An algorithm for solving least-squares problems with a Helmholtz block and multiple right-hand-sides

Bas Peters, Chen Greif, Felix J. Herrmann

International Conference On Preconditioning Techniques For Scientific And Industrial Applications
June 18, 2015

SLIM
University of British Columbia

Problem of interest

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

Originates from the 'discretize-then-optimize' framework for PDE-constrained optimization:

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

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

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

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

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

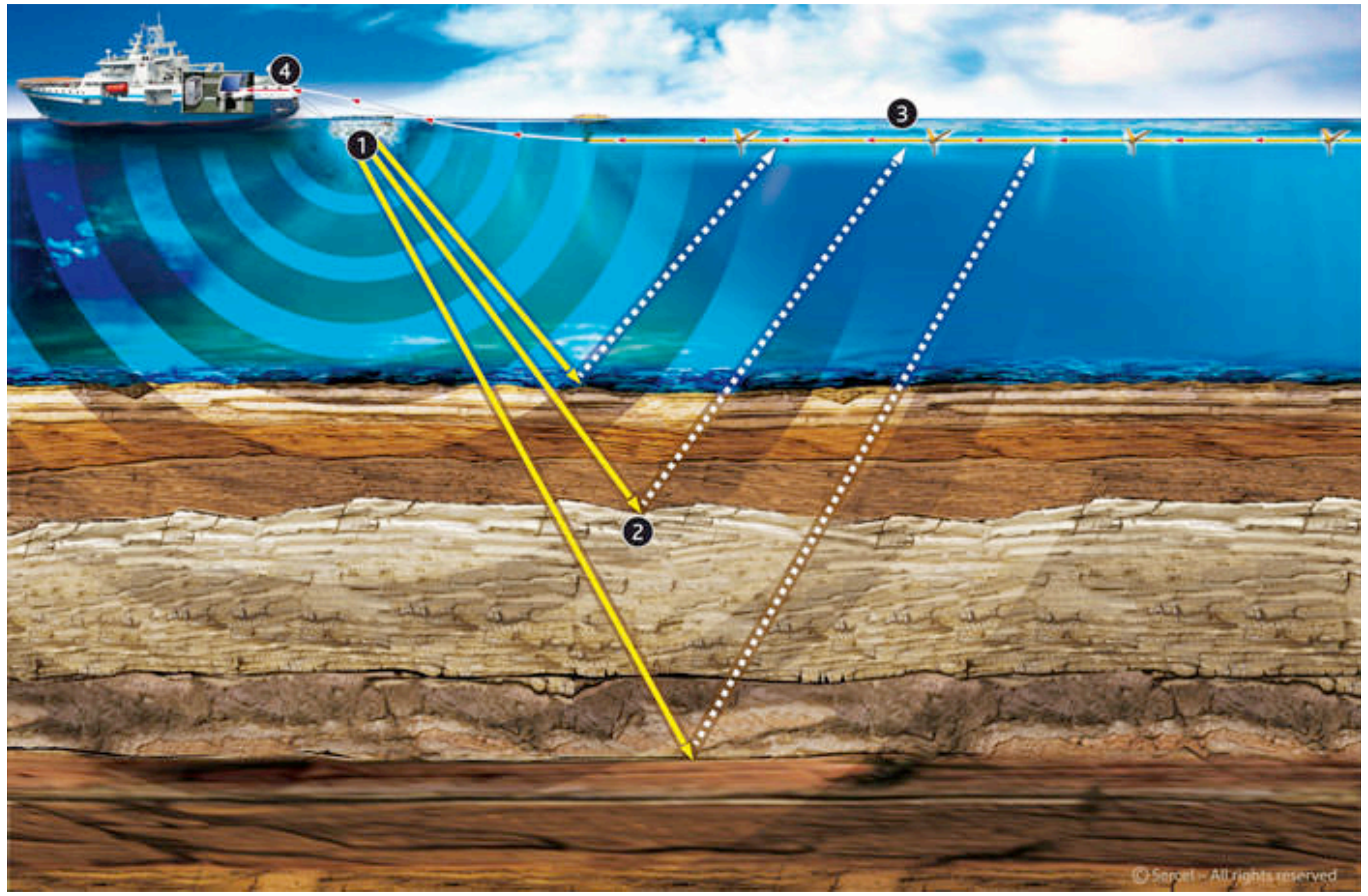
$\mathbf{u} \in \mathbb{C}^N$ field

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

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

PDE-constrained optimization

The PDE of interest in this talk is the scalar Helmholtz equation



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

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

- 1 PDE: *N* ~ [1e6 - 1e9] grid points
- [1 - 100] right-hand-sides (*k* sources)
- [1 - 100] *m* receivers ($P \in \mathbb{R}^{m\times N}$)

