Iterative solution strategy for least-squares problems with a PDE-block

Bas Peters
SINBAD Fall Consortium meeting, 2014.
Whistler, BC.

SLIM
University of British Columbia

Goal

Develop a large-scale iterative solver for the least-squares problem in Wavefield Reconstruction Inversion (WRI):

- solve for sources & frequencies in parallel
- low-memory requirements
- no incomplete factorizations
- use existing optimized Helmholtz solvers

WRI

Constrained formulation:

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

Unconstrained quadratic penalty formulation:

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

WRI

Unconstrained quadratic penalty formulation:

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

Gradient w.r.t. the field:

$$\nabla_{\mathbf{u}}\phi_\lambda(\mathbf{b}, \boldsymbol{\kappa}, \mathbf{u}) = P^*(P\mathbf{u} - \mathbf{d}) + \lambda^2 H^*(H(\mathbf{b}, \boldsymbol{\kappa})\mathbf{u} - \mathbf{q})$$

Wavefield reconstruction inversion

Set field gradient to 0 – i.e, project out the fields: $\nabla_{\mathbf{u}}\phi_\lambda(\mathbf{b},\boldsymbol{\kappa},\mathbf{u}) = 0$

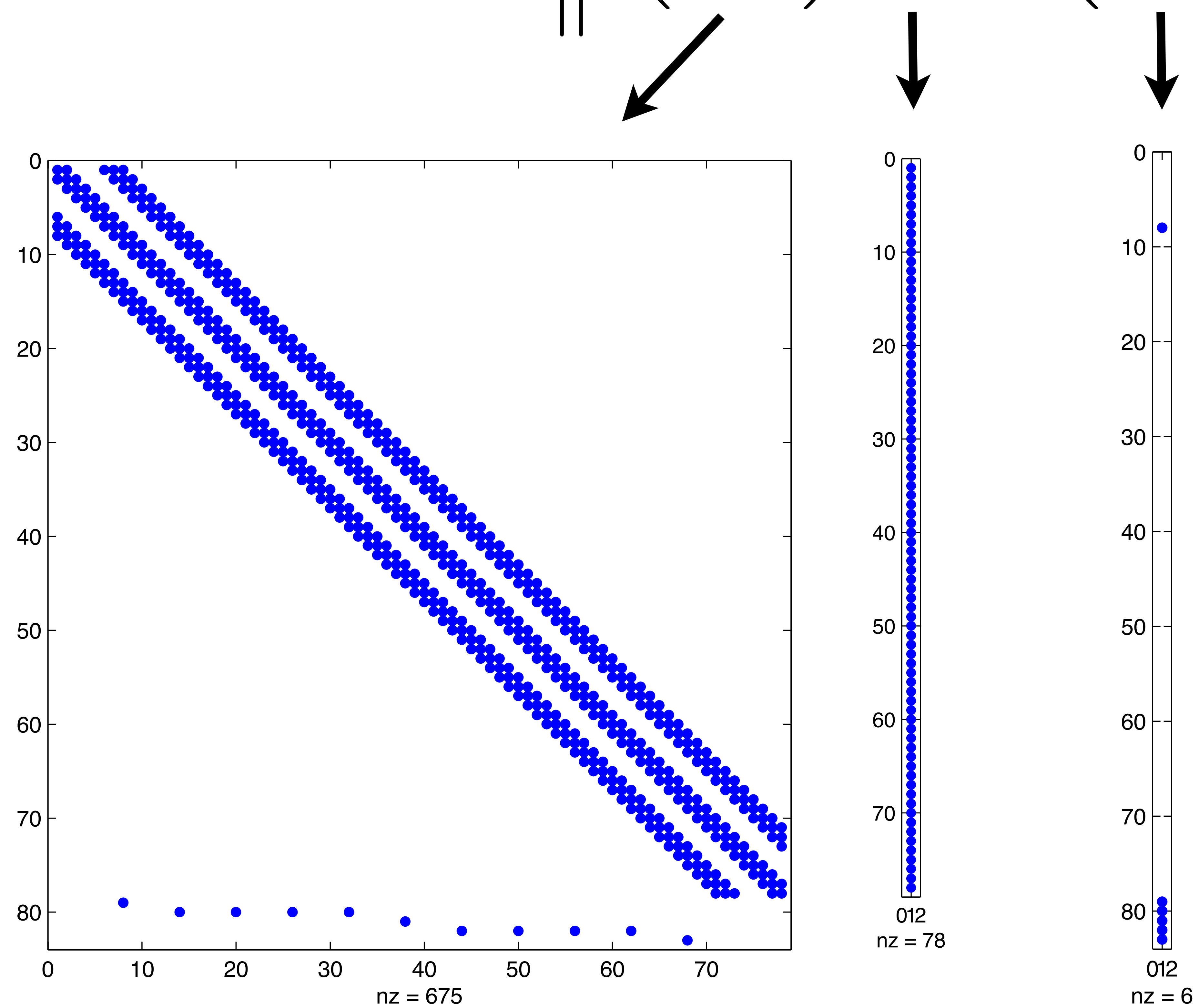
Solve $$(P^*P + \lambda^2 H^*H)\mathbf{u} = P^*\mathbf{d} + \lambda^2 H^*\mathbf{q}$$

