Compressive least squares migration with on-the-fly Fourier transforms

Philipp A. Witte, Mathias Louboutin, Fabio Luporini, Gerard J. Gorman and Felix J. Herrmann

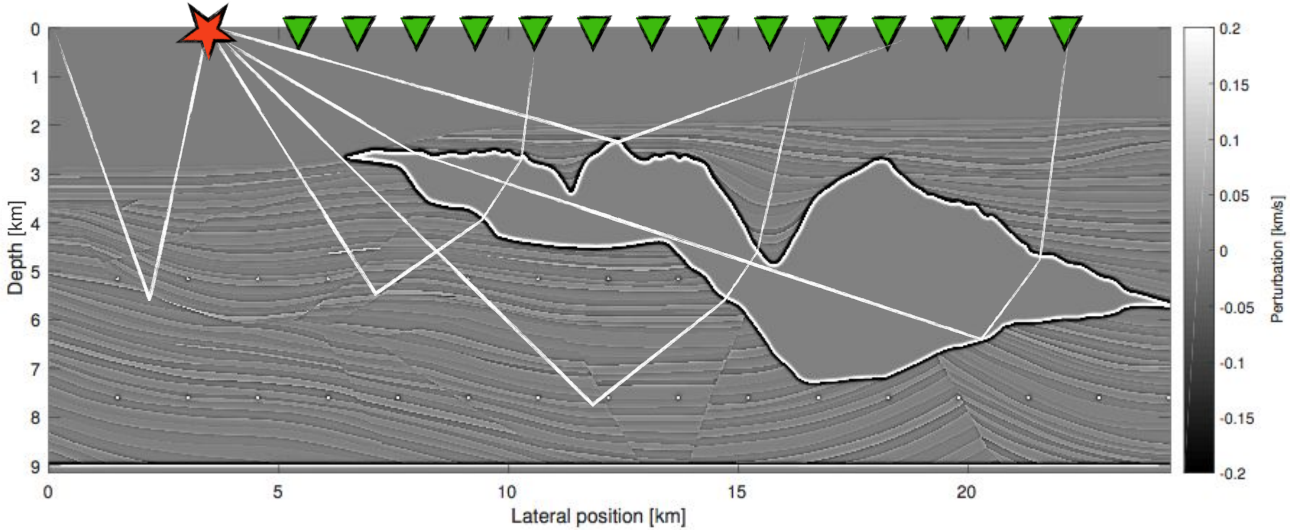
SLIM

Georgia Institute of Technology

Motivation

Geophysical exploration using seismic imaging:

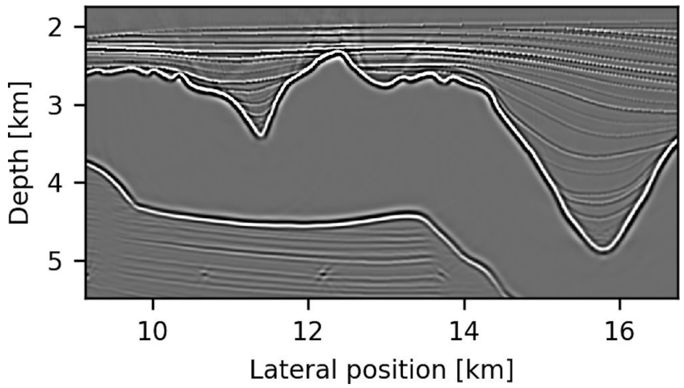
- Image velocity/impedance contrasts in the subsurface

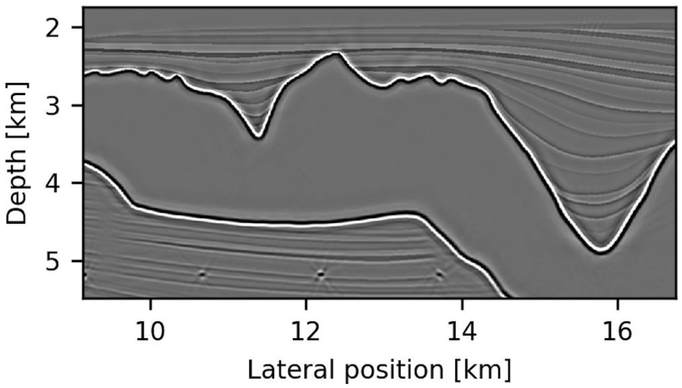


Motivation

Reverse time migration (RTM) vs. least squares imaging (LS-RTM):

- RTM: adjoint of Born scattering operator (1 data pass)
- LS-RTM: iterative inversion (many data passes, usually ~20)

