Matrix-free quadratic-penalty methods for PDE-constrained optimization

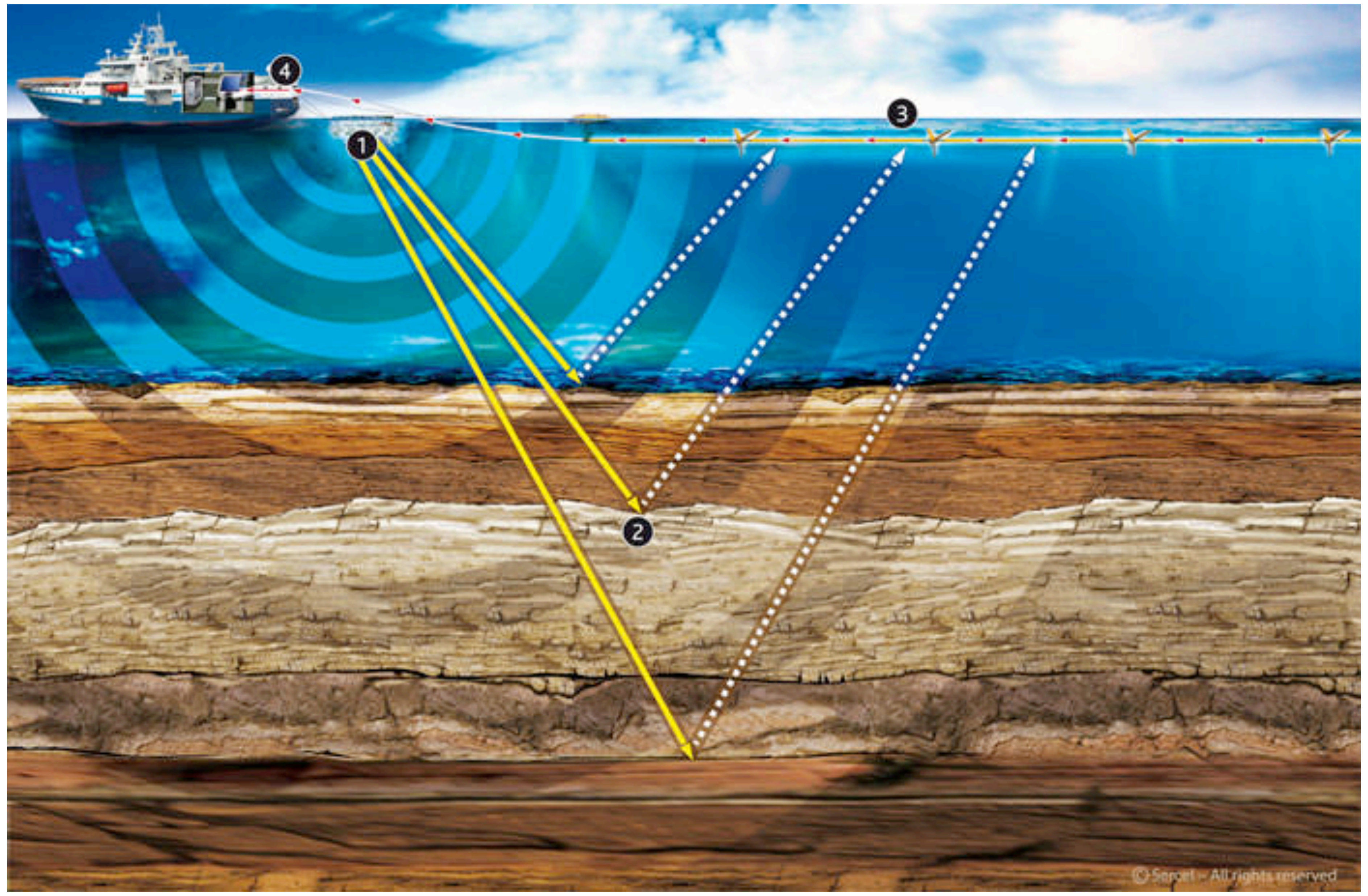
Bas Peters, Felix J. Herrmann, Chen Greif

Matrix-Free Methods for Large-Scale Optimization and Applications
2015 SIAM Conference on Computational Science and Engineering
March 14, 2015

SLIM
University of British Columbia

PDE-constrained optimization

This talk is about parameter estimation with wavefields.



[from:http://www.sercel.com/about/Pages/what-is-geophysics.aspx]

PDE-constrained optimization

This talk is about parameter estimation with the Helmholtz equation.

Challenging because:

- oscillatory data and predicted fields
- non-convex
- local minimizers often unacceptable
- 1 PDE: ~ [1e6 - 1e9] grid points
- working with multiple [10 - 1000] PDE's simultaneously is very challenging

PDE-constrained optimization

known:

- source/receiver locations
- source function (sometimes)
- the PDE (usually simplified physics)

unknown:

- PDE-coefficients (acoustic velocity)

notation:

- fields (‘state variables’)
- medium parameters (‘control variables’)

[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

PDE-constrained optimization

Multi-experiment structure:

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

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

$k \times N$ field parameters

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

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

eliminate field variables

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

reduced space

[E Haber et al., 2000 ; I Epanomeritakis et al., 2008]
[T. van Leeuwen & F.J. Herrmann, 2014]

- storage as low as two fields at a time
- highly nonlinear function value computation is
  - expensive
  - inexact when sub-problems are solved iteratively
- dense reduced-Hessian
- requires extra safeguards/accuracy control

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

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

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

eliminate field variables

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

- few algorithms are based on the quadratic-penalty form
- interchanging objective and constraints lead to same algorithm
- also works with fixed λ

[R.E. Kleinman & P.M.van den Berg, 1992 ; T. van Leeuwen & F.J. Herrmann, 2013]

