3D frequency domain FWI

using a row-projected Helmholtz solver

Tristan van Leeuwen, Art Petrenko

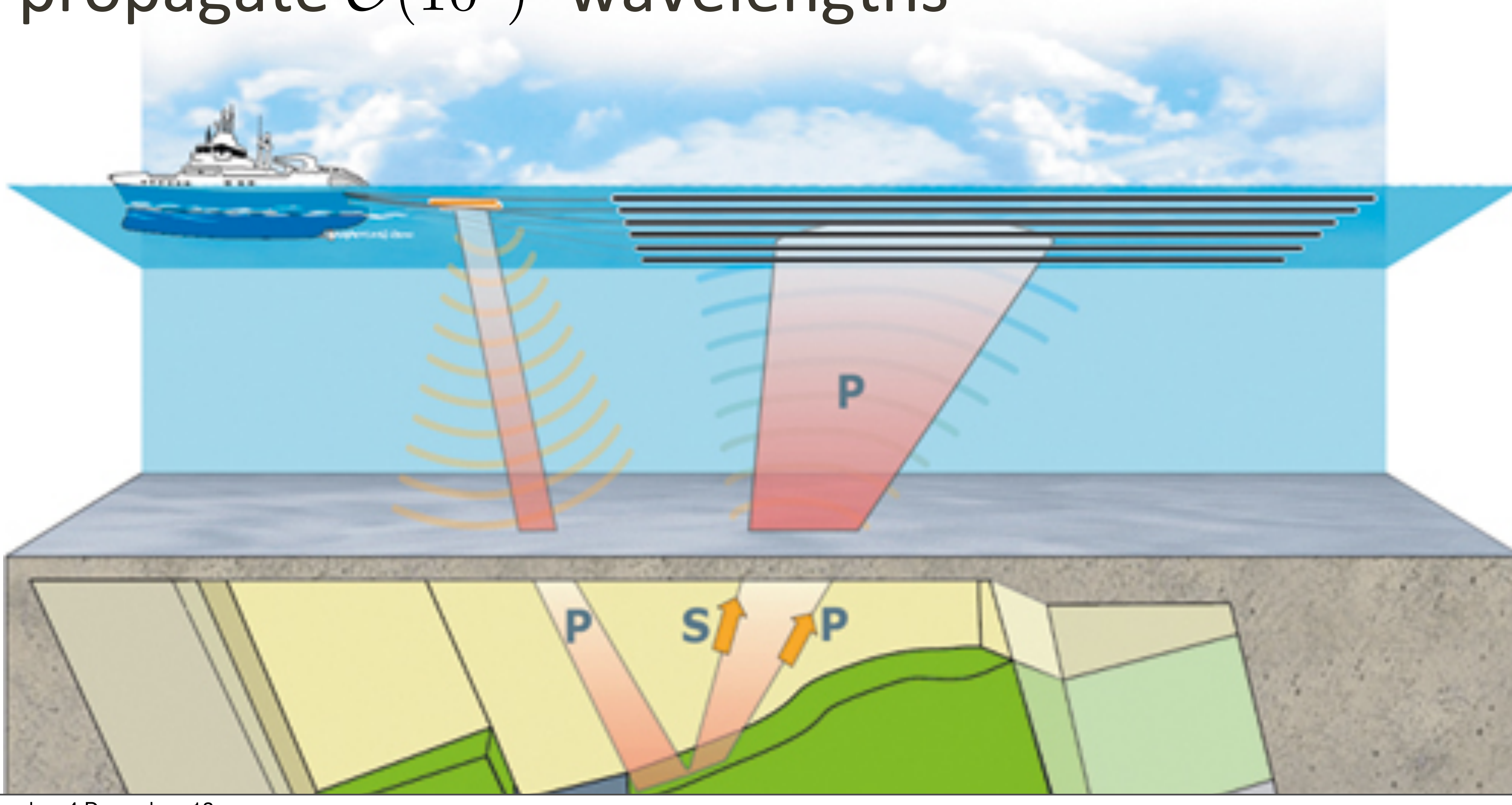
joint work with:

D. Gordon (univ. Haifa) & R. Gordon (Technion)

SLIM
University of British Columbia

Infer 3D velocity model from *multi-experiment* data:

- $\mathcal{O}(10^9)$ unknowns
- $\mathcal{O}(10^{15})$ datapoints
- propagate $\mathcal{O}(10^2)$ wavelengths



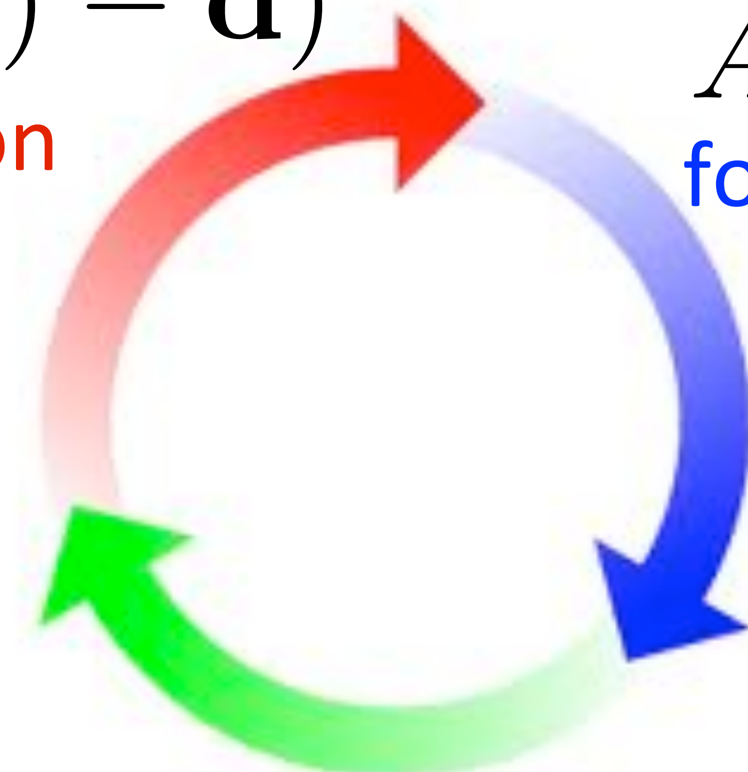
SLIM

$$\min_{\mathbf{m}} \rho(F(\mathbf{m}) - \mathbf{d})$$

formulation

$$A(\mathbf{m})\mathbf{u} = \mathbf{q}$$

forward modelling



$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{s}_k$$

optimization strategies

computational framework

Formulation

non-linear least-squares problem:

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) = \sum_{i=1}^{M} ||\mathbf{d}_i - P_i \mathbf{u}_i||_2^2$$

gradient:

$$\frac{\partial \Phi}{\partial m_k} = \sum_{i=1}^{M} \mathbf{u}_i^H \left(\frac{\partial A(\mathbf{m})}{\partial m_k} \right)^H \mathbf{v}_i$$

where:

$$A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

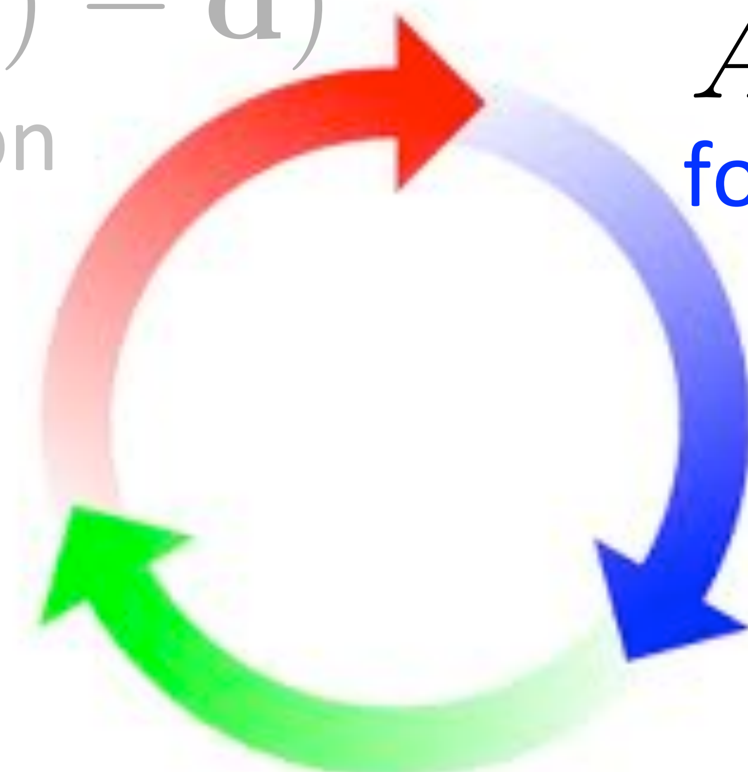
$$A(\mathbf{m})^H \mathbf{v}_i = P_i^T (\mathbf{d}_i - P_i \mathbf{u}_i)$$

$$\min_{\mathbf{m}} \rho(F(\mathbf{m}) - \mathbf{d})$$

formulation

$$A(\mathbf{m})\mathbf{u} = \mathbf{q}$$

forward modelling



$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{s}_k$$

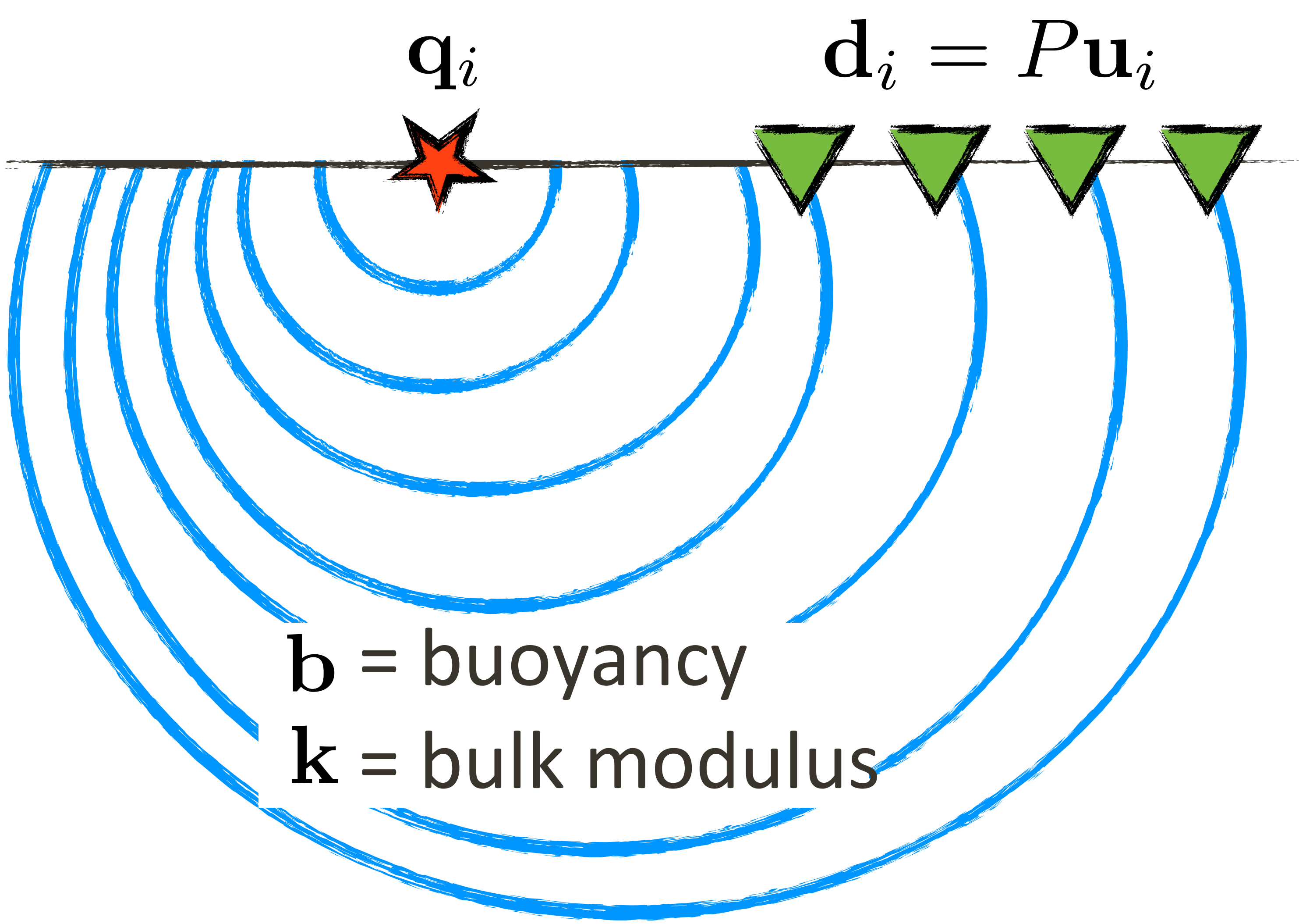
optimization strategies

computational framework

Forward modelling

We model the data in the *acoustic* approximation

$$\big(\omega^2 + \mathbf{k}\nabla \cdot (\mathbf{b}\nabla)\big)\mathbf{u}_i = \mathbf{q}_i$$



Preconditioning

Replace original system $A\mathbf{x} = \mathbf{b}$ by

$$M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$$

where $M^{-1} \approx A^{-1}$ *and* can be applied cheaply.

