A *penalty* method for *PDE*-constrained optimization with applications to *wave*-equation based *seismic* inversion

Tristan van Leeuwen* and Felix J. Herrmann

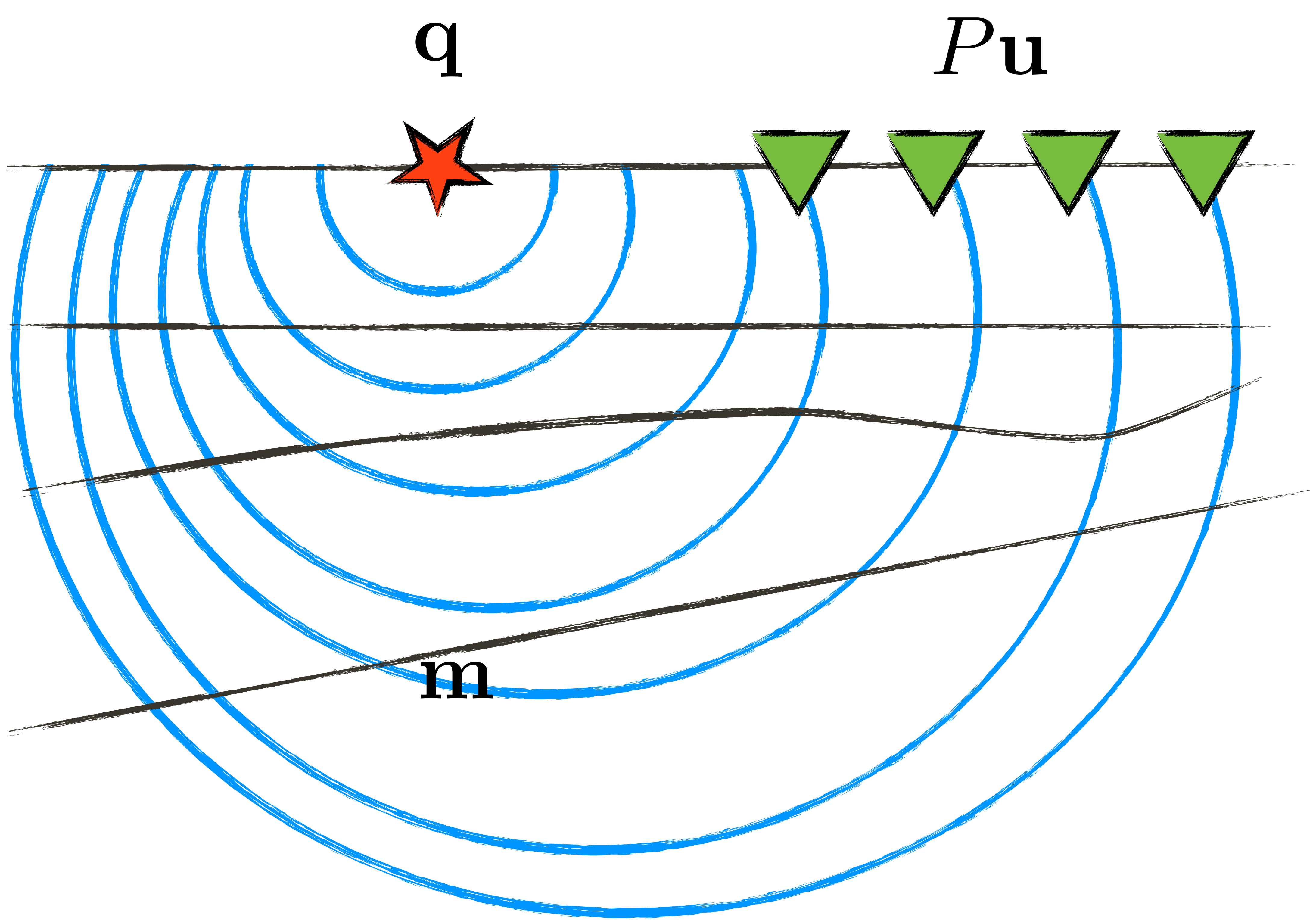
* now @ CWI Amsterdam

SLIM
University of British Columbia

van Leeuwen, T and Herrmann, F J (2013). Mitigating local minima in full-waveform inversion by expanding the search space. Geophysical Journal International.
van Leeuwen, T and Herrmann, F J (2013).A penalty method for PDE-constrained optimization. Submitted for publication
van Leeuwen, T and Herrmann, F J (2013). US Provisional Patent Application No. 61/815,533. A Penalty Method for PDE-Constrained Optimization with Applications to Wave-Equation Based Seismic Inversion

Seismic waveform inversion

Given observed data, $\mathbf{d}$, find *model* parameters, $\mathbf{m}$, and a *wavefield*, $\mathbf{u}$, such that, $P\mathbf{u} \approx \mathbf{d}$ and $A(\mathbf{m})\mathbf{u} \approx \mathbf{q}$



Observation

If we measure the wavefield *everywhere*, we can *trivially* solve

$$P\mathbf{u} = \mathbf{d}$$

Given $\mathbf{u}$ we *can* calculate the *medium* parameters by *solving*

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

for $\mathbf{m}$ *knowing* $\mathbf{q}$.

This leads to $$\min_{\mathbf{m}} ||A(\mathbf{m})P^{-1}\mathbf{d} - \mathbf{q}||_2^2$$

compared to $$\min_{\mathbf{m}} ||PA(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}||_2^2$$

Question

When P is *not* invertible, can we *reconstruct* a *wavefield* from the *observed* data that can act as a *proxy*?

For a given $\mathbf{m}$, we want the *proxy* wavefield to

I. *fit* the data–i.e., $P\mathbf{u} \approx \mathbf{d}$, and

II. *obey* the *wave*-equation–i.e., $A(\mathbf{m})\mathbf{u} \approx \mathbf{q}$

Why *not* use the least-squares solution of the *data*-augmented system?

$$\begin{pmatrix} P \\ A(\mathbf{m}) \end{pmatrix} \mathbf{u} \approx \begin{pmatrix} \mathbf{d} \\ \mathbf{q} \end{pmatrix}$$

Naive algorithm

1. Reconstruct the *unknown* wavefield by *solving*

$$\mathbf{u}_k = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} P \\ A(\mathbf{m}_k) \end{pmatrix} \mathbf{u} - \begin{pmatrix} \mathbf{d} \\ \mathbf{q} \end{pmatrix} \right\|_2^2$$

2. Reconstruct the *model* by *solving*

$$\mathbf{m}_{k+1} = \arg\min_{\mathbf{m}} \|A(\mathbf{m})\mathbf{u}_k - \mathbf{q}\|_2^2$$

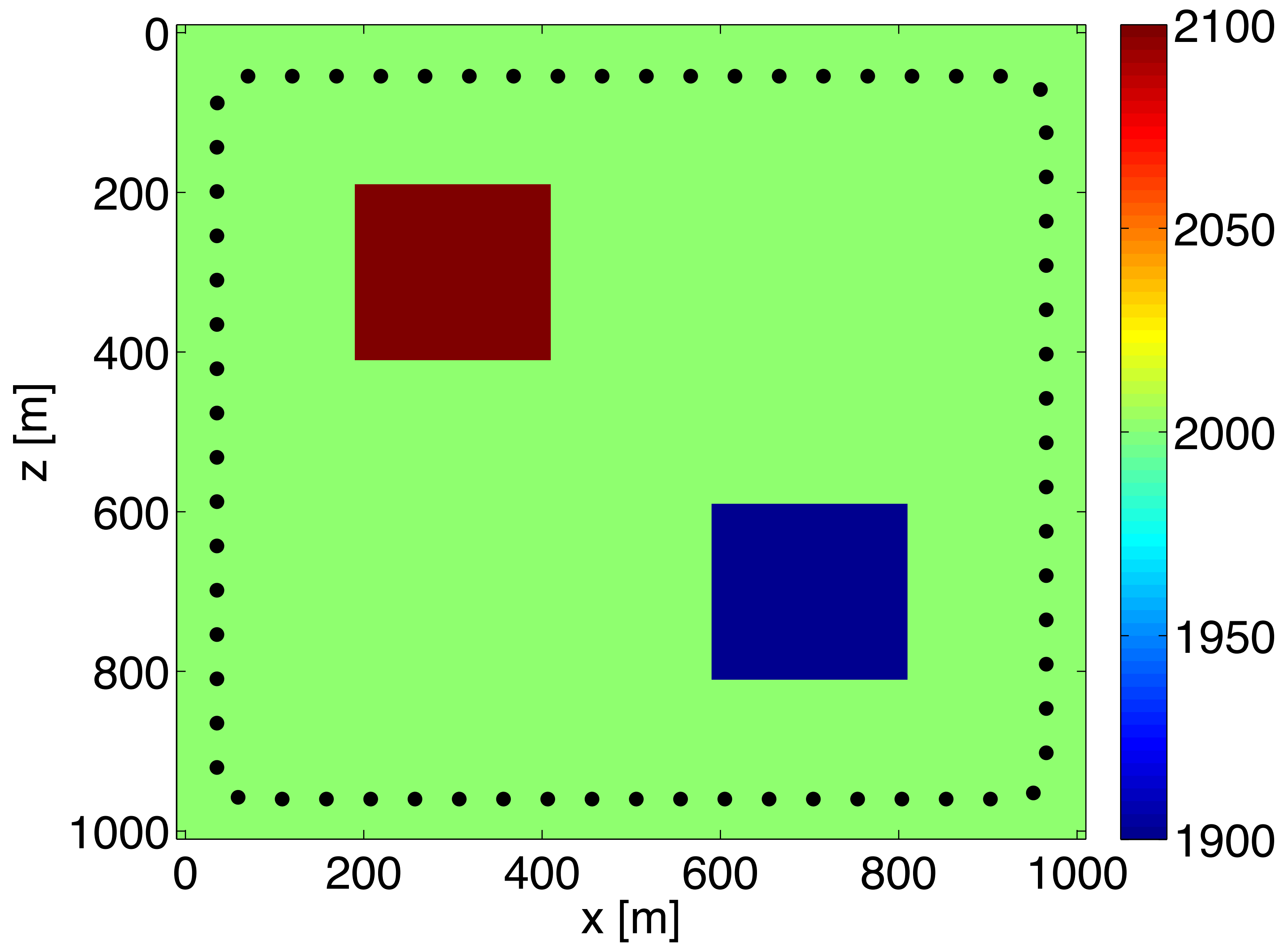
3. Repeat

Simple example

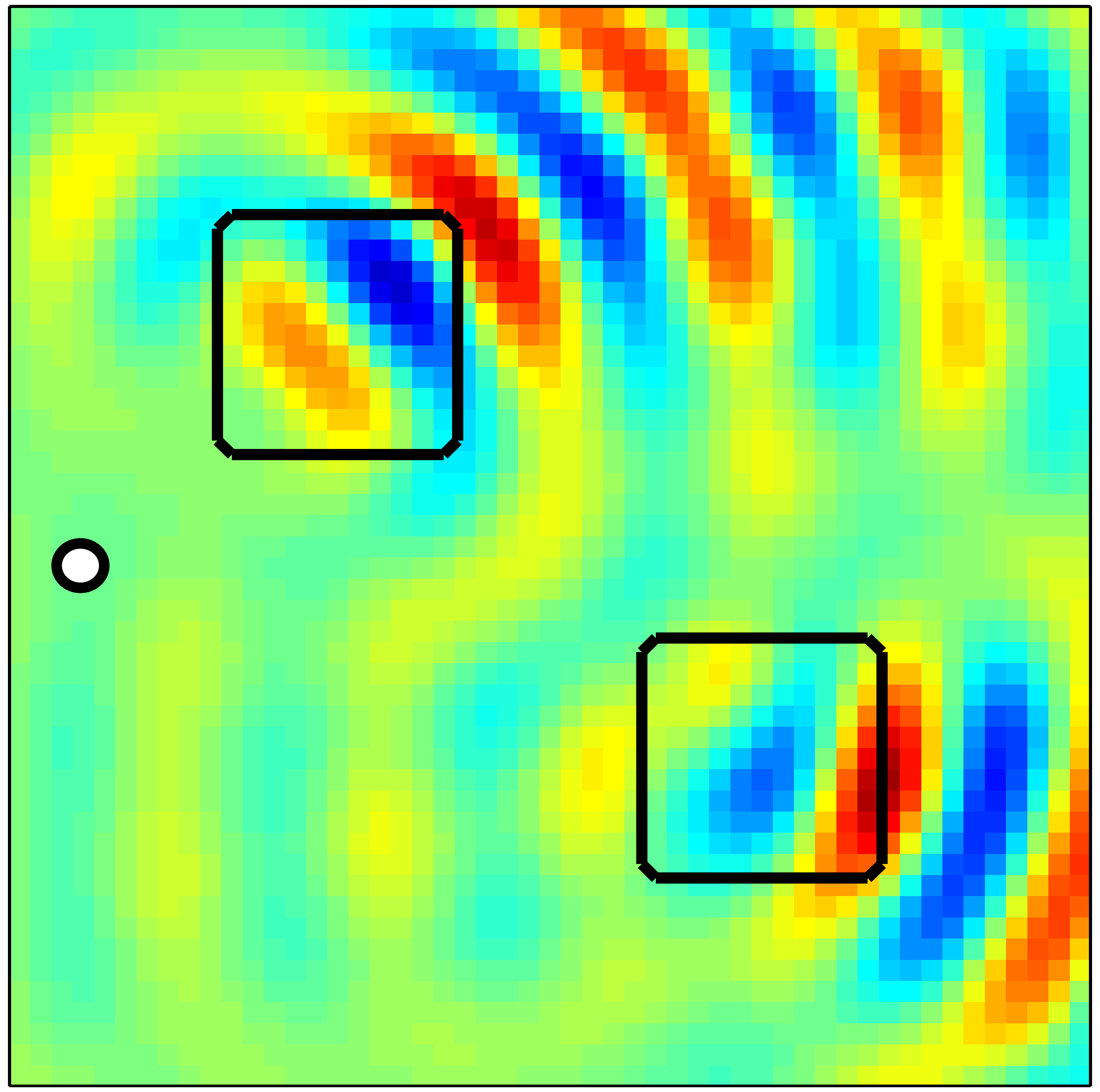
- Invert *monochromatic* data (10 Hz) with transducers *all* around
- *constant* velocity *initial* velocity *model*
- *5*-point discretization of Helmholtz system

$$A(\mathbf{m}) = \omega^2 \mathrm{diag}(\mathbf{m}) + L$$

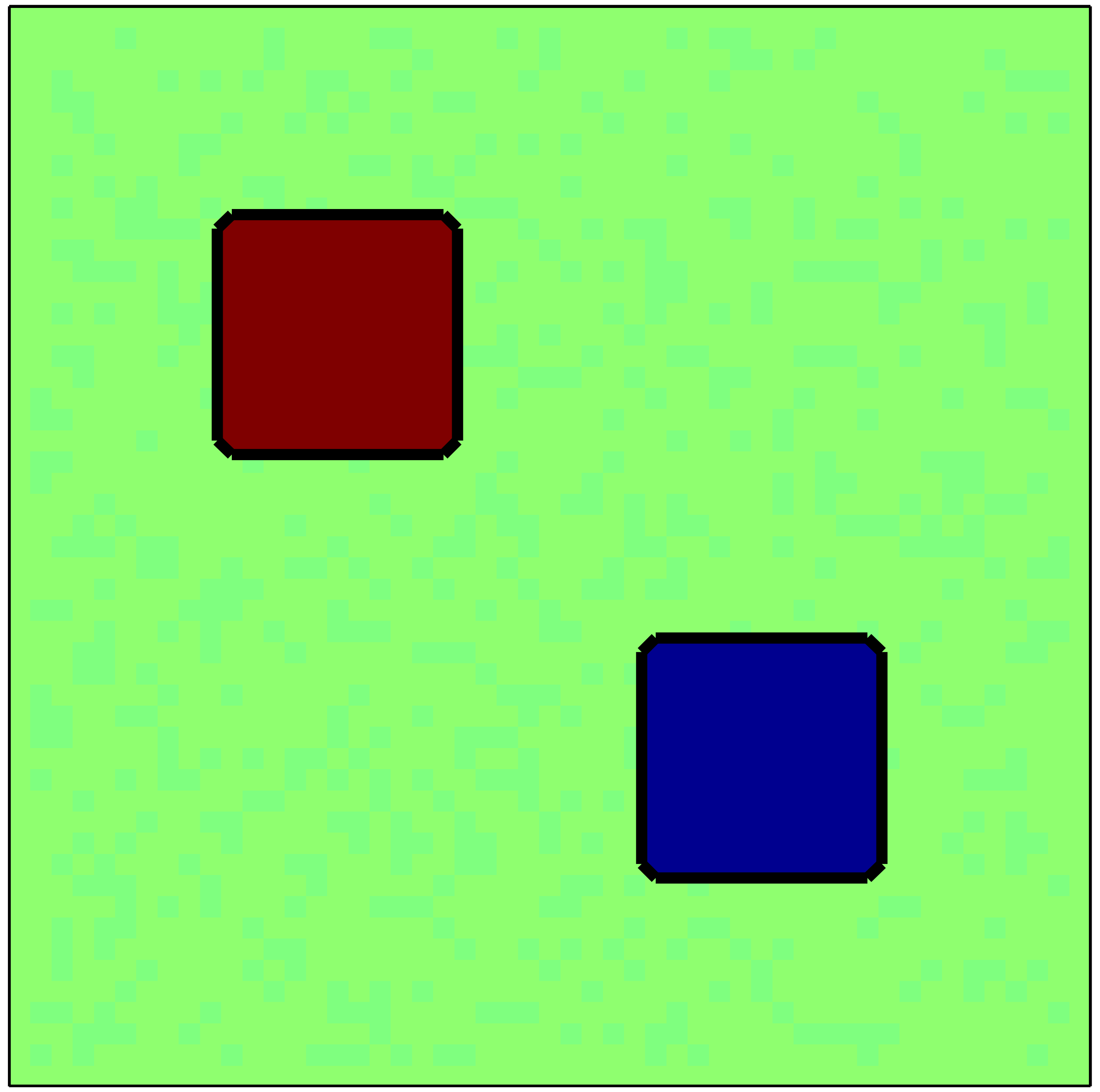
- two *subproblems* are both *linear* (in $\mathbf{m}$ and $\mathbf{u}$)



