A quadratic-penalty full-space method for waveform inversion

Bas Peters, Felix J. Herrmann

2015 SINBAD Fall Consortium meeting. October 27.

SLIM
University of British Columbia

Motivation

Full-Waveform Inversion (FWI) works well when a good start model and/or low-frequency data is available.

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\mathrm{pred}}(\mathbf{m})-\mathbf{d}_{\mathrm{obs}}\|_2^2$$

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

Motivation

Full-Waveform Inversion (FWI) works well when a good start model and/or low-frequency data is available.

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\mathrm{pred}}(\mathbf{m})-\mathbf{d}_{\mathrm{obs}}\|_2^2$$

If iterative solvers in frequency domain are used, we cannot compute:

$$H(\mathbf{m})^{-1}\mathbf{q} = \mathbf{u}$$

Instead we obtain:

$$\hat{\mathbf{u}} = H(\mathbf{m})^{-1}(\mathbf{q}+\mathbf{r_u}) = H(\mathbf{m})^{-1}\mathbf{r_u} + \mathbf{u}$$

Motivation

Full-Waveform Inversion (FWI) works well when a good start model and/or low-frequency data is available.

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\text{pred}}(\mathbf{m}) - \mathbf{d}_{\text{obs}}\|_2^2$$

If iterative solvers in frequency domain are used, we cannot compute:

$$H(\mathbf{m})^{-1}\mathbf{q} = \mathbf{u}$$

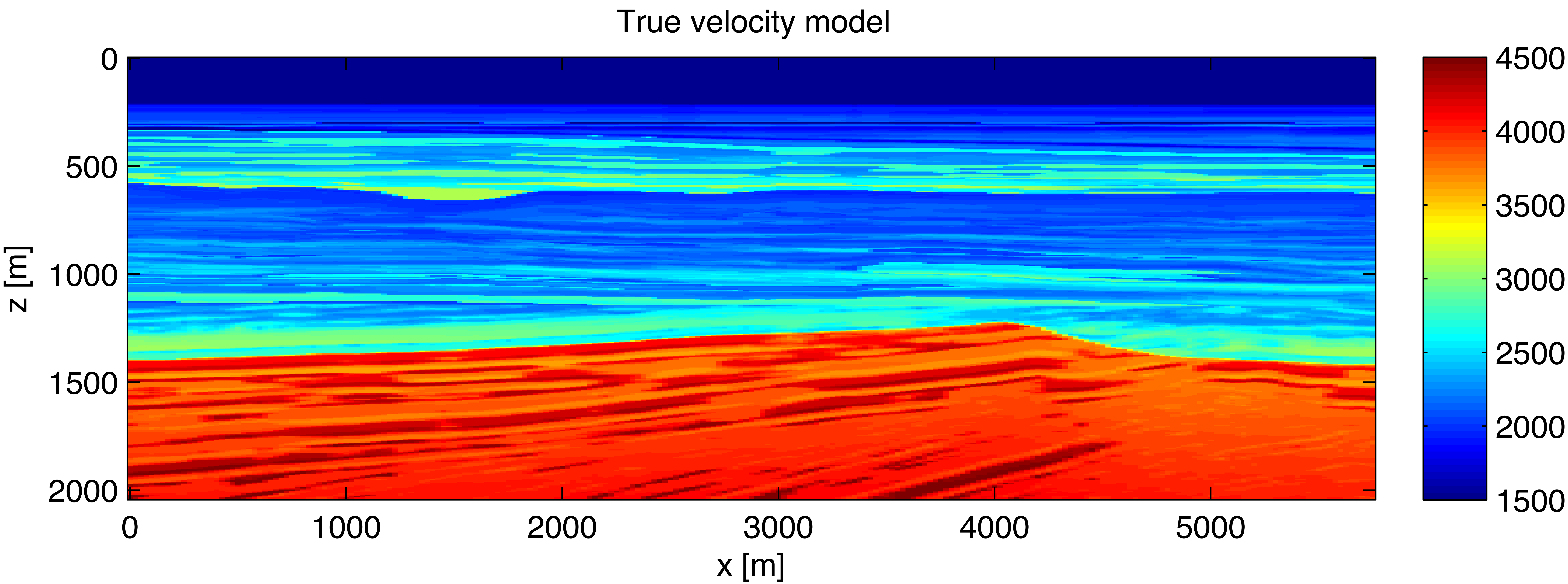
Instead we obtain:

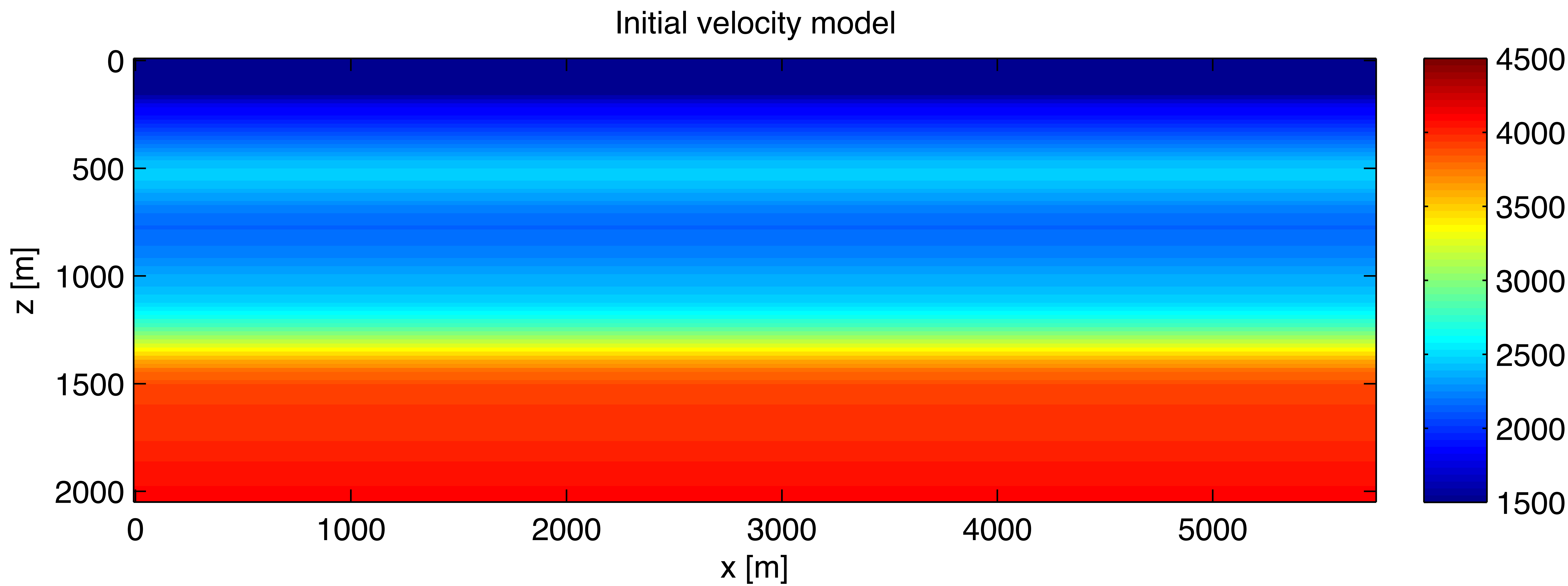
$$\hat{\mathbf{u}} = H(\mathbf{m})^{-1}(\mathbf{q} + \mathbf{r_u}) = \underbrace{H(\mathbf{m})^{-1}\overbrace{\mathbf{r_u}}^{\text{residual}}}_{\text{error}} + \mathbf{u}$$