Popular approaches:

- *Asymptotic/one-way approximations*
- *Multi-grid on damped equation*

Our approach:

- *Generic, simple and robust*
- *Easily parallelized and optimized*

[Ernst & Martin, 2012]

Kaczmarz

The Kaczmarz method solves a system

$$A\mathbf{x} = \mathbf{b}$$

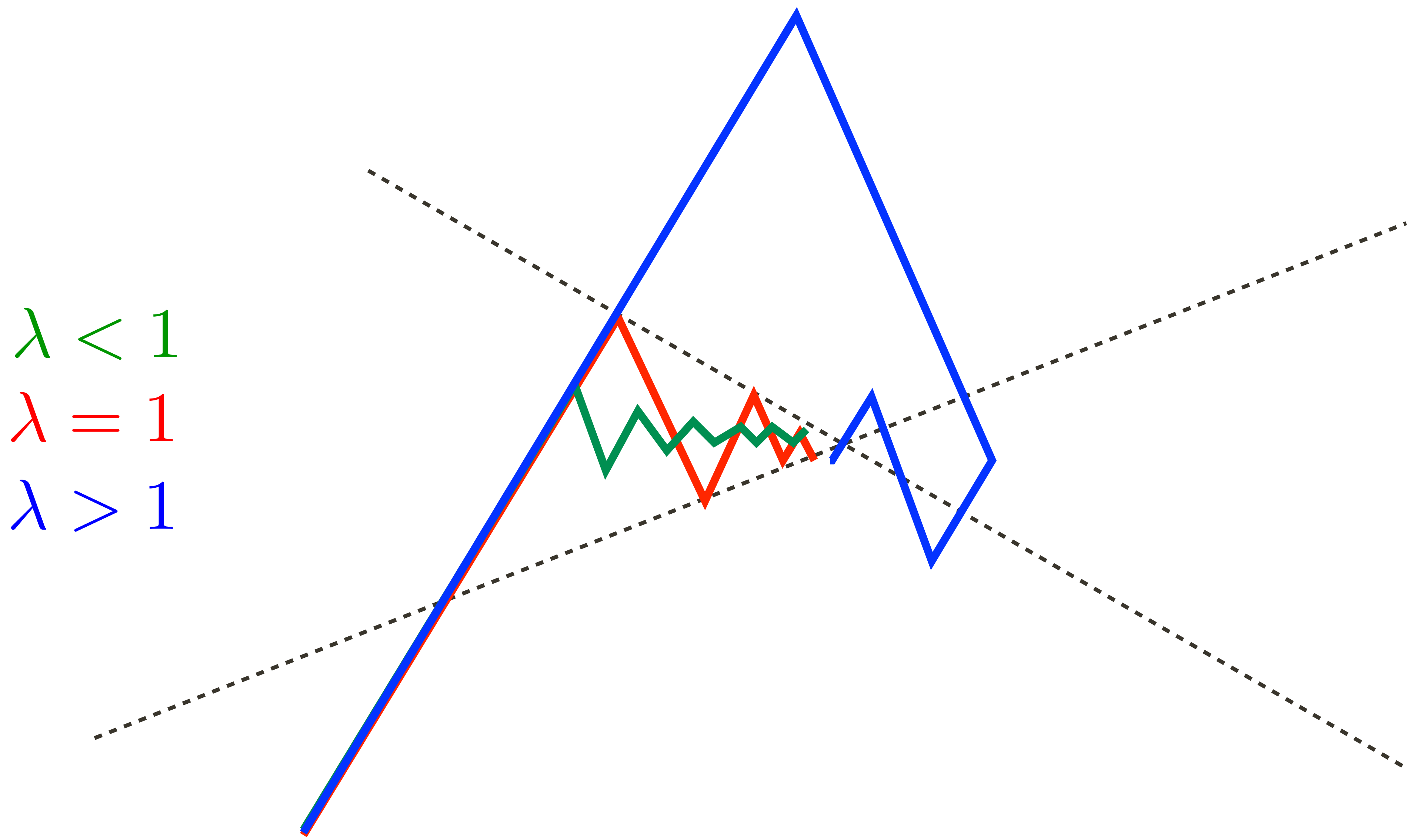
by successive row projections

$$\mathbf{x} := \mathbf{x} + \frac{\lambda}{||\mathbf{a}_i||_2^2}\left(b_i - \mathbf{a}_i^T\mathbf{x}\right)\mathbf{a}_i,$$

with relaxation parameter $0 < \lambda < 2$

[Kaczmarz, ’37]

Kaczmarz



Kaczmarz

Split matrix: $AA^T = D + L + L^T$

The corresponding preconditioner is given by

$$M^{-1} = A^T H$$

where

$$H = \lambda(2-\lambda)(D+\lambda L^T)^{-1} D (D+\lambda L)^{-1}$$

The matrix $M^{-1}A$ is *symmetric* and *positive semidefinite*; we can use CG.

Error propagation

analyze error propagation:

$$\mathbf{e}_{k+1} = Q(\lambda)\mathbf{e}_k$$

with $\quad Q(\lambda) = I - A_\lambda^T H_\lambda A_\lambda$

Fourier analysis: $\quad e_j(\theta) = \exp(\imath j\theta)$

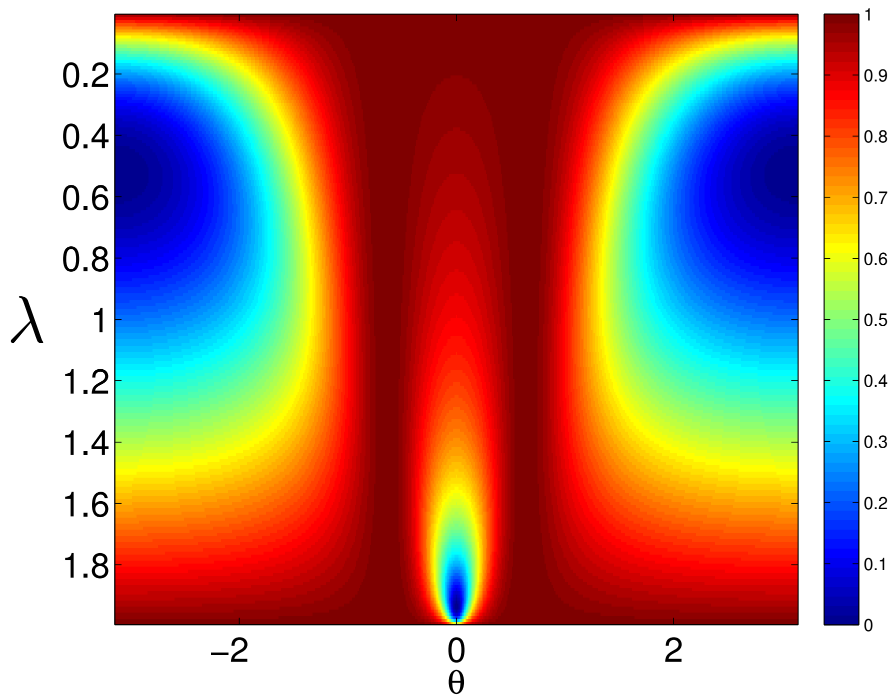
$$a(\lambda,\theta) = |Q(\lambda)\mathbf{e}(\theta)|/|\mathbf{e}(\theta)|$$