PDE-constrained optimization

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

↓

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

↓

eliminate field variables

↓

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

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

reduced gradient method / adjoint-state / reduced Lagrangian

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

↓

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

↓

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

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

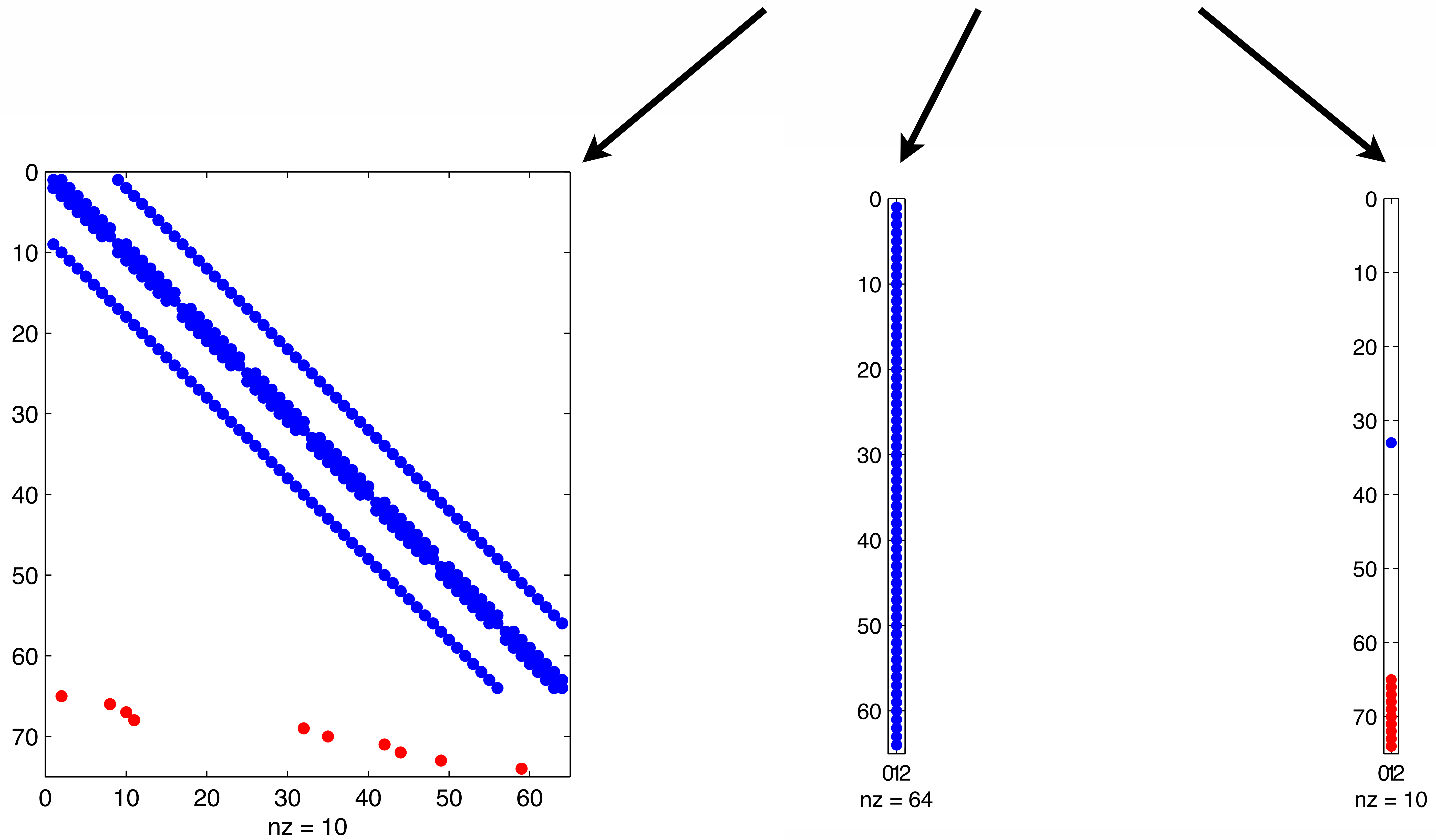
↓

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

reduced quadratic-penalty

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

$$\nabla_{\mathbf{u}}\phi(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = 0 \longleftrightarrow \bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$



[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 outer iteration:

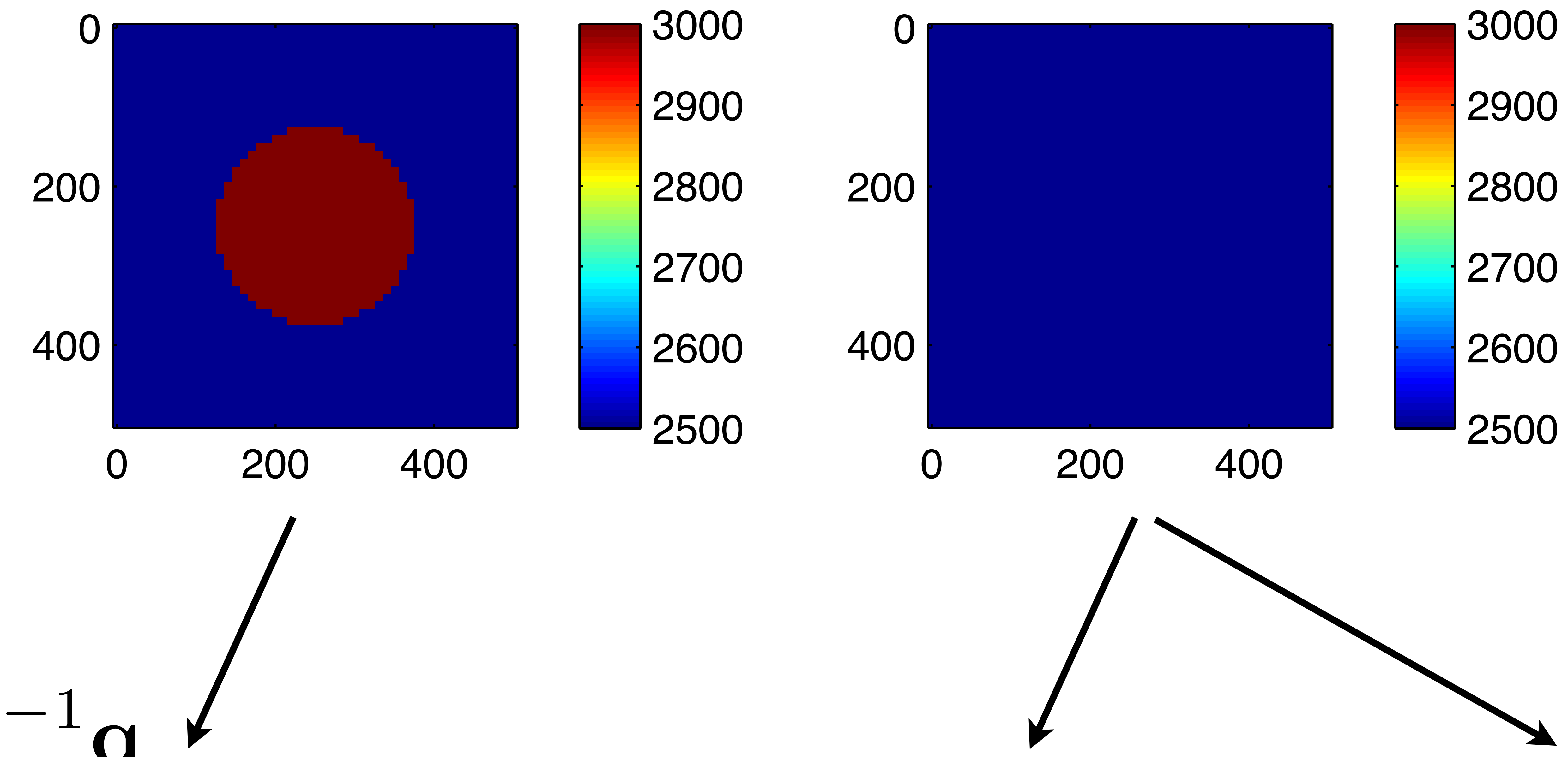
- compute $$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$
- evaluate $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$ & $\nabla_{\mathbf{m}}\bar{\phi}(\mathbf{m},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

Properties of the problem

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

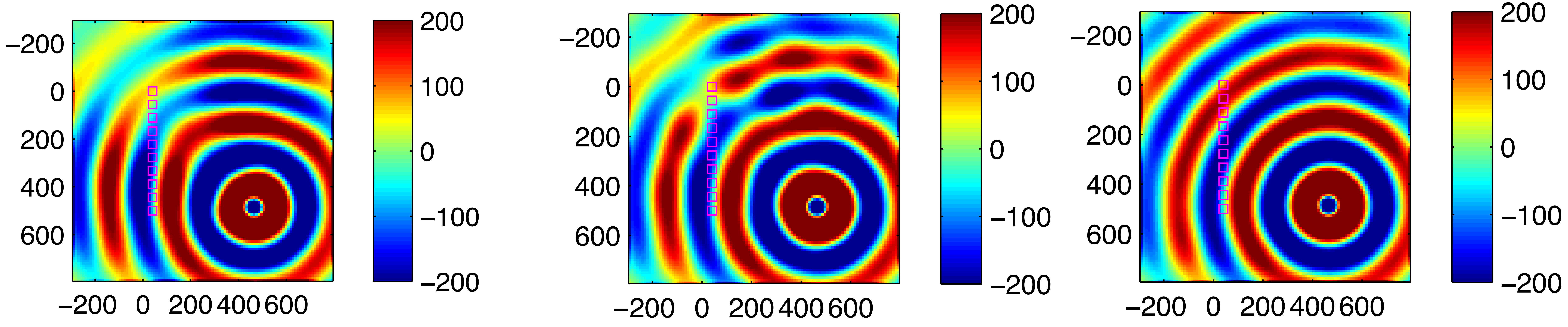
- H is indefinite, not Hermitian
- inconsistent
- full column rank
- may lose symmetry when using a Perfectly Matched Layer (PML)