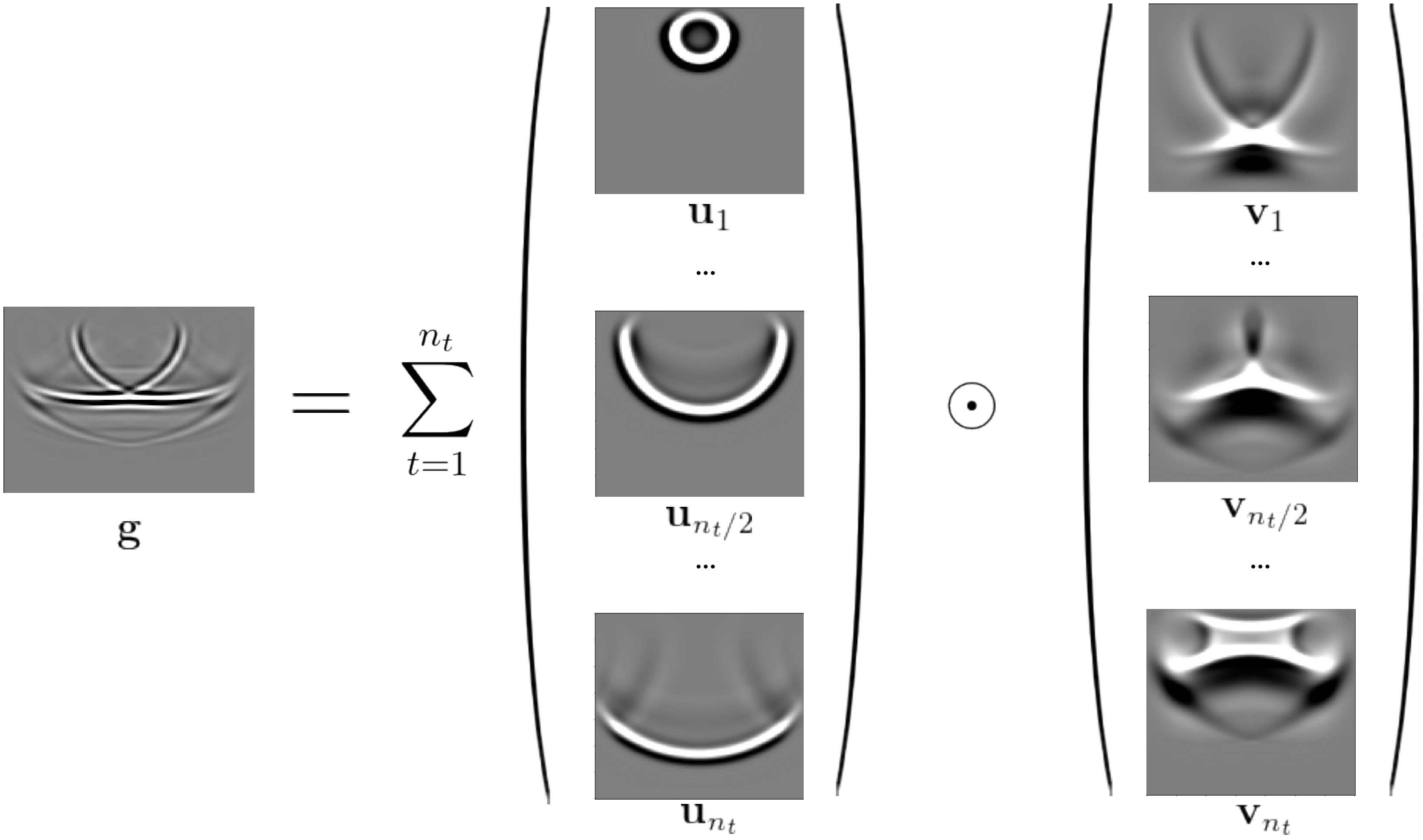


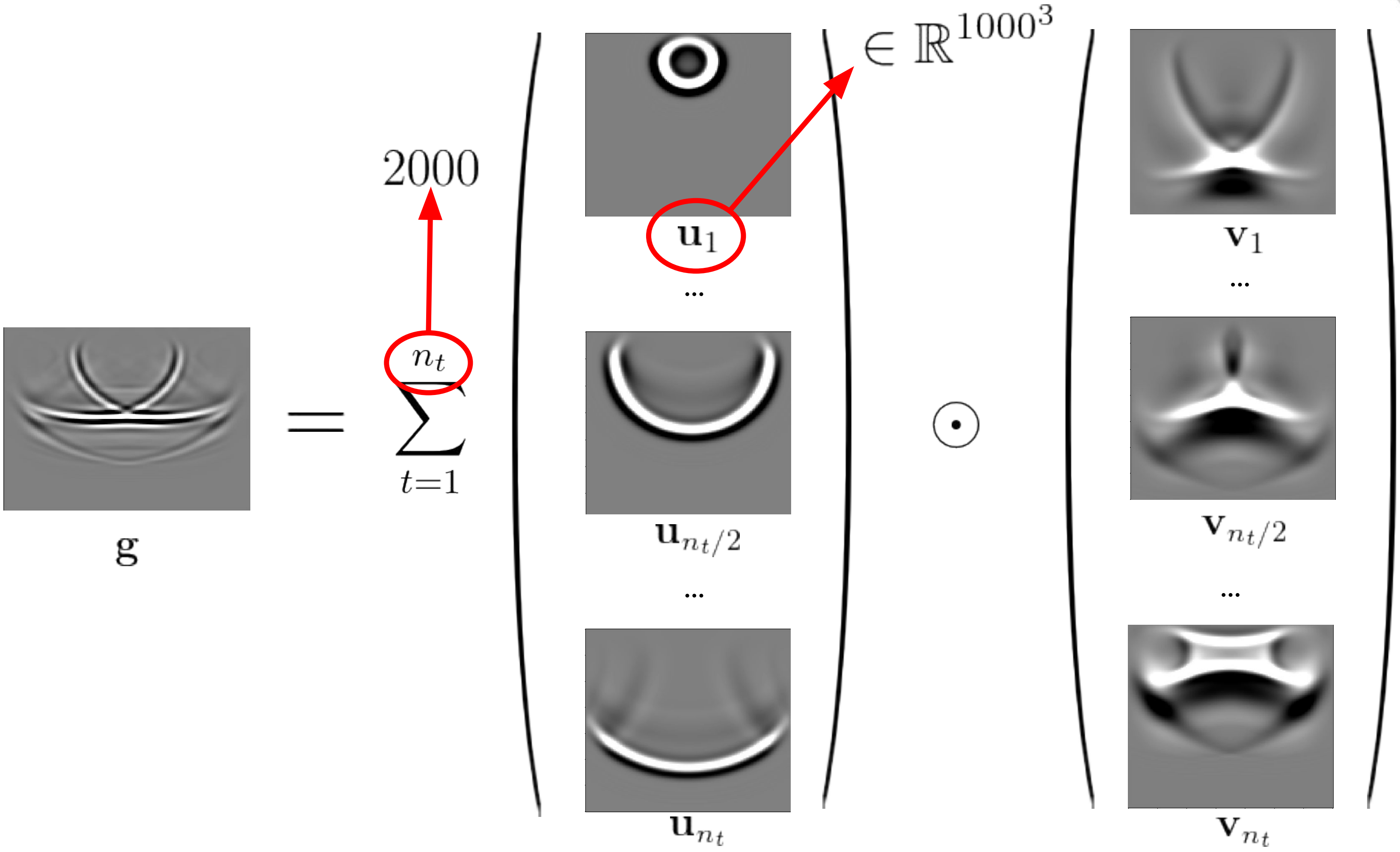


Motivation

Computational challenges of seismic imaging:

- LS-RTM: Need to solve large number of PDEs at every iteration
- Expensive PDE solves: propagate waves over ~20,000 time steps
- Large data sets (easily up to ~100 TB)
- Large number of variables (between 1e6 and 1e10)
- Large memory requirements for backpropagation


$\mathbf{g}$
$=$
$\sum_{t=1}^{n_t}$
$\mathbf{u}_1$
...
$\mathbf{u}_{n_t/2}$
...
$\mathbf{u}_{n_t}$
$\odot$
$\mathbf{v}_1$
...
$\mathbf{v}_{n_t/2}$
...
$\mathbf{v}_{n_t}$



$\mathbf{g} = \sum_{t=1}^{n_t} \left(\mathbf{u}_1 \ldots \mathbf{u}_{n_t/2} \ldots \mathbf{u}_{n_t} \right) \odot \left(\mathbf{v}_1 \ldots \mathbf{v}_{n_t/2} \ldots \mathbf{v}_{n_t} \right)$
2000
$\in \mathbb{R}^{1000^3}$

Motivation

Strategies for backpropagation:

- Wavefield subsampling
- Wavefield reconstruction from boundary (e.g. McMechan, 1983)
- Optimal checkpointing (Griewank and Walther, 2000)
- Domain decomposition (MPI based)
- Elasticache, memory optimized instances (AWS)
- Frequency domain imaging (solve Helmholtz)
- Time-to-frequency conversion (e.g. DFTs) (Overview in Furse, 1998)

Motivation

Low cost least squares RTM with on-the-fly DFTs:

- Time domain modeling w/ on-the-fly DFTs
- Work w/ random subsets of frequencies/sources
- Benefits of LS-RTM at a fraction of the cost
- Sparsity-promoting minimization to address subsampling artifacts

Time-to-frequency conversion

Least squares RTM objective function in frequency domain:

$$\underset{\delta\mathbf{m}}{\text{minimize}} \quad \sum_{j=1}^{n_s}\sum_{k=1}^{n_f}\frac{1}{2}\left\|\mathbf{J}(\mathbf{m}_0,\bar{q}_{jk})\,\delta\mathbf{m}-\bar{\mathbf{d}}_{jk}^{\text{obs}}\right\|_2^2.$$

With:

$\mathbf{J} \in \mathbb{C}^{n_r \times n}$	linearized Born scattering operator
$\delta\mathbf{m} \in \mathbb{C}^{n}$	unknown image
$\mathbf{m}_0 \in \mathbb{C}^{n}$	migration velocity (assumed to be known)
$\bar{\mathbf{d}}_{\text{jk}}^{\text{obs}} \in \mathbb{C}^{n_r}$	observed seismic data
$\bar{q}_{jk} \in \mathbb{C}$	source wavelet (assumed to be known)

SLIM

Time-to-frequency conversion

Model linearized data with time modeling code:

$$\bar{\mathbf{d}}_{jk}^{\mathrm{pred}} = -\mathbf{P}_r \mathbf{R}_k \mathbf{F} \mathbf{A}(\mathbf{m}_0)^{-1} \mathrm{diag}\left[\frac{\partial \mathbf{A}(\mathbf{m}_0)}{\partial \mathbf{m}} \mathbf{A}(\mathbf{m}_0)^{-1} \mathbf{F}^* \mathbf{R}_k^* \mathbf{p}_s^* \bar{q}_{jk} \right] \delta \mathbf{m}$$

With:

$\mathbf{P}_r \in \mathbb{C}^{n_r \times n}$ receiver restriction

$\mathbf{R}_k \in \mathbb{C}^{n \times n \cdot n_t}$ frequency restriction

$\mathbf{F} \in \mathbb{C}^{n \cdot n_t \times n \cdot n_t}$ discrete Fourier transform

$\mathbf{A}(\mathbf{m}_0) \in \mathbb{R}^{n \cdot n_t \times n \cdot n_t}$ discretized wave equation

Time-to-frequency conversion

Compute gradient of LS-RTM objective function:

$$\bar{\mathbf{g}}_{jk} = -\mathrm{Re}\Big[\mathrm{diag}\big(\omega_k^2 \bar{\mathbf{u}}_{jk}\big)^* \bar{\mathbf{v}}_{jk}\Big]$$

With: $\bar{\mathbf{u}}_{jk} = \mathbf{R}_k \mathbf{F} \mathbf{A}(\mathbf{m}_0)^{-1} \mathbf{F}^* \mathbf{R}_k^* \mathbf{p}_s^* \bar{q}_{jk}$

$$\bar{\mathbf{v}}_{jk} = \mathbf{R}_k \mathbf{F} \mathbf{A}(\mathbf{m}_0)^{-*} \mathbf{F}^* \mathbf{R}_k^* \mathbf{P}_r^* (\bar{\mathbf{d}}_{jk}^{\mathrm{pred}} - \bar{\mathbf{d}}_{jk}^{\mathrm{obs}})$$

$$\bar{\mathbf{u}}_{jk}, \bar{\mathbf{v}}_{jk} \in \mathbb{C}^n$$

Time-to-frequency conversion

So far: explicit DFTs in modeling expressions

- replace w/ on-the-fly DFTs
- during time stepping compute:

$$\bar{\mathbf{u}}_{jk}^{\text{real}} = \sum_{i=1}^{n_t} \cos(2\pi f_k i \Delta t)\mathbf{u}_i,$$

$$\bar{\mathbf{u}}_{jk}^{\text{imag}} = -\sum_{i=1}^{n_t} \sin(2\pi f_k i \Delta t)\mathbf{u}_i$$

$$\bar{\mathbf{u}}_{jk}^{\text{real}}, \bar{\mathbf{u}}_{jk}^{\text{imag}}, \mathbf{u}_i \in \mathbb{R}^n$$