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

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

eliminate field variables

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

reduced gradient method/
reduced Lagrangian

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

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

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

reduced quadratic-penalty

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

A reduced-space quadratic-penalty method

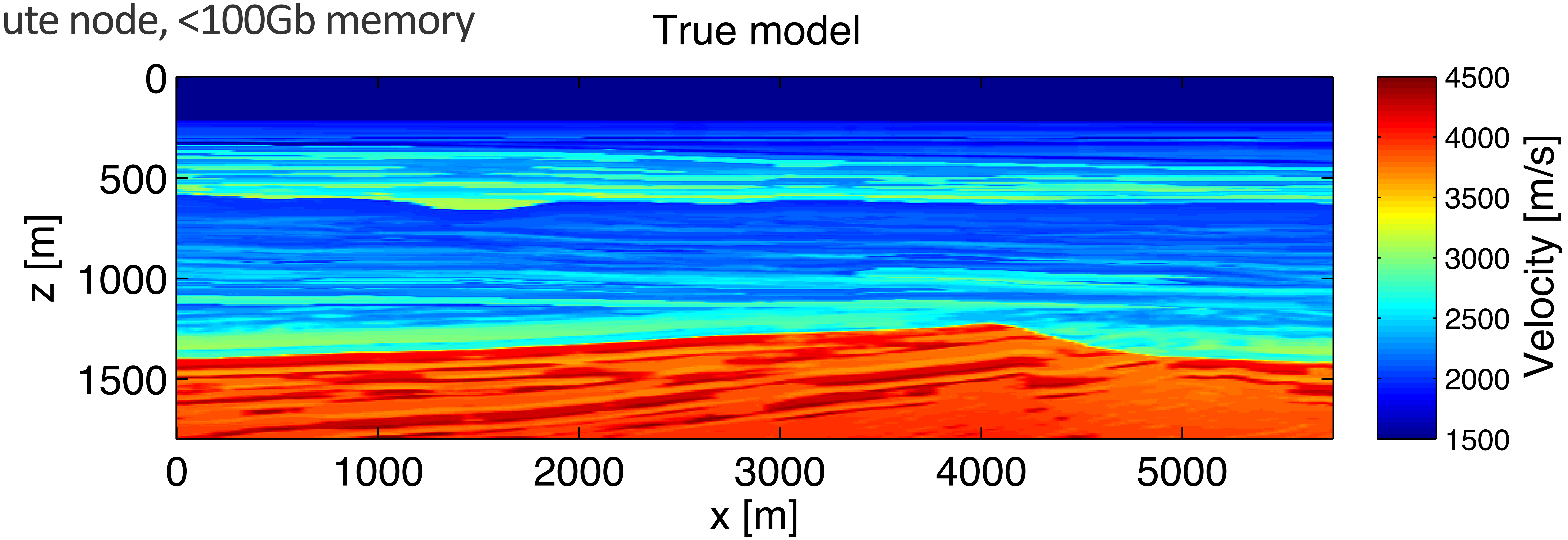
To minimize: $$\min_{\mathbf{m}} \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$$

at every iteration:

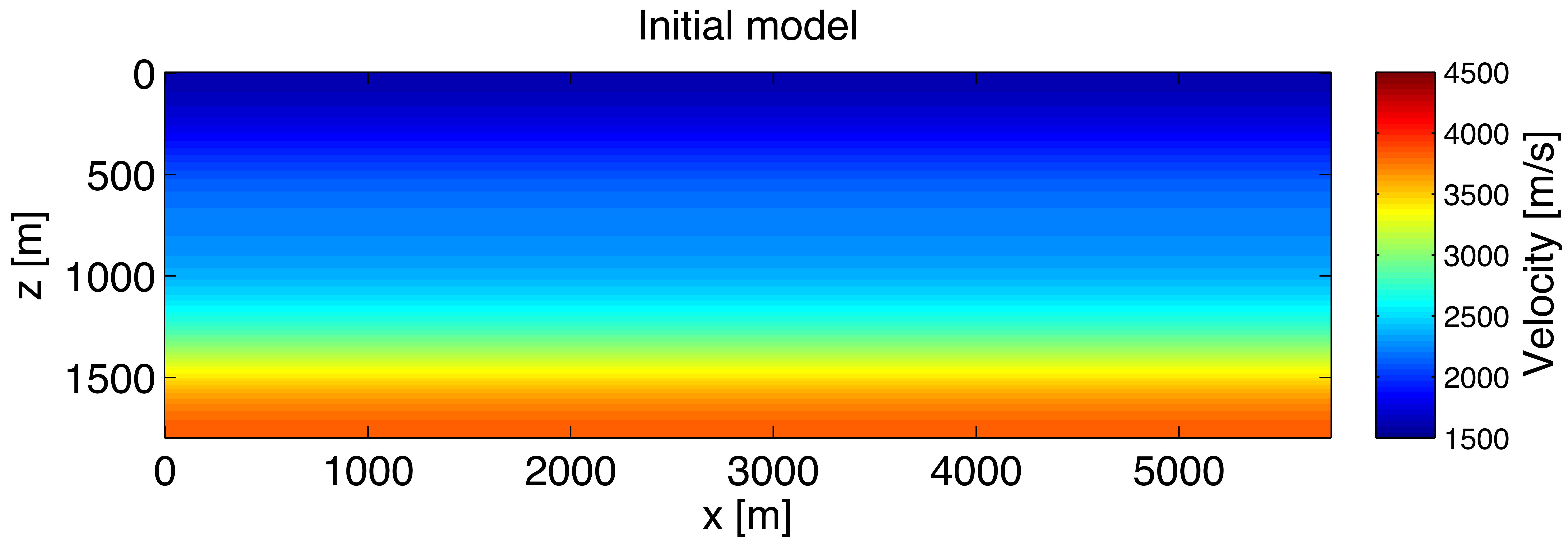
- compute $$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$
- evaluate $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$ & $\nabla_{\mathbf{m}}\bar{\phi}(\mathbf{m},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

~ 2e8 field variables

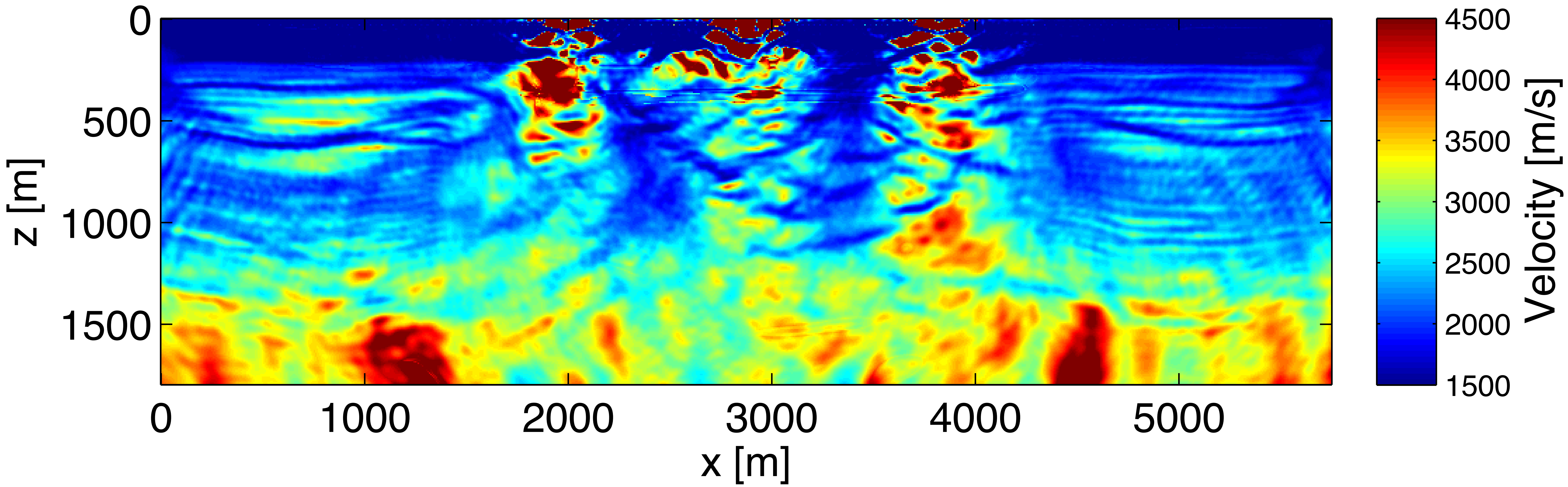
1 compute node, <100Gb memory



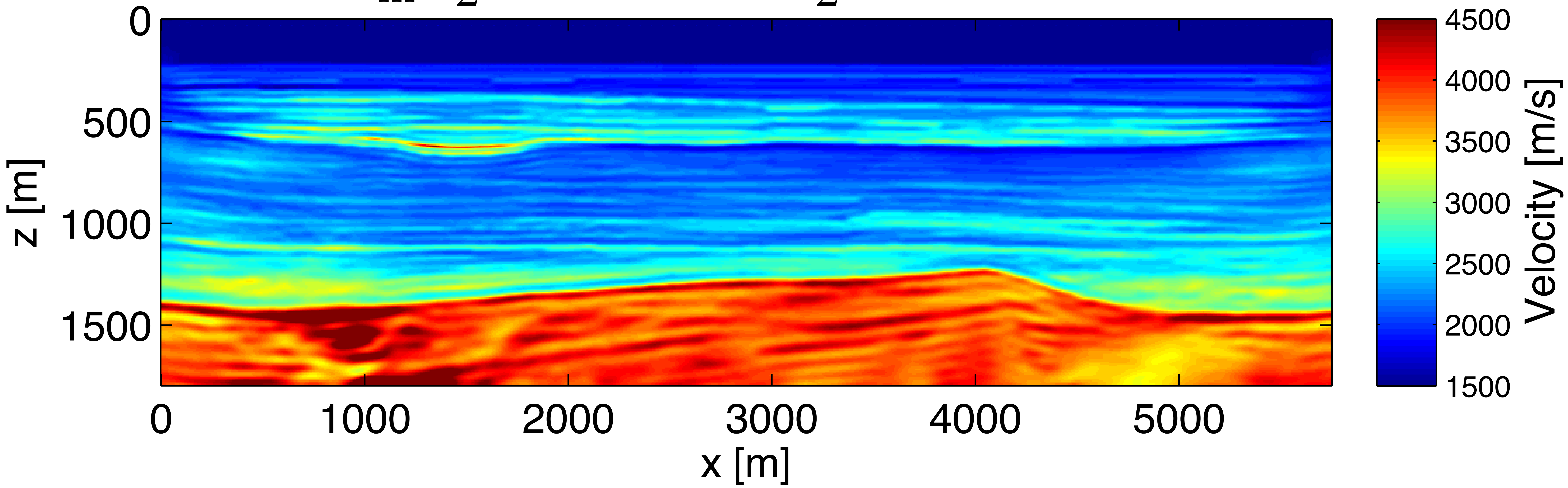
Example from [Peters et al. 2014]



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



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



Solution of the sub-problem

Main challenge: solve $\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$

- iteratively & matrix-free
- no QR or LU factorizations
- at cost cost of a few PDE solves

Solution of the sub-problem

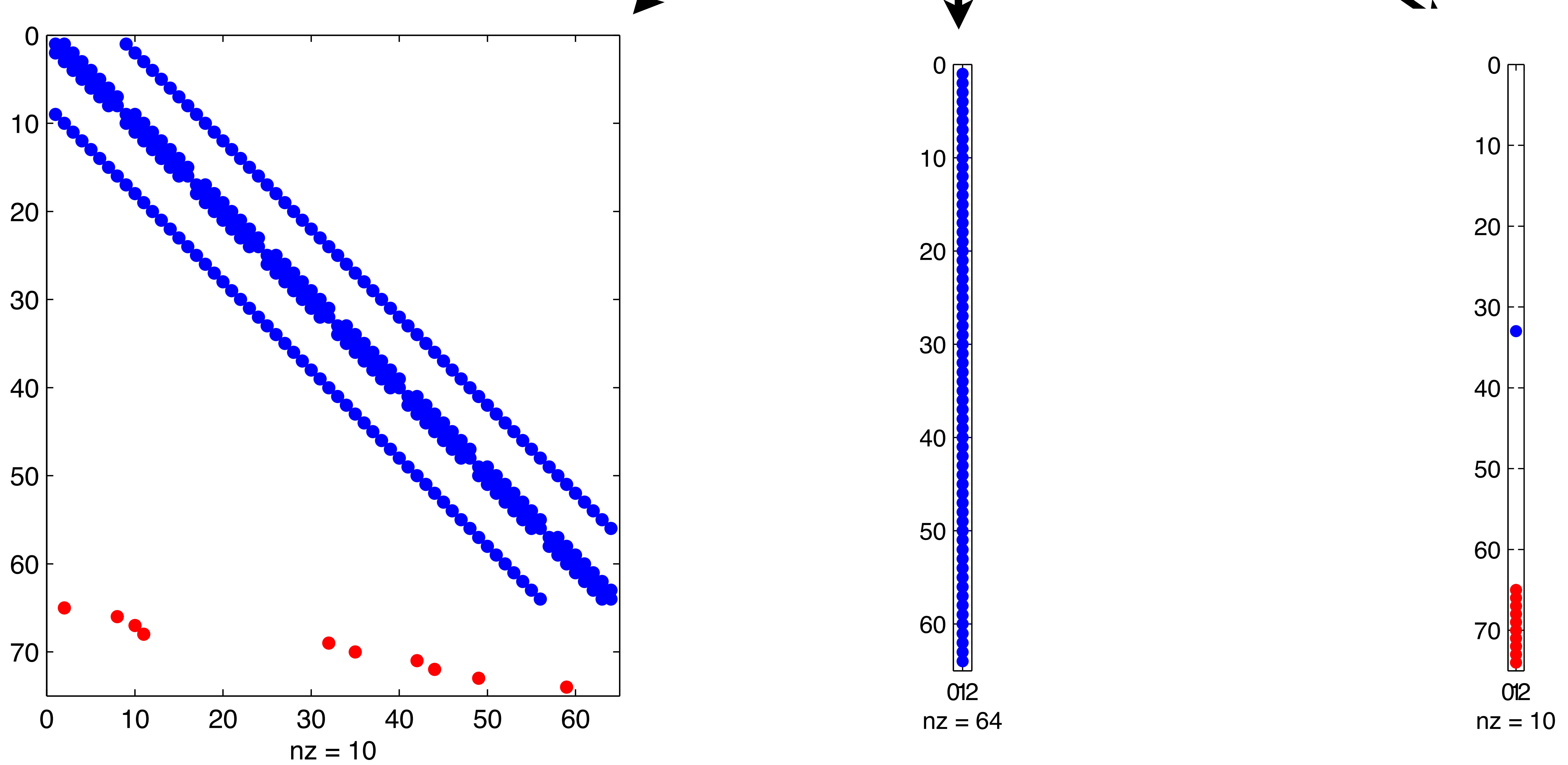
Properties of the sub-problem:

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

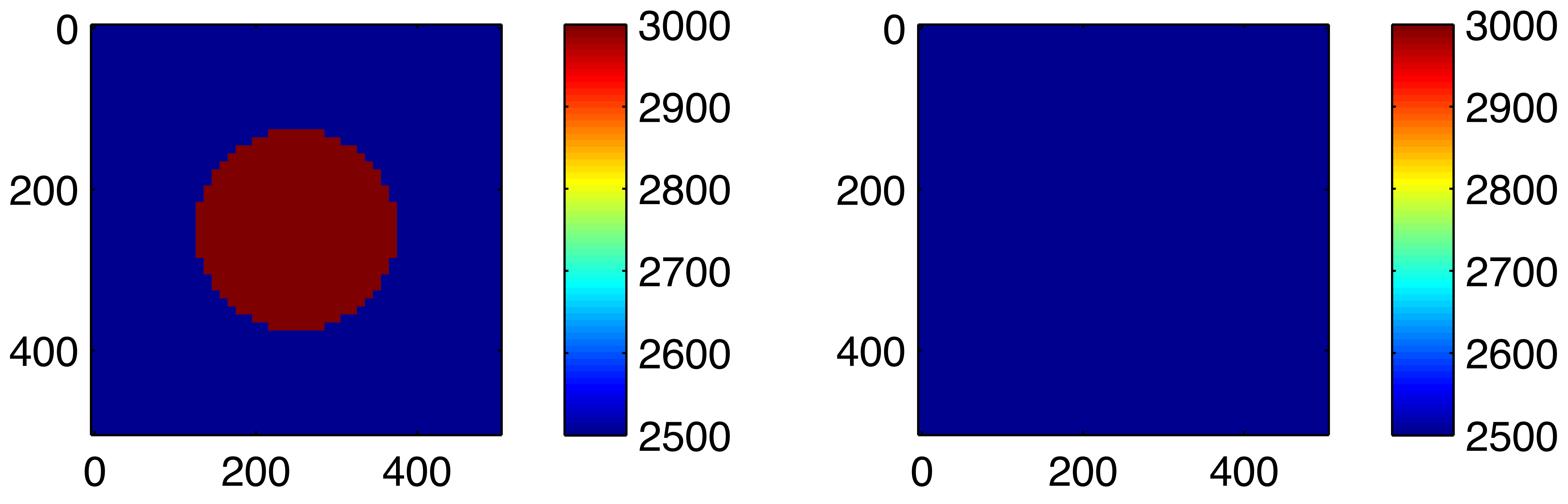
- H is indefinite, asymmetric & very large
- inconsistent
- full column rank
- very large condition number (squared) of the H^*H block

Solution of the sub-problem

$$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$

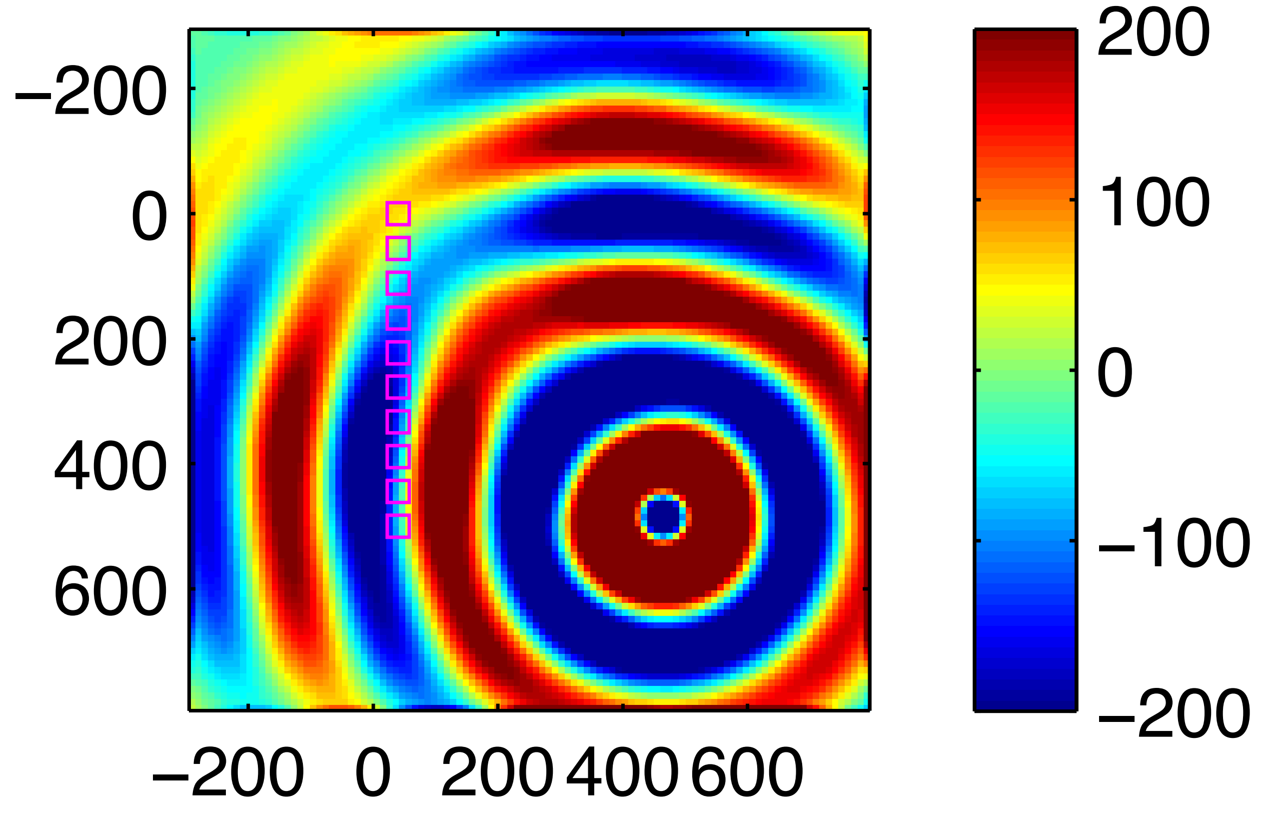


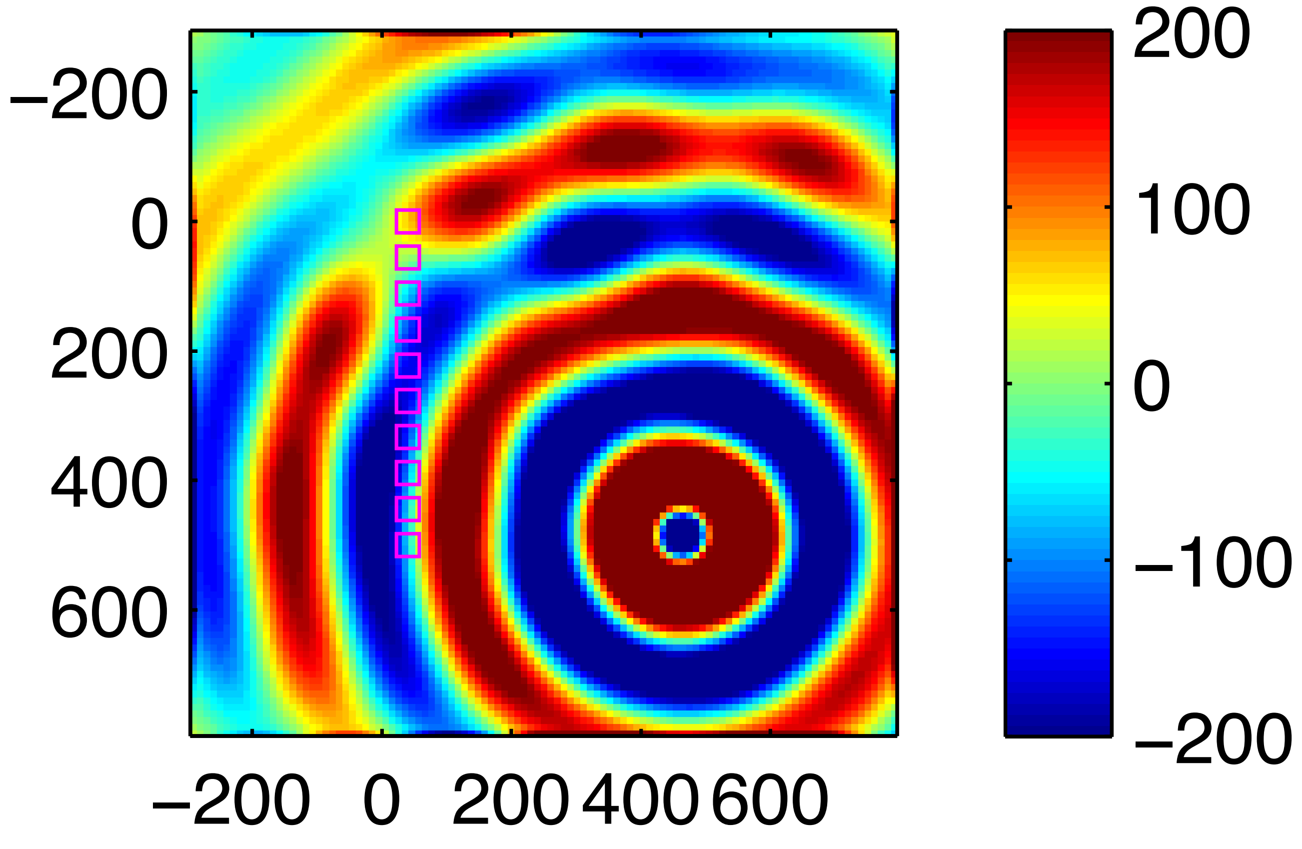
2D example

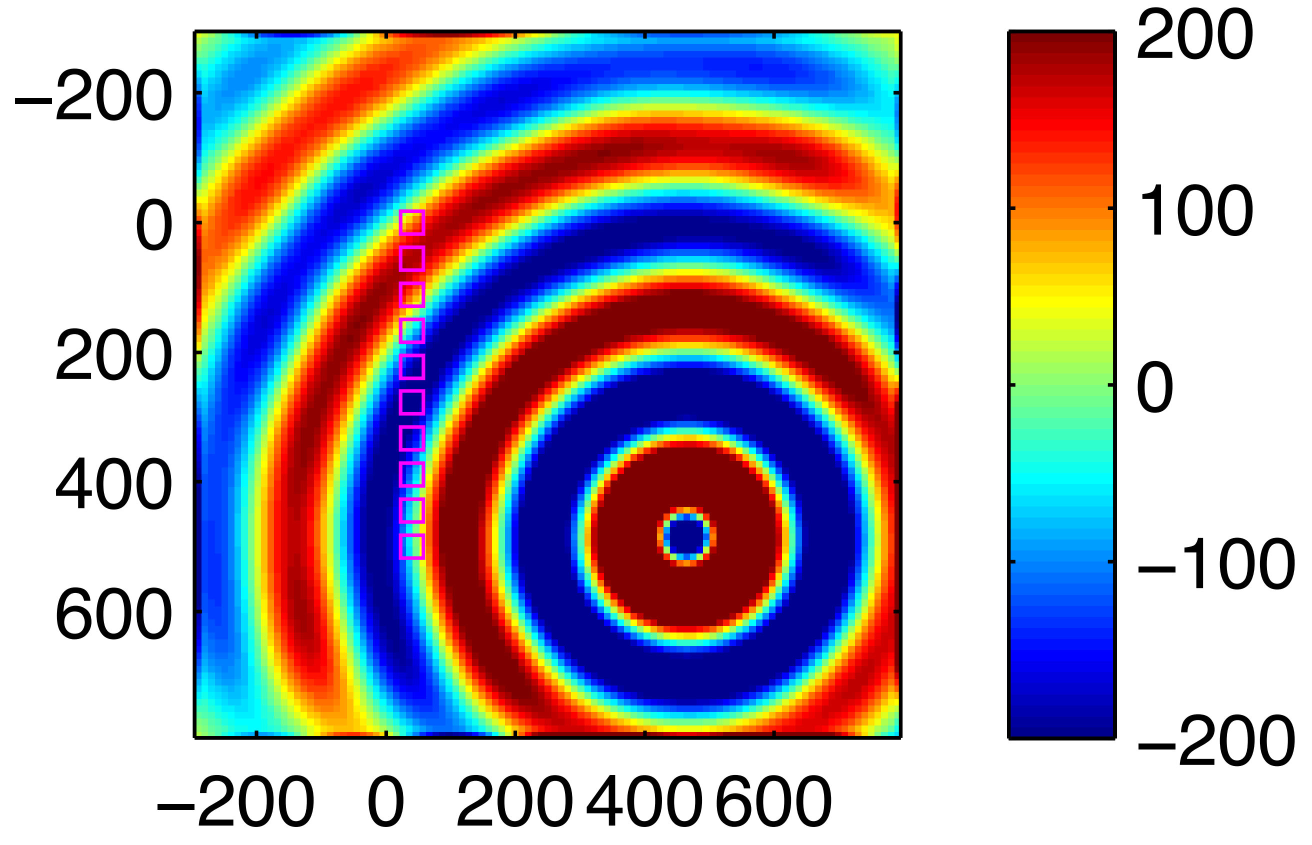


$$\mathbf{u} = H(\mathbf{m}_*)^{-1}\mathbf{q}$$

$$\mathbf{u} = H(\mathbf{m}_0)^{-1}\mathbf{q}$$







$$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}_0) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$

Proposed algorithm

LS-problem in normal-equation form:

$$\left(\lambda^2 H(\mathbf{m})^* H(\mathbf{m}) + P^* P\right)\bar{\mathbf{u}} = \lambda^2 H(\mathbf{m})\mathbf{q} + P^*\mathbf{d}$$

Split-preconditioning by λH w/o computations

$$(I + H_\lambda^{-*} P^* P H_\lambda^{-1})\mathbf{y} = \lambda\mathbf{q} + (H_\lambda^*)^{-1} P^*\mathbf{d}, \quad \text{with} \quad H_\lambda \bar{\mathbf{u}} = \mathbf{y}$$

- $m+1$ distinct eigenvalues (identity + low-rank)
- even for inexact Helmholtz

Proposed algorithm

Exploit identity + low-rank structure:

$$(I + H_\lambda^{-*} P^* P H_\lambda^{-1})\mathbf{y} = \lambda \mathbf{q} + (H_\lambda^*)^{-1} P^* \mathbf{d}, \quad \text{with} \quad H_\lambda \bar{\mathbf{u}} = \mathbf{y}$$

by solving $H^{-*}P^* = W$

- n_{rec} Helmholtz problems (inexactly)
- low-rank factorization

Proposed algorithm

Leverage low-rank factorization:

$$(I + WW^*)\mathbf{y} = \lambda\mathbf{q} + W\mathbf{d}, \quad \text{with} \quad H_\lambda \bar{\mathbf{u}} = \mathbf{y}$$

and invert system matrix as

$$\mathbf{y} = (I - W(I + W^*W)^{-1}W^*)(\lambda\mathbf{q} + W\mathbf{d}), \quad \text{with} \quad H_\lambda \bar{\mathbf{u}} = \mathbf{y}$$

so we only need to invert $(I + W^*W) \in \mathbb{C}^{m \times m}$

Proposed algorithm

for angular frequency ω **do**
  // solve m Helmholtz problems
  $H_\lambda^* W = P^*$
  $M = (I + W^* W)^{-1}$
  for right hand side i **do**
    $\mathbf{y}_i = \big(I - WMW^*\big)\big(\lambda \mathbf{q}_i + W\mathbf{d}_i\big)$
    // solve for $\bar{\mathbf{u}}_i$
    $H_\lambda \bar{\mathbf{u}}_i = \mathbf{y}_i$
  end for
end for

Proposed algorithm

Matrix-free algorithm

- no direct solves
- related mildly overdetermined systems **[L. M. Delves & I. Barrodale, 1979]**

Computational cost:

- 1 PDE per receiver
- 1 PDE per source

Memory requirements:

- 1 vector per receiver (W)
- system matrix (H)
- storage for solving systems with H

Proposed algorithm

Inexact solutions to the linear systems:

---

for angular frequency ω **do**

  // solve m Helmholtz problems inexactly

→ $H_\lambda^* \hat{W} = P^* + R_w$

  $M = (I + \hat{W}^* \hat{W})^{-1}$

  for right hand side i **do**

    $\mathbf{y}_i = \left(I - \hat{W} M \hat{W}^*\right)\left(\lambda \mathbf{q}_i + \hat{W} \mathbf{d}_i\right)$

    // solve for $\bar{\mathbf{u}}_i$ inexactly

→ $H_\lambda \hat{\mathbf{u}}_i = \mathbf{y}_i + r_u$

  end for

end for

Proposed algorithm

(preliminary) error bound on inexactly computed solution:

$$\frac{\|\hat{\mathbf{u}} - \bar{\mathbf{u}}\|}{\|\bar{\mathbf{u}}\|} \leq \kappa(H) \frac{\|(I + \mathbf{w}\mathbf{w}^*)^{-1}(H_\lambda^{-*}\mathbf{r}_w \mathbf{d} + \mathbf{r}_y - (\mathbf{w}(H_\lambda^{-*}\mathbf{r}_w)^* + (H_\lambda^{-*}\mathbf{r}_w)\mathbf{w}^*)\hat{\mathbf{y}}) + \mathbf{r}_u\|}{\|\mathbf{y}\|}$$

Proposed algorithm

(preliminary) error bound on inexactly computed solution:

$$\frac{\|\hat{\mathbf{u}} - \bar{\mathbf{u}}\|}{\|\bar{\mathbf{u}}\|} \leq \kappa(H) \frac{\|(I + \mathbf{w}\mathbf{w}^*)^{-1}(H_\lambda^{-*}\mathbf{r}_w\mathbf{d} + \mathbf{r}_y - (\mathbf{w}(H_\lambda^{-*}\mathbf{r}_w)^* + (H_\lambda^{-*}\mathbf{r}_w)\mathbf{w}^*)\hat{\mathbf{y}}) + \mathbf{r}_u\|}{\|\mathbf{y}\|}$$

↑

H instead of H^*H

Proposed algorithm

(preliminary) error bound on inexactly computed solution:

residual of solving a system with H

$$\frac{\|\hat{\mathbf{u}}-\bar{\mathbf{u}}\|}{\|\bar{\mathbf{u}}\|} \leq \kappa(H) \frac{\|(I+\mathbf{w}\mathbf{w}^*)^{-1}(H_\lambda^{-*}\mathbf{r}_w\mathbf{d}+\mathbf{r}_y-(\mathbf{w}(H_\lambda^{-*}\mathbf{r}_w)^*+(H_\lambda^{-*}\mathbf{r}_w)\mathbf{w}^*)\hat{\mathbf{y}})+\mathbf{r}_u\|}{\|\mathbf{y}\|}$$

Suggested PDE-solver

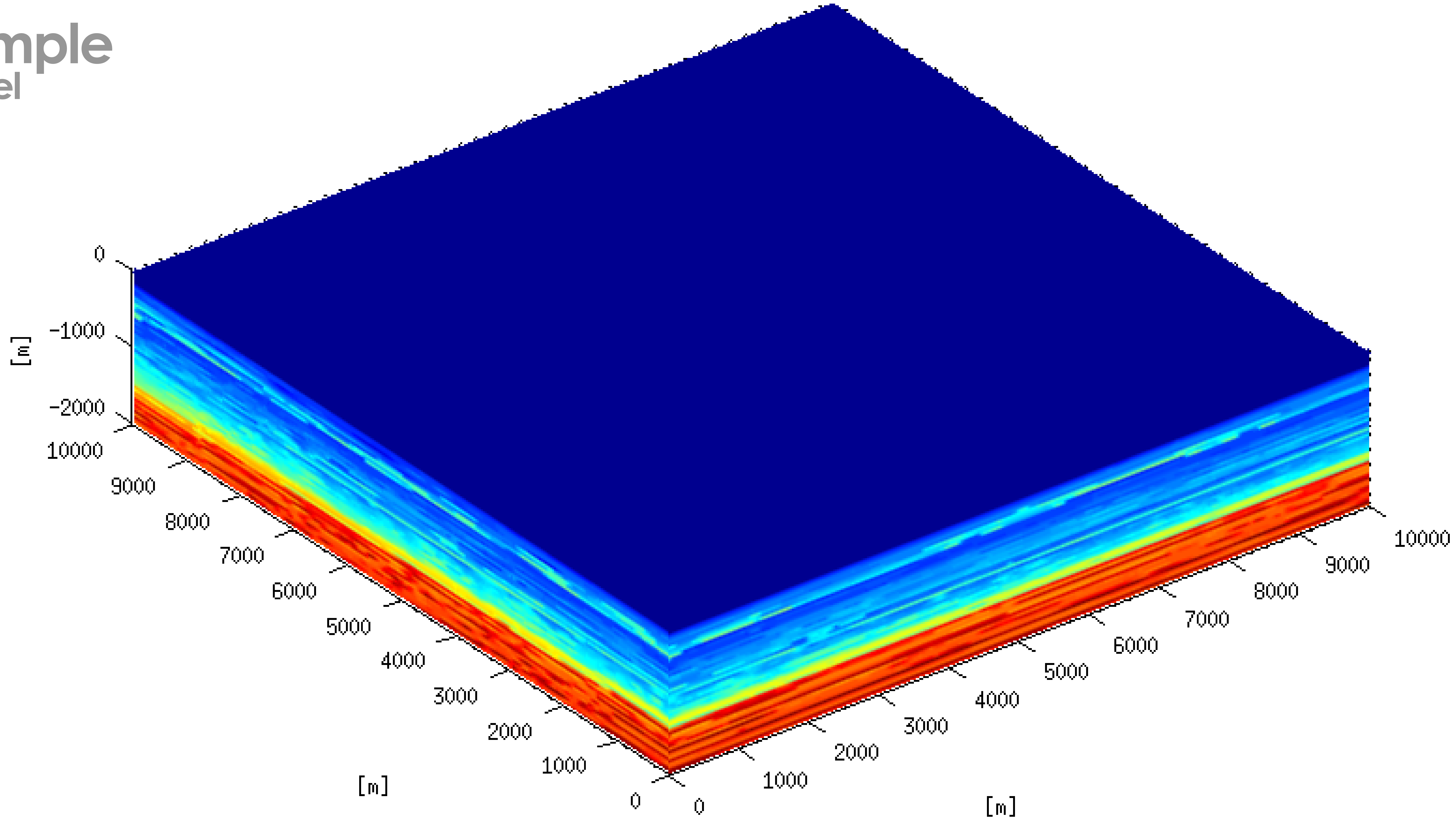
Need to store 1 vector per receiver

-> use PDE-solver with low-memory & setup requirements

Helmholtz:

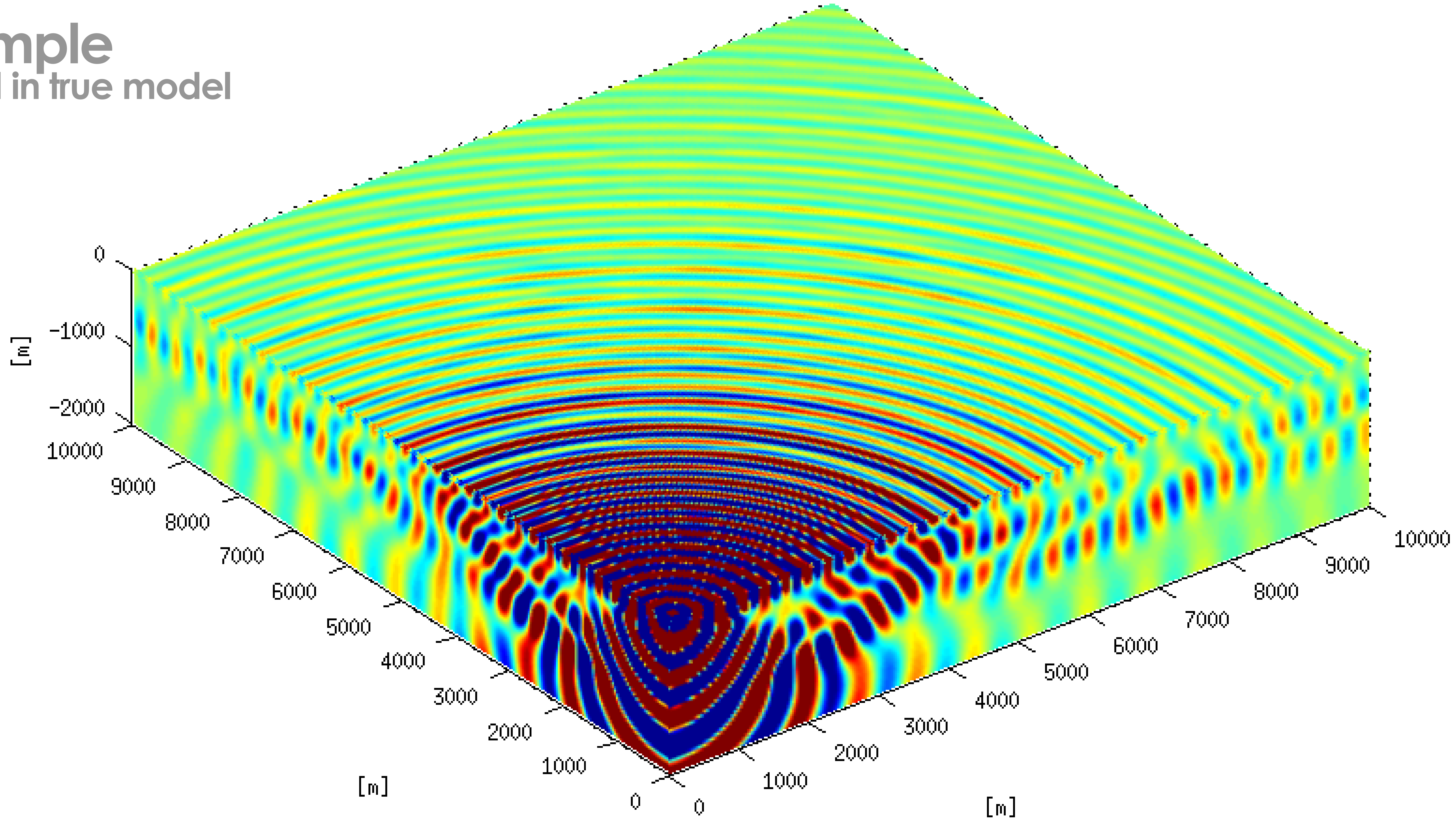
- CGMN (only 4 vectors) / CARP-CG [A. Bjorck & T. Elfving, 1979; D. Gordon & R. Gordon, 2010; T. van Leeuwen & F.J. Herrmann, 2014]
- Shifted-Laplacian w/ multi-grid [Y.A. Erlangga, 2008; H. Calandra et al., 2013]
- combination of the above [R. Lago & F.J. Herrmann, 2015]

3D Example
– true model

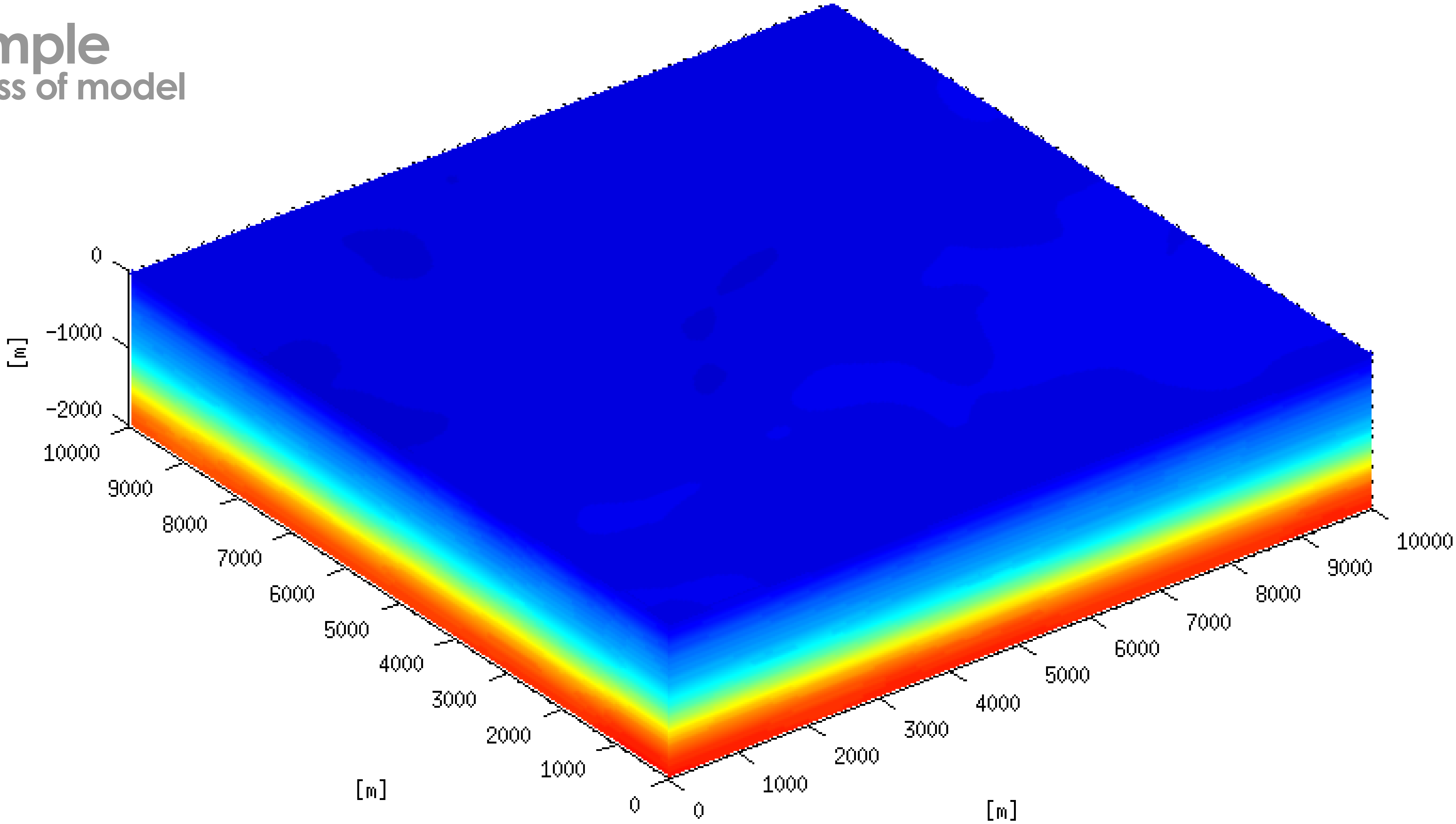


10 x 10 x 2 km, 5 Hz, 27-point discretization, ~1e7 grid points, source at [0,0,0]

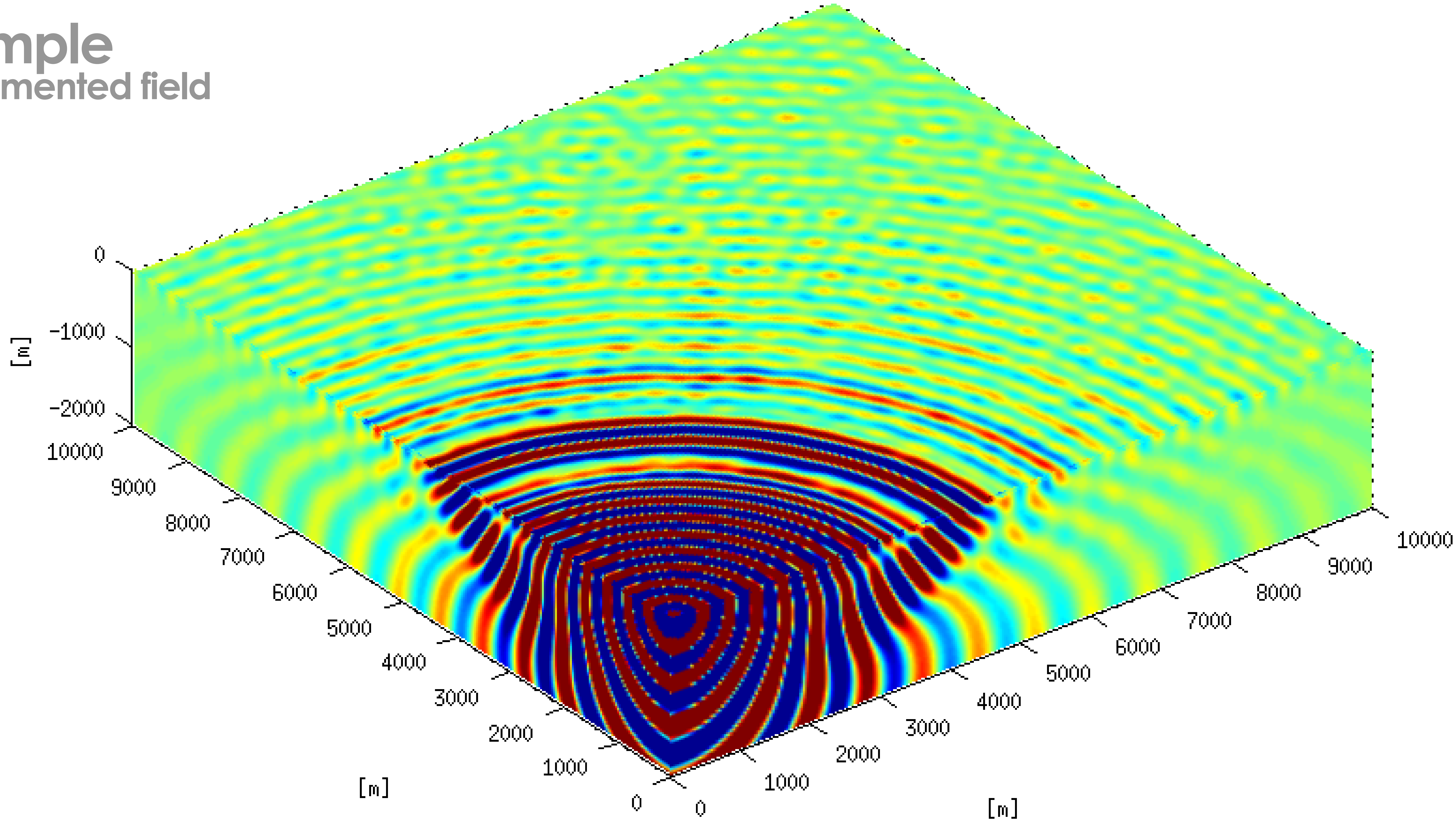
3D Example
– wavefield in true model



3D Example
– initial guess of model



3D Example
– data-augmented field



Conclusions

- Developed matrix-free version of a reduced-space quadratic-penalty method.
- Enabler for 3D parameter estimation w/ the quadratic-penalty method.
- Proposed algorithm might be used for other large-scale mildly overdetermined problems w/ many variables & few constraints.

Current & future work

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

$$\begin{pmatrix} \nabla^2_{\mathbf{u},\mathbf{u}}\phi & \nabla^2_{\mathbf{u},\mathbf{m}}\phi \\ \nabla^2_{\mathbf{m},\mathbf{u}}\phi & \nabla^2_{\mathbf{m},\mathbf{m}}\phi \end{pmatrix} \begin{pmatrix} \delta\mathbf{u} \\ \delta\mathbf{m} \end{pmatrix} = -\begin{pmatrix} \nabla_{\mathbf{u}}\phi \\ \nabla_{\mathbf{m}}\phi \end{pmatrix}$$

$$\lambda^2 H^* H + P^* P$$

Developed algorithm is also a key building block for a full-space algorithm

Penalty approach avoids storing multipliers

Acknowledgements

Tristan van Leeuwen, Art Petrenko & Rafael Lago for the CGMN & CARP-CG implementation

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, Sub Salt Solutions, WesternGeco, and Woodside.

References

A. Bjorck and T. Elfving, Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations, BIT, 19 (1979), pp. 145–163.

1. D. Gordon and R. Gordon, CARP-CG: A robust and efficient parallel solver for linear systems, applied to strongly convection dominated pdes, Parallel Computing, 36 (2010), pp. 495– 515.
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.

References (2)

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.

10. Calandra, H., Gratton, S., Pinel, X. and Vasseur, X. [2013] An improved two-grid preconditioner for the solution of three-dimensional Helmholtz problems in heterogeneous media. Numerical Linear Algebra with Applications.

11. Erlangga, Y.A. [2008] Advances in iterative methods and preconditioners for the Helmholtz equation. Archives of Computational Methods in Engineering, 15, 37–66.

12. 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.

13. Rafael Lago, Felix J. Herrmann. Towards a robust geometric multigrid scheme for {Helmholtz} equation, Tech Report, UBC, 2015.