Example

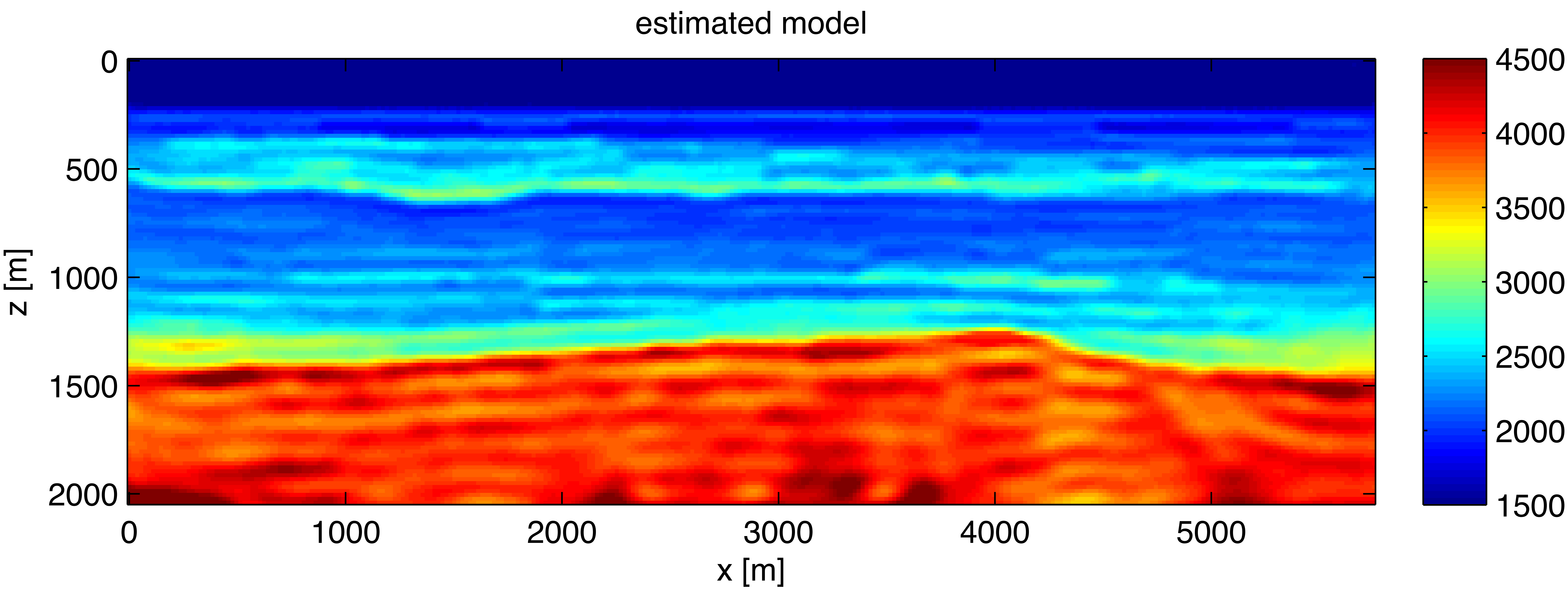
Illustration of what happens when PDE's are solved inaccurately

- 3-15Hz
- good start model
- full-offset source & receiver array
- noise free data

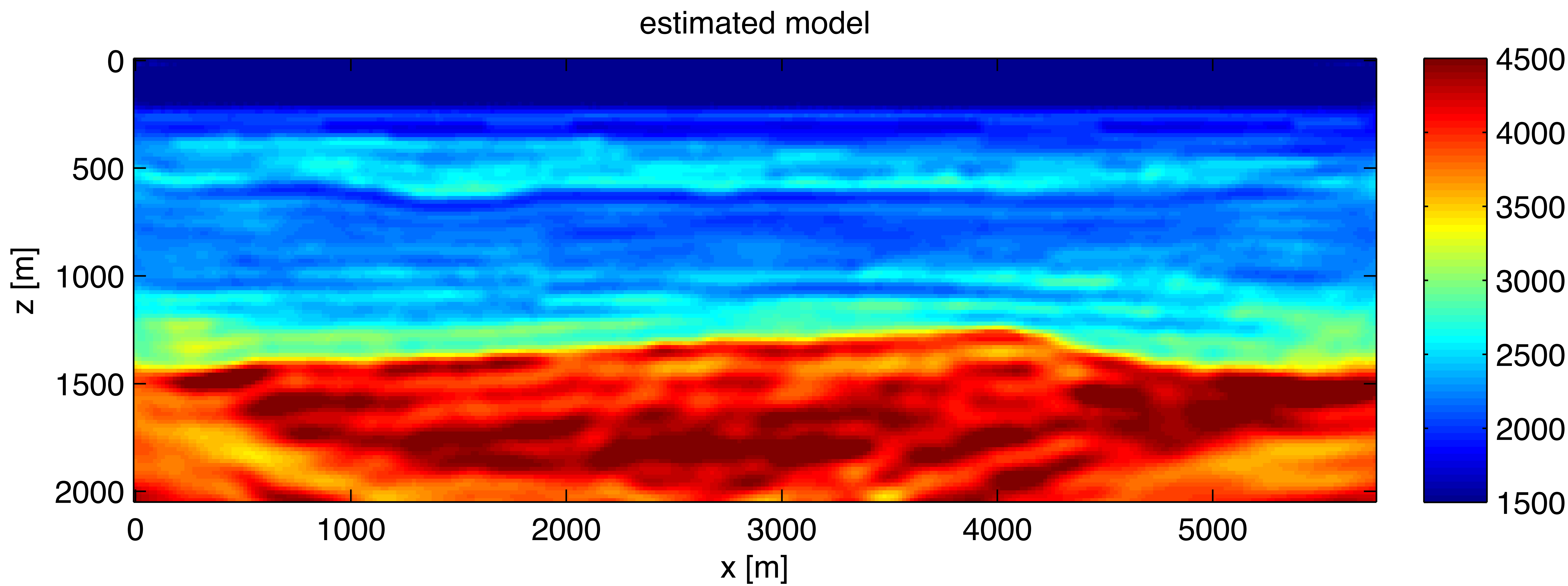

True velocity model
0
500
1000
1500
2000
z [m]
0
1000
2000
3000
4000
5000
x [m]
4500
4000
3500
3000
2500
2000
1500


