Relax the *physics* & *expand* the *search* space - *FWI* via *W*avefield *R*econstruction *I*nversion

Felix J. Herrmann

SLIM

University of British Columbia

van Leeuwen, T and Herrmann, F J (2013). Mitigating local minima in full-waveform inversion by expanding the search space. Geophysical Journal International.
van Leeuwen, T and Herrmann, F J (2013). A penalty method for PDE-constrained optimization. Submitted for publication
van Leeuwen, T and Herrmann, F J (2013). US Provisional Patent Application No. 61/815,533. A Penalty Method for PDE-Constrained Optimization with Applications to Wave-Equation Based Seismic Inversion

Relax the *physics* & *expand* the *search* space - *FWI* via *W*avefield *R*econstruction *I*nversion

Ernier Esser, Tristan van Leeuwen*, and Bas Peters

* now @ CWI Amsterdam

SLIM

University of British Columbia

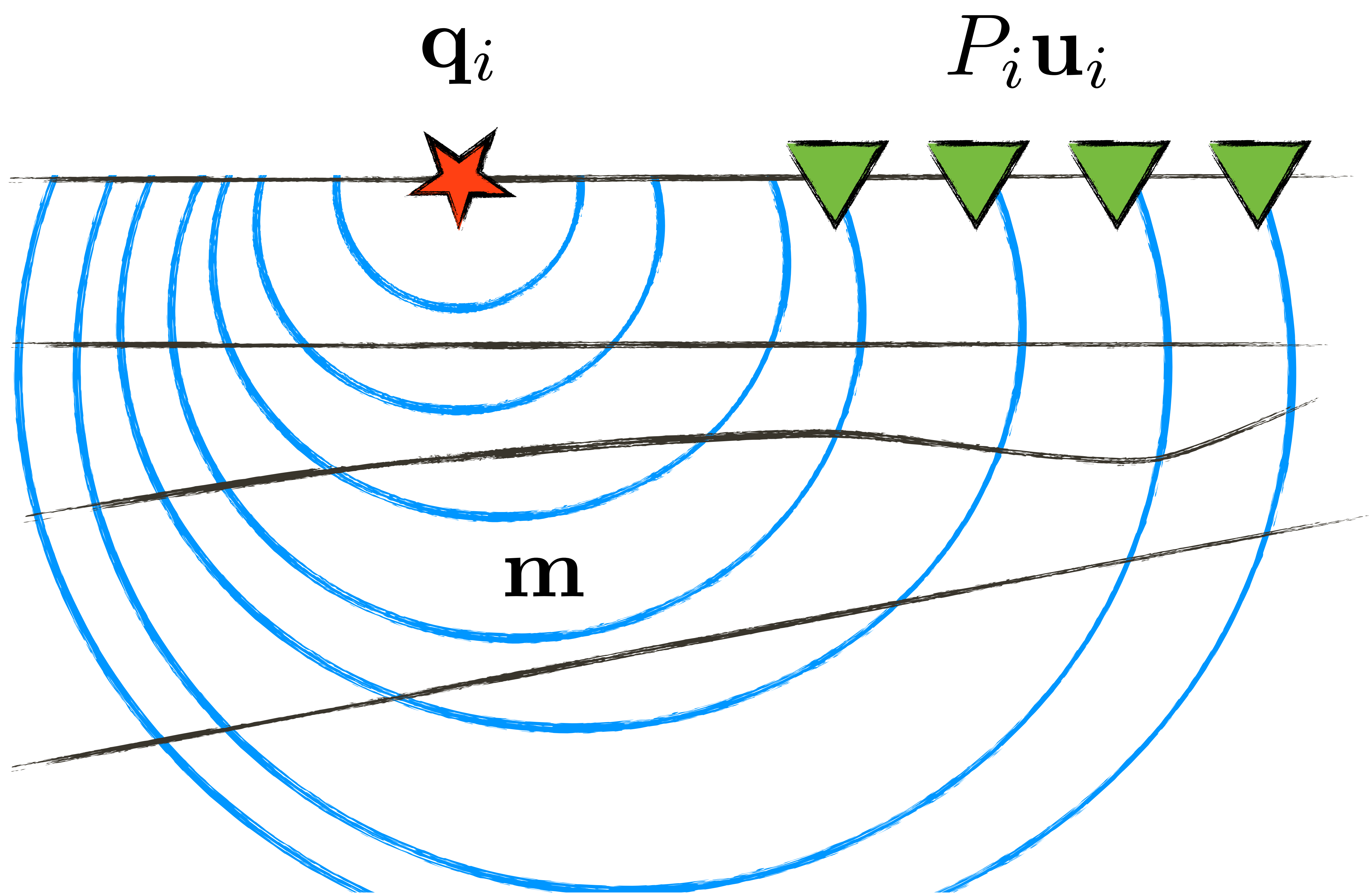
van Leeuwen, T and Herrmann, F J (2013). Mitigating local minima in full-waveform inversion by expanding the search space. Geophysical Journal International.

van Leeuwen, T and Herrmann, F J (2013). A penalty method for PDE-constrained optimization. Submitted for publication

van Leeuwen, T and Herrmann, F J (2013). US Provisional Patent Application No. 61/815,533. A Penalty Method for PDE-Constrained Optimization with Applications to Wave-Equation Based Seismic Inversion

Waveform inversion

Retrieve the medium parameters from *partial* measurements of the *solution* of the *wave*-equation: $A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$



*Wave*form inversion

- If we know the wavefields *everywhere*, we solve for $\mathbf{m}$ from

$$A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

- The challenge is *reconstructing* the wavefields from *partial* measurements

- Repeat in *alternating* fashion

[equation error approach]

WRI–*W*avefield *R*econstruction *I*nversion

- Jointly *fit* observed data

$$P\mathbf{u}_i \approx \mathbf{d}_i$$

- and *wave*-equation

$$A(\mathbf{m})\mathbf{u}_i \approx \mathbf{q}_i$$

- via *least-squares* solution of *data*-augmented *wave*-equation for *fixed* $\mathbf{m}$

$$\begin{pmatrix} P_i \\ A(\mathbf{m}) \end{pmatrix} \mathbf{u}_i \approx \begin{pmatrix} \mathbf{d}_i \\ \mathbf{q}_i \end{pmatrix}$$

- Now, *fix* $\mathbf{u}_i$ and solve

$$\min_{\mathbf{m}} \|A(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\|_2^2$$

PDE-constrained optimization

all-at-once approach

simulated data

simulated wavefied

$$\min_{\mathbf{m},\mathbf{u}} \sum_{i}^{M} ||P_i\mathbf{u}_i - \mathbf{d}_i||_2^2 \quad \text{s.t.} \quad A_i(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

observed data

source

Helmholtz equation

PDE-constrained optimization

Elimination of the *constraint* leads for *all sources* to

$$\min_{\mathbf{m}} \phi_{\text{red}}(\mathbf{m}) = ||PA(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}||_2^2$$

- ***no*** need to *store* ***all*** wavefields (block-elimination)
- suitable for black-box optimization (e.g., LBFGS)
- *need* to *solve* forward & *adjoint* PDEs
- very *non*-linear
- *dense* (GN) Hessian, involves PDE solves

[Tarantola '84]

Penalty formulation
WRI–*W*avefield *R*econstruction *I*nversion

The *traditional* formulation:

- accounts for *errors* in the data
- considers physics as *infallible*

The *penalty* formulation:

- accounts for *errors* in *both* data *and* physics

Penalty formulation

Add *constraint* as *penalty*

$$\min_{\mathbf{m},\mathbf{u}} \phi_\lambda(\mathbf{m},\mathbf{u}) = ||P\mathbf{u}-\mathbf{d}||_2^2 + \lambda^2||A(\mathbf{m})\mathbf{u}-\mathbf{q}||_2^2$$

coincides with *original* problem when $\lambda \uparrow \infty$

[Wright ’00]

Penalty formulation
gradient & Hessian

Eliminate the wavefield by *variable projection*

$$\nabla_{\mathbf{u}}\phi_\lambda(\mathbf{m},\mathbf{u}) = 0$$

Define a *reduced* penalty *objective*:

$$\phi_{\text{pen}}(\mathbf{m}) = \phi_\lambda(\mathbf{m},\mathbf{u}(\mathbf{m}))$$

with

$$\nabla\phi_{\text{pen}} = \nabla_{\mathbf{m}}\phi_\lambda$$
$$\nabla^2\phi_{\text{pen}} = \nabla^2_{\mathbf{m}}\phi_\lambda - \nabla^2_{\mathbf{m},\mathbf{u}}\phi_\lambda\left(\nabla^2_{\mathbf{u}}\phi_\lambda\right)^{-1}\nabla^2_{\mathbf{u},\mathbf{m}}\phi_\lambda$$

[Golub ’73, Aravkin ’12]

Penalty approach

- ***no*** need to store *all* the fields
- ***no*** *adjoint* solves
- *sparse* approximation of GN Hessian for small λ
- *less non*-linear in $\mathbf{m}$
- need to solve *overdetermined* PDE
- *not* clear how to *pick* λ
- ...

Penalty vs. reduced

Penalty method

for each source i

$$\textit{solve} \begin{pmatrix} P \\ \lambda A(\mathbf{m}) \end{pmatrix} \mathbf{u} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \mathsf{diag}(\mathbf{u}_i)^* (A(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i)$$

end

Conventional method

for each source i

$$\textit{solve}\ A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

$$\textit{solve}\ A(\mathbf{m})^* \mathbf{v}_i = P^*(P\mathbf{u}_i - \mathbf{d}_i)$$

$$\mathbf{g} = \mathbf{g} + \omega^2 \mathsf{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

end

Penalty vs. reduced

Penalty method

for each source i

$$\textit{solve} \begin{pmatrix} P \\ \lambda A(\mathbf{m}) \end{pmatrix} \mathbf{u} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \mathsf{diag}(\mathbf{u}_i)^* (A(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i)$$

end

correlation augmented wavefield & *PDE* residual

Conventional method

for each source i

$$\textit{solve}\ A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

$$\textit{solve}\ A(\mathbf{m})^* \mathbf{v}_i = P^*(P\mathbf{u}_i - \mathbf{d}_i)$$

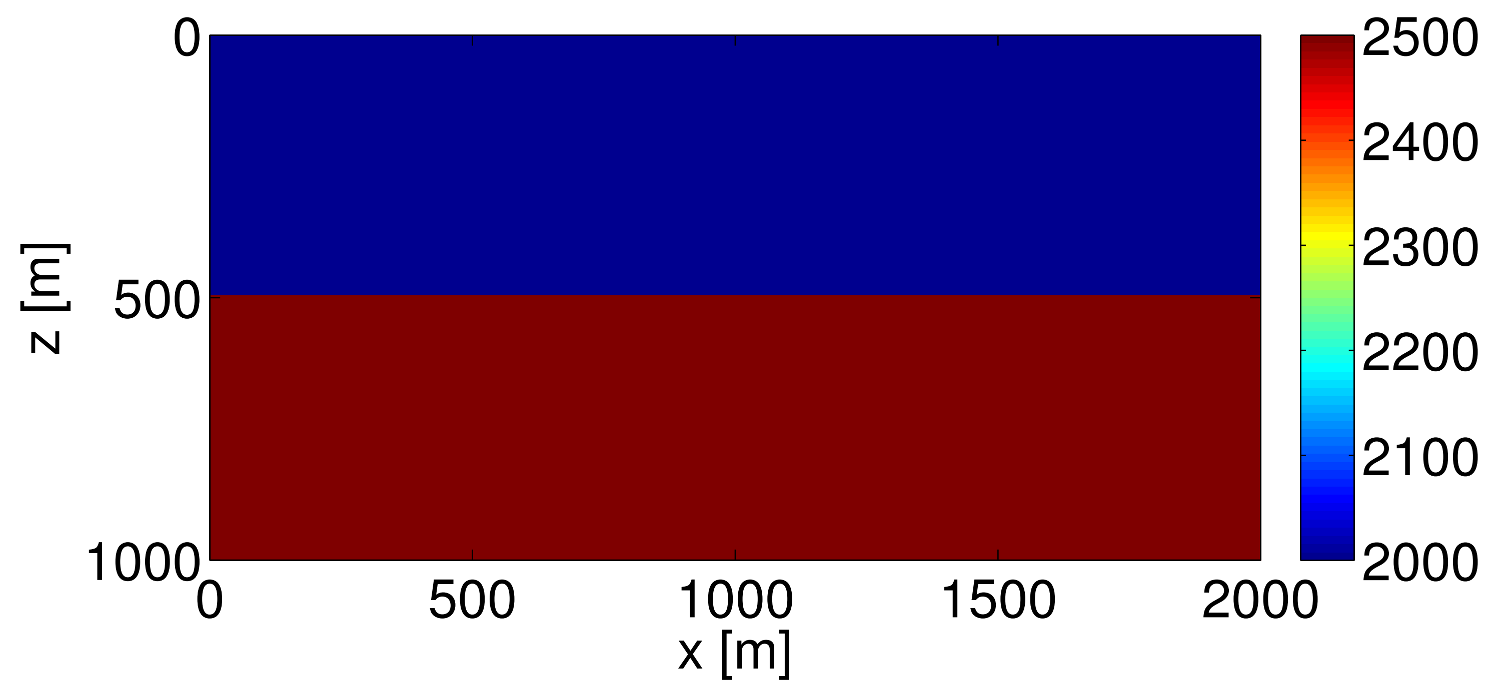
$$\mathbf{g} = \mathbf{g} + \omega^2 \mathsf{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

