A sparse reduced Hessian approximation for multi-parameter wavefield reconstruction inversion

Bas Peters & Felix J. Herrmann

October, 2014
SEG annual meeting

SLIM
University of British Columbia

Multi-parameter PDE-constrained optimization

Main issues with multi-parameter inverse problems:

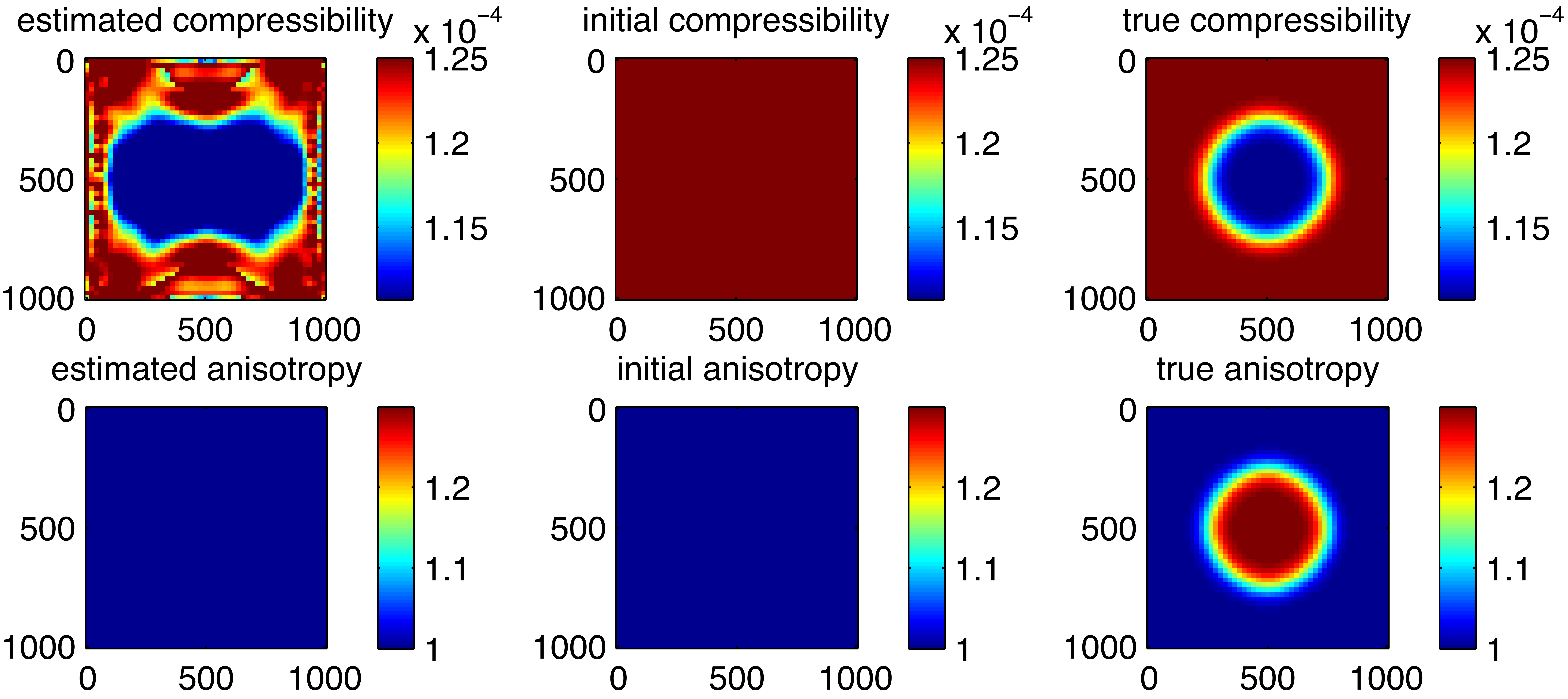
- Non uniqueness
- Parameter scaling

Multi-parameter PDE-constrained optimization

Cross-well toy example

Update parameters simultaneously using a quasi-Newton method.

Multi-parameter PDE-constrained optimization



Multi-parameter PDE-constrained optimization

Data fit > 99%

Proposed solutions include:

- find ‘best’ parameterization
- sequential/alternating inversion
- regularization
- manual scaling of gradients

Problems:

- sensitive to parameter choices
- manual fine-tuning

Multi-parameter PDE-constrained optimization

Observation:

- The Hessian naturally provides information about ‘scaling’ and ‘coupling’.

Using the Hessian is also proposed by Lavoué et al. 2014.