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

Can be very different numerically!

Problem properties

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



Problem properties

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

$\begin{pmatrix} \lambda H \\ P \end{pmatrix} \rightarrow A$, $\begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \rightarrow b$

Problem properties

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

- slightly overdetermined
- unique solution
- inconsistent, nonzero residual at solution
- full column rank
- very large

Direct solvers

Sparse QR factorization, for example SuiteSparseQR.

Sparse Cholesky factorization on the normal equations.

Works for 2D problems.

Requires too much memory for 3D problems.

Basic Iterative methods

Iterative solvers for least-squares problems:

- CGLS
- LSQR
- LSMR
- AB-GMRES / BA-GMRES
- various methods for extremely overdetermined/underdetermined problems

Common features:

- do not form normal equations
- avoid numerical instability
- convergence depends on condition number squared
- efficient when condition number is low

Convergence of basic iterative methods

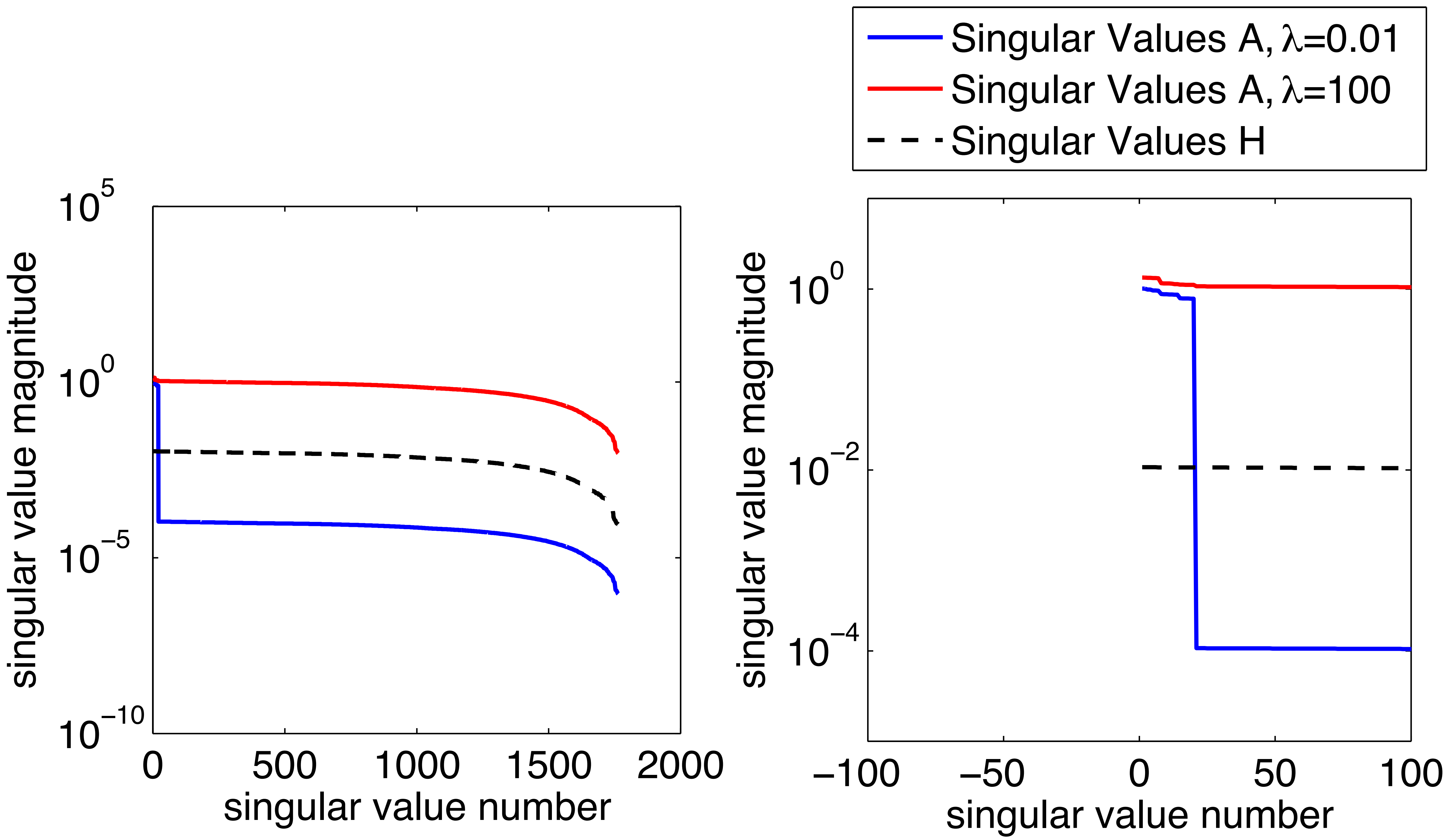
fast convergence if:

- low condition number
- also for high condition number when eigenvalues are clustered

Challenges

Very large condition number (squared) of the H^*H block

Outliers in the eigenvalue spectrum, due to P^*P



Proposed algorithm

Combine

- preconditioning with a Helmholtz solver
- low-rank decomposition

Main properties:

- bulk ‘work’ consists of solving Helmholtz problems inexactly
- low-memory imprint
- computationally feasible
- scalable

Proposed algorithm

For each frequency:

1. solve 1 Helmholtz problem per receiver $H^*W = P^*$

For each source

1. solve 1 Helmholtz problem $H^*\mathbf{g}_i = A^*\mathbf{b_i}$
2. solve exactly (easy) $(I + WW^*)\mathbf{y_i} = \mathbf{g_i}$
3. solve 1 Helmholtz problem $H\bar{\mathbf{u}}_i = \mathbf{y_i}$

Proposed algorithm

For each frequency:

1. solve 1 Helmholtz problem per receiver $H^*W = P^*$

For each source

1. solve 1 Helmholtz problem $H^*\mathbf{g}_i = A^*\mathbf{b_i}$
2. solve exactly (easy)* $(I + WW^*)\mathbf{y_i} = \mathbf{g_i}$
3. solve 1 Helmholtz problem $H\bar{\mathbf{u}}_i = \mathbf{y_i}$

* analytic solution, includes inverse of a $n_{\text{rec}} \times n_{\text{rec}}$ matrix

Proposed algorithm, derivation

problem:

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

precondition by λH , no computations

$$\left(\frac{1}{\lambda^2} H^{-*} P^* P H^{-1} + I\right)\mathbf{y} = (\lambda H^*)^{-1}(P^*\mathbf{d} + \lambda^2 H^*\mathbf{q})$$

Clusters most eigenvalues.

Even in case of inexact Helmholtz solves!