end

correlation wavefield & *data* residual

One reflector example

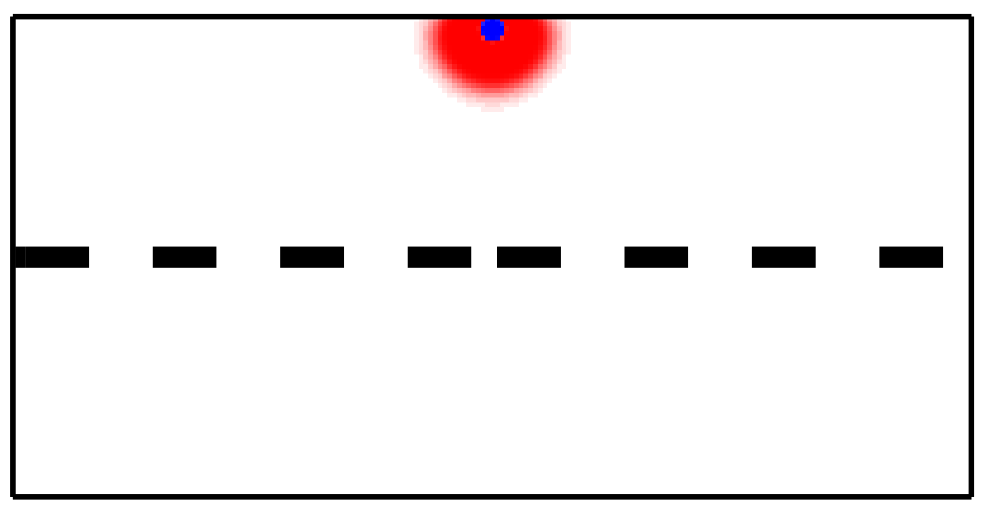
true model



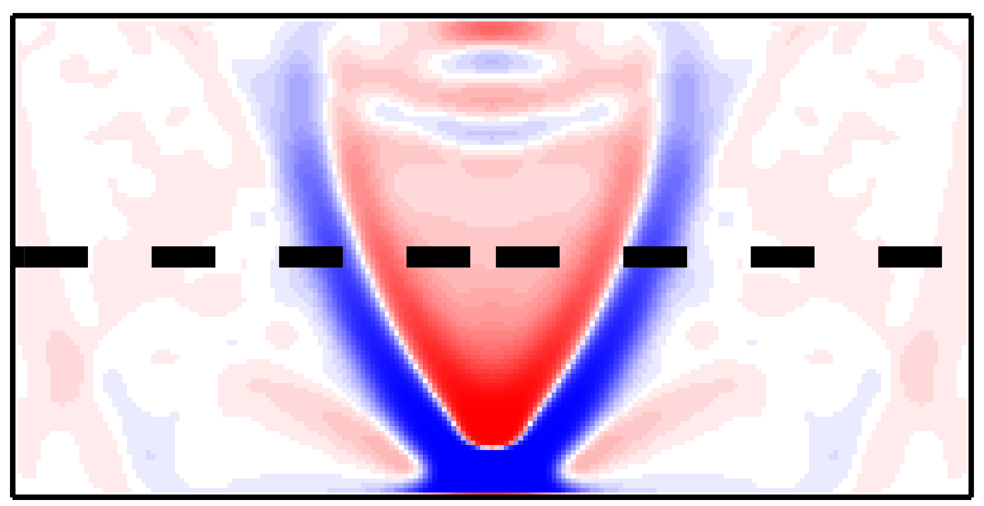
Wavefields in *homogeneous* background

FWI

forward

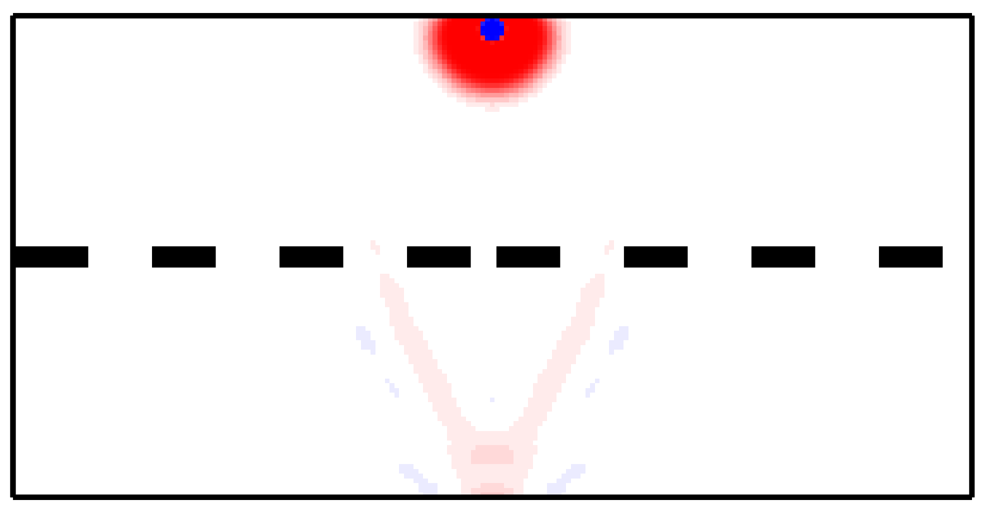


adjoint

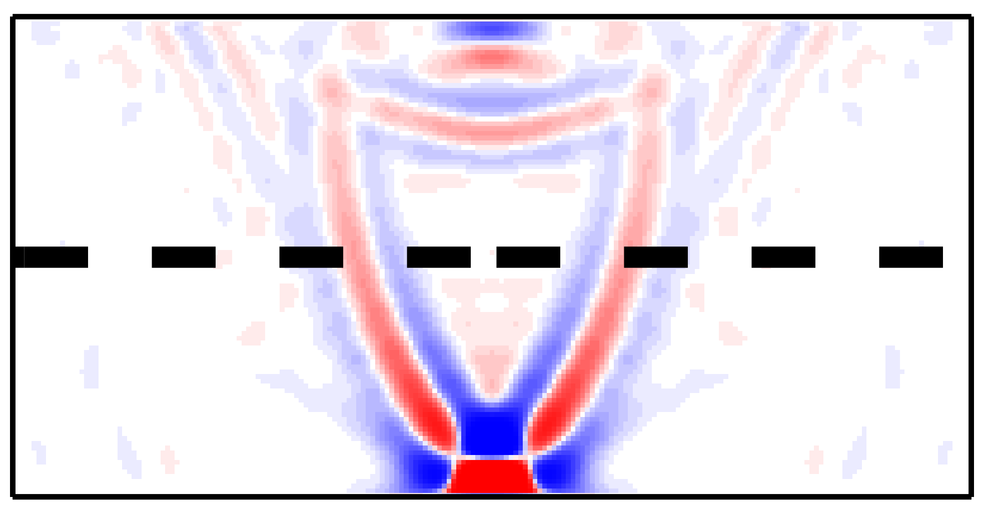


WRI

reconstructed wavefield



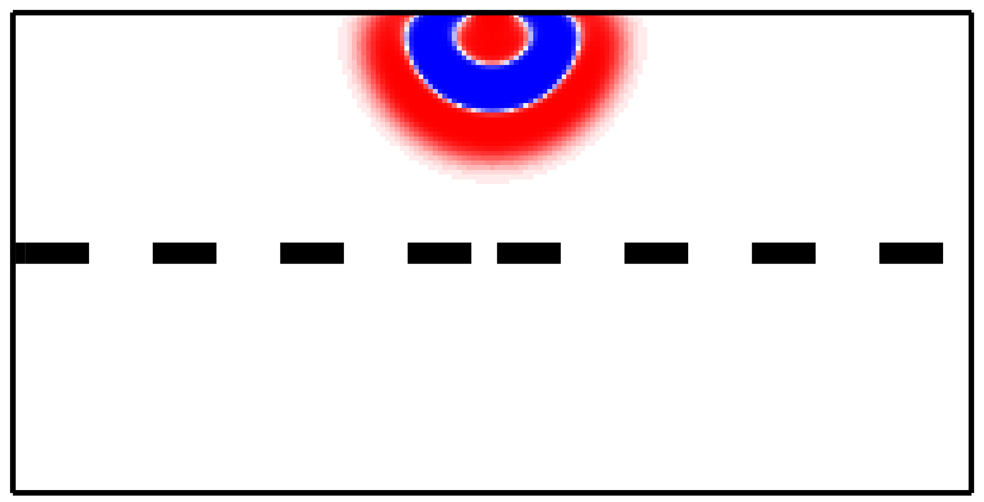
PDE residual



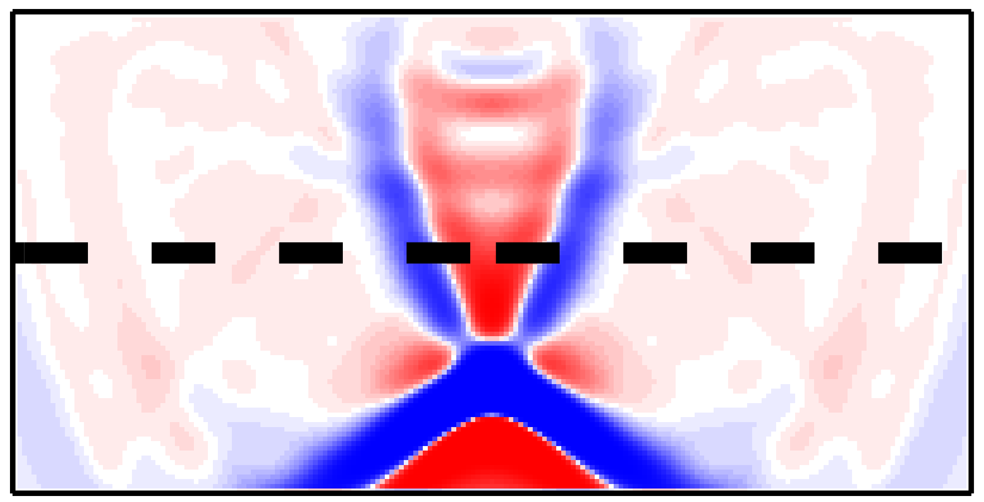
Wavefields in *homogeneous* background

FWI

forward

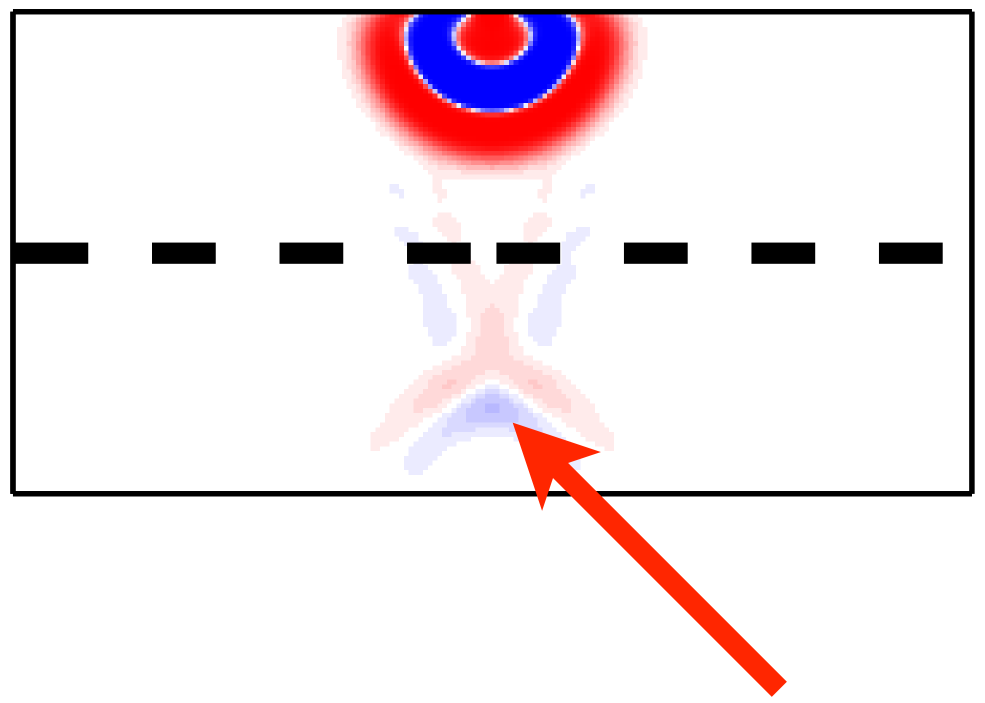


adjoint

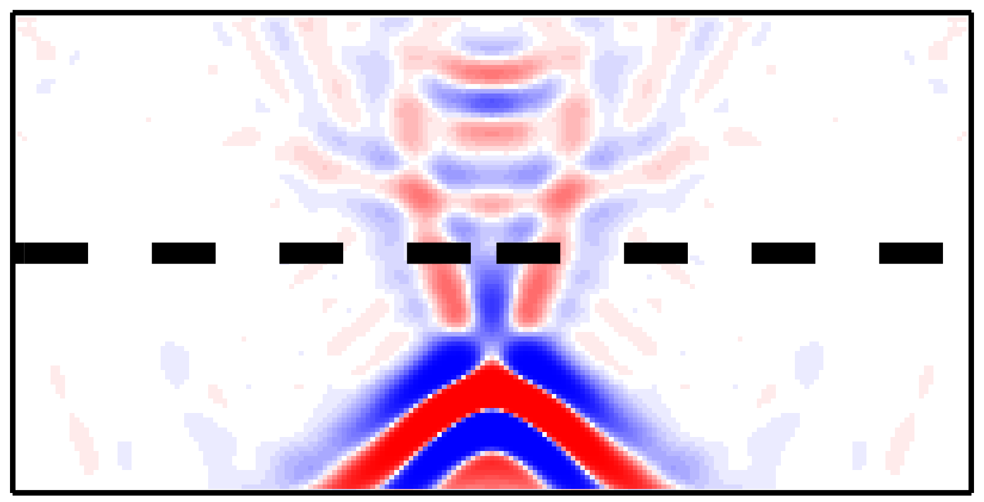


WRI

reconstructed wavefield



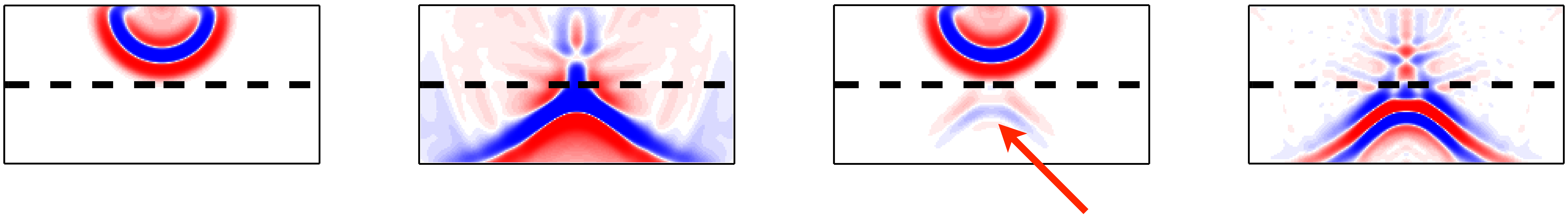
PDE residual



Wavefields in *homogeneous* background

FWI **WRI**

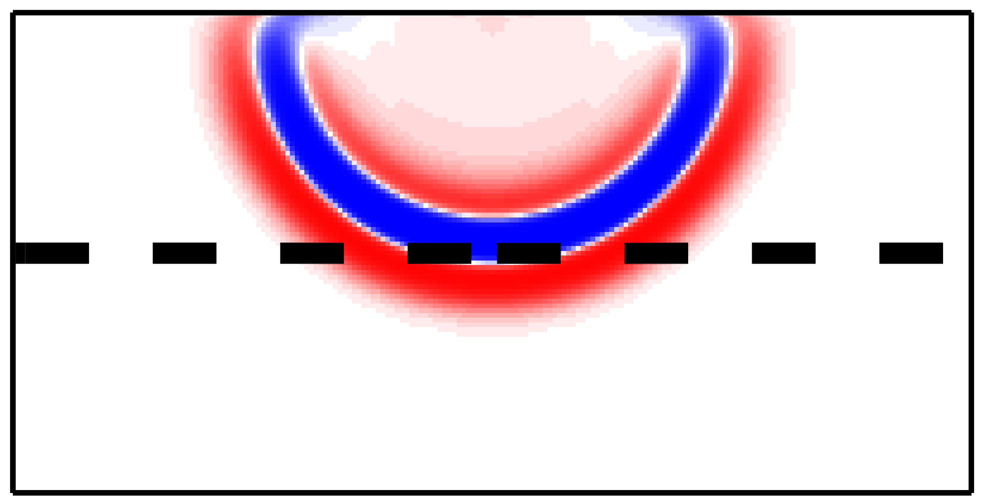
forward adjoint reconstructed wavefield PDE residual



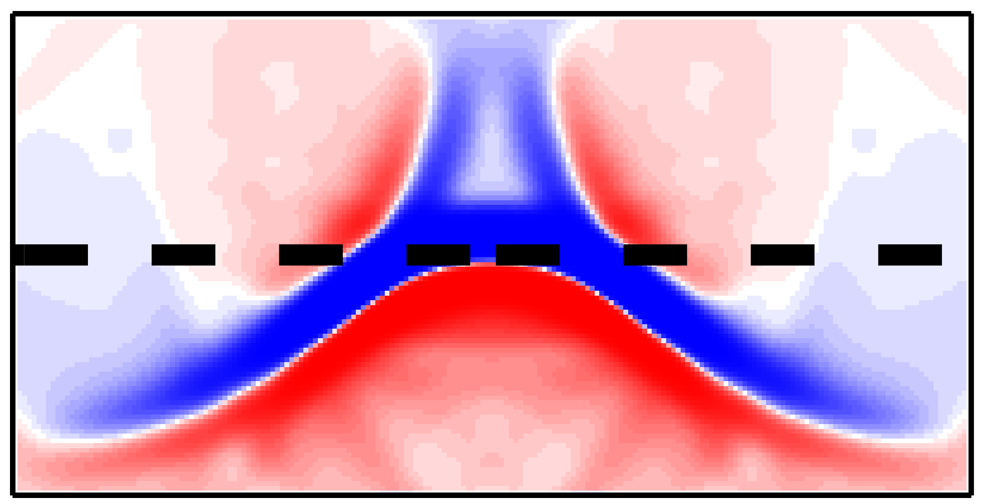
Wavefields in *homogeneous* background

FWI

forward

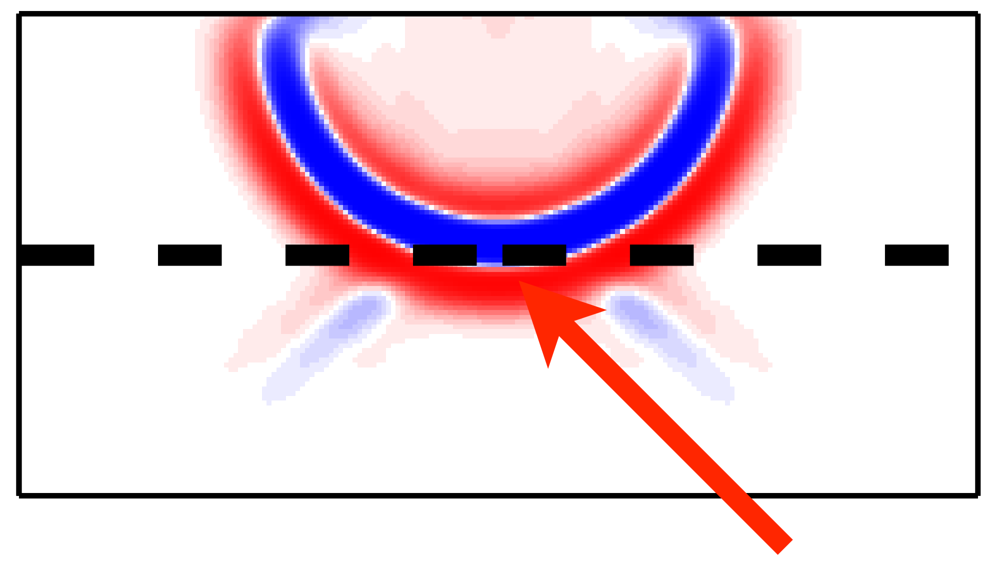


adjoint

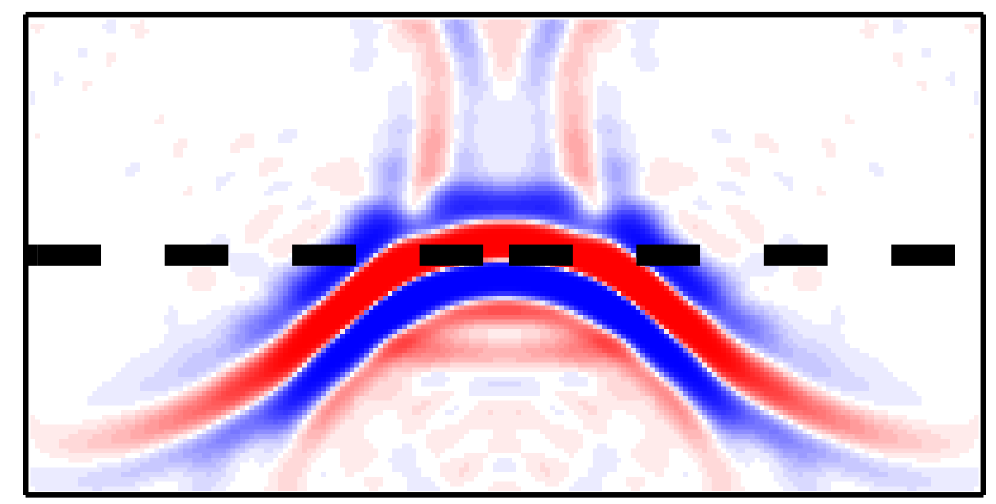


WRI

reconstructed wavefield

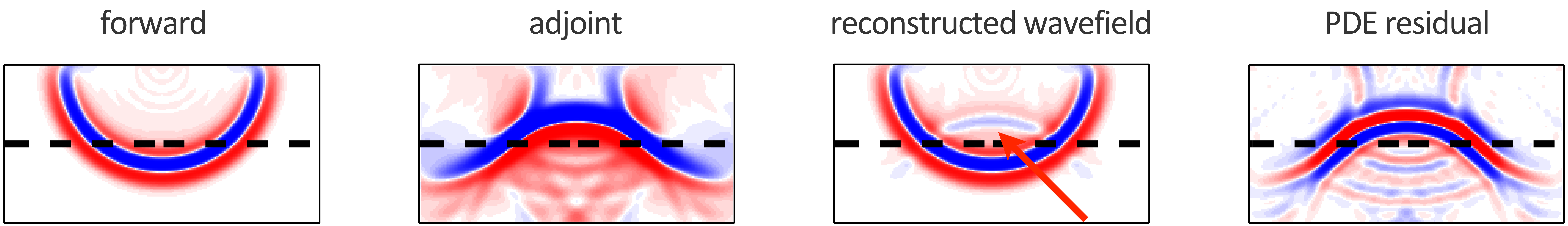


PDE residual



Wavefields in *homogeneous* background

FWI **WRI**



Stylized examples

Penalty vs. reduced

Penalty method

for each source i

$$\textit{solve} \begin{pmatrix} P \\ \lambda A(\mathbf{m}) \end{pmatrix} \mathbf{u} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \mathsf{diag}(\mathbf{u}_i)^* (A(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i)$$