Initial velocity model
0
500
1000
1500
2000
z [m]
0
1000
2000
3000
4000
5000
x [m]
4500
4000
3500
3000
2500
2000
1500

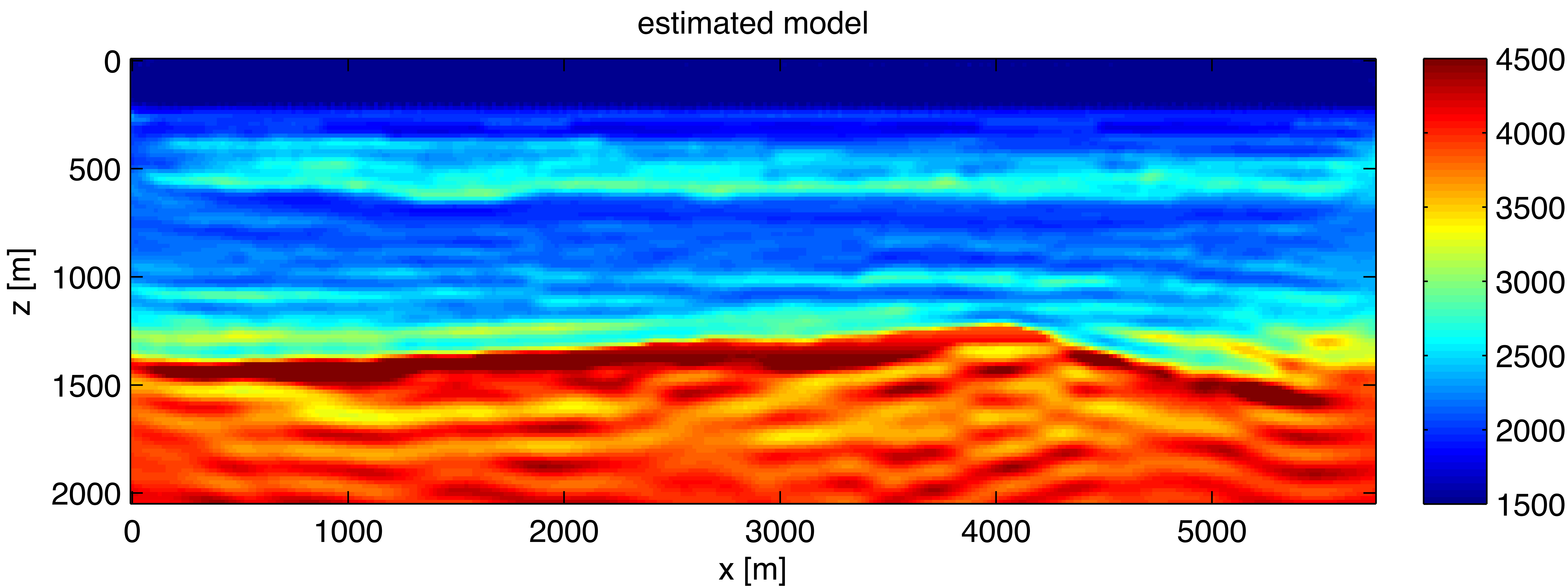
FWI using **very accurate iterative** solver & LBFGS



