Pros and Cons of Full- and Reduced-space Methods for Wavefield Reconstruction Inversion

Felix J. Herrmann & Bas Peters

SLIM
University of British Columbia

SIAM Conference on Mathematical and Computational Issues in the Geosciences
June 29, 2015

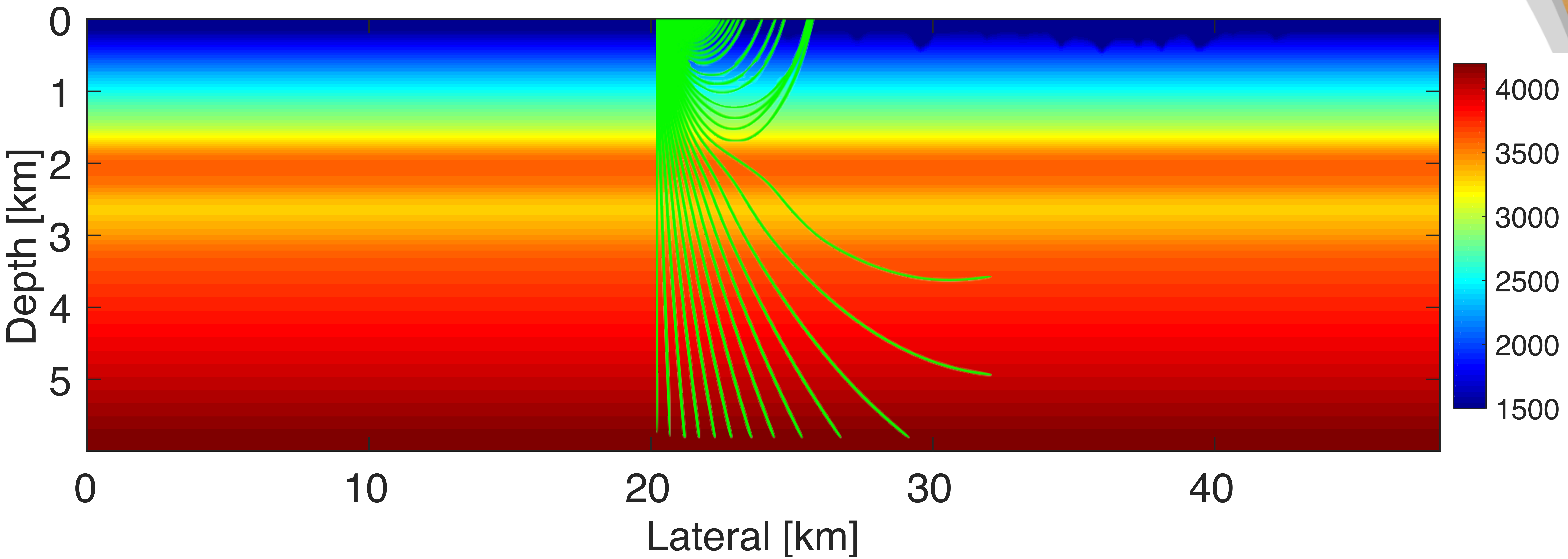
Pros and Cons of Full- and Reduced-space Methods for Wavefield Reconstruction Inversion

Felix J. Herrmann & Bas Peters

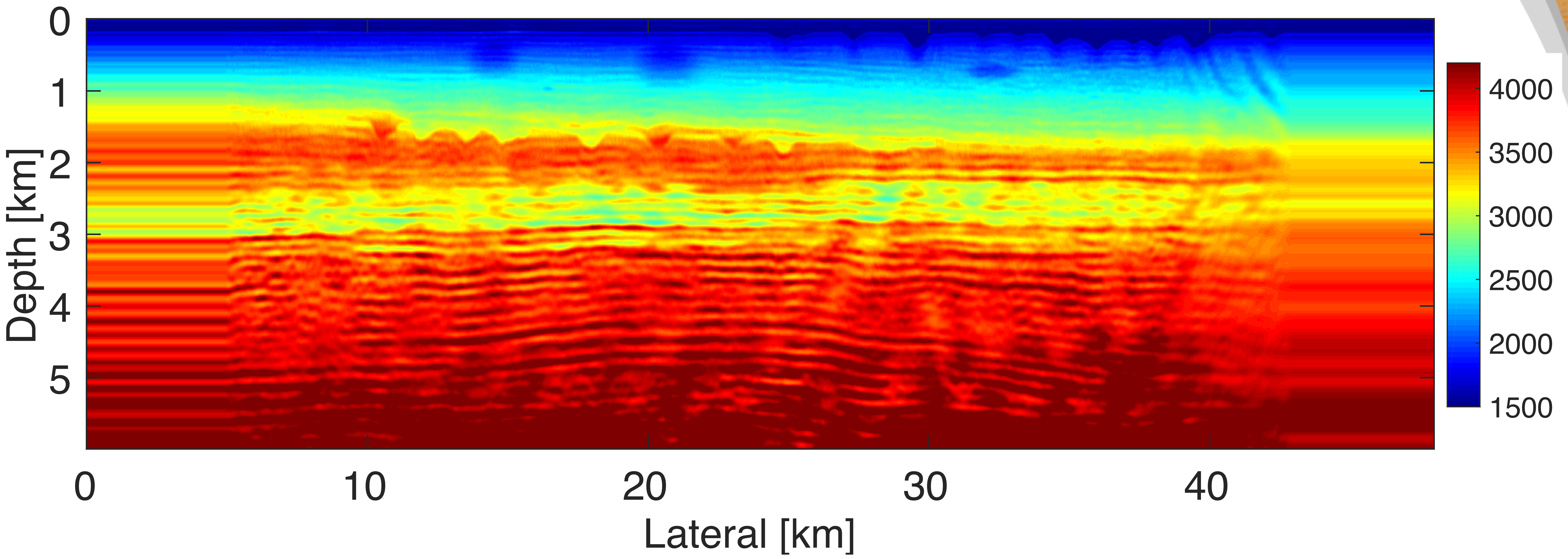
SLIM
University of British Columbia

SIAM Conference on Mathematical and Computational Issues in the Geosciences
June 29, 2015

