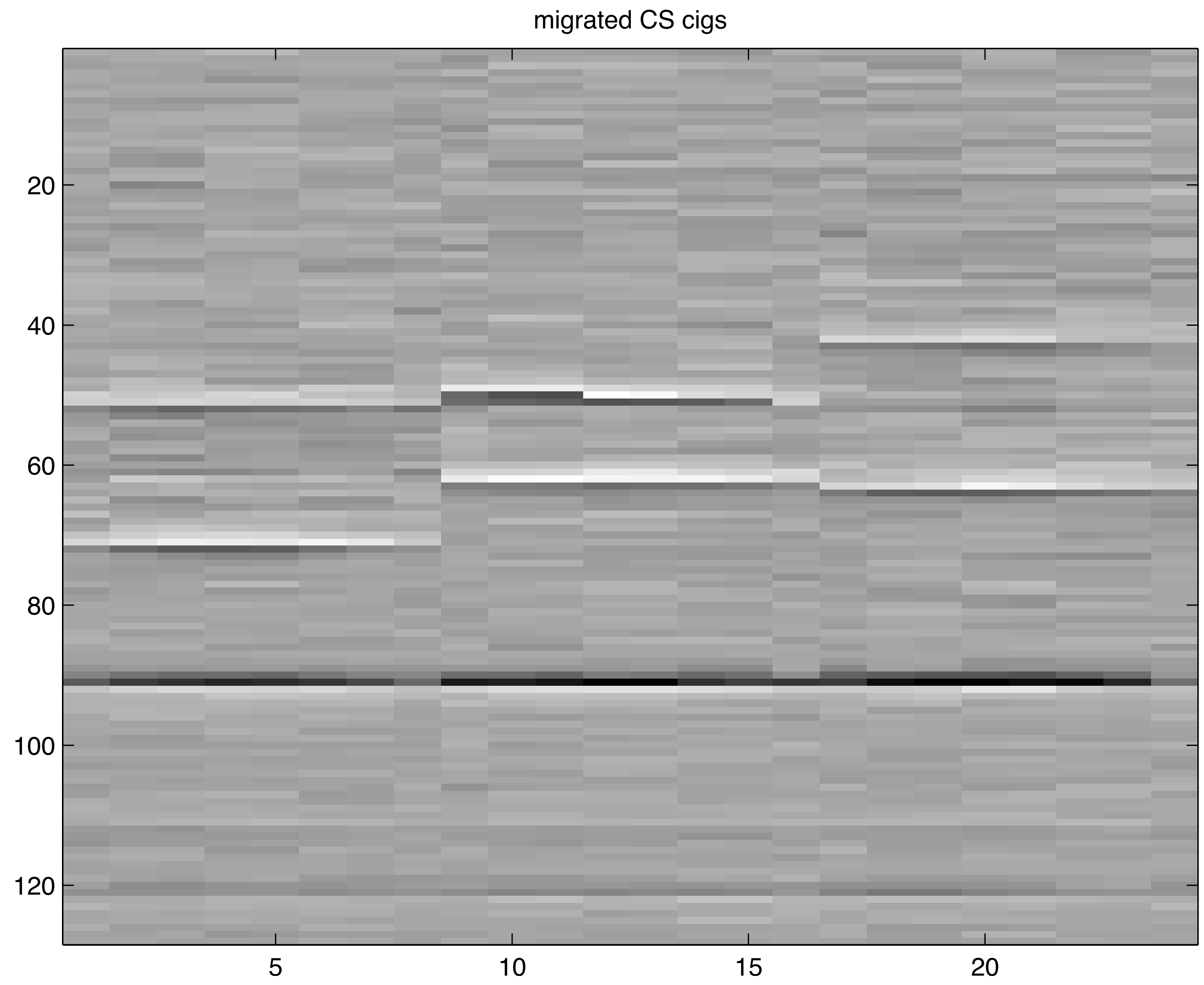




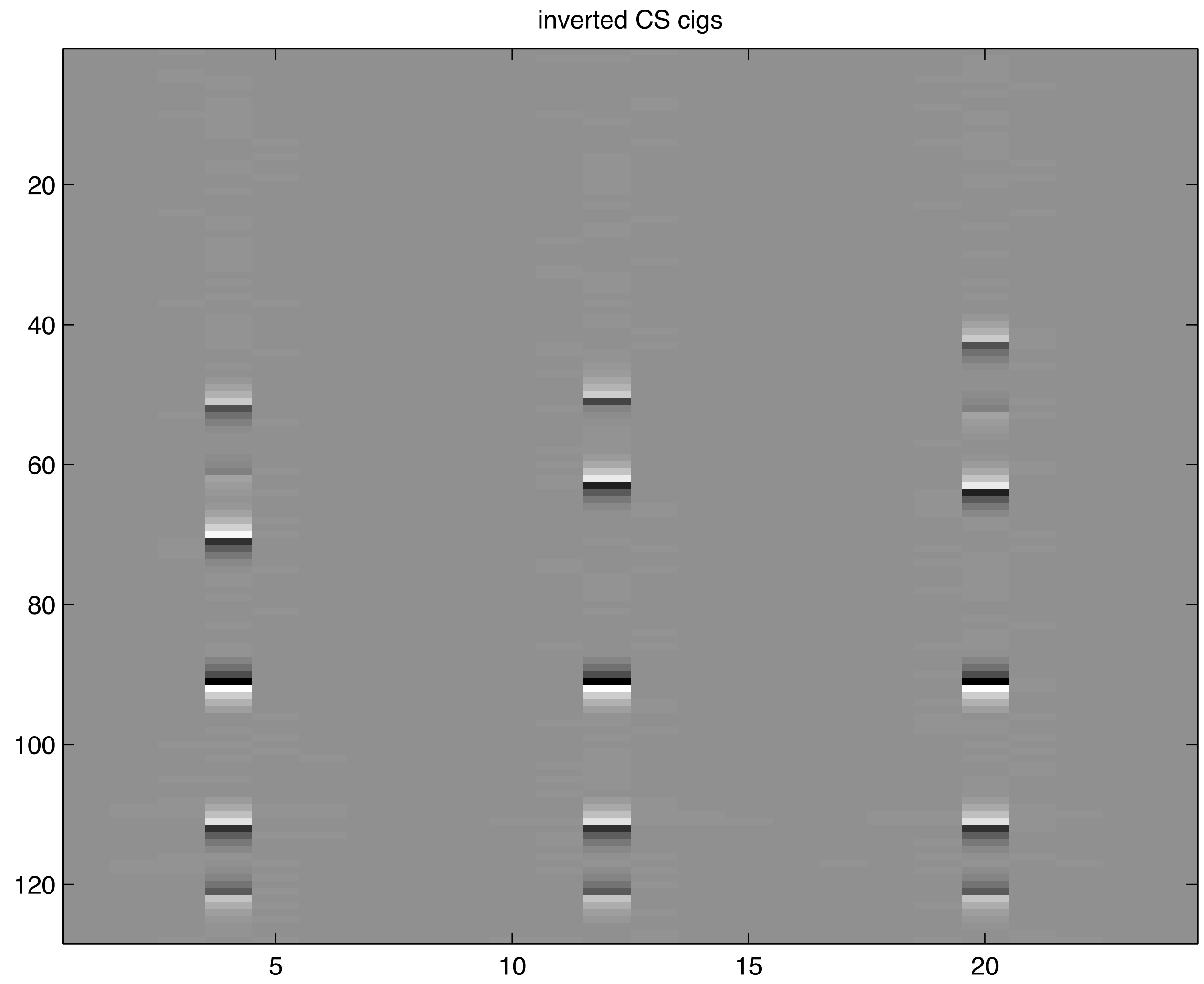
matched filter **sparsity promotion**

Recovery from 64-fold subsampling ...

Example



matched filter



sparsity promotion

Common-image gathers are focussed.

Implications

- Model-space CS leads to a significant reduction of
  - simulation costs (reduction of the number of right-hand sides & frequencies)
  - storage costs
  - matrix-matrix multiply costs

- Opens enticing perspective to solve Symes' nonlocal extension, i.e.
  - map model variables to a functional $\mathbf{m}(x) \mapsto \boldsymbol{\mathcal{M}}(x, x')$
  - solve inverse problem with extra constraint:

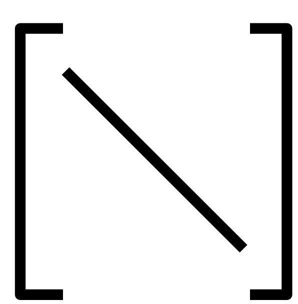
$$\min_{\mathbf{U}\in\boldsymbol{\mathcal{U}},\, \mathbf{X}\in\boldsymbol{\mathcal{X}}} \frac{1}{2}\|\mathbf{P} - \mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \begin{cases} \mathbf{H}[\mathbf{S}^*\mathbf{X}]\mathbf{U} & = \mathbf{Q} \\ \mathrm{P}_h\mathbf{X} & = \mathbf{0} \end{cases} \quad \wedge \quad \|\mathbf{X}\|_{1,2} \leq \tau$$

- Open problem
  - size functionals is prohibitive => need CS techniques to compress

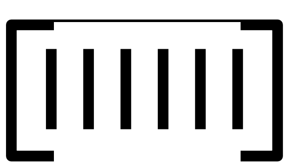
Why we are doing this

Invariance & Sparsity

- Compressing operators using an alternative notion of invariance