Adjoint-state based FWI leads to dense Hessians

Problems:

- Cannot store dense Hessian
- Matrix-vector product cost extra PDE solves

Wavefield reconstruction inversion

Intuitively:

1. Reconstruct the wavefield, based on partial measurements and model estimate
2. Use this field to update the medium parameters

Wavefield reconstruction inversion

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

Objective:

$$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\sum_{kl} \|P\bar{\mathbf{u}}_{kl} \overset{\substack{\text{Data-misfit}\\ \Big\downarrow}}{-} \mathbf{d}_{kl}\|_2^2 + \frac{\lambda^2}{2}\|H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} \overset{\substack{\text{PDE-misfit}\\ \Big\downarrow}}{-} \mathbf{q}_{kl}\|_2^2$$

where $$\bar{\mathbf{u}}_{kl} = \arg\min_{\mathbf{u}_{kl}} \left\| \begin{pmatrix} \lambda H_k(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u}_{kl} - \begin{pmatrix} \lambda \mathbf{q}_{kl} \\ \mathbf{d}_{kl} \end{pmatrix} \right\|_2$$

and λ is a tradeoff parameter between PDE-fit and data-fit

Wavefield reconstruction inversion

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

Objective:

Data-misfit ⬇ PDE-misfit ⬇

$$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\sum_{kl} \|P\bar{\mathbf{u}}_{kl} - \mathbf{d}_{kl}\|_2^2 + \frac{\lambda^2}{2}\|H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\|_2^2$$

with gradient:

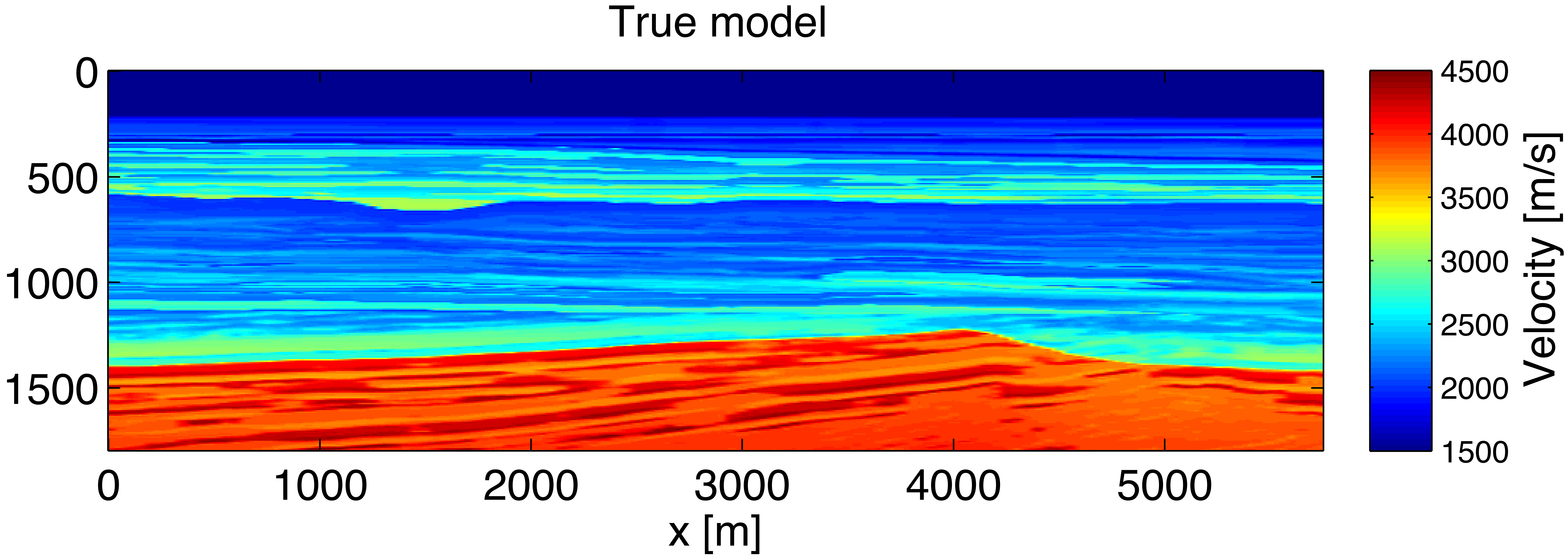
$$\nabla_{\mathbf{m}}\bar{\phi}_\lambda = \sum_{kl} \lambda^2 G_{kl}(\mathbf{m}, \bar{\mathbf{u}}_{kl})^* \left(H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\right)$$

and G_{kl}^* is the partial derivative of the discrete Helmholtz system.