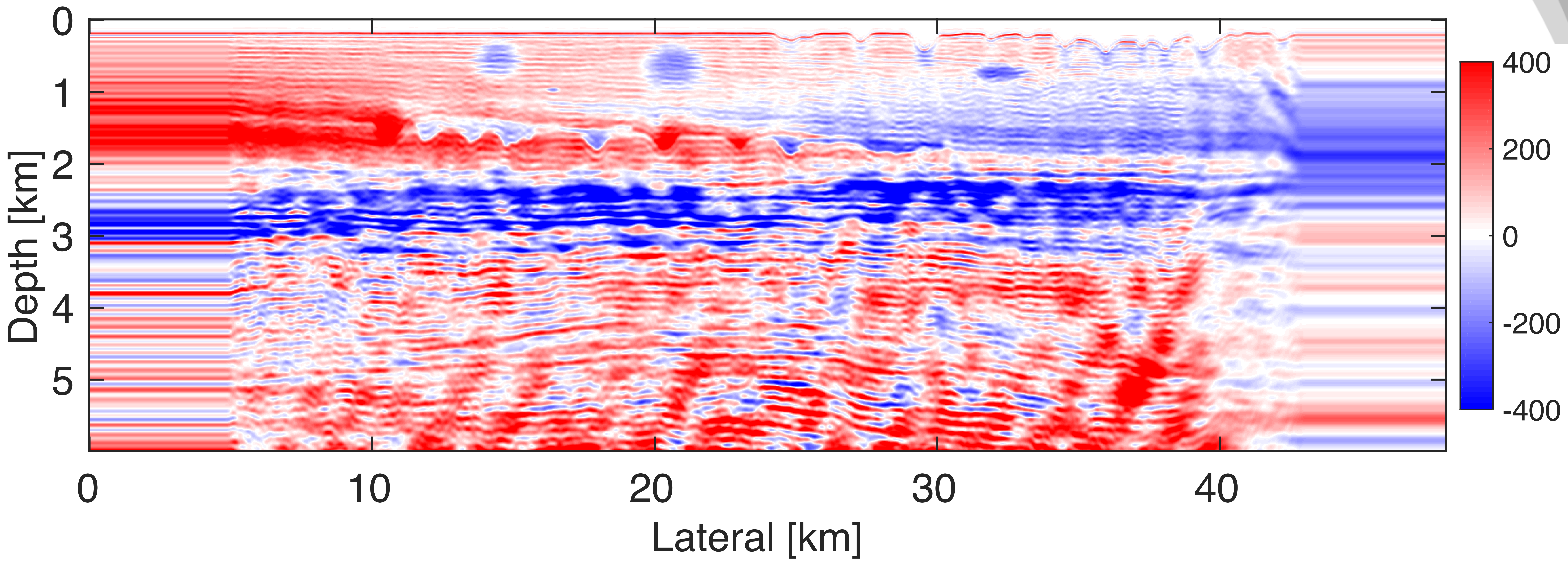
Initial model



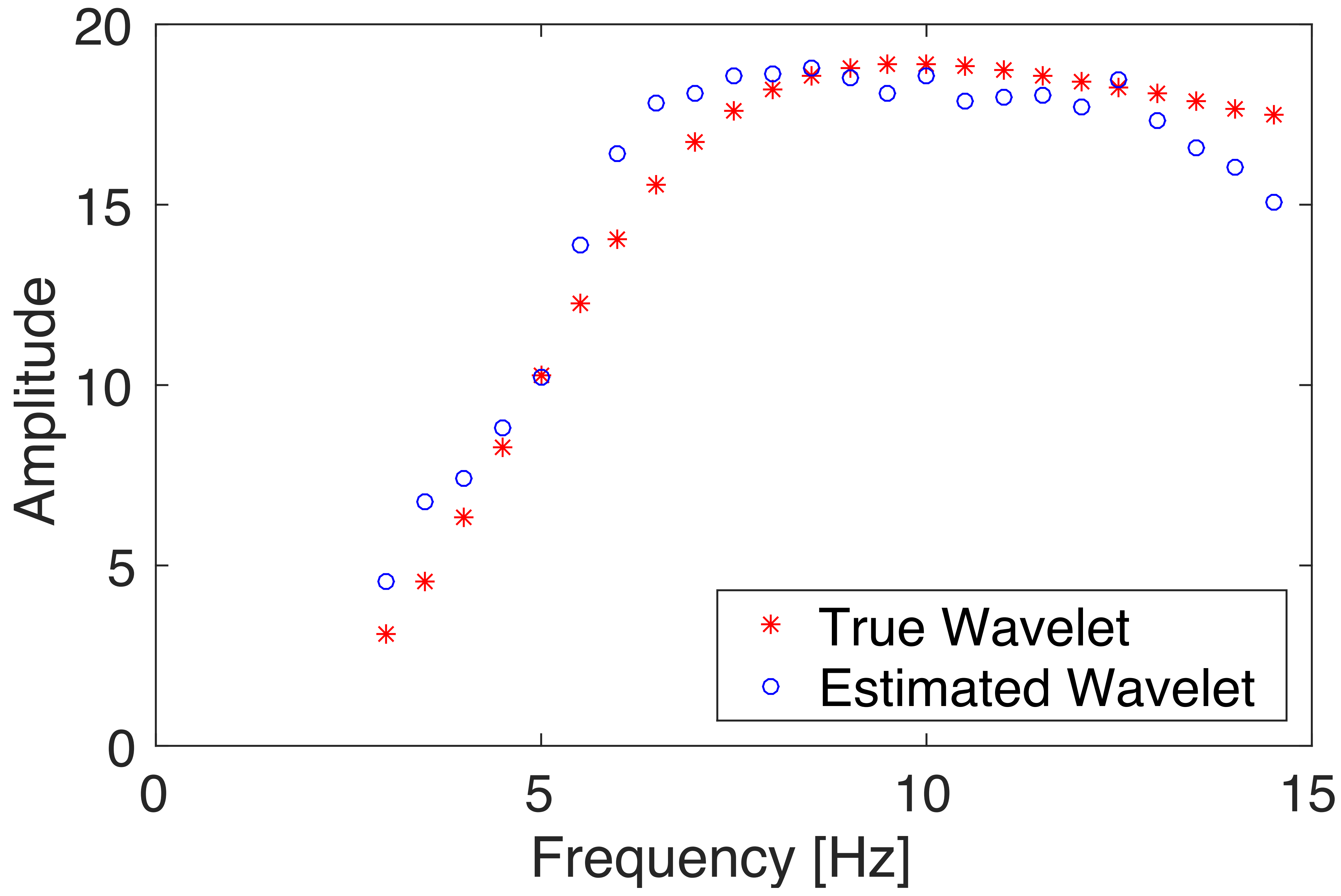
Inversion result



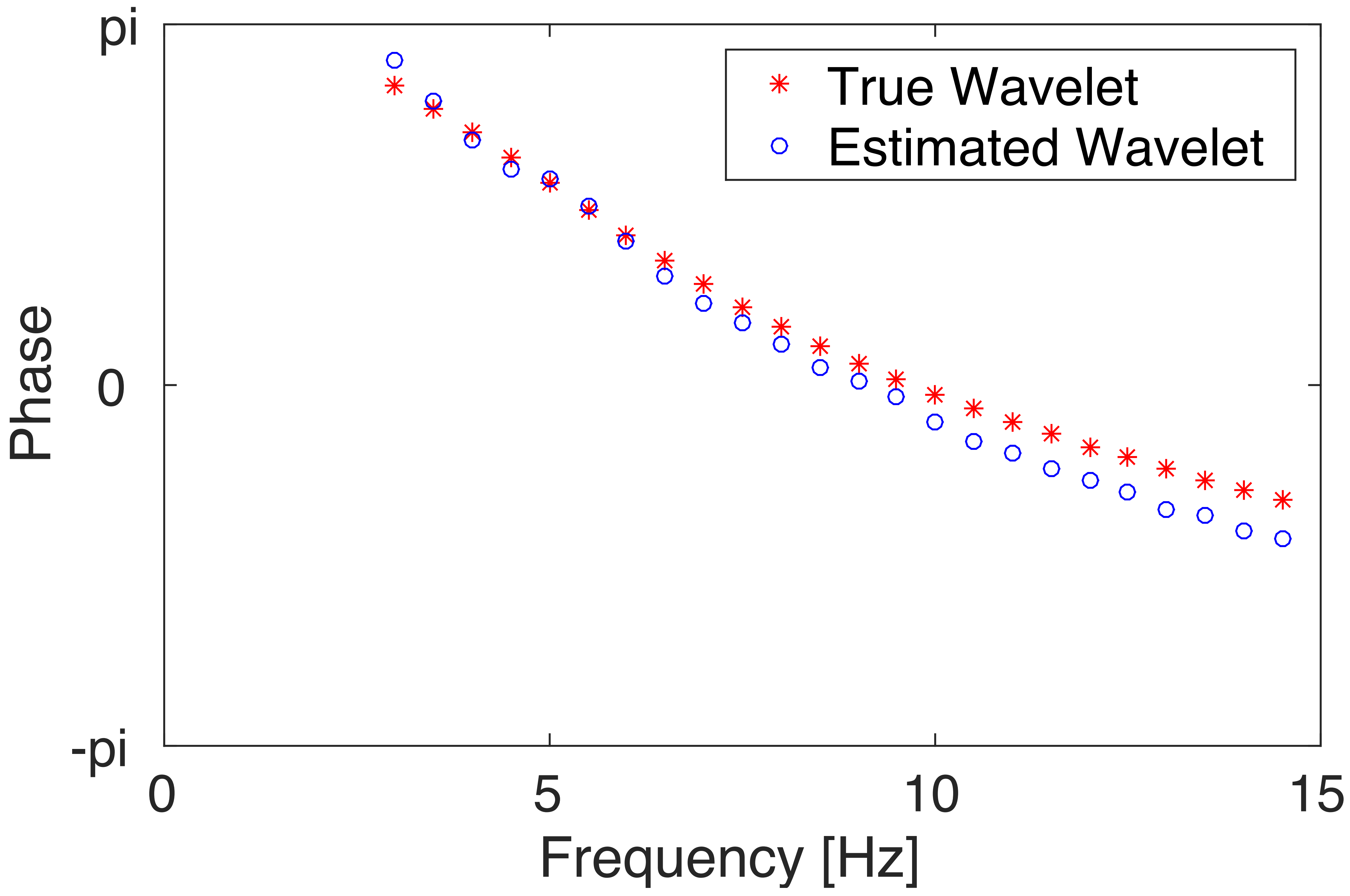
Model update



Source wavelet comparison



Amplitude



Phase

[E. Haber & U.M. Ascher, 2001 ; G. Biros & O. Ghattas , 2005 ; Grote et. al., 2011]

PDE-constrained optimization

Use the 'discretize-then-optimize' framework:

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

$H(\mathbf{m}) \in \mathbb{C}^{N\times N}$ discrete PDE

$\mathbf{m} \in \mathbb{R}^N$ medium parameters

$P \in \mathbb{R}^{m\times N}$ selects field at receivers

$\mathbf{u} \in \mathbb{C}^N$ field

$\mathbf{d} \in \mathbb{C}^m$ observed data

$\mathbf{q} \in \mathbb{C}^N$ source

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

$$\mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 + \mathbf{v}^*\left(H(\mathbf{m})\mathbf{u}-\mathbf{q}\right)$$

All-at-once full-space: [E. Haber & U.M. Ascher, 2001 ; G. Biros & O. Ghattas , 2005 ; Grote et. al., 2011]

- update fields, medium parameters & multipliers simultaneously
- function value, gradient, Hessian evaluation is ~free & exact
- sparse Hessian
- requires storage of all fields & multipliers + working memory (gradients, Hessian & update direction)
- updates are computationally demanding