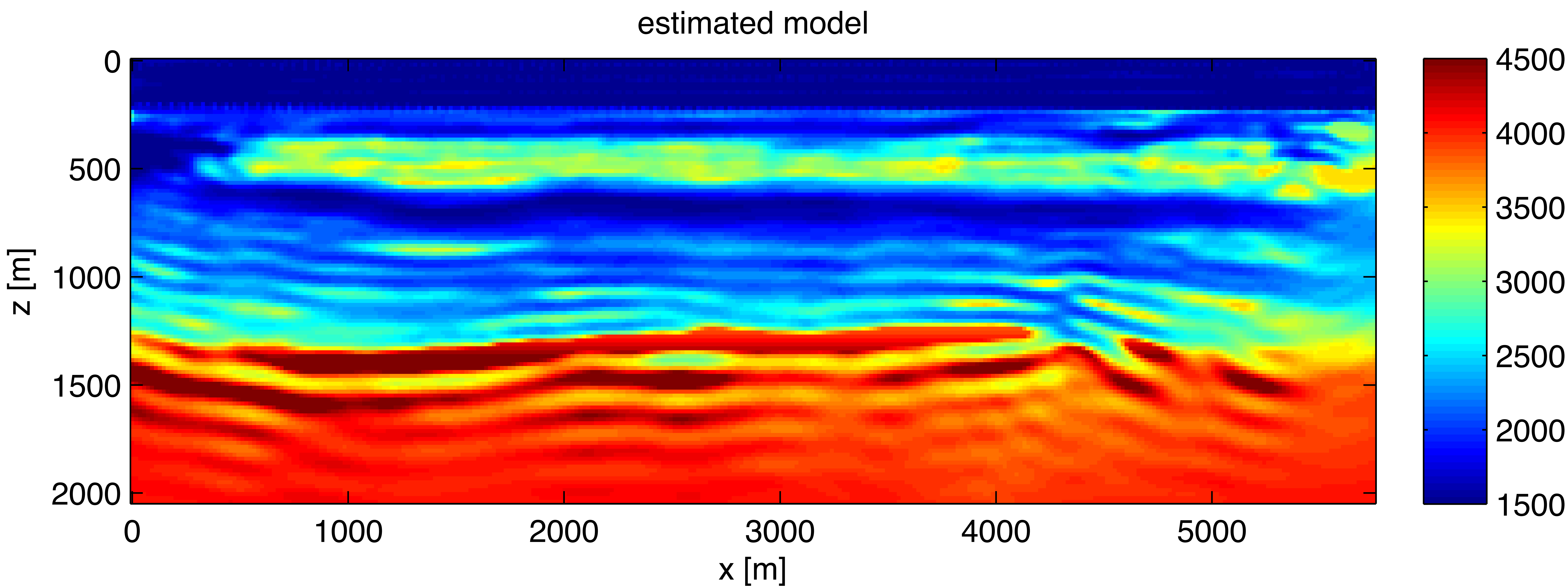
FWI using **accurate iterative** solver & LBFGS



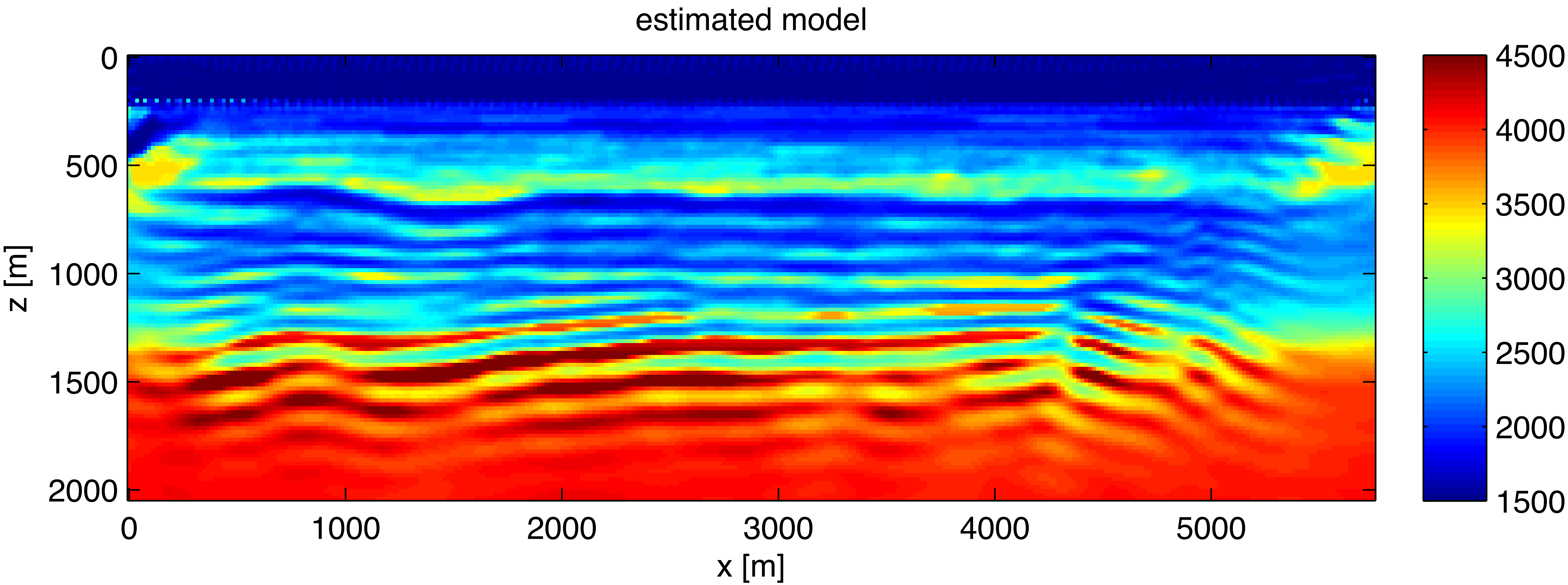
FWI using **less accurate iterative** solver & LBFGS



