Examples from the Penalty-method

Bas Peters, Felix J. Herrmann & Tristan van Leeuwen
December 2, 2013

SLIM
University of British Columbia

Motivation

- There are indications that the Penalty method is much less sensitive to local minima
- Penalty method only requires 1 least squares solution, no forward and adjoint-PDE's
- Penalty method can outperform the reduced Lagrangian approach when both methods converge

Reduced Lagrangian method

Least-squares objective:

$$\phi_{\text{red}}(\mathbf{m}) = \frac{1}{2}\sum_{kl} \|PH_k(\mathbf{m})^{-1}\mathbf{q}_{kl} - \mathbf{d}_{kl}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\text{pred}} - \mathbf{d}_{\text{obs}}\|_2^2$$

$\mathbf{m}$: model
P : Restriction to receiver locations
k, l : frequency and source index
H_k : discrete Helmholtz system
$\mathbf{q}_{kl}$: source term
$\mathbf{d}_{kl}$: observed data

Reduced Lagrangian method

Least-squares objective:

$$\phi_{\mathrm{red}}(\mathbf{m}) = \frac{1}{2}\sum_{kl} \|PH_k(\mathbf{m})^{-1}\mathbf{q}_{kl} - \mathbf{d}_{kl}\|_2^2 = \frac{1}{2}\|\mathbf{d}_{\mathrm{pred}} - \mathbf{d}_{\mathrm{obs}}\|_2^2$$

with the gradient (via the adjoint-state method):

$$\nabla_{\mathbf{m}}\phi_{\mathrm{red}} = \sum_{kl} G_{kl}^{*}\mathbf{v}_{kl}$$

where

G_{kl}^{*} is the partial derivative of the discrete Helmholtz system

$\mathbf{v}_{kl}$ is the adjoint field/back propagated data residue

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

Objective:

Data-misfit ⬇ PDE-misfit ⬇

$$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\sum_{kl} \underbrace{\|P\bar{\mathbf{u}}_{kl} - \mathbf{d}_{kl}\|_2^2}_{} + \frac{\lambda^2}{2}\|H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\|_2^2$$

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

and λ is a tradeoff parameter between PDE-fit and data-fit

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

Objective:

$$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\sum_{kl} \underbrace{\|P\bar{\mathbf{u}}_{kl} - \mathbf{d}_{kl}\|_2^2}^{\text{Data-misfit}} + \frac{\lambda^2}{2}\overbrace{\|H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\|_2^2}^{\text{PDE-misfit}}$$

with gradient:

$$\nabla_{\mathbf{m}}\bar{\phi}_\lambda = \sum_{kl} \lambda^2 G_{kl}(\mathbf{m}, \bar{\mathbf{u}}_{kl})^* \left(H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\right)$$