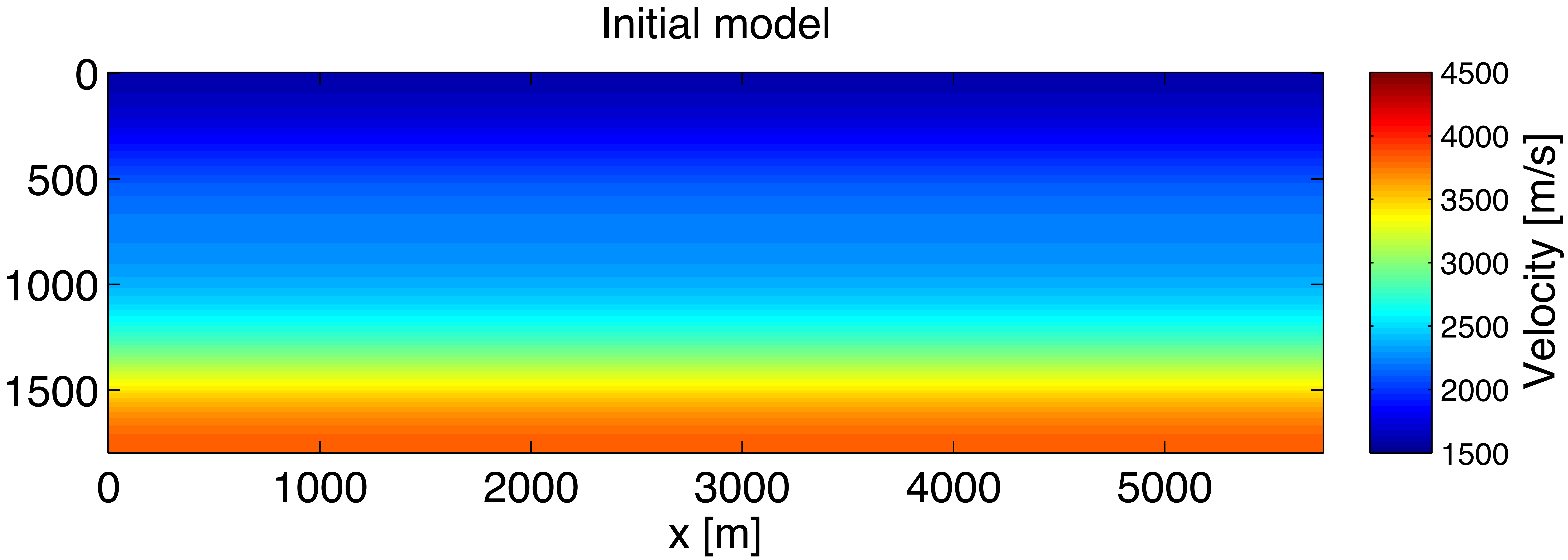
Wavefield reconstruction inversion

$$\bar{\mathbf{u}}_{kl} = \arg\min_{\mathbf{u}_{kl}} \left\| \begin{pmatrix} \lambda H_k(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u}_{kl} - \begin{pmatrix} \lambda \mathbf{q}_{kl} \\ \mathbf{d}_{kl} \end{pmatrix} \right\|_2$$

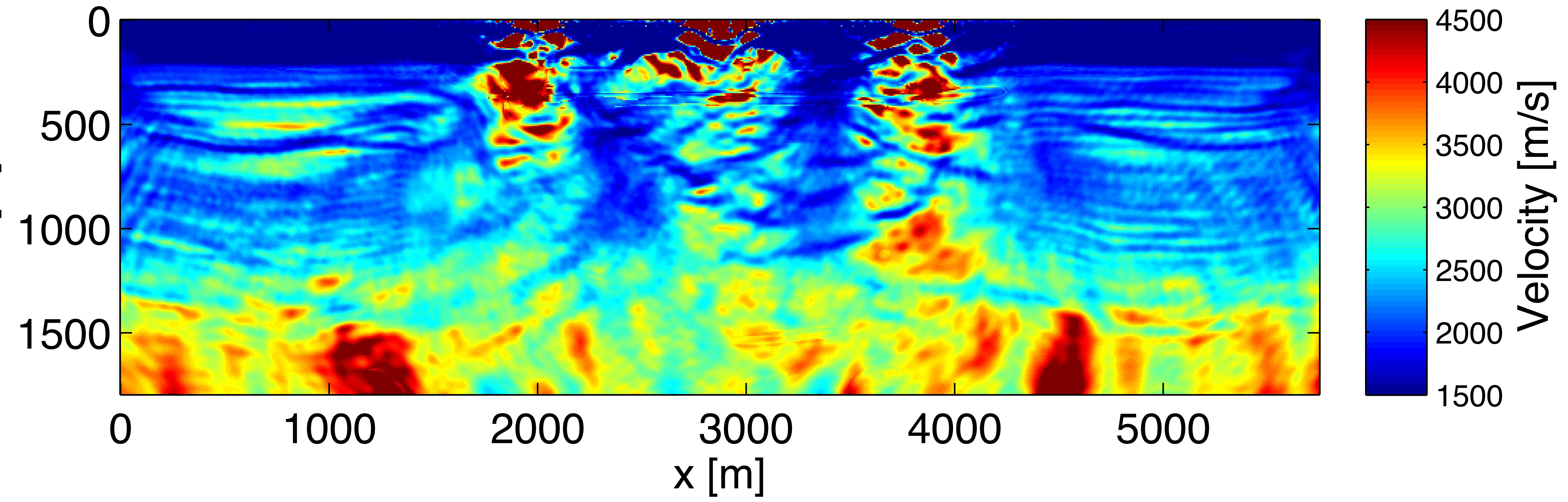
- not exactly a wavefield
- field can fit the data
- no cycle-skipping
- objective function still non-convex