FWI using **inaccurate accurate iterative** solver & LBFGS



FWI using **very inaccurate accurate iterative** solver &LBFGS



Inexact PDE solves

reduced-space: (includes FWI) **[A Tarantola, 1984; E Haber et al., 2000; I Epanomeritakis et al., 2008]**

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

→ error in medium parameter update

- storage as low as two fields at a time
- dense reduced-Hessian
- requires at least some extra safeguards/accuracy control **[T. van Leeuwen & F.J. Herrmann, 2014]**

Goal

This talk is about deriving an algorithm which:

- allows for inexact solutions of linear systems
- enjoys similar parallelism and memory requirements as FWI

Data, PDE's and constraints

For the true medium parameters and true fields we know that:

$$H(\mathbf{m})\mathbf{u} = \mathbf{q} \ \& \ P\mathbf{u} = \mathbf{d}$$

Many ways to use these equations to form:

- objectives,
- constraints
- algorithms

Adjoint-state based FWI is just one algorithm.

Data, PDE's and constraints

objective

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

constraint

Data, PDE's and constraints

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

objective

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

constraint

SLIM

Data, PDE's and constraints

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

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

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\text{pred}}(\mathbf{m})-\mathbf{d}_{\text{obs}}\|_2^2$$

SLIM

Data, PDE's and constraints

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

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

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\text{pred}}(\mathbf{m})-\mathbf{d}_{\text{obs}}\|_2^2$$

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

Data, PDE's and constraints

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

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

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\text{pred}}(\mathbf{m})-\mathbf{d}_{\text{obs}}\|_2^2$$

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

Multi-experiment structure:

$$P\mathbf{u} - \mathbf{d} \qquad\qquad H(\mathbf{m})\mathbf{u} - \mathbf{q}$$

$$\begin{pmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_k \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_k \end{pmatrix} - \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_k \end{pmatrix} \qquad \begin{pmatrix} H_1 & & & \\ & H_2 & & \\ & & \ddots & \\ & & & H_k \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_k \end{pmatrix} - \begin{pmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \vdots \\ \mathbf{q}_k \end{pmatrix}$$

$$k = n_{\mathrm{src}} \times n_{\mathrm{freq}} \qquad\qquad k \times N \quad \text{field parameters}$$

A quadratic-penalty based full space method

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

σ - λ relation is known
[W. Gander, 1980; A. Bjork, 1996]

$$\downarrow$$

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

A quadratic-penalty based full space method

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

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 for working subset of fields & medium parameters

$$G = \frac{\partial H(\mathbf{m})\mathbf{u}}{\partial \mathbf{m}}$$

A quadratic-penalty based full space method

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

- update fields & medium parameters simultaneously
- function value, gradient, Hessian evaluation is ~free & exact
- sparse Hessian
- theory allows for inexact updates computations
- requires storage of working subset of fields
+ working memory (gradients, Hessian & update)
- update computation is challenging

A quadratic-penalty based full space method

Approximate: block diagonal & positive (semi) definite

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

- give up some of Newton’s method properties
- update computation intrinsically parallel per field
- no need to form off-diagonal blocks

philosophy: more & cheaper iterations

Memory requirements

save fields for the working subset of frequencies & sources

can be distributed over multiple nodes

feasible? need

- parallel computing
- **simultaneous sources** (redrawing is possible)
- small frequency batches