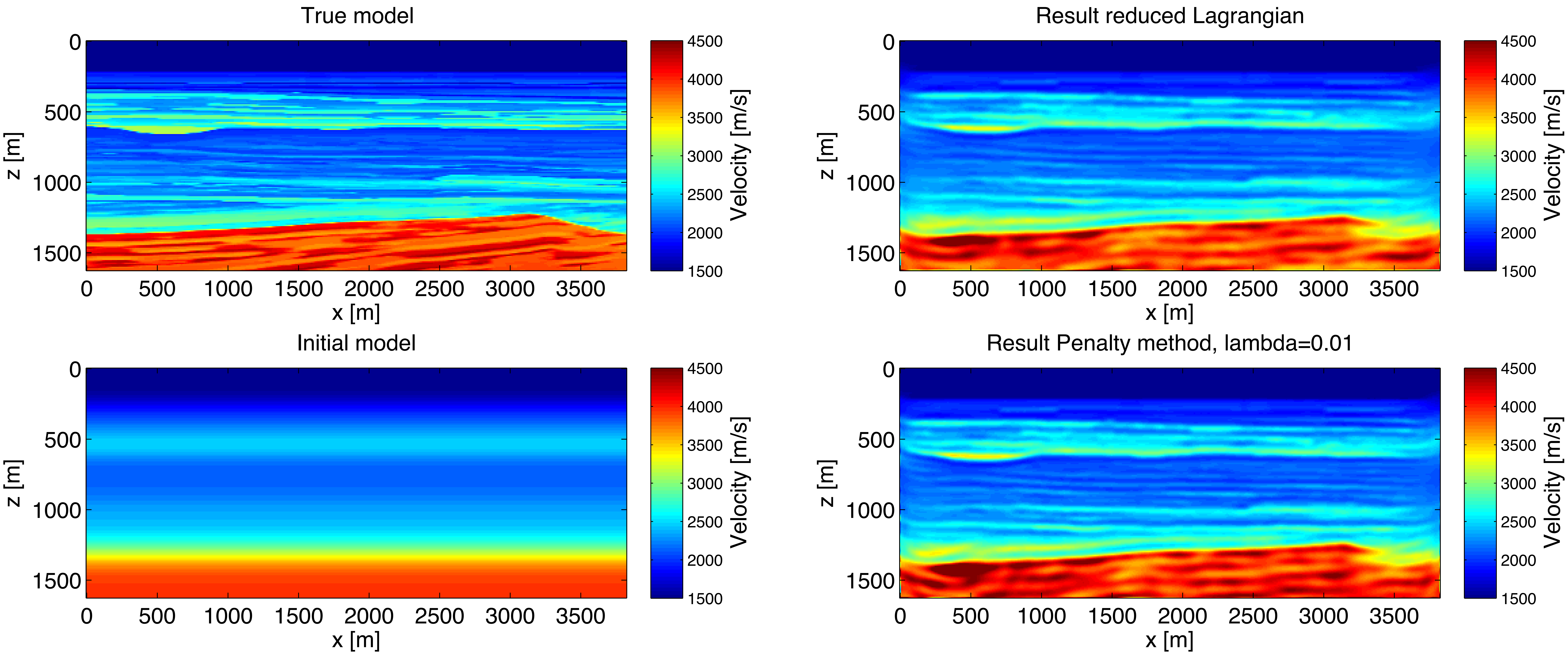
Non-linear waveform inversion

- Used the L-BFGS algorithm
- 64 equally distributed sources and receivers near the surface
- Ricker waveform with 30Hz peak frequency

Example 1a (easy):

- 18 frequency batches (10 iterations each) as {2 3}, {3 4} ,... ,{19 20} Hertz
- No noise