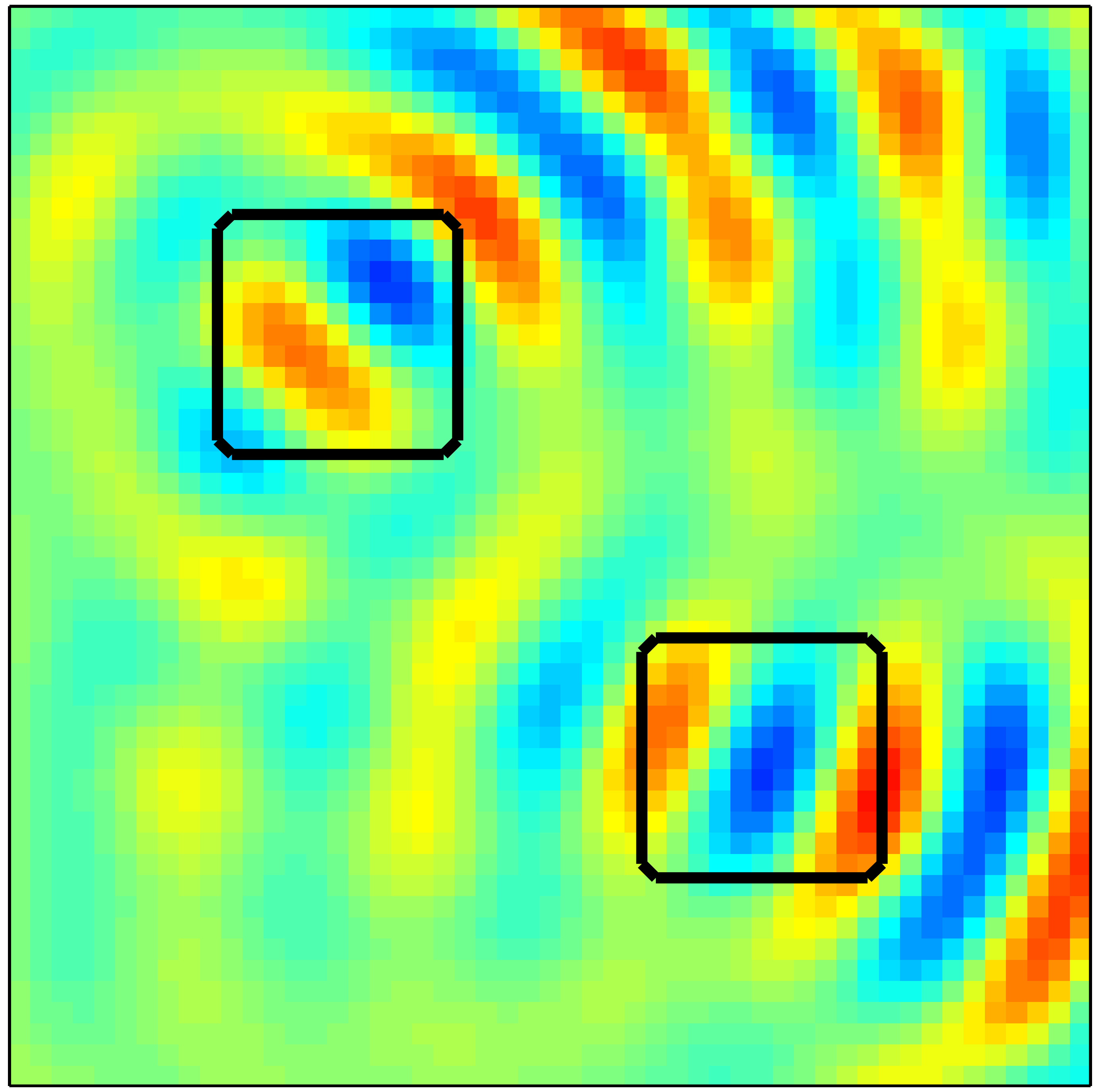
true wavefield



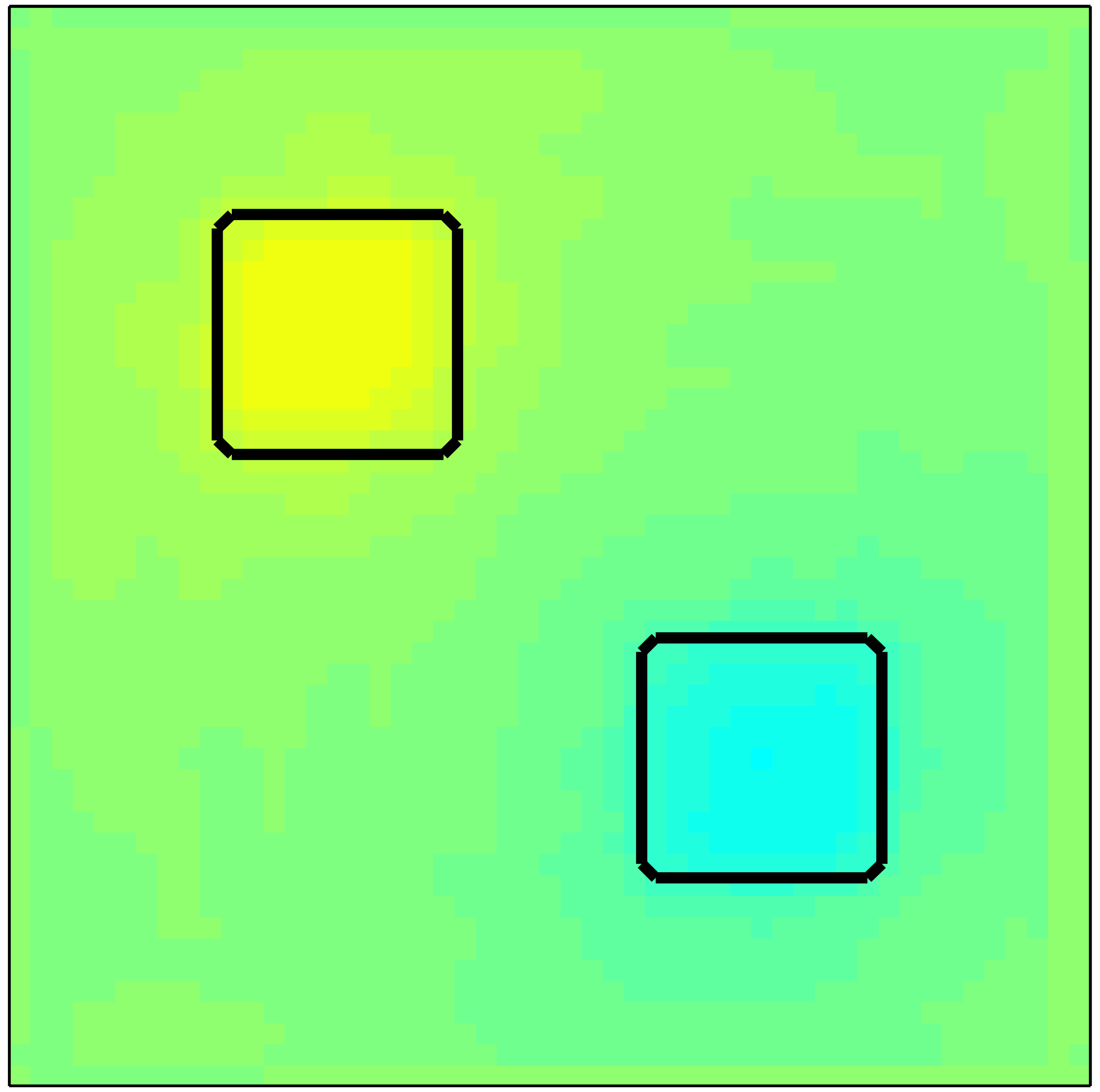
true model



iteration 1

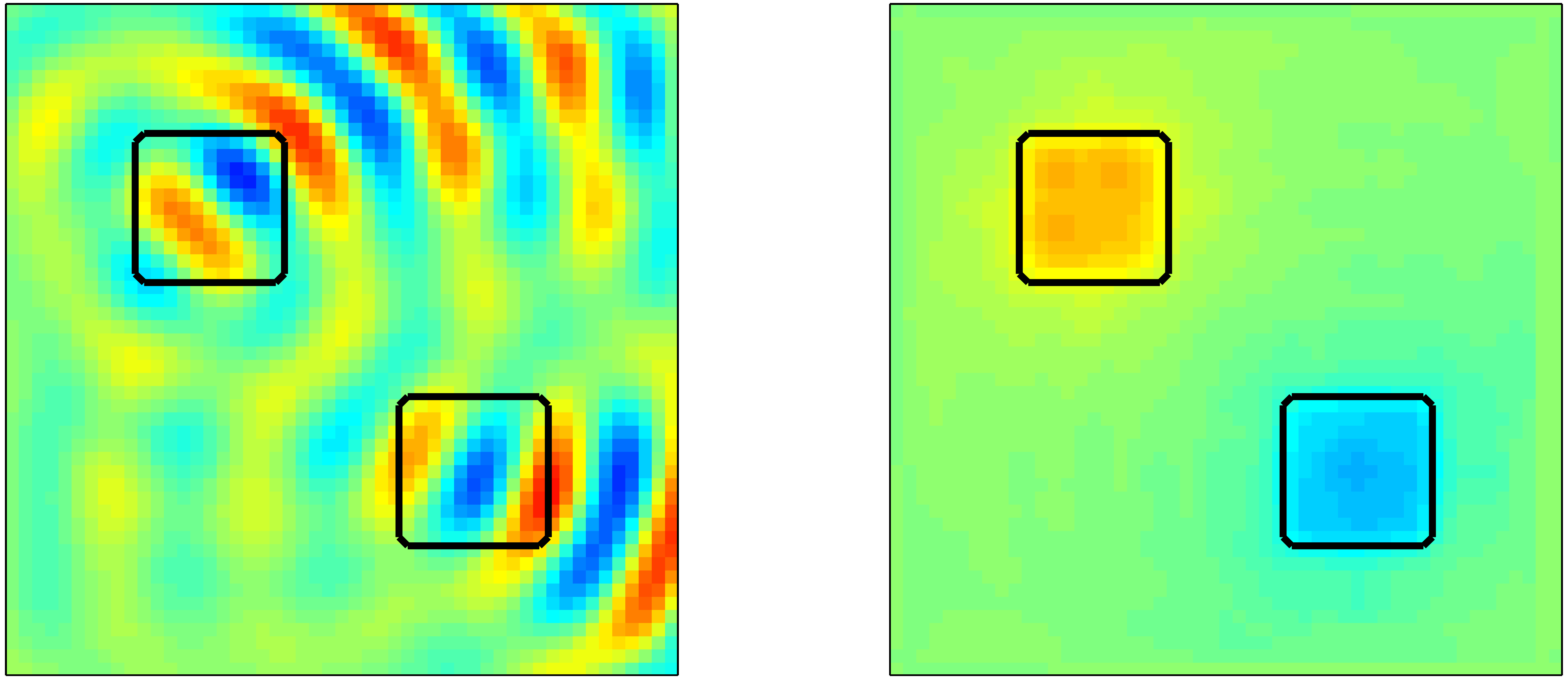


$$\mathbf{u}_0 = (P^*P + A_0^*A_0)^{-1}(P^*\mathbf{d} + A_0^*\mathbf{q})$$



$$\mathbf{m}_1 = \mathrm{diag}(|\mathbf{u}_0|^2)^{-1}\mathrm{diag}(\mathbf{u}_0)^*(\mathbf{q} - L\mathbf{u}_0)$$

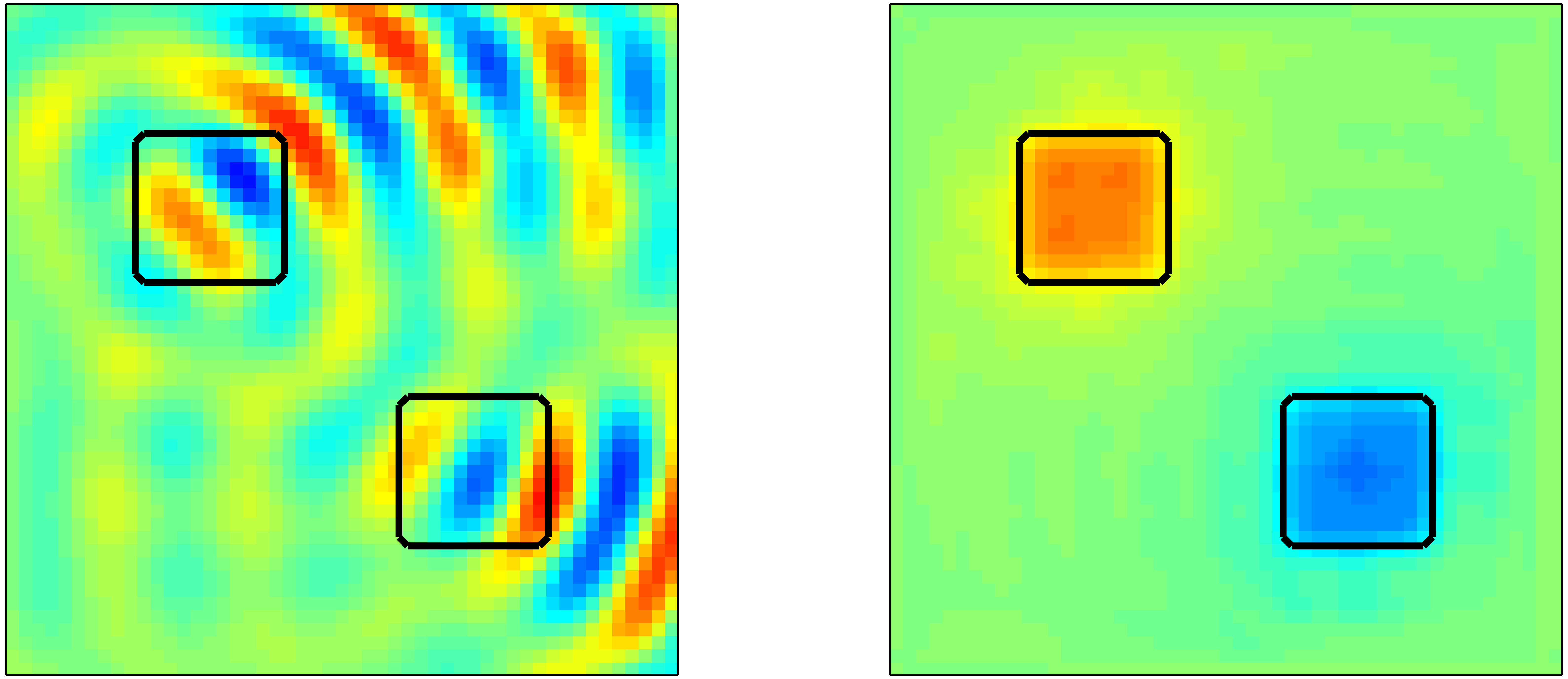
iteration 2



$$\mathbf{u}_1 = (P^*P + A_1^*A_1)^{-1}(P^*\mathbf{d} + A_1^*\mathbf{q})$$

$$\mathbf{m}_2 = \mathrm{diag}(|\mathbf{u}_1|^2)^{-1}\mathrm{diag}(\mathbf{u}_1)^*(\mathbf{q} - L\mathbf{u}_1)$$

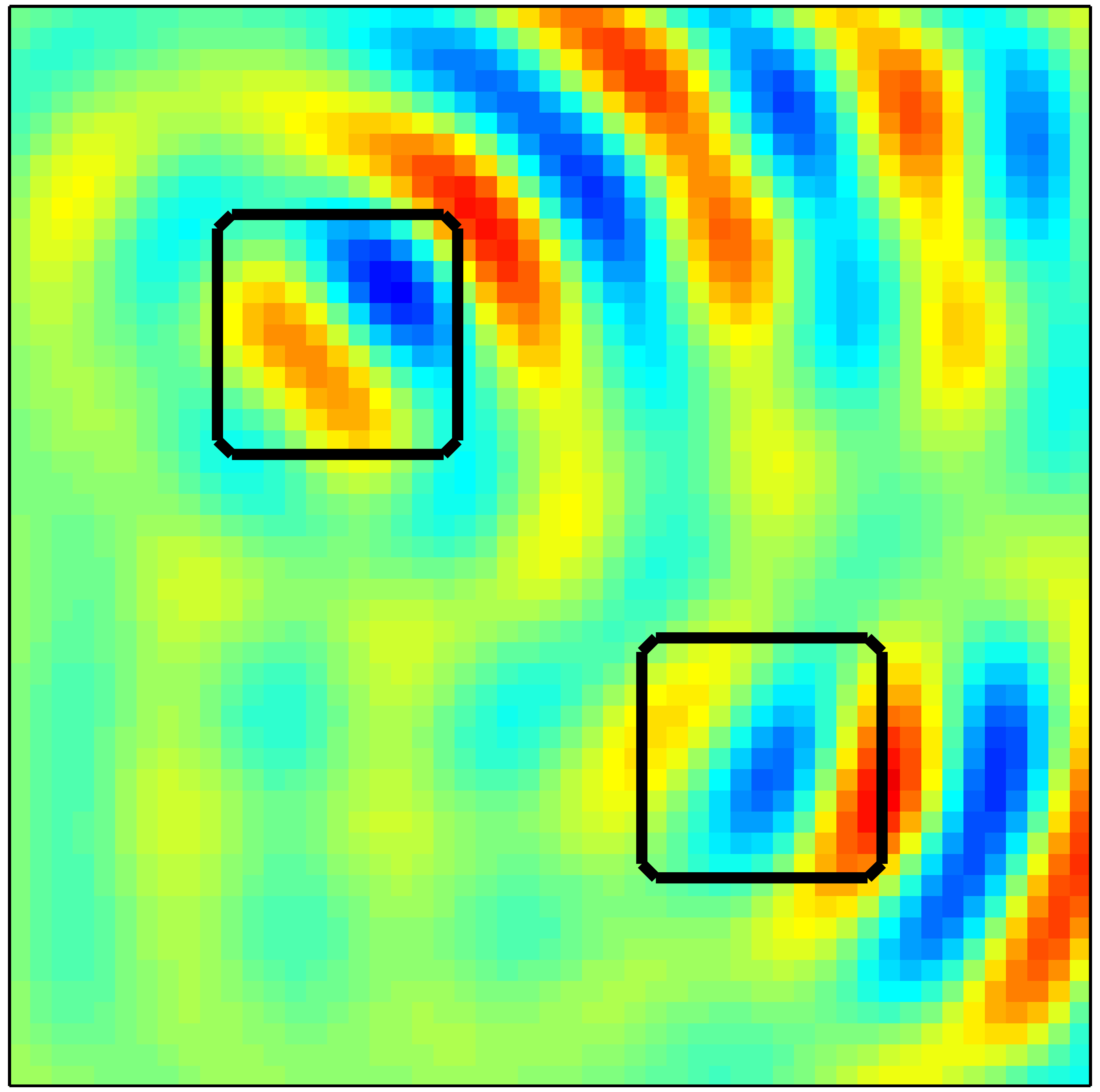
iteration 3

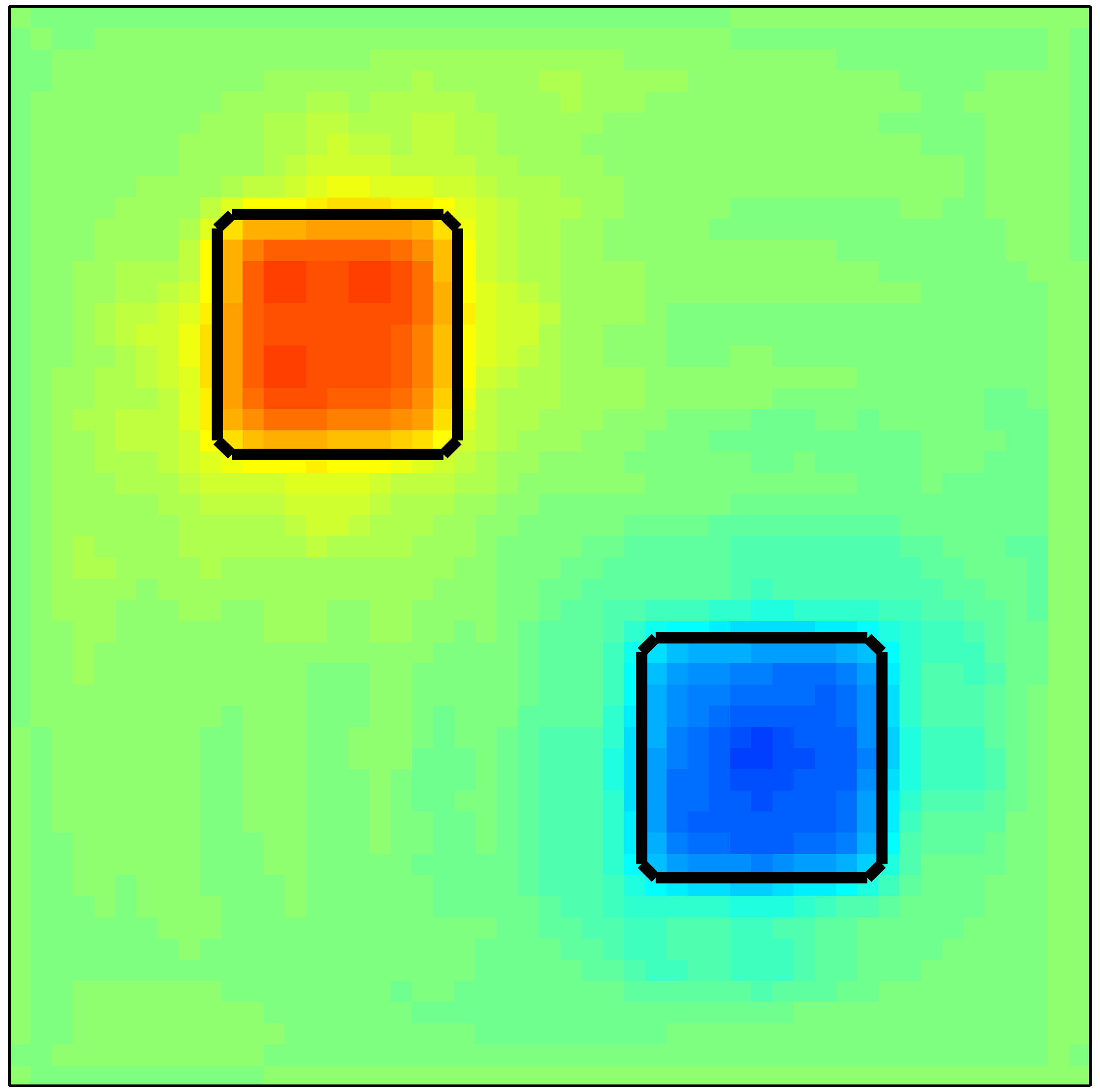


$$\mathbf{u}_2 = \left(P^*P + A_2^*A_2\right)^{-1}\left(P^*\mathbf{d} + A_2^*\mathbf{q}\right)$$

$$\mathbf{m}_3 = \mathrm{diag}(|\mathbf{u}_2|^2)^{-1}\mathrm{diag}(\mathbf{u}_2)^*\left(\mathbf{q} - L\mathbf{u}_2\right)$$

iteration 4

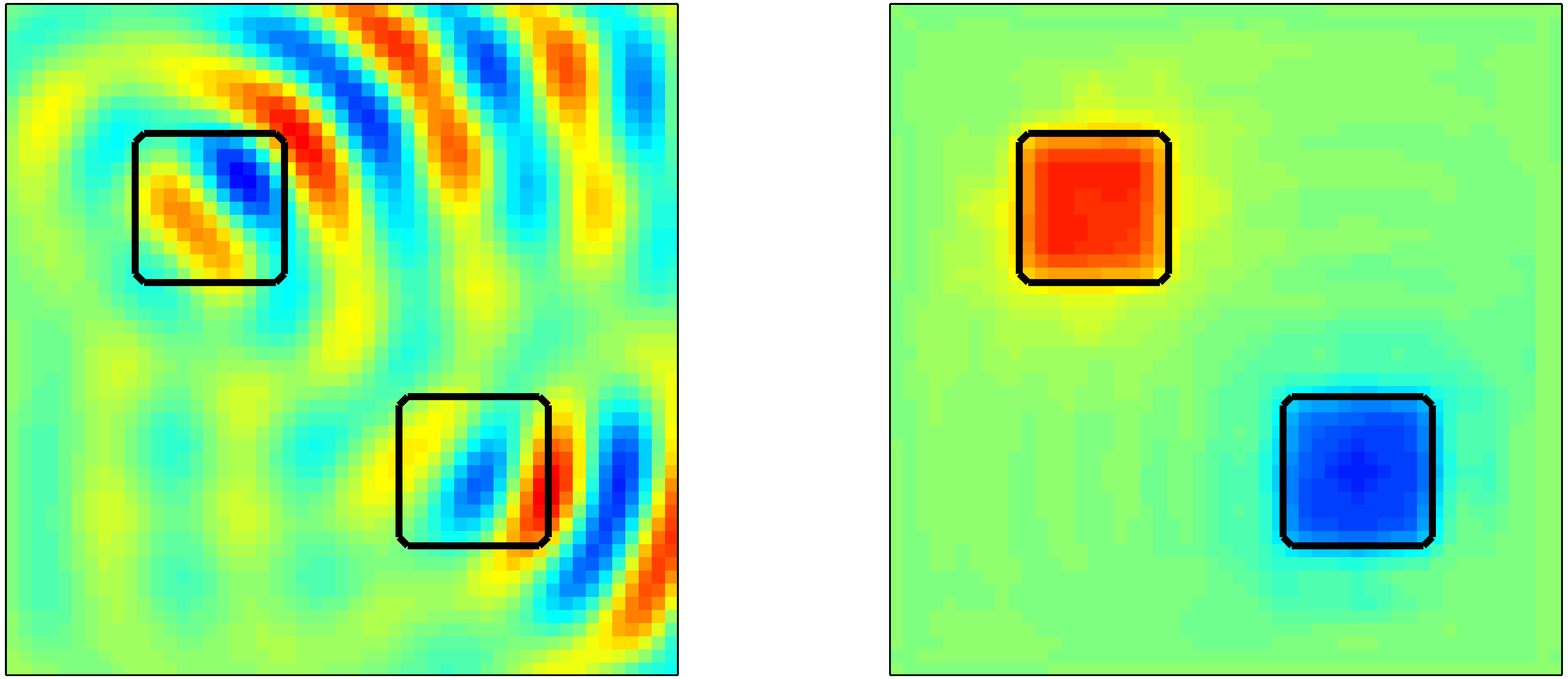




$$\mathbf{u}_3 = (P^*P + A_3^*A_3)^{-1}(P^*\mathbf{d} + A_3^*\mathbf{q})$$

$$\mathbf{m}_4 = \mathrm{diag}(|\mathbf{u}_3|^2)^{-1}\mathrm{diag}(\mathbf{u}_3)^*(\mathbf{q} - L\mathbf{u}_3)$$

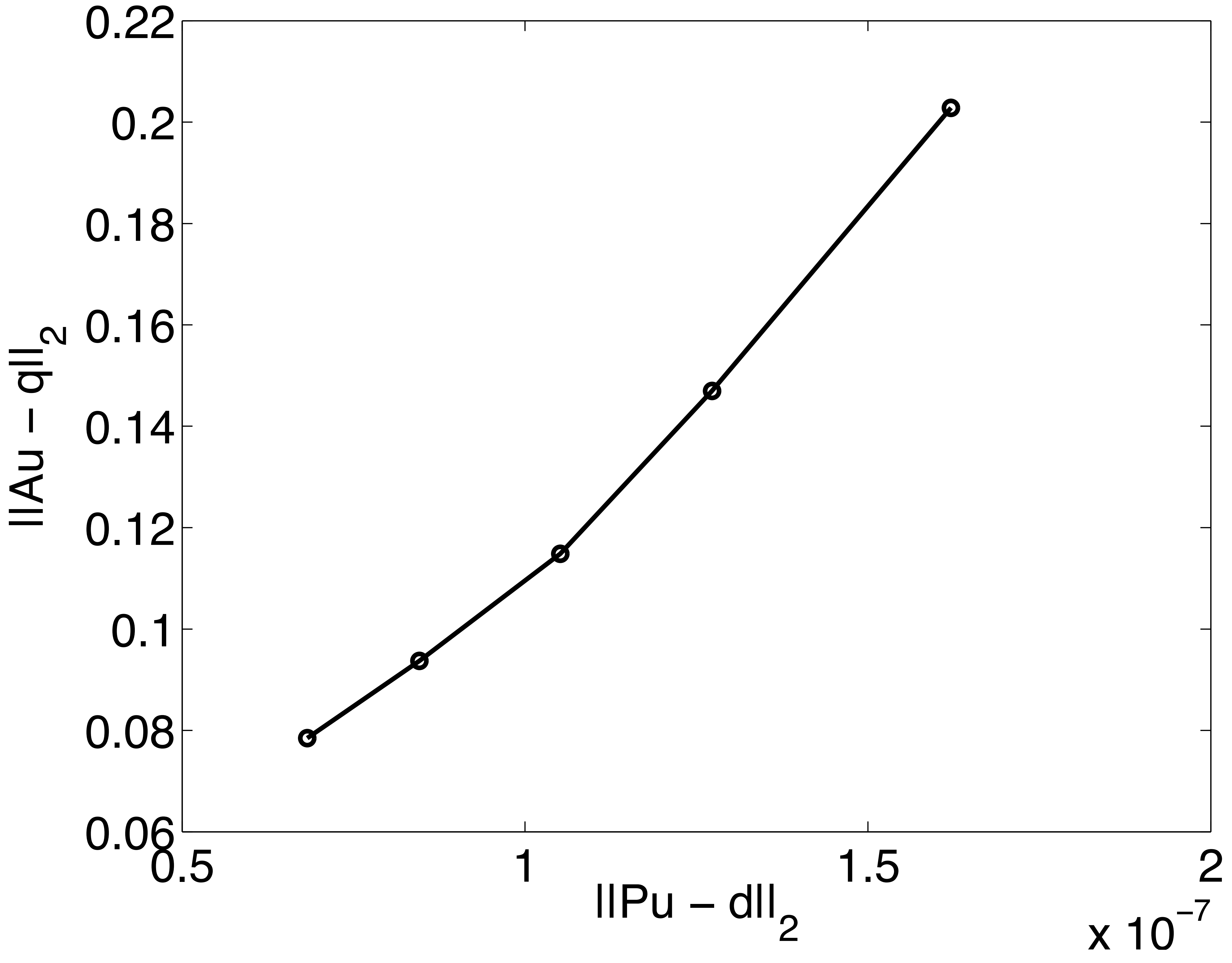
iteration 5



$$\mathbf{u}_4 = \left(P^*P + A_4^*A_4\right)^{-1}\left(P^*\mathbf{d} + A_4^*\mathbf{q}\right)$$

$$\mathbf{m}_5 = \mathrm{diag}(|\mathbf{u}_4|^2)^{-1}\mathrm{diag}(\mathbf{u}_4)^*\left(\mathbf{q} - L\mathbf{u}_4\right)$$

Convergence
in terms of *data*-fit & *PDE*-fit



Outline

PDE-constrained optimization

- *full* & *reduced*-space approaches
- KKT system & adjoint formulation

Penalty formulation

- *augmented* system
- gradient & GN Hessian

Numerical examples

Future work & conclusions

PDE-*constrained* optimization
multi-experiment data

$$\min_{\mathbf{m},\mathbf{u}} \sum_{i=1}^{M} \frac{1}{2}\|P_i\mathbf{u}_i - \mathbf{d}_i\|_2^2 \quad \text{subject to} \quad A_i(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i, \quad i = 1\cdots M,$$

- wavefields collected in $\mathbf{u} = [\mathbf{u}_1; \cdots ; \mathbf{u}_M]$ are typically *much bigger* than $\mathbf{m}$
- *solving* these problems for *large* models & large *M* is
  - *computationally* expensive
  - requires *lots* of *RAM* to store wavefields for *each* source

All-at-once approach [Haber & Asher ’01]

Cast into *Lagrangian form*

adjoint wavefields

$$\min_{\mathbf{m},\mathbf{u},\mathbf{v}} \mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = \sum_{i=1}^{M} \frac{1}{2}\|P_i\mathbf{u}_i - \mathbf{d}_i\|_2^2 + \langle \mathbf{v}_i, A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\rangle$$

Solve by *iterations* on *sparse* KKT system

$$\begin{pmatrix} * & * & G^* \\ * & P^*P + * & A \\ G & A & 0 \end{pmatrix} \begin{pmatrix} \delta\mathbf{m} \\ \delta\mathbf{u} \\ \delta\mathbf{v} \end{pmatrix} = - \begin{pmatrix} G^*\mathbf{v} \\ A^*\mathbf{v} + P^*(P\mathbf{u} - \mathbf{d}) \\ A\mathbf{u} - \mathbf{q} \end{pmatrix}$$

* higher order terms

All-at-once approach

Involves actions of *block*-diagonal matrices $A,\ P,\ G$ with

$$G_i(\mathbf{m}, \mathbf{u}_i) = \frac{\partial A_i(\mathbf{m})\mathbf{u}_i}{\partial \mathbf{m}}$$

Pros:

- *no* explicit PDE solves
- iterations w/ *sparse* (KKT) system
- GN Hessian is *sparse*
- search-space *expansion* – *mitigates* local *minima*
- *"more"* linear in $\mathbf{m}$

Cons:

- introduces *extra* variable
- *indefinite* system so *challenge* to *precondition*
- requires *simultaneous* access= storage of *forward & adjoint wavefields* for *all* sources
- That's a show *stopper*...