Example from **[Peters et al. 2013]**

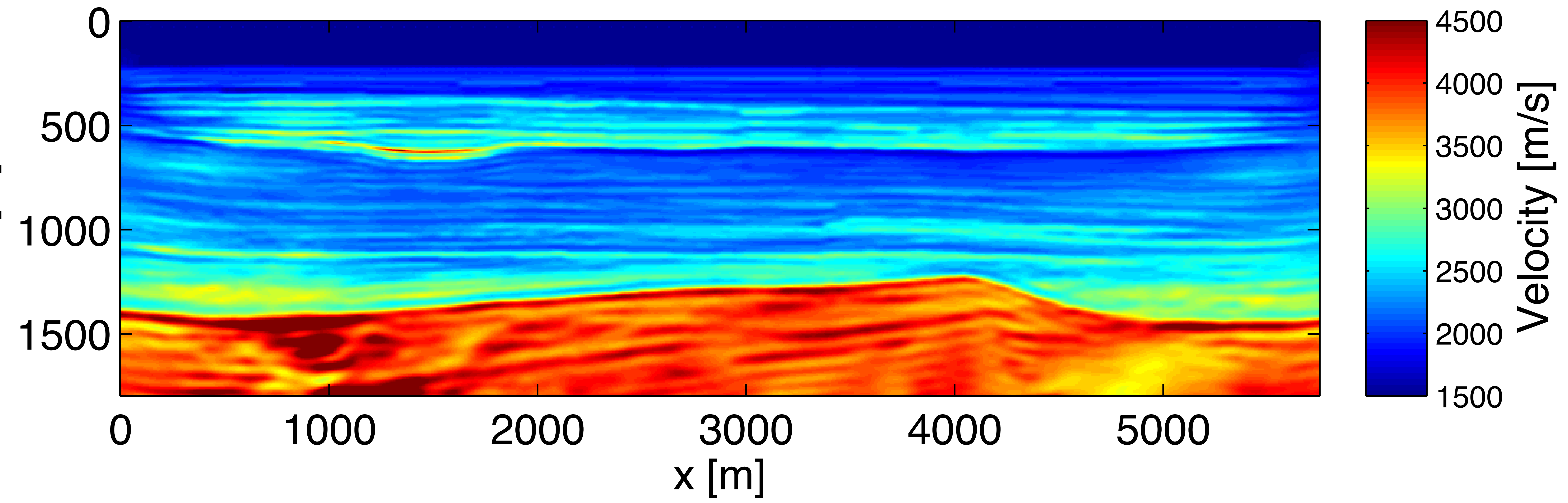


Result FWI



5- 28Hz data

Result WRI, λ =1



Wavefield reconstruction inversion

Main benefits of WRI:

- WRI is less sensitive to the initial modal
- Can show that the Gauss-Newton Hessian is diagonal

Wavefield reconstruction inversion

WRI

```
for each source $i$
    solve $\begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_{\lambda,i} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$
    $\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \mathrm{diag}(\bar{\mathbf{u}}_{i,\lambda})^* (A(\mathbf{m}) \bar{\mathbf{u}}_{i,\lambda} - \mathbf{q}_i)$
    $H_{GN} = H_{GN} + \lambda^2 \omega^4 \mathrm{diag}(\mathbf{u}_i)^* \mathrm{diag}(\mathbf{u}_i)$
    $\mathbf{m} = \mathbf{m} - \alpha H_{GN}^{-1} \mathbf{g}$
end
```

diagonal
=
pseudo Hessian

Conventional FWI

```
for each source $i$
    solve $A(\mathbf{m}) \mathbf{u}_i = \mathbf{q}_i$
    solve $A(\mathbf{m})^* \mathbf{v}_i = P_i^* (P_i \mathbf{u}_i - \mathbf{d}_i)$
    $\mathbf{g} = \mathbf{g} + \omega^2 \mathrm{diag}(\mathbf{u}_i)^* \mathbf{v}_i$
    $\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$
end
```

Goals

- multi-parameter WRI

- Obtain an approximation of the Hessian without solving extra PDE's.
- Approximation must be sparse and available in memory explicitly.