Error propagation

study the amplitude as a function of $n_g = 2\pi/(kh)$

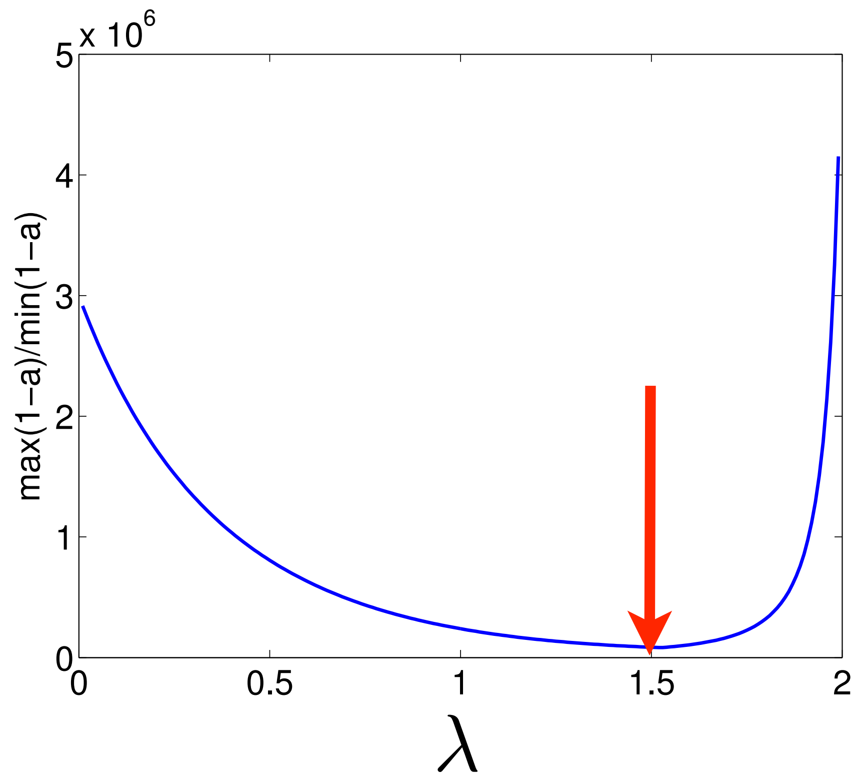
$n_g = 10$

$a(\lambda, \theta)$

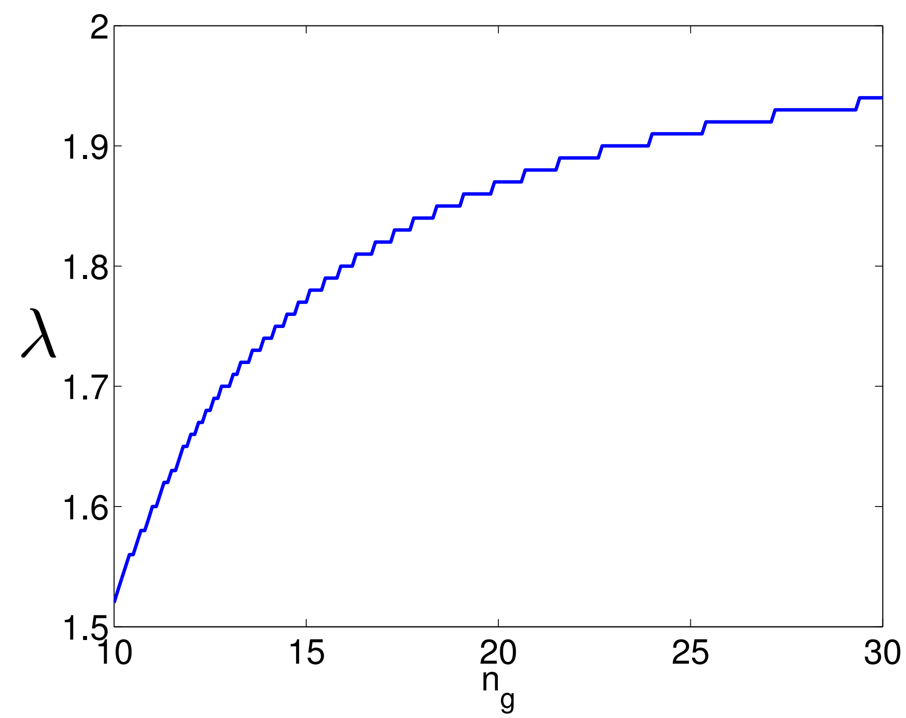


$n_g = 10$

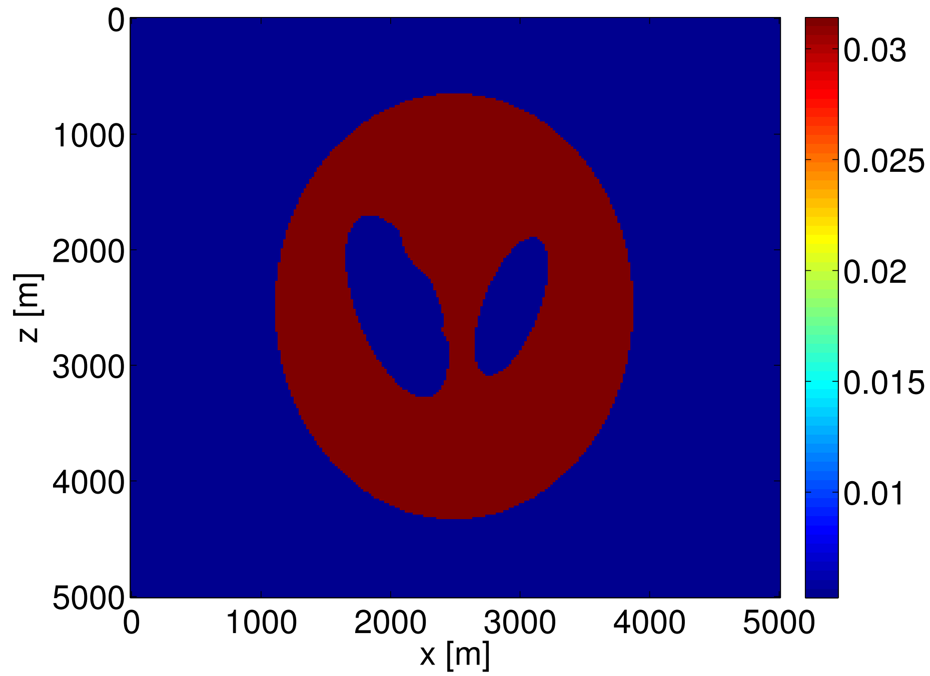
`condition number'

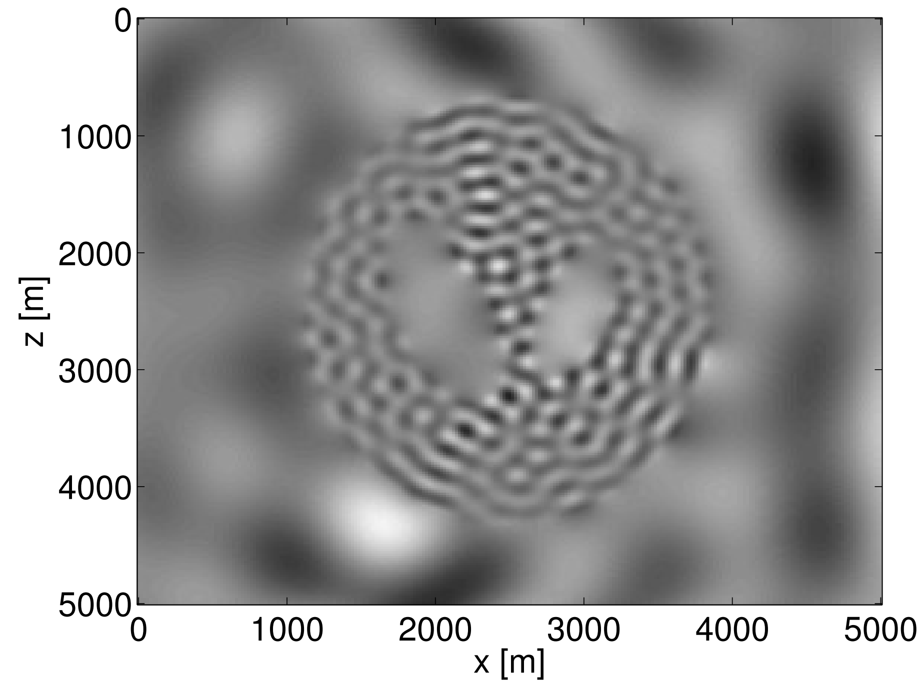


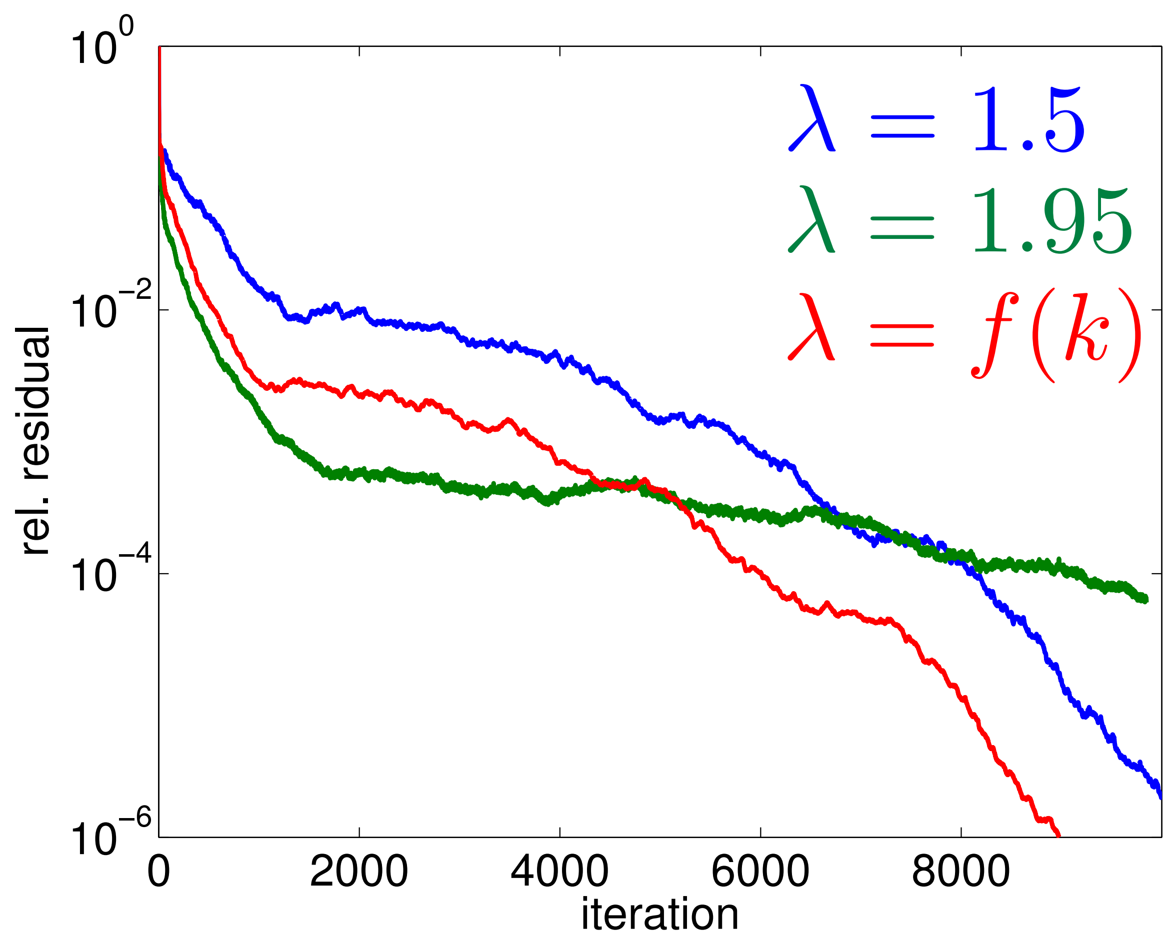
optimal λ



Error propagation







Kaczmarz

We never form the matrix explicitly, but compute its action:

Algorithm 1 $\mathsf{DKSWP}(A, \mathbf{r}, \lambda) = M^{-1}\mathbf{r}$

$\mathbf{x} = 0$
forward sweep
for $i = 1$ to n **do**
  $\mathbf{x} := \mathbf{x} + \lambda(r_i - \mathbf{a}_i^T\mathbf{x})\mathbf{a}_i / ||\mathbf{a}_i||_2^2$
end for
backward sweep
for $i = n$ to 1 **do**
  $\mathbf{x} := \mathbf{x} + \lambda(r_i - \mathbf{a}_i^T\mathbf{x})\mathbf{a}_i / ||\mathbf{a}_i||_2^2$
end for
return x