Algorithm

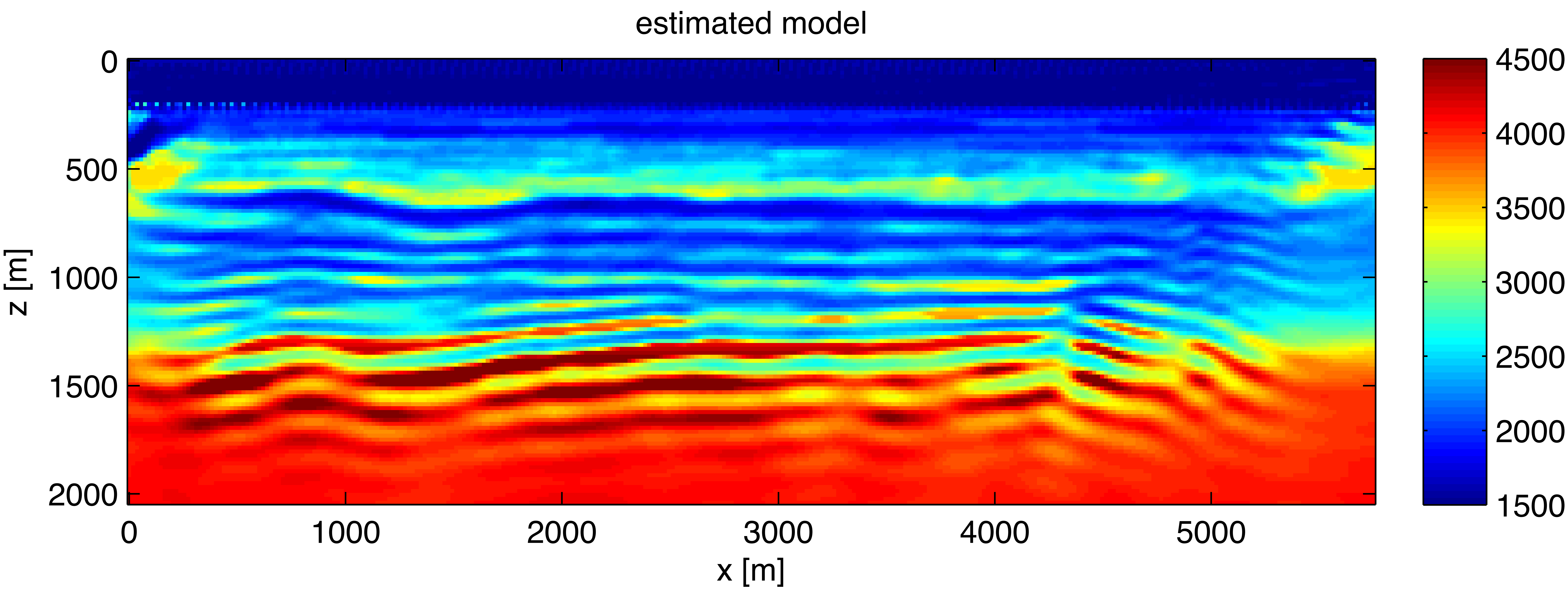
0. construct initial guess $\mathbf{m}$ for medium and $\mathbf{u}_i$ for each field

while not converged **do**

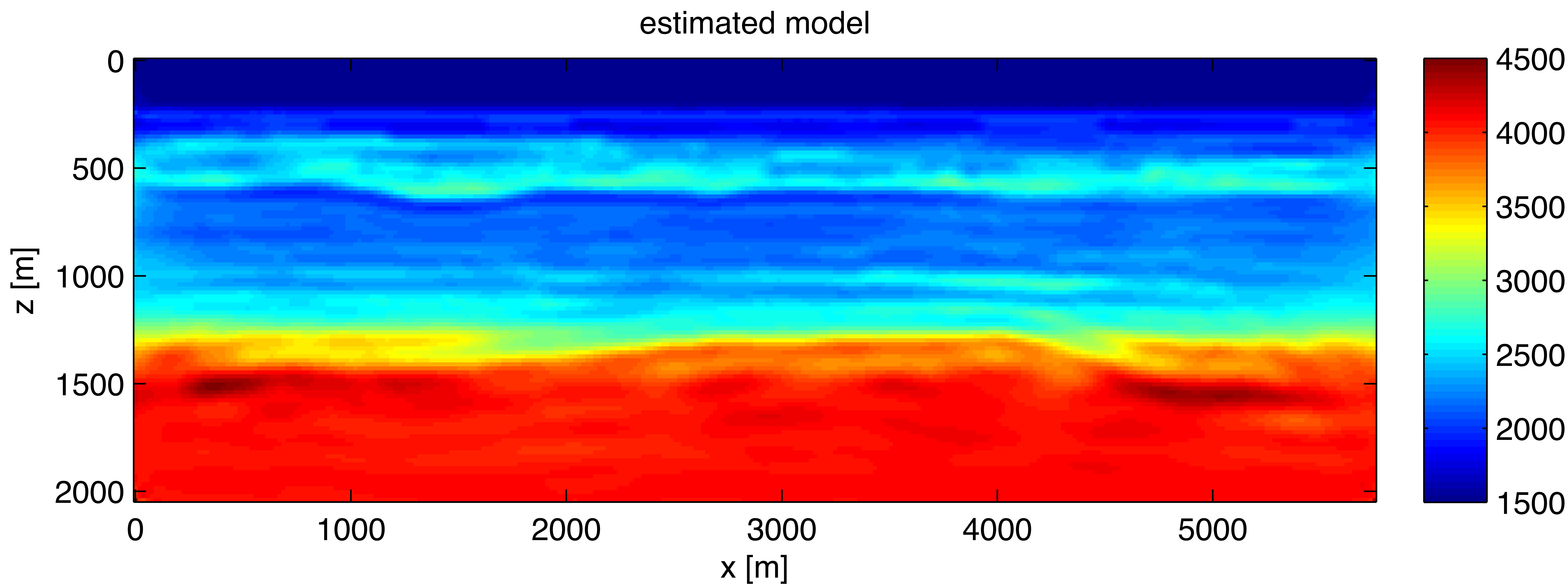
1. form Hessian and gradient `// form (~free)`
2. ignore the $\nabla^2_{\mathbf{u},\mathbf{m}}\phi, \nabla^2_{\mathbf{m},\mathbf{u}}\phi$ blocks `// approximate`
3. find $\delta\mathbf{m}$ & each $\delta\mathbf{u}_i$ in parallel `// solve`
4. find steplength α using linesearch `// evaluate (~free)`
5. $\mathbf{m} = \mathbf{m} + \alpha\delta\mathbf{m}$ & $\mathbf{u} = \mathbf{u} + \alpha\delta\mathbf{u}$ `// update model and fields`