Reduced approach

[A. Tarantola & Valette ‘82; Plessix ’06; Epanomeritakis et al. ‘08]

For each $i = 1 \cdots M$ experiments *eliminate* the PDE *constraint*–i.e,.

$$\min_{\mathbf{m}} \phi_{\mathrm{red}}(\mathbf{m}) = \sum_{i=1}^{M} \frac{1}{2} \| P_i A_i(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i \|_2^2$$

Solve by *accumulating* gradients

$$\nabla \phi_{\mathrm{red}}(\mathbf{m}) = \sum_{i=1}^{M} G_i(\mathbf{m}, \mathbf{u}_i)^* \mathbf{v}_i$$

$$= \sum_{i=1}^{M} \omega_i^2 \mathrm{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

Reduced approach

Involves solutions of *forward* & *adjoint* for $i = 1 \cdots M$

$$\begin{aligned} A_i(\mathbf{m})\mathbf{u}_i &= \mathbf{q}_i \\ A_i(\mathbf{m})^*\mathbf{v}_i &= -P_i^*(P_i\mathbf{u}_i - \mathbf{d}_i) \end{aligned}$$

Accumulate reduced GN Hessian

$$\nabla^2\phi_{\mathrm{red}}(\mathbf{m}) \approx H_{GN} = \sum_{i=1}^{M} G_i^* A_i^{-*} P_i^T P_i A_i^{-1} G_i$$

Pros:

- *no* need to have *simultaneous* access to *all* wavefields
- *accumulate* gradients & Hessians
- *adjoint*-state method has *intuitive* interpretation

Cons:

- PDEs need to be solved *explicitly*
- Jacobian & GN Hessian are *dense* matrices
- *smaller* search space – *local* minima
- *highly* nonlinear in $\mathbf{m}$

Challenge

Can we find an *alternative* approach that has *best* of both *worlds*?

- *expanded* search space as in all-at-once method w/o needing *simultaneous* access to *all* wavefields for each *iteration*
- *sparse* matrices
- *mundane* nonlinearity
- *accumulation* of *gradients* & GN *Hessians*

Probabilistic argument

In Tarantola's formulation the least-squares misfit ϕ_{reduced}

- allows for *errors* in the data *residual*
- *but,* considers wave-equation *physics* as *infallible*

Instead, we consider both *physics* & *observed* data as *fallible*...

- *least-squares* penalties for *model* & *data* errors

New approach [van Leeuwen & FJH '13]

Penalty formulation:

$$\min_{\mathbf{m},\mathbf{u}} \phi_\lambda(\mathbf{m},\mathbf{u}) = \sum_{i=1}^{M} \frac{1}{2}\|P_i\mathbf{u}_i - \mathbf{d}_i\|_2^2 + \frac{\lambda^2}{2}\|A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\|_2^2$$

Solution coincides [Bertsekas '96] with solution of original *constrained* formulation as $\lambda \uparrow \infty$

Penalty approach

For each $i = 1 \cdots M$ *eliminate* the wavefield by solving $\nabla_{\mathbf{u}_i} \phi_\lambda(\mathbf{m}, \mathbf{u}_i) = 0$

Define *modified* objective

$$\phi_{\text{pen}}(\mathbf{m}) = \overline{\phi}_\lambda(\mathbf{m}, \overline{\mathbf{u}}(\mathbf{m}))$$

Calculate derivates

$$\nabla \phi_{\text{pen}} = \nabla_{\mathbf{m}} \overline{\phi}_\lambda$$

$$\nabla^2 \phi_{\text{pen}} = \nabla^2_{\mathbf{m}} \overline{\phi}_\lambda - \nabla^2_{\mathbf{m},\mathbf{u}} \overline{\phi}_\lambda \left(\nabla^2_{\mathbf{u}} \overline{\phi}_\lambda \right)^{-1} \nabla^2_{\mathbf{u},\mathbf{m}} \overline{\phi}_\lambda$$

(*all* evaluated at *optimal* $\overline{\mathbf{u}}$)

SLIM

Penalty approach

A *local* minimizer of $\phi_{\mathrm{pen}}(\mathbf{m})$, together with the corresponding $\overline{\mathbf{u}}$ are also *local* minimizers of $\overline{\phi}_{\lambda}(\mathbf{m}, \overline{\mathbf{u}}(\mathbf{m}))$ and *vice versa*.

Penalty approach

Solve for *wavefields* by *inverting*

$$\min_{\mathbf{u}_i} \left\| \begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_i - \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix} \right\|_2^2$$

Compute gradient–***no*** adjoint wavefield required!

$$\nabla \phi_{\mathrm{pen}}(\mathbf{m}) = G(\mathbf{m}, \overline{\mathbf{u}})^* (A(\mathbf{m})\overline{\mathbf{u}} - \mathbf{q})$$

GN–Hessian

$$\begin{aligned} \nabla^2 \phi_{\mathrm{pen}}(\mathbf{m}) = \lambda^2 G^* G \\ + \lambda^4 G^* A \left(P^* P + \lambda^2 A^* A\right)^{-1} A^* G \end{aligned}$$

Penalty approach

Accumulate gradients

$$\nabla_{\mathbf{m}}\phi_{\text{pen}}(\mathbf{m},\overline{\mathbf{u}}) = \sum_{i=1}^{M} G_i(\mathbf{m},\overline{\mathbf{u}}_i)^* \left(A_i(\mathbf{m})\overline{\mathbf{u}}_i - \mathbf{q}_i\right)$$

Accumulate (approximate for small λ) Gauss-Newton Hessians:

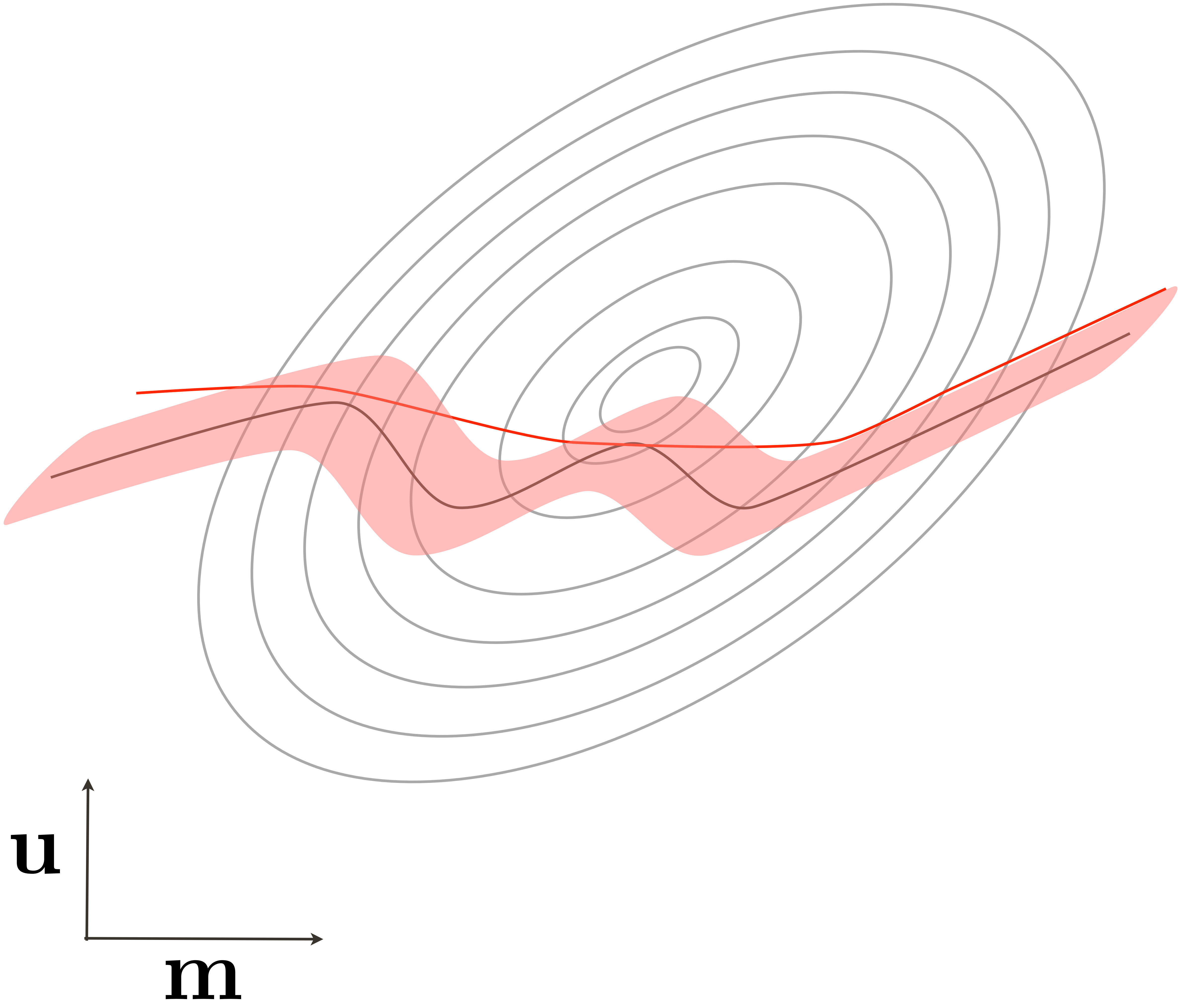
$$\nabla^2\phi_{\text{pen}}(\mathbf{m}) \approx \sum_{i=1}^{M} \lambda^2 G_i^* G_i$$

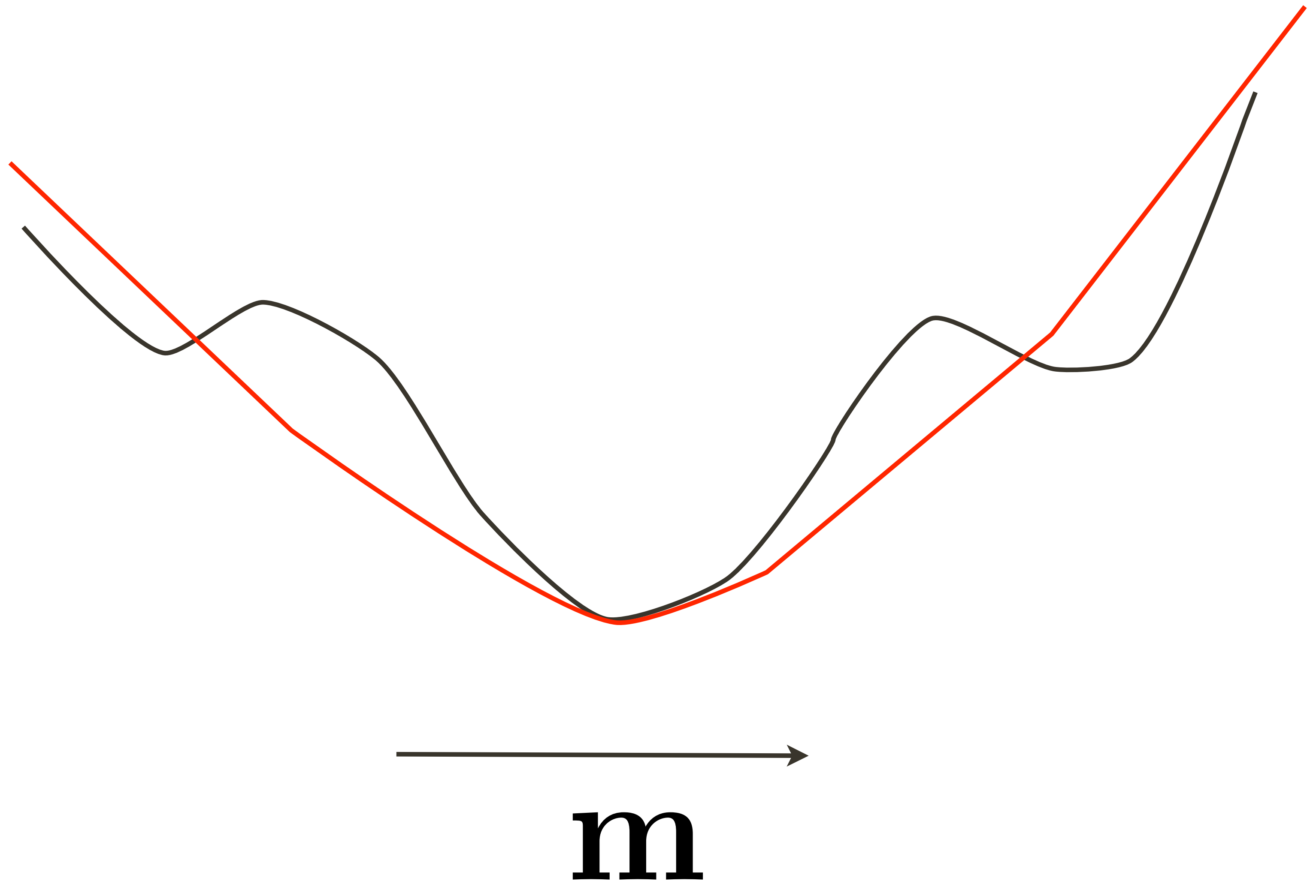
Penalty approach

Algorithm 1 FWI algorithm based on the penalty formulation

for $t = 0$ to N **do**
  $\mathbf{g}_t = 0$
  $H_t = 0$
  for $k = 1$ to N_f **do**
    for $l = 1$ to N_s **do**
      solve $\begin{pmatrix} \lambda_t A_k(\mathbf{m}_t) \\ P \end{pmatrix} \mathbf{u}_{kl} = \begin{pmatrix} \lambda_t \mathbf{q}_{kl} \\ \mathbf{d}_{kl} \end{pmatrix}$
      $\mathbf{g}_t = \mathbf{g}_t + \lambda_t^2 G_k(\mathbf{m}_t, \mathbf{u}_{kl})^* (A_k(\mathbf{m}_t)\mathbf{u}_{kl} - \mathbf{q}_{kl})$
      $H_t = H_t + \lambda_t^2 G_k(\mathbf{m}_t, \mathbf{u}_{kl})^* G_k(\mathbf{m}_t, \mathbf{u}_{kl})$
    end for
  end for
  solve $H_t \Delta\mathbf{m}_t = -\mathbf{g}_t$
  update $\mathbf{m}_{t+1} = \mathbf{m}_t + \Delta\mathbf{m}_t$
  update λ_t.
end for

Penalty vs. Reduced





Penalty vs. Reduced

Pros:

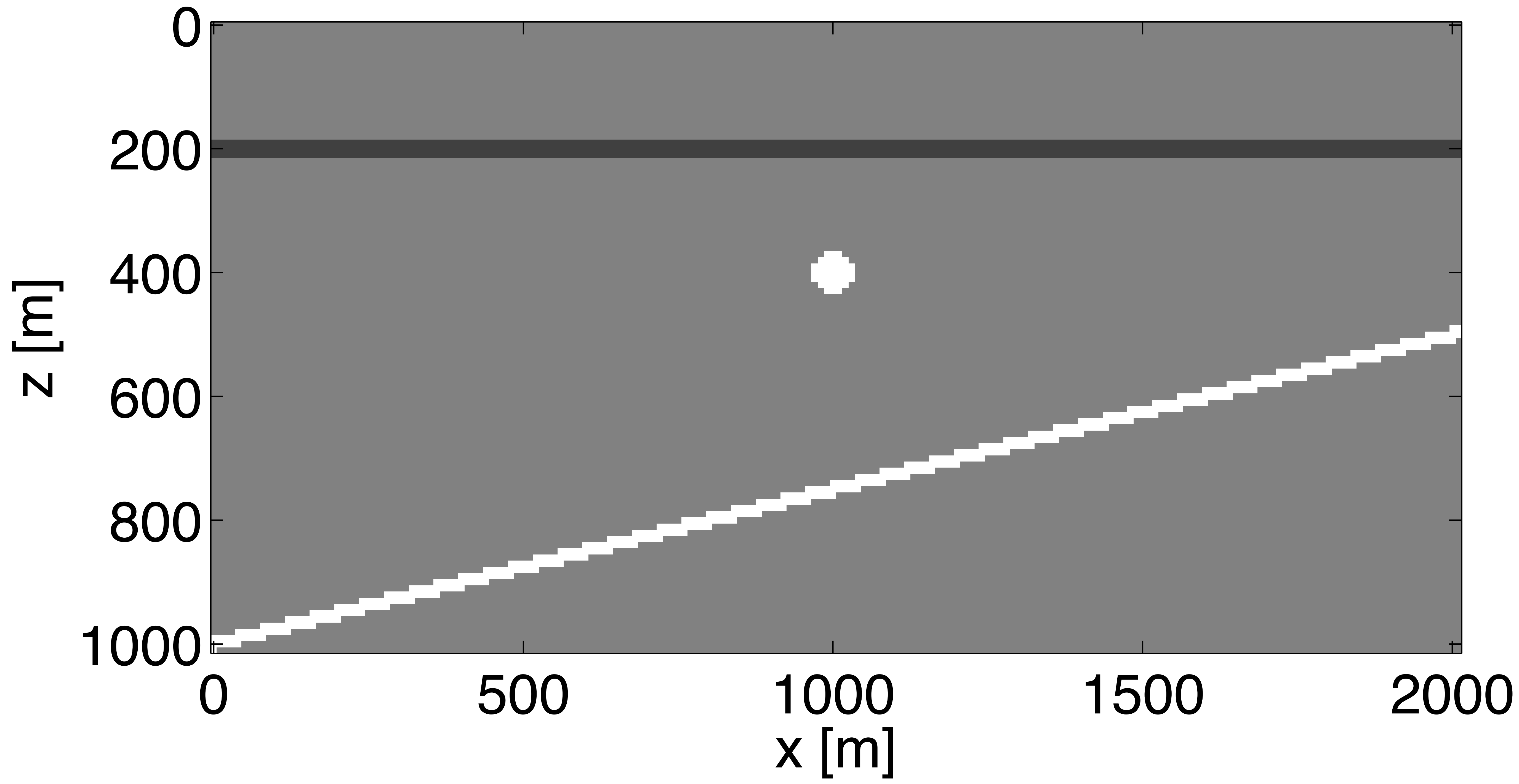
- *expansion* of *search* space
- *no* need to have *simultaneous* access to *all* wavefields for *each* iteration
- *no* adjoint wavefields
- *sparse* approximation of Hessian at *small* λ
- less *non*-linear in $\mathbf{m}$

Cons:

- *accurate* solutions of overdetermined *data*-augmented system
- cooling *strategy* for λ

Numerical examples

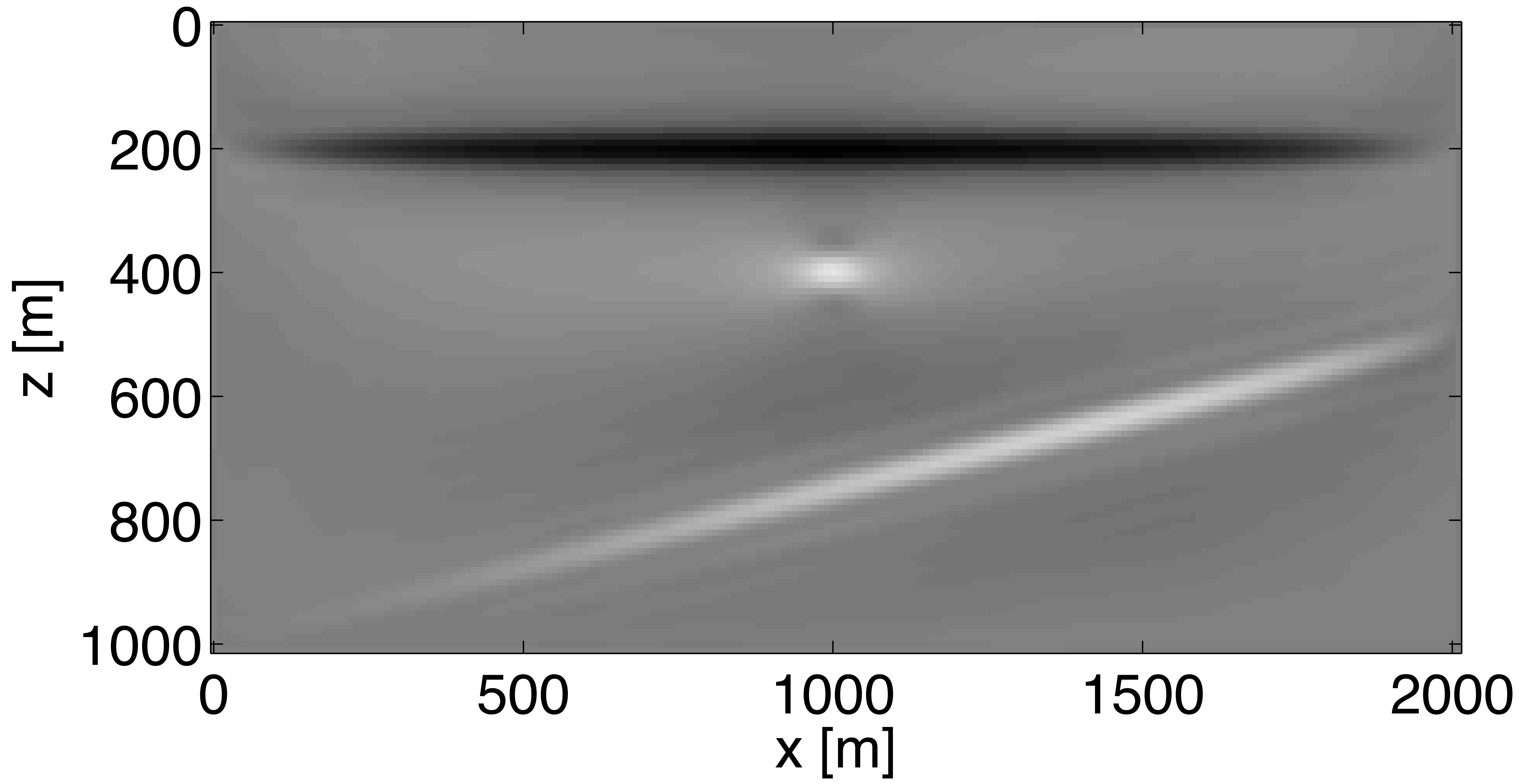
- Imaging
- Local minima
- Full-waveform inversion
- DC resistivity
- Optical tomography



Original image

Imaging

The *gradient* of the reduced objective yields an *image* of the subsurface..