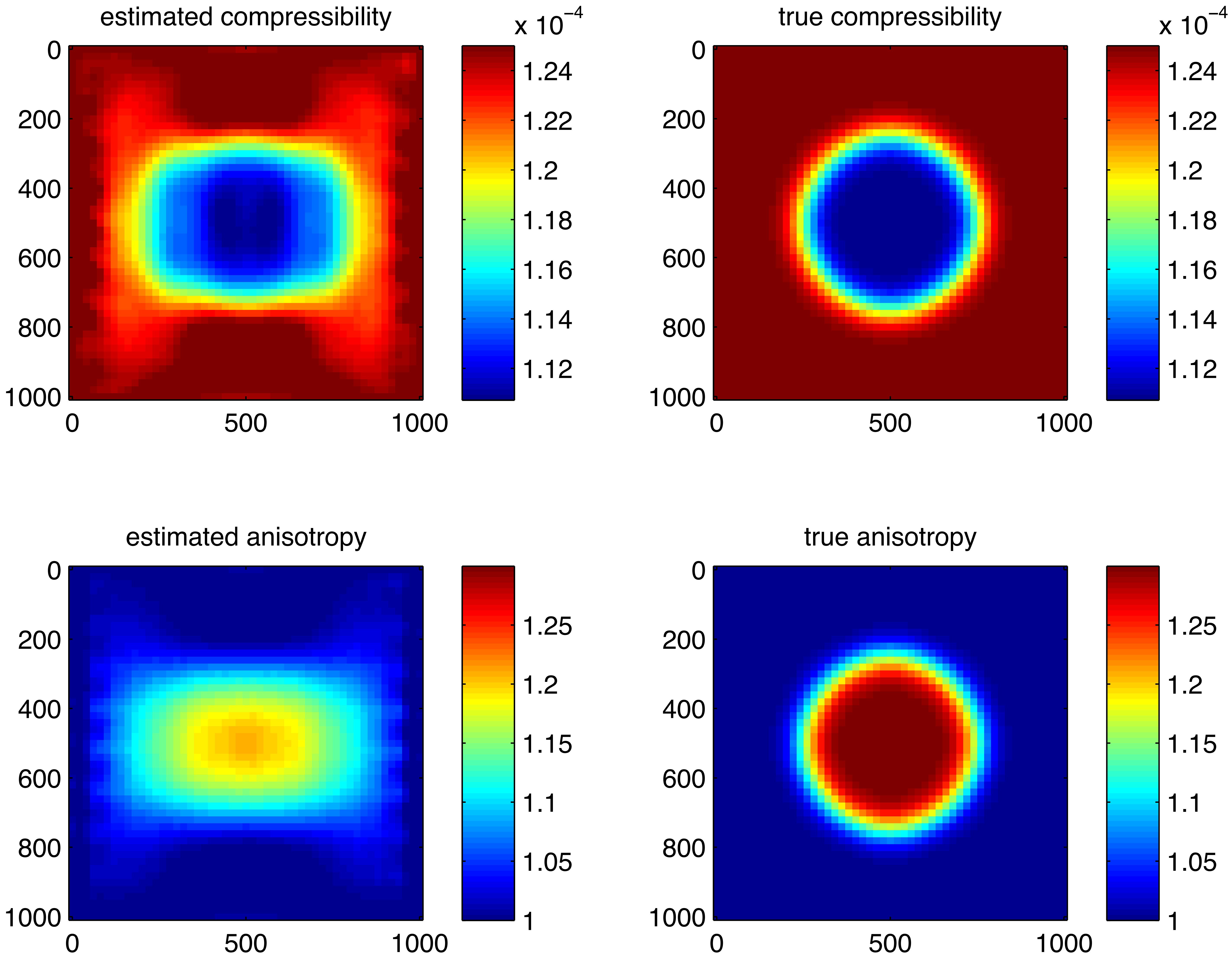
Goals

- multi-parameter WRI

- Obtain an approximation of the Hessian without solving extra PDE's.
- Approximation must be sparse and available in memory explicitly.

This is possible!

Multi-parameter WRI example



Multi-parameter PDE-constrained optimization

Constrained multi-parameter problem:

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

Unconstrained quadratic penalty formulation:

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

A quadratic penalty form was also the starting point of other, very different algorithms **[Kleinman & van den Berg, 1992, Banerjee et al. 2013]**

Multi-parameter PDE-constrained optimization

Constrained multi-parameter problem:

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

Unconstrained quadratic penalty formulation:

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

Corresponding Newton-system (sparse)

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

A reduced-space algorithm

Corresponding Newton-system (sparse)

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

write: $\begin{pmatrix} E & B \\ C & D \end{pmatrix}$

Solve for the fields, $\nabla_{\mathbf{u}}\phi_\lambda(\mathbf{u}, \boldsymbol{\kappa}, \mathbf{b}) = 0$, then

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

$$\left(\overset{\substack{\text{sparse}\\ \Downarrow}}{D} - \overset{\substack{\text{dense}\\ \Downarrow}}{C}E^{-1}B\right) \begin{pmatrix}\delta_{\boldsymbol{\kappa}} \\ \delta_{\mathbf{u}}\end{pmatrix} = \begin{pmatrix}\nabla_{\boldsymbol{\kappa}}\phi_\lambda \\ \nabla_{\mathbf{b}}\phi_\lambda\end{pmatrix} - CE^{-1}\overset{\substack{=0\\ \Downarrow}}{\nabla_{\mathbf{u}}\phi_\lambda(\bar{\mathbf{u}}, \boldsymbol{\kappa}, \mathbf{b})}$$

Sparse-Dense splitting does not happen in the Lagrangian form.

A reduced-space algorithm

$$\left(\overset{\substack{\text{sparse}\\ \Downarrow}}{D} - \overset{\substack{\text{dense}\\ \Downarrow}}{CE^{-1}B}\right)\begin{pmatrix}\delta_{\boldsymbol{\kappa}}\\ \delta_{\mathbf{u}}\end{pmatrix} = \begin{pmatrix}\nabla_{\boldsymbol{\kappa}}\phi_\lambda\\ \nabla_{\mathbf{b}}\phi_\lambda\end{pmatrix} - CE^{-1}\overset{\substack{=0\\ \Downarrow}}{\nabla_{\mathbf{u}}\phi_\lambda}(\bar{\mathbf{u}}, \boldsymbol{\kappa}, \mathbf{b})$$

Use the sparse part of the reduced Hessian **[T. van Leeuwen & F.J. Herrmann, 2013]**

This is the approximate Newton step for a reduced objective:

$$\bar{\phi}_\lambda(\mathbf{b}, \boldsymbol{\kappa}) = \frac{1}{2}\overset{\substack{\text{Data-misfit}\\ \Downarrow}}{\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2} + \frac{\lambda^2}{2}\overset{\substack{\text{PDE-misfit}\\ \Downarrow}}{\|H(\mathbf{b}, \boldsymbol{\kappa})\bar{\mathbf{u}} - \mathbf{q}\|_2^2}$$

A reduced-space algorithm

$$\left(\overset{\substack{\text{sparse}\\ \downarrow}}{D} - \overset{\substack{\text{dense}\\ \downarrow}}{CE^{-1}B}\right)\begin{pmatrix}\delta_{\boldsymbol{\kappa}}\\ \delta_{\mathbf{u}}\end{pmatrix} = \begin{pmatrix}\nabla_{\boldsymbol{\kappa}}\phi_\lambda\\ \nabla_{\mathbf{b}}\phi_\lambda\end{pmatrix} - CE^{-1}\overset{\substack{=0\\ \downarrow}}{\nabla_{\mathbf{u}}\phi_\lambda(\bar{\mathbf{u}},\boldsymbol{\kappa},\mathbf{b})}$$

So far I used only the diagonal blocks, yielding a sparse, SPD matrix.