(Furse, 1998; Sirgue et al., 2010)

Time-to-frequency conversion

Computing gradients $\bar{\mathbf{g}}_{jk} \in \mathbb{R}^n$ w/ on-the-fly DFTs

- perform DFT in adjoint loop
- compute frequency-domain adjoint wavefield + gradient in single step:

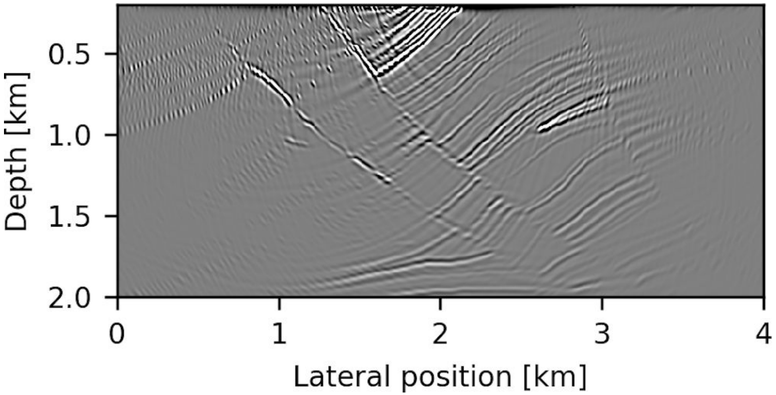
$$\bar{\mathbf{g}}_{jk} = -\sum_{i=1}^{n_t} (2\pi f_k)^2 \mathrm{diag}\Big[\bar{\mathbf{u}}_{jk}^{\mathrm{real}} \cos(2\pi f_k i \Delta t) - \bar{\mathbf{u}}_{jk}^{\mathrm{imag}} \sin(2\pi f_k i \Delta)\Big] \mathbf{v}_i$$

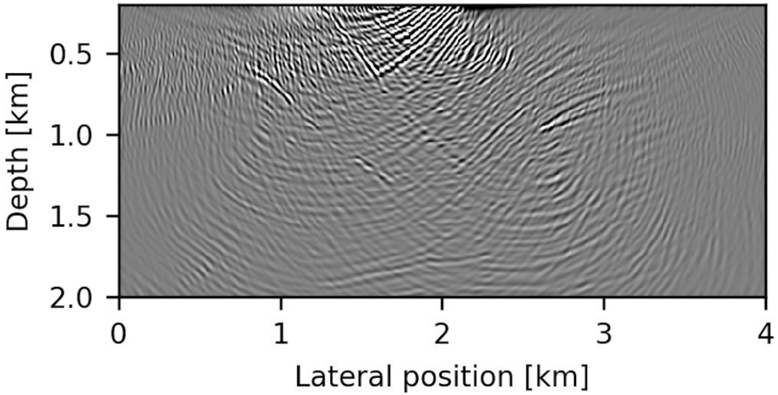
- alternative imaging conditions possible as well

Sparsity-promoting LS-RTM

So far: frequency-domain imaging w/ time modeling

- need fine frequency sampling for imaging
- periodic subsampling causes aliasing/coherent artifacts
- compressive sensing: random sampling

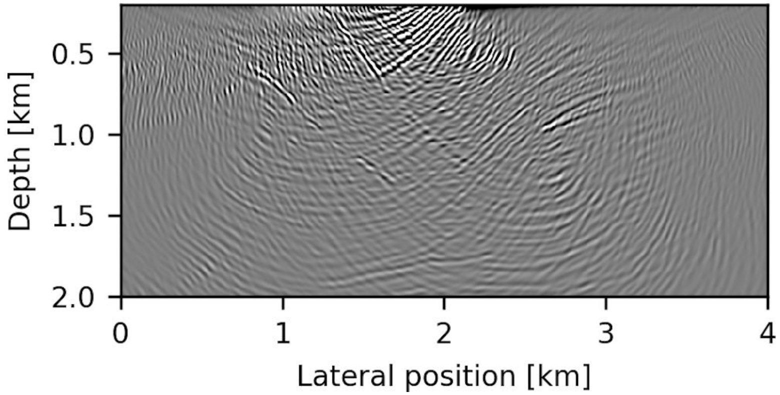


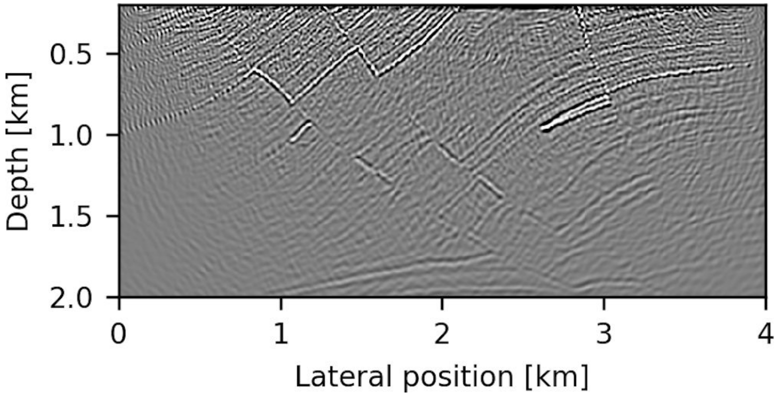


Sparsity-promoting LS-RTM

Compressive sensing (CS) inspired imaging:

- linear system becomes underdetermined
- different random frequencies for each experiment
- coherent reflectors + incoherent noise





Sparsity-promoting LS-RTM

Ingredients for CS inspired seismic imaging:

- random sampling to break coherency of noise
- sparsifying transform (e.g. curvelets)
- sparsity-promoting optimization to recover noise-free image

(Candes et al., 2006; Donoho, 2006)

(Tu et al., 2015; Herrmann et al., 2015)

SLIM

Sparsity-promoting LS-RTM

LS-RTM as sparsity-promoting minimization problem:

- elastic net objective function (strongly convex)

$$\underset{\delta\mathbf{z}}{\text{minimize}} \quad \lambda||\mathbf{C}\,\delta\mathbf{z}||_1 + \frac{1}{2}||\mathbf{C}\,\delta\mathbf{z}||_2^2$$

$$\text{subject to:} \quad \sum_{j=1}^{n_s}\sum_{k=1}^{n_f}\left\|\mathbf{M}_l^{-1}\mathbf{J}(\mathbf{m}_0,\bar{q}_{jk})\mathbf{M}_r^{-1}\delta\mathbf{z} - \mathbf{M}_l^{-1}\bar{\mathbf{d}}_{jk}^{\text{obs}}\right\|_2 \leq \sigma$$

$\mathbf{C} \in \mathbb{C}^{n_c \times n}$ curvelet transform

$\delta\mathbf{z} \in \mathbb{R}^n$ seismic impedance

$\mathbf{M}_l \in \mathbb{R}^{n_r \times n_r}$, $\mathbf{M}_r \in \mathbb{R}^{n \times n}$ left/right-hand preconditioners

$\lambda, \sigma \in \mathbb{R}$ penalty parameter and noise in ℓ_2-ball

Sparsity-promoting LS-RTM

Algorithm: the linearized Bregman method

1. Initialize $\mathbf{x}_1 = \mathbf{0}$, $\mathbf{z}_1 = \mathbf{0}$, q, λ, batch sizes $\hat{n}_s \ll n_s$ and $\hat{n}_f \ll n_f$
2. **for** $i = 1, ..., n$
3. Select subset of shots and frequencies $\mathcal{S} = (\int_{\text{shot}}, \int_{\text{freq}}), |\int_{\text{shot}}| = \hat{n}_s, |\int_{\text{freq}}| = \hat{n}_f$
4. $\bar{\mathbf{d}}_{\mathcal{S}}^{\text{pred}} = \mathbf{M}_l^{-1}\mathbf{J}_{\mathcal{S}}\mathbf{M}_r^{-1}\mathbf{x}$
5. $\bar{\mathbf{g}}_{\mathcal{S}} = \mathbf{M}_r^{-\top}\mathbf{J}_{\mathcal{S}}^{\top}\mathbf{M}_l^{-\top}\mathcal{P}_{\sigma}(\bar{\mathbf{d}}_{\mathcal{S}}^{\text{pred}} - \bar{\mathbf{d}}_{\mathcal{S}}^{\text{obs}})$
6. $\mathbf{z}_{i+1} = \mathbf{z}_i - t_i\bar{\mathbf{g}}_{\mathcal{S}}$
7. $\mathbf{x}_{i+1} = S_{\lambda}(\mathbf{z}_{i+1})$
8. **end**