$$H_{GN} = H_{GN} + \lambda^2 \omega^4 \mathsf{diag}(\mathbf{u}_i)^* \mathsf{diag}(\mathbf{u}_i)$$

$$\mathbf{m} := \mathbf{m} - \alpha H_{GN}^{-1} \mathbf{g}$$

end

Conventional method

for each source i

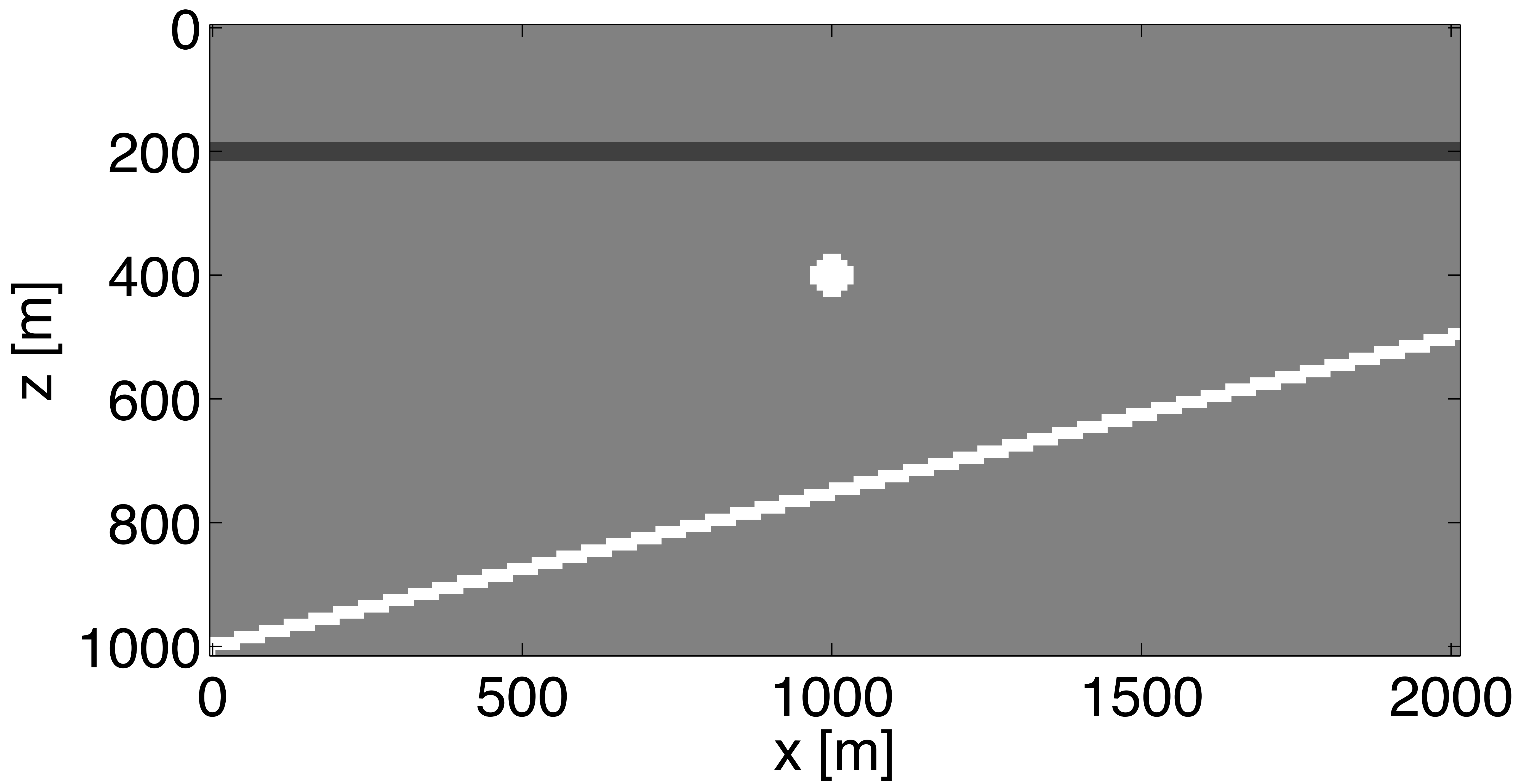
$$\textit{solve}\ A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

$$\textit{solve}\ A(\mathbf{m})^* \mathbf{v}_i = P^*(P\mathbf{u}_i - \mathbf{d}_i)$$

$$\mathbf{g} = \mathbf{g} + \omega^2 \mathsf{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

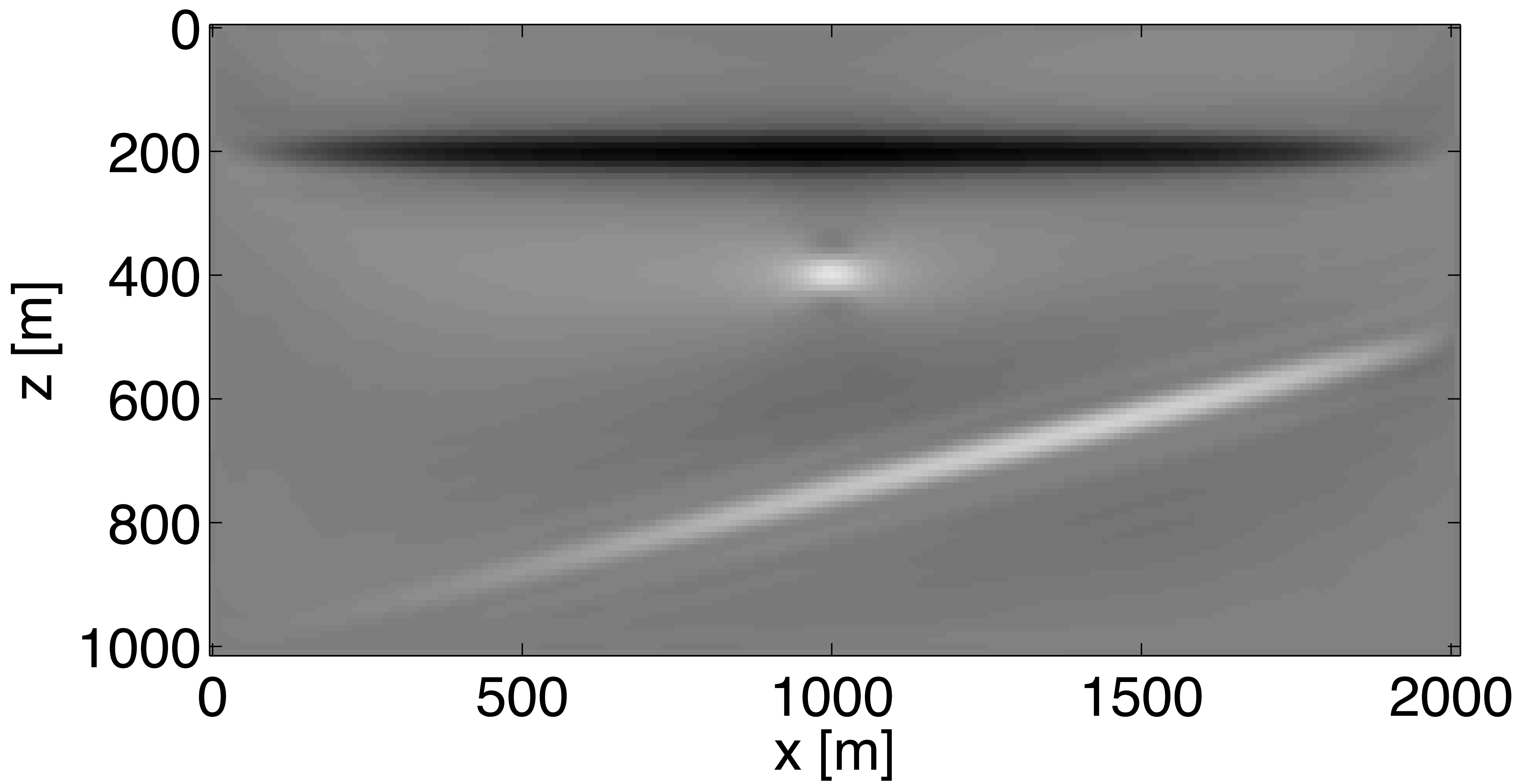
end



Original **image**

Imaging

The *gradient* of the reduced objective yields an *image* of the subsurface..



RTM-migration

Imaging

Conventional **RTM:**

1. solve forward wave-equation
2. solve adjoint wave-equation
3. apply `imaging condition’