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

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

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

A reduced-space algorithm

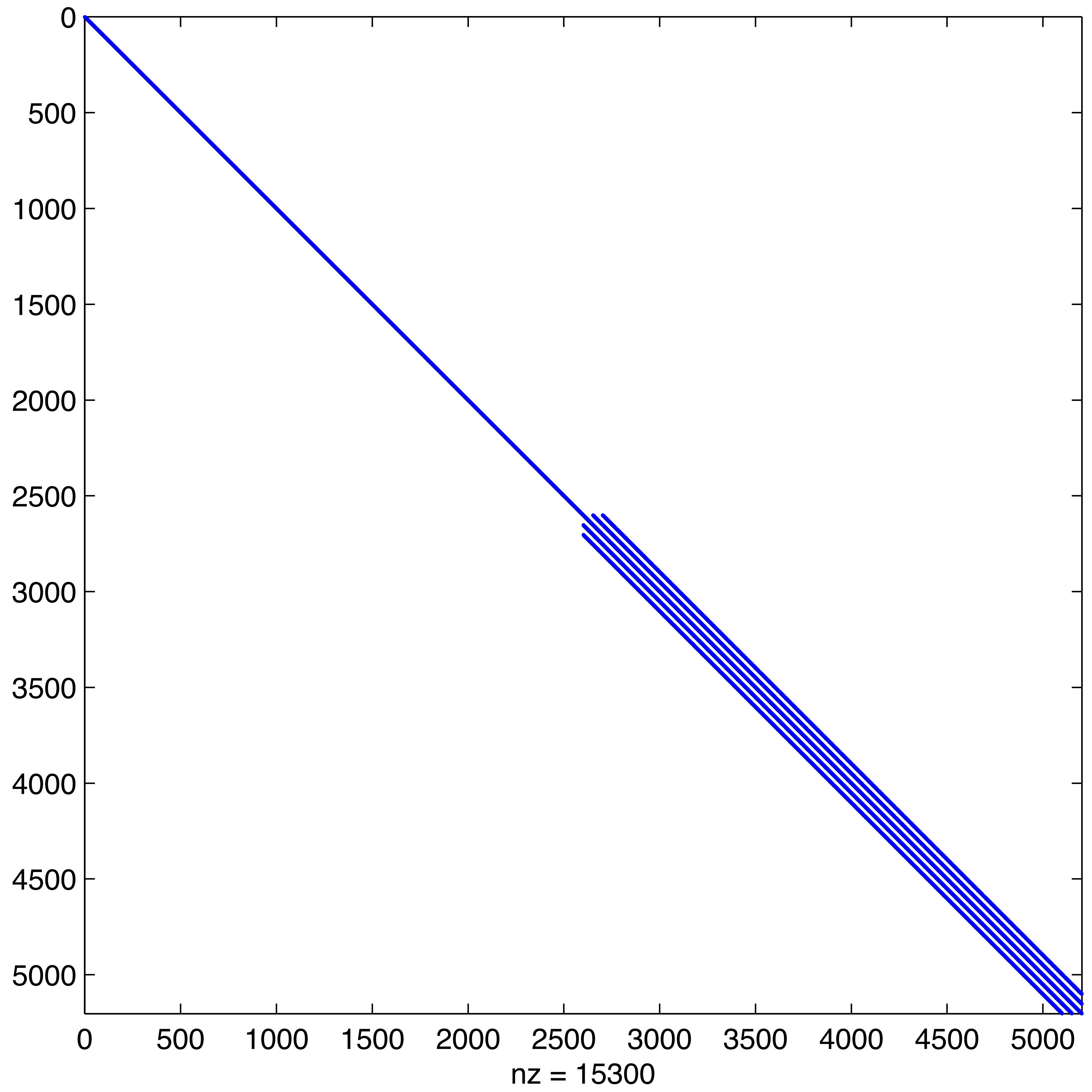
$$\Big(\overset{\substack{\text{sparse}\\ \Downarrow}}{D} - \overset{\substack{\text{dense}\\ \Downarrow}}{C}E^{-1}B\Big)\begin{pmatrix}\delta_{\boldsymbol{\kappa}}\\ \delta_{\mathbf{u}}\end{pmatrix} = \begin{pmatrix}\nabla_{\boldsymbol{\kappa}}\phi_\lambda\\ \nabla_{\mathbf{b}}\phi_\lambda\end{pmatrix} - CE^{-1}\overset{\substack{=0\\ \Downarrow}}{\nabla_{\mathbf{u}}\phi_\lambda}(\bar{\mathbf{u}}, \boldsymbol{\kappa}, \mathbf{b})$$