True, initial and final models



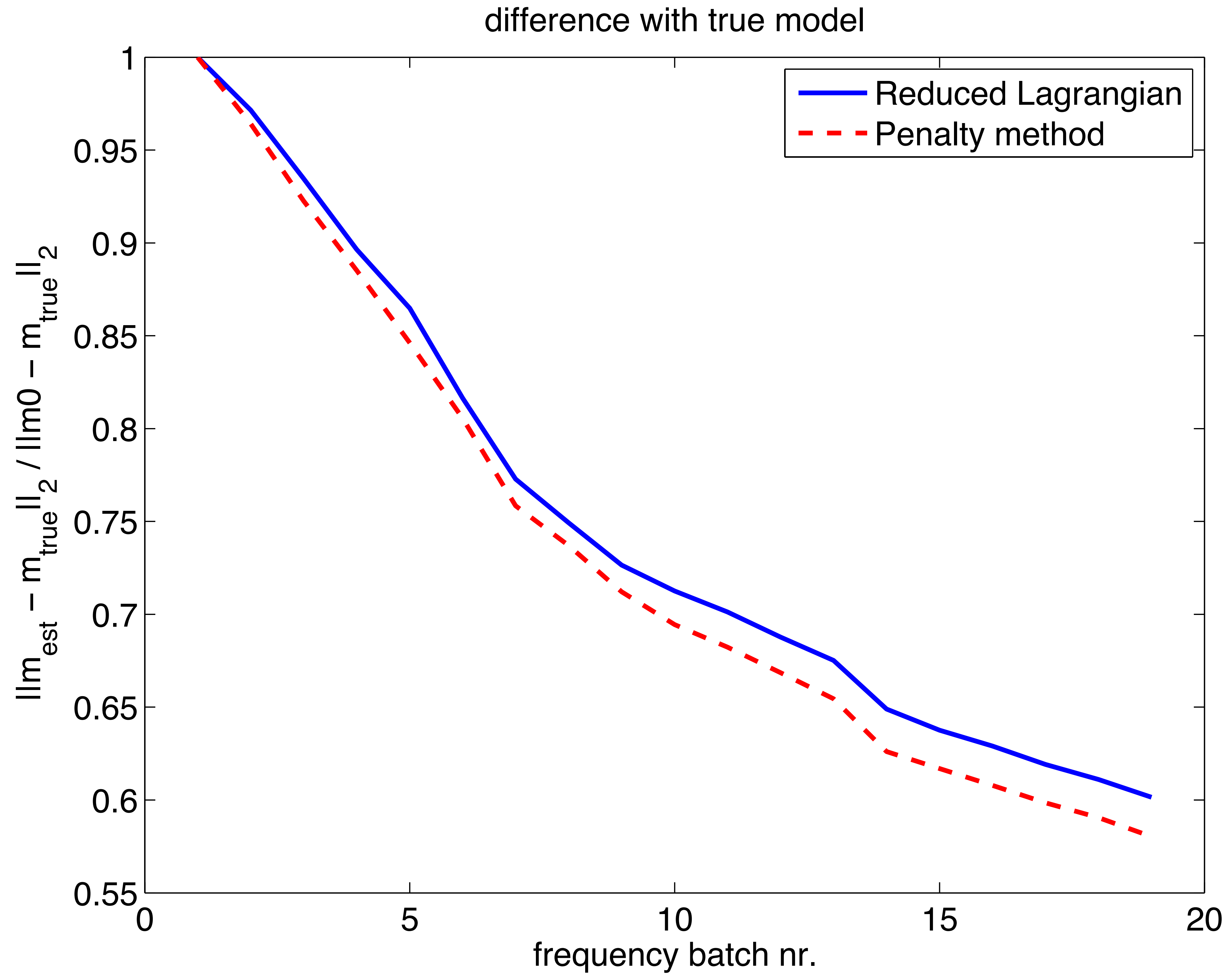
True model

Result reduced Lagrangian

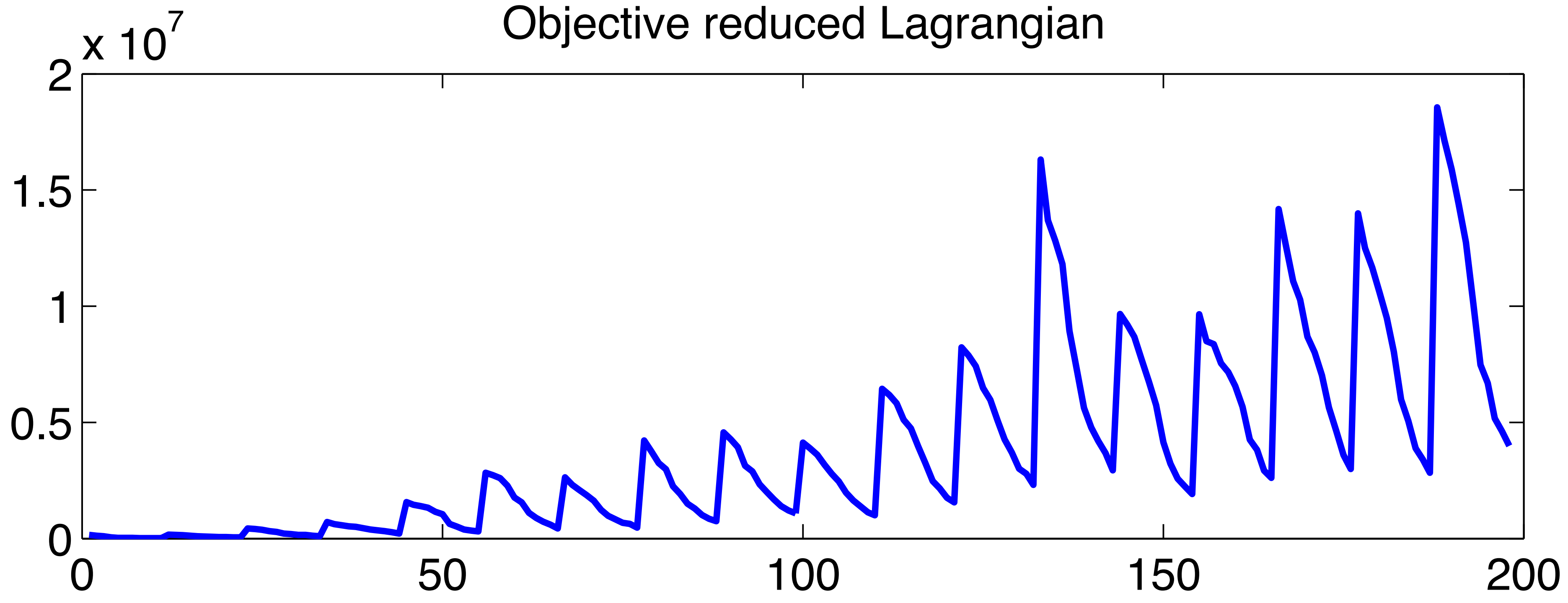
Initial model

Result Penalty method, lambda=0.01

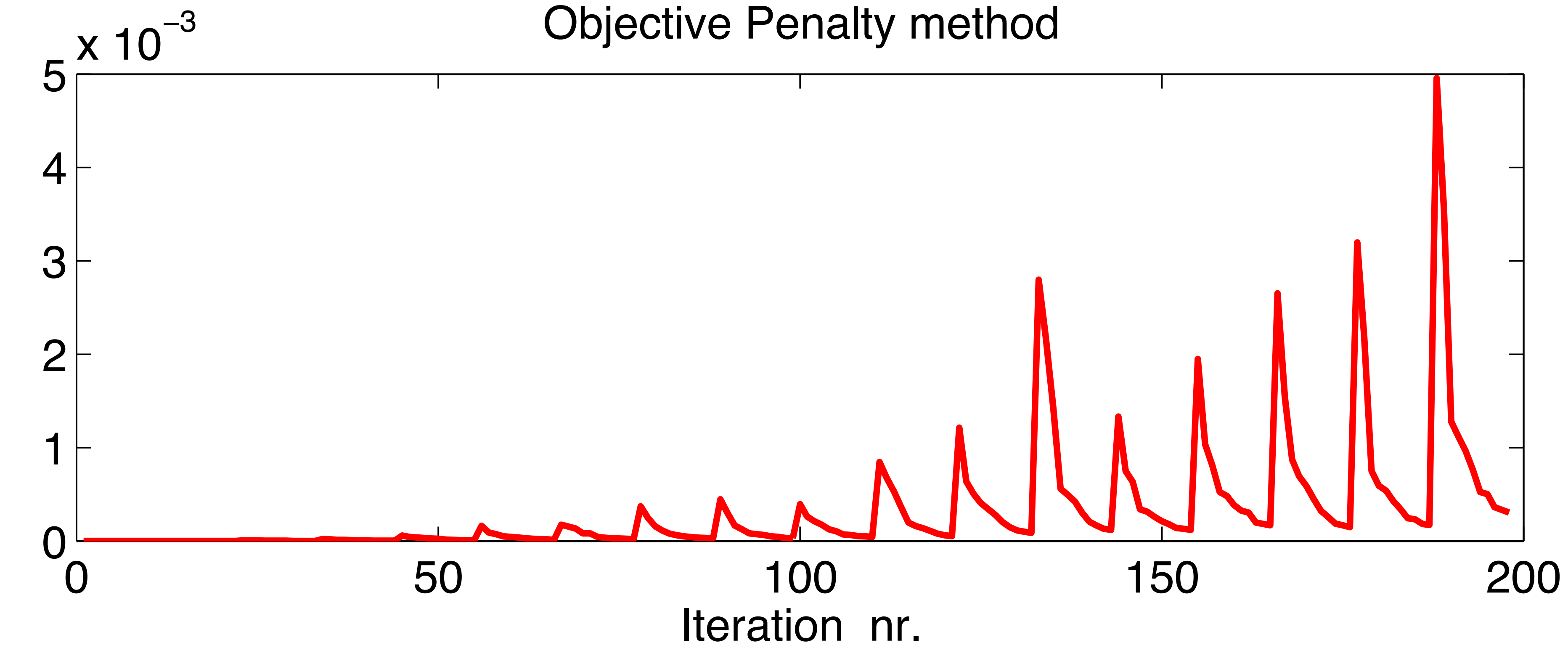
Objective and model error



$$\phi_{\text{red}}(\mathbf{m}) = \frac{1}{2}\|\mathbf{d}_{\text{pred}} - \mathbf{d}_{\text{obs}}\|_2^2$$



Do not compare 1 to 1



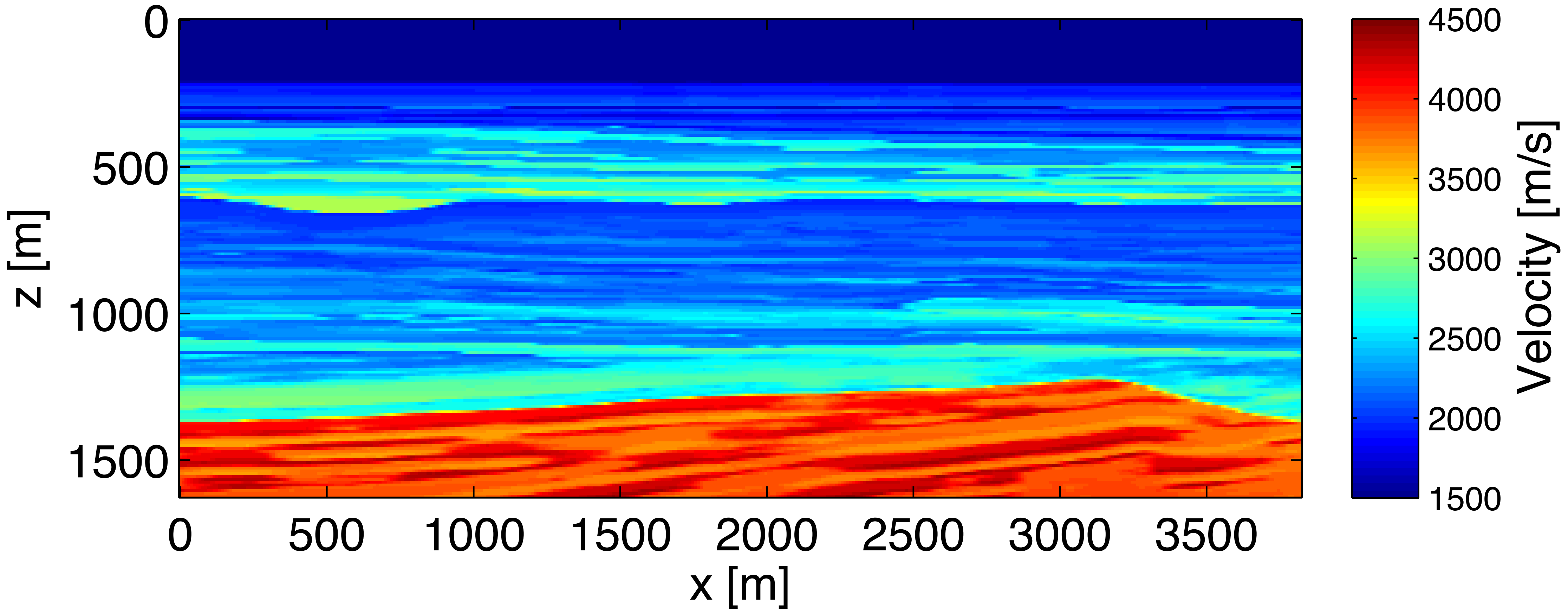
$$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\sum_{kl}\|P\bar{\mathbf{u}}_{kl} - \mathbf{d}_{kl}\|_2^2 + \frac{\lambda^2}{2}\|H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\|_2^2$$