Proposed algorithm, derivation

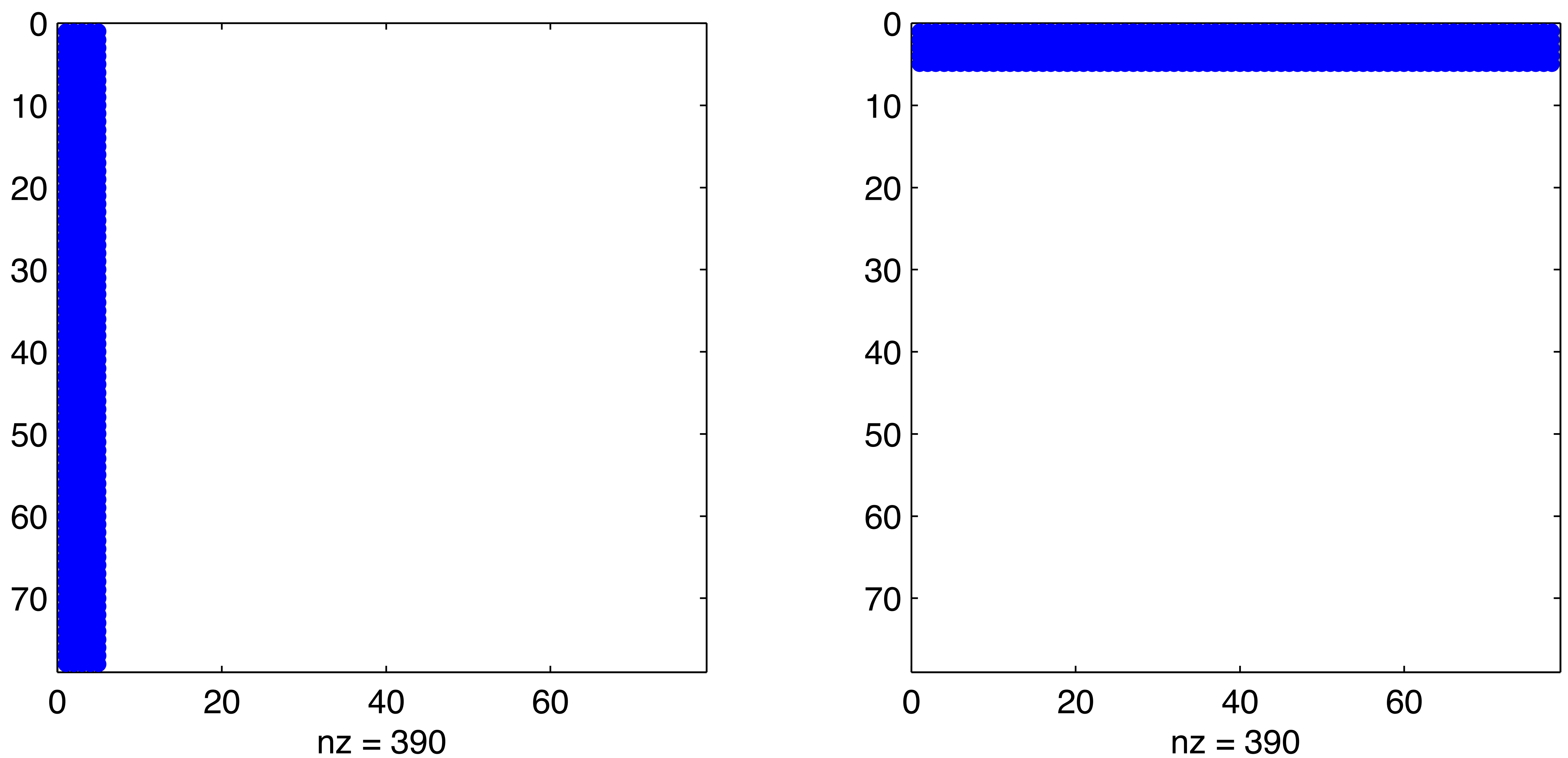
$$\left(\frac{1}{\lambda^2}\underline{H^{-*}P^*}PH^{-1}+I\right)\mathbf{y}=(\lambda H^*)^{-1}(P^*\mathbf{d}+\lambda^2H^*\mathbf{q})$$

solve explicitly

cost: n_{rec} Helmholtz problem inexactly

Proposed algorithm, derivation

$$\left(\frac{1}{\lambda^2}\underline{H^{-*}P^*}\,\underline{PH^{-1}} + I\right)\mathbf{y} = (\lambda H^*)^{-1}(P^*\mathbf{d} + \lambda^2 H^*\mathbf{q})$$



inverses transformed into low-rank factorization

Proposed algorithm, derivation

$$\left(\frac{1}{\lambda^2}WW^* + I\right)\mathbf{y} = (\lambda H^*)^{-1}(P^*\mathbf{d} + \lambda^2 H^*\mathbf{q})$$

Solve using analytic expression for the new system matrix, exactly

Solve one Helmholtz problem for the right-hand-side, inexactly

Solve one Helmholtz problem to retrieve original variables, inexactly

Proposed algorithm

For each frequency:

1. solve 1 Helmholtz problem per receiver $H^*W = P^*$

For each source

1. solve 1 Helmholtz problem $H^*\mathbf{g}_i = A^*\mathbf{b_i}$
2. solve exactly (easy) $(I + WW^*)\mathbf{y_i} = \mathbf{g_i}$
3. solve 1 Helmholtz problem $H\bar{\mathbf{u}}_i = \mathbf{y_i}$

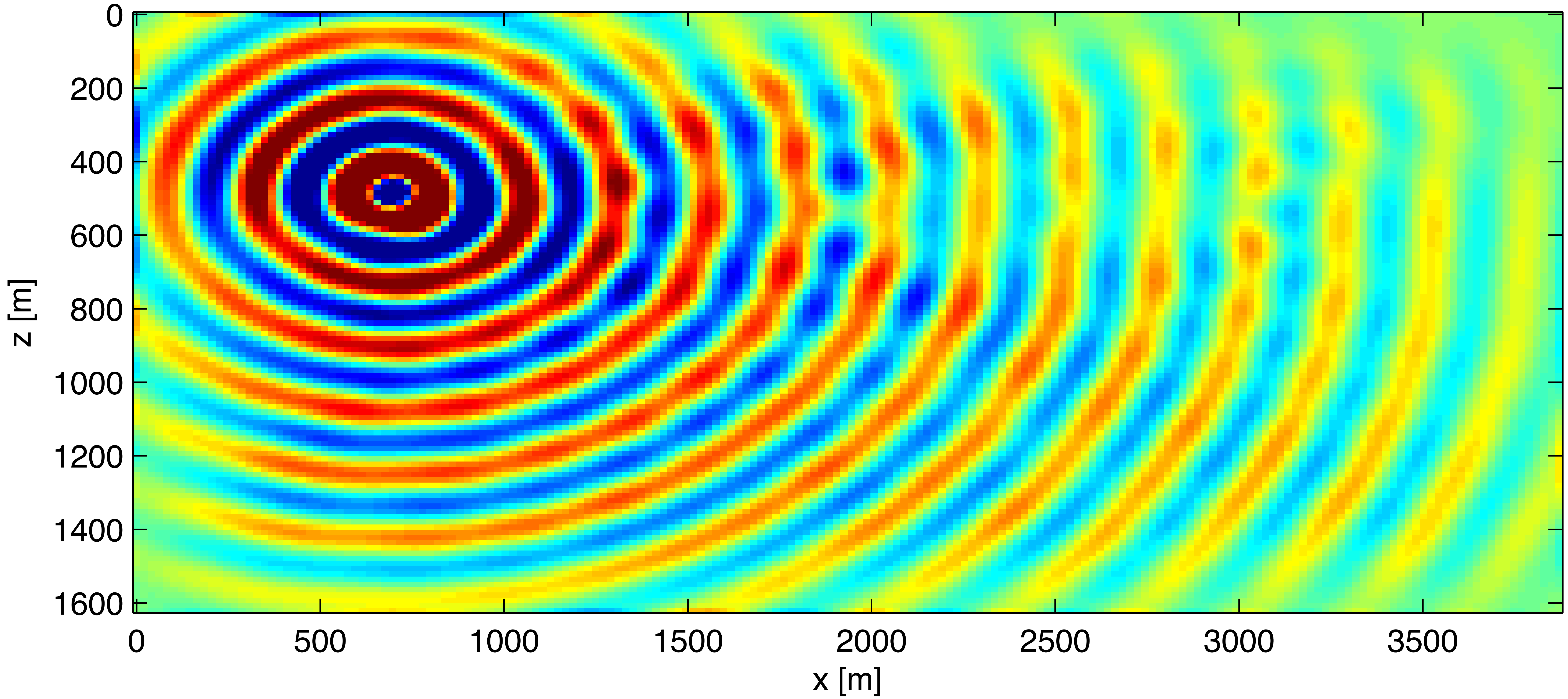
Example

direct solver

inhomogeneous true medium

used homogeneous medium to compute this field

includes PML