2D example



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

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



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

Algorithms

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

Algorithms

What about preconditioned LSQR, CGLS? (preconditioner: λH)

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

using exact preconditioning this solves

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

- $m+1$ distinct eigenvalues (identity + low-rank) $\qquad (m = n_{\text{rec}})$

Algorithms

What about preconditioned LSQR, CGLS? (preconditioner: λH)

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

- expected computational cost: $n_{\text{src}} \times 2(1 + n_{\text{rec}})$ PDE solves
- not competitive with Lagrangian based reduced-space algorithms which require $2n_{\text{src}}$ PDE solves
- more PDE solves required in case of inexact PDE solves

SLIM

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 without 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 PDE solves
- Exploit identity + low-rank structure by solving $H^{-*}P^* = W$

Proposed algorithm

identity + 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 (Sherman-Morrison)

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

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

(this is alway small enough to do explicitly, $m \leq 100$)

Proposed algorithm

identity + low-rank factorization:

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

Stability of Sherman-Morrison is a concern in general, but was found to be sufficiently accurate for some Helmholtz test problems.

In case Sherman-Morrison is not accurate enough:

$$\arg\min_{\mathbf{y}} \left\| \begin{pmatrix} I \\ W^* \end{pmatrix} \mathbf{y} - \begin{pmatrix} \lambda\mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$

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**

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

    // 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
→ $H_\lambda^* \hat{W} = P^* + R_W$
  $\hat{M} = (I + \hat{W}^* \hat{W})^{-1}$
  for right hand side $\mathbf{b}_i$ **do**
    $\hat{\mathbf{y}}_i = \left(I - \hat{W}\hat{M}\hat{W}^*\right)\left(\lambda \mathbf{q}_i + \hat{W}\mathbf{d}_i\right)$
    // solve for $\bar{\mathbf{u}}_i$ inexactly
→ $H_\lambda \hat{\mathbf{u}}_i = \hat{\mathbf{y}}_i + \mathbf{r}_\mathbf{u}$
  end for
end for

Proposed algorithm

error propagation (1 right-hand-side, 1 receiver case):

$$H_\lambda^* \hat{\mathbf{w}} = \mathbf{p}^* + \mathbf{r}_{\mathbf{w}}$$

$$(I + \hat{\mathbf{w}}\hat{\mathbf{w}}^*)\hat{\mathbf{y}} = \lambda \mathbf{q} + \hat{\mathbf{w}} d$$

$$H_\lambda \hat{\mathbf{u}} = \hat{\mathbf{y}} + \mathbf{r}_{\mathbf{u}}$$

Proposed algorithm

error propagation (1 right-hand-side, 1 receiver case):

$$H_\lambda^* \hat{\mathbf{w}} = \mathbf{p}^* + \mathbf{r}_\mathbf{w}$$

$$(I + \hat{\mathbf{w}}\hat{\mathbf{w}}^*)\hat{\mathbf{y}} = \lambda \mathbf{q} + \hat{\mathbf{w}} d \quad \longrightarrow \quad \text{solve as: } \hat{\mathbf{y}} = (I - \hat{m}\hat{\mathbf{w}}\hat{\mathbf{w}}^*)(\lambda \mathbf{q} + \hat{\mathbf{w}} d)$$

with $\hat{m} = \dfrac{1}{1 + \hat{\mathbf{w}}^* \hat{\mathbf{w}}}$

$$H_\lambda \hat{\mathbf{u}} = \hat{\mathbf{y}} + \mathbf{r}_\mathbf{u}$$

derivation of error bounds based on observable quantities is work in progress

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]

Randomization and subsampling

What is the number of receivers is too large, storage wise?

Can we approximate the least-squares problem using randomization & subsampling?

Use ideas from algorithms such as

- **[V Rokhlin & M Tygert, 2008]**
- Blendenpik **[H. Avron et. al., 2010]**
- LSRN **[X. Meng, M. A. Saunders, M. W. Mahoney, 2014]**

Randomization and subsampling

Initial attempt in this work:

apply randomization and subsampling to the receiver block only for a one-step approximation:

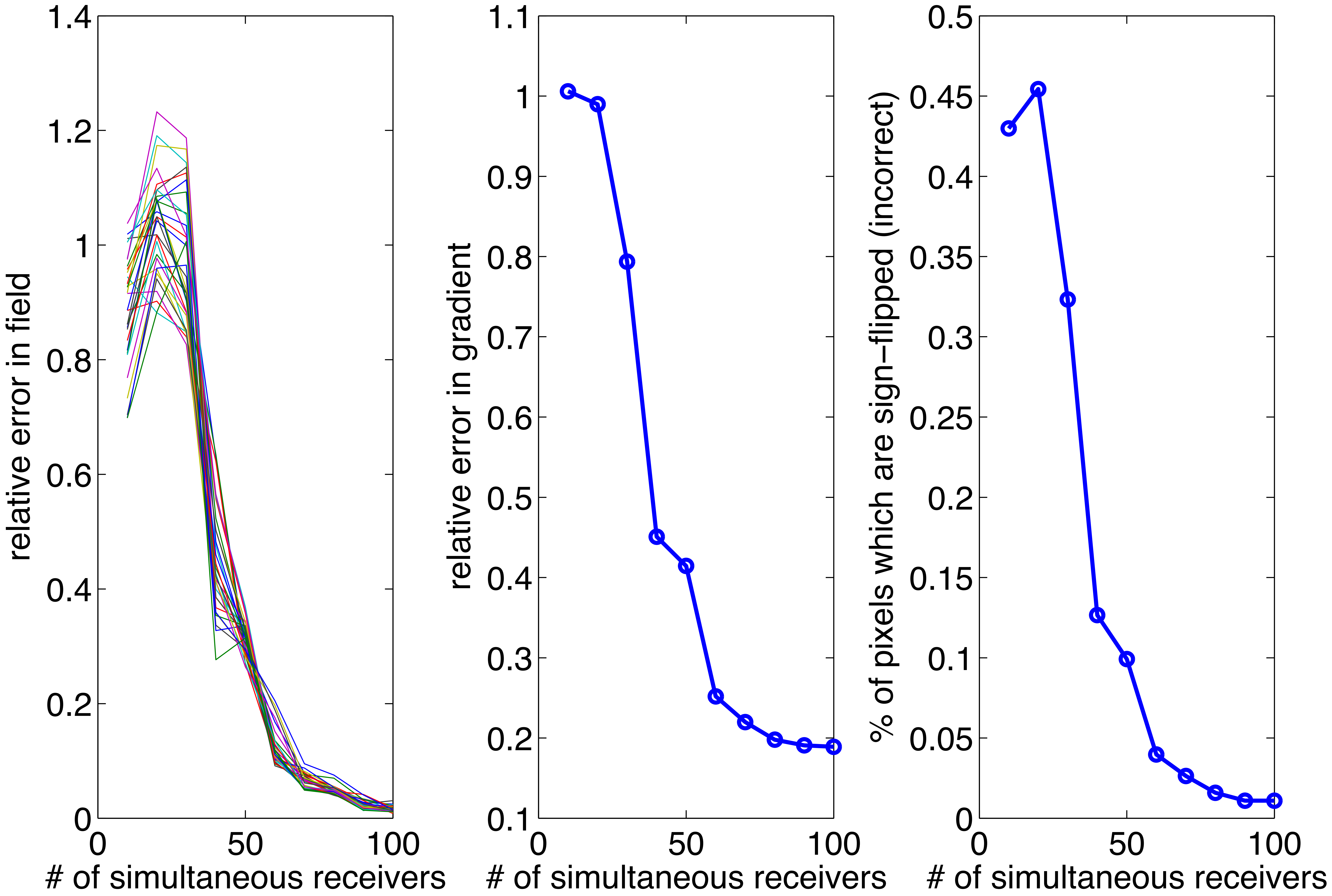
$$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ VP \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ V\mathbf{d} \end{pmatrix} \right\|_2$$

$V \in \mathbb{C}^{l \times m}, \quad l < m$ (complex, random, flat)