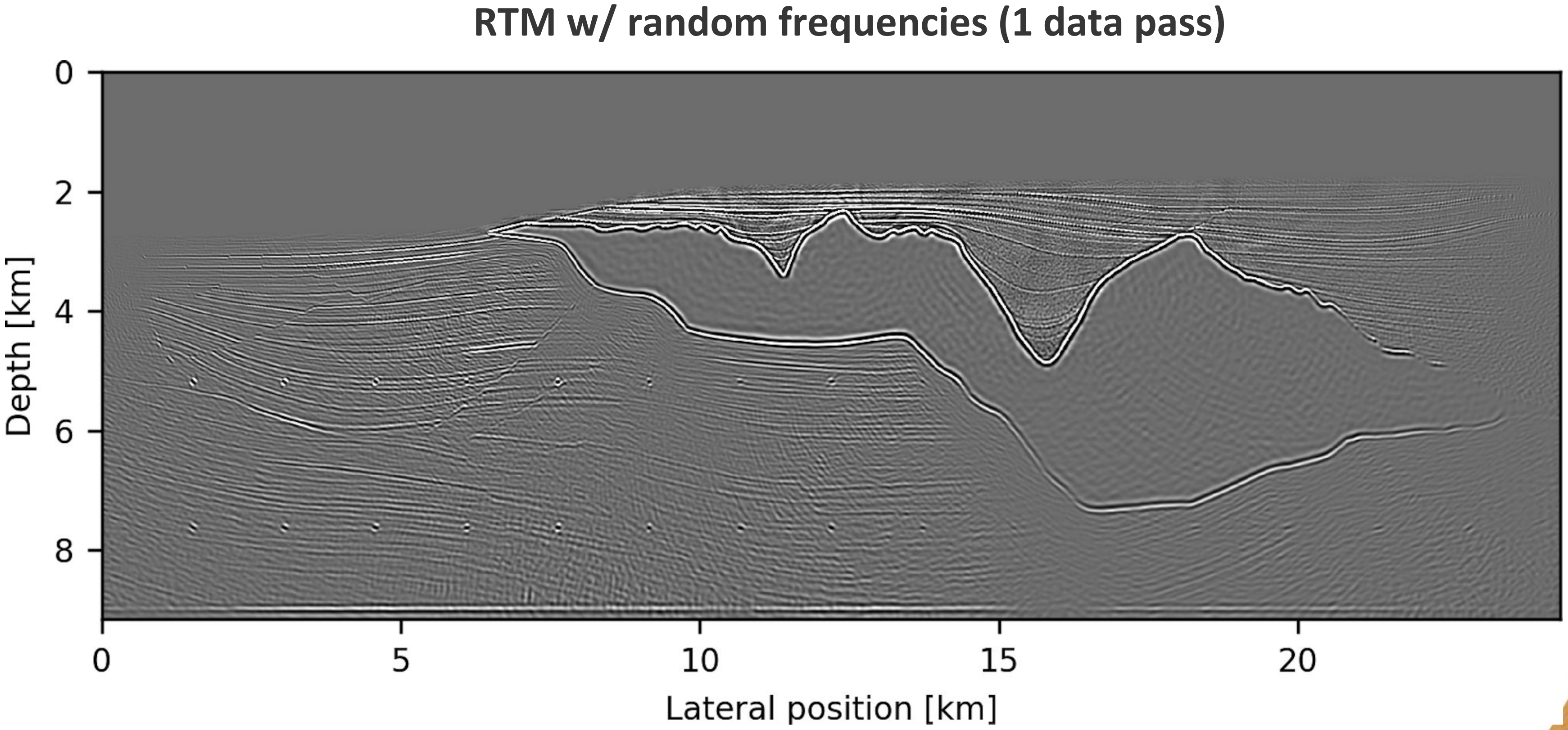
(Yin et al., 2008; Lorenz et al., 2014)

Sigsbee 2A example

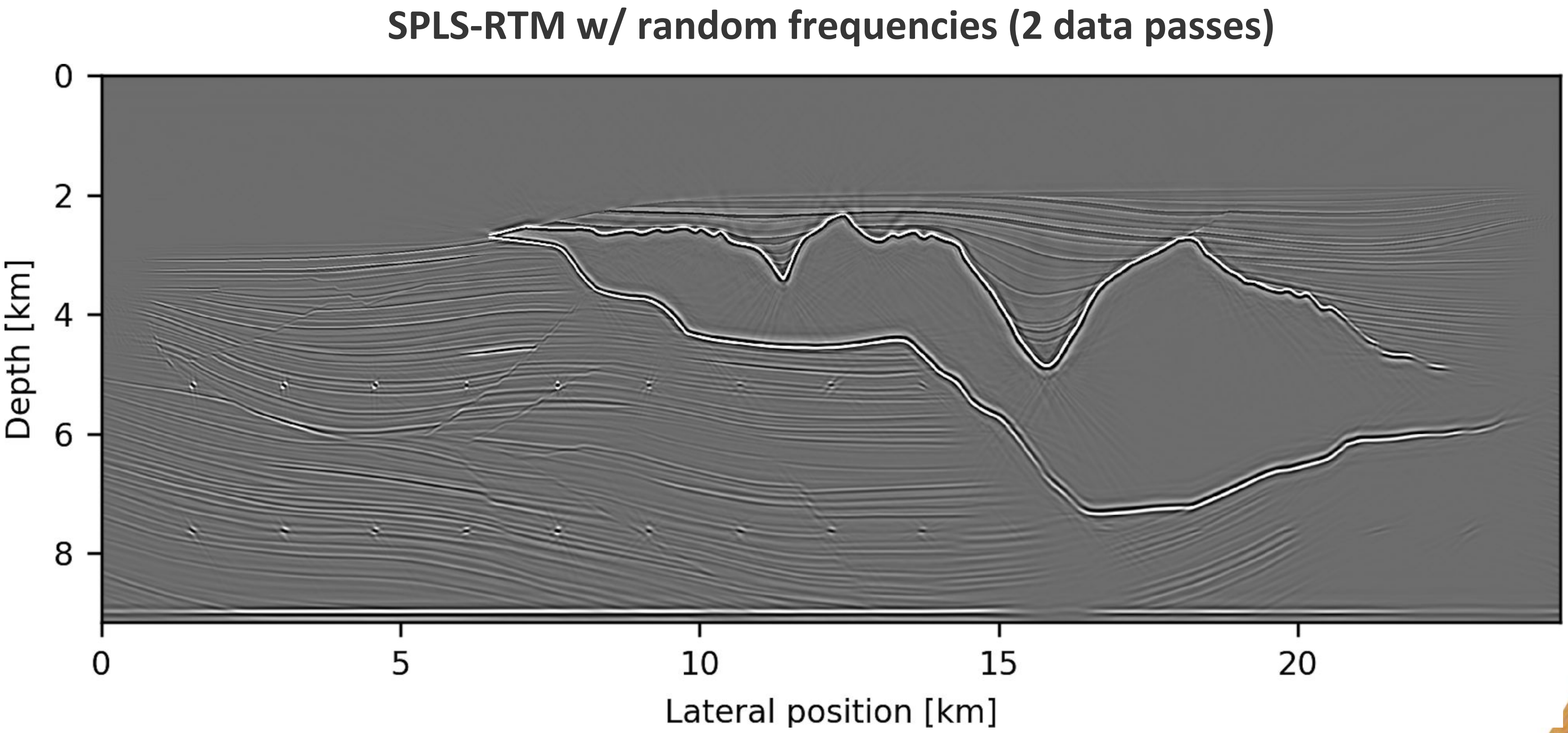
Sigsbee 2A example:

- 24.6 by 9.2 km 2D model (3201 x 1201 grid points)
- 14,095 time steps
- 935 experiments (shot locations)
- comparison of RTM, SPLS-RTM
- on-the-fly DFTs w/ subsampling vs optimal checkpointing
- 20 iterations w/ 100 shots and 20 frequencies per shot