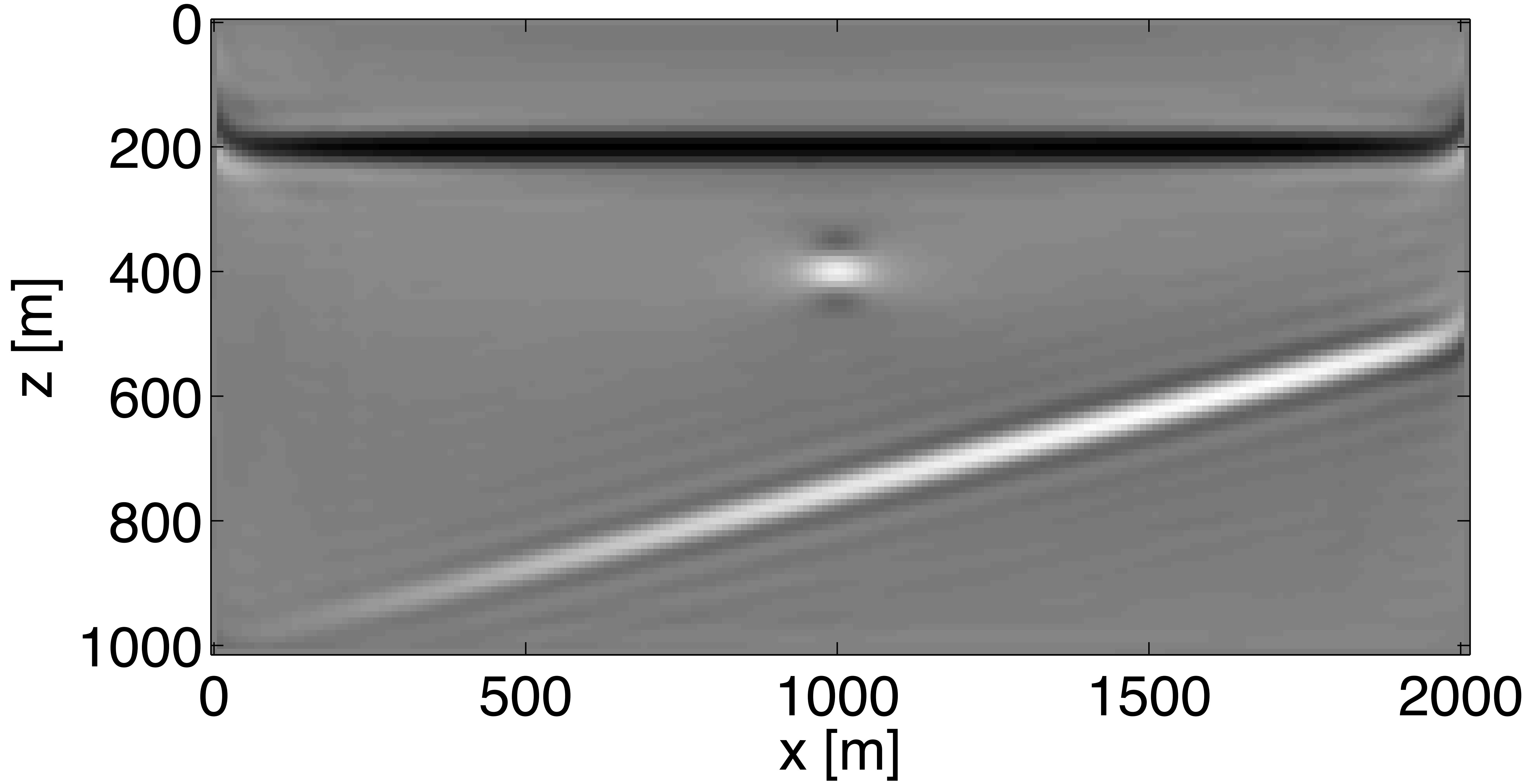


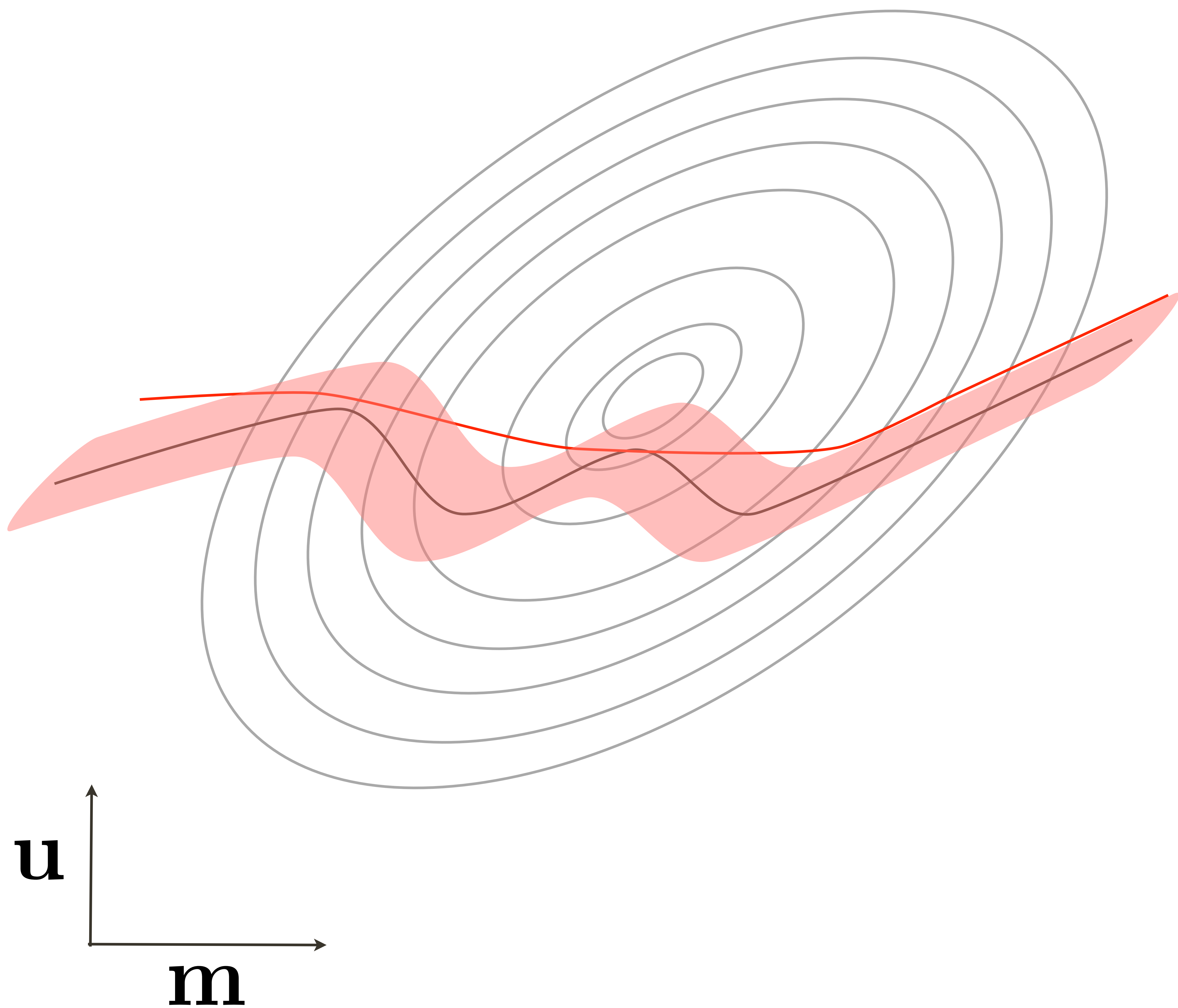
Image w/ penalty method

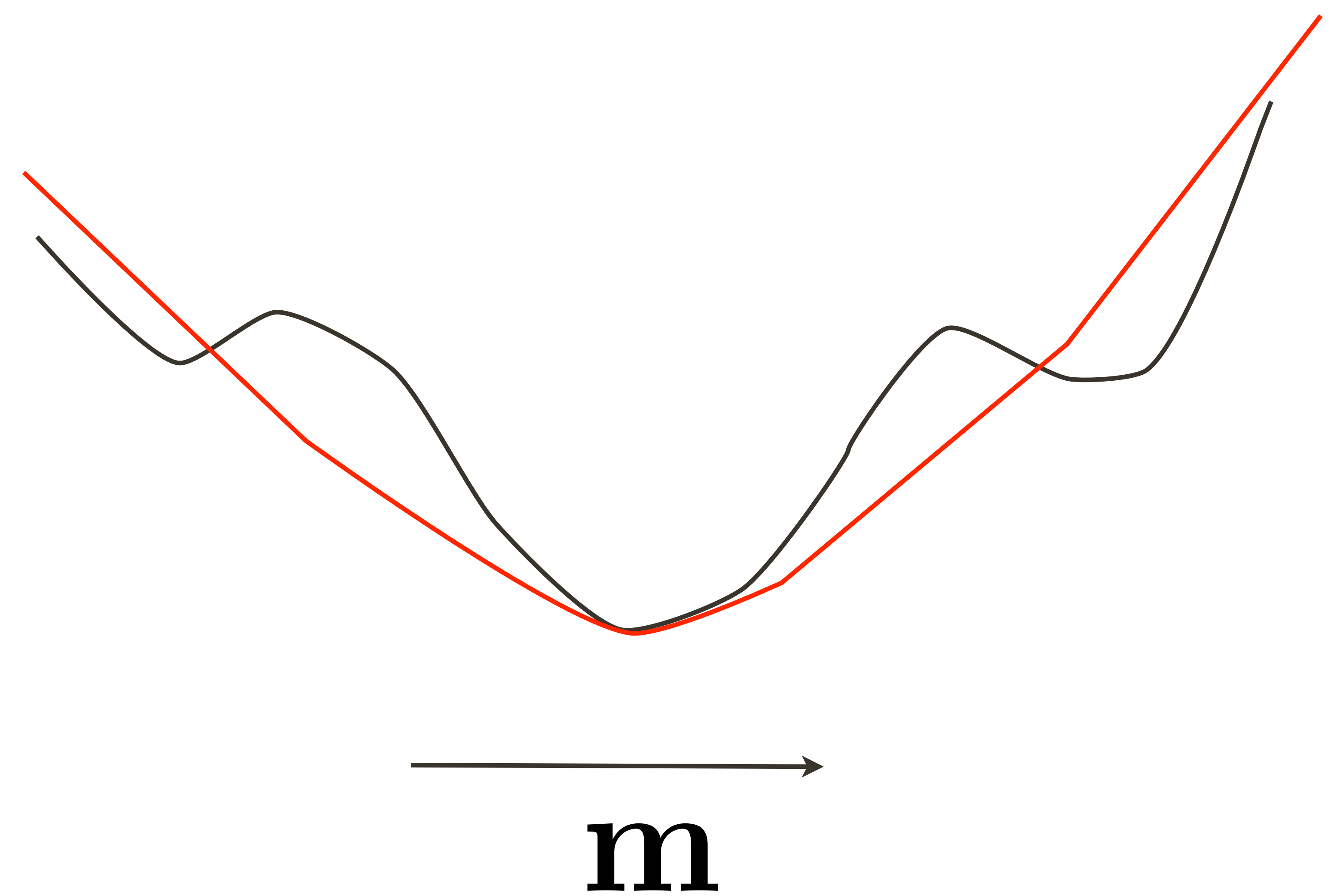
Imaging

***Penalty*-method reverse-time migration:**

1. solve overdetermined wave-equation
2. go for lunch
3. apply `imaging condition’

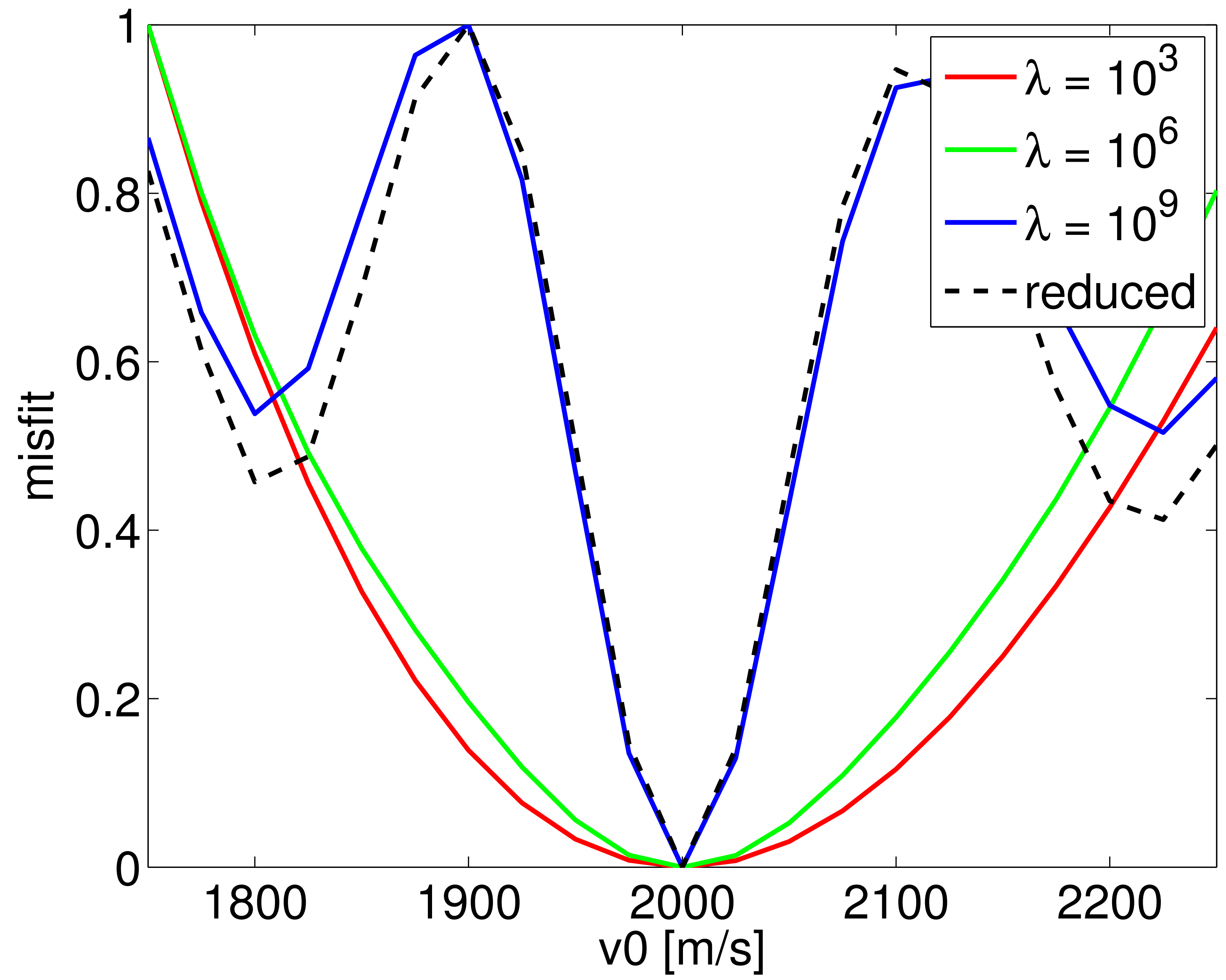
Penalty vs. reduced



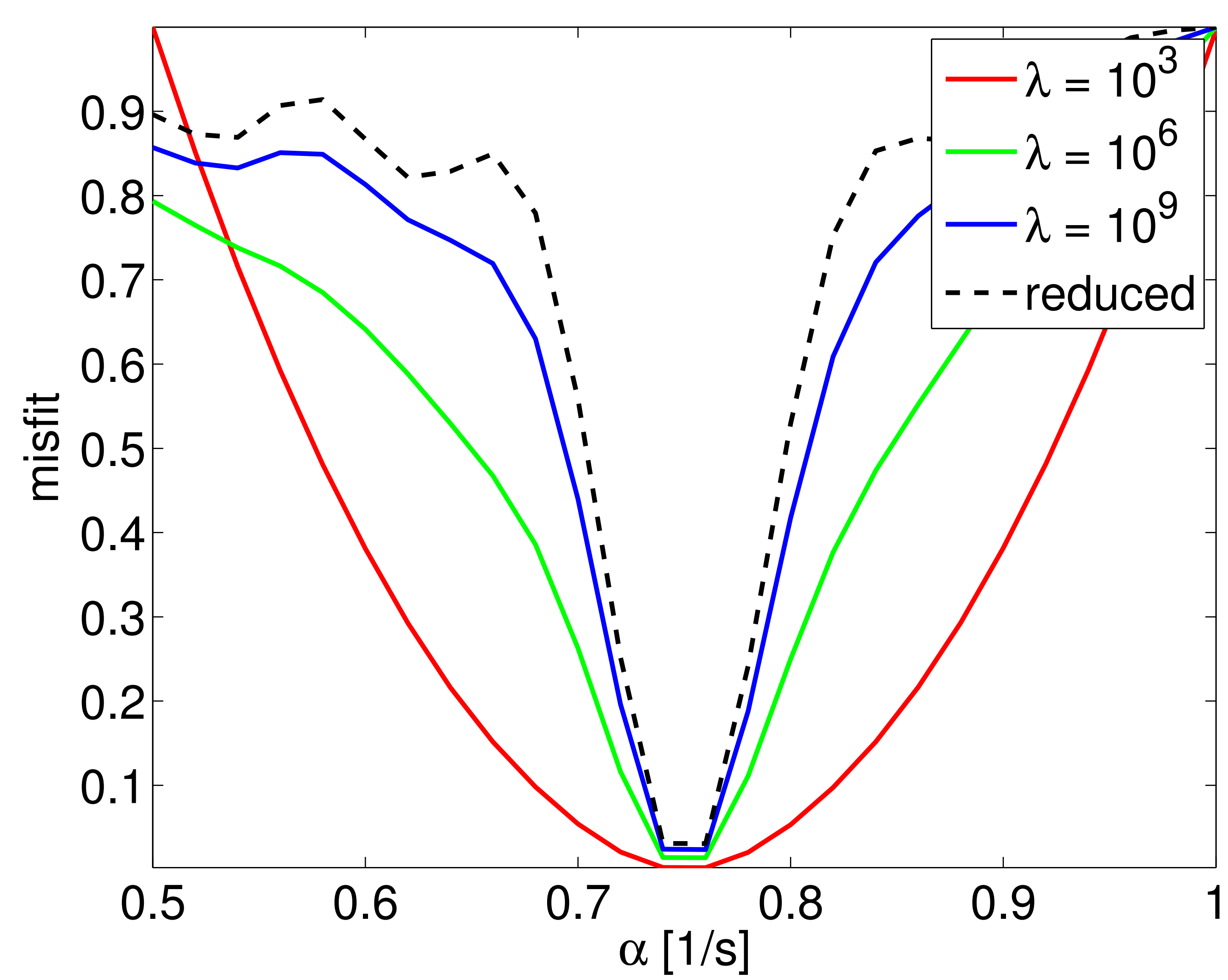


Local minima

single shot, single frequency data for *linear* velocity profile $v(z) = v_0 + \alpha z$,
misfit as function of (v_0, α)