Accuracy of the approximate Newton step is important:

$$\begin{pmatrix}\delta_{\boldsymbol{\kappa}}\\ \delta_{\mathbf{u}}\end{pmatrix} = D^{-1}\begin{pmatrix}\nabla_{\boldsymbol{\kappa}}\phi_\lambda\\ \nabla_{\mathbf{b}}\phi_\lambda\end{pmatrix}$$

not the accuracy of the Hessian approximation, D ,itself

A reduced-space algorithm



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

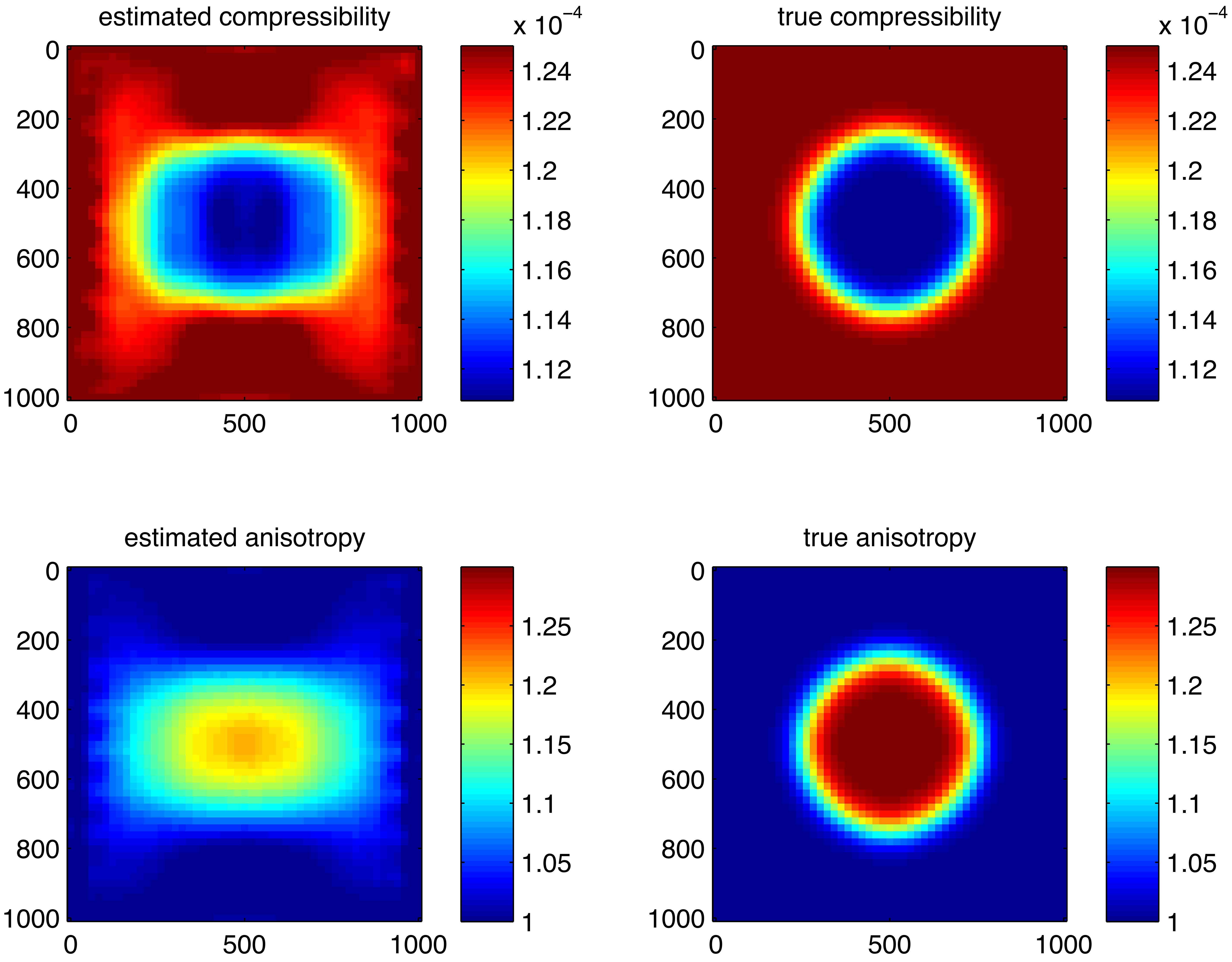
Approximate reduced Hessian sparsity pattern