- NOT going for invariance of support in a certain basis (Diagonalization)
- Instead going for ***sparsity invariance*** under a sparsifying transform (Compressive Sensing)



diagonalization



compressive sensing

SLIM
Seismic Laboratory for Imaging and Modeling

Observations & outlook

- CS allows for a ***compression*** of data volumes without *significant* ***loss*** of ***information*** yielding a ***reduction*** in *computational* costs

- CS has ***direct*** implications for seismic acquisition--from ***sequential*** to ***simultaneous*** acquisition

- ***Joint*** sparsity promotion allows for ***focusing***

- **Speculation:** Proposed approach may be suitable to handle Symes's proposal to add a degree of freedom yielding a ***nonlocal*** *forward* model in tandem with an inverse problem that penalizes ***nonlocality*** through ***focusing ...***

Acknowledgments

- E. van den Berg and M. P. Friedlander for *SPGL1* (www.cs.ubc.ca/labs/scl/spgl1) & *Sparco* (www.cs.ubc.ca/labs/scl/sparco)
- Sergey Fomel and Yang Liu for Madagascar (rsf.sf.net)
- E. Candes and the Curvelab team
- My (ex) students Gilles Hennenfent, Tim Lin, Gang Tang, and PDFs Reza Shahidi & Yogi Erlangga

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.

slim.eos.ubc.ca

and... Thank you!