CARP CG

Algorithm 1 $\mathsf{CARP}-\mathsf{BCG}(A, U_0, S, \gamma)$

$P_0 = R_0 = \mathsf{DKSWP}(A, U_0, B, \gamma) - U_0$

while not converged **do**

$Q_k = P_k - \mathsf{DKSWP}(A, P_k, 0, \gamma)$

$\alpha_k = ({P_k}^* Q_k)^{-1}({R_k}^* R_k)$

$X_{k+1} = X_k + P_k \alpha_k$

$R_{k+1} = R_k - Q_k \alpha_k$

$\beta_k = ({R_k}^* R_k)^{-1}({R_{k+1}}^* R_{k+1})$

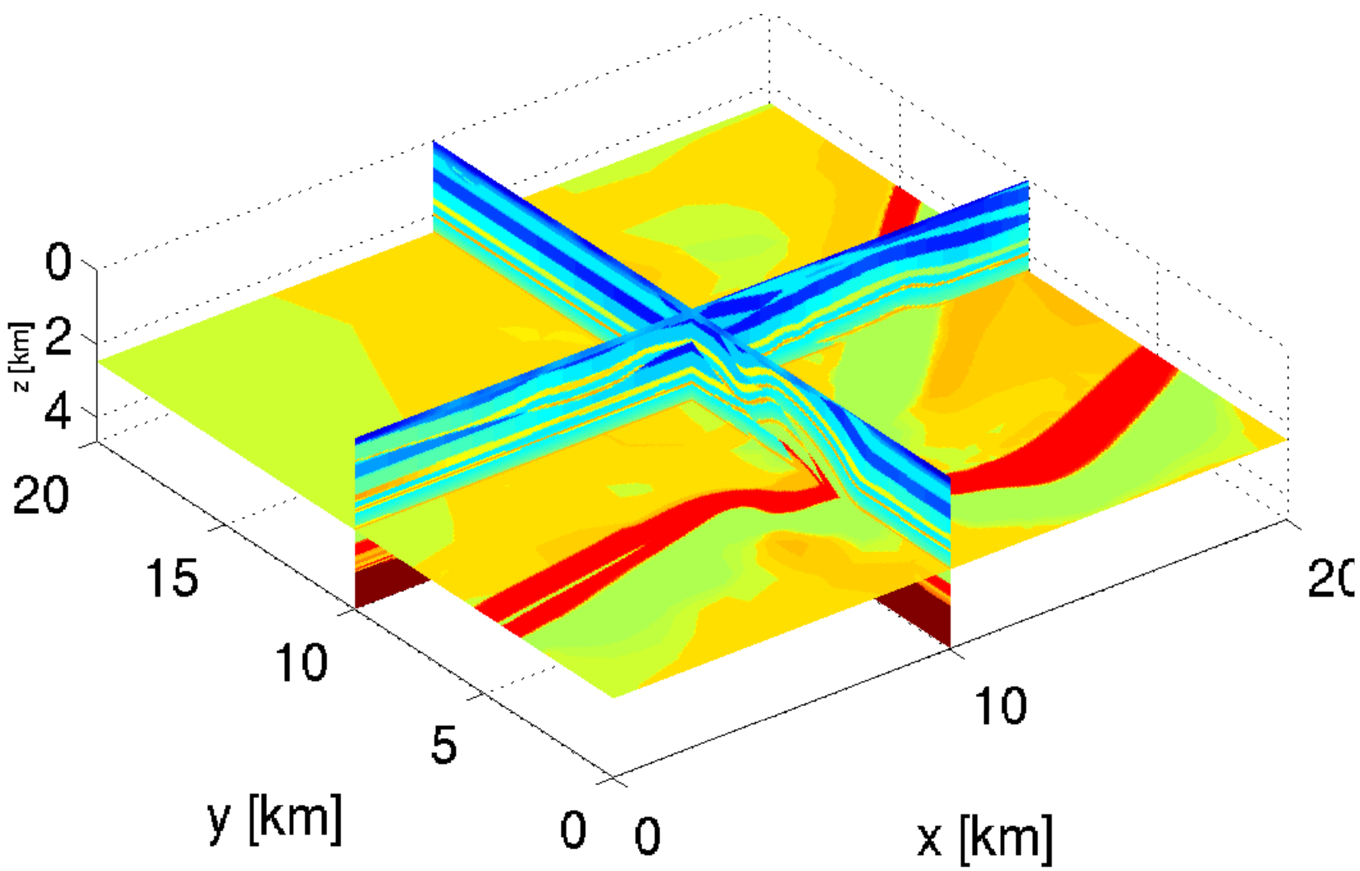
$P_{k+1} = R_k + P_k \beta_k$

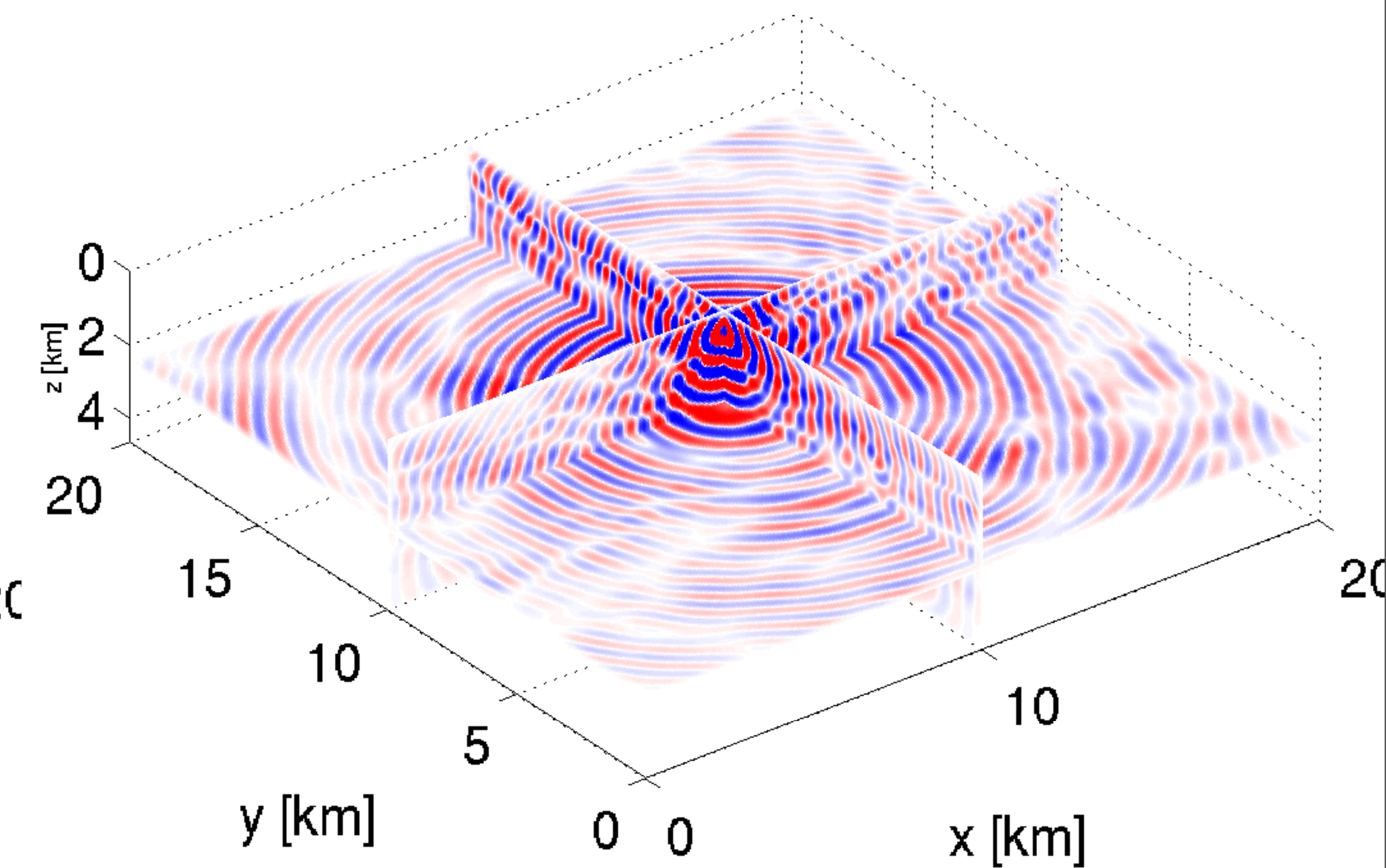
$k = k + 1$

end while

Example

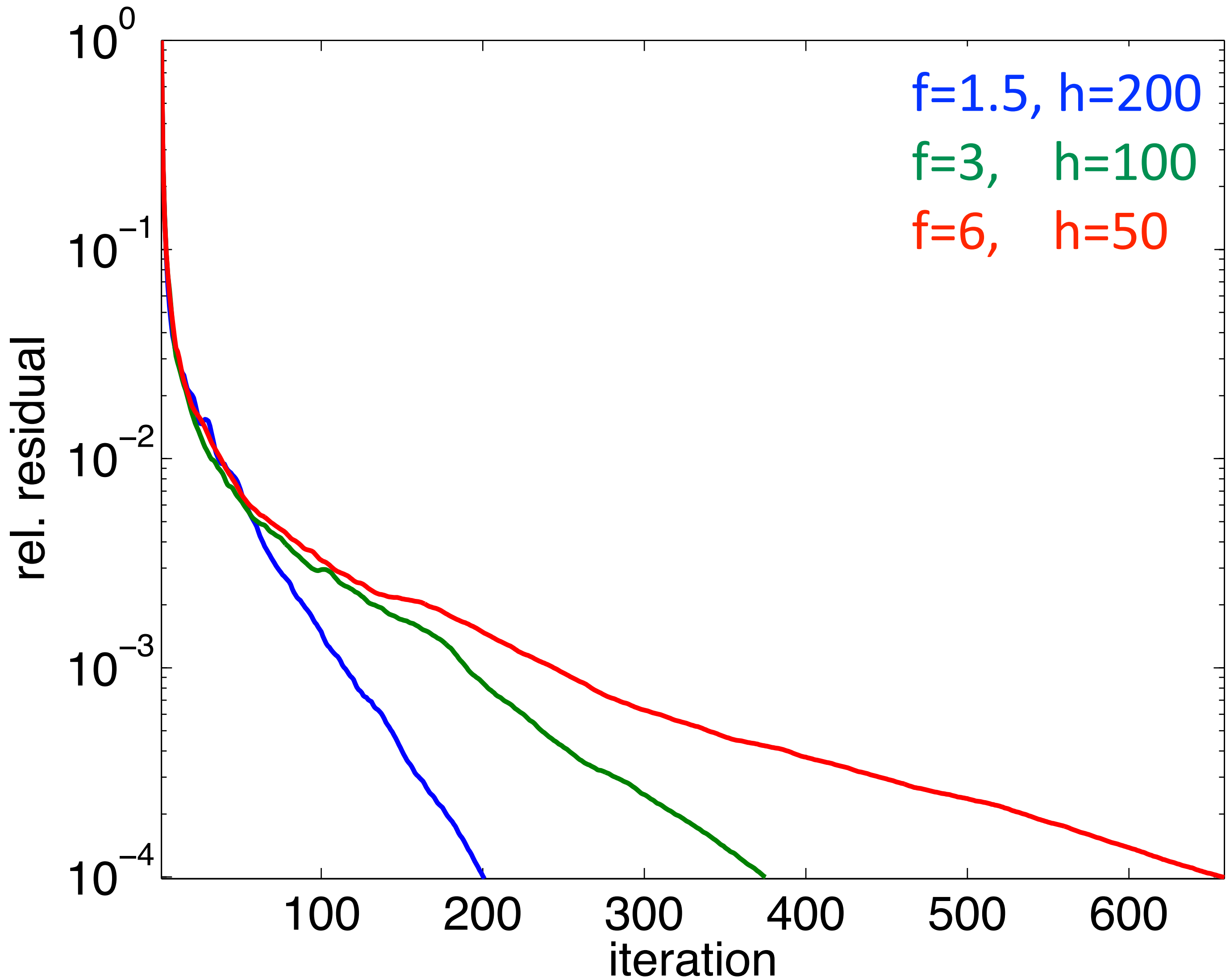
27 point stencil (2nd order), PML





Example

one r.h.s.



on 1 processor

Example

Parallelization via domain-decomposition:

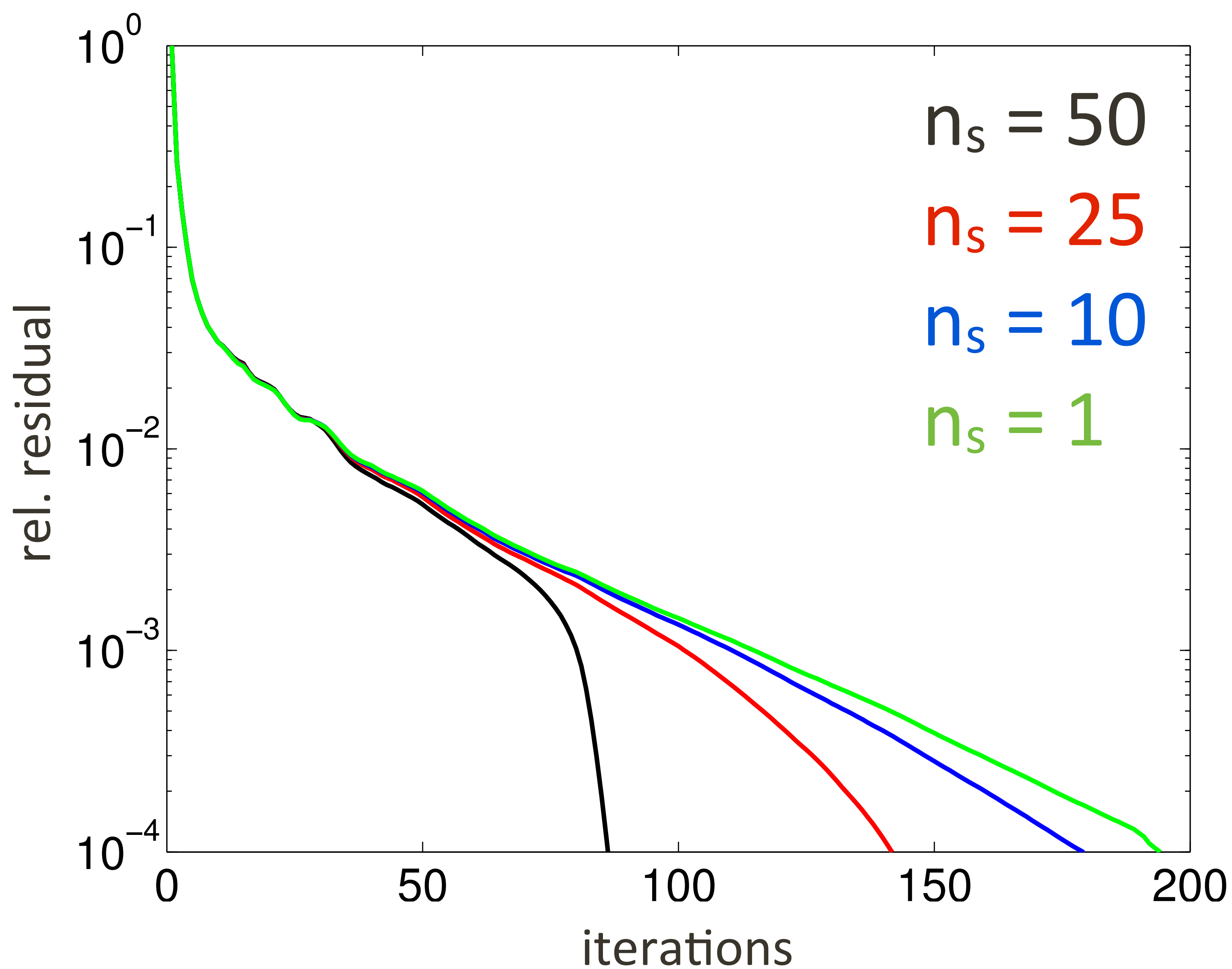
- sweeps applied on subdomains
- `additive Schwarz'
- convergence guaranteed

np	iter	time	efficiency
1	659	20785.40	1.00
2	657	11306.90	0.92
4	596	4882.50	0.96
8	603	3960.10	0.60

grid: 47x201x201, h=100, f=3 Hz, $\epsilon = 10^{-4}$

Example

multiple r.h.s.