Sigsbee 2A example

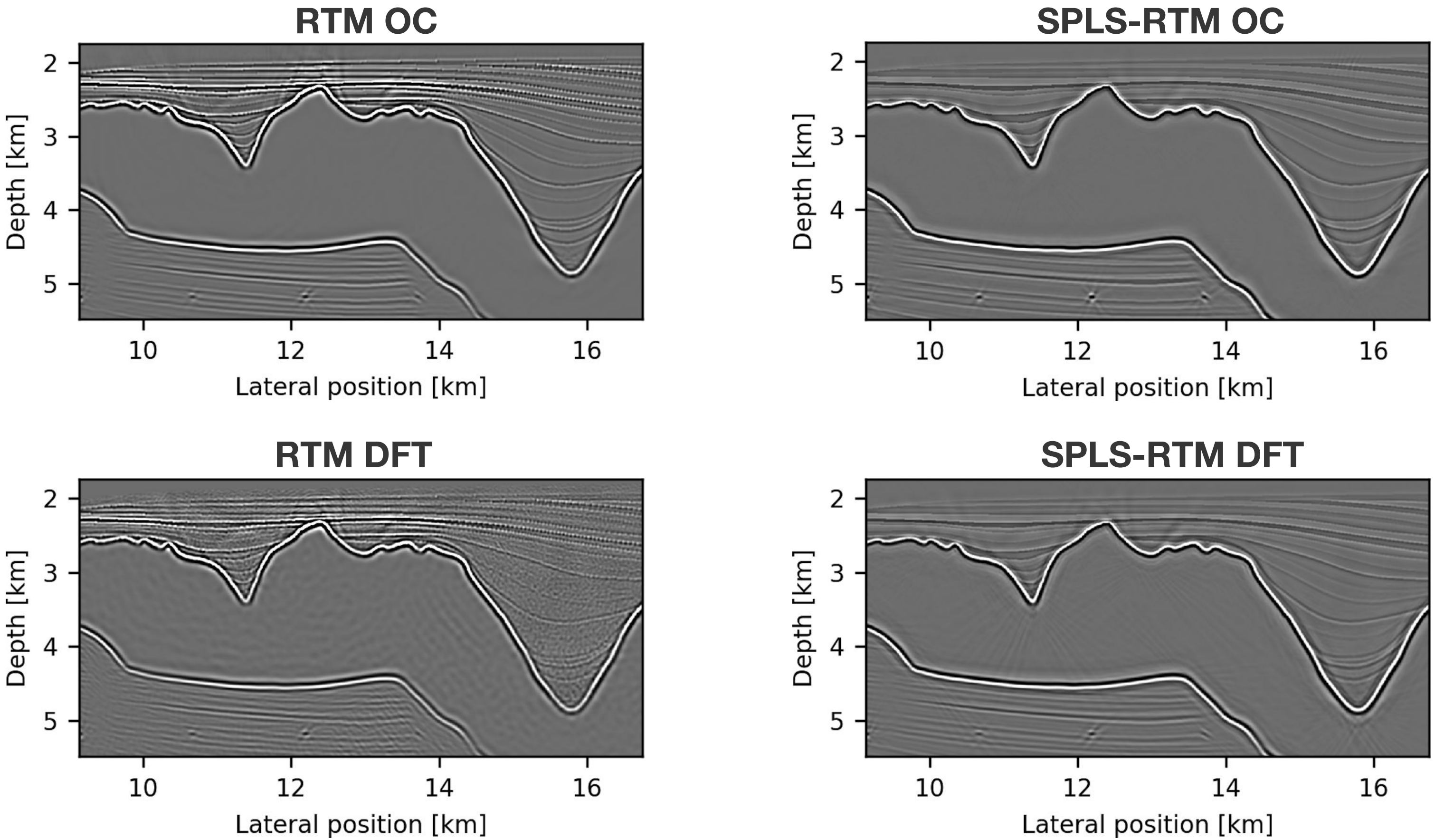


Sigsbee 2A example

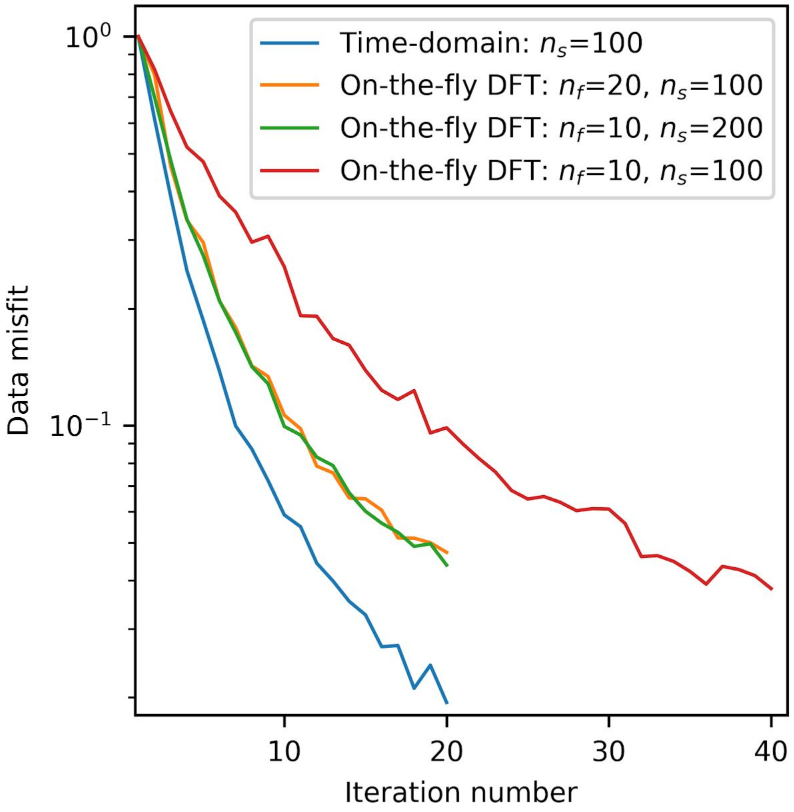


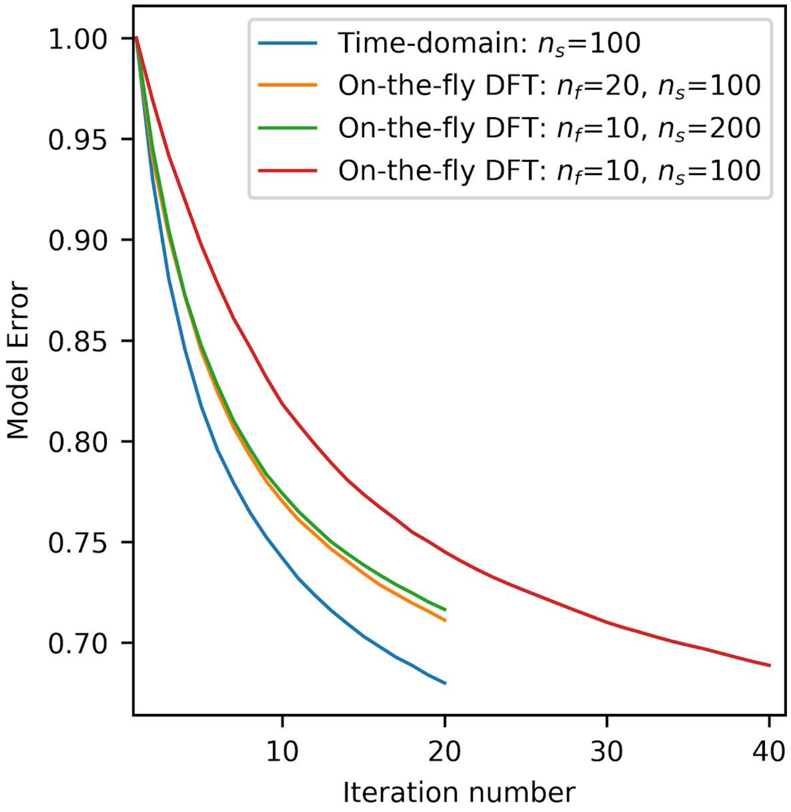
Sigsbee 2A example

OC: optimal checkpointing
DFT: on-the-fly Fourier transforms

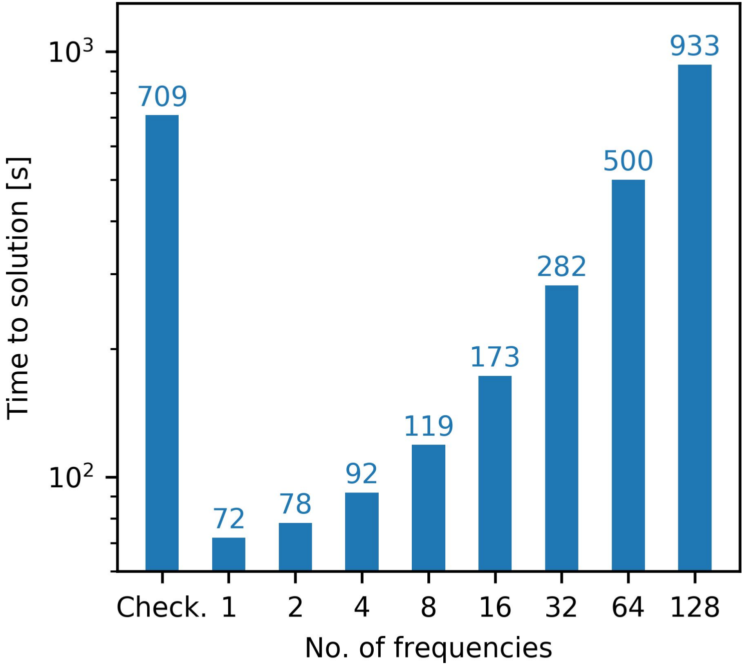


Sigsbee 2A example





Sigsbee 2A example



Strategy	Memory	Additional cost
TD: save all wavefields	$\mathcal{O}(n_t)$	–
TD: optimal checkpointing	$\mathcal{O}(\log n_t)$	$\mathcal{O}(\log n_t)$
TD: boundary reconstruction	$\mathcal{O}(n_t)$	$\mathcal{O}(n_t)$
FD: on-the-fly DFT	$\mathcal{O}(n_f)$	$\mathcal{O}(n_f)$