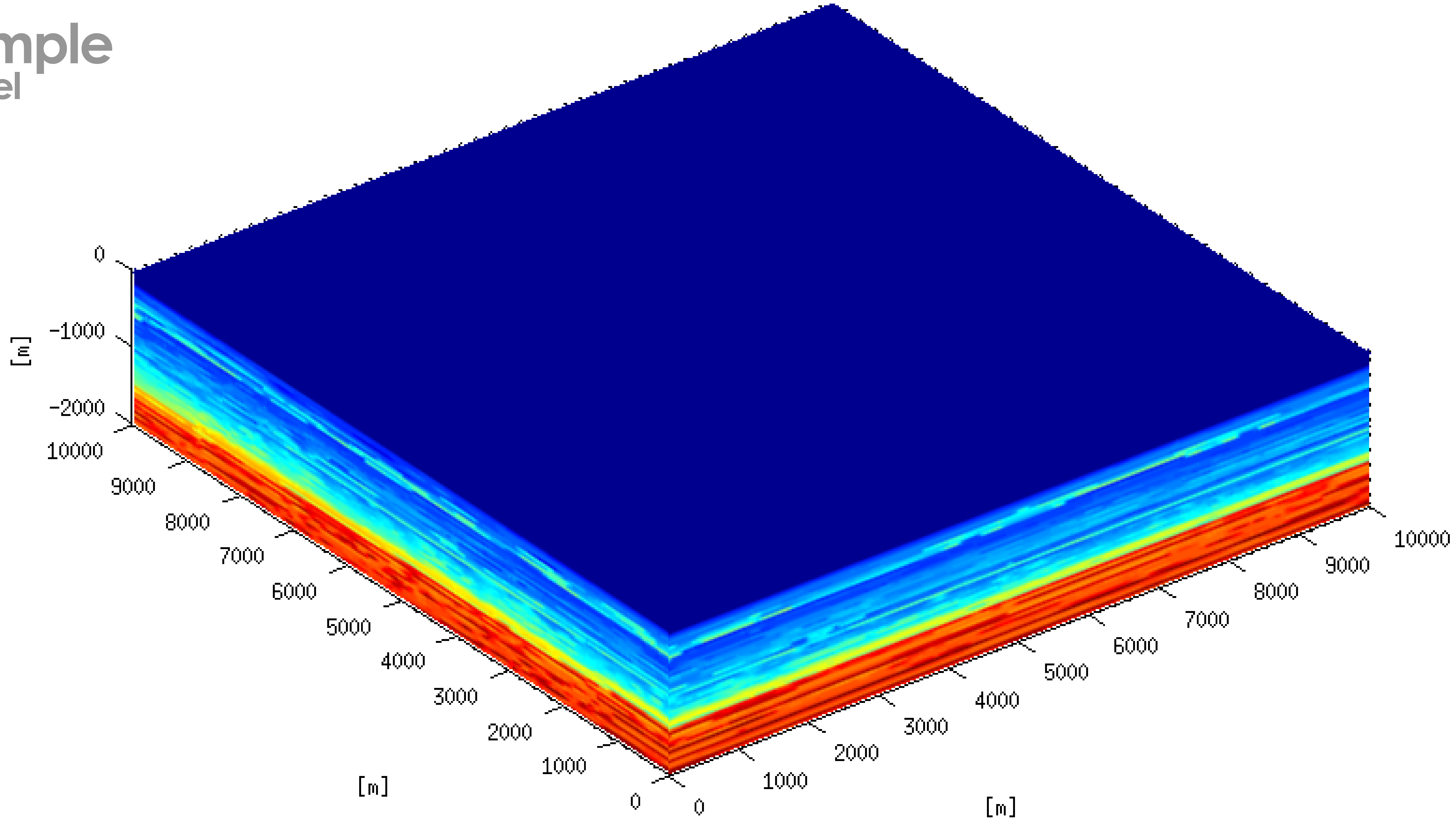
reduces

- # of PDE solves
- # vectors to be stored

Simultaneous receivers

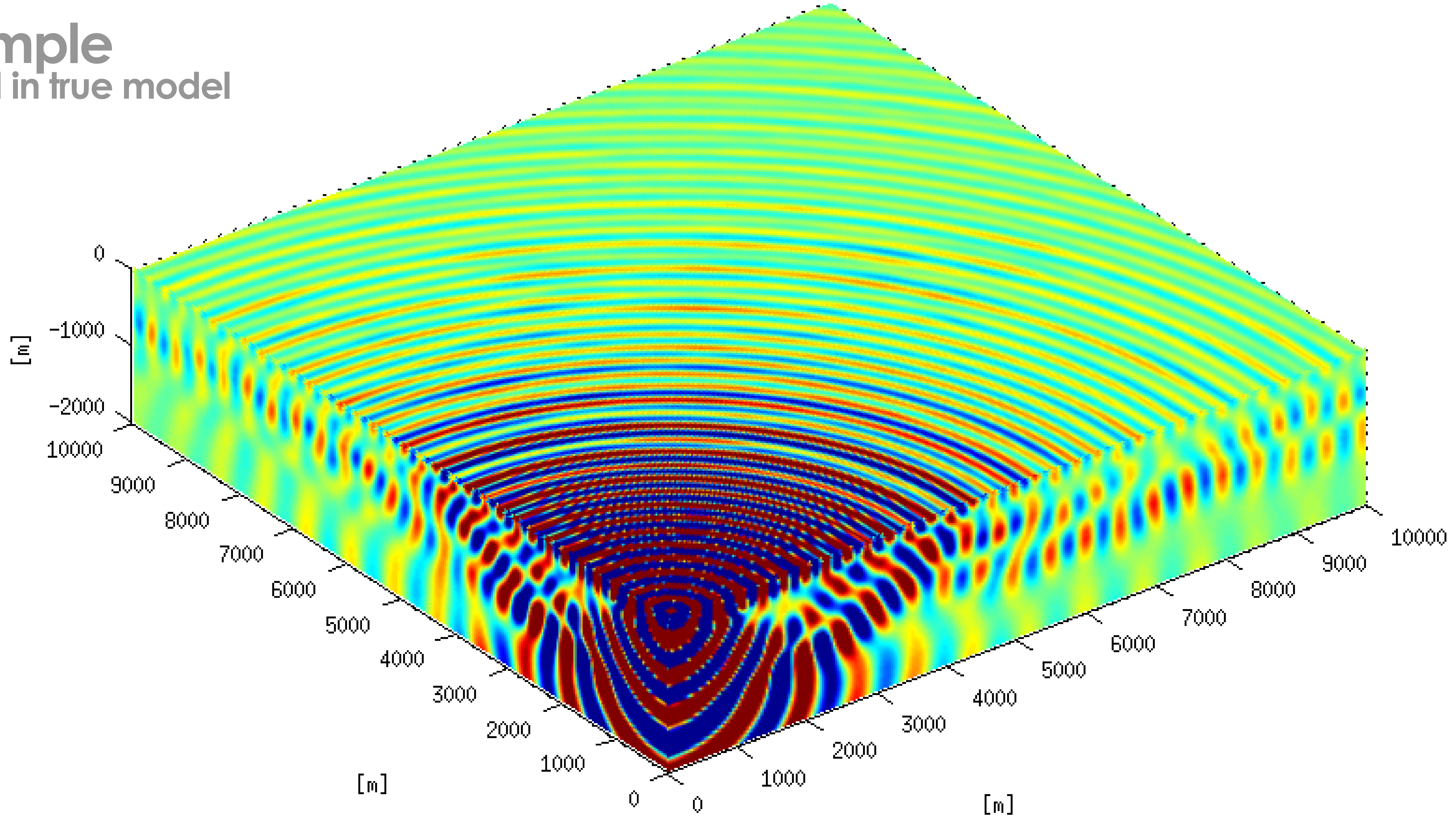