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

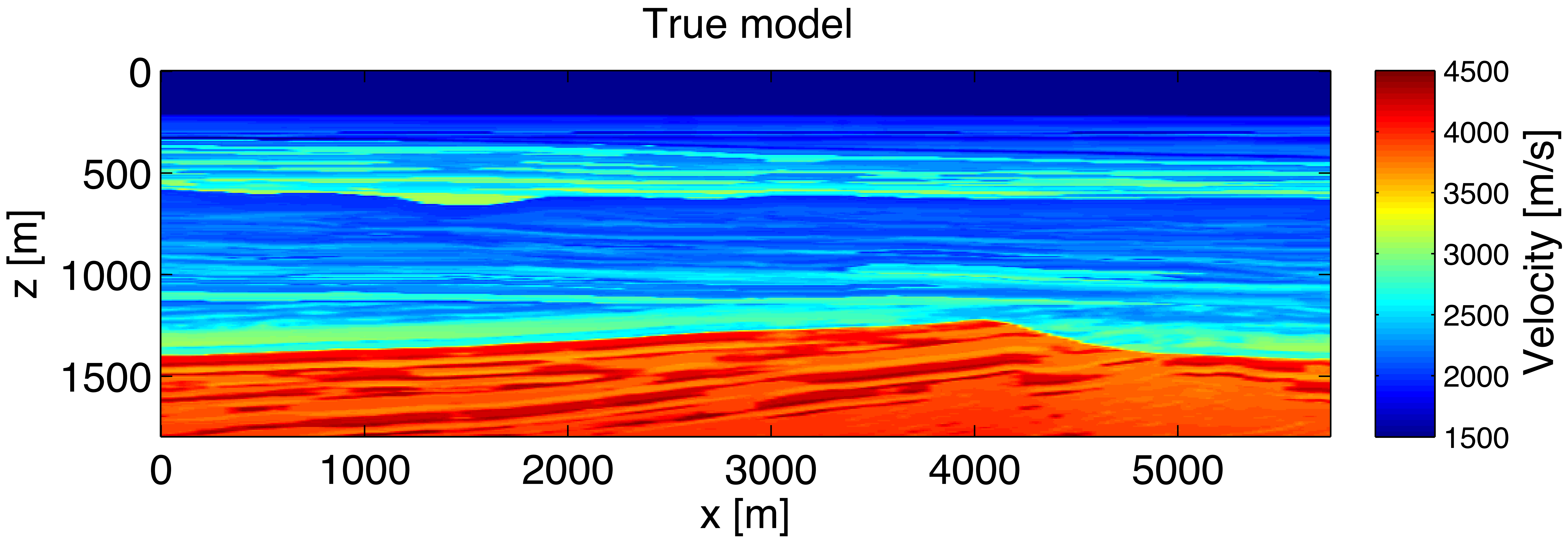
$$\mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 + \mathbf{v}^*\left(H(\mathbf{m})\mathbf{u}-\mathbf{q}\right)$$

eliminate field variables

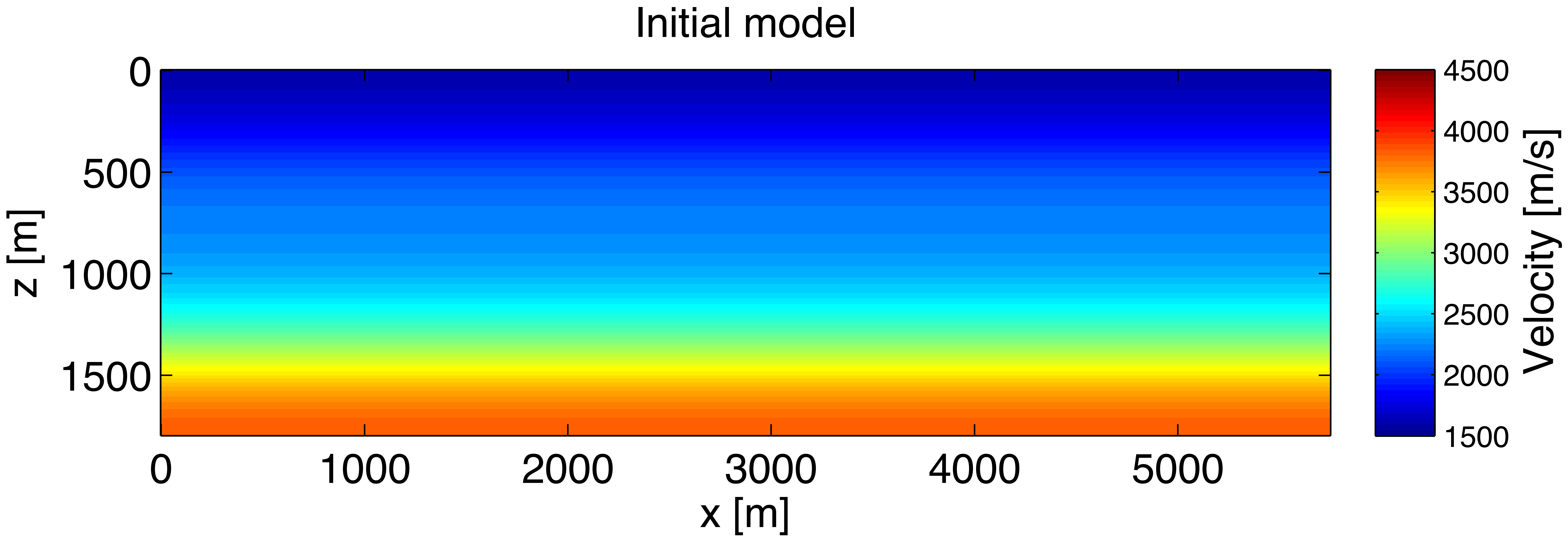
$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2$$

Reduced space: [Tarantola, ’84; Haber et al., ‘00; Epanomeritakis et al., ‘08]

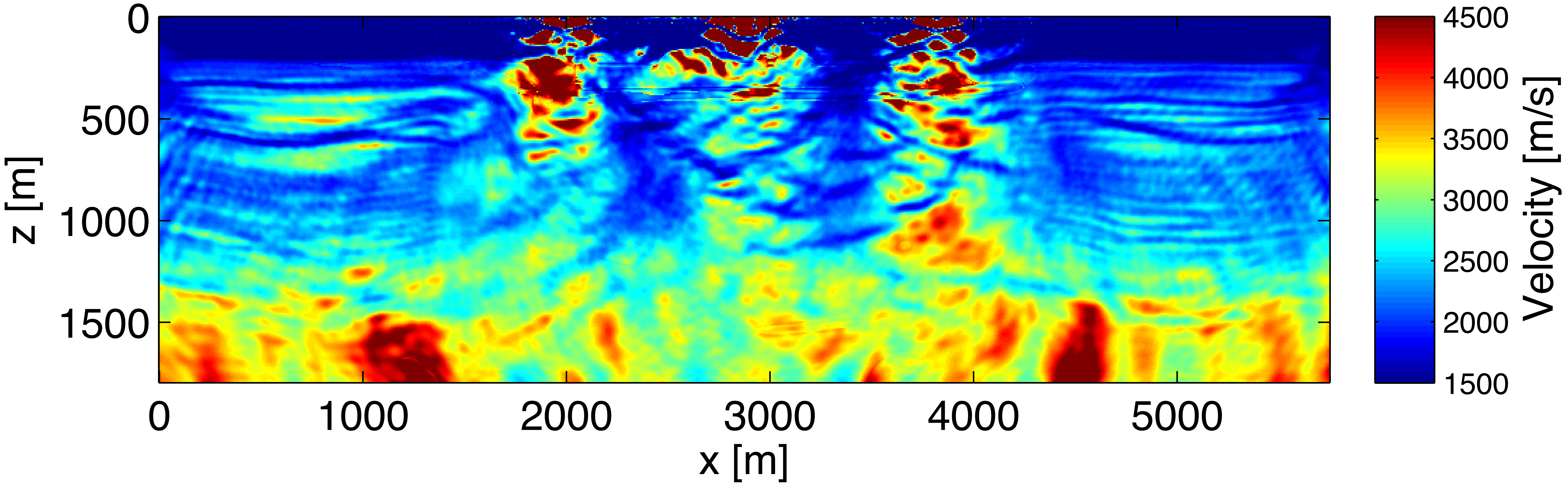
- storage as low as two fields at a time
- highly nonlinear in **m** & non-convex
  - expensive “exact” forward & adjoint solves for each iteration
  - inexact when sub-problems are solved iteratively
- dense reduced-Hessian
- requires extra safeguards/accuracy control [T. van Leeuwen & F.J. H ‘14]
- **reliance on accurate starting models to avoid cycle skipping**



Example from **[Peters et al. 2014]**



$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2$$



Few algorithms have quadratic-penalty form:
[R.E. Kleinman & P.M. van den Berg, '92 ; T. van Leeuwen & F.J.H, '13]

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

$$\downarrow$$

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\mathbf{u}-\mathbf{q}\|_2^2$$

$$\downarrow$$

eliminate field variables: solve $\nabla_{\mathbf{u}}\phi(\mathbf{m},\bar{\mathbf{u}},\lambda)=0$

[T. van Leeuwen & F.J.H. '13,'15]

$$\downarrow$$

$$\min_{\mathbf{m}} \frac{1}{2}\|P\bar{\mathbf{u}}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}}-\mathbf{q}\|_2^2$$

reduced quadratic-penalty

Penalty method:

- no need to store all the fields (u)
- no adjoint solves
- sparse approximation of Gauss-Newton Hessian for small λ
- less non-linear in **m**
- **need to solve data-augmented wave equation**