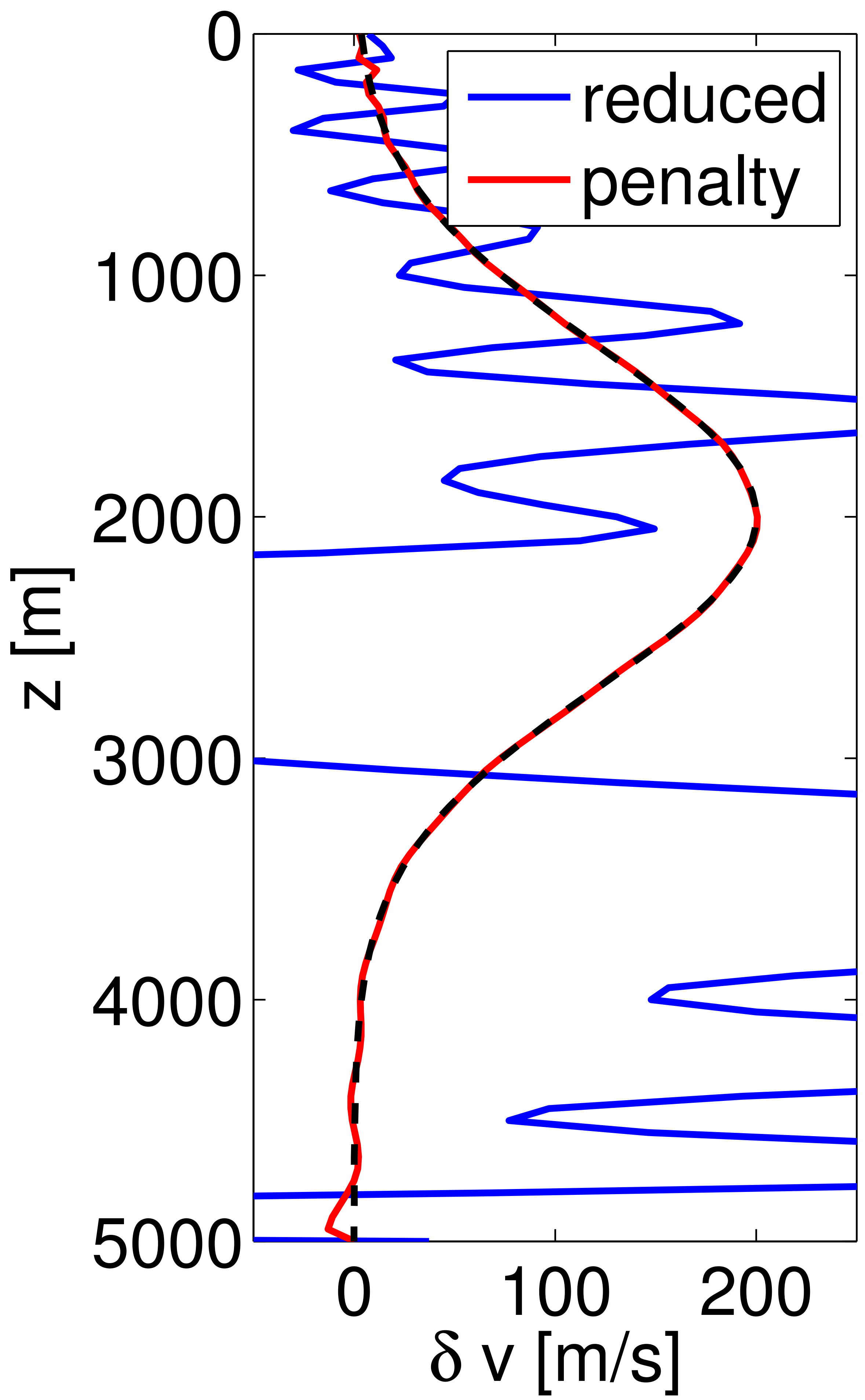


constant

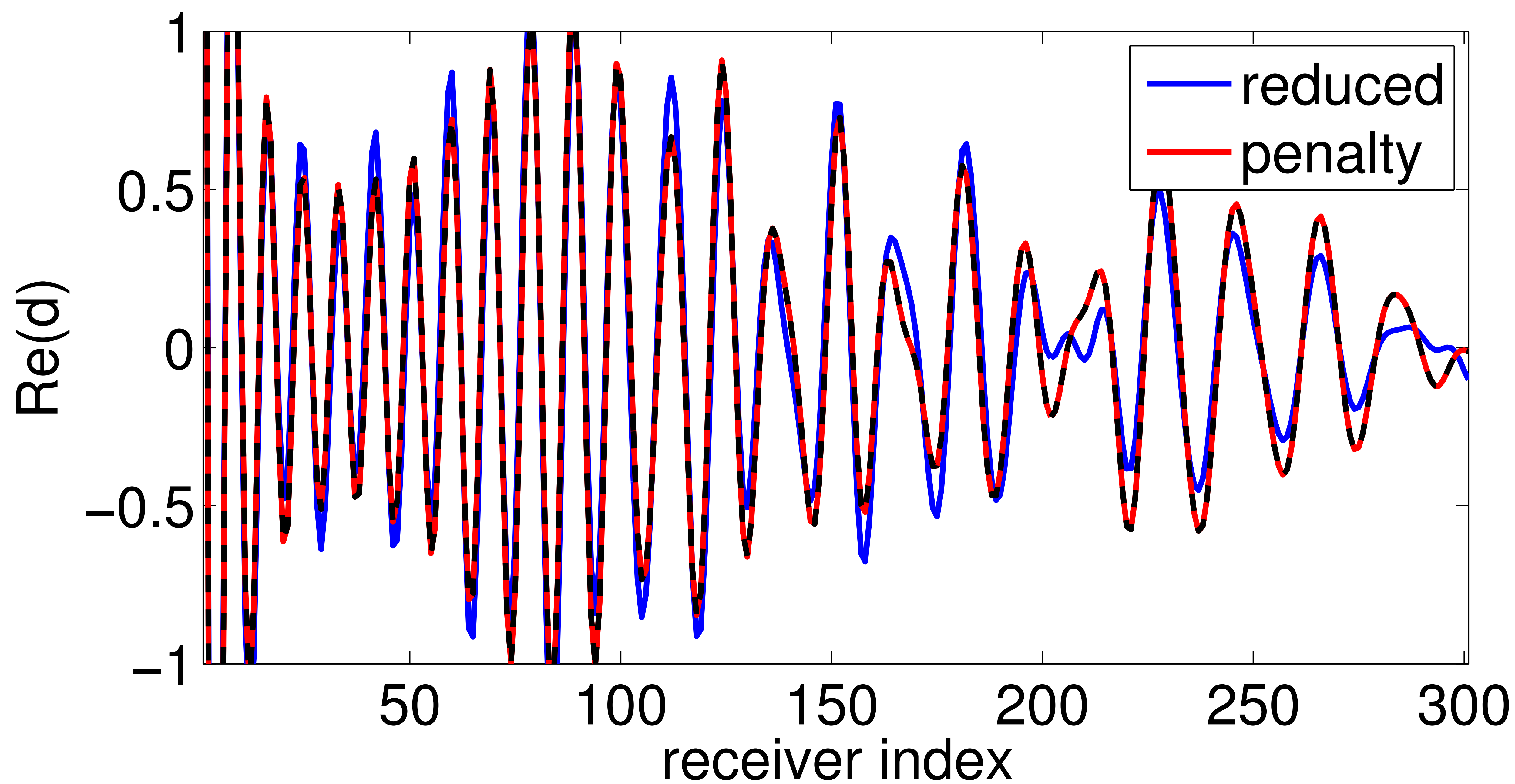


slope

Local minima



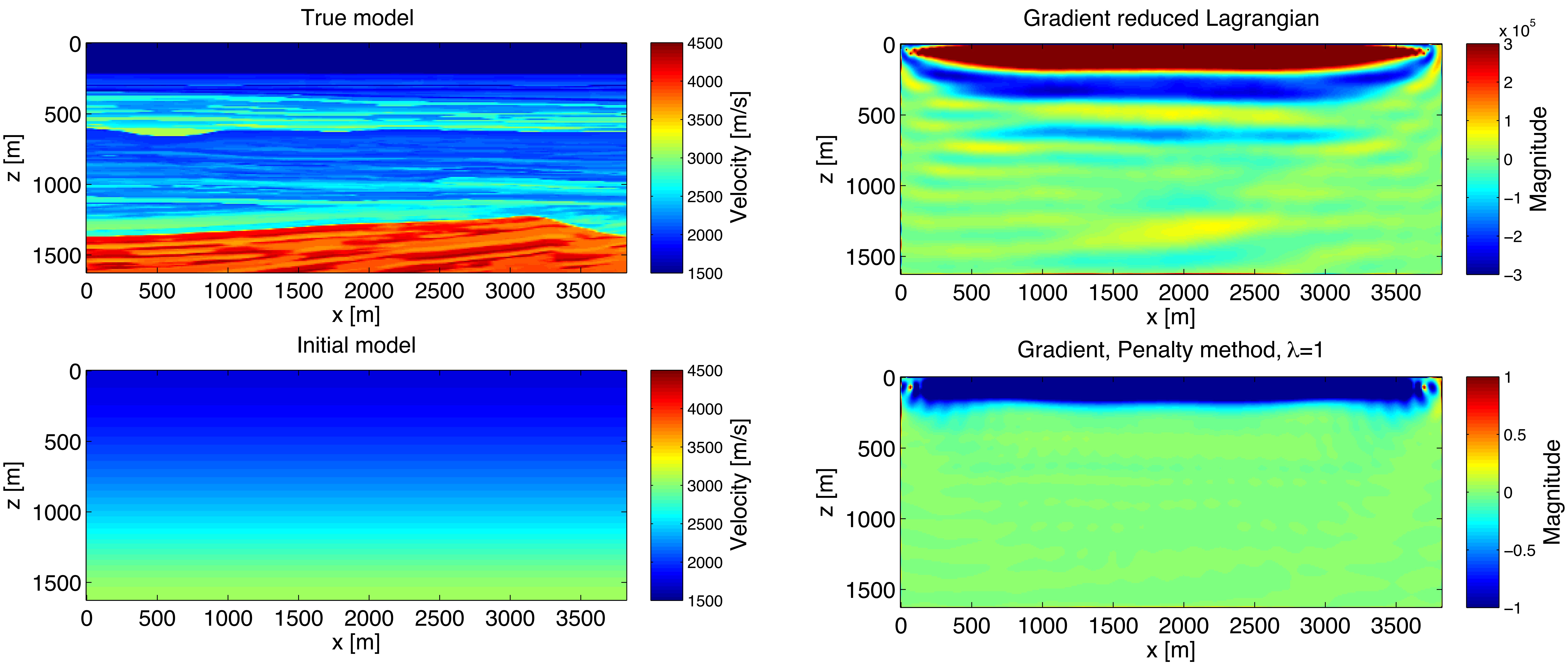
After 1000 iterations



Example – BG Compass model

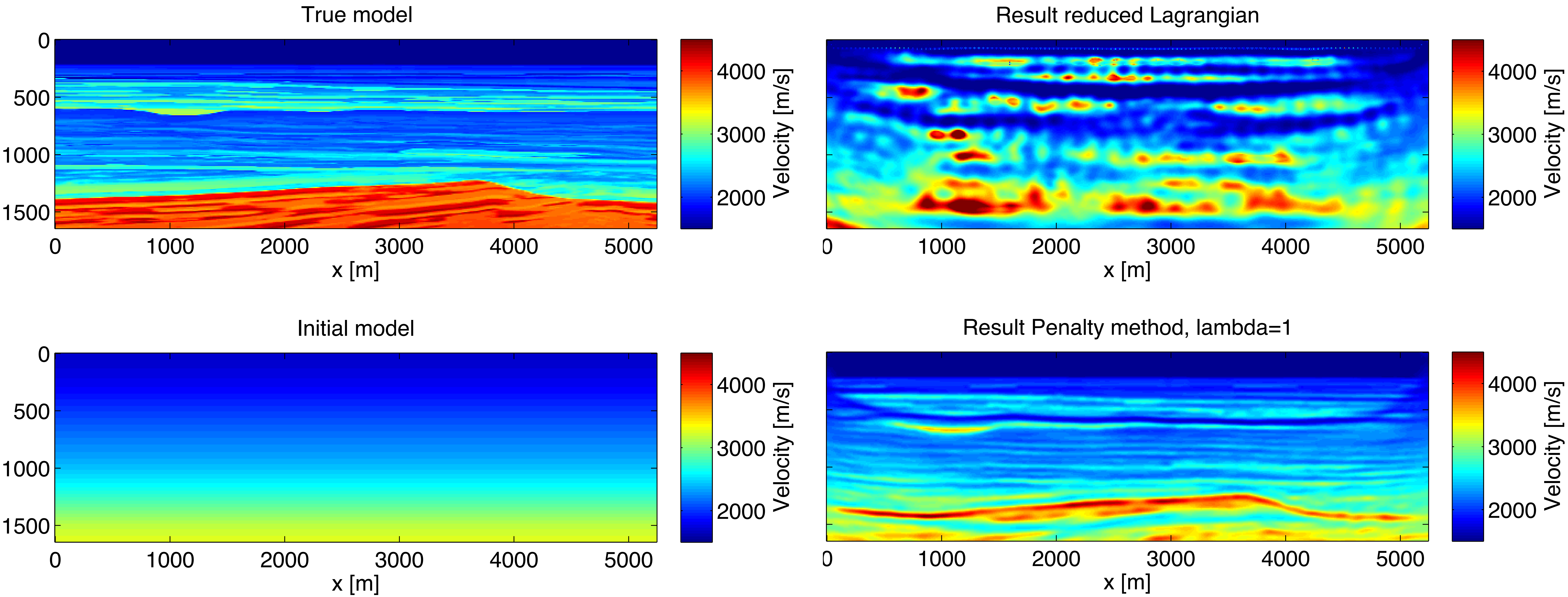
- *Low* frequencies missing, 15 frequency batches (10 iterations each) {5 6} ,{6 7},… ,{19 20} Hertz
- Ricker wavelet with 30Hz peak frequency
- Data contains random noise
- Inaccurate initial model
- Max offset: 5 km, 135 sources & receivers

A *look* at the *first* gradient...