3D Example
– true model

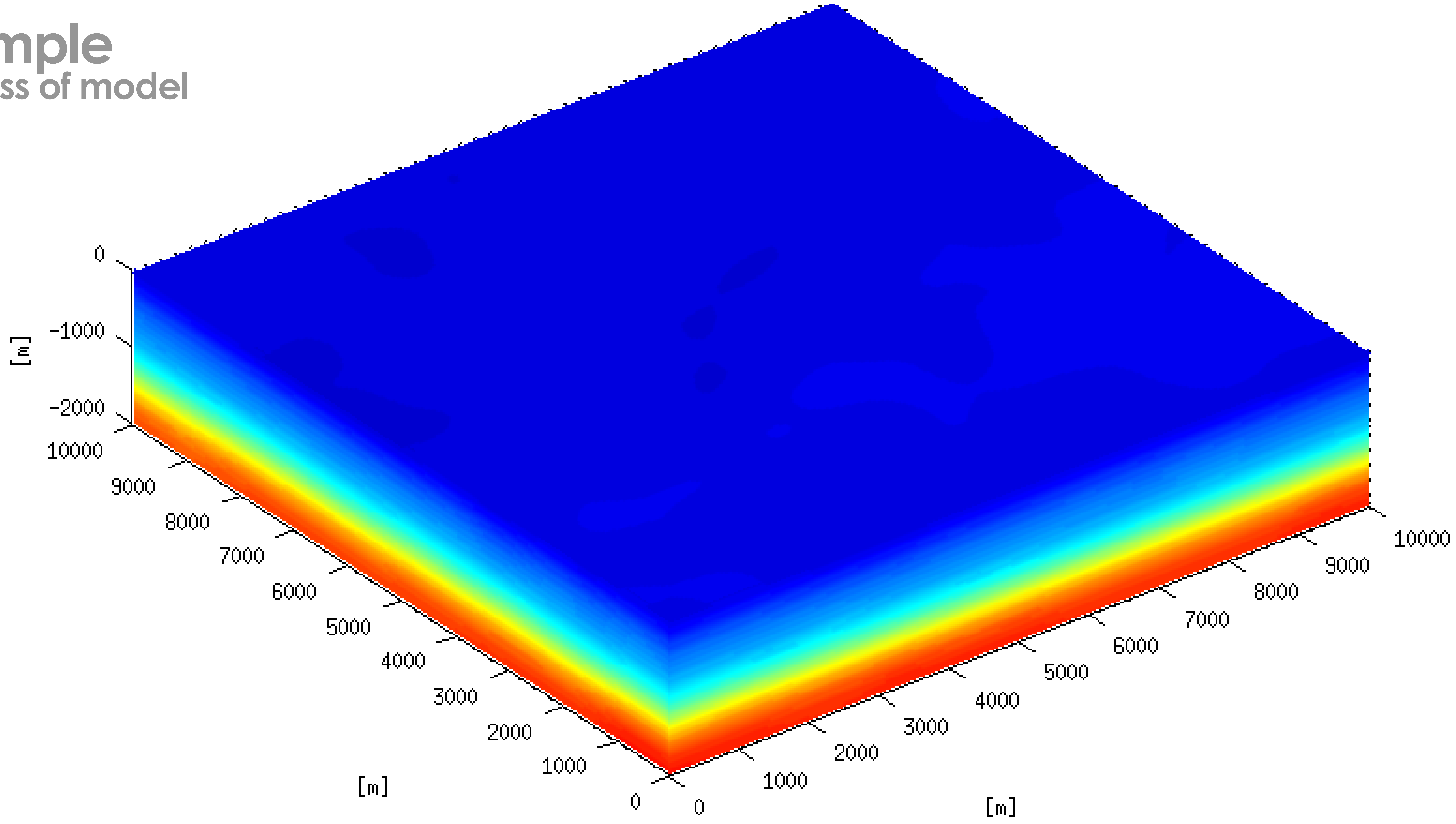


10 x 10 x 2 km, 5 Hz, 27-point discretization, ~1e7 grid points, source at [0,0,0]