[T. van Leeuwen & F.J. Herrmann, 2013]

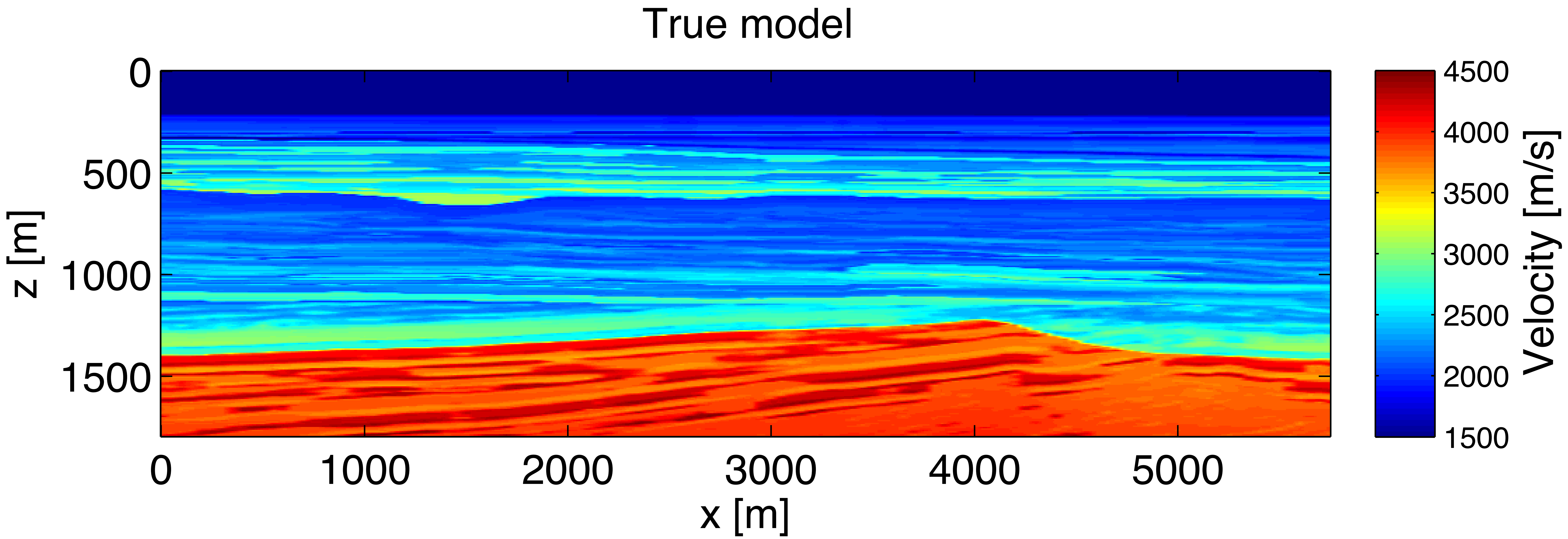
Reduced-space quadratic-penalty method
– Wavefield Reconstruction Inversion (WRI)

Minimize: $$\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$$

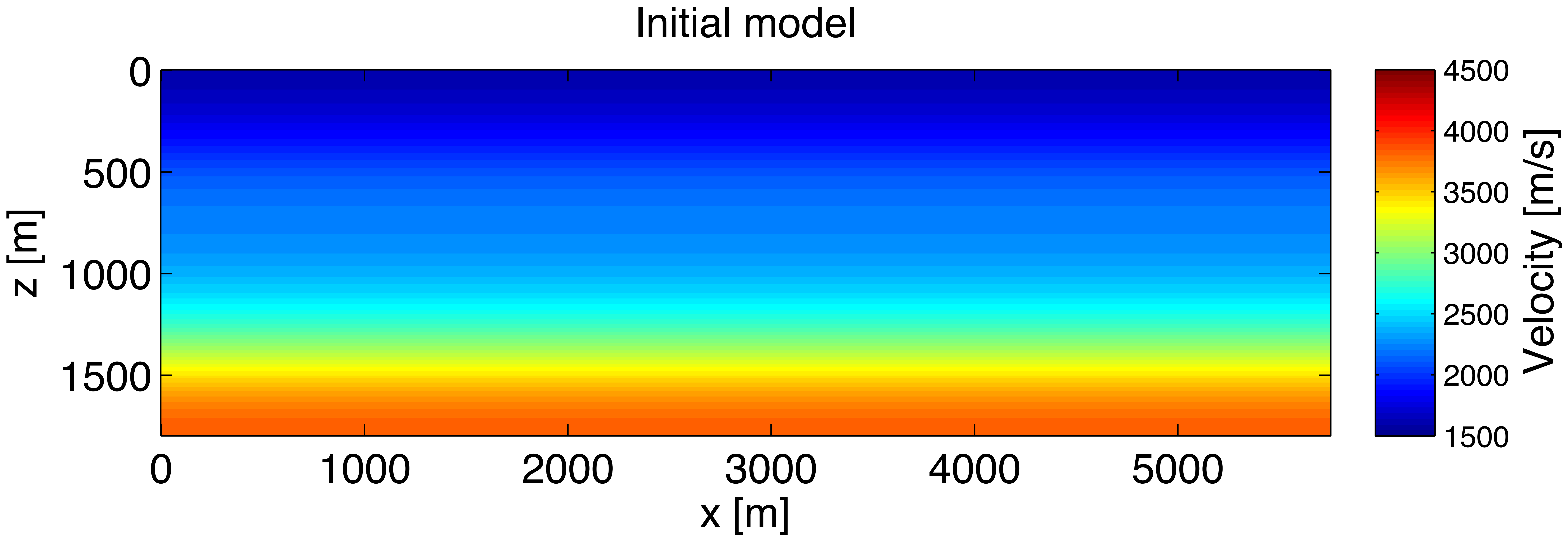
at every iteration

- compute $$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$
- evaluate $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$ & $\nabla_{\mathbf{m}}\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$
- update $\mathbf{m}$