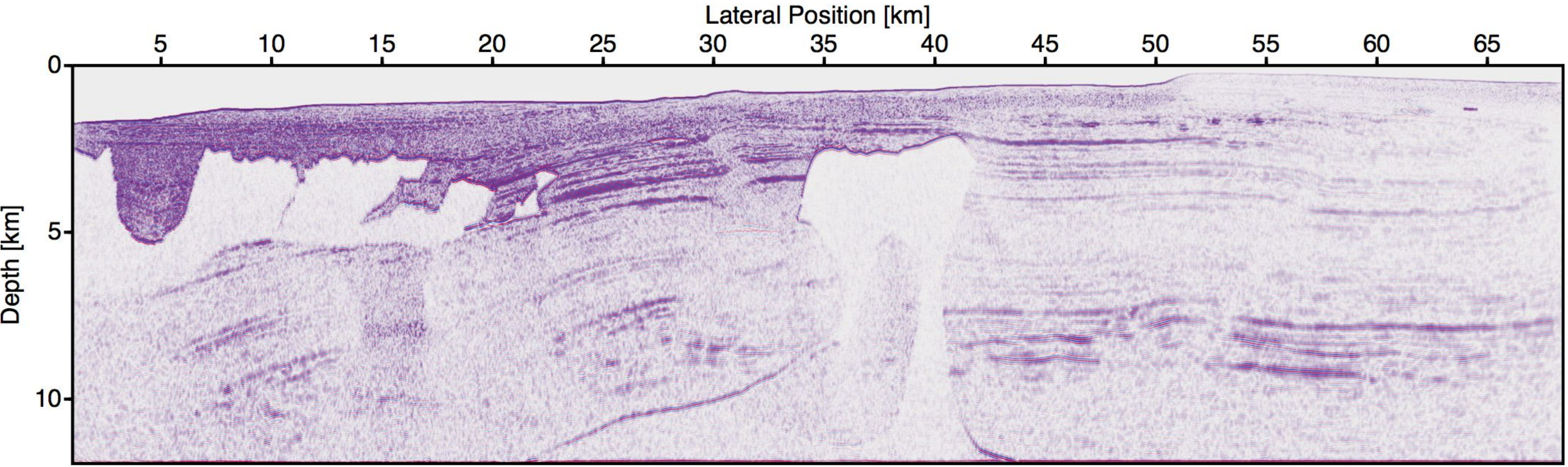
- DFT approach independent of number of time steps
- But: how many frequencies are required?

BP Synthetic 2004

SPLS-RTM enables imaging at large-scales

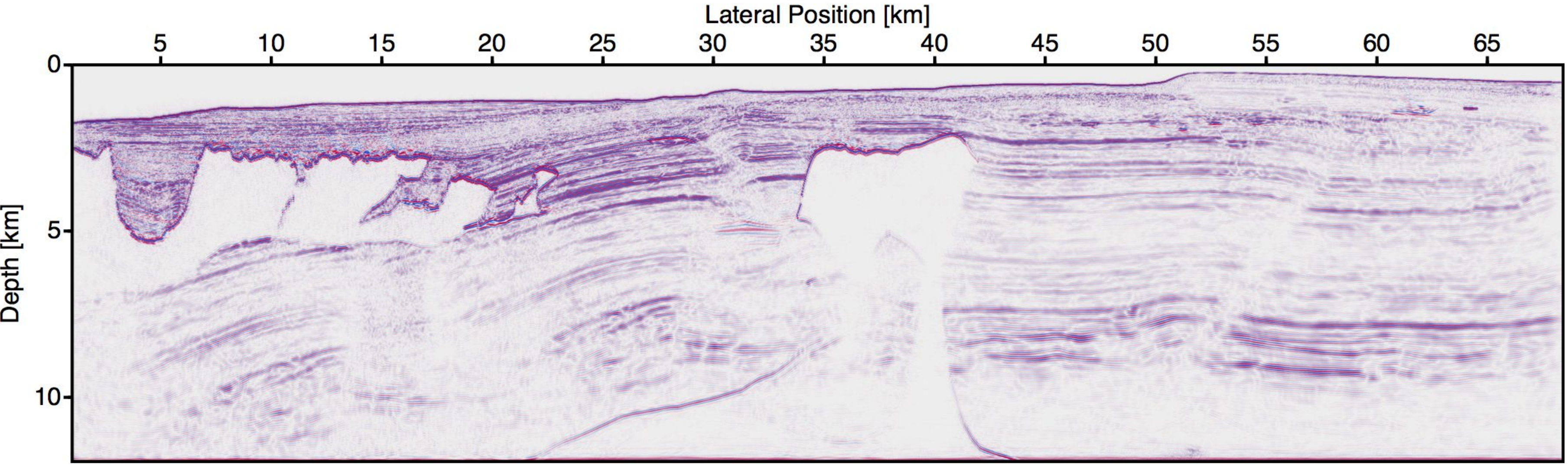
- 2D BP synthetic model 2004
- 67.4 by 11.9 km image (10,789 x 1,911 grid points)
- 1340 shot locations
- 16,920 time steps
- 20 iterations w/ 200 shots, 20 frequencies

BP Synthetic 2004



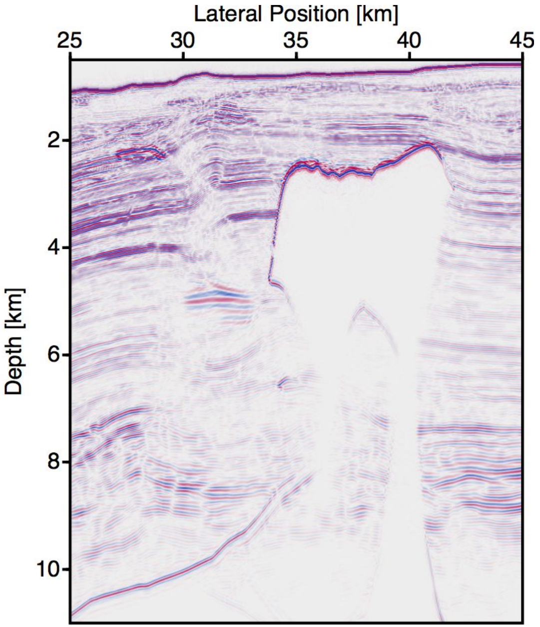
RTM w/ 20 random frequencies per shot

BP Synthetic 2004

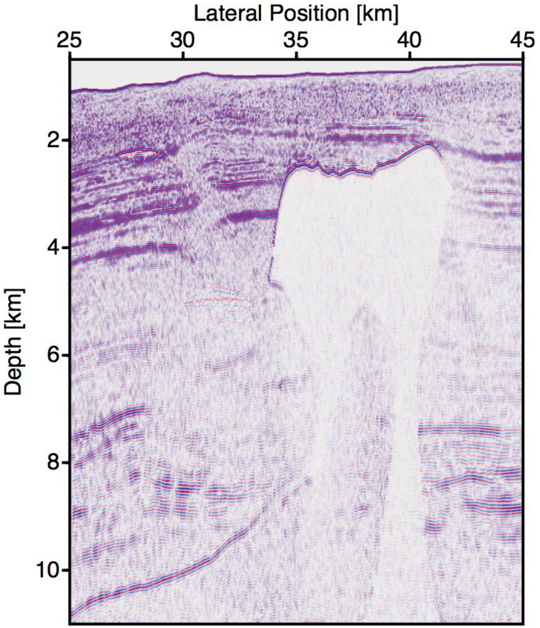


SPLS-RTM w/ 20 random frequencies per shot

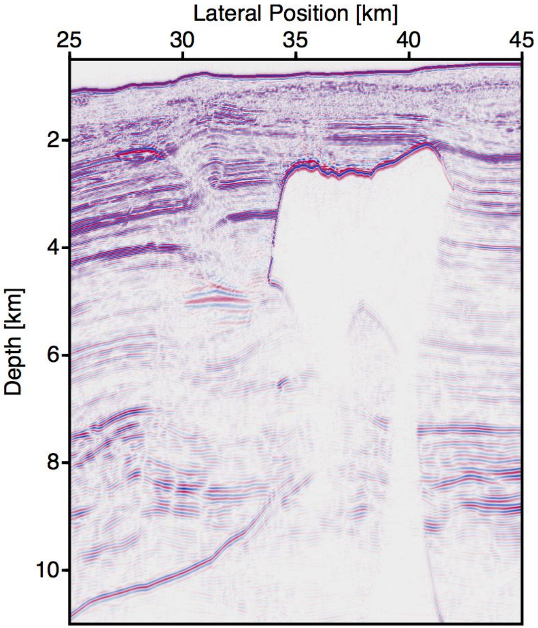
BP Synthetic 2004



SPLS-RTM (OC)