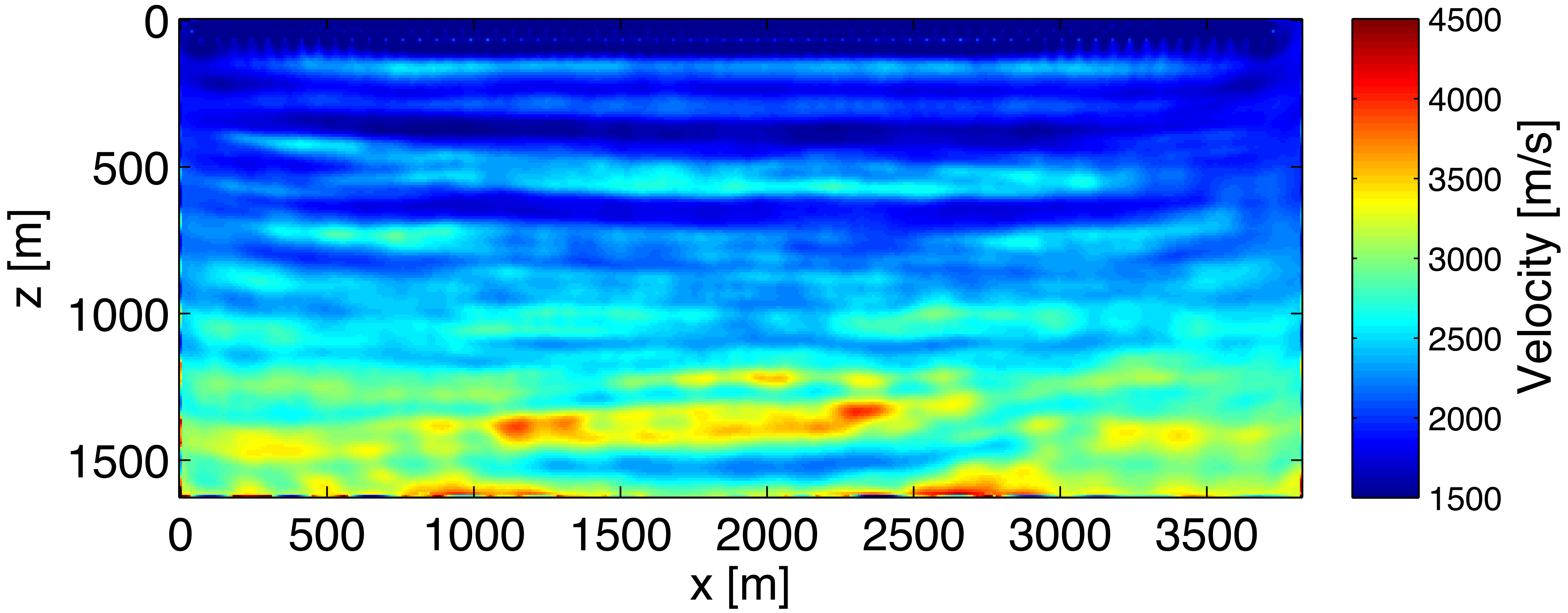
Non-linear waveform inversion

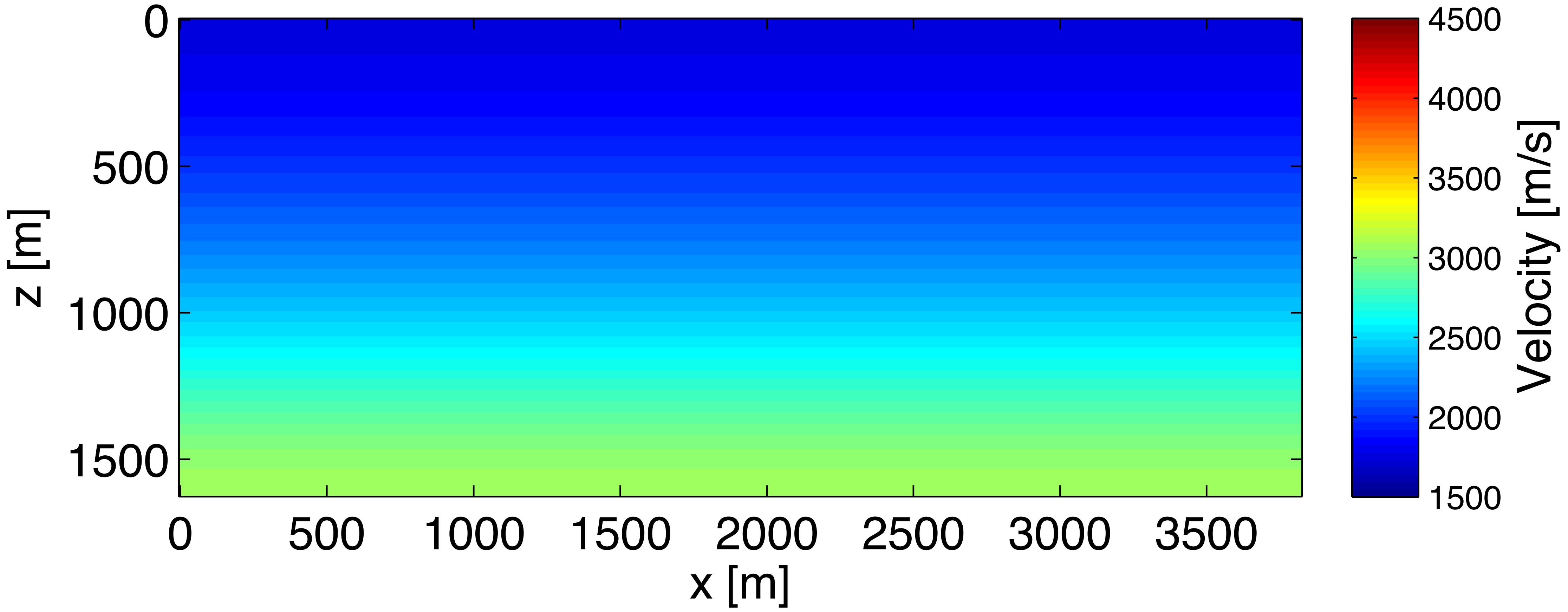
Example 1b (difficult):

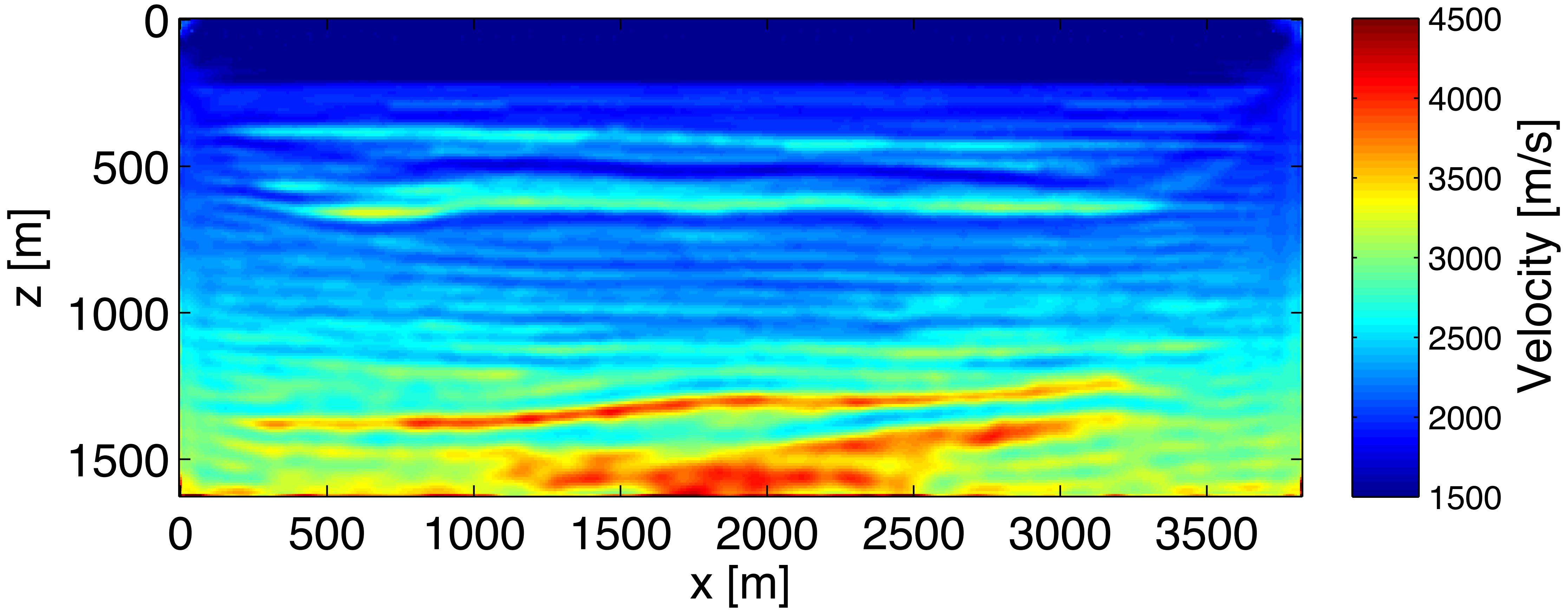
- Lots of low frequencies missing, 14 frequency batches (10 iterations each) as {7}, {7 8} ,{8 9},… ,{19 20} Hertz
- Data contains random noise
- Inaccurate initial model