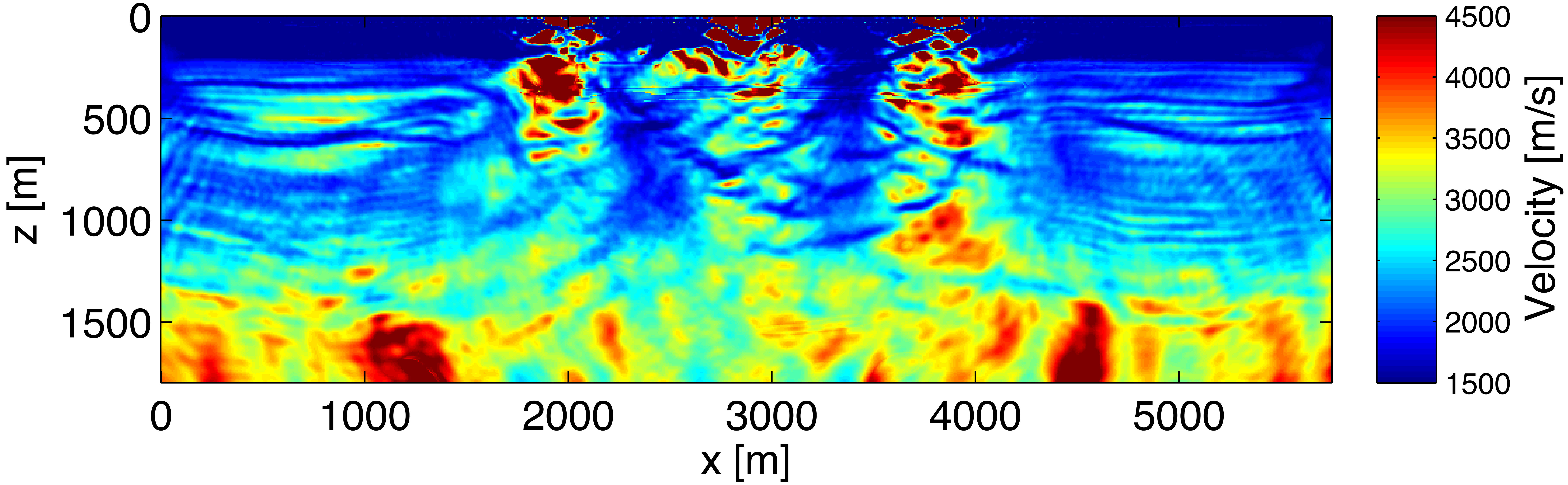
+ trust-region / line-search



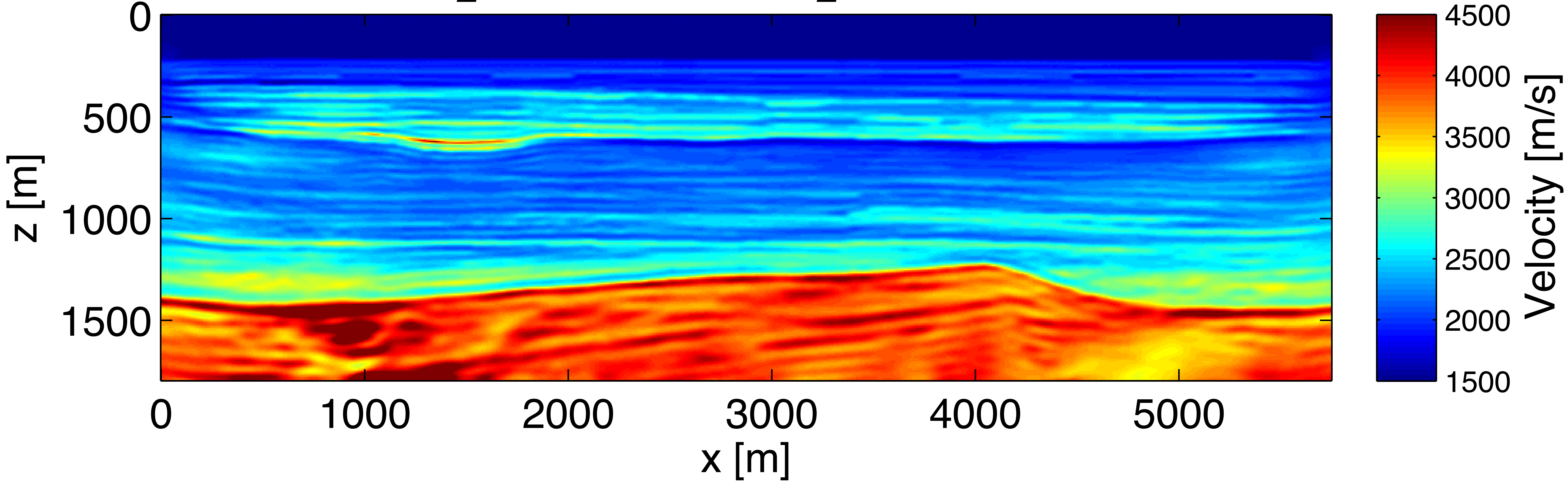
Example from [Peters et al. 2014]



$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}\|_2^2$$



$$\min_{\mathbf{m}} \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2 \qquad \text{(small } \lambda\text{)}$$



Reduced-space sub-problems

Solve $\mathbf{u} = H^{-1}\mathbf{q}$ or $\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$

In 3D we have iterative & inexact solves:

- variable elimination & projection are assumed exact
- inaccuracy may cause problems in reduced-space methods
  - how accurate do we need to be?
  - can we change the accuracy per iteration?
- costs are dominated by accuracy & # of source experiments
- heuristic solutions proposed in frugal approaches
  [T. van Leeuwen & F.J. Herrmann, 2014]

Inexact PDE solves
– full-space vs reduced-space

reduced-space:

- error in objective function value
- error in gradient
- error in Hessian

→ error in medium parameter update

full-space:

- objective function value always exact
- gradient always exact
- Hessian always exact

→ globally convergent inexact Newton methods
[S.C. Eisenstat & H.F. Walker, 1994]

SLIM

Quadratic-penalty full space methods

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\mathbf{u}-\mathbf{q}\|_2^2$$

Newton's method:

$$\begin{pmatrix} P^*P + \lambda^2 H^* H & \nabla^2_{\mathbf{u},\mathbf{m}}\phi \\ \nabla^2_{\mathbf{m},\mathbf{u}}\phi & \lambda^2 G^*_{\mathbf{m}} G_{\mathbf{m}} \end{pmatrix} \begin{pmatrix} \delta\mathbf{u} \\ \delta\mathbf{m} \end{pmatrix} = -\begin{pmatrix} P^*(P\mathbf{u}-\mathbf{d}) + \lambda^2 H^*(H\mathbf{u}-\mathbf{q}) \\ \lambda^2 G^*_{\mathbf{m}}(H\mathbf{u}-\mathbf{q}) \end{pmatrix}$$

updates for medium parameters

updates for all fields

How to solve?

Quadratic-penalty based full space methods

Initial attempt: block diagonal approximation

$$\begin{pmatrix} P^*P + \lambda^2 H^*H & 0 \\ 0 & \lambda^2 G_{\mathbf{m}}^* G_{\mathbf{m}} \end{pmatrix} \begin{pmatrix} \delta\mathbf{u} \\ \delta\mathbf{m} \end{pmatrix} = - \begin{pmatrix} P^*(P\mathbf{u} - \mathbf{d}) + \lambda^2 H^*(H\mathbf{u} - \mathbf{q}) \\ \lambda^2 G_{\mathbf{m}}^*(H\mathbf{u} - \mathbf{q}) \end{pmatrix}$$

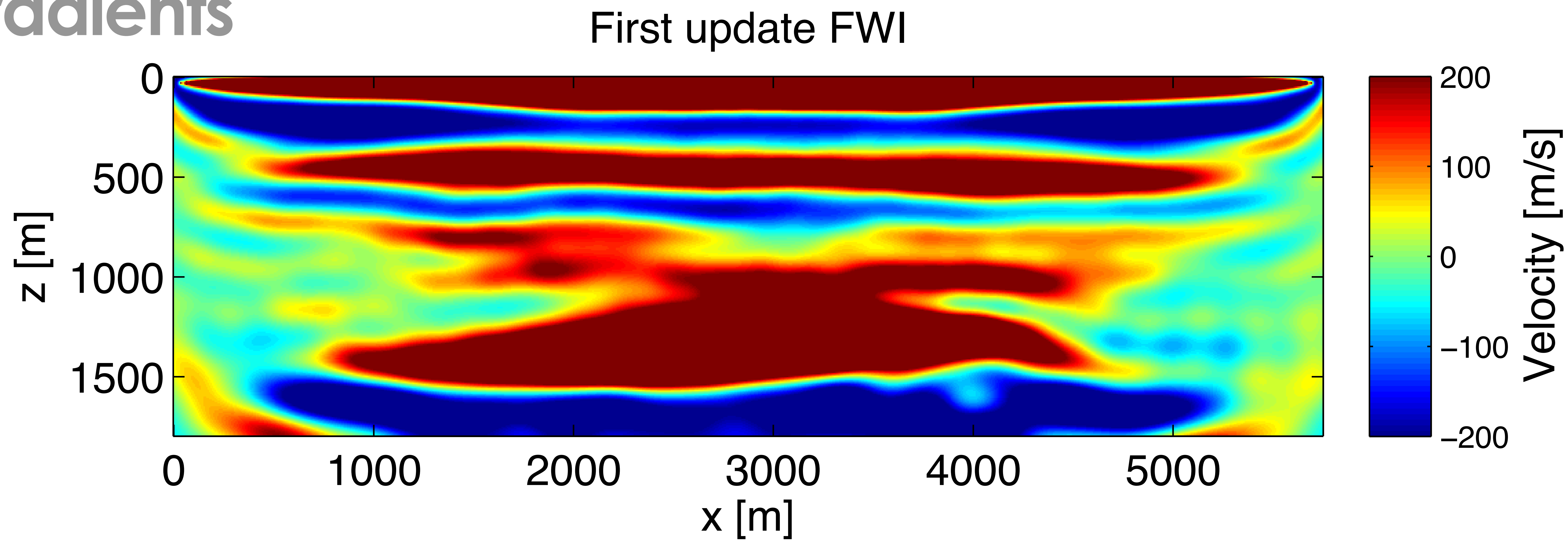
Give up some of Newton's method properties.

Update computation intrinsically parallel per field.

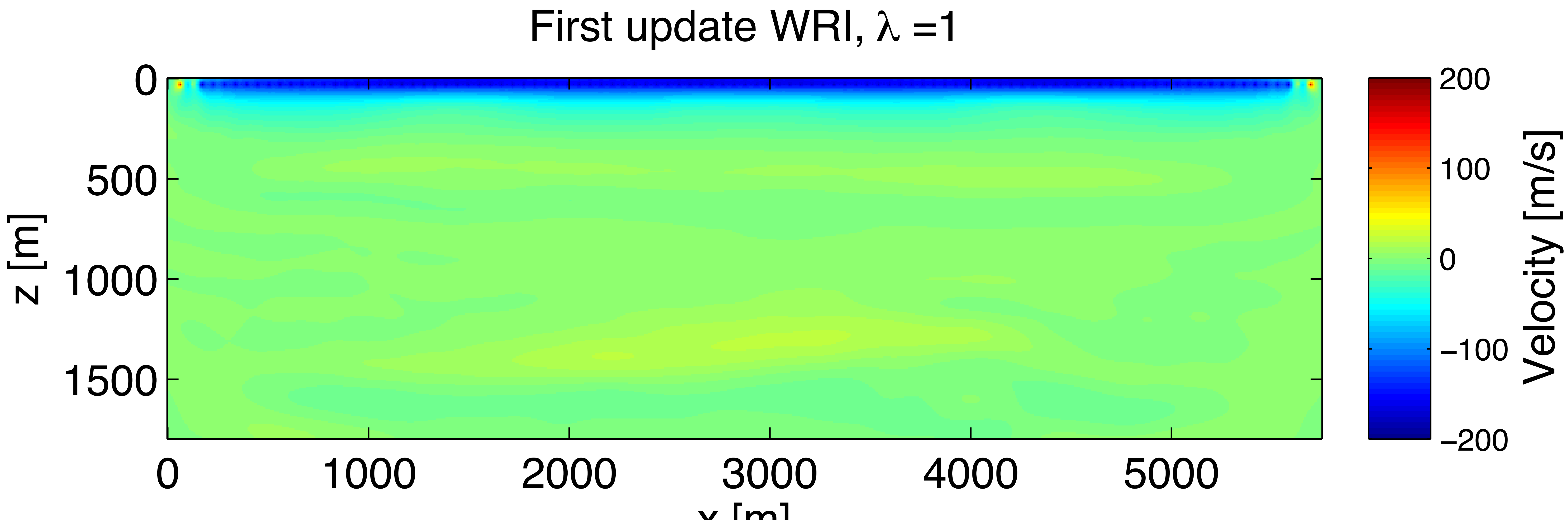
Same as WRI

- if system is solved exactly, and
- fields are initialized by data fit via exact solve data-augmented system