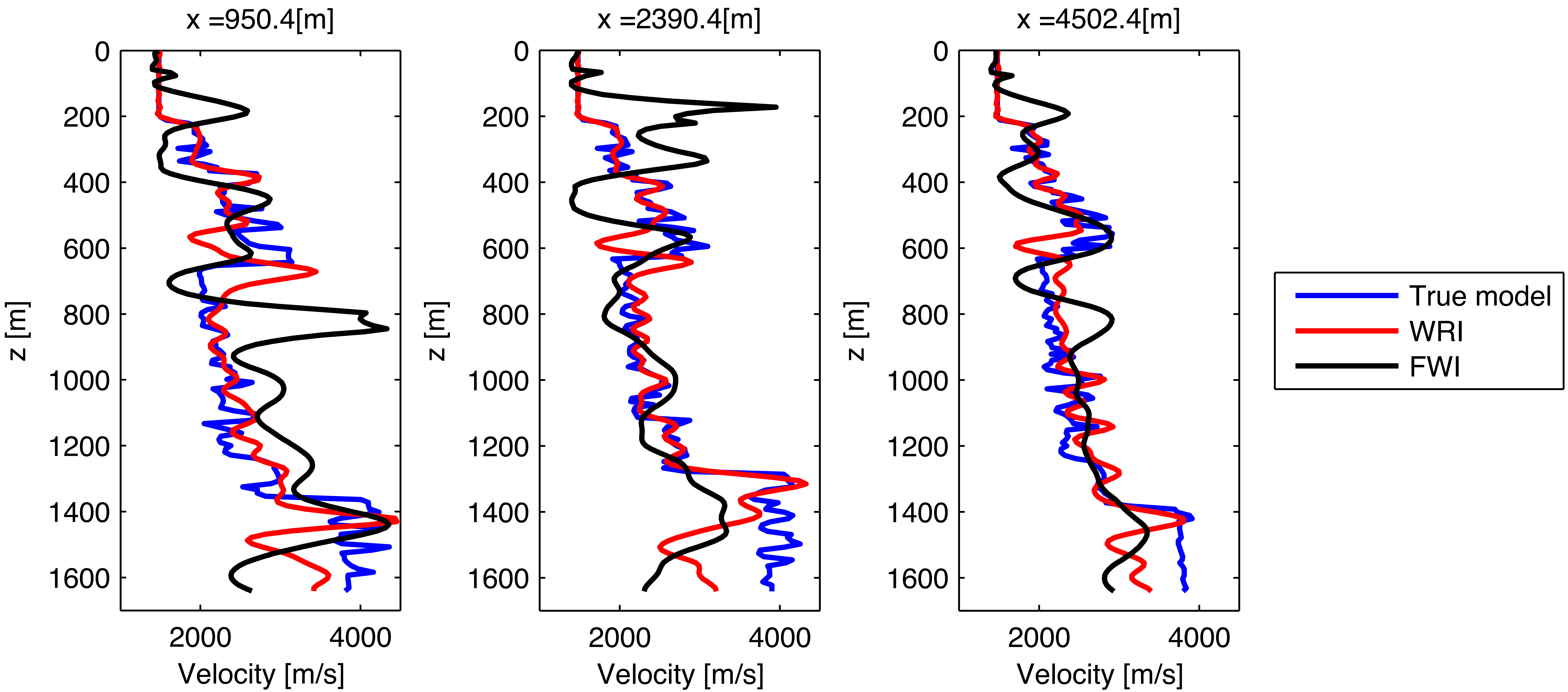


Gradients which will be the first updates, Frequency band: 7 Hz

True, initial & final models

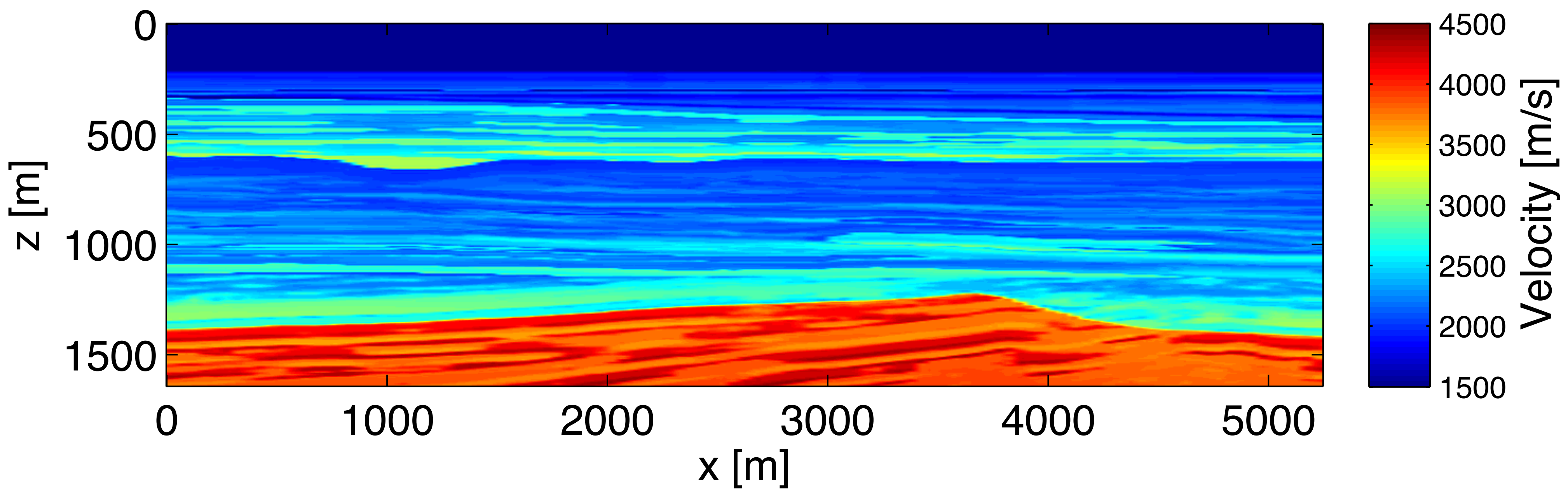


Cross-sections

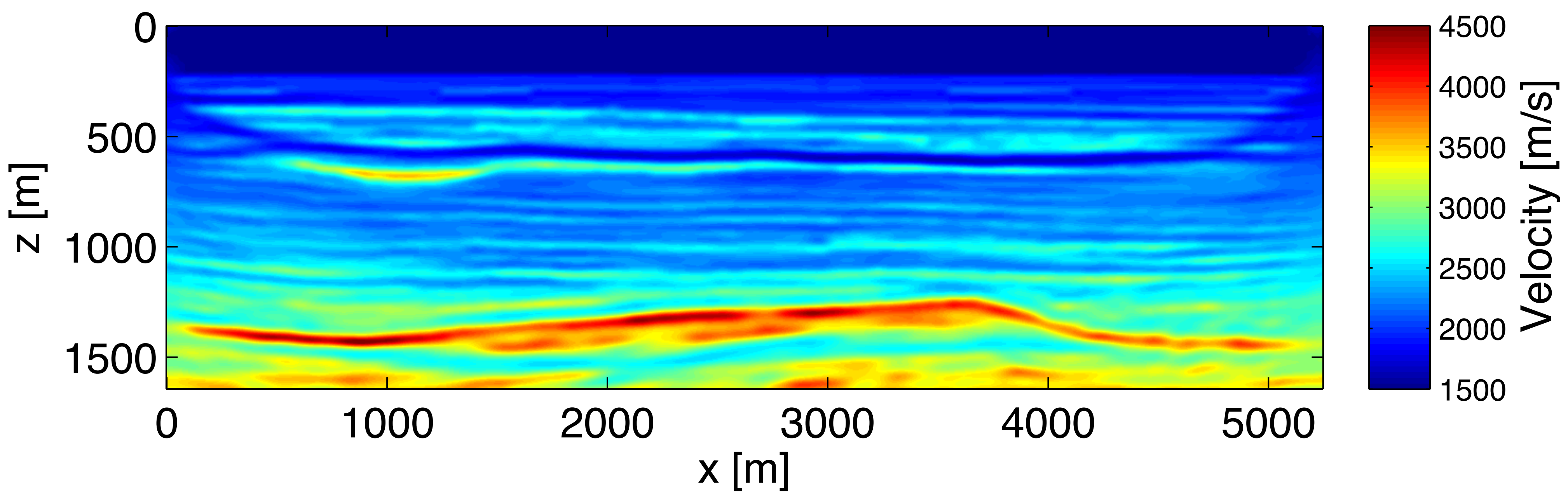


Imaging

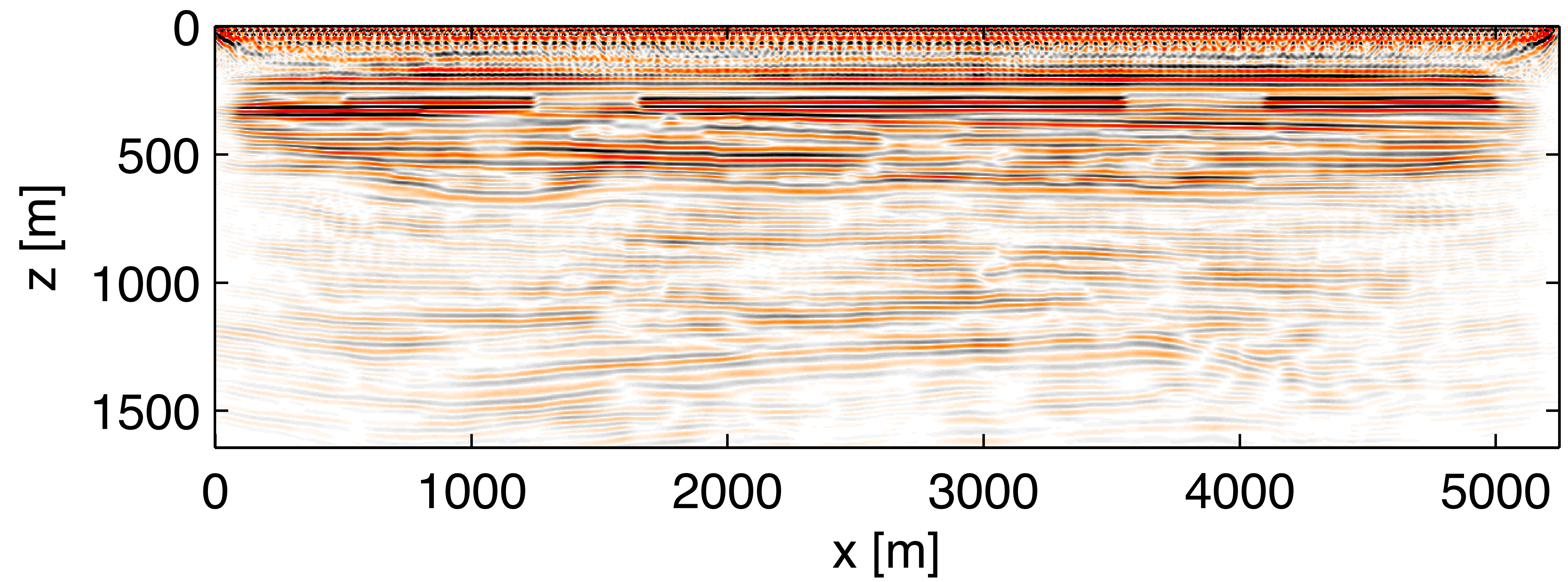
True model



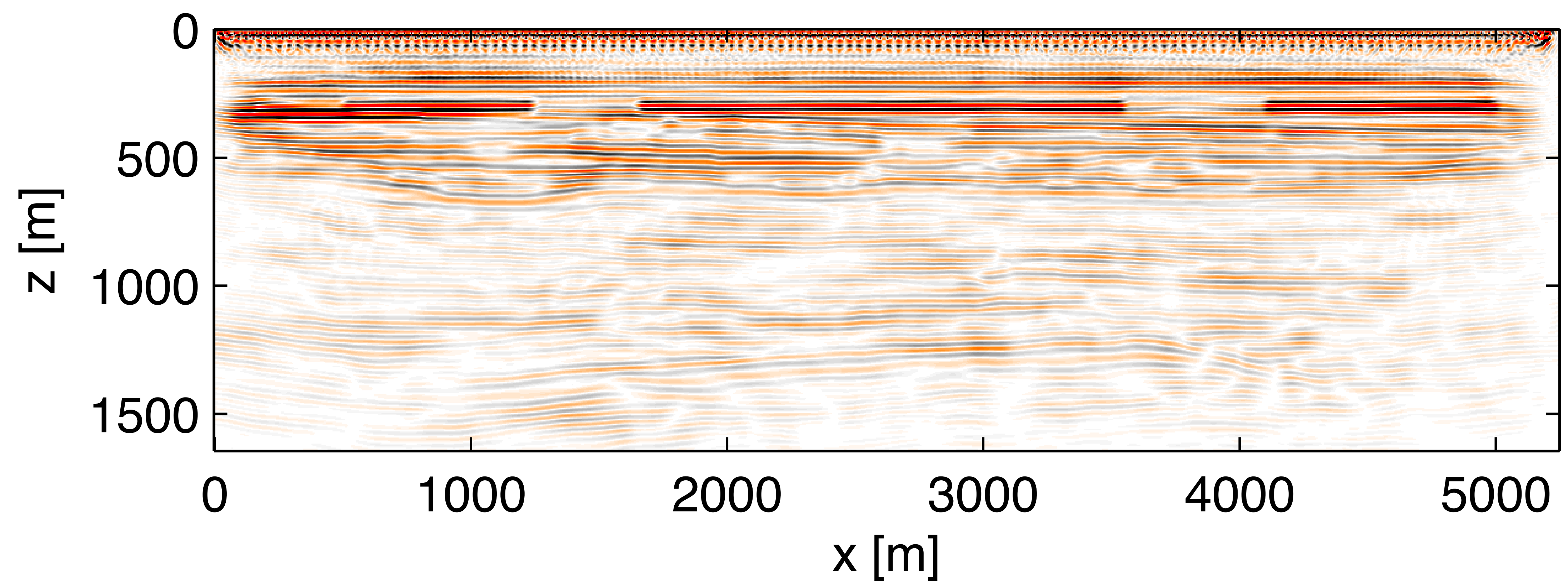
Initial model



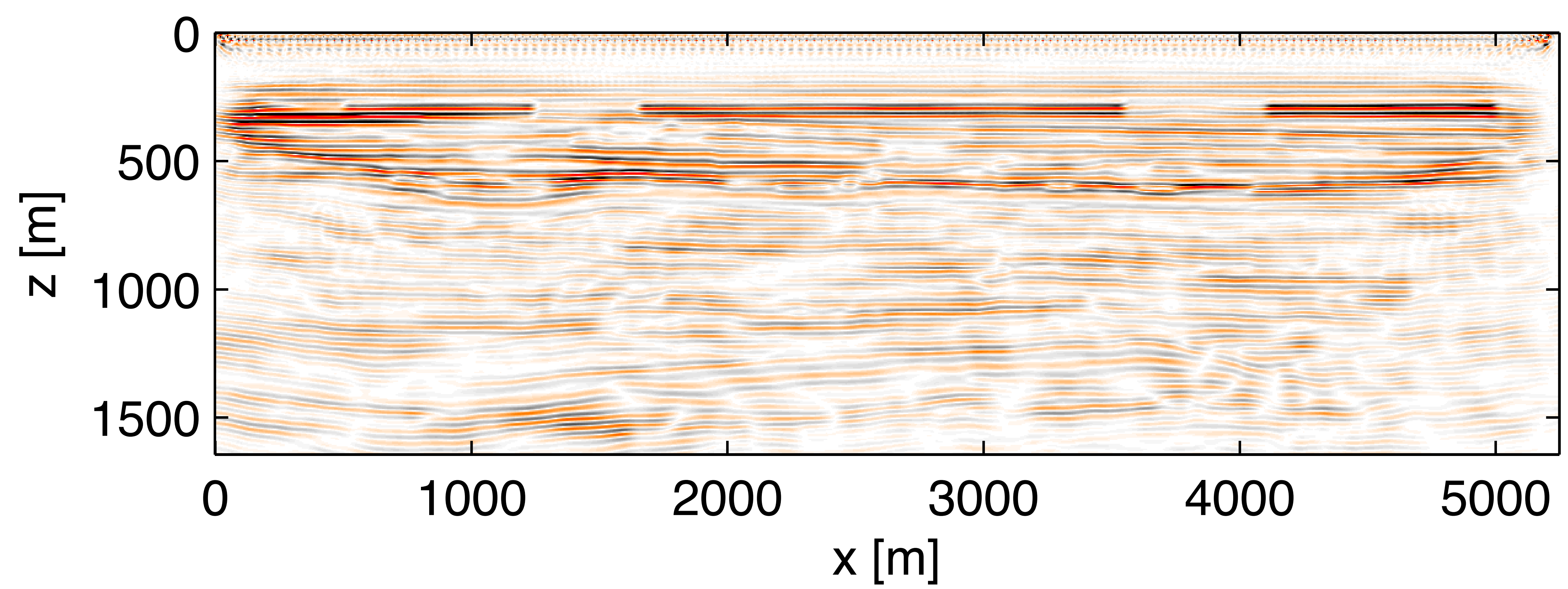
Result reduced Lagrangian



Result WRI, λ=1



Result WRI (GN), λ=1



Ernie Esser, Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann, "**A scaled gradient projection method for total variation regularized full waveform inversion**". 2014.
Bas Peters and Felix J. Herrmann, "**A sparse reduced Hessian approximation for multi-parameter Wavefield Reconstruction Inversion**". 2014.

Extensions

Total-variation regularization

via *scaled* gradient *projections* & *bound* constraints

Multi-parameter case

via *sparse* approximate Gauss-Newton *scaling*

Scaled Gradient Projection w/ *Total* Variation & Bound Constraints

Solve

$$\min_{\mathbf{m}} \mathbf{g}(\mathbf{m}) \quad \text{subject to} \quad \begin{cases} m_i \in [B_1, B_2] \\ \|\mathbf{m}\|_{TV} \leq \tau \end{cases}$$

with $\|\mathbf{m}\|_{TV} = \sum_{ij} \frac{1}{h} \left\| \begin{bmatrix} (m_{i,j+1} - m_{i,j}) \\ (m_{i+1,j} - m_{i,j}) \end{bmatrix} \right\|$

by iterating (*outer* loop)

$$\Delta\mathbf{m} = \arg\min_{\Delta\mathbf{m}} \Delta\mathbf{m}^T \nabla\mathbf{g}(\mathbf{m}) + \frac{1}{2}\Delta\mathbf{m}^T H_{GN} \Delta\mathbf{m} + c\Delta\mathbf{m}^T\Delta\mathbf{m}$$

$$\text{subject to} \;\; \mathbf{m} + \Delta m_i \in [B_1, B_2] \text{ and } \|\mathbf{m} + \Delta\mathbf{m}\|_{TV} \leq \tau$$

$$\mathbf{m} = \mathbf{m} + \Delta\mathbf{m}$$

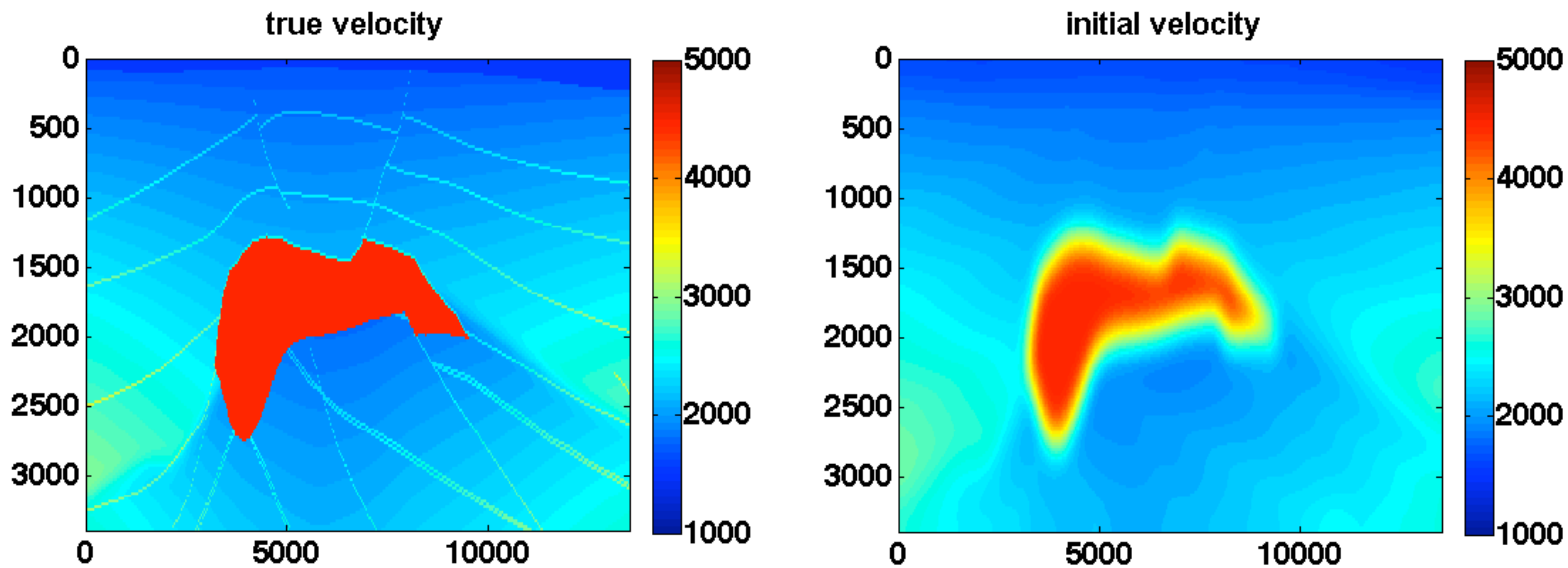
Scaled Gradient Projection
convex subproblems

Solve *convex* subproblems (*inner* loop) for $\Delta\mathbf{m}$ by

- a *primal-dual* method that *alternates* projections *onto* an ℓ_1 *ball* & the *bound* constraints

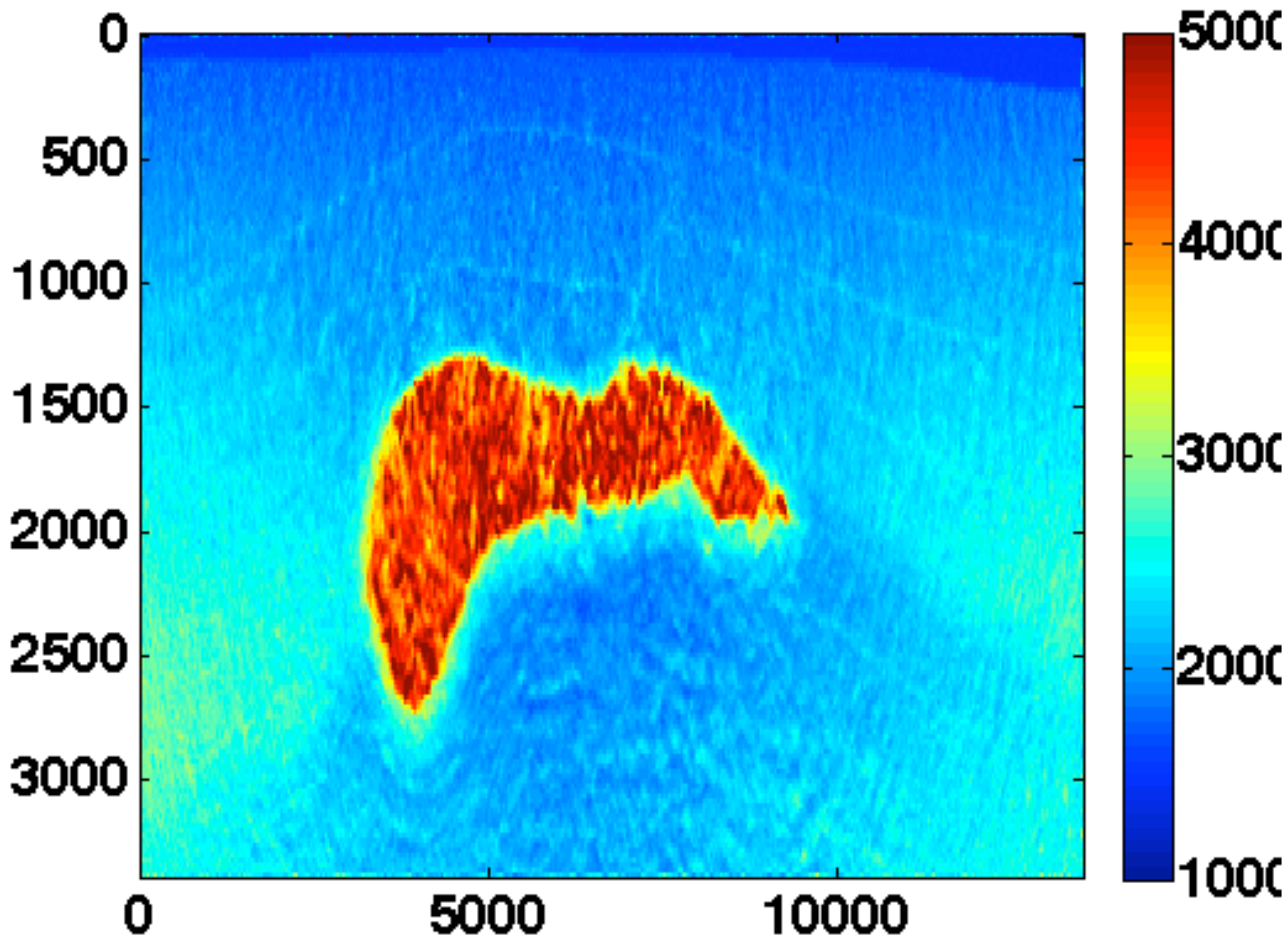
Computationally *efficient* because

- H_{GN} is *diagonal*
- ***no*** need to recompute *gradients* ($\mathbf{g}$) or *Hessian* approximations