RTM-migration

Imaging

***Conventional* RTM:**

1. solve forward wave-equation
2. solve adjoint wave-equation
3. apply `imaging condition’

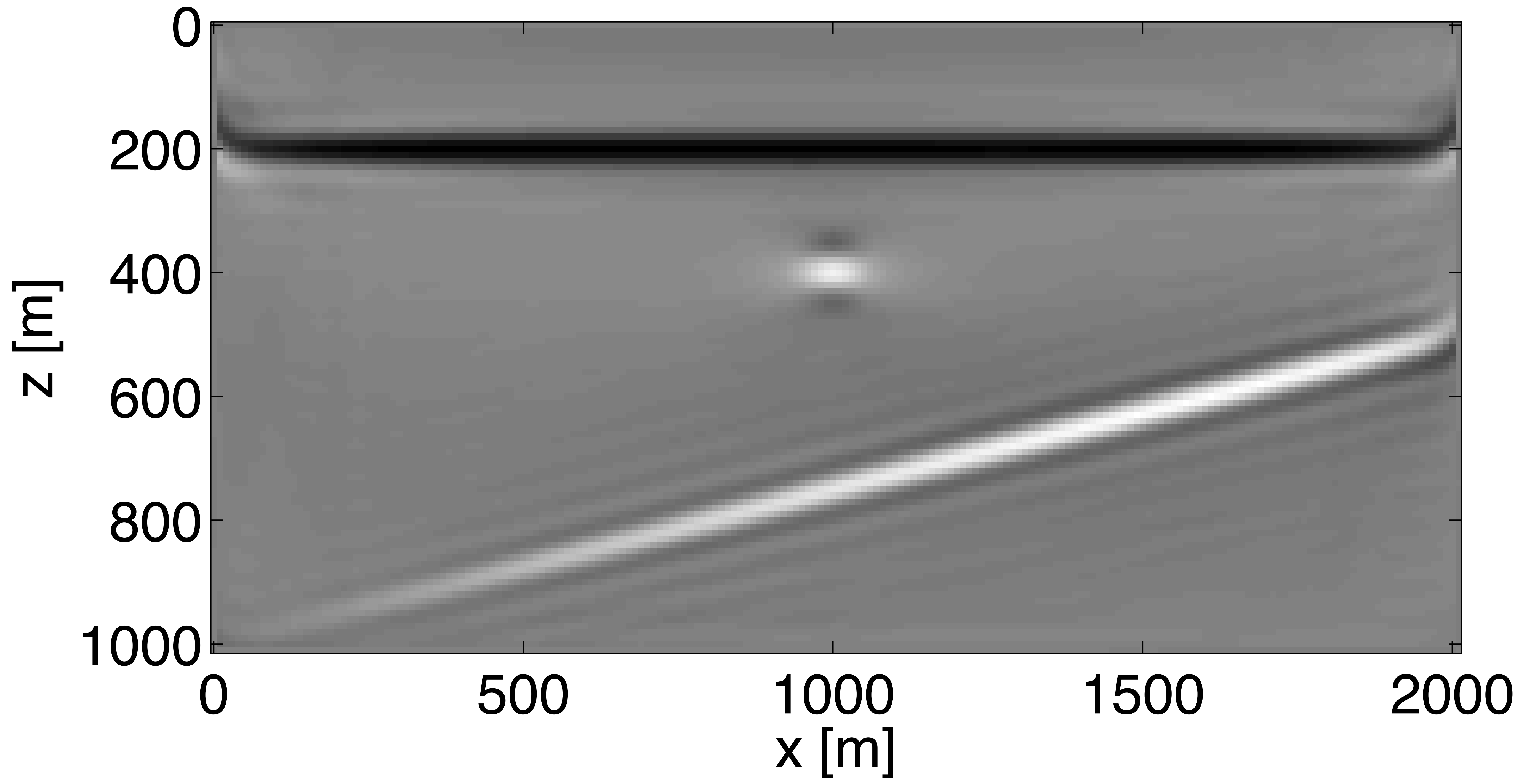


Image w/ penalty method

Imaging

***Penalty*-method reverse-time migration:**

1. solve overdetermined wave-equation
2. go for lunch
3. apply `imaging condition’

Penalty vs. reverse-time

Reverse-time migration w/ *adjoint* state:

$$\delta \mathbf{m}_{\mathrm{red}} = \sum_{i=1}^{M} \omega_i^2 \mathrm{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

$$\mathbf{v}_i := \overbrace{A_i(\mathbf{m})^{-*}\left(P_i^*(\mathbf{d}_i - P_i\mathbf{u}_i)\right)}^{\text{back prop. data residue}}$$

$$H_{\mathrm{red}}^{GN} = \sum_{i=1}^{M} G_i^* A_i^{-*} P_i^T P_i A_i^{-1} G_i$$

"*Forward*"–time migration w/ penalty formulation

$$\delta \mathbf{m}_{\mathrm{pen}} = \lambda^2 \sum_{i=1}^{M} \omega_i^2 \mathrm{diag}(\mathbf{u}_i)^* \delta \mathbf{u}_i$$

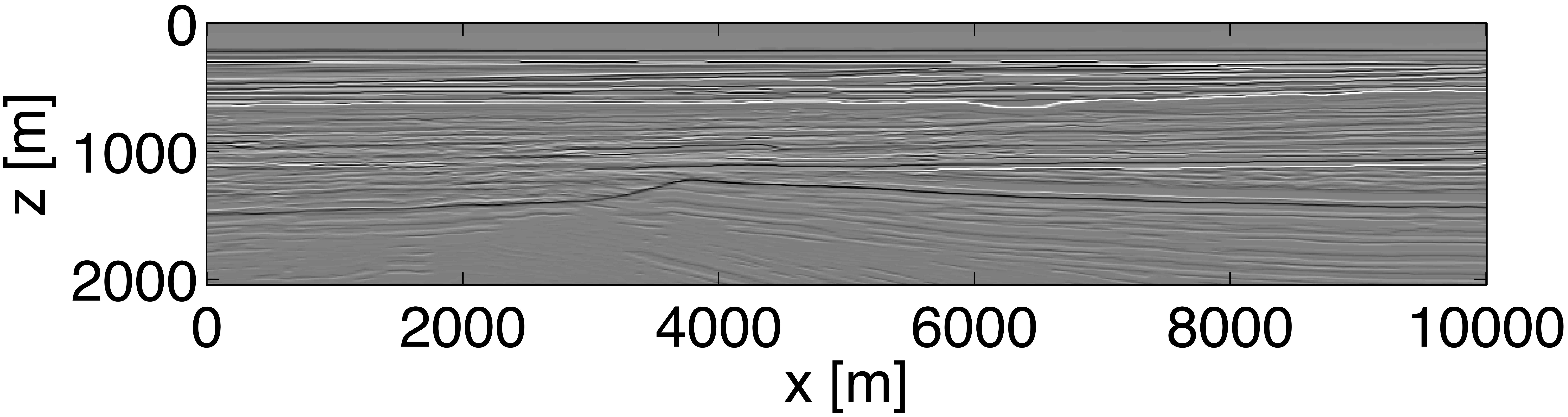
$$\delta \mathbf{u}_i := \overbrace{\left(A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\right)}^{\text{PDE residue}}$$

with *trivial* sparse GN–Hessian

$$H_{\mathrm{pen}}^{GN} = \lambda^2 \sum_{i=1}^{M} \omega_i^4 \mathrm{diag}(\mathbf{u}_i)^* \mathrm{diag}(\mathbf{u}_i)$$

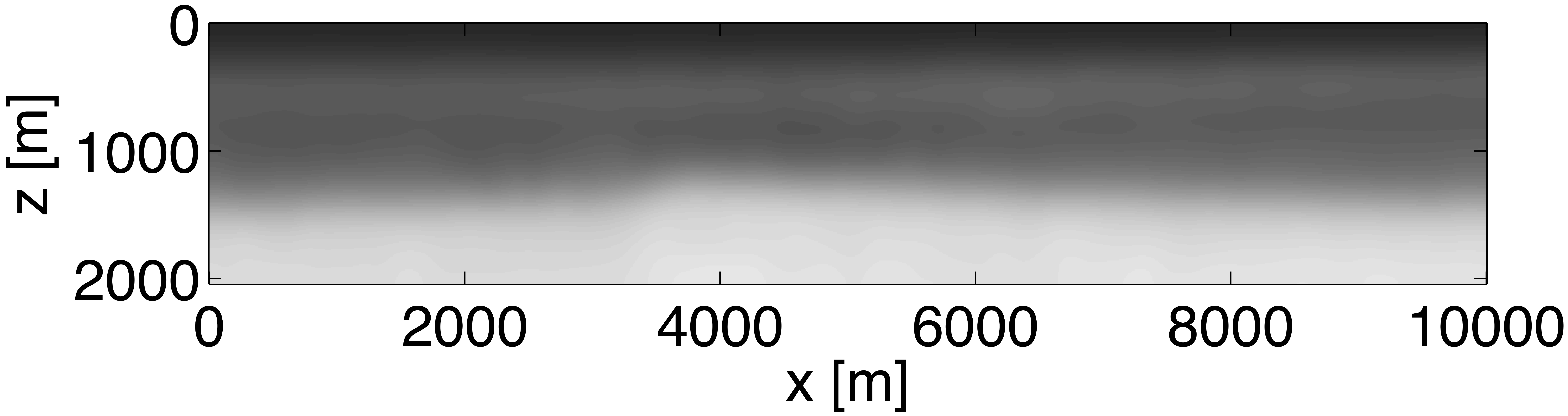
Imaging
[BG Compass model]

Perturbation
model grid spacing: 10m
inversion crime
split spread w/ 50m source & receiver spacing
57 frequencies 2-30Hz



Imaging

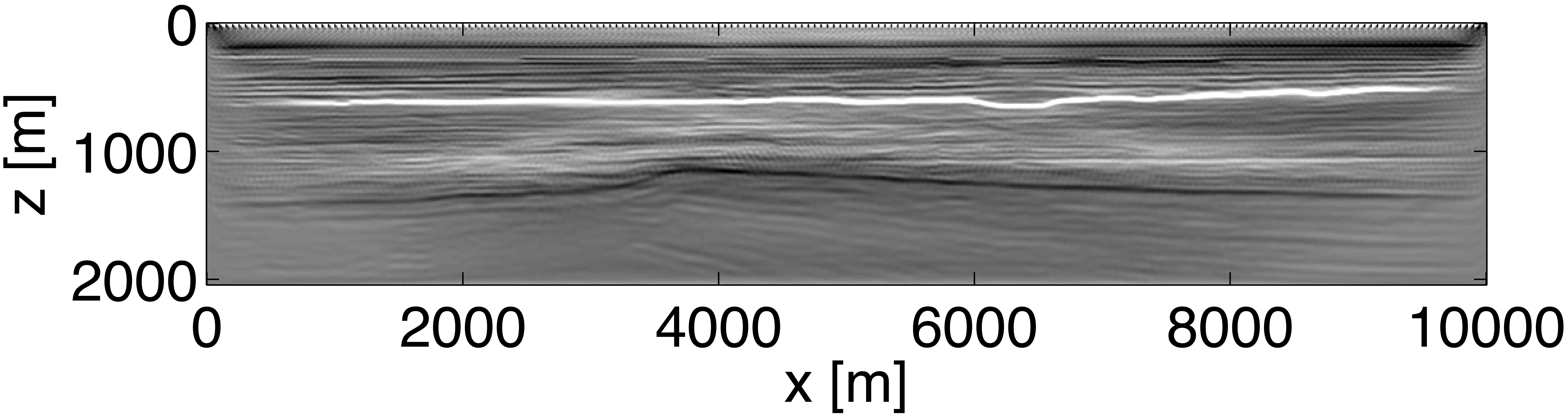
Background model



Imaging

Conventional **RTM:**

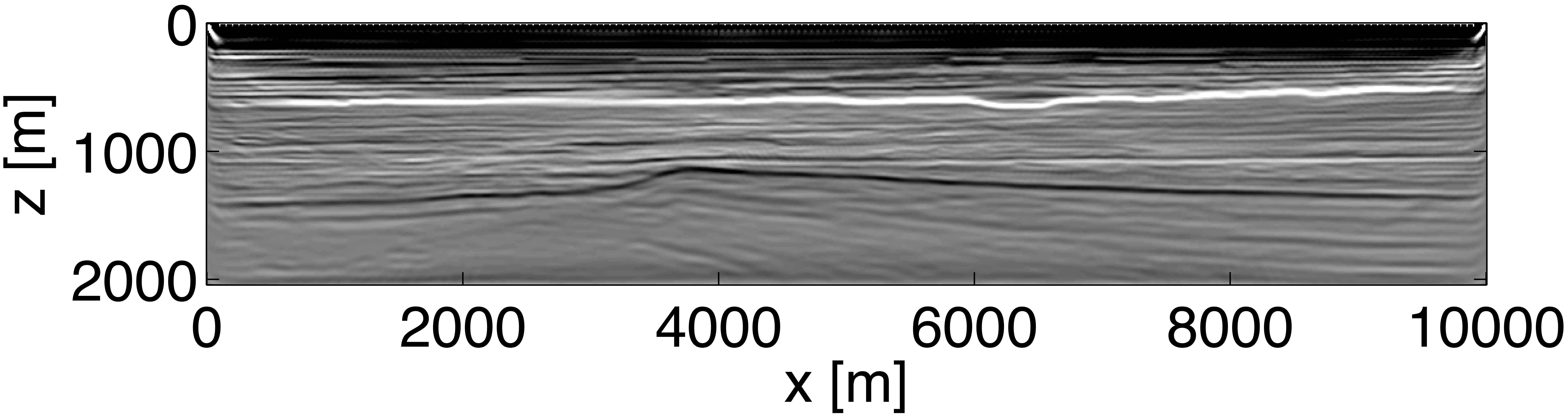
1. solve forward wave-equation
2. solve adjoint wave-equation
3. apply `imaging condition’



Imaging

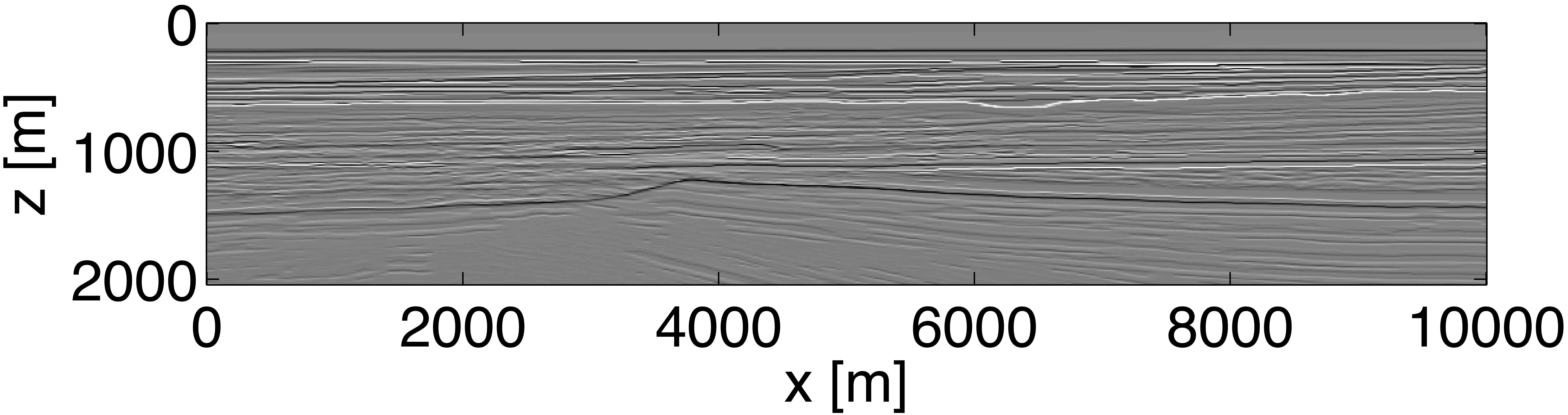
***Penalty*-method reverse-time migration:**

1. solve overdetermined wave-equation
2. go for lunch
3. apply `imaging condition’



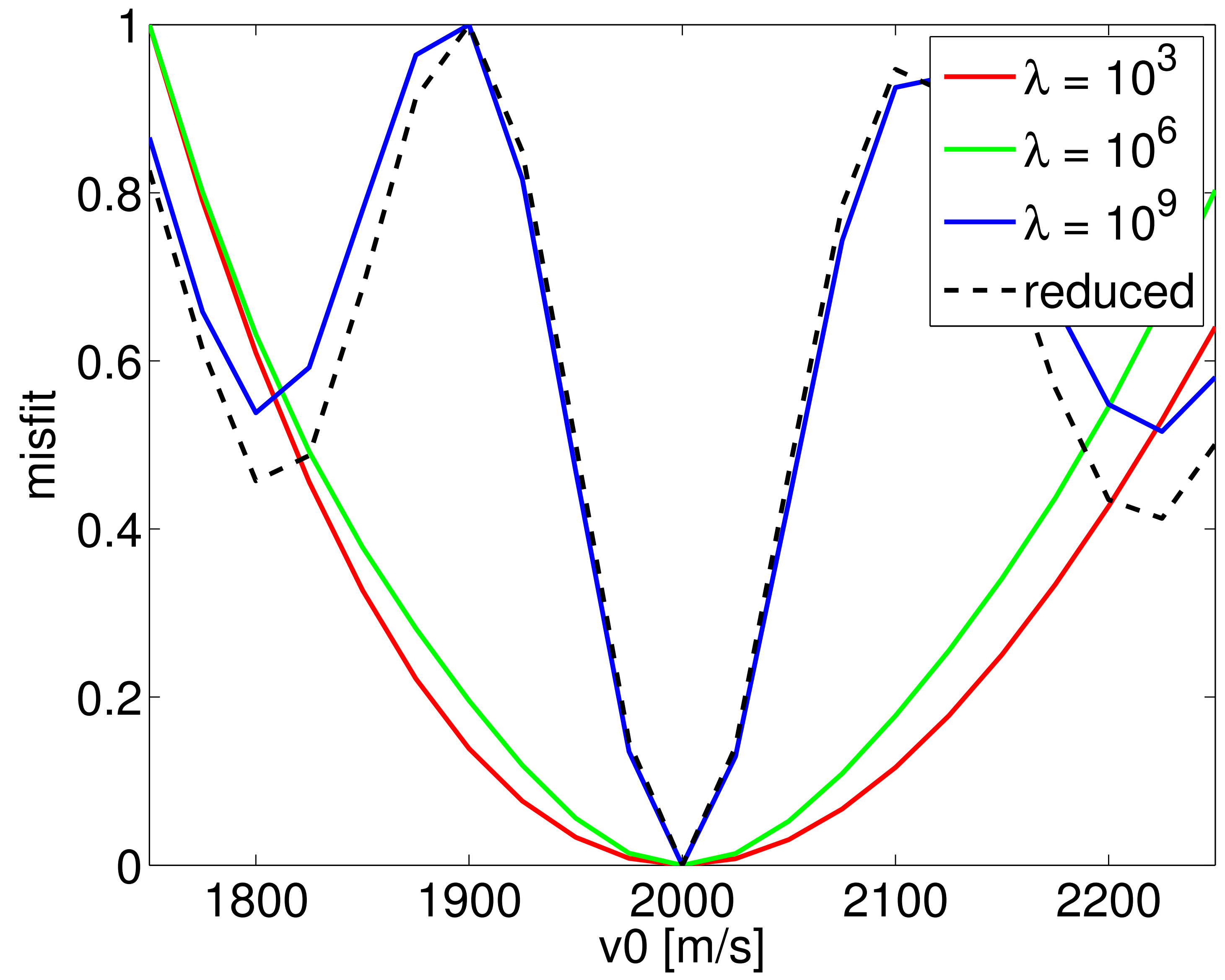
Imaging
[BG Compass model]

Perturbation
model grid spacing: 10m
inversion crime
split spread w/ 50m source & receiver spacing
57 frequencies 2-30Hz

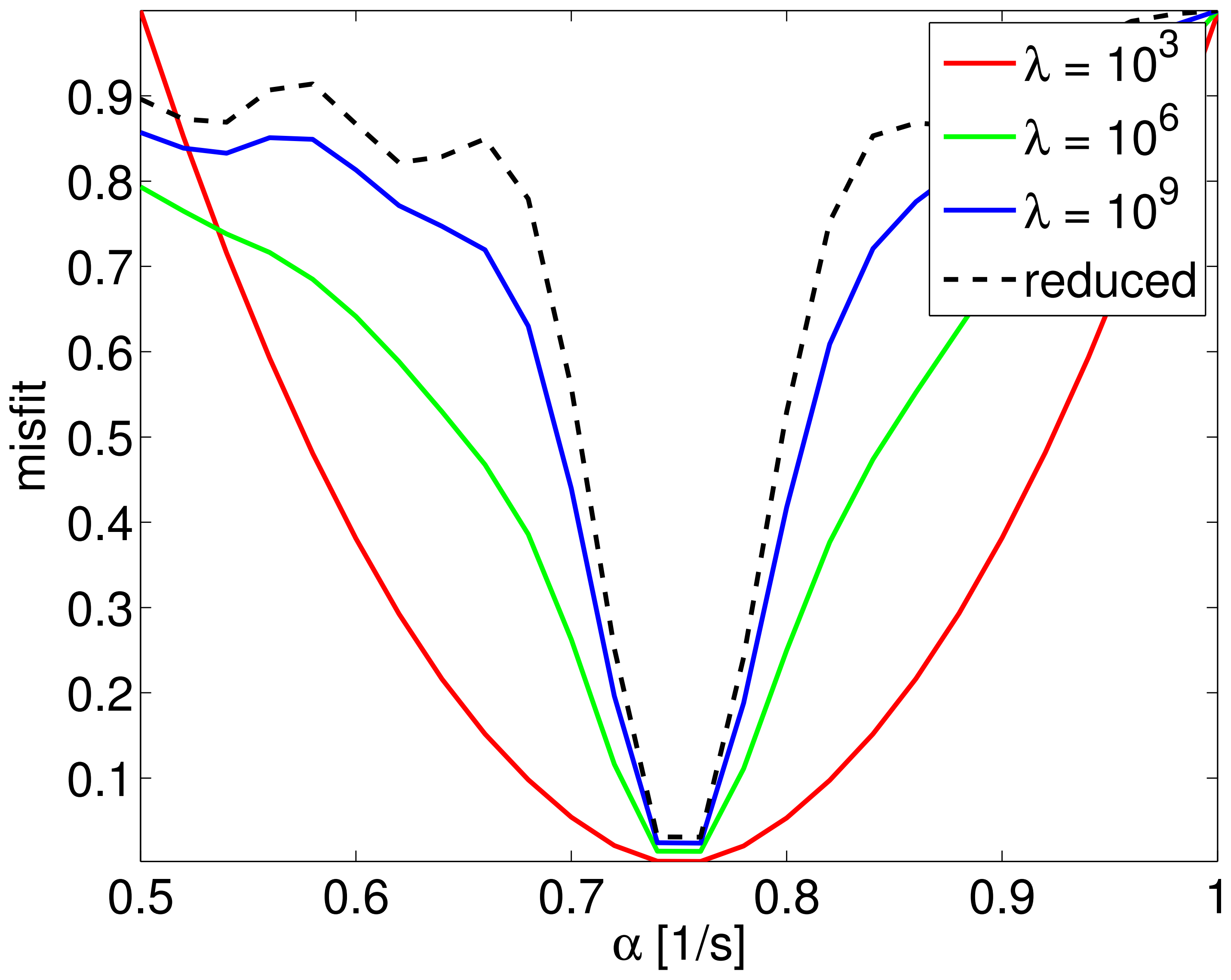


Local minima

single shot, single frequency data for *linear* velocity profile $v(z) = v_0 + \alpha z$, misfit as function of (v_0, α)



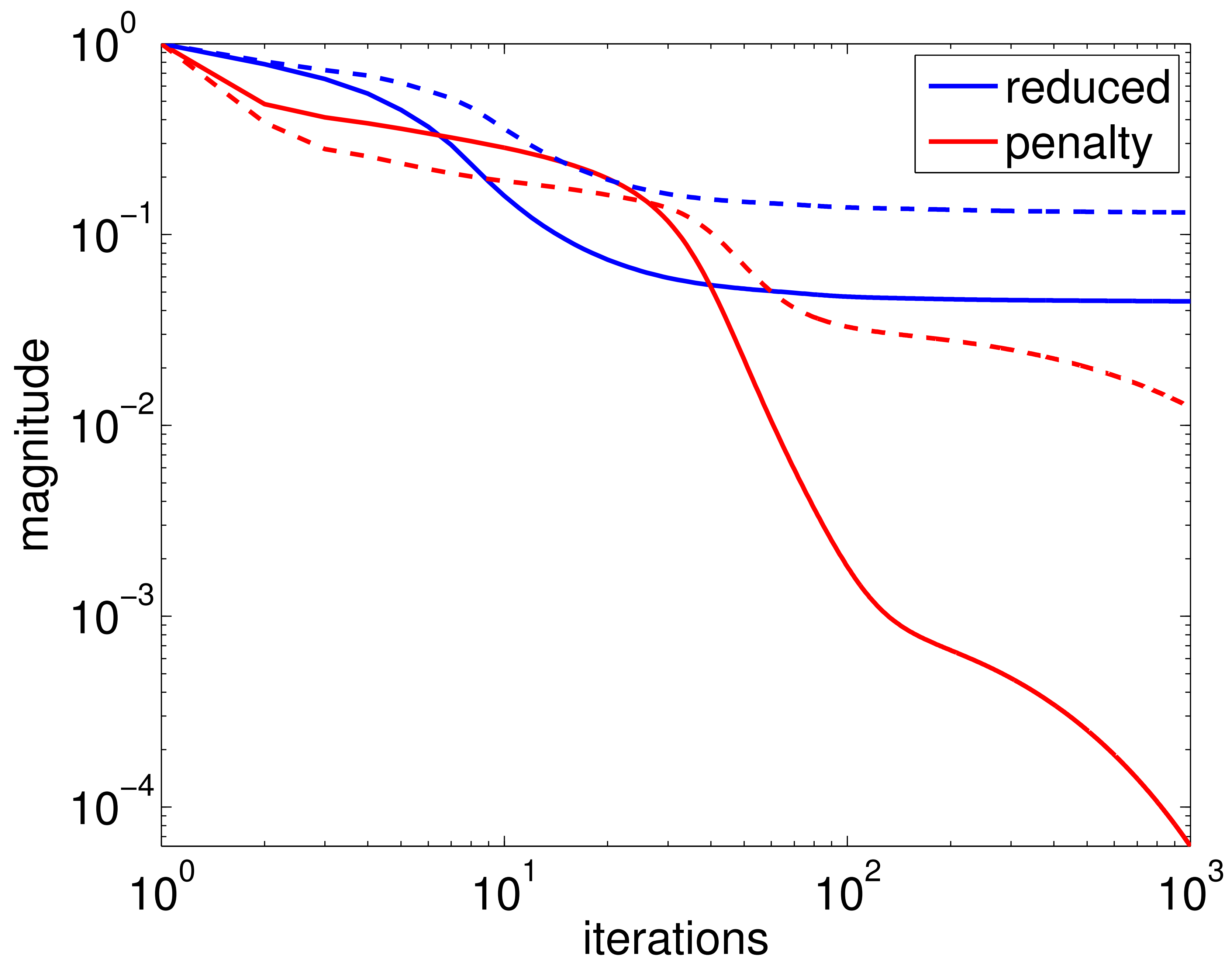
constant



slope

Local minima

Invert *single* shot, *single* frequency data for 1D velocity profile using *steepest* descent.