Gradients



First update FWI
z [m]
x [m]
0
500
1000
1500
0
1000
2000
3000
4000
5000
200
100
0
−100
−200
Velocity [m/s]

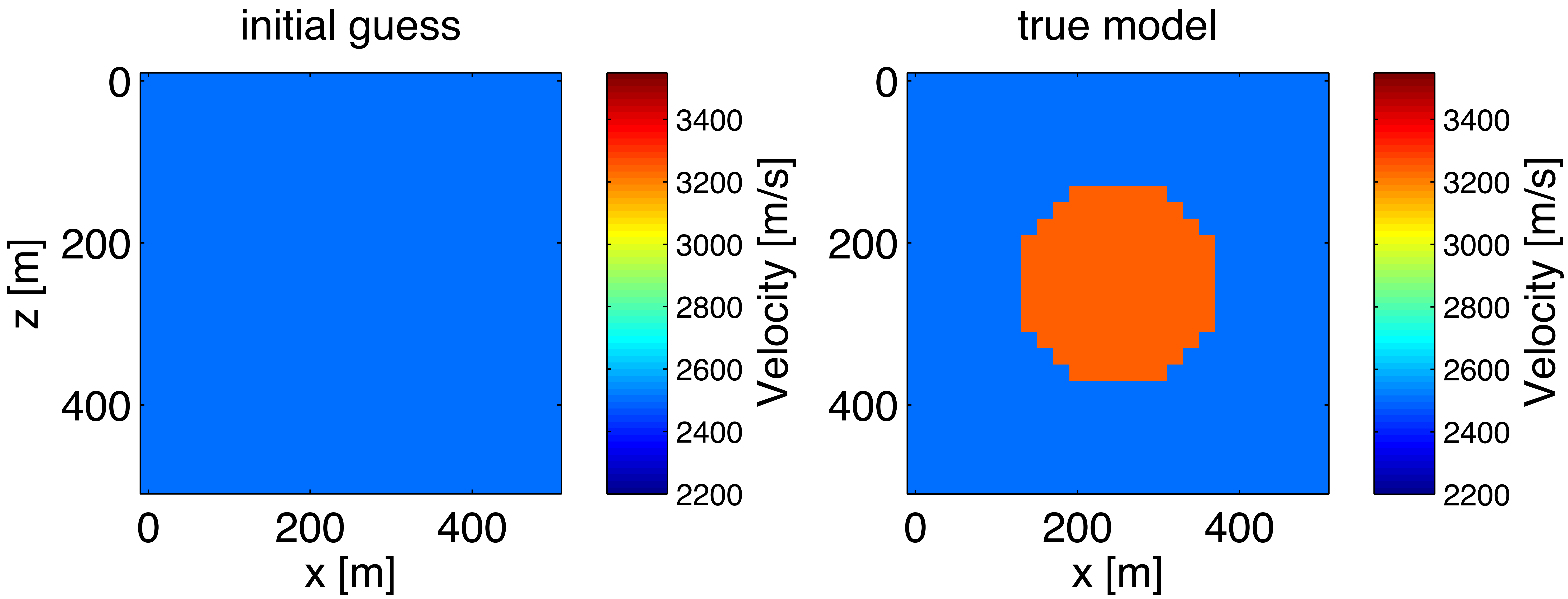


First update WRI, λ =1
z [m]
x [m]
0
500
1000
1500
0
1000
2000
3000
4000
5000
200
100
0
−100
−200
Velocity [m/s]

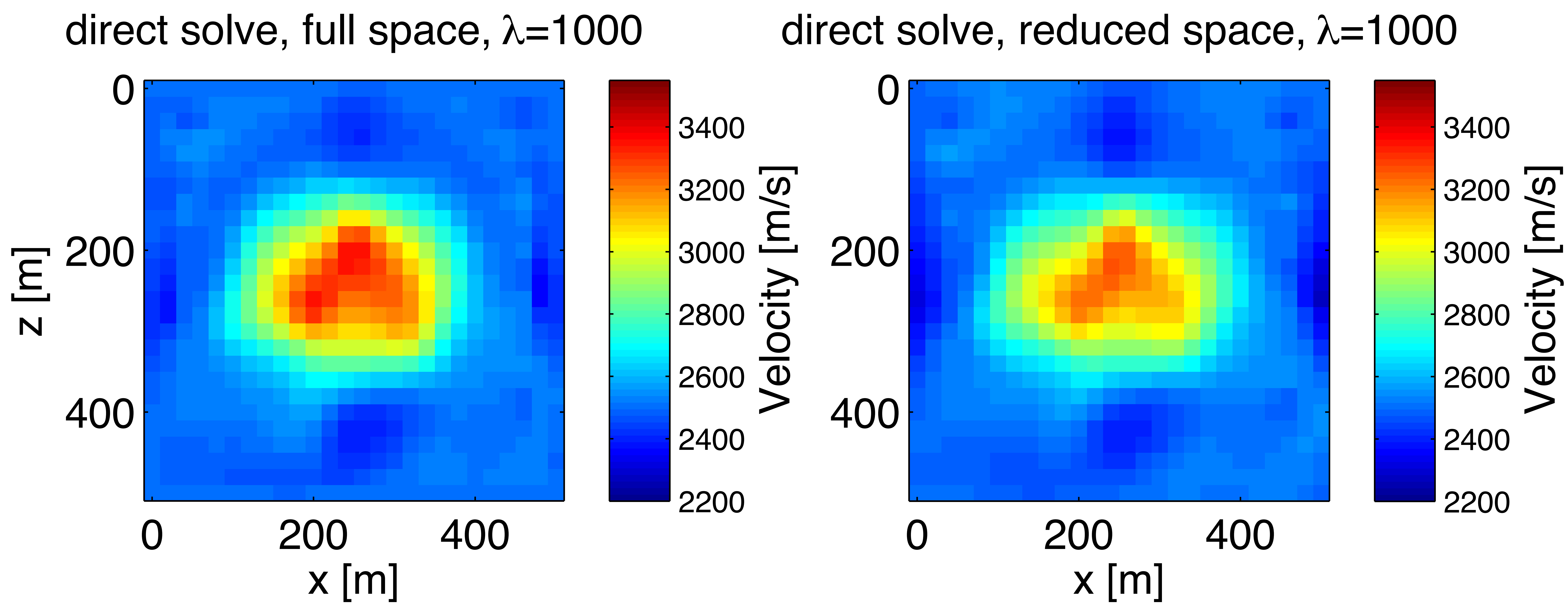
Toy problem

- cross-well setting
- 4 frequencies [6-10] Hz
- 5 simultaneous sources
- 5 receivers