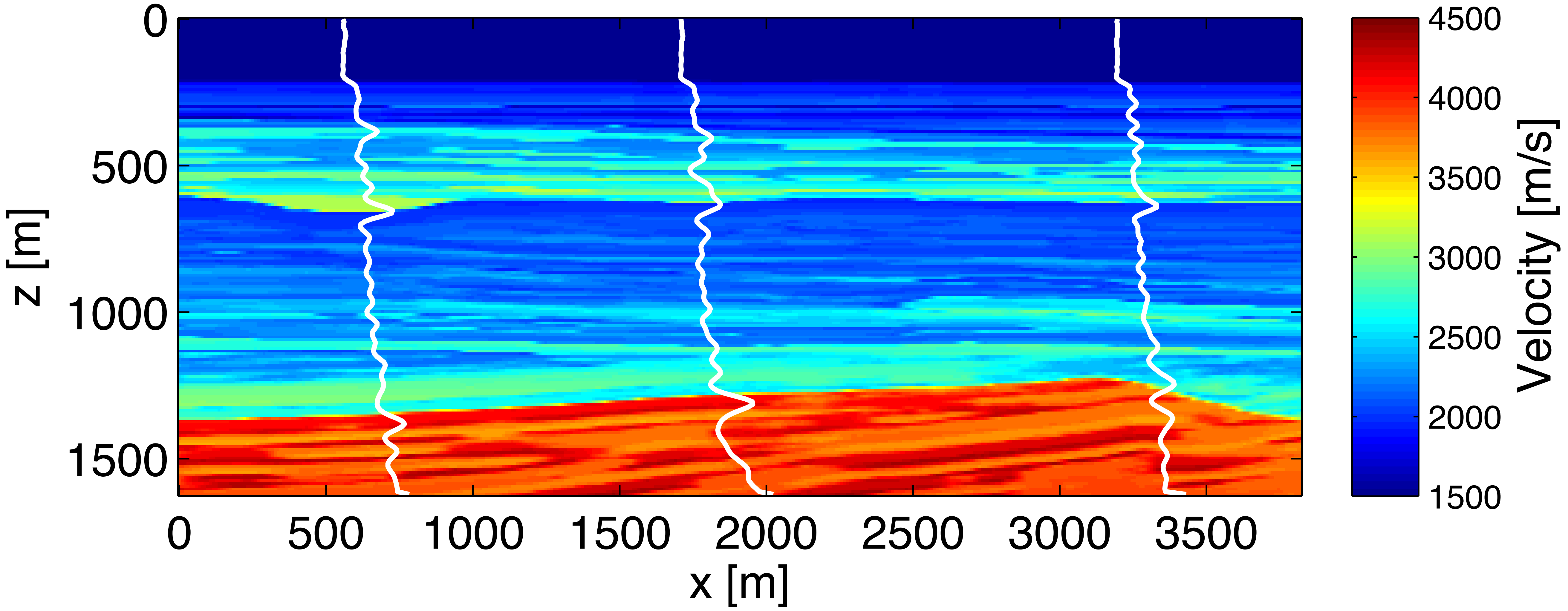
True, initial and final models


True model

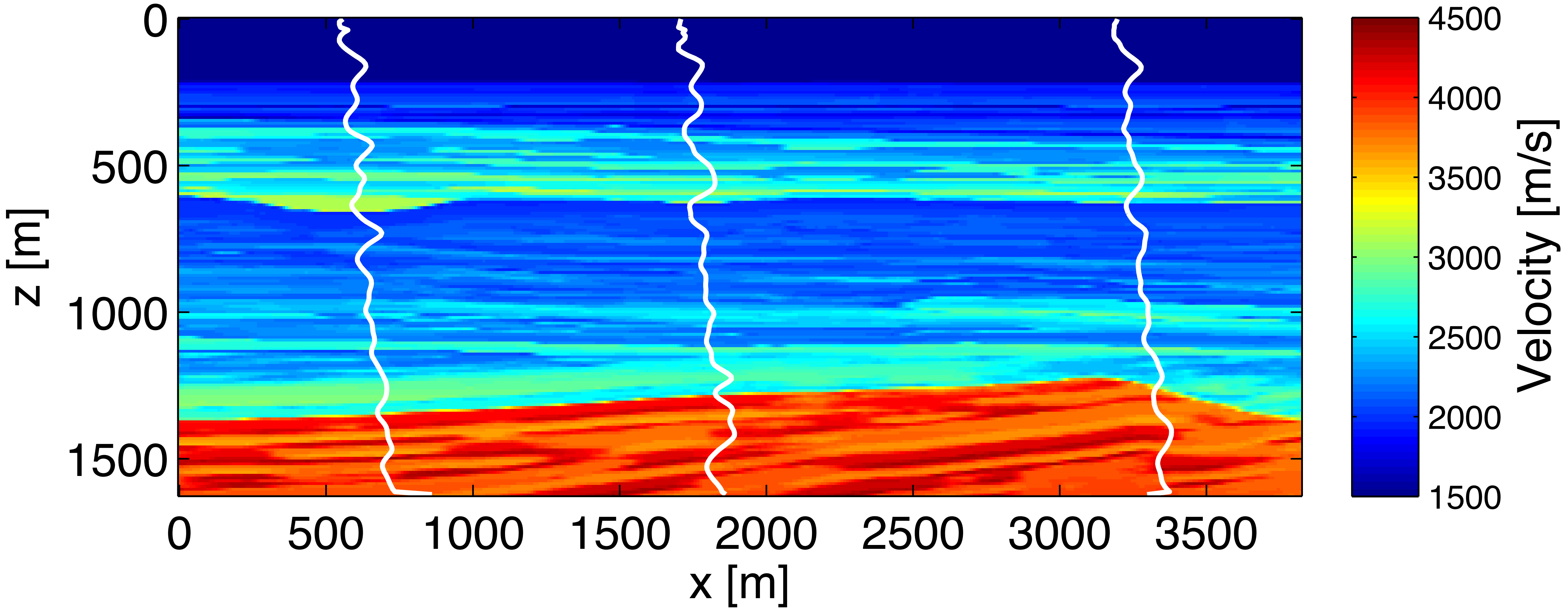