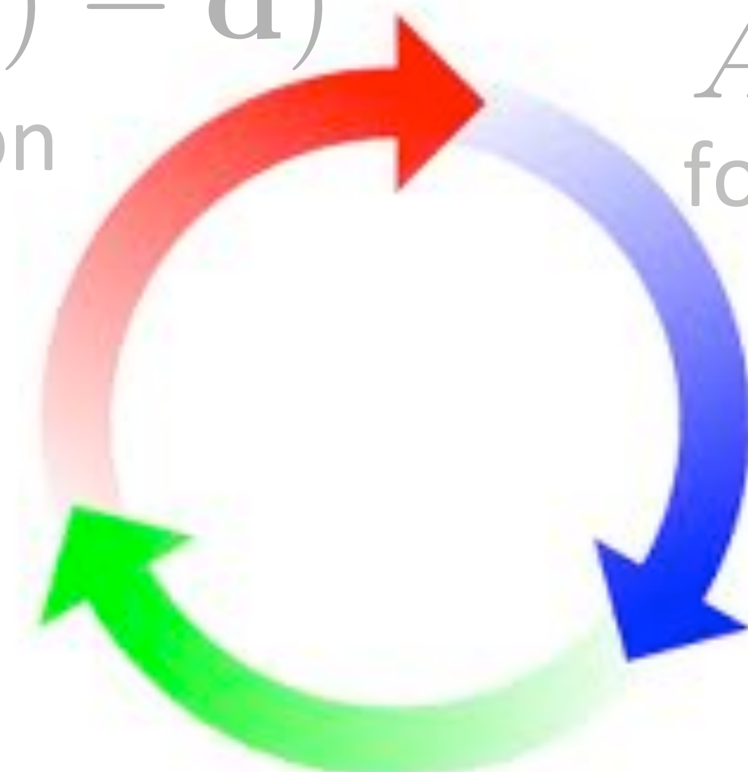


$$\min_{\mathbf{m}} \rho(F(\mathbf{m}) - \mathbf{d})$$

formulation

$$A(\mathbf{m})\mathbf{u} = \mathbf{q}$$

forward modelling



$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{s}_k$$

optimization strategies

computational framework

Fast optimization

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) = \frac{1}{M}\sum_{i=1}^{M} \phi_i(\mathbf{m})$$

steepest descent

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \lambda_k \nabla\Phi(\mathbf{m}_k)$$

But: evaluation of *full* misfit and gradient is very expensive.

Fast optimization

Gradient descent with errors:

$$\nabla\widetilde{\Phi}_k = \nabla\Phi_k + \mathbf{e}_k$$

stochastic/incremental gradient require an *unbiased* estimate*:*

$$\mathbb{E}(\mathbf{e}_k) = 0$$

and have *sublinear* convergence rate

[Bertsekas ’96,’08; Nemirovski ’00]

Fast optimization

Instead, require that the error is *bounded*

$||\mathbf{e}_k|| \leq B_k$ and $\lim_{k\to\infty} B_{k+1}/B_k \leq 1$

Then

$$\Phi(\mathbf{m}_k) - \Phi(\mathbf{m}^*) = c^k(\Phi(\mathbf{m}_k) - \Phi(\mathbf{m}_0)) + \mathcal{O}(\max\{B_k, c^k\})$$

where $0 < c < 1$

Linear convergence rate if $B_k = \mathcal{O}(\gamma^k)$

[Friedlander & Schmidt '12]

SLIM

Fast optimization

The gradient is an *average*

$$\nabla \Phi = \frac{1}{M} \sum_{i=1}^{M} \nabla \phi_i$$

which we can approximate by

$$\nabla \Phi \approx \nabla \widetilde{\Phi} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \nabla \phi_i$$

Express the error as

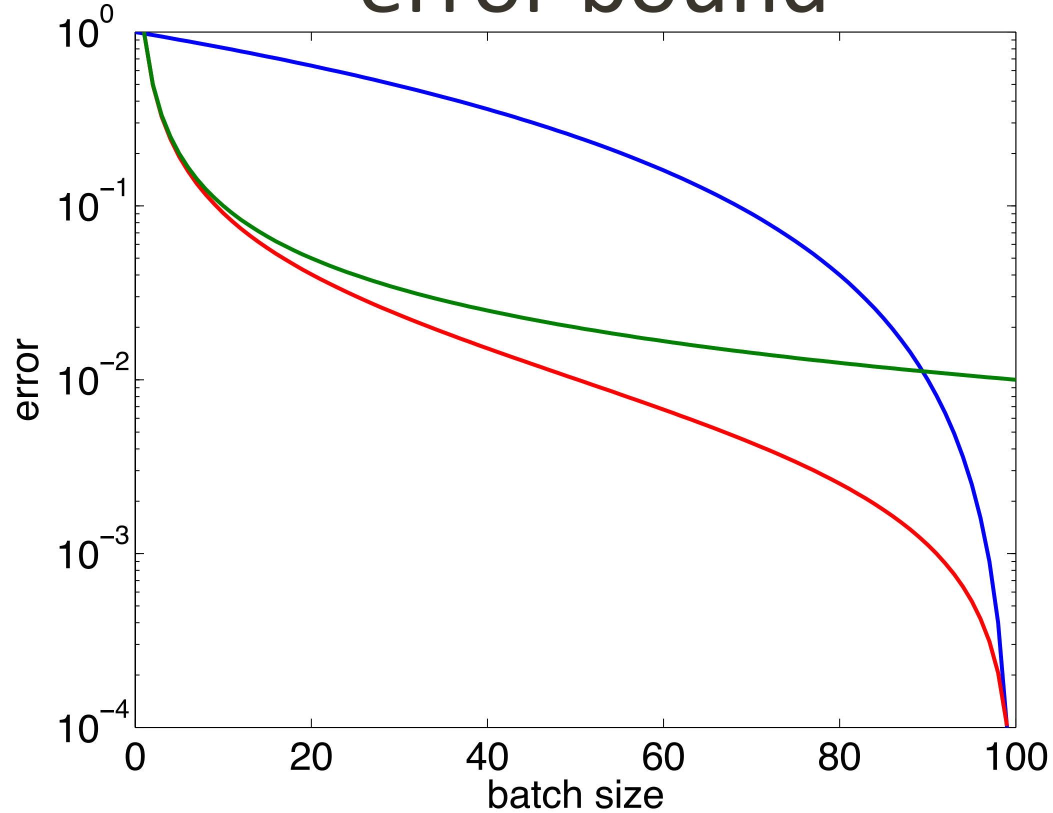
$$\mathbf{e}_k = \frac{M - |\mathcal{I}_k|}{M|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \nabla \phi_i + \frac{1}{M} \sum_{i \notin |\mathcal{I}|} \nabla \phi_i$$

Fast optimization

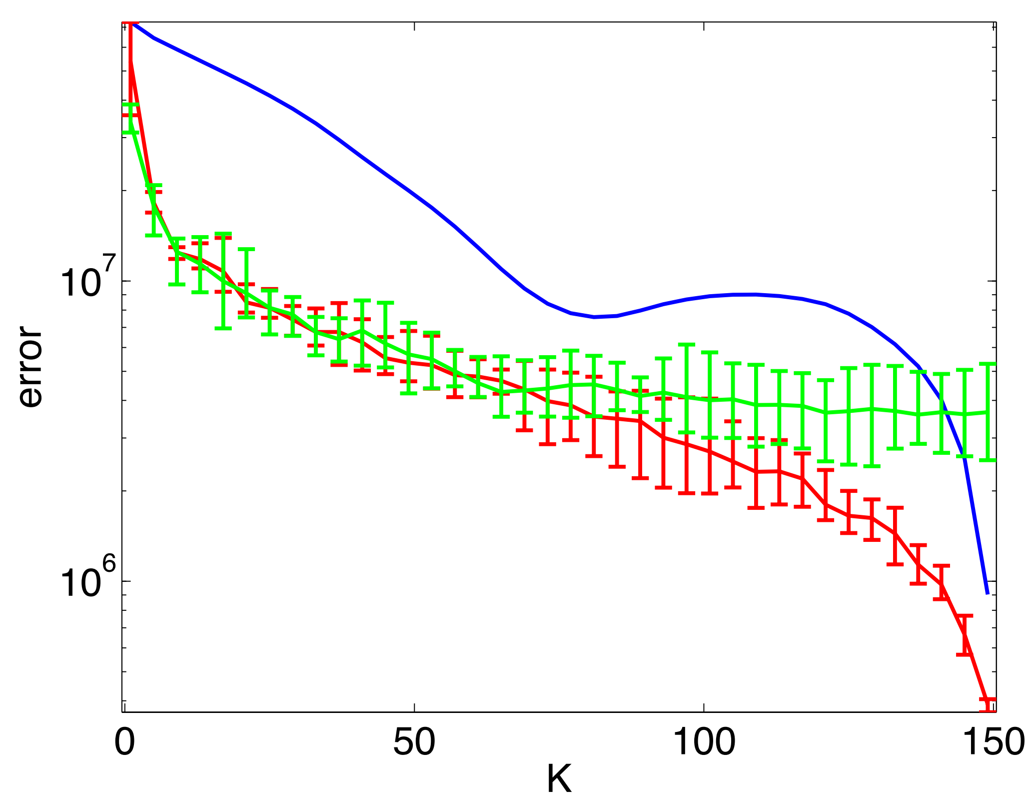
Decrease the error by adding elements to the batch