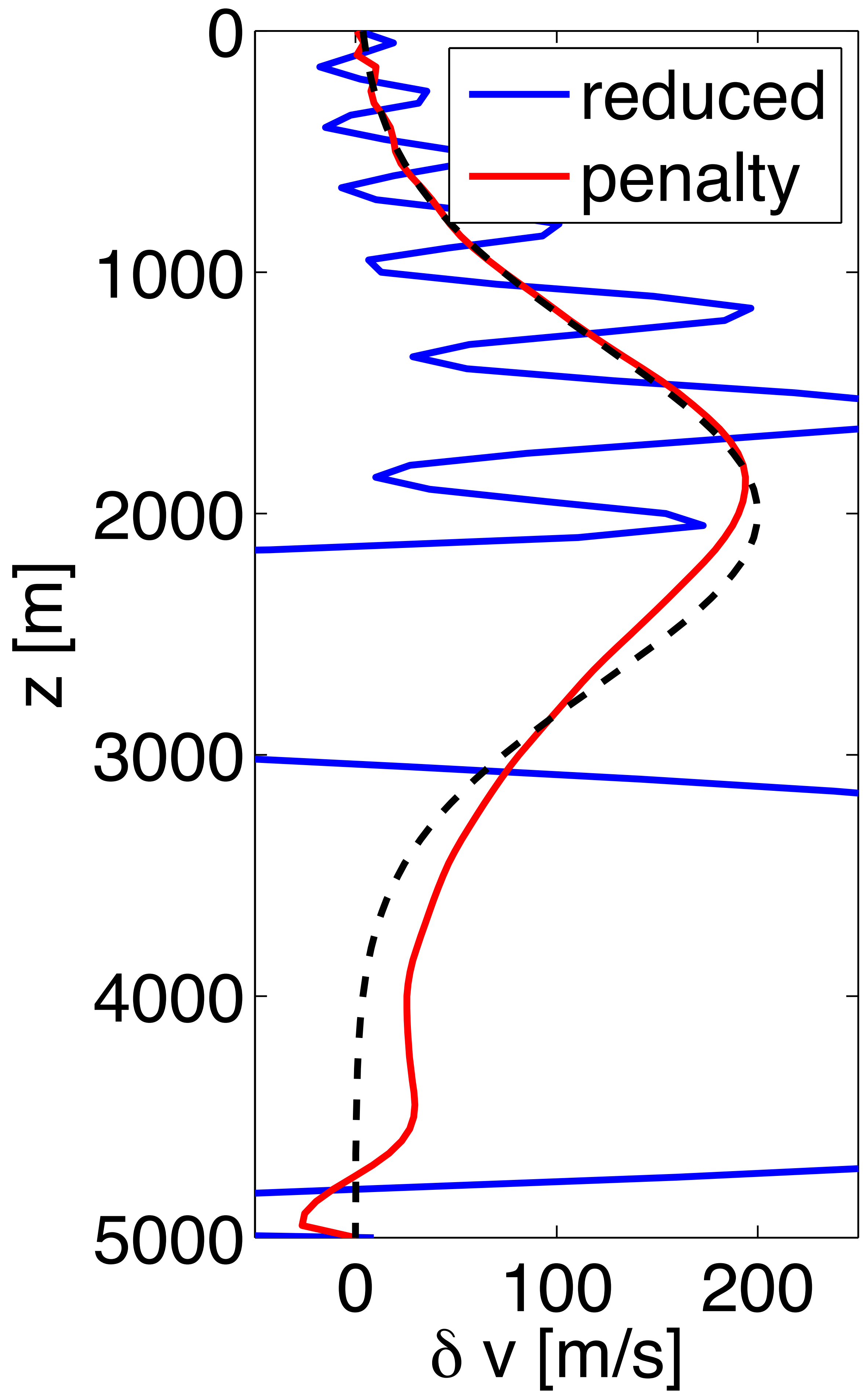


- misfit

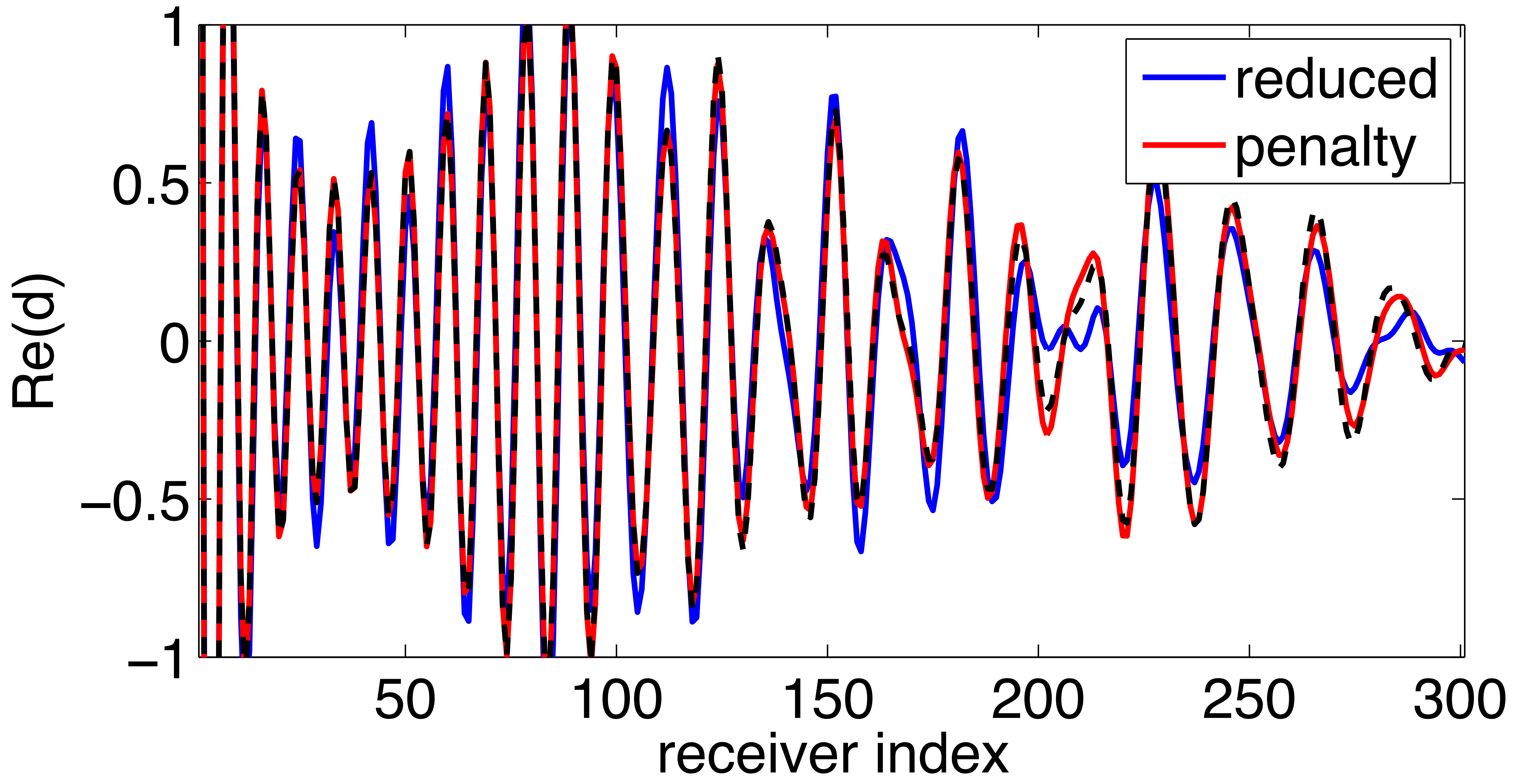
-- gradient

misfit & gradient as function # iterations

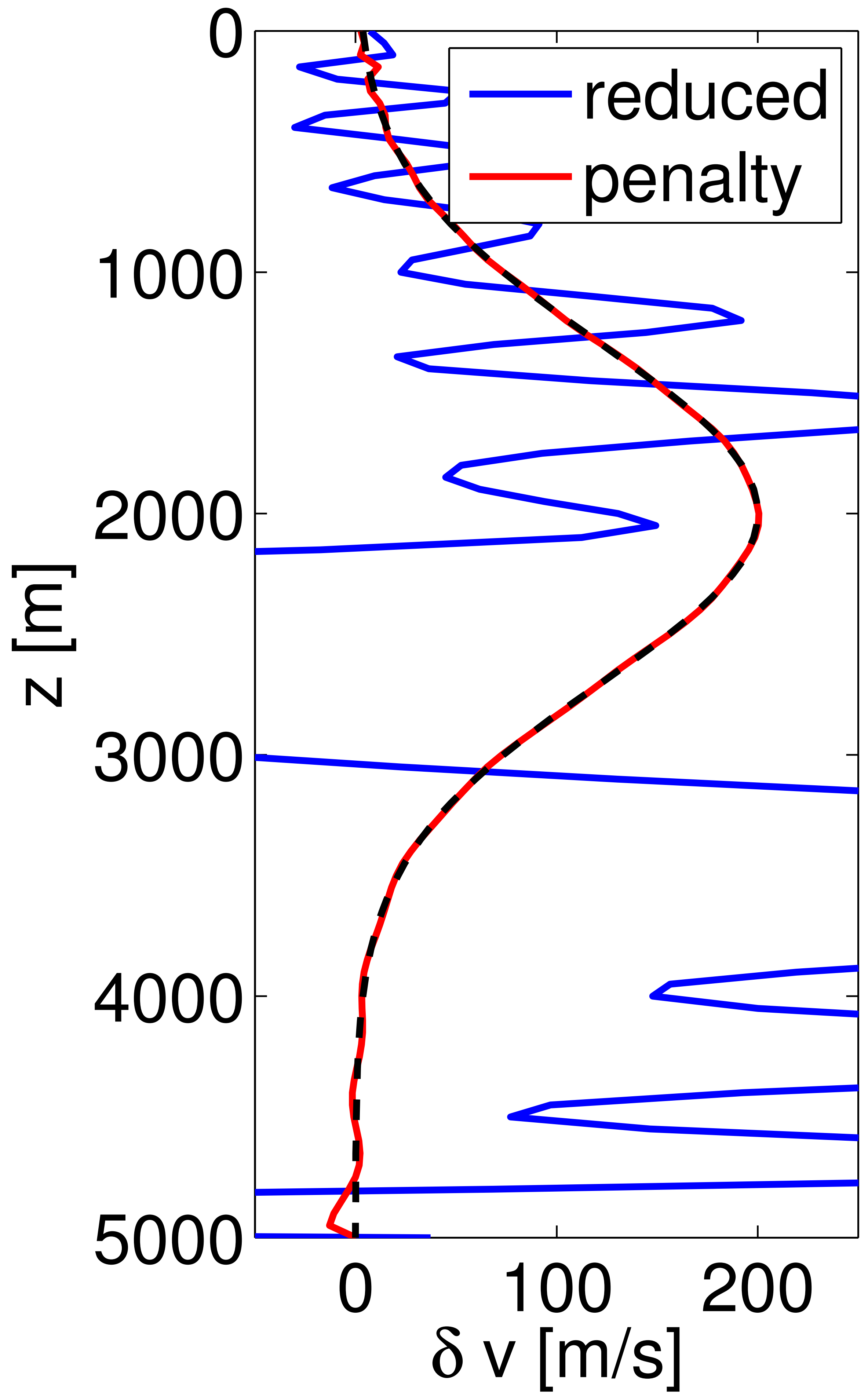
Local minima



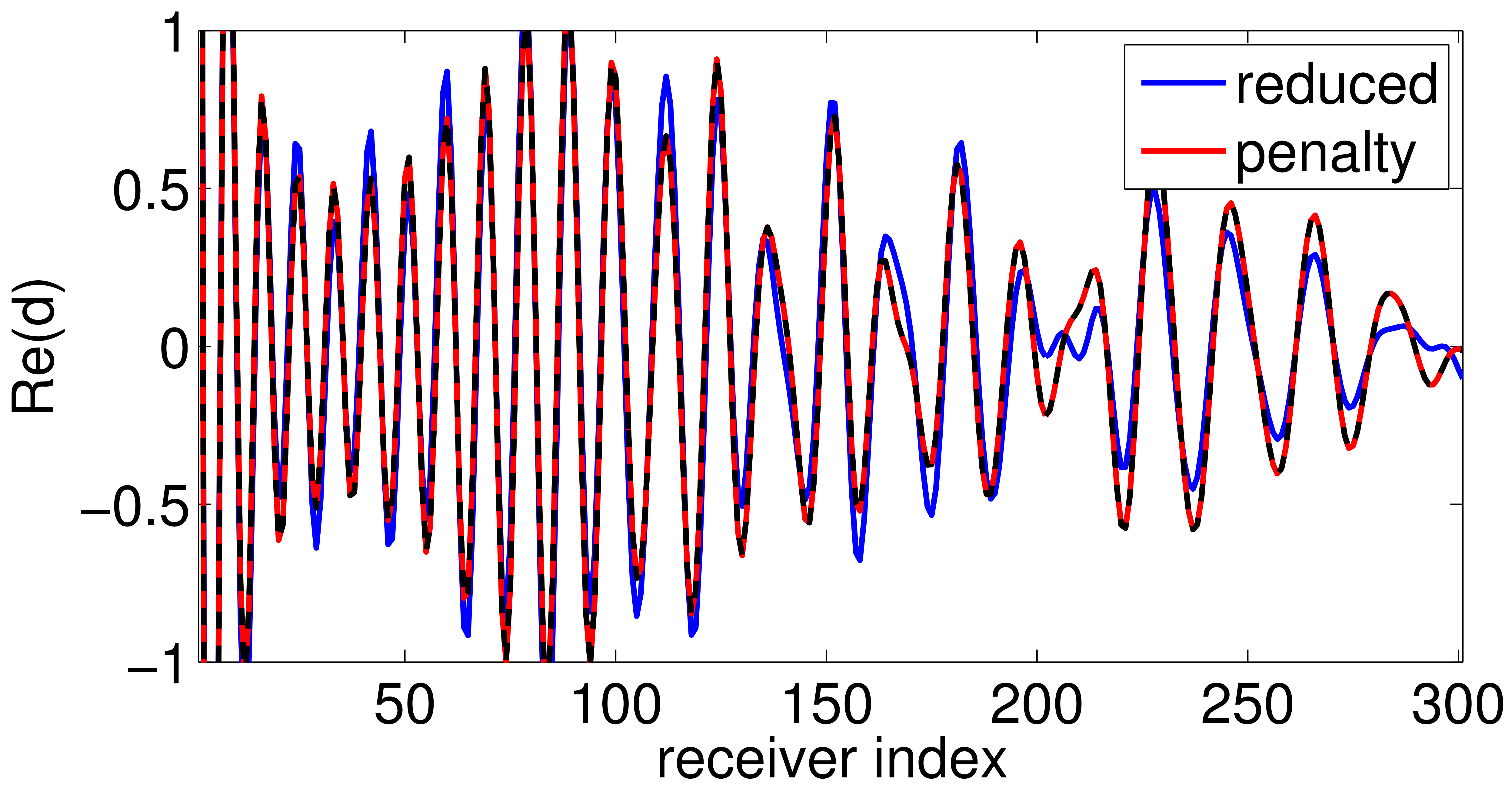
After 100 iterations



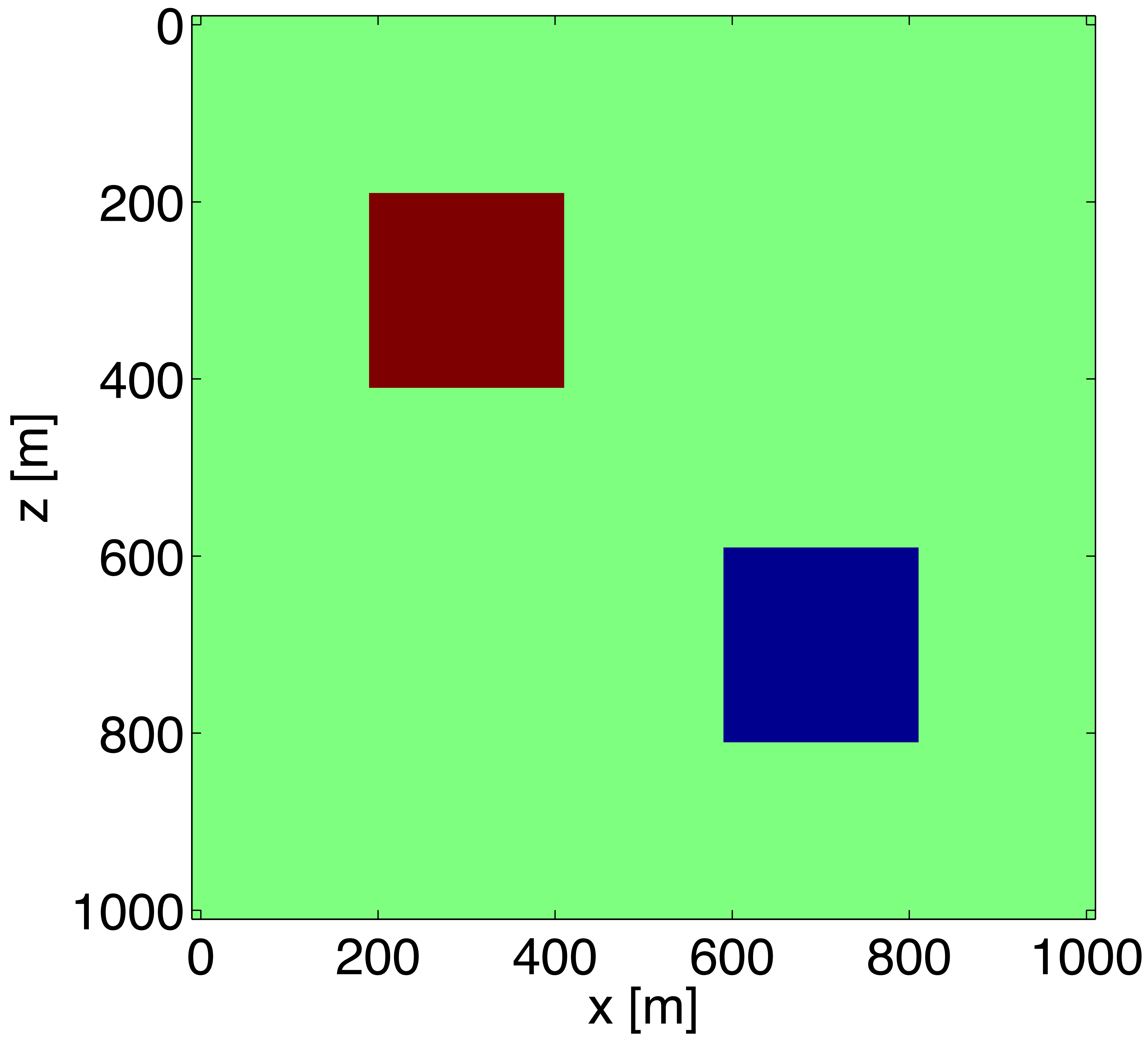
Local minima



After 1000 iterations



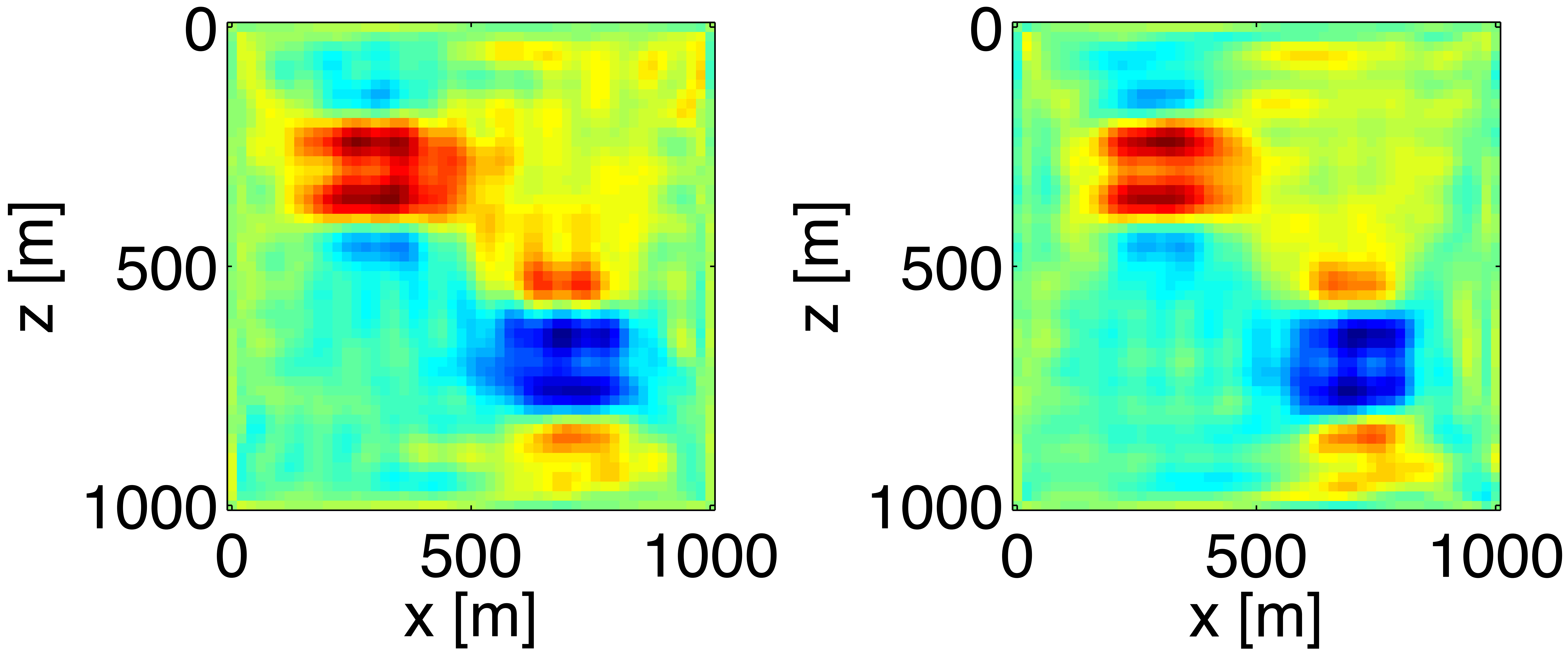
Cross well



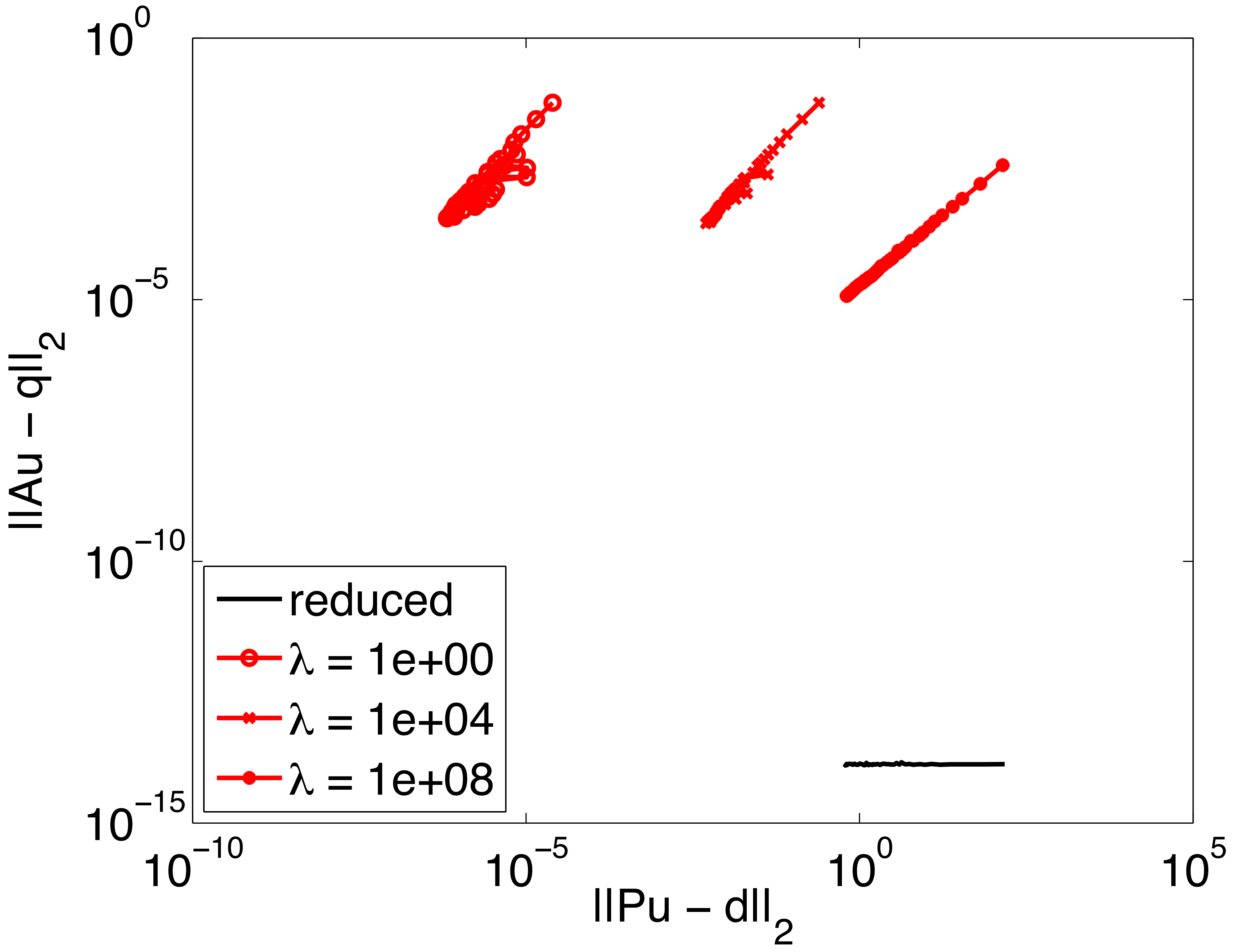
original medium