end

FWI using **very inaccurate accurate iterative** solver &LBFGS



Full-space method of this talk using very inaccurate accurate iterative solver



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

The presented algorithm is a quadratic-penalty version of Lagrangian-based all-at-once algorithms:

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

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

solve (inexactly) at every iteration: Newton-KKT system

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

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

Lagrangian based full-space methods also store the multipliers + corresponding gradient & Hessian blocks.

no intrinsic parallel structure

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

* higher order terms

number of field variables: $2 \times n_{\mathrm{src}} \times n_{\mathrm{freq}} \times n_{\mathrm{grid}}$

Inexact PDE solves
– full-space vs reduced-space

reduced-space (FWI):

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

→ error in medium parameter update

full-space (this talk):

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

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

Full vs Reduced-space

	Reduced-space	Full-space
Hessian, gradient & function evaluation	solve PDE's	~free
Hessian, gradient & function evaluation	inexact	exact
Hessian	dense	sparse
memory for fields	2 fields per parallel process	working subset of simultaneous source fields in memory
working memory	1 gradient & update direction	update directions & gradients in memory

~free = sparse matrix-vector products

Computational details

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

Optimizing the fields means the fields including PML region.

G may be diagonal. Example: if $\mathbf{m}$ is slowness squared and discretization is standard finite-difference.

Main computational challenge is to inexactly solve:

$$\big(P^*P + \lambda^2 H^*H\big)\delta\mathbf{u} = P^*(P\mathbf{u} - \mathbf{d}) + \lambda^2 H^*\big(H\mathbf{u} - \mathbf{q}\big)$$

-> has same structure as least-squares problem in WRI, use same solver.

SLIM

Conclusions

Constructed a quadratic-penalty based full-space method which:

- updates fields & medium parameters simultaneously
- main computations are intrinsically parallel
- suitable for frequency domain waveform inversion with iterative solvers
- con: need to store working subset of simultaneous source fields
- but, less storage needed compared to Lagrangian full-space methods

Acknowledgements

Thanks to our sponsors

This work was financially supported by SINBAD Consortium members BG Group, BGP, CGG, Chevron, ConocoPhillips, DownUnder GeoSolutions, Hess, Petrobras, PGS, Schlumberger, Statoil, Sub Salt Solutions and Woodside;
and by the Natural Sciences and Engineering Research Council of Canada via NSERC Collaborative Research and Development Grant DNOISEII (CRDPJ 375142--08).

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. 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.
4. Eldad Haber, Uri M Ascher, and Doug Oldenburg, On optimization techniques for solving nonlinear inverse problems, Inverse Problems, 16 (2000), pp. 1263–1280.
5. E Haber and U M Ascher, Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Problems, 17 (2001), p. 1847.
6. 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.
7. 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.
8. 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.
9. Haber, Eldad, and Uri M. Ascher. "Preconditioned all-at-once methods for large, sparse parameter estimation problems." Inverse Problems 17.6 (2001): 1847.

References

10. Ake Bjork, Numerical methods for least squares problems. siam, 1996.

11. Walter Gander, Least squares with a quadratic constraint, Numerische Mathematik, 36 (1980), pp. 291–307.