Result reduced Lagrangian


Initial model


Result Penalty method, lambda=1

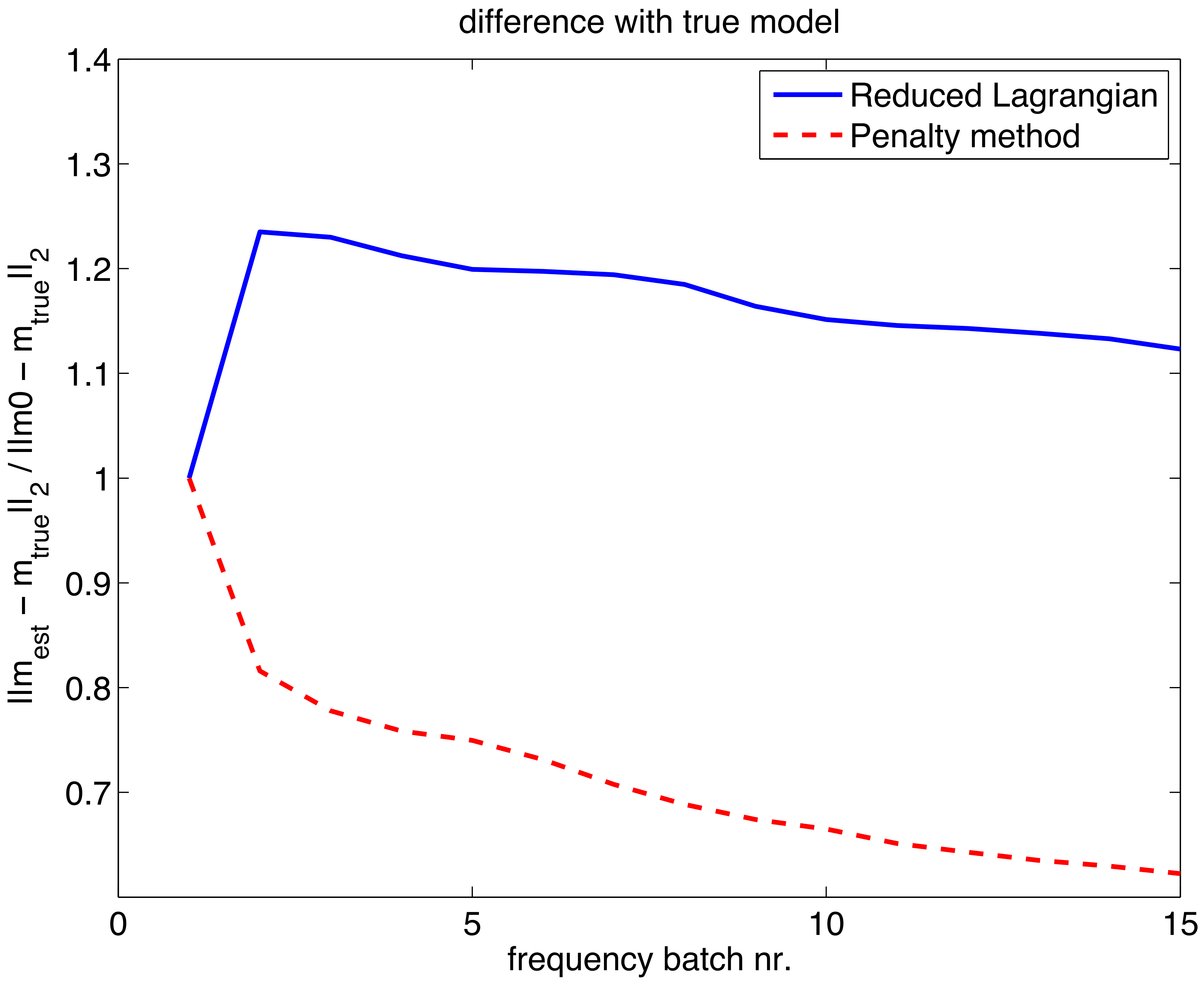