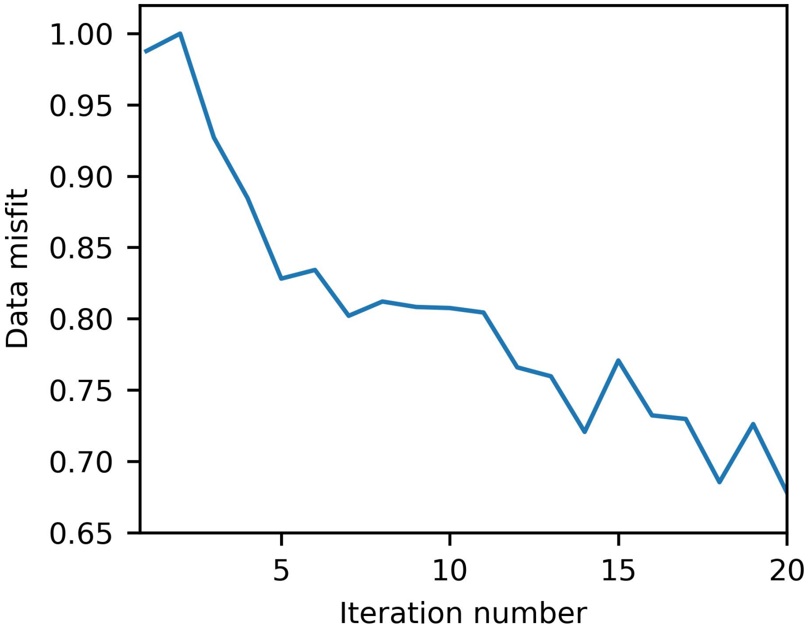


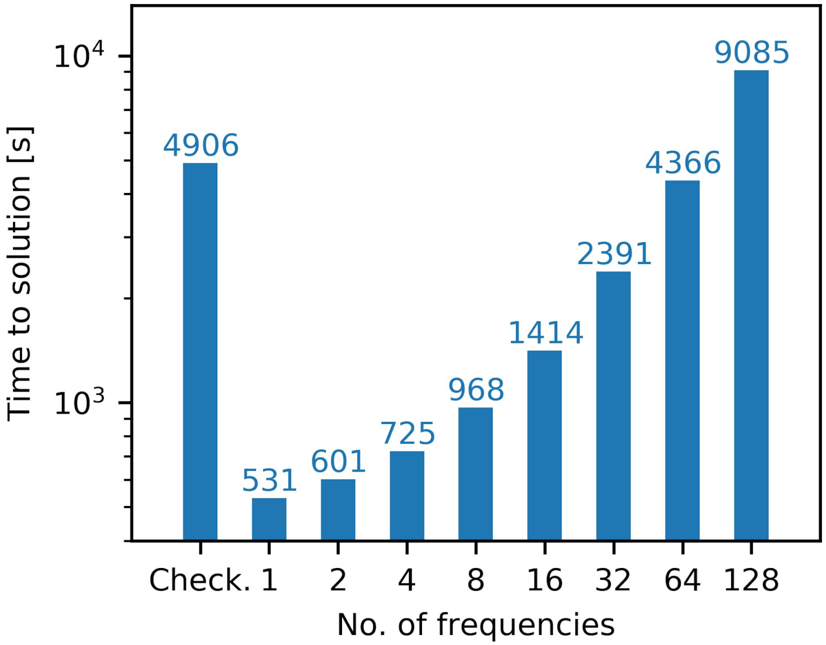
RTM (DFT)



SPLS-RTM (DFT)

BP Synthetic 2004





SLIM

Summary

Iterative least squares imaging at a fraction of the cost:

- on-the-fly DFTs
- comparable quality to time-domain **IF** formulated as sparsity-promoting LS problem
- memory independent of no. of time steps
- cost and memory depend of number of frequencies
- flexibility in terms of no. of iterations and batchsizes
- possible to trade cost for quality

Acknowledgements

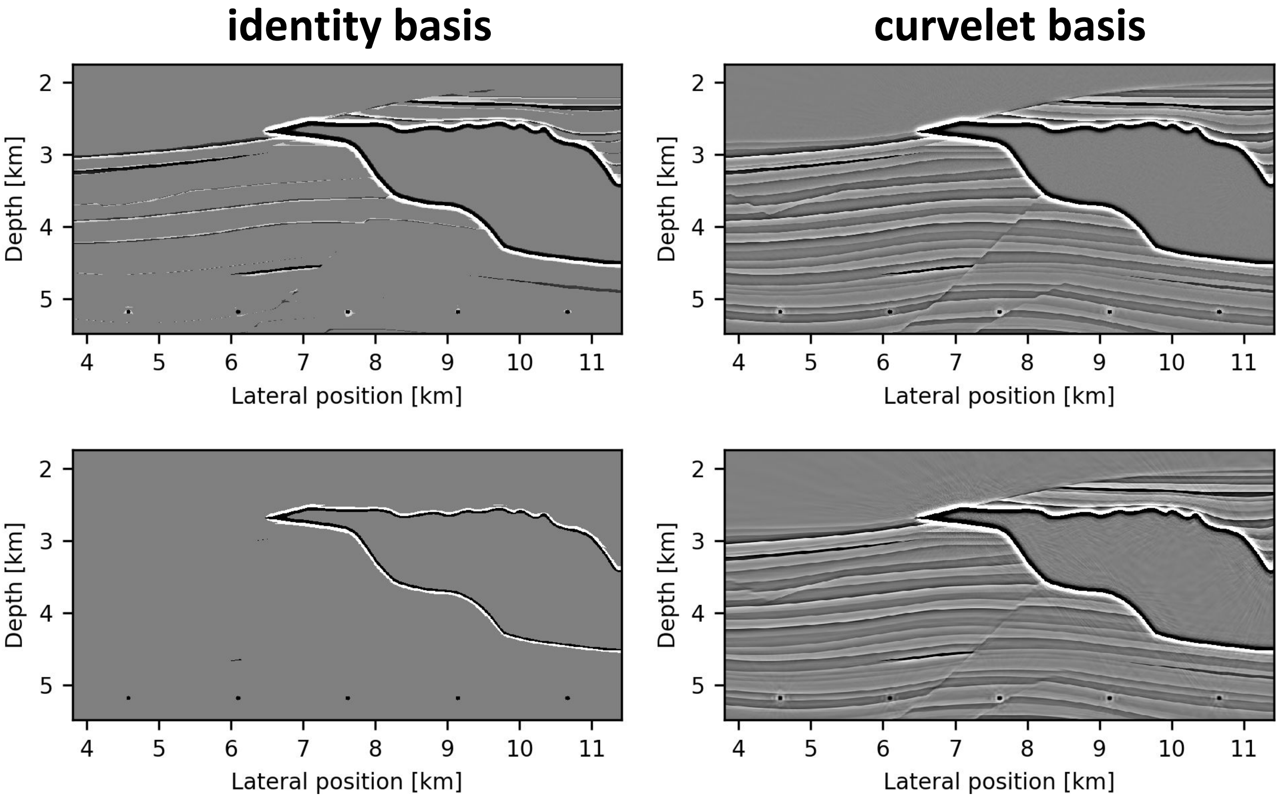
Many thanks to:

Eileen R. Martin (Virginia Tech) for organizing this session
Members of the SLIM group (Georgia Tech)
Henryk Modzelewski (The University of British Columbia)
Charles Jones (Osokey Ltd.)

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.