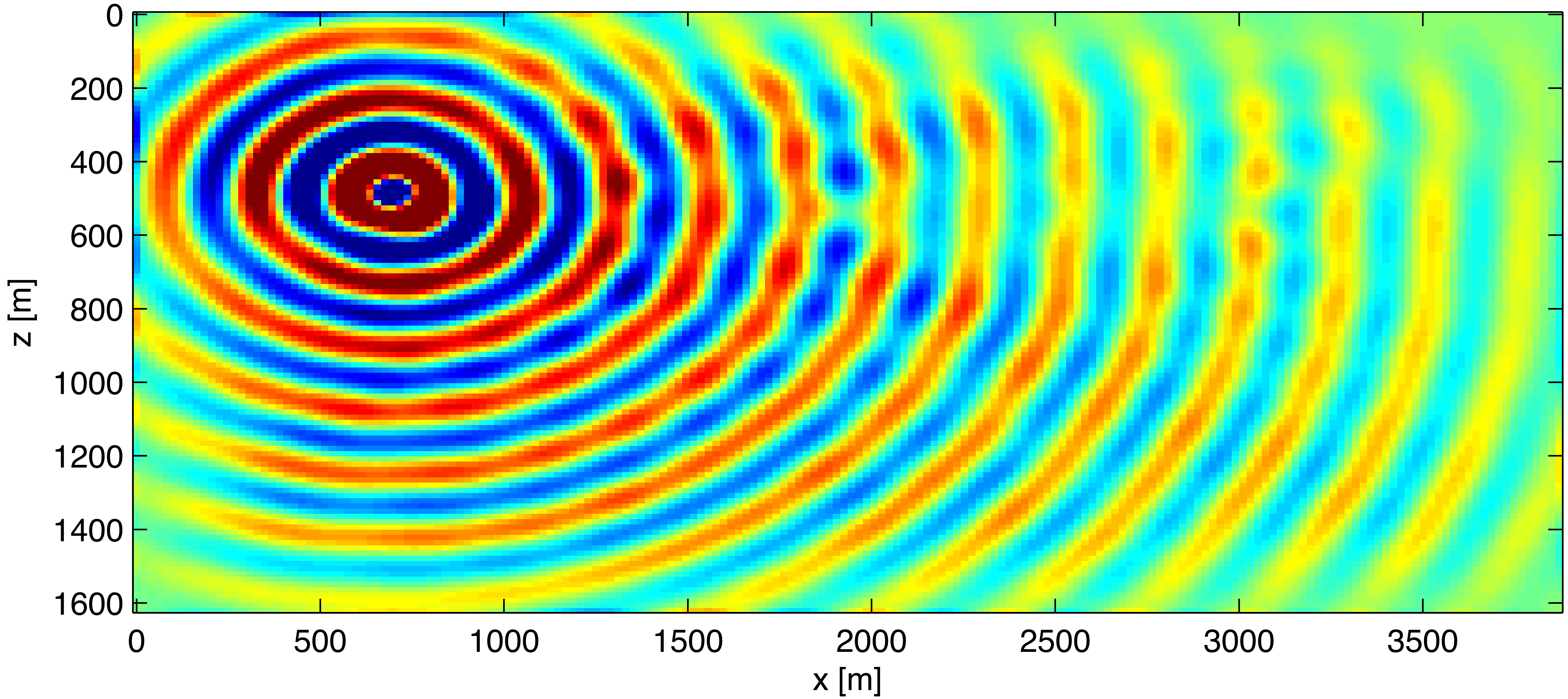
Example

iterative solver, relative error = 0.0057

inhomogeneous true medium

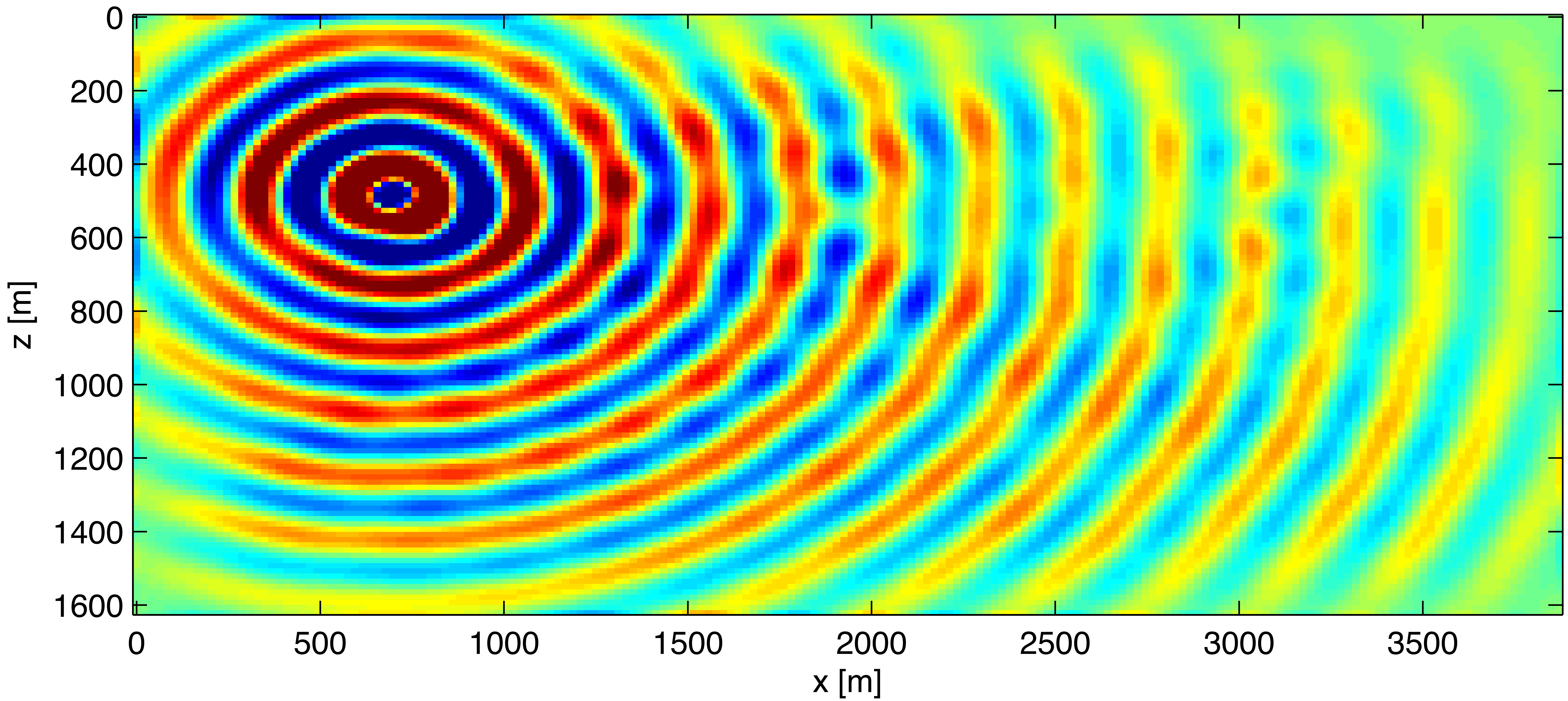
used homogeneous medium to compute this field

includes PML



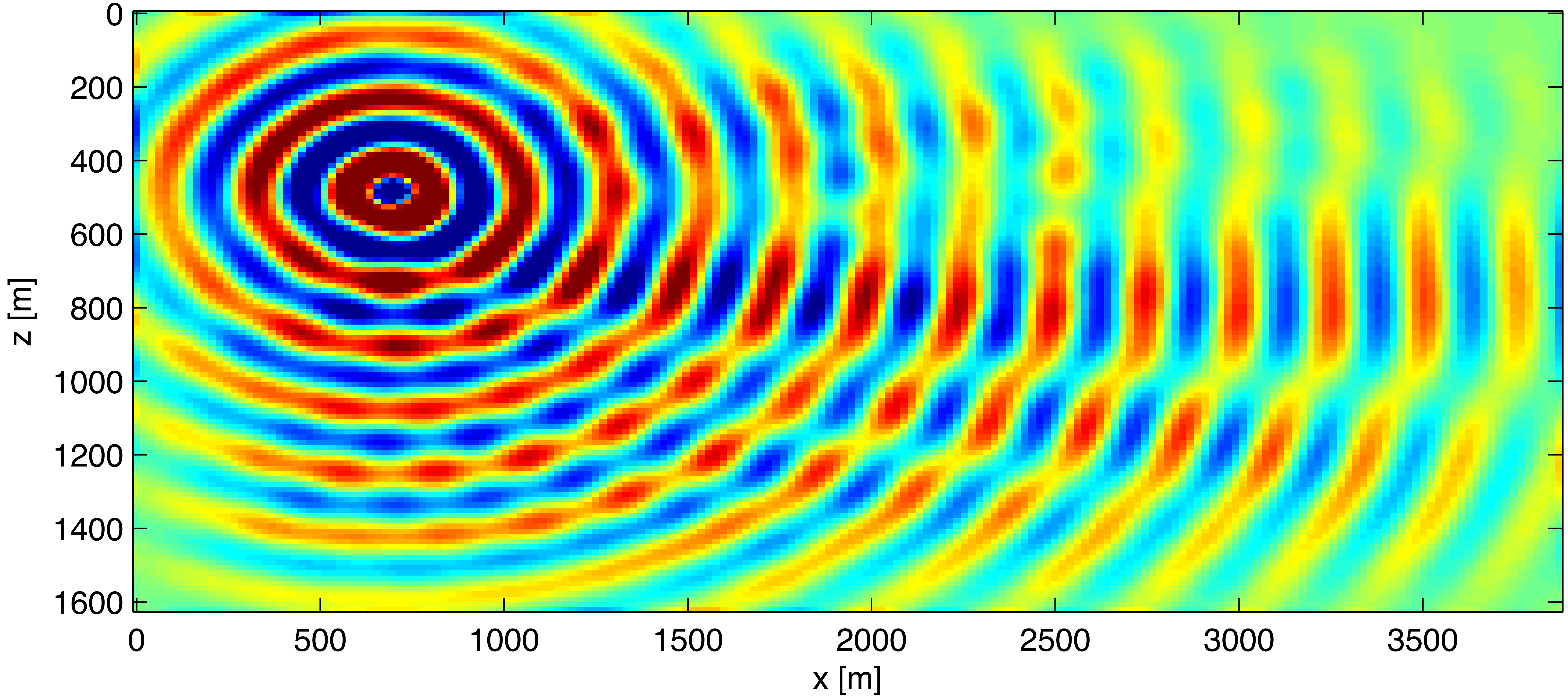
Example

iterative solver, relative error = 0.04
inhomogeneous true medium
used homogeneous medium to compute this field
includes PML



Example

iterative solver, relative error = 0.29
inhomogeneous true medium
used homogeneous medium to compute this field
includes PML



Parallelism

All expensive components are intrinsically parallel:

- solve for sources in parallel.
- solve for receivers in parallel.

Computational cost

Total inverse problem solving cost depends on

- method (FWI, WRI, full-space)
- optimization (gradient-descent, quasi-Newton, Newton-type)
- cost of solving Helmholtz, data-augmented system

Cost of FWI vs WRI vs full-space methods is future work

For now we look at the cost of the proposed algorithm

Computational cost

Per frequency

- 1 Helmholtz problem per receiver, inexactly
- 2 Helmholtz problems per source, inexactly

Cost is predictable.

Computational cost

$\kappa(H)$ contains:

- scaling effects with frequency
- domain size
- points-per-wavelength

Computational cost

Performance of the proposed algorithm?

How does it interact with the Helmholtz solver?

Look at error bounds

Computational cost

Error bound (work in progress)

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

↑

relative error
in solution

Computational cost

Error bound (work in progress)

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

↑ relative error in solution

↑ Helmholtz condition number

Computational cost

Error bound (work in progress)

(observable) residuals of the 3 Helmholtz solves

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

relative error in solution

Helmholtz condition number

SLIM

Computational cost

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

Conclusion based on error analysis so far:

- error depends on $\kappa(H)$ and H^{-1} on a vector inside the norm
- Helmholtz solver/preconditioner does not enter the bound (nice)
- depends on the residual of the inexact Helmholtz solves

- proposed algorithm brings down the $\kappa(H)^2$ dependence of standard least-squares algorithms

Memory requirements

use low-memory Helmholtz solvers (CGMN, shifted Laplacian)

peak memory: $\approx n_{rec} N$

N : number of grid points

this only is the peak if $n_{\text{rec}} \lessapprox n_{\text{stencil points}}$

Conclusions

3D WRI is feasible

same parallelism as in FWI

can use already available Helmholtz solvers

Acknowledgements

Chen Greif & my colleagues from SLIM

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, ION, Petrobras, PGS, Statoil, Total SA, Sub Salt Solutions, WesternGeco, and Woodside.