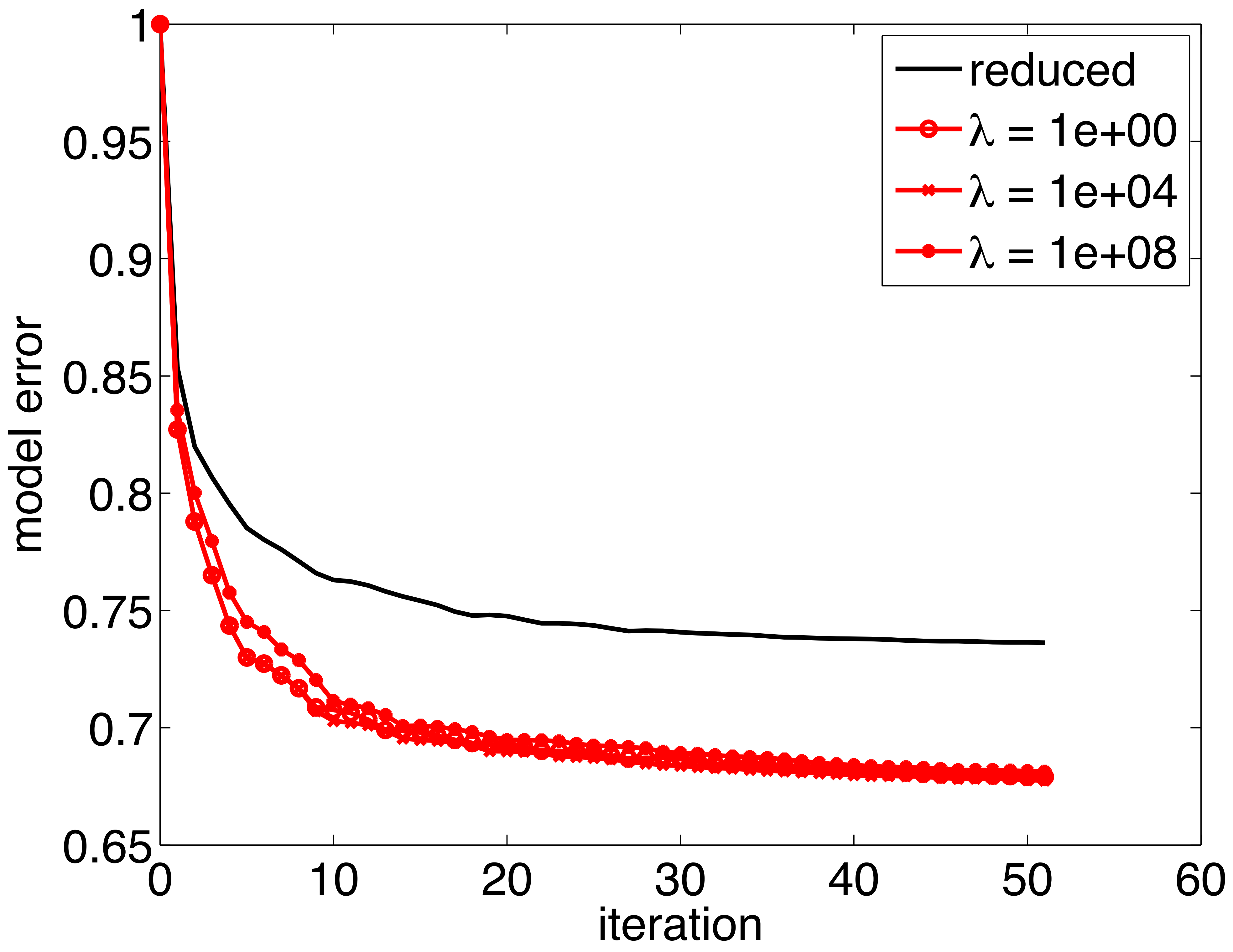
- 2D cross-well configuration
- 5 point Helmholtz FD operator with abc
- Direct solver for PDEs (A\b)
- invert a single frequency (10 Hz)
- L-BFGS

Cross well FWI

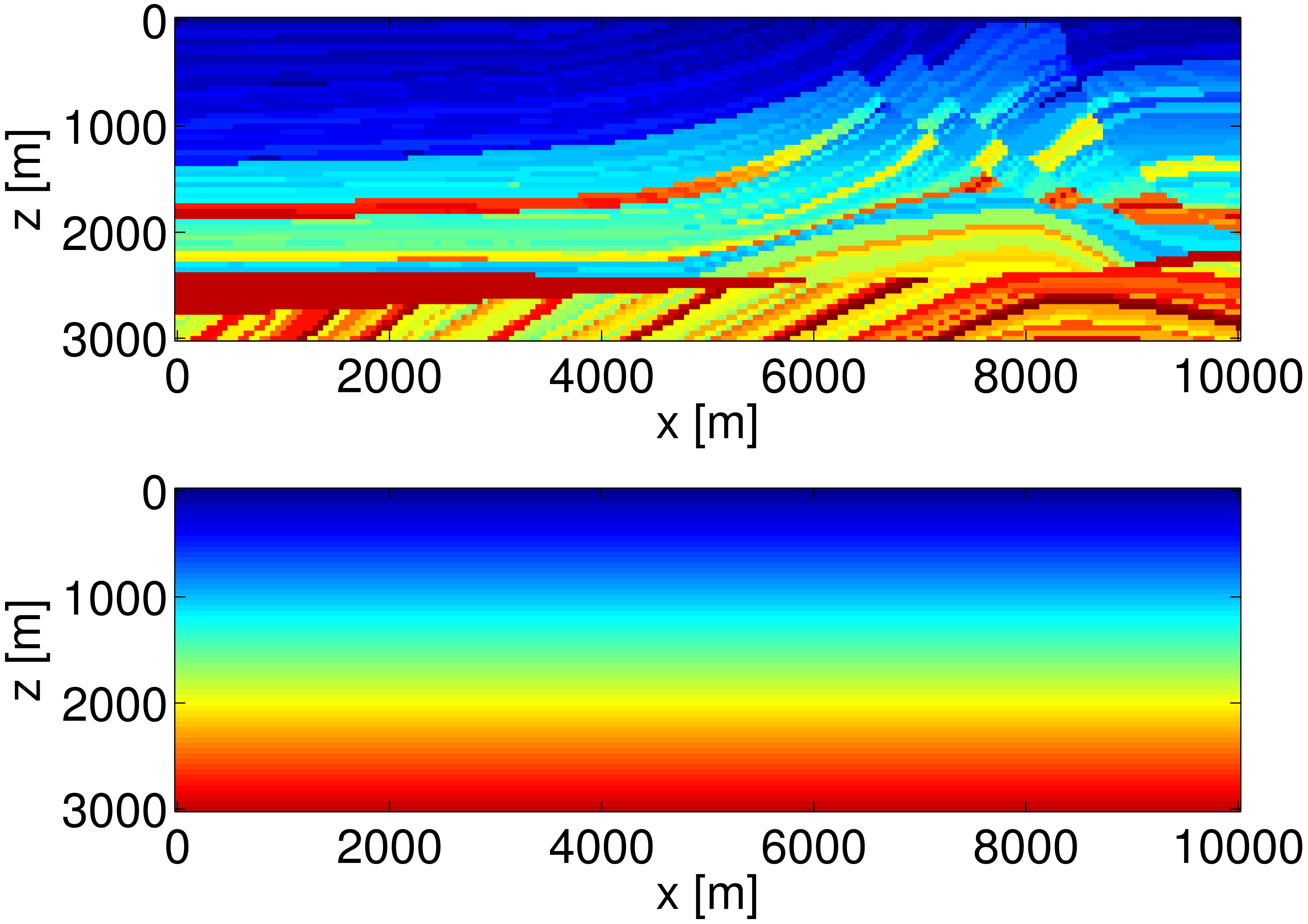


Cross well FWI





Waveform inversion



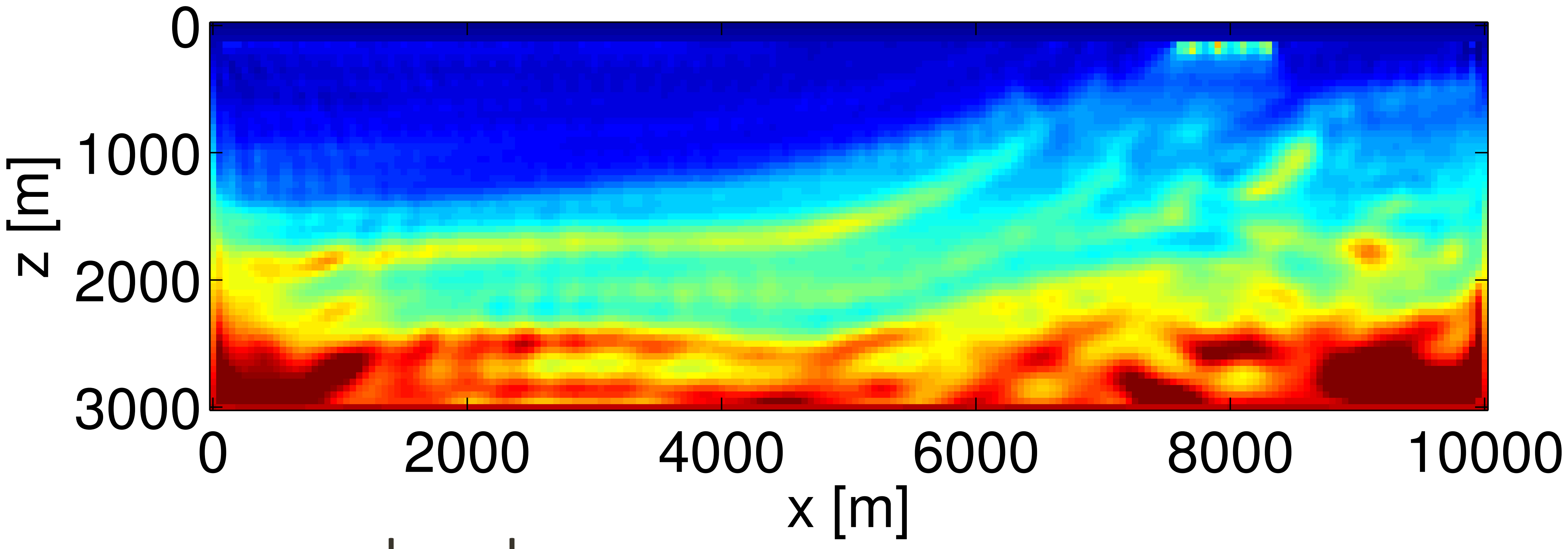
original + initial model

- 2D reflection configuration
- 5 point Helmholtz FD operator with abc
- Direct solver for PDEs (A\b)
- frequency continuation
- L-BFGS