- in a pre-scribed order
- random *without* replacement
- random *with* replacement

error bound



numerical result

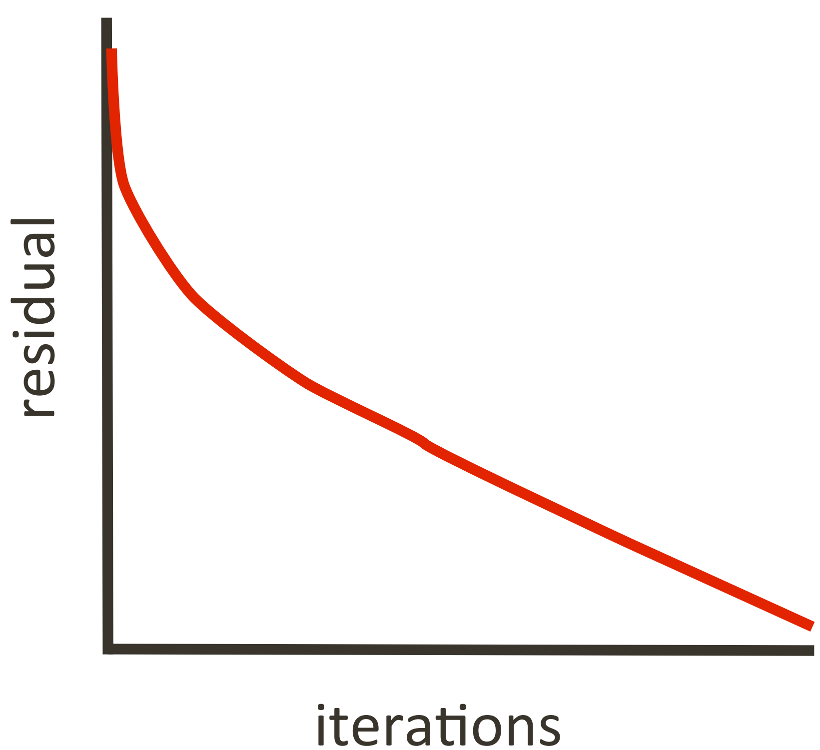


Fast optimization

use approximate PDE solves:

$$A(\mathbf{m})\mathbf{u}_i \approx \mathbf{q}_i$$
$$A(\mathbf{m})^H\mathbf{v}_i \approx P_i^T(\mathbf{d}_i - P_i\mathbf{u}_i)$$

control the accuracy by the number of iterations



Examples

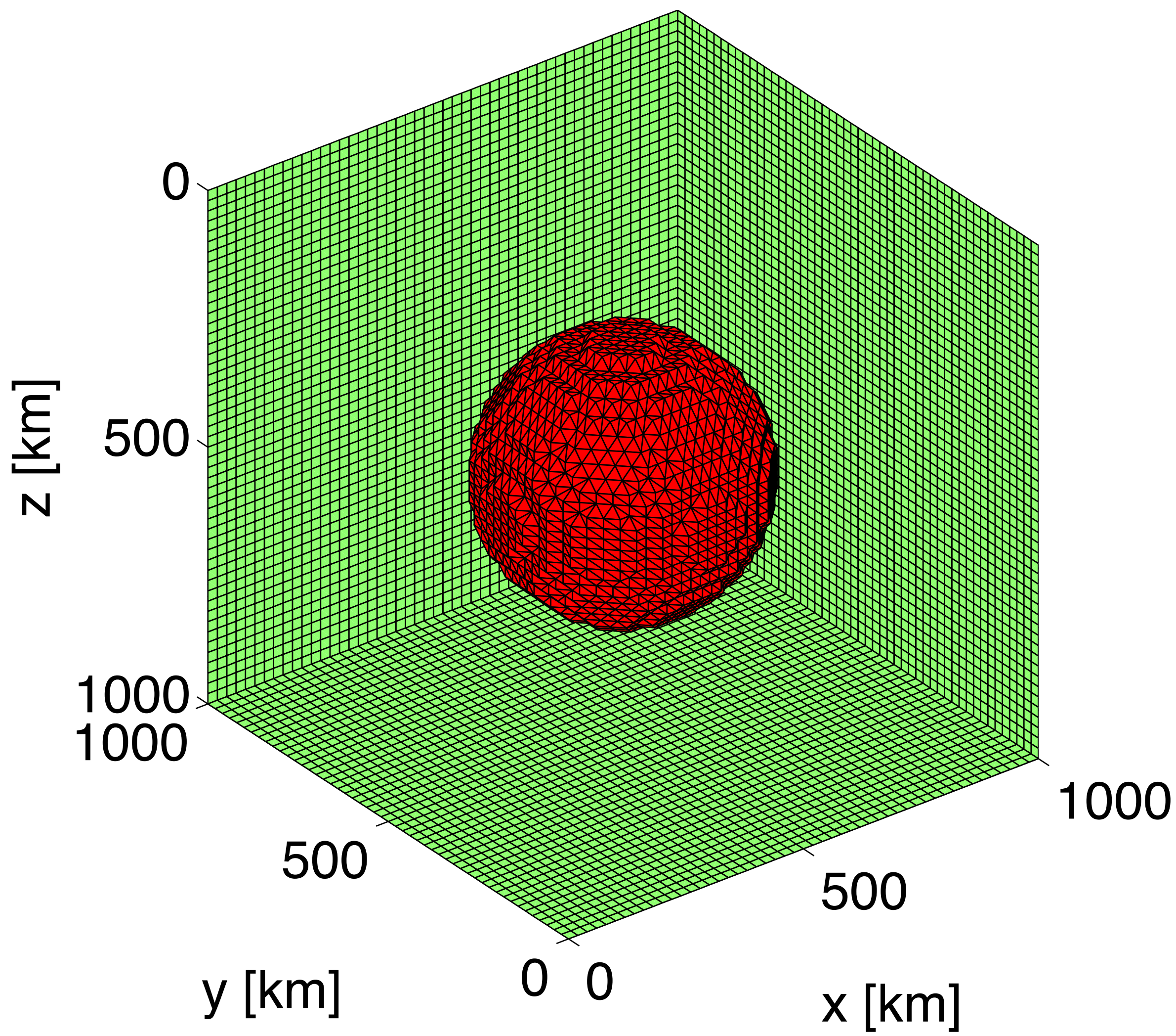
Invert multi-source, multi-frequency data:

$$\min_{\mathbf{m},\mathbf{w}} \sum_{i=1}^{M} ||w_i F_i(\mathbf{m}) - \mathbf{d}_i||_2^2$$

- source calibration via variable projection
- increase sources as: $b_{k+1} = \min\{b_k + \beta, b_{\max}\}$
- increase accuracy as: $\epsilon_{k+1} = \max\{\alpha\epsilon_k, \epsilon_{\min}\}$
- L-BFGS with 5 vectors

Edam model

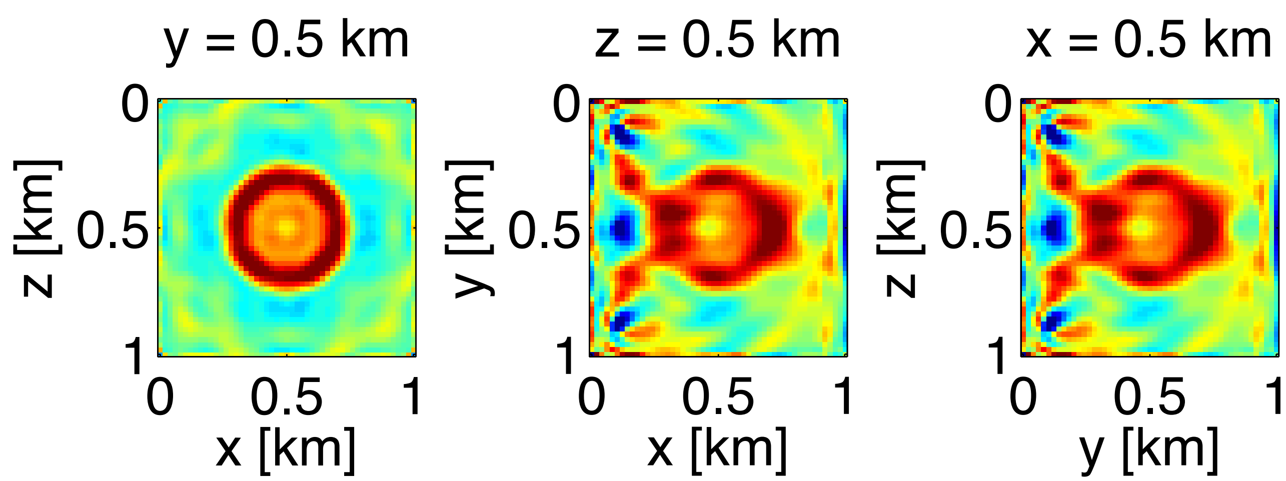
51^3 gridpoints, f = [5,10,15]