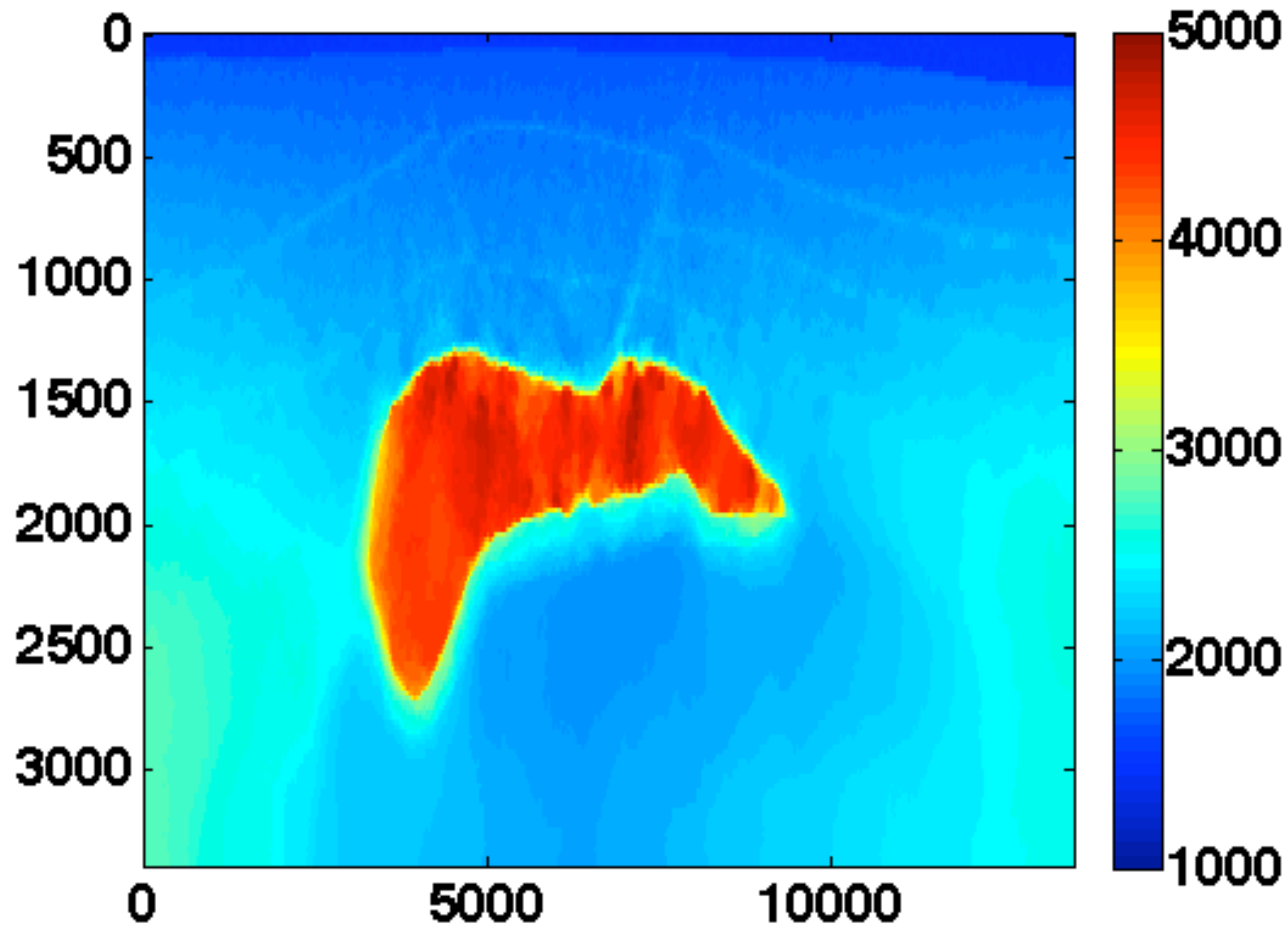


SEG/EAGE model

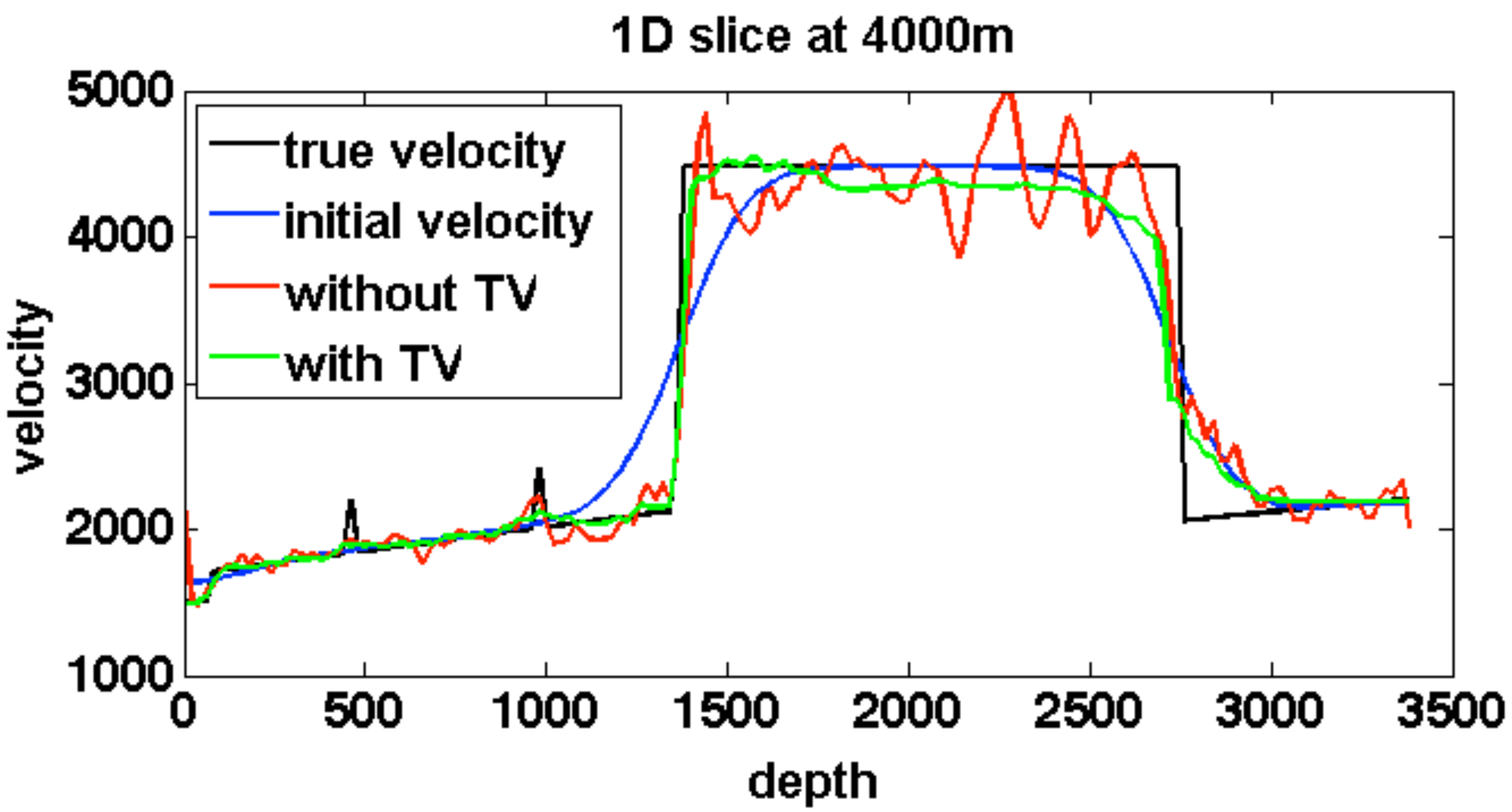
- # of sources= 116, # of receivers= 673
- 2 *simultaneous* sources
- frequency range: 3-33Hz in *overlapping* batches of 2
- maximum number of outer iterations per frequency batch: 25
- maximum number of inner iterations for convex subproblems: 2000
- known Ricker wavelet sources with 30Hz peak frequency

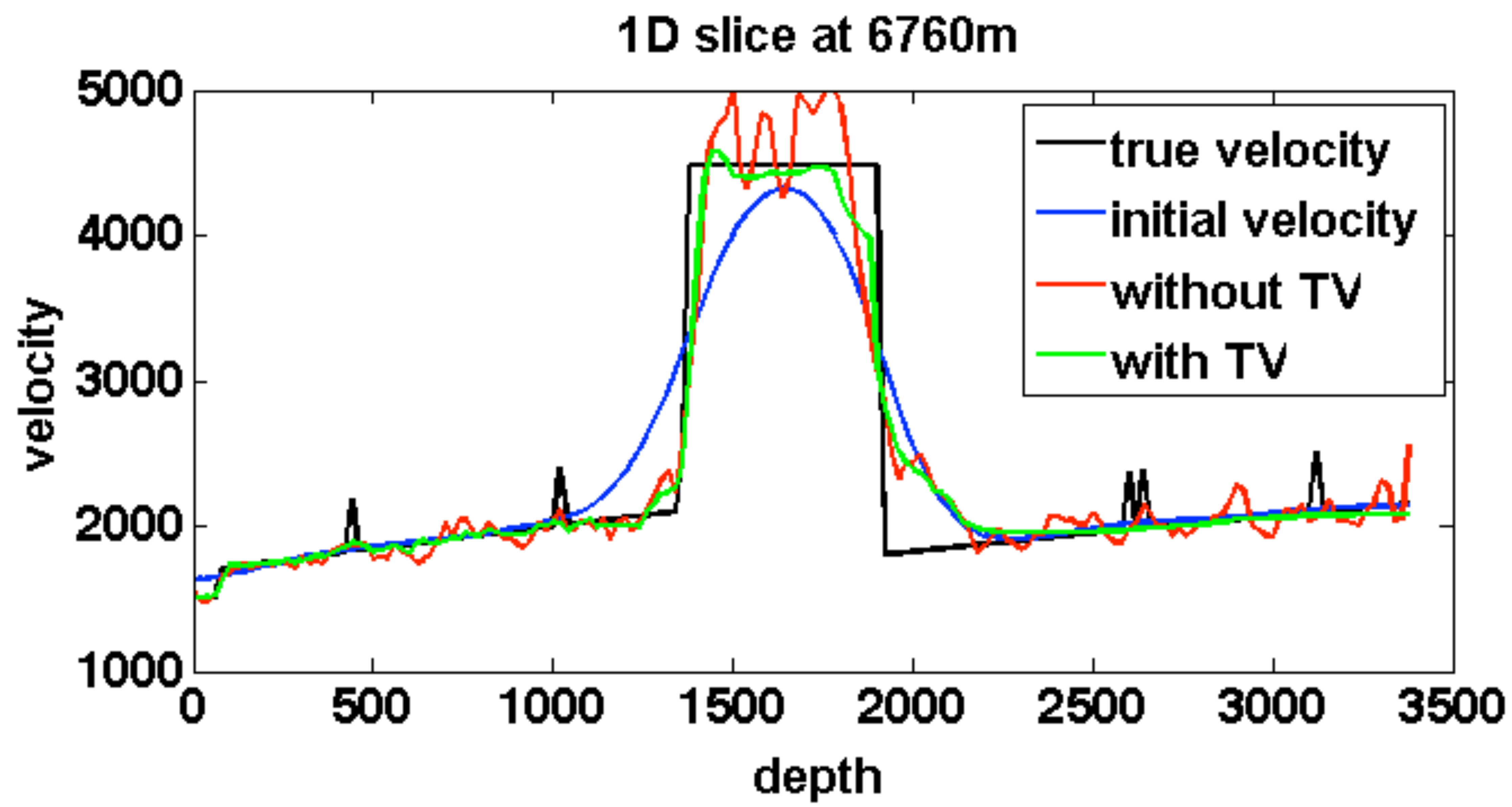


w/o TV constraint

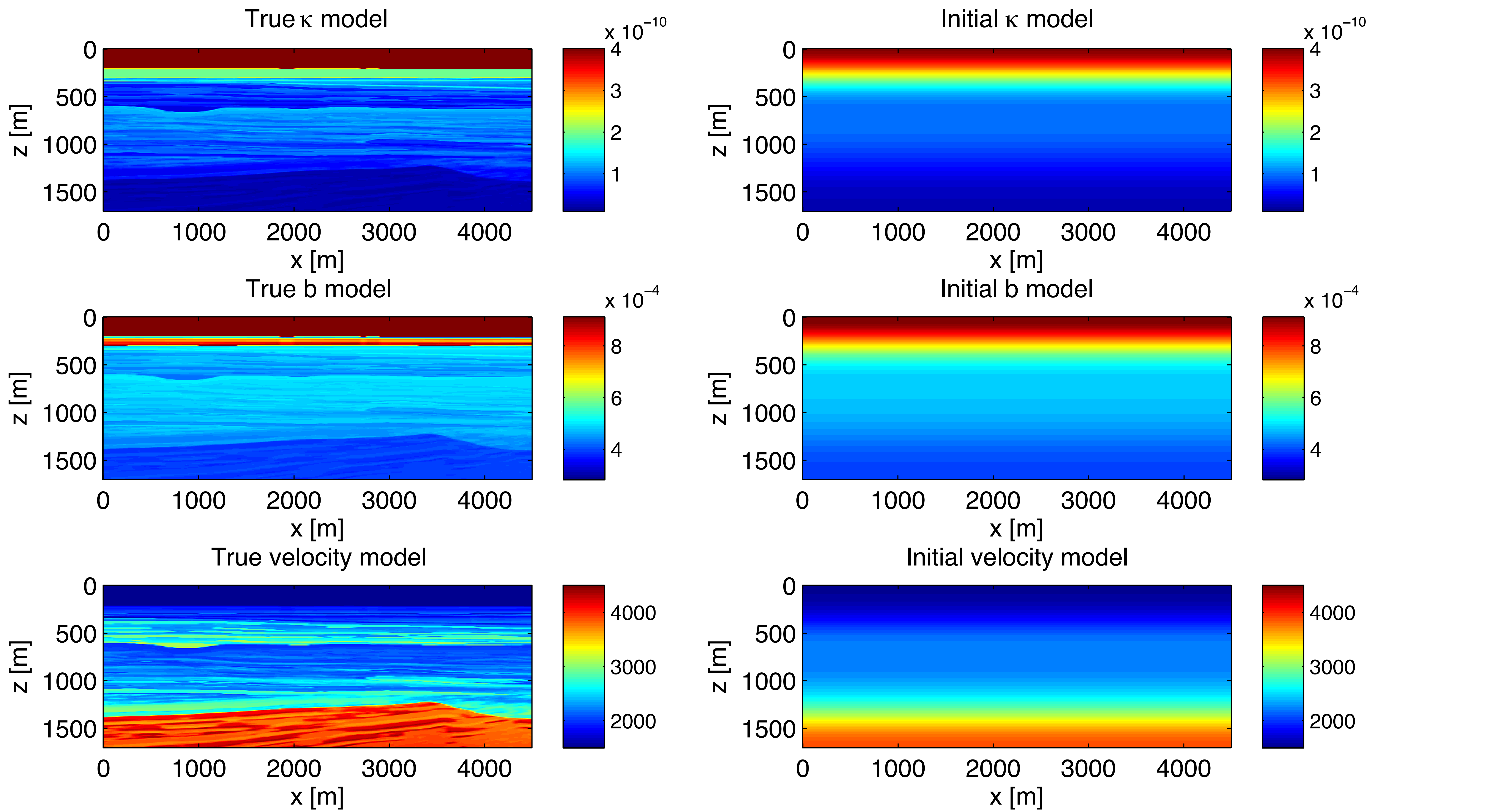


w/ TV constraint

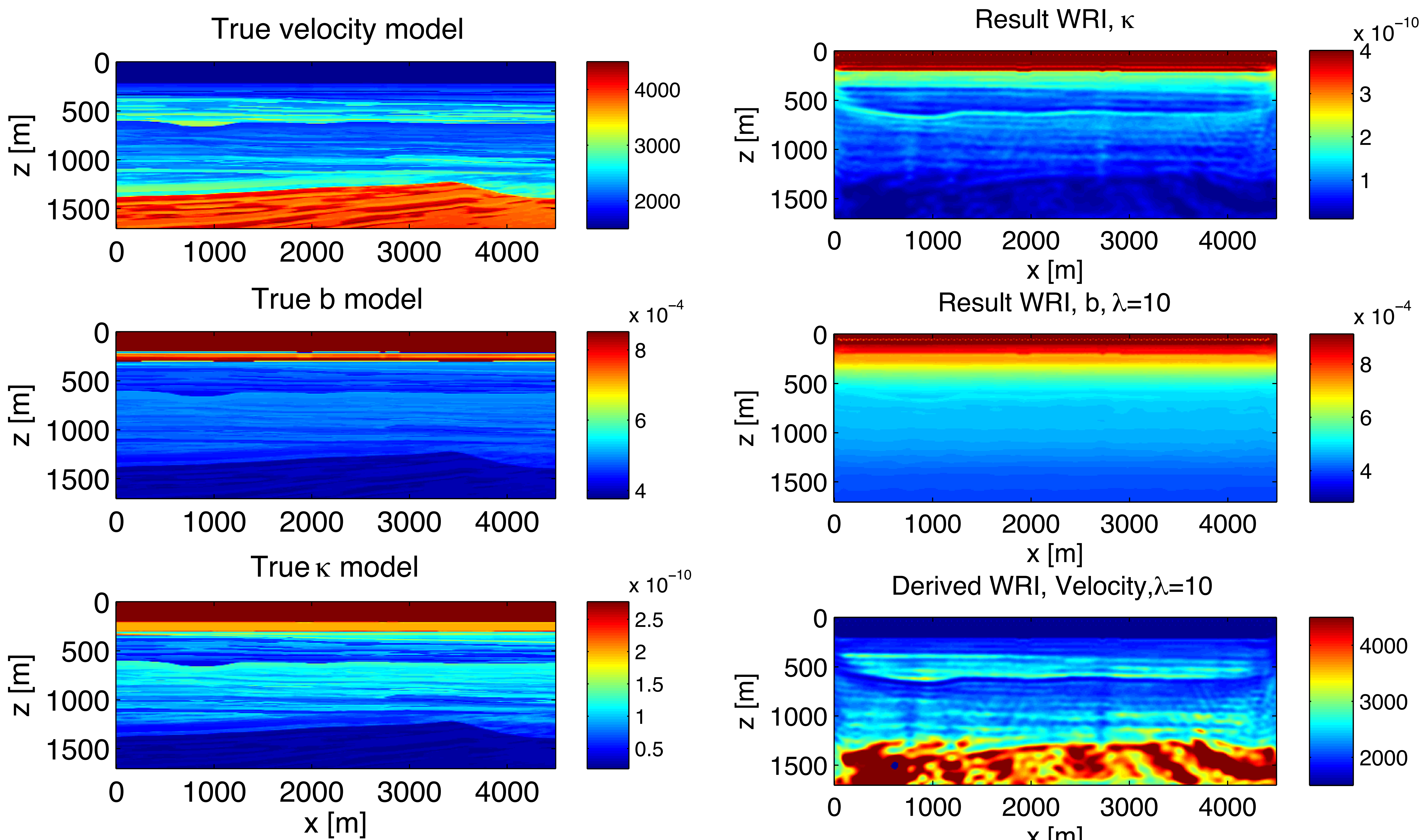

1D slice at 4000m
5000
4000
3000
2000
1000
velocity
0
500
1000
1500
2000
2500
3000
3500
depth
true velocity
initial velocity
without TV
with TV