Sparsity-promoting LS-RTM

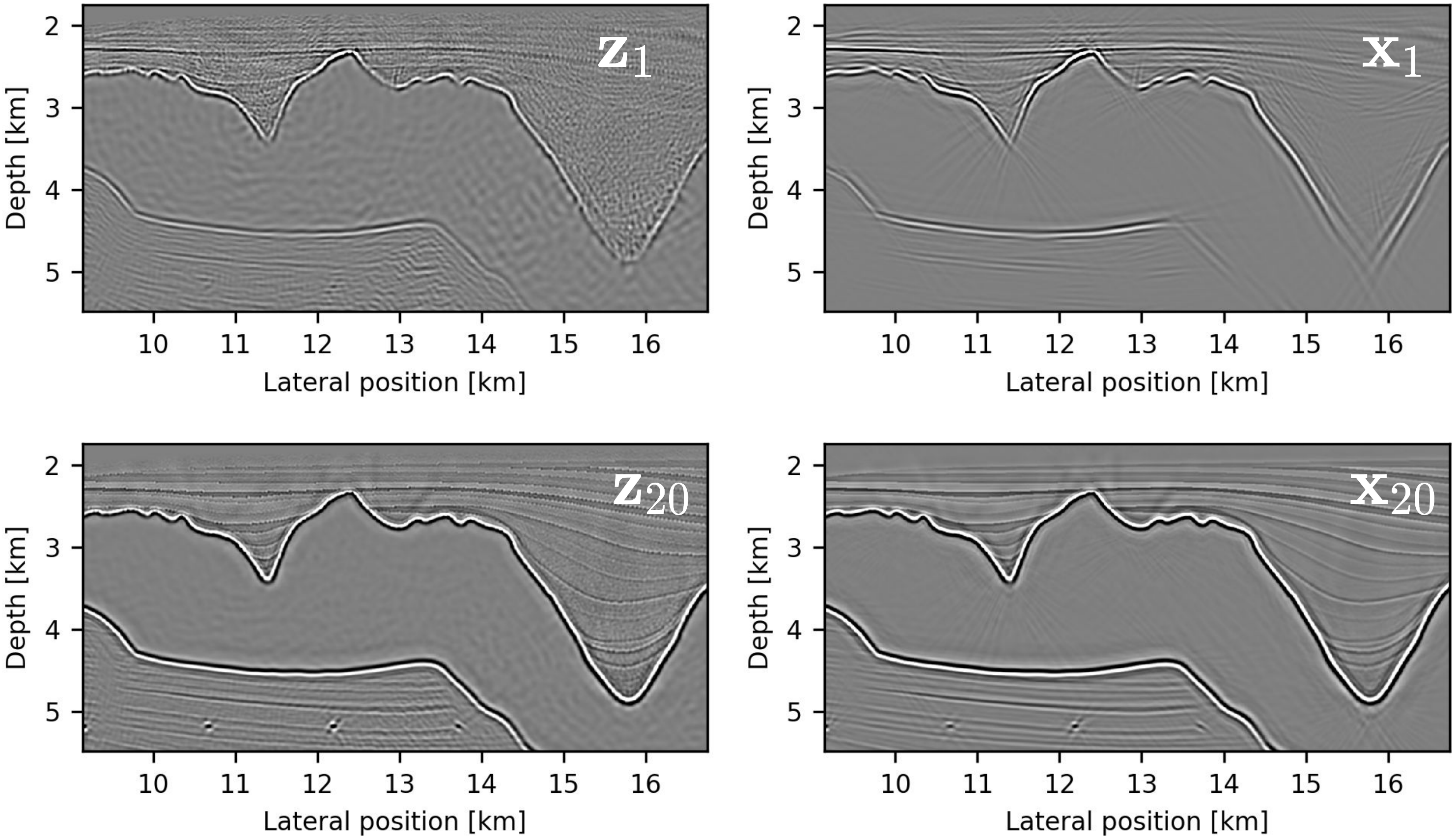
Why curvelets?



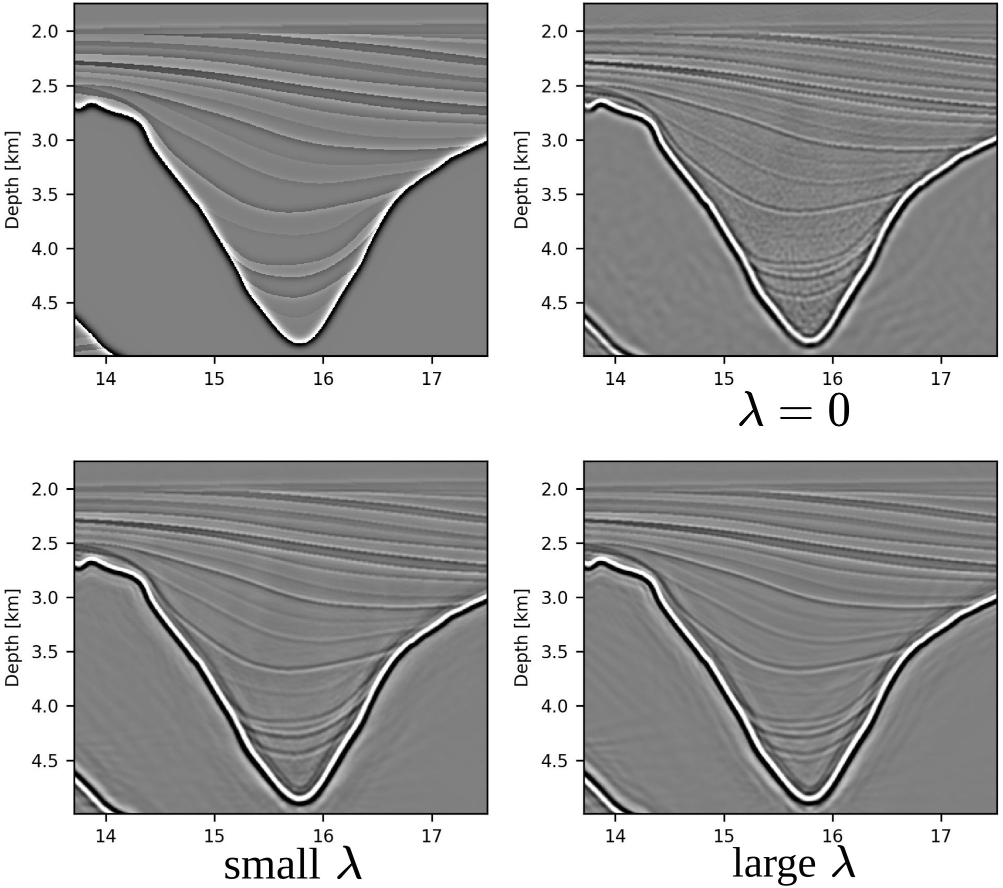
Largest 5% of coefficients

Largest 1% of coefficients

Sigsbee 2A example



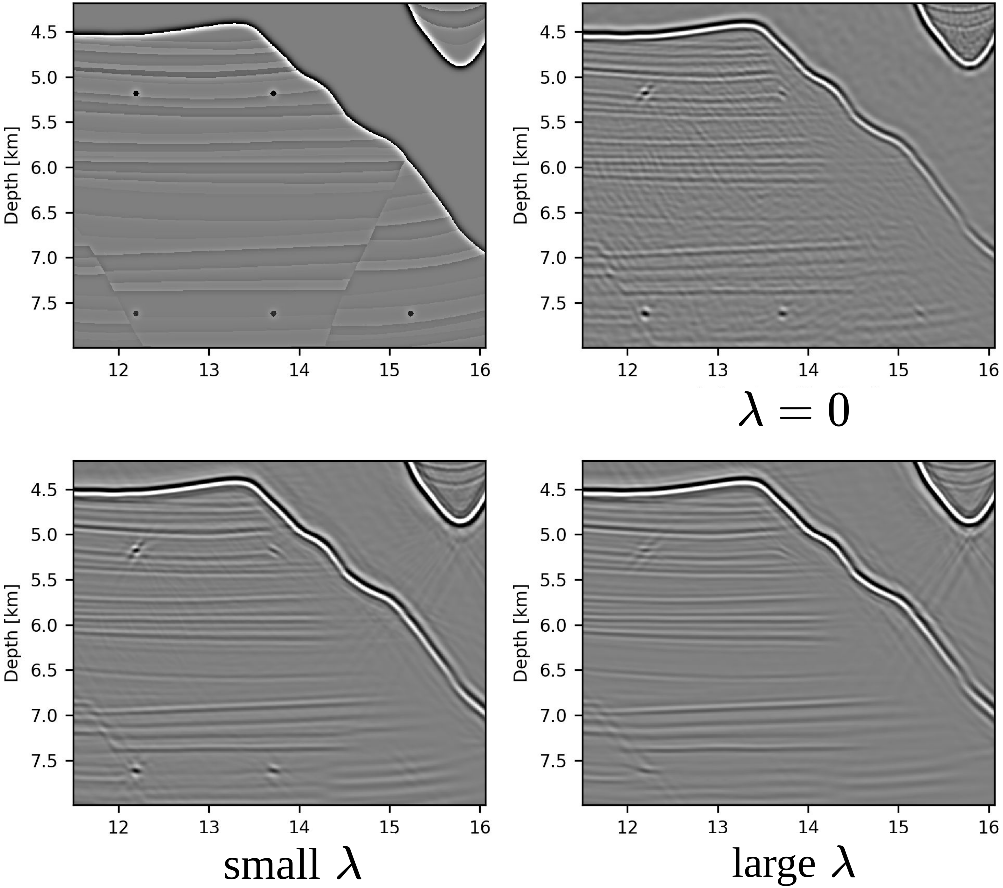
Supplementary slides



Role of the penalty parameter:

- determines the amount of thresholding
- areas of poor illumination more sensitive to its choice

Supplementary slides



Role of the penalty parameter:

- determines the amount of thresholding
- areas of poor illumination more sensitive to its choice