Toy problem

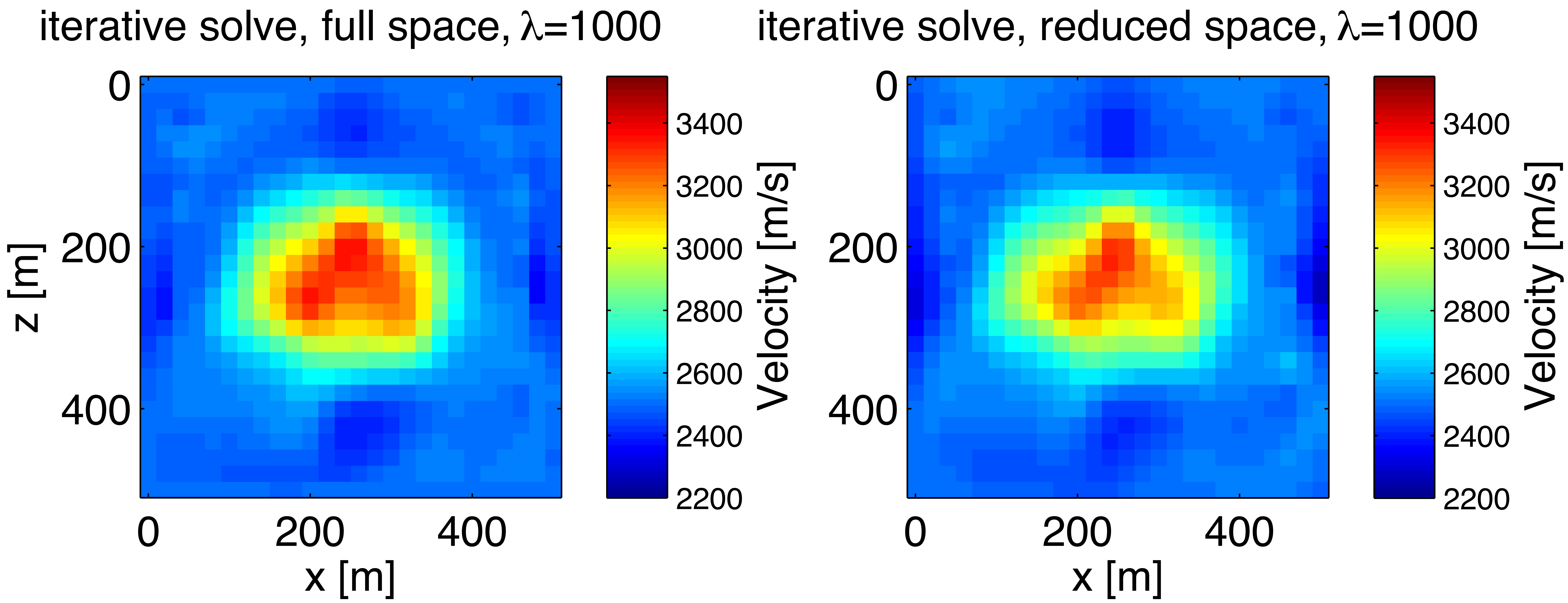

initial guess
true model
z [m]
0
200
400
x [m]
0
200
400
Velocity [m/s]
3400
3200
3000
2800
2600
2400
2200

Toy problem



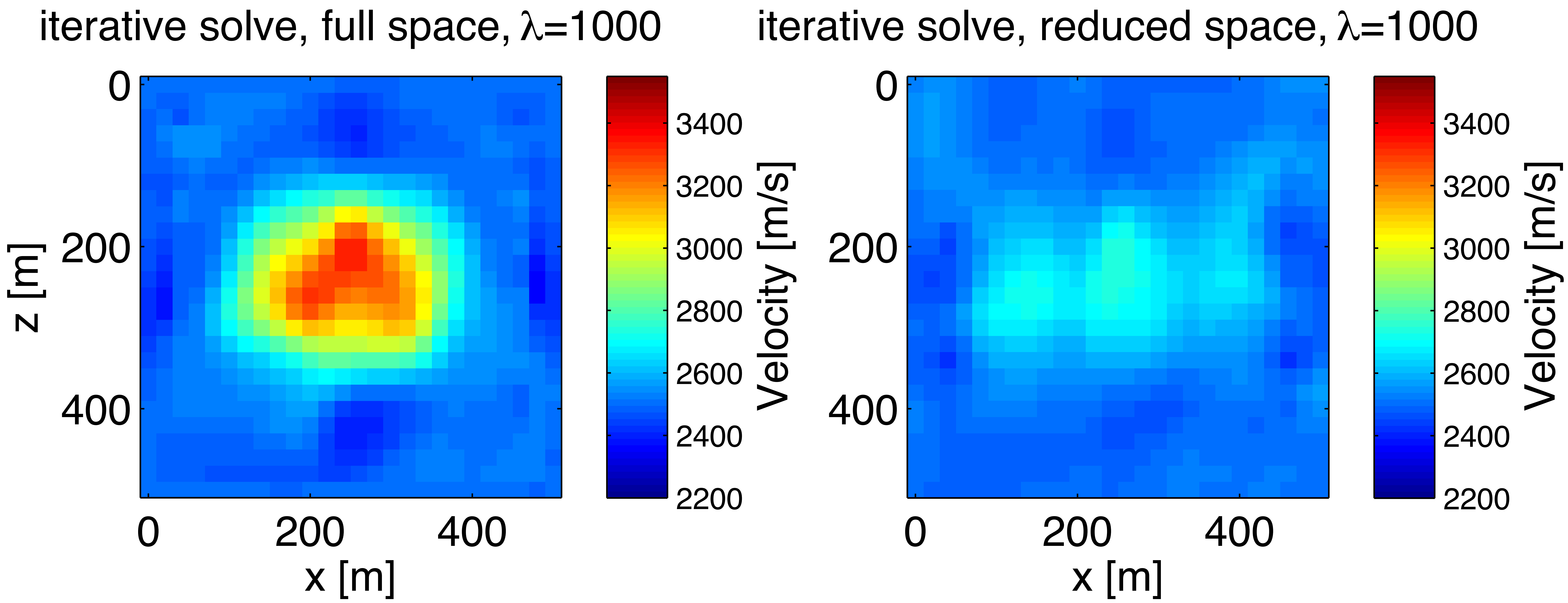
direct solution for least-squares problems