1D slice at 6760m
5000
4000
3000
2000
1000
velocity
0
500
1000
1500
2000
2500
3000
3500
depth
true velocity
initial velocity
without TV
with TV

Multi-parameter WRI
true & initial models



Multi-parameter WRI
l-BFGS



Multi-parameter WRI

PDE-contrained formulation:

$$\phi_\lambda(\mathbf{b}, \boldsymbol{\kappa}, \mathbf{u}) = \frac{1}{2}\|P\mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|A(\mathbf{b}, \boldsymbol{\kappa})\mathbf{u} - \mathbf{q}\|_2^2$$

Penalty formulation:

$$\min_{\mathbf{b}, \boldsymbol{\kappa}, \mathbf{u}} \frac{1}{2}\|P\mathbf{u} - \mathbf{d}\|_2^2 \quad \text{s.t.} \quad A(\mathbf{b}, \boldsymbol{\kappa})\mathbf{u} = \mathbf{q}$$

Newton system:

$$\left(\begin{array}{c|cc} \nabla^2_{\mathbf{u},\mathbf{u}}\phi_\lambda & \nabla^2_{\mathbf{u},\boldsymbol{\kappa}}\phi_\lambda & \nabla^2_{\mathbf{u},\mathbf{b}}\phi_\lambda \\ \hline \nabla^2_{\boldsymbol{\kappa},\mathbf{u}}\phi_\lambda & \nabla^2_{\boldsymbol{\kappa},\boldsymbol{\kappa}}\phi_\lambda & \nabla^2_{\boldsymbol{\kappa},\mathbf{b}}\phi_\lambda \\ \nabla^2_{\mathbf{b},\mathbf{u}}\phi_\lambda & \nabla^2_{\mathbf{b},\boldsymbol{\kappa}}\phi_\lambda & \nabla^2_{\mathbf{b},\mathbf{b}}\phi_\lambda \end{array}\right) \begin{pmatrix} \delta_{\mathbf{u}} \\ \delta_{\boldsymbol{\kappa}} \\ \delta_{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{u}}\phi_\lambda \\ \nabla_{\boldsymbol{\kappa}}\phi_\lambda \\ \nabla_{\mathbf{b}}\phi_\lambda \end{pmatrix}$$

Multi-parameter WRI
sparse GN Hessian

Solve for wavefield ($\nabla_{\mathbf{u}}\phi_\lambda(\bar{\mathbf{u}},\boldsymbol{\kappa},\mathbf{b}) = 0$)

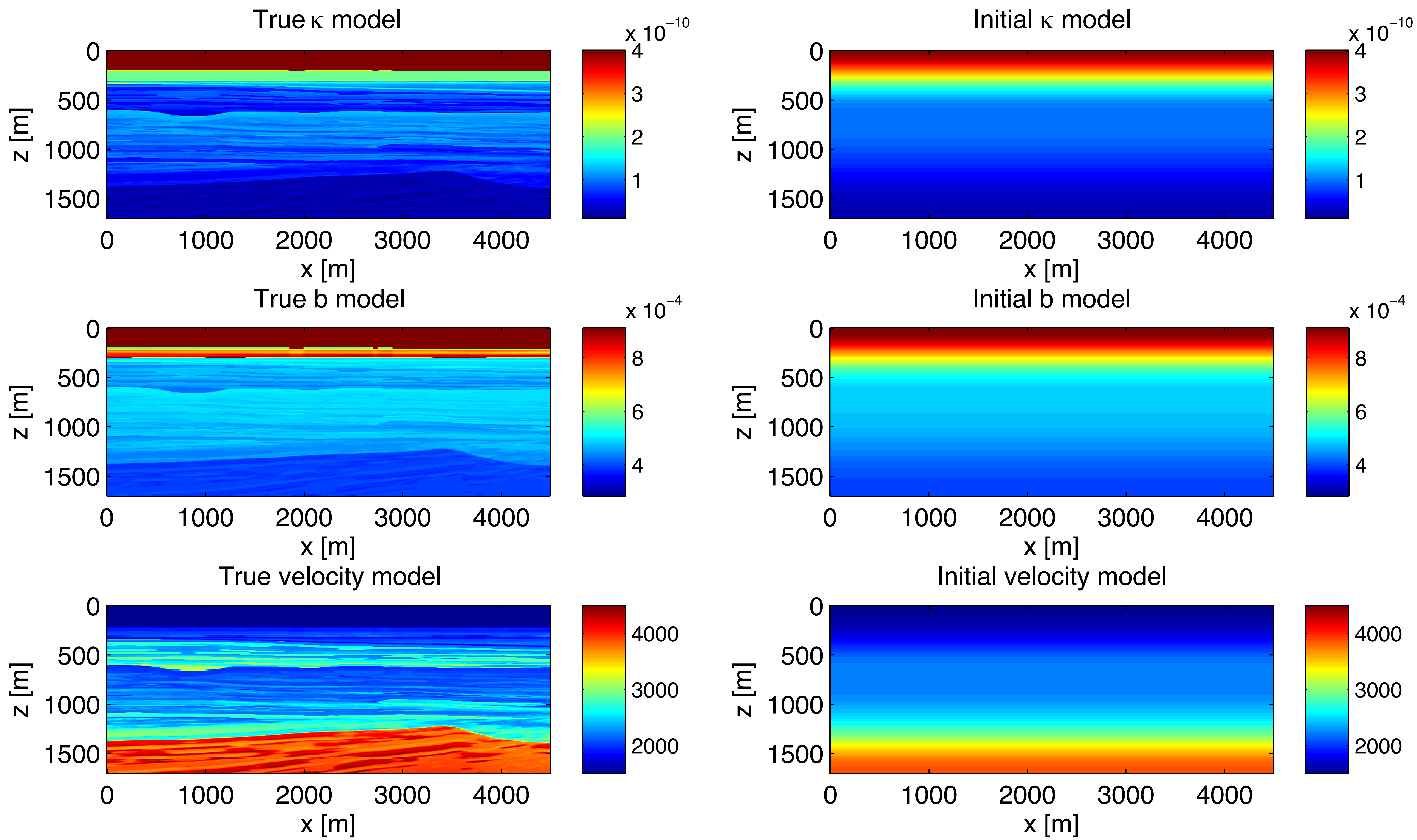
Ignore *dense* & *off*-diagonal blocks:

$$\tilde{H}_{GN} = \begin{pmatrix} \nabla^2_{\boldsymbol{\kappa},\boldsymbol{\kappa}}\phi_\lambda & 0 \\ 0 & \nabla^2_{\mathbf{b},\mathbf{b}}\phi_\lambda \end{pmatrix} = \begin{pmatrix} G^*_{\boldsymbol{\kappa}} G_{\boldsymbol{\kappa}} & 0 \\ 0 & G^*_{\mathbf{b}} G_{\mathbf{b}} \end{pmatrix}$$

with

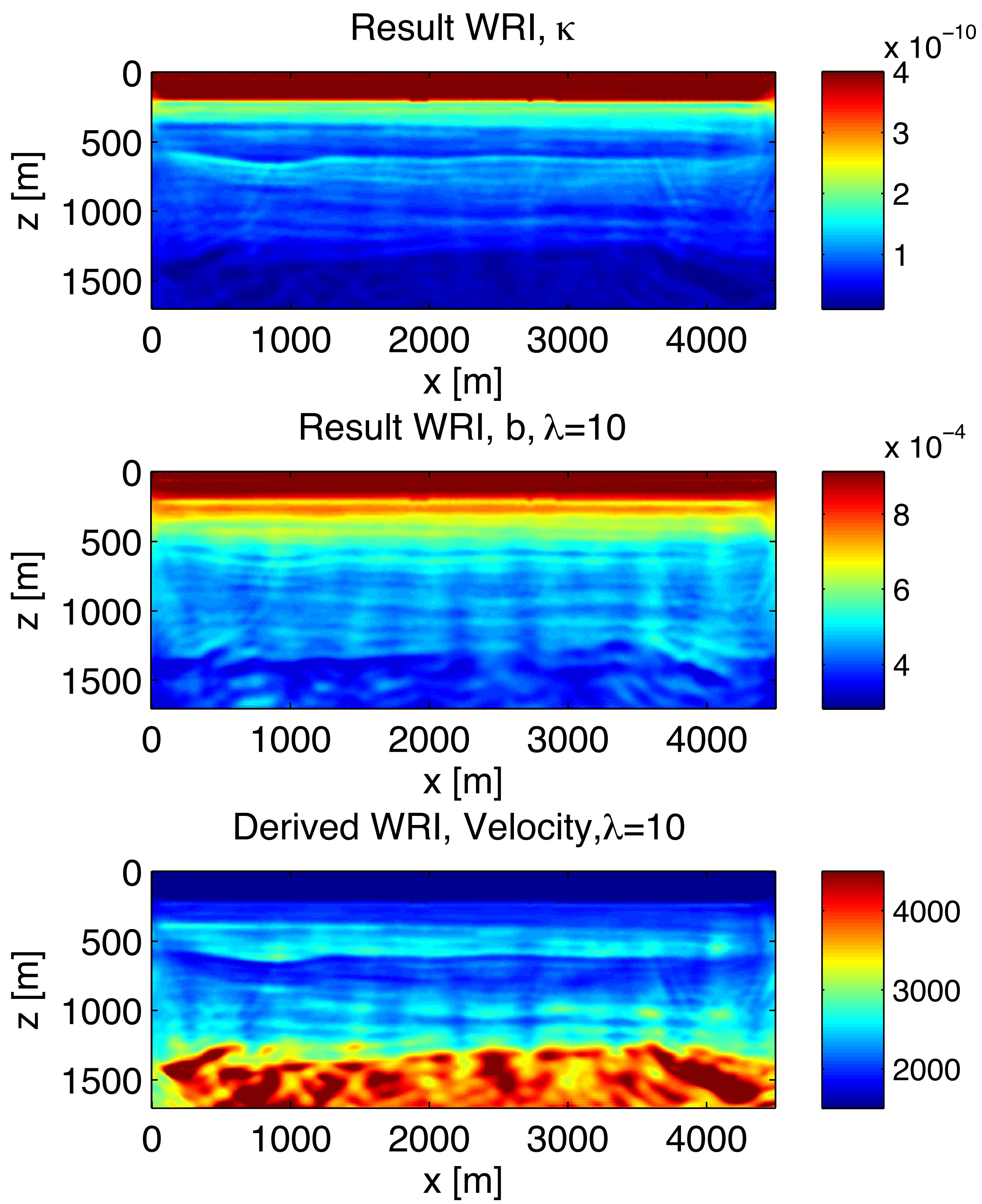
$$G_{\boldsymbol{\kappa}} = \partial H(\mathbf{b},\boldsymbol{\kappa})\bar{\mathbf{u}}/\partial\boldsymbol{\kappa} \qquad G_{\mathbf{b}} = \partial H(\mathbf{b},\boldsymbol{\kappa})\bar{\mathbf{u}}/\partial\mathbf{b}$$

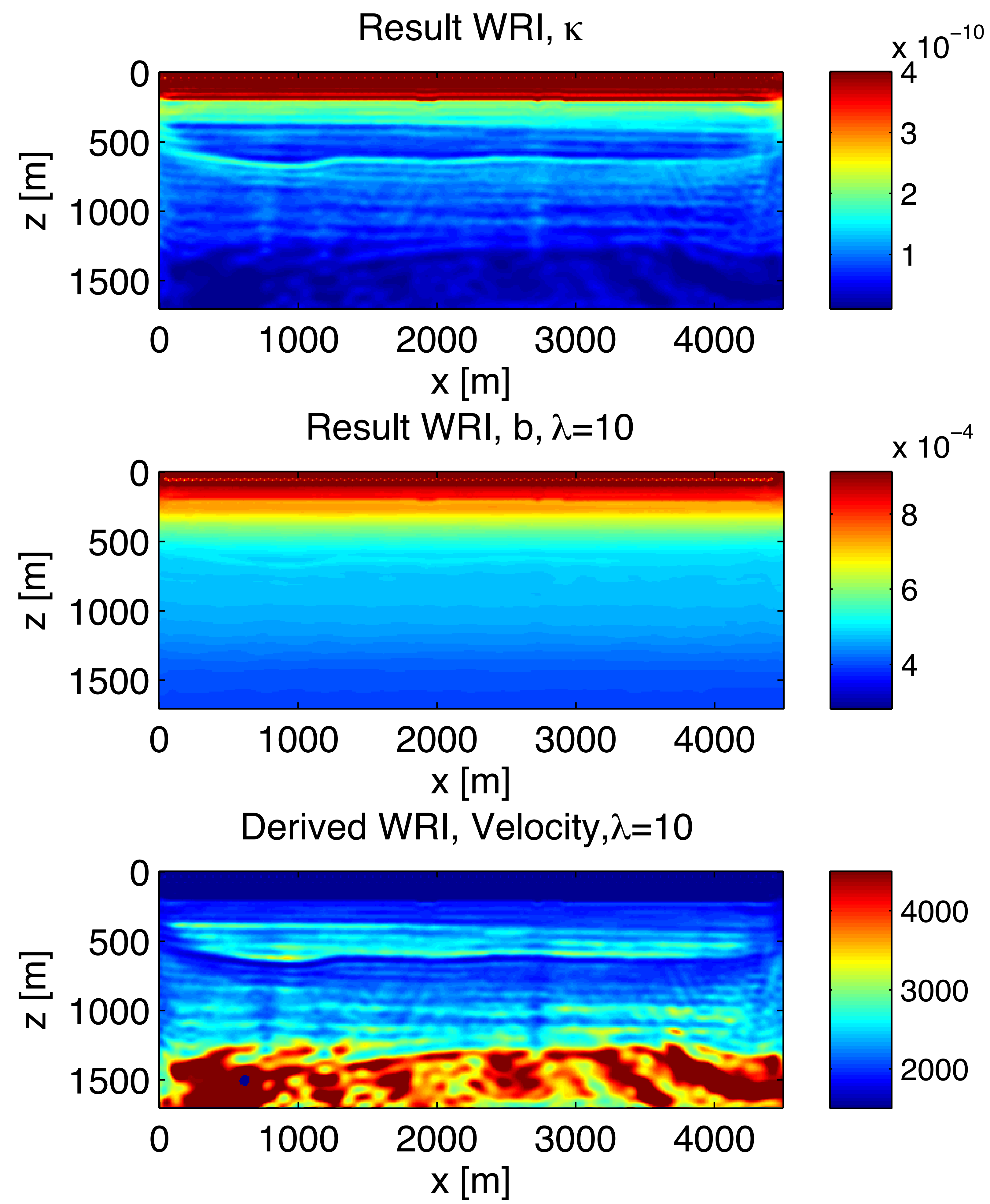
Multi-parameter WRI
true & initial models



Multi-parameter WRI vs constrained formulation

New algorithm





basic L-BFGS

Conclusions

New method for *wave*-equation based inversion:

- *extended* search space as in *all-at-once* similar but with memory & CPU requirements as in *reduced* approach
- *no* adjoints & *sparse* GN-Hessian approximation, `*less* non-linear’
- *less* susceptible to getting *trapped* by *local* minima

Still *early* days in the development:

- *encouraging* 2-D results w/ ***less*** *sensitivity* to *initial* model
- TV regularization & bound constraints via ***cheap*** *convex sub*problems
- ***cheap*** *scalings* of the *inverse* of the GN Hessian

Acknowledgements

Thank you for your attention !

https://www.slim.eos.ubc.ca/

SINBAD

NSERC CRSNG

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, WesternGeco, and Woodside.