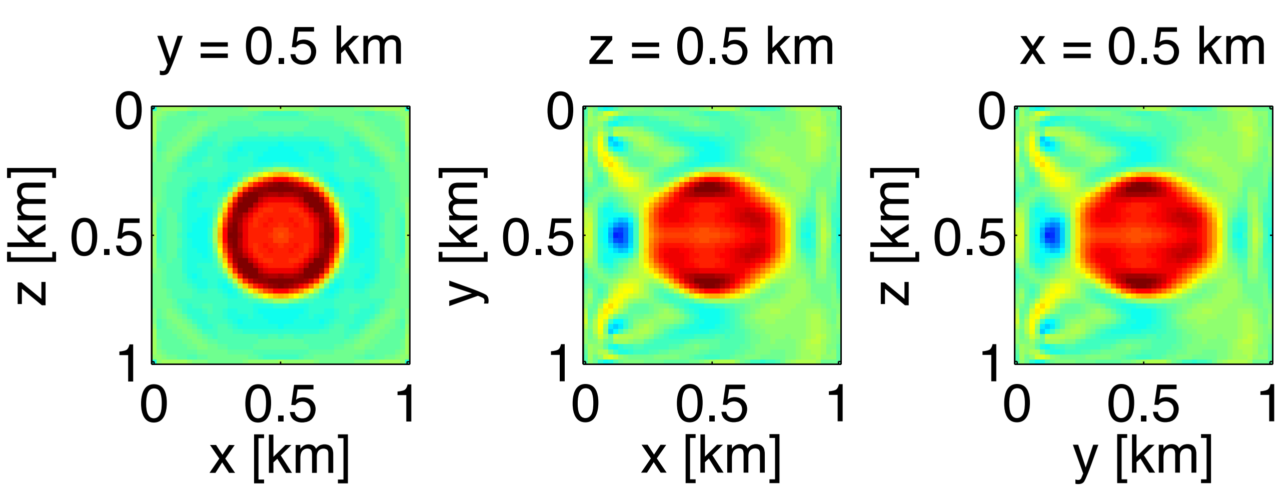


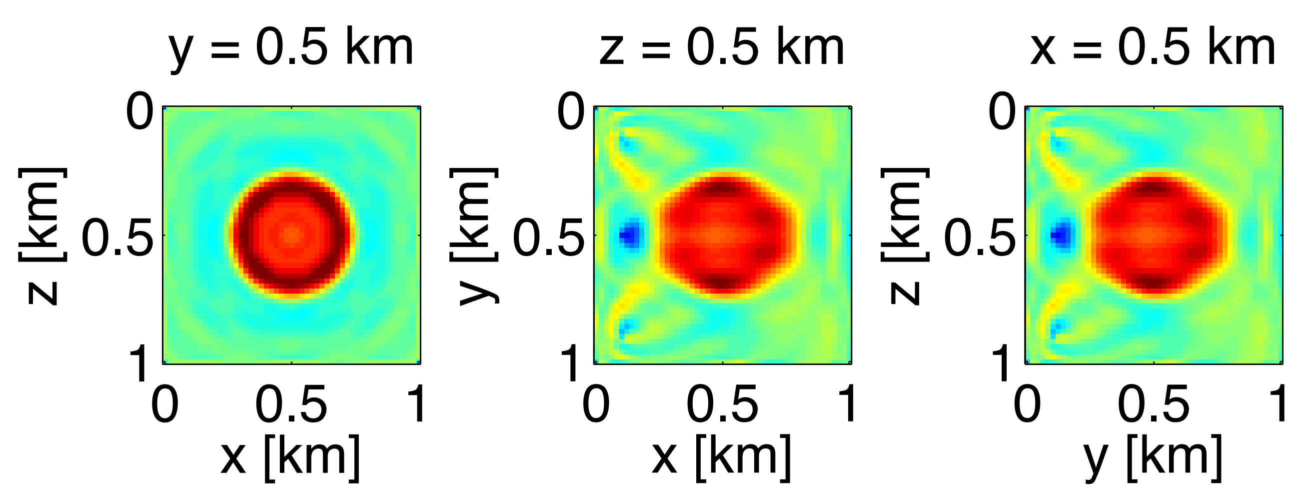
Edam model

$$b_0 = 9, \beta = 0, \epsilon_{\min} = 10^{-6}, n_{\text{iter}} = 10$$

$$\epsilon_0 = 10^{-3}, \alpha = 1,$$



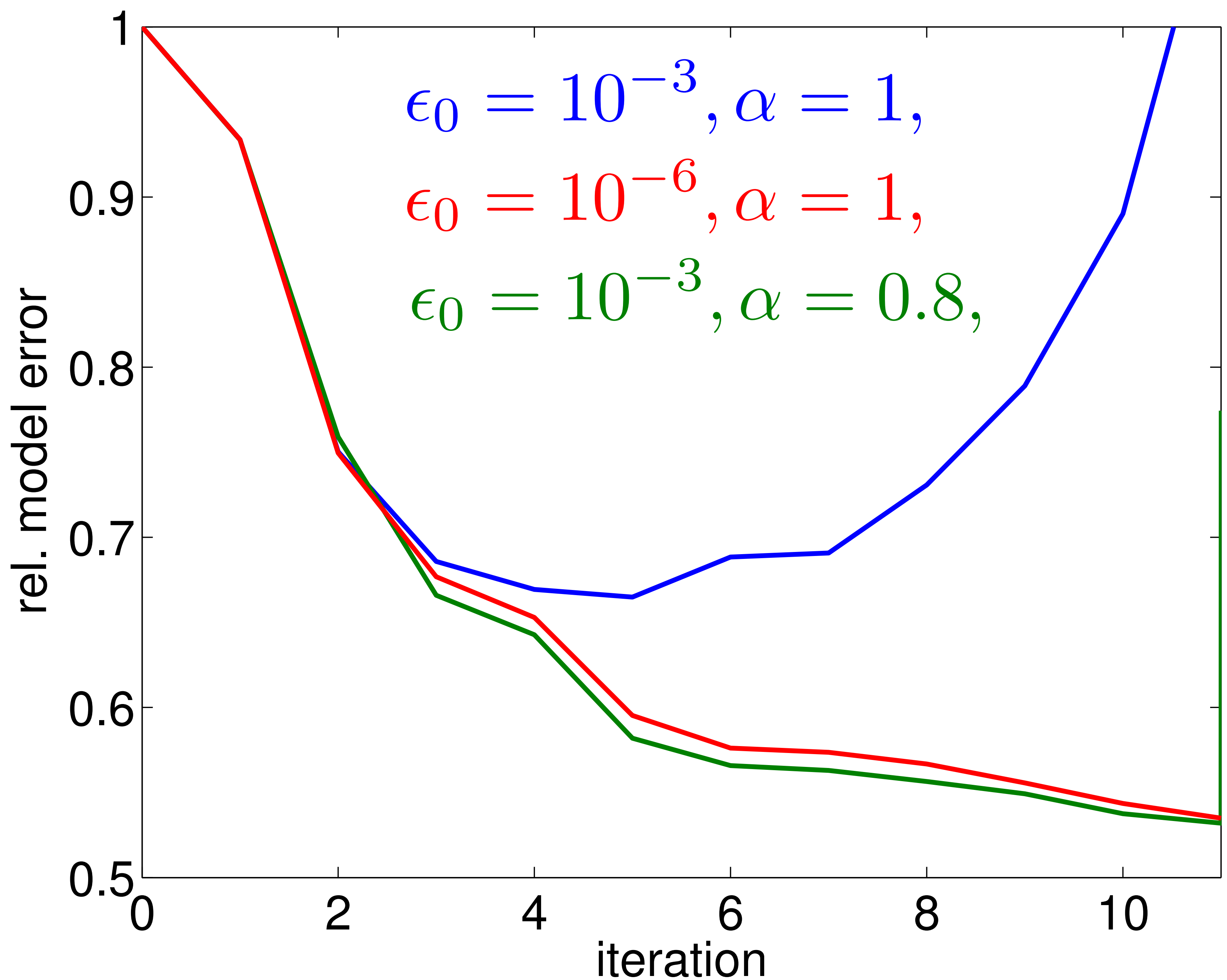




$$\epsilon_0 = 10^{-6}, \alpha = 1,$$

$$\epsilon_0 = 10^{-3}, \alpha = 0.8,$$

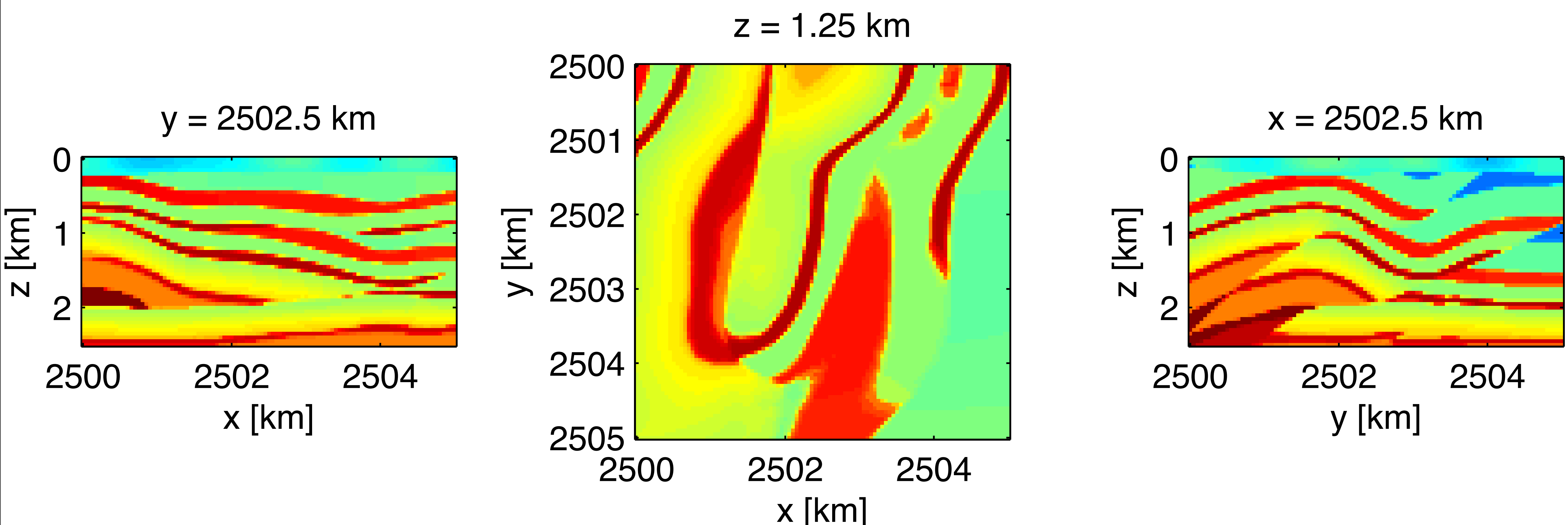
Edam model



$\epsilon_0 = 10^{-3}, \alpha = 1,$
$\epsilon_0 = 10^{-6}, \alpha = 1,$
$\epsilon_0 = 10^{-3}, \alpha = 0.8,$
rel. model error
iteration
1
0.9
0.8
0.7
0.6
0.5
0
2
4
6
8
10