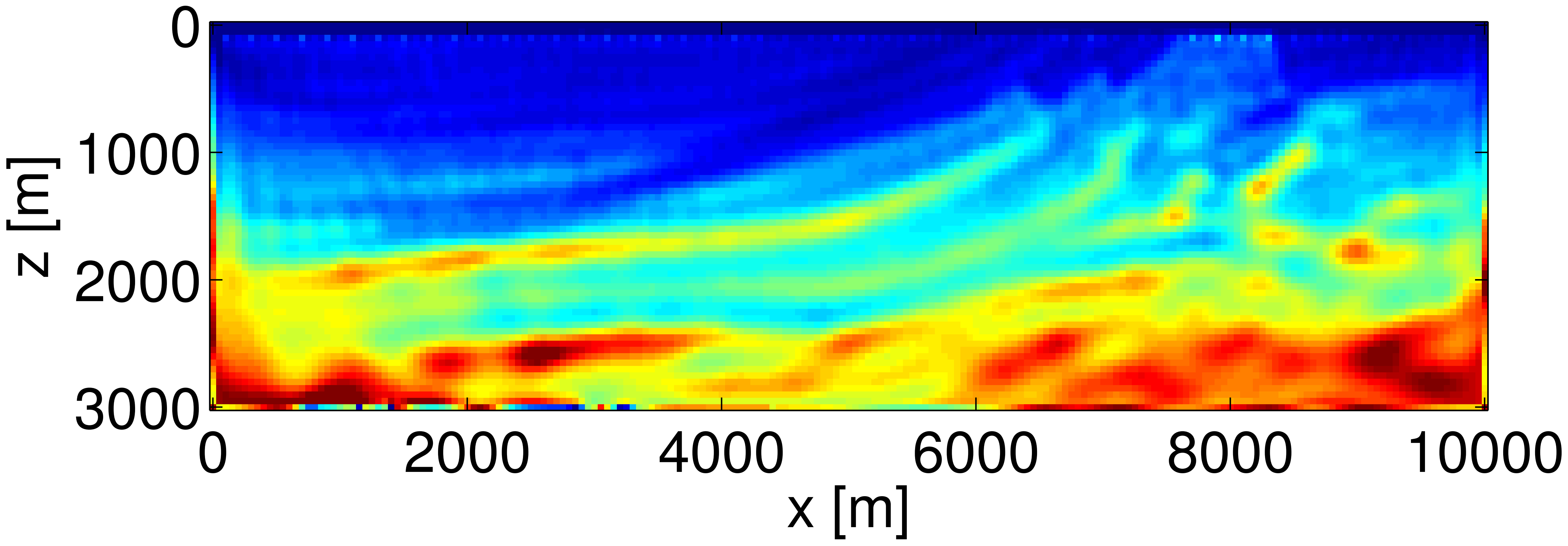
Waveform inversion

frequencies 1-5 Hz, no noise

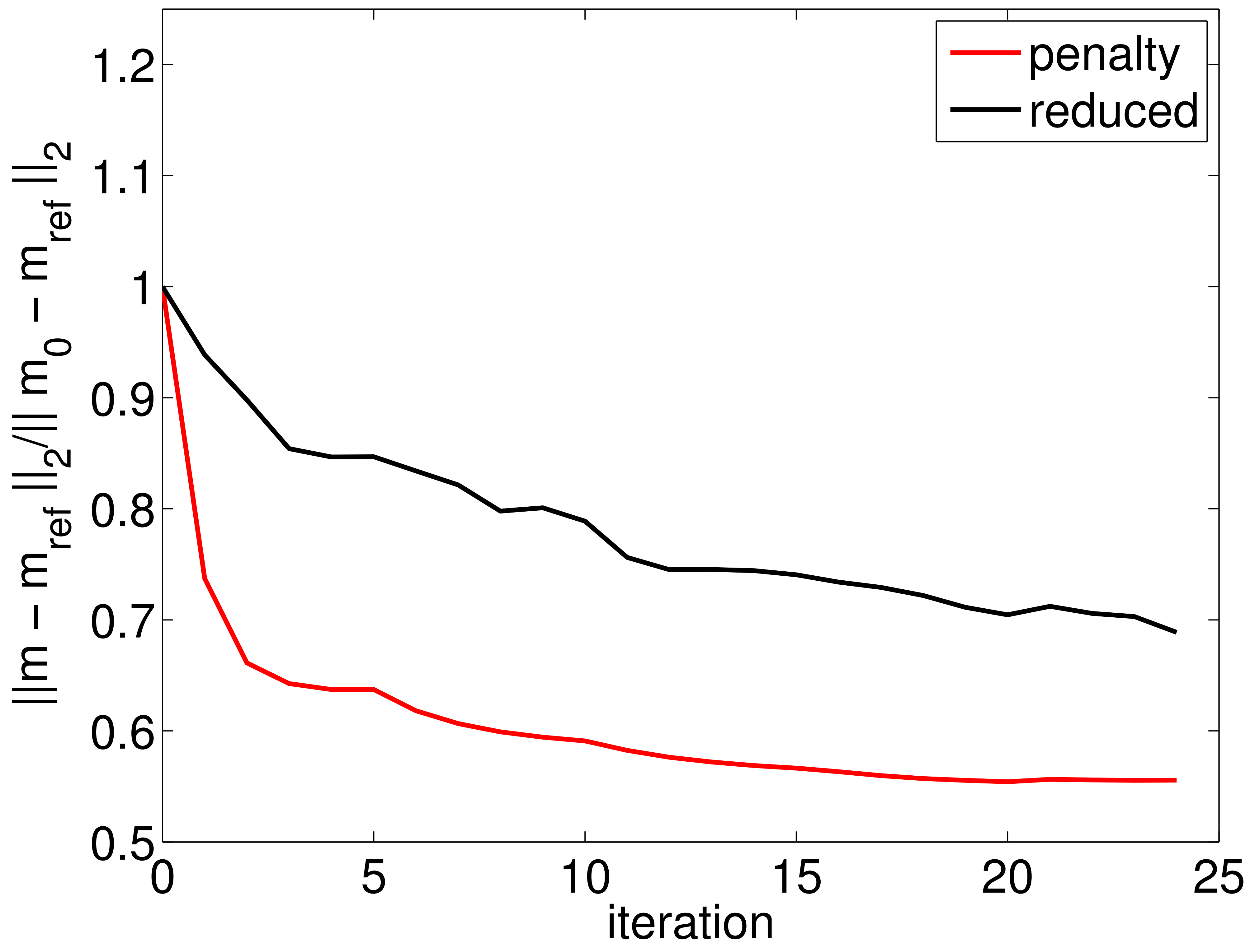
penalty



reduced

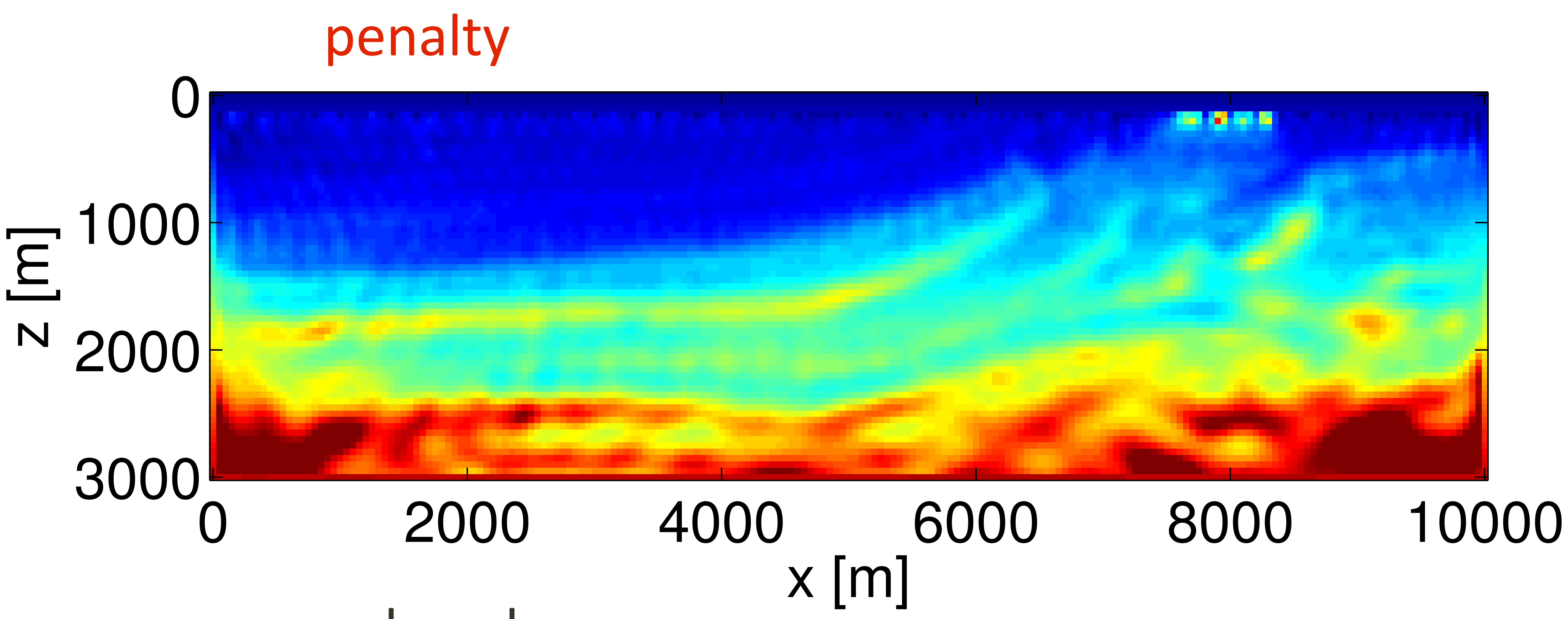


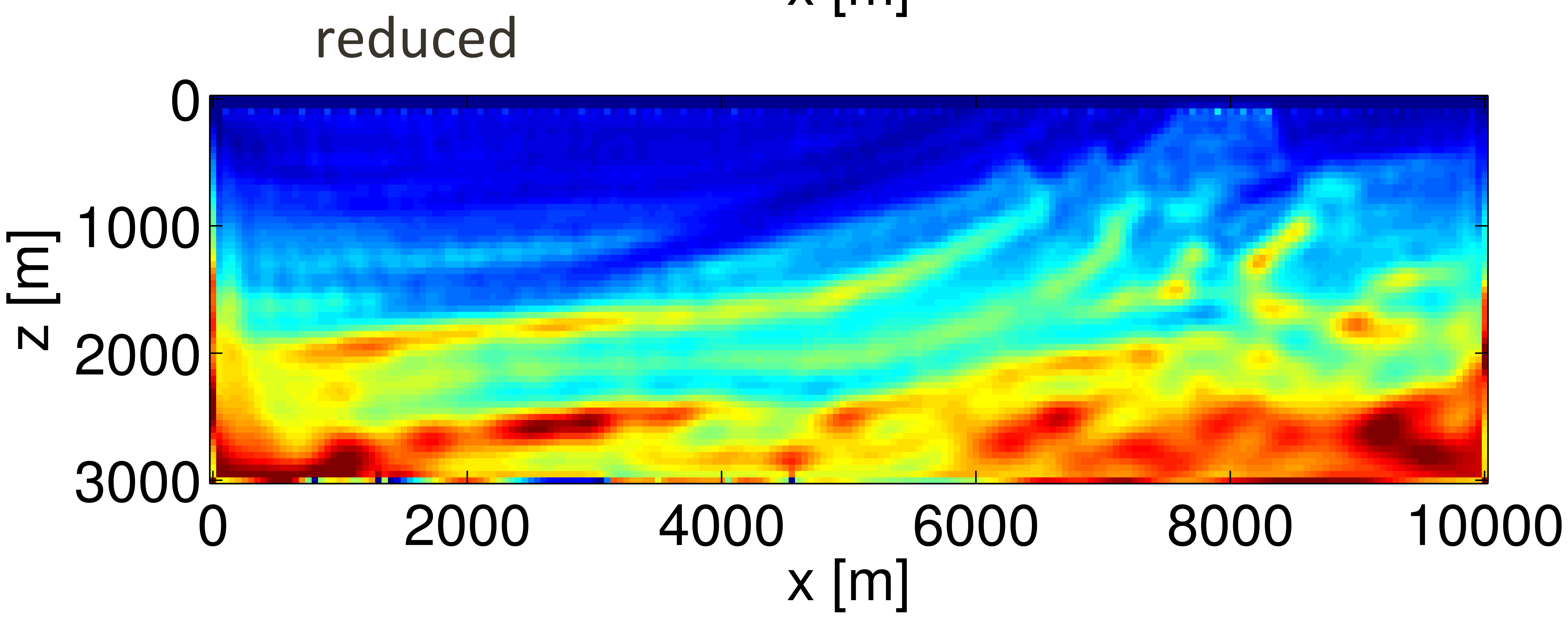
model error

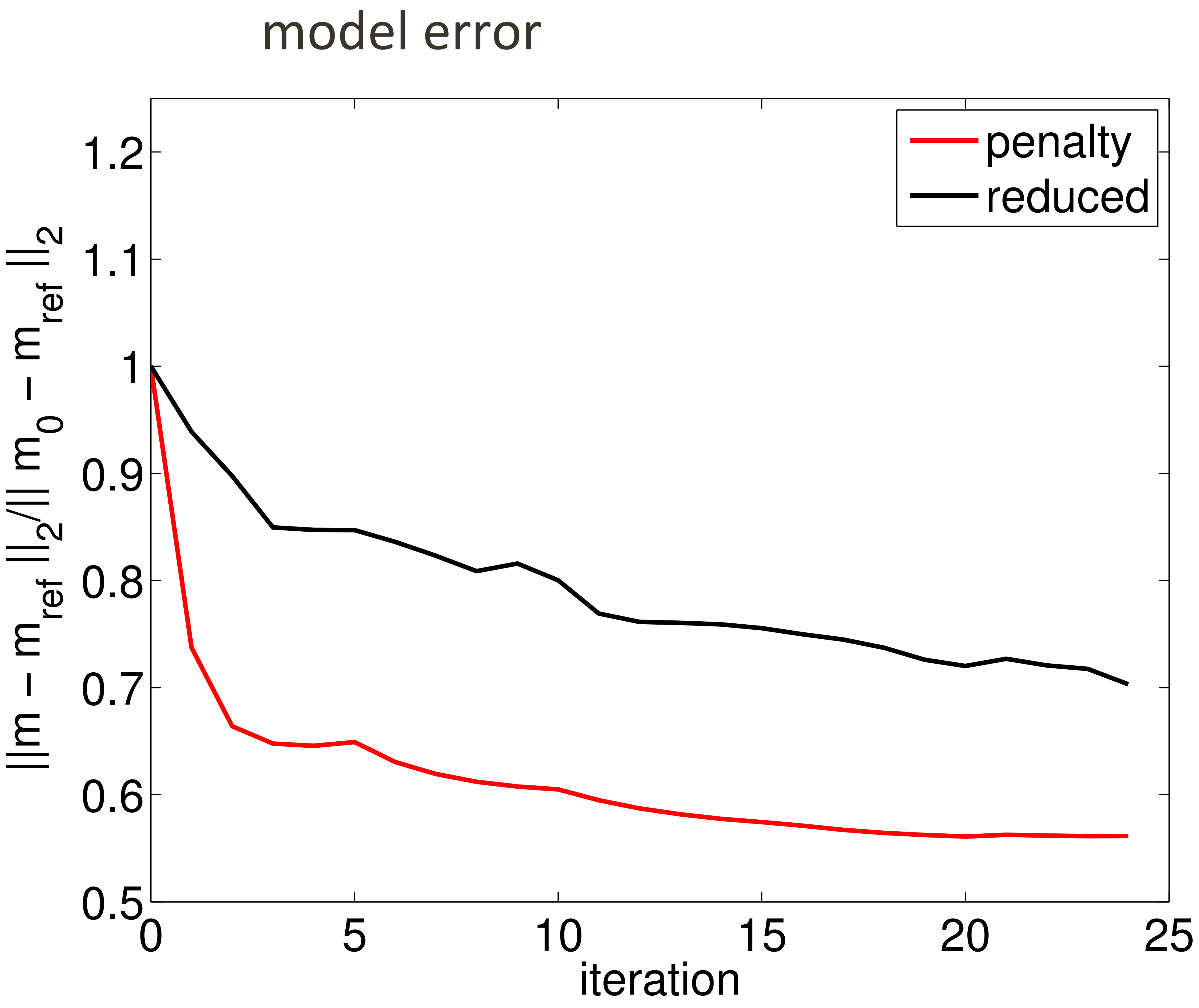


Waveform inversion

frequencies 1-5 Hz, 10 % Gaussian noise

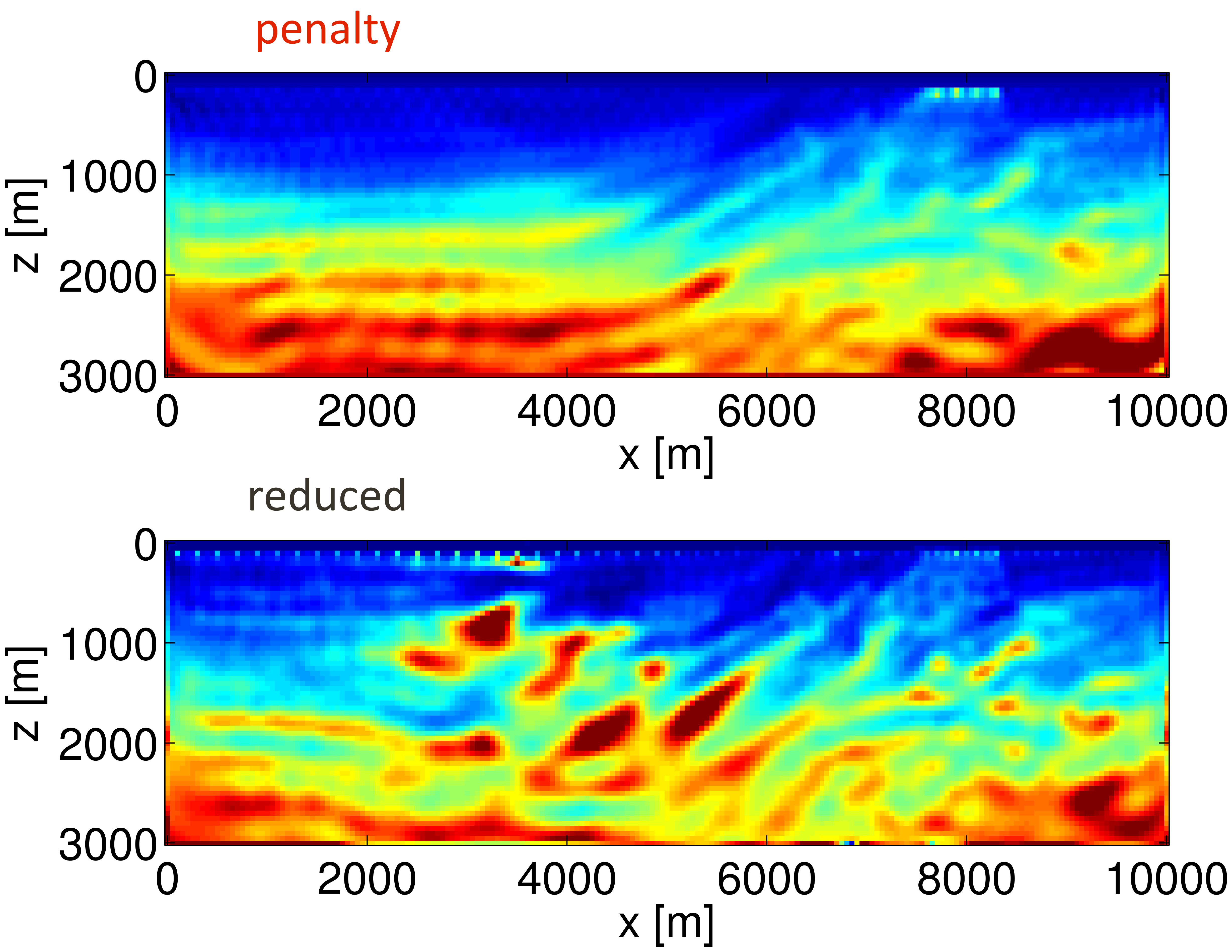


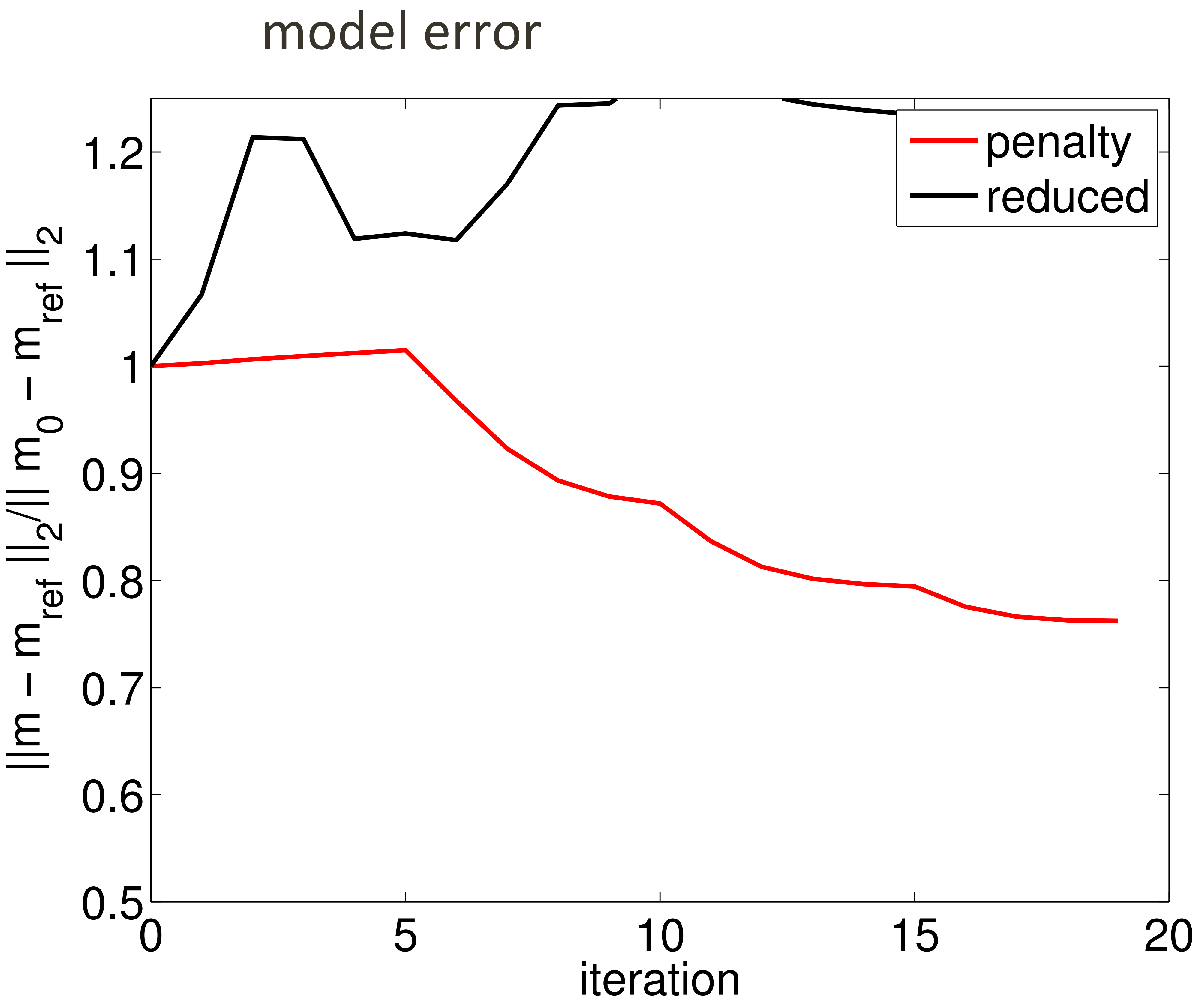




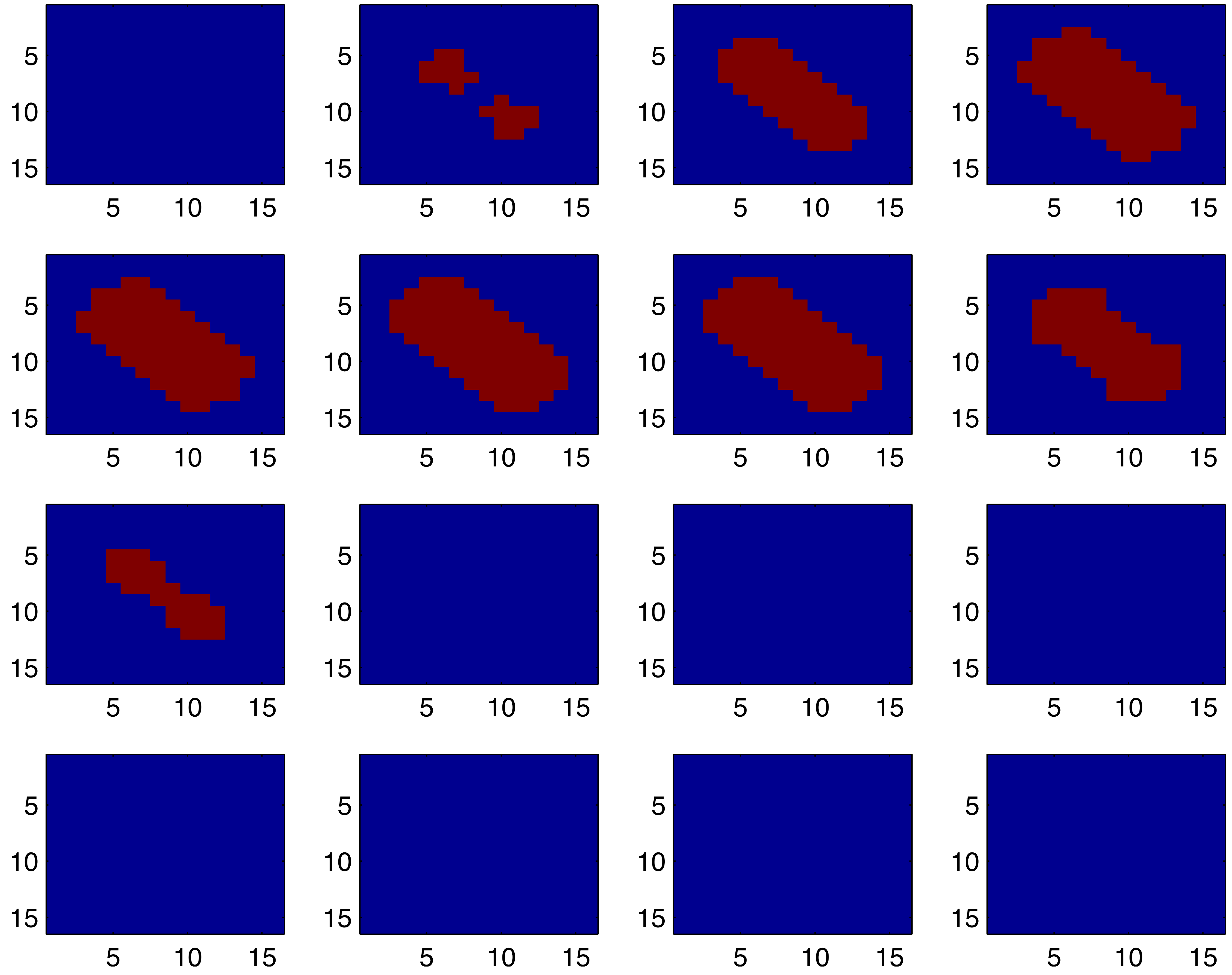
Waveform inversion

frequencies 2-5 Hz, 10 % Gaussian noise





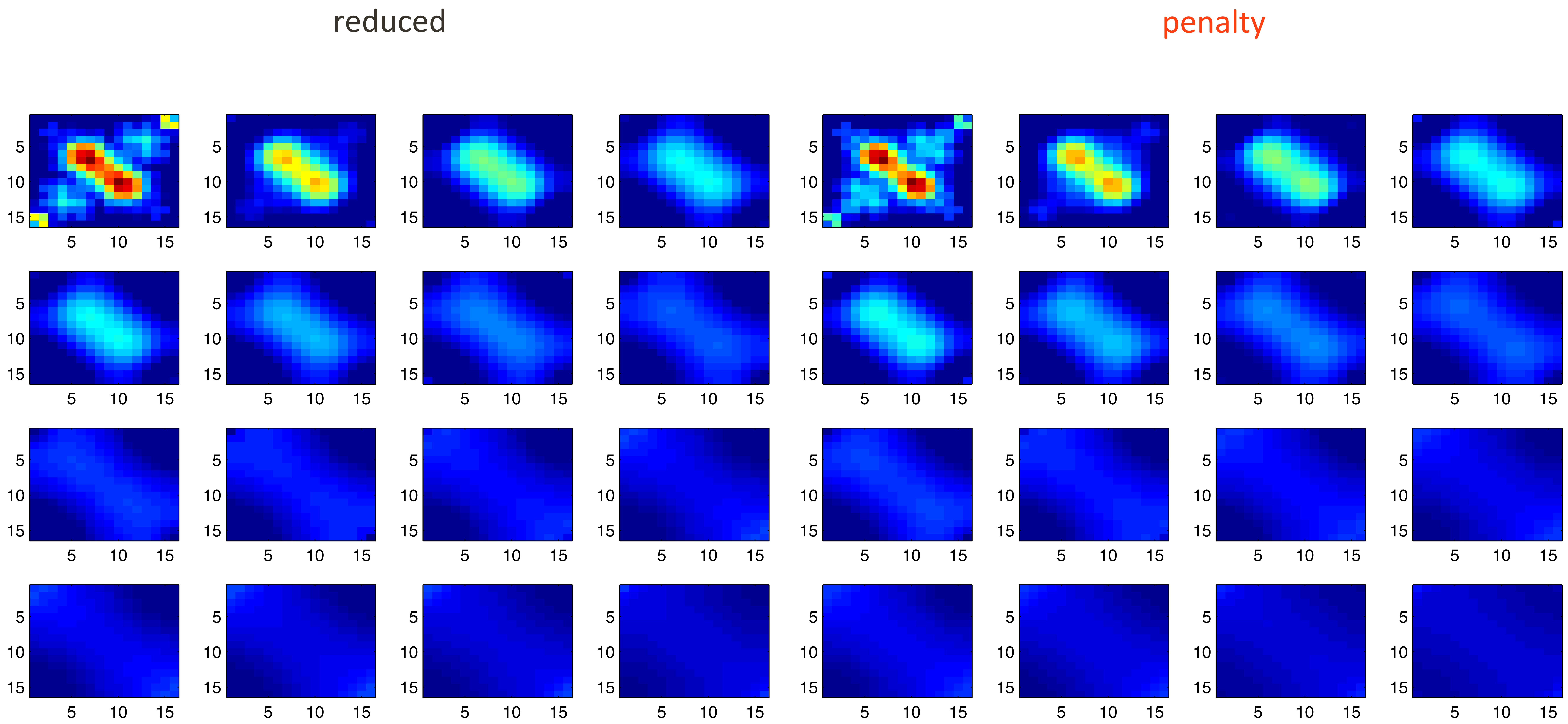
3D DC resistivity



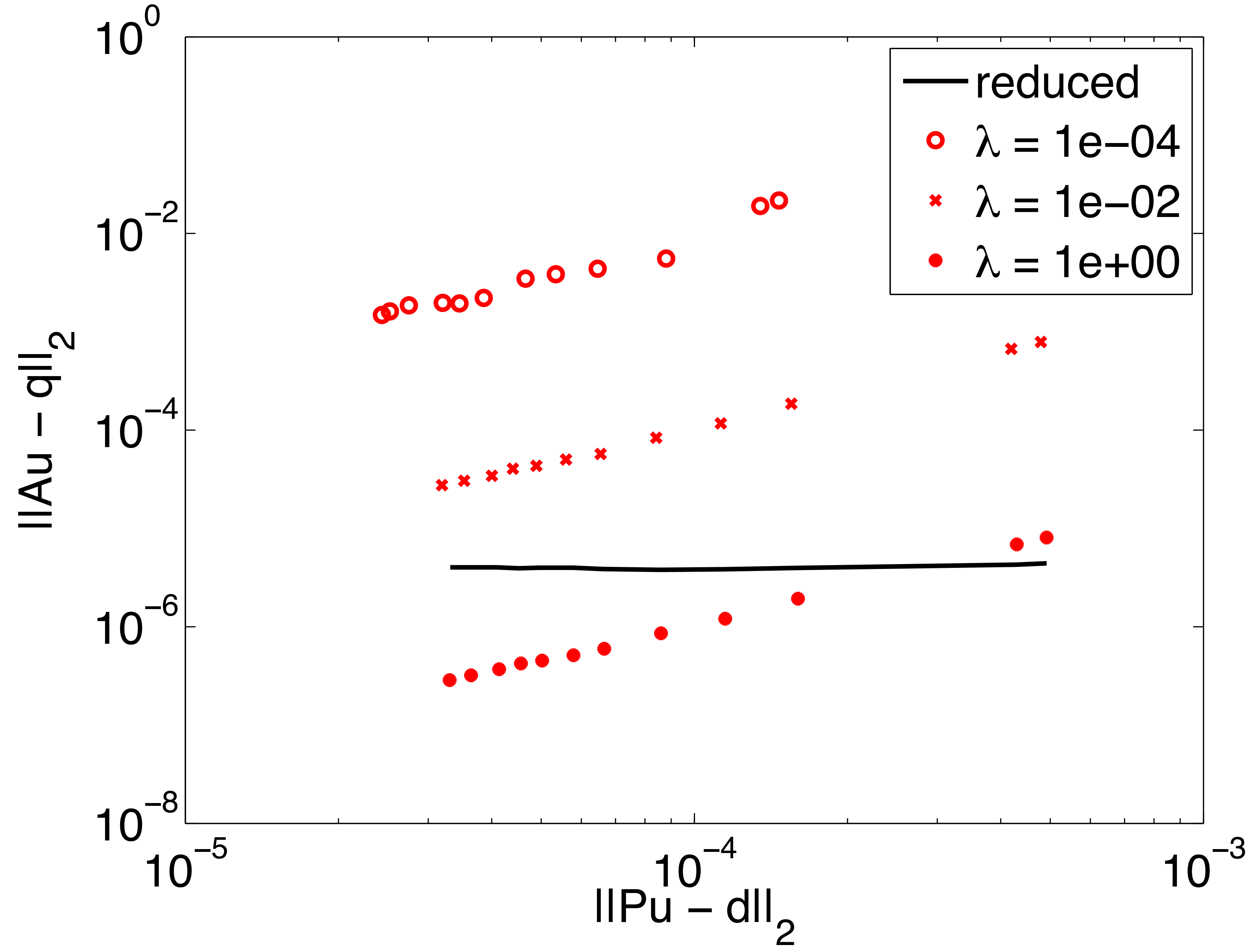
conductivity

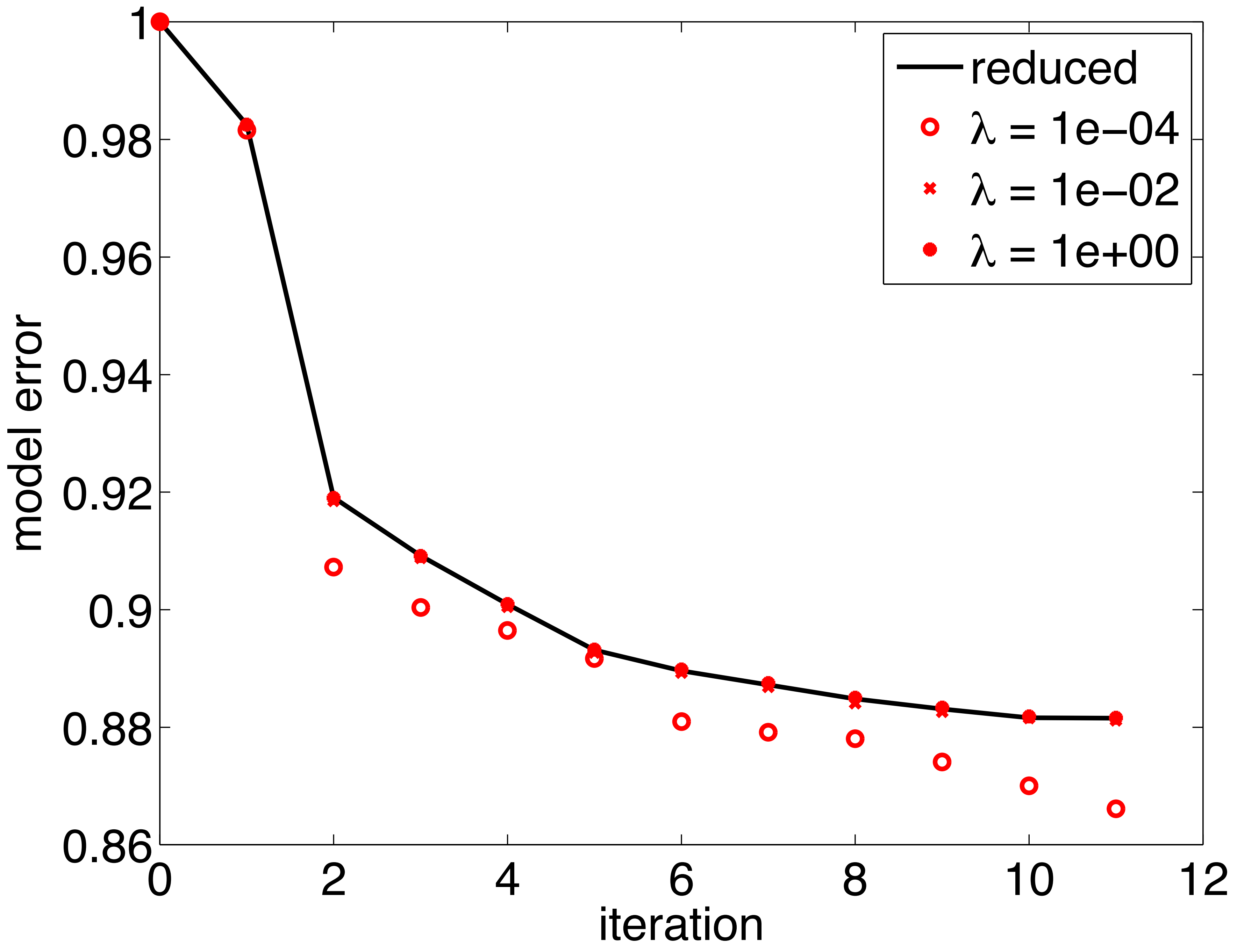
- PDE: $\nabla \cdot \sigma \nabla u = q$
- 16x16x16 grid
- 83 sources (crosswell)
- rec. on the surface
- FD discretization [Haber '07]
- iterative solvers for PDEs
- L-BFGS

3D DC resistivity

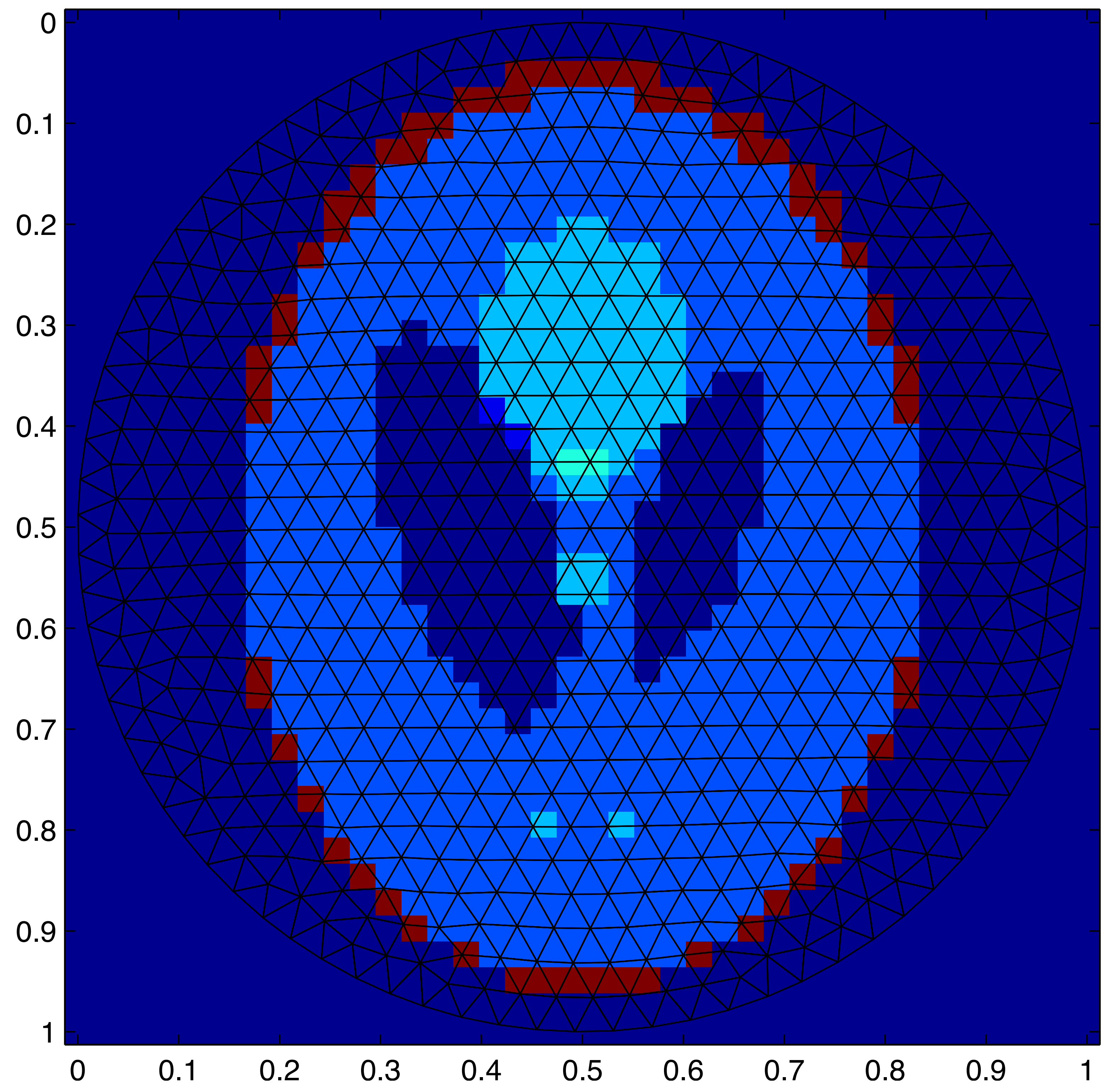


3D DC resistivity



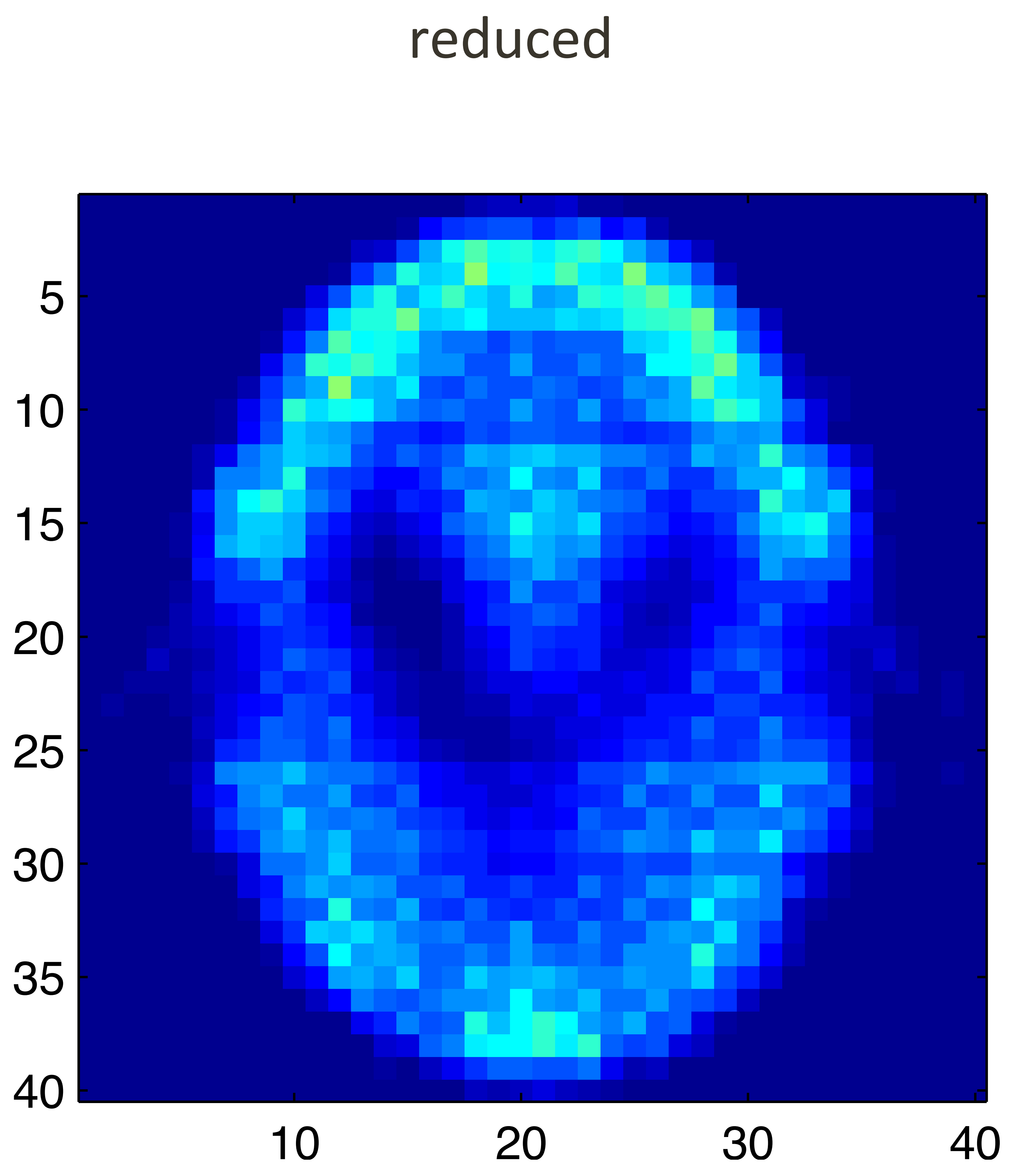