Toy problem



accurate iterative solution for least-squares problems

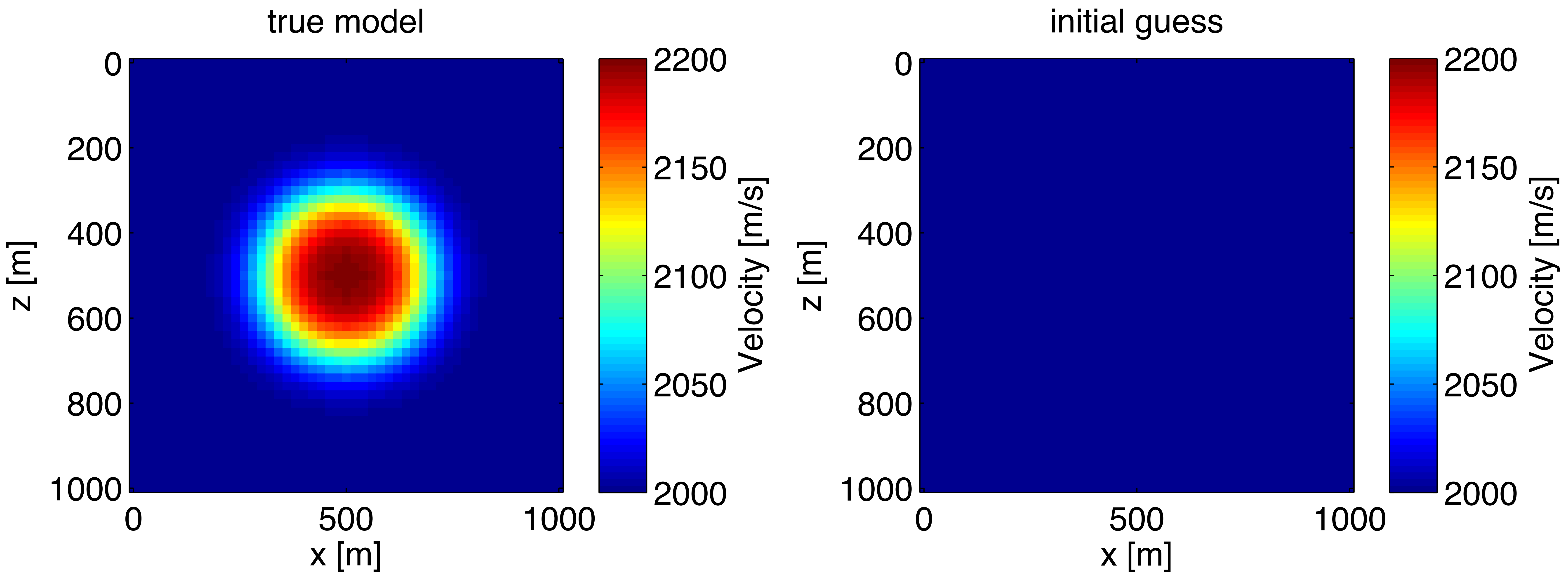
Toy problem



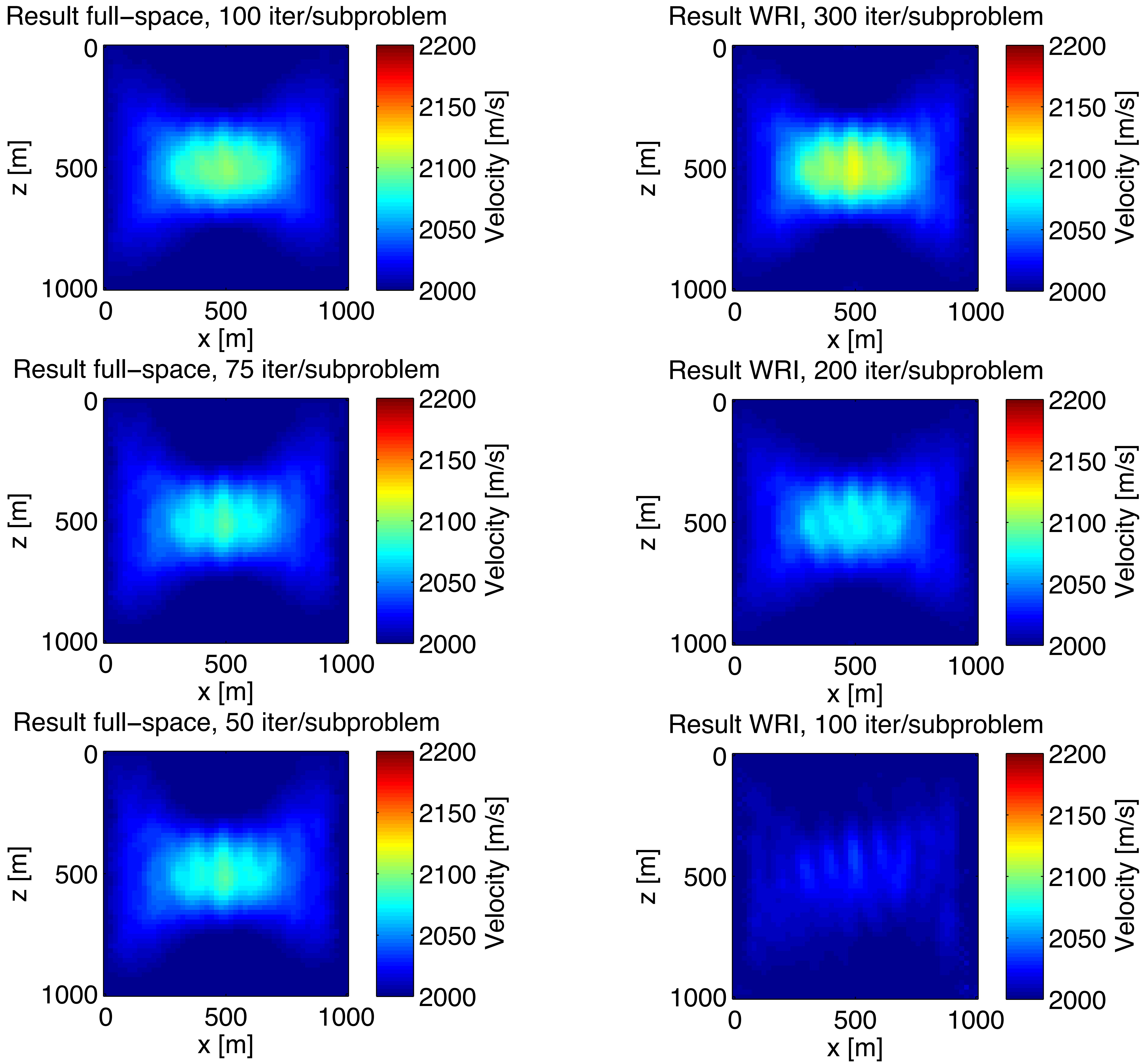
inaccurate iterative solution for least-squares problems

Toy problem

- cross-well setting
- 4 frequencies [4-10] Hz
- 20 sources, 9 receivers

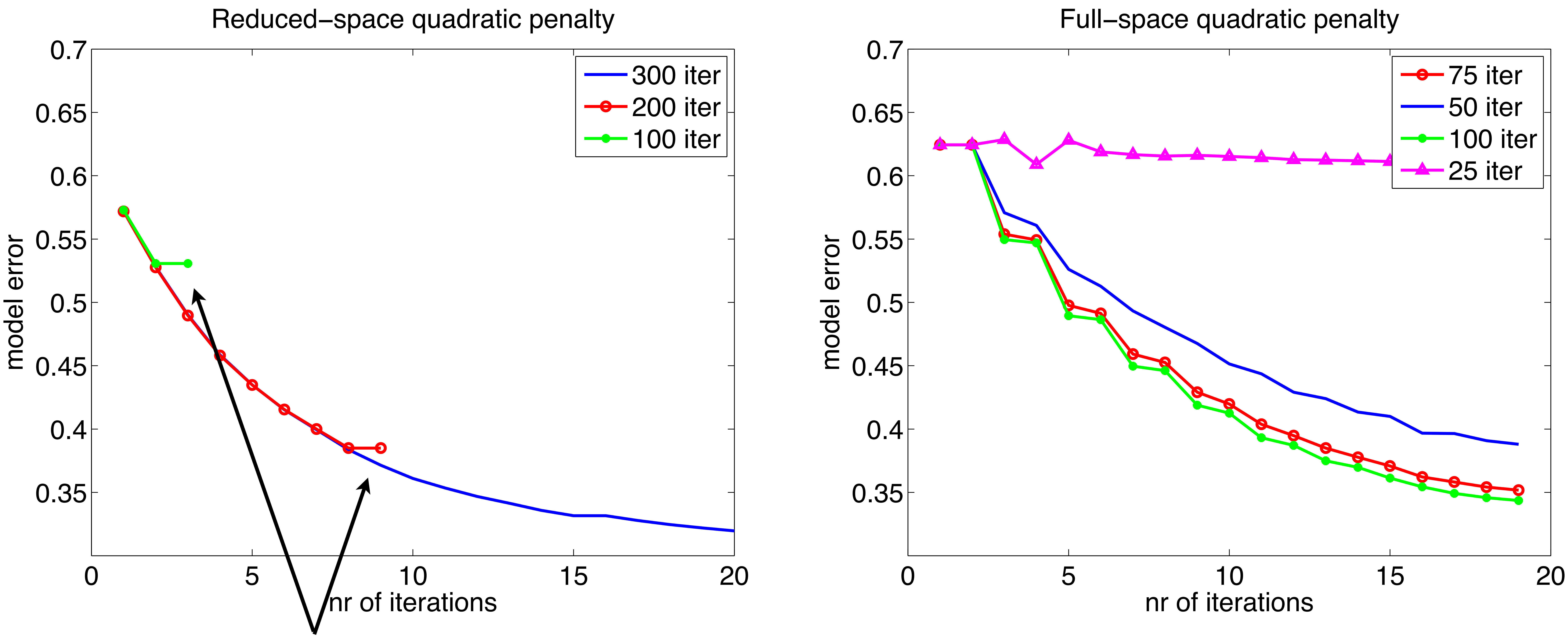


Results



Result full–space, 100 iter/subproblem
Result WRI, 300 iter/subproblem
Result full–space, 75 iter/subproblem
Result WRI, 200 iter/subproblem
Result full–space, 50 iter/subproblem
Result WRI, 100 iter/subproblem
z [m]
x [m]
Velocity [m/s]
0
500
1000
2000
2050
2100
2150
2200

Results:
model error for various iteration budgets for sub-problems



no descent direction found for inaccurate sub-problem solves

Memory requirements

Keep in memory all fields for all frequencies & sources

Can be distributed over multiple nodes

Quadratic-penalty based full-space avoids storage of multipliers, but may not be enough by itself

Feasible? Need

- parallel computing
- simultaneous sources
- stochastic optimization to reduce the working memory
- small frequency batches

Full vs Reduced-space

	Reduced-space FWI	Reduced-space WRI	Full-space
Hessian, gradient & function evaluation	solve PDE's	solve 1 least-squares problem	~free
Hessian, gradient & function evaluation	inexact	inexact	exact
Hessian	dense	sparse + dense	sparse
memory for fields	2 fields per parallel process	1 field per parallel process	all fields in memory (can be distributed over nodes)
working memory	1 gradient & update direction	1 gradient & update direction	update directions & gradients in memory

~free = sparse matrix-vector products

Conclusions

WRI's extends the search space

- less reliant on starting models
- but requires accurate solves

All-at-once methods remain memory intensive

- but less reliant on accurate solves
- still need accurate initialization wavefields by guaranteeing data fits

Bottom line. Not clear which one will perform better... yet.

SLIM

Acknowledgements

Thanks to Chevron and

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, DownUnder GeoSolutions, Hess, Petrobras, PGS, Subsalt Ltd, WesternGeco, and Woodside.

References

1. Eisenstat, Stanley C., and Homer F. Walker. "Globally convergent inexact Newton methods." SIAM Journal on Optimization 4.2 (1994): 393-422.

2. Tristan van Leeuwen and Felix J. Herrmann, frequency-domain seismic inversion with controlled sloppiness, SIAM Journal on Scientific Computing, 36 (2014), pp. S192–S217.

3. Delves, L. M., and I. Barrodale. "A fast direct method for the least squares solution of slightly overdetermined sets of linear equations." IMA Journal of Applied Mathematics 24.2 (1979): 149-156.

4. B Peters, FJ Herrmann, T van Leeuwen. Wave-equation Based Inversion with the Penalty Method-Adjoint-state Versus Wavefield-reconstruction Inversion. 76th EAGE Conference, 2014.

5. M.J. Grote, J. Huber, and O. Schenk, Interior point methods for the inverse medium problem on massively parallel architectures, Procedia Computer Science, 4 (2011), pp. 1466 – 1474. Proceedings of the International Conference on Computational Science, {ICCS} 2011.

6. Eldad Haber, Uri M Ascher, and Doug Oldenburg, On optimization techniques for solving nonlinear inverse problems, Inverse Problems, 16 (2000), pp. 1263–1280.

7. E Haber and U M Ascher, Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Problems, 17 (2001), p. 1847.

8. I Epanomeritakis, V Akcelik, O Ghattas, and J Bielak. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. Inverse Problems, 24(3):034015, June 2008.

9. George Biros and Omar Ghattas, Parallel lagrange–newton–krylov– schur methods for pde-constrained optimization. part i: The krylov–schur solver, SIAM Journal on Scientific Computing, 27 (2005), pp. 687–713.

10. R.E. Kleinman and P.M.van den Berg, A modified gradient method for two- dimensional problems in tomography, Journal of Computational and Applied Mathematics, 42 (1992), pp. 17 – 35.