A reduced-space algorithm

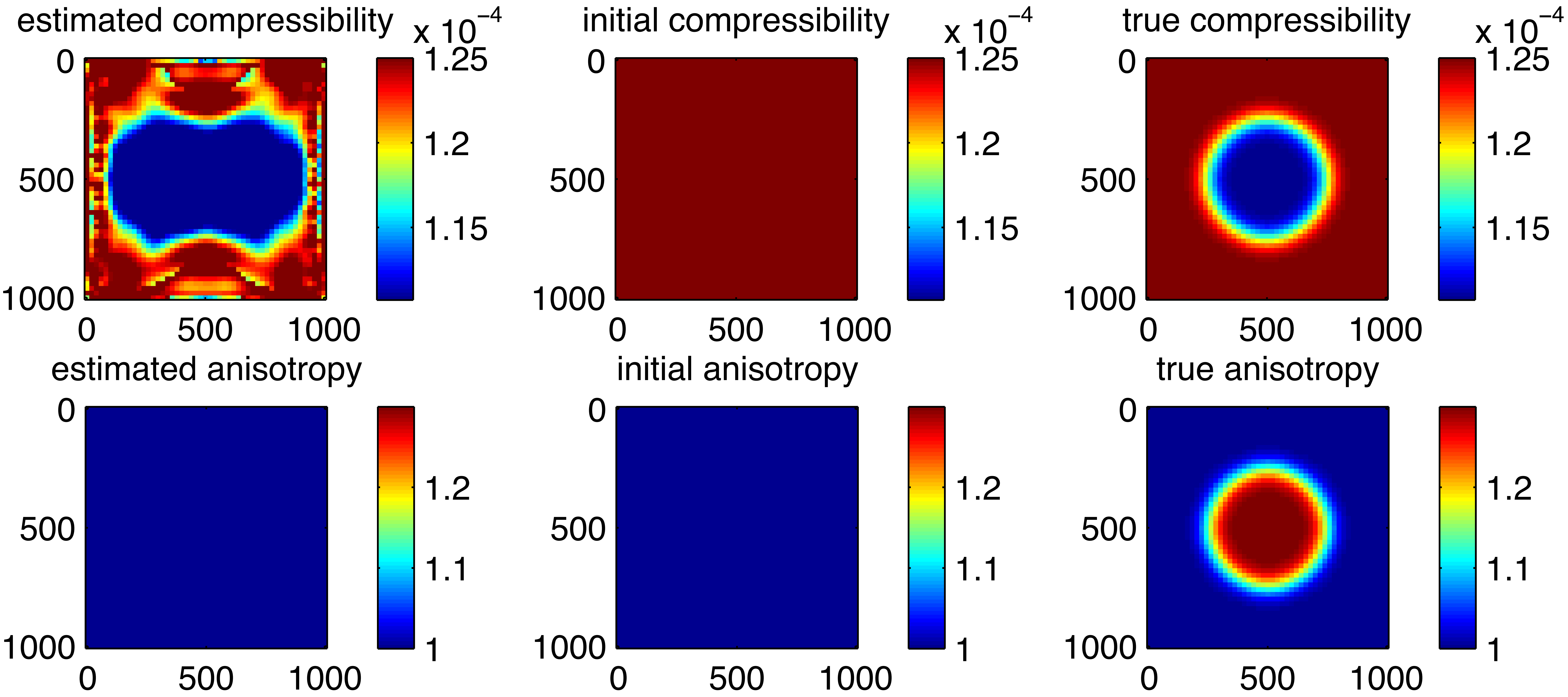
Algorithm 1 Waveform inversion with a sparse Hessian approximation.

```
while Not converged do
1.  $\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda \mathbf{A}(\mathbf{b}, \kappa) \\ \mathbf{P} \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$   // Solve
2.  $\mathbf{G}_{\kappa}, \mathbf{G}_{\mathbf{b}}, \nabla_{\mathbf{b}} \bar{\phi}_{\lambda}, \nabla_{\kappa} \bar{\phi}_{\lambda}$   // Form
3.  $\mathbf{p}_{gn} = \tilde{\mathbf{H}}^{-1} \mathbf{g}$ // Solve
4.  find steplength $\alpha$   // Linesearch
5.  $\mathbf{m} = \mathbf{m} + \alpha \mathbf{p}_{gn}$   // update model
end
```

A reduced-space algorithm; example



standard quasi-Newton result



Summary & Conclusions

WRI provides access to a reduced Hessian approximation which is:

- sparse
- easy to invert
- scales different parameter classes based on the optimization, data and PDE itself

Can extend to more than two parameters.

Different optimization strategies do not remove non uniqueness.

no theoretical guarantees yet

Acknowledgements

The SLIM students & postdocs

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, BP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Total SA, WesternGeco, and Woodside.

References

Peters, Bas, and Felix J. Herrmann. 2014. “A Sparse Reduced Hessian Approximation for Multi-Parameter Wavefield Reconstruction Inversion.” EAGE 2014.

R.E. Kleinman and P.M.van den Berg . 1992. A modified gradient method for two- dimensional problems in tomography. Journal of Computational and Applied Mathematics.

Biswanath Banerjee and Timothy F. Walsh and Wilkins Aquino and Marc Bonnet. 2013. Large scale parameter estimation problems in frequency-domain elastodynamics using an error in constitutive equation functional. Computer Methods in Applied Mechanics and Engineering.

Leeuwen, Tristan van, and Felix J. Herrmann. 2013. “Mitigating Local Minima in Full-Waveform Inversion by Expanding the Search Space.” Geophysical Journal International .

Leeuwen, Tristan van, and Felix J. Herrmann. 2013. “A Penalty Method for PDE-Constrained Optimization.” UBC.

Bas Peters, Felix J. Herrmann, and Tristan van Leeuwen. 2014. “Wave-equation based inversion with the penalty method: adjoint-state versus wavefield-reconstruction inversion”. EAGE.