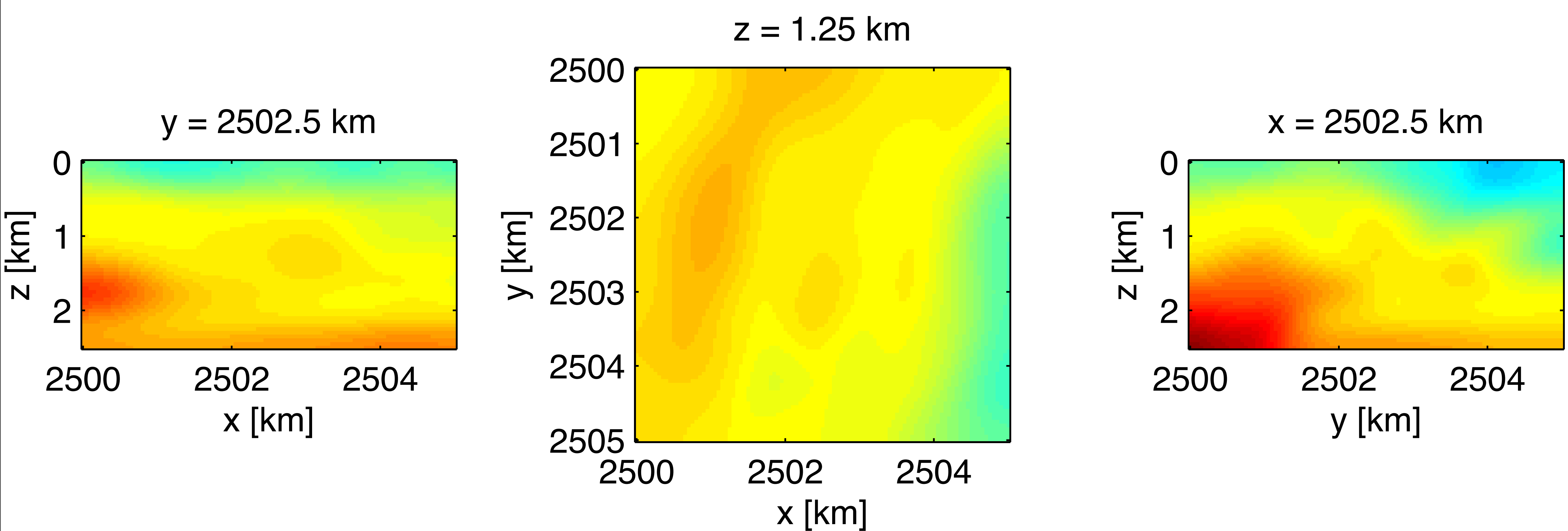
Overthrust model

101x101x51 gridpoints, f = 4



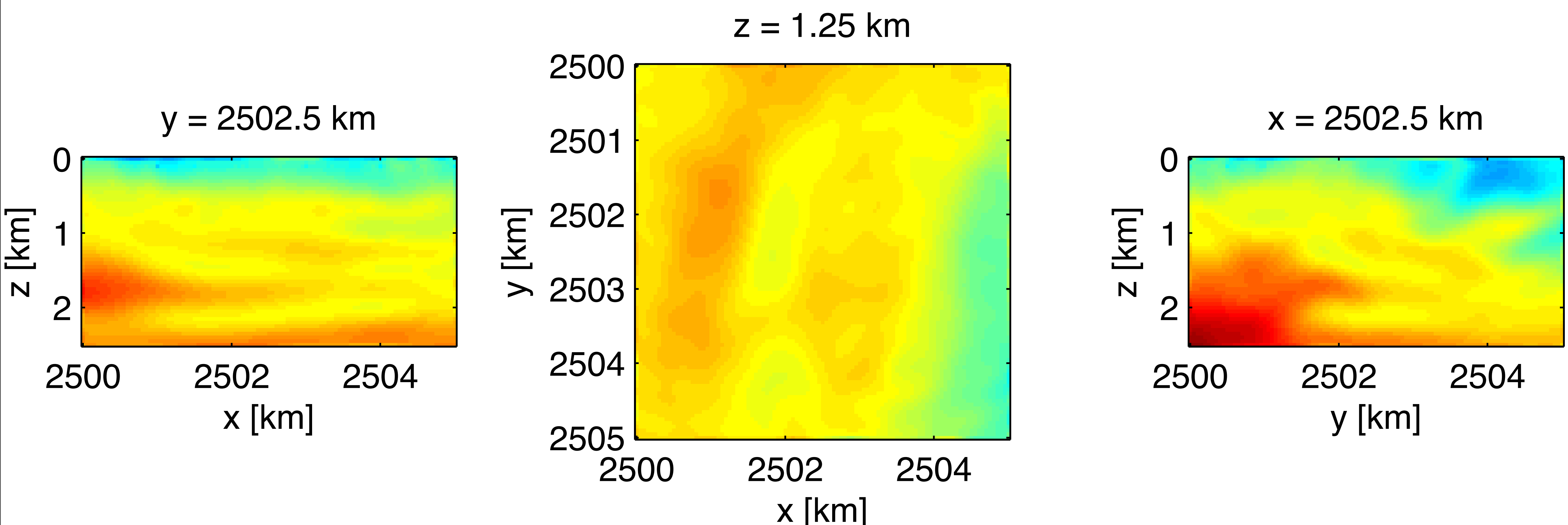
Overthrust model

initial model



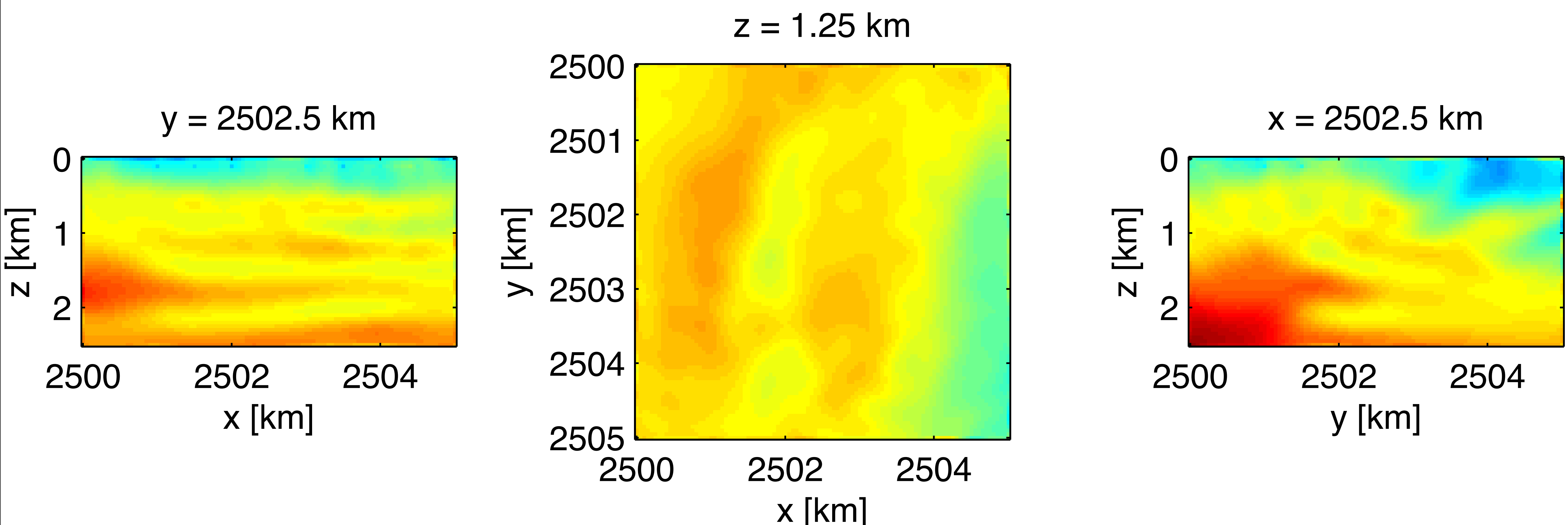
Overthrust model

$$b_0 = 1, \beta = 0, \epsilon_0 = 10^{-3}, \epsilon_{\min} = 10^{-6}, \alpha = 1, n_{\text{iter}} = 100$$



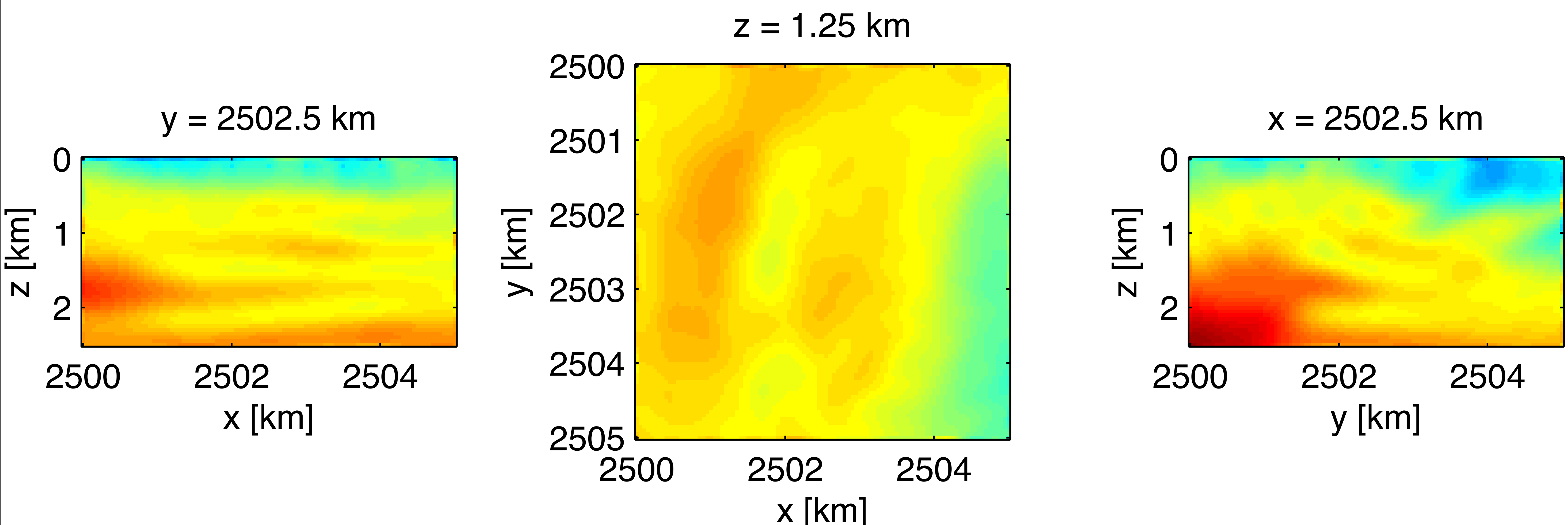
Overthrust model

$$b_0 = 5, \beta = 0, \epsilon_0 = 10^{-3}, \epsilon_{\min} = 10^{-6}, \alpha = 1, n_{\text{iter}} = 20$$

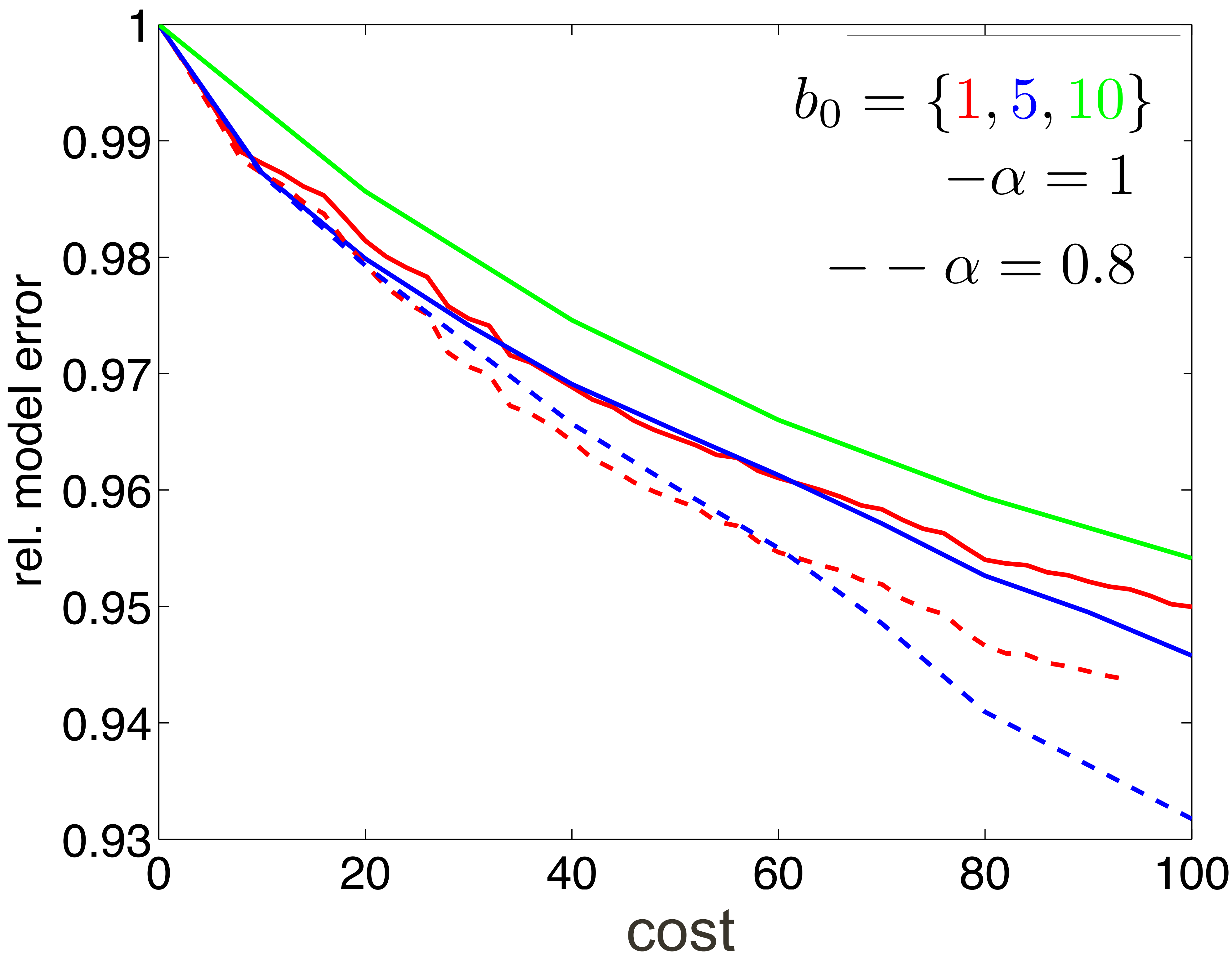


Overthrust model

$$b_0 = 10, \beta = 0, \epsilon_0 = 10^{-3}, \epsilon_{\min} = 10^{-6}, \alpha = 1, n_{\text{iter}} = 10$$



Overthrust model



Summary

- Generic preconditioner for Helmholtz with controllable accuracy
- efficient optimization strategy that can exploit inaccurate gradients
- good results by working on small (random) subsets of the data

- further computational aspects discussed by Art Petrenko (now) and me (tomorrow)
- Extension to elastic discussed by Bas Peters

SLIM

Acknowledgements

Thank you!

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, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

Further reading

T. van Leeuwen, 2012 - Fourier analysis of the CGMN method for solving the Helmholtz equation. ArXiv:1210:2644.

T. van Leeuwen, 2012 - A parallel matrix-free framework for frequency-domain seismic modelling, imaging and inversion in Matlab. Submitted to SISC.

A.Y. Aravkin and T. van Leeuwen, 2012 - Estimating nuisance parameters in inverse problems. Inverse Problems.

A. Aravkin, M.P. Friedlander, F.J. Herrmann and T. van Leeuwen, 2012 - Robust inversion, dimensionality reduction and randomized sampling. Mathematical Programming.

T. van Leeuwen and F.J. Herrmann, 2012 - Fast waveform inversion without source encoding. Geophysical Prospecting.

B.M. Bell and J.V. Burke. Algorithmic differentiation of implicit functions and optimal values.Advances in Automatic Differentiation, 2008.
B.M. Bell, J.V. Burke, and A. Schumitzky. A relative weighting method for estimating parameters and variances in multiple data sets. Comp. Stat. & Data Analysis, 1996.
Gene Golub and Victor Pereyra. Separable nonlinear least squares: the variable projection method and its applications. Inverse Problems, 19(2):R1, 2003.
G.H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares which variables separate. SIAM J. Numer. Anal., 10(2):413–432, 1973.
M. R. Osborne. Separable least squares, variable projection, and the Gauss-Newton algorithm. Electronic Transactions on Numerical Analysis, 28(2):1–15, 2007.
R G Pratt, C Shin, and Gj Hicks. Gauss-newton and full newton methods in frequency-space seismic waveform inversion. Geophysical Journal International, 1998.
M. P. Friedlander and M. Schmidt. Hybrid deterministic-stochastic methods for data fitting. SIAM J. Scientific Computing, 34(3), 2012 (submitted April 2011)
D. P. Bertsekas and J. Tsitsiklis, Neuro-Dynamic Programming, Athena Scientific, Belmont,Mass, USA, 1st edition, 1996.
A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," SIAM J. Opt., 19(4): pp. 1574–1609, 2008.
D. P. Bertsekas and J. N. Tsitsiklis, "Gradient convergence in gradient methods with errors," SIAM Journal on Optimization, 10(3):pp. 627–642, 2000.