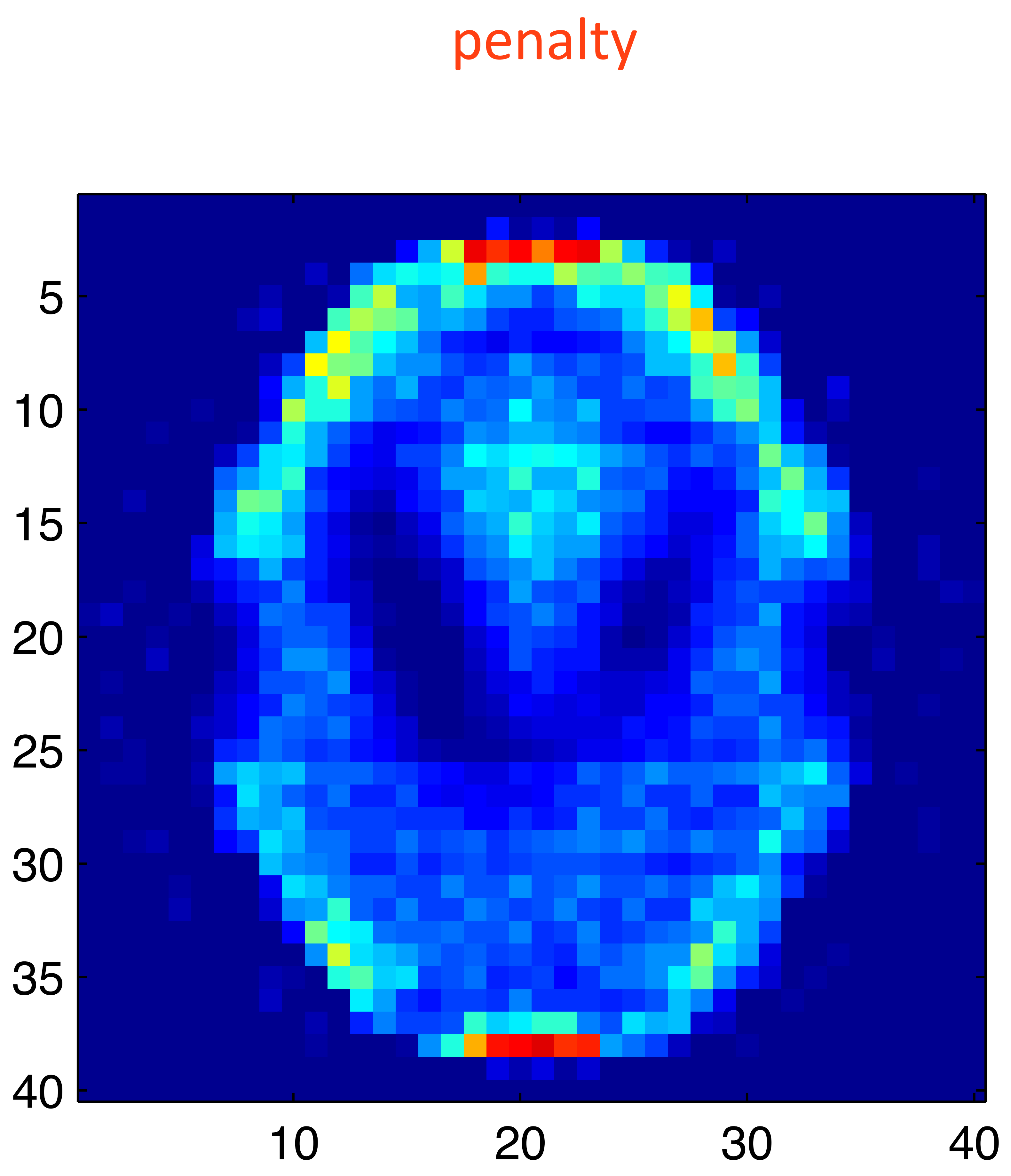
Optical tomography



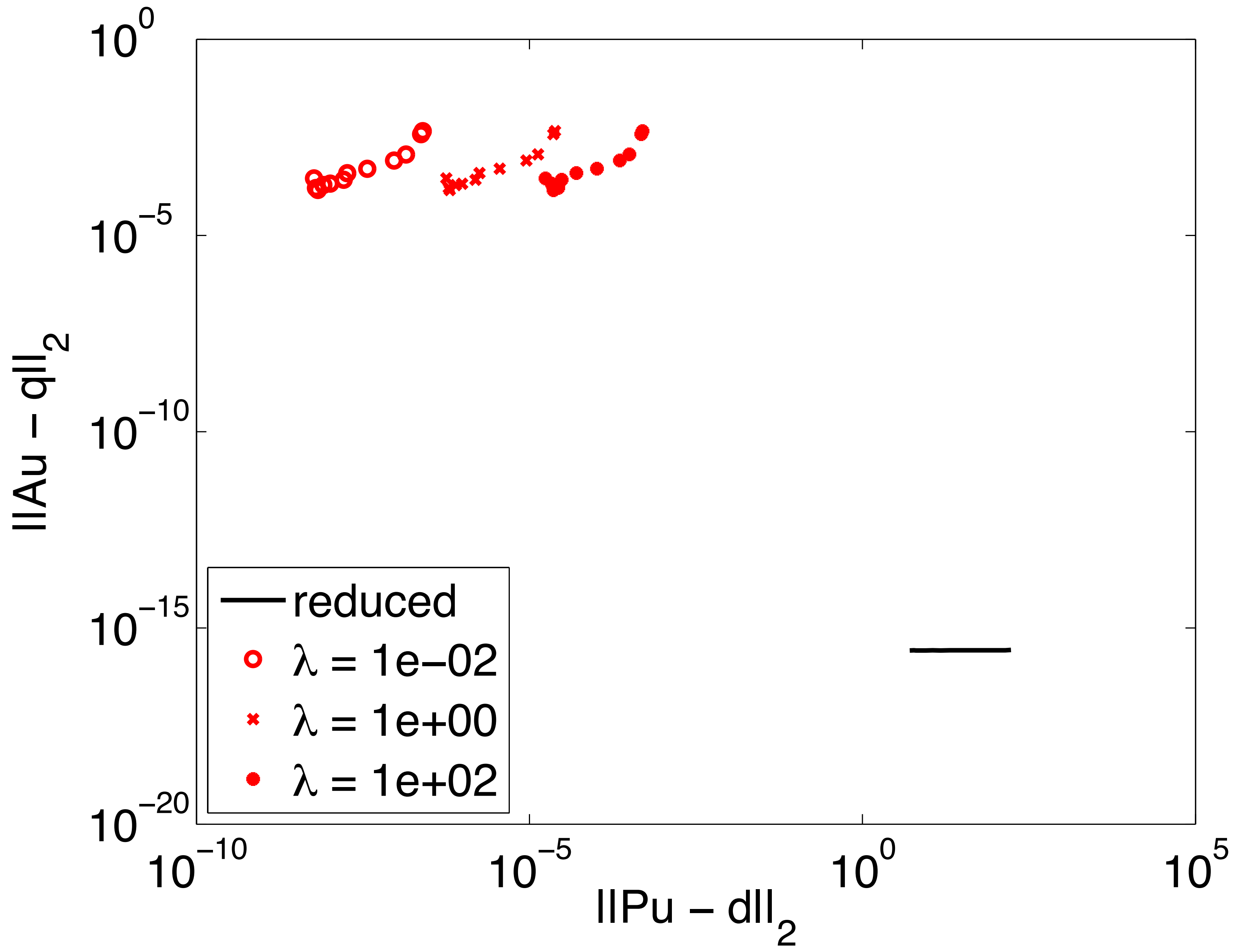
- PDE: $(-\imath\omega/c + \Omega \cdot \nabla + \mu_t)u = 0$
- FV discretization [Abdoulaev '05]
- 1 frequency (400 MHz), 21 angles
- Direct solvers for PDEs
- L-BFGS

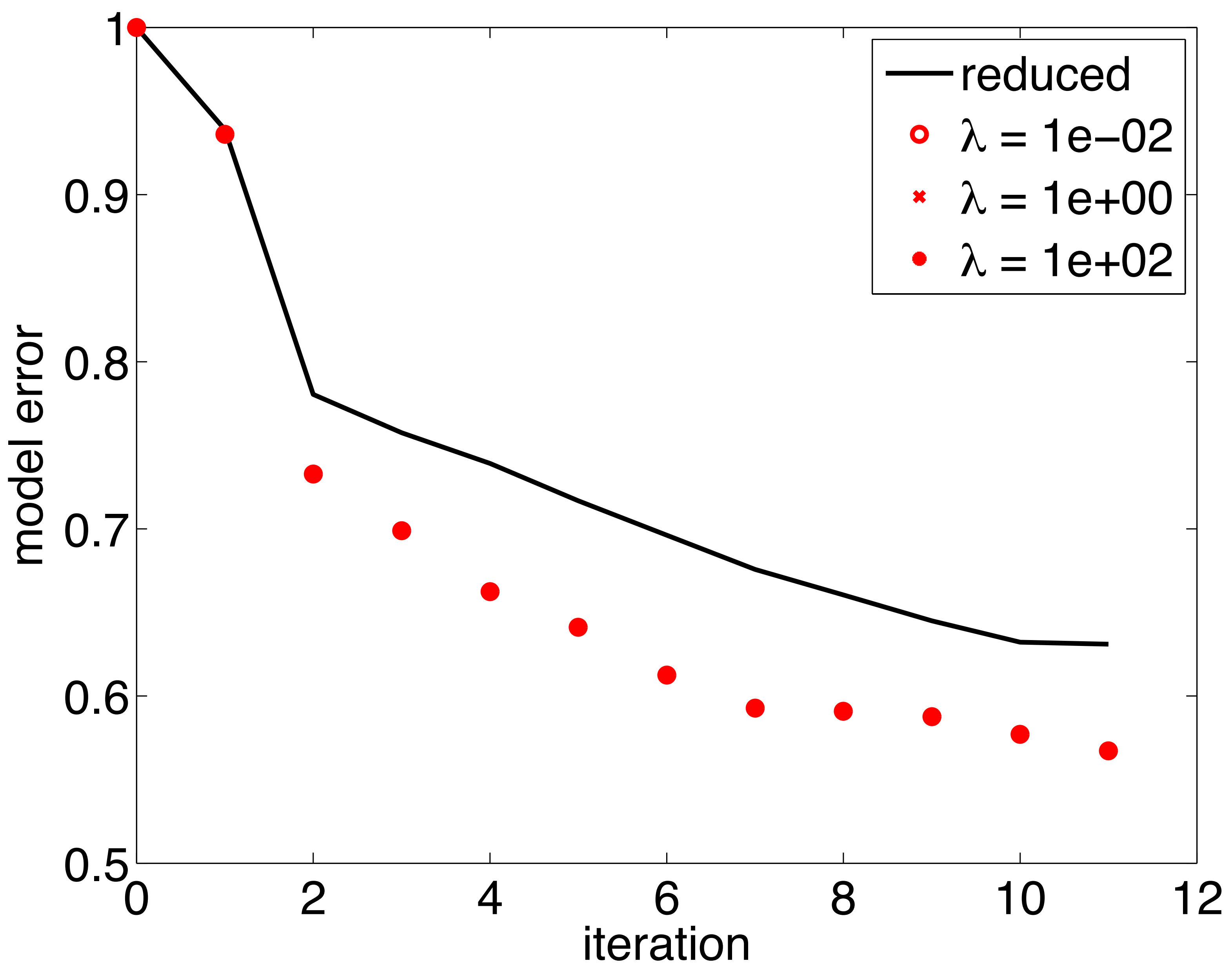
Optical tomography





Optical tomography





Conclusions

New method for *wave*-equation based inversion:

- *extended* search space as in *all-at-once* similar but with memory & CPU requirements as in *reduced* approach
- *no* adjoints & *sparse* GN-Hessian approximation
- `*less* non-linear' problem

Applicable to other PDE-constrained problems:

- DC resistivity
- Optical tomography

Future work

- *fast* solves for *augmented* PDE, preconditioning
- (GN) Hessian approximations
- misfit penalties & (sparse) regularization
- other PDEs
- *extensions* to combat *non*-uniqueness
- ...

Acknowledgements

Thank you for your attention !

https://www.slim.eos.ubc.ca/

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