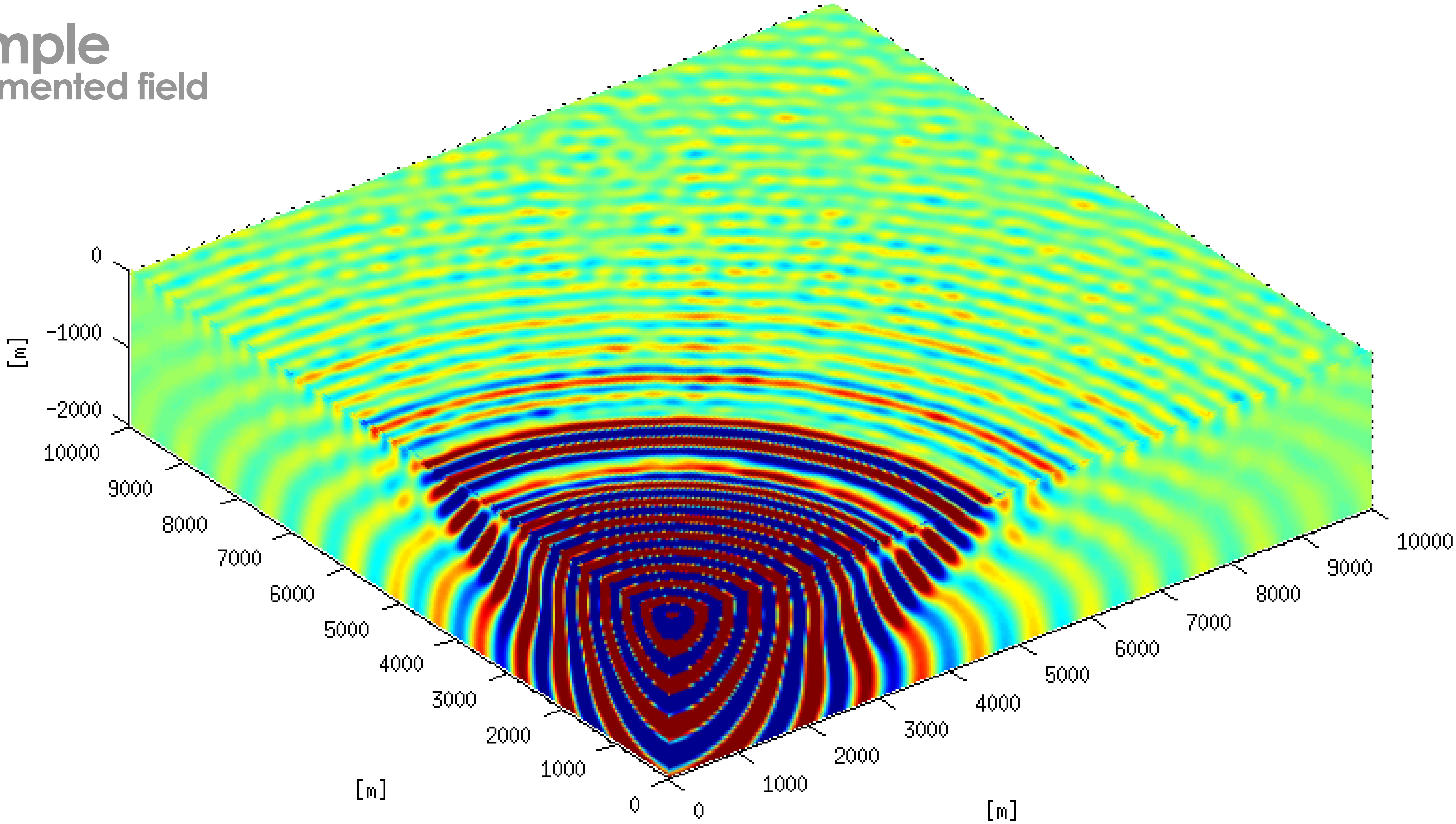
3D Example
– wavefield in true model



3D Example
– initial guess of model



3D Example
– data-augmented field



Conclusions

- Enabler for 3D parameter estimation using a quadratic-penalty method.

- There is potential for randomization and subsampling to reduce the computational cost and memory requirements.

- Proposed algorithm might be used for other large-scale mildly overdetermined problems with many variables & few constraints.

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

1. A. Bjorck and T. Elfving, Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations, BIT, 19 (1979), pp. 145–163.
2. 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.
3. Tristan van Leeuwen and Felix J. Herrmann, frequency-domain seismic inversion with controlled sloppiness, SIAM Journal on Scientific Computing, 36 (2014), pp. S192–S217.
4. 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.
5. Eldad Haber, Uri M Ascher, and Doug Oldenburg, On optimization techniques for solving nonlinear inverse problems, Inverse Problems, 16 (2000), pp. 1263–1280.
6. E Haber and U M Ascher, Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Problems, 17 (2001), p. 1847.
7. 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.
8. 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.
9. 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)

10. 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.

11. 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.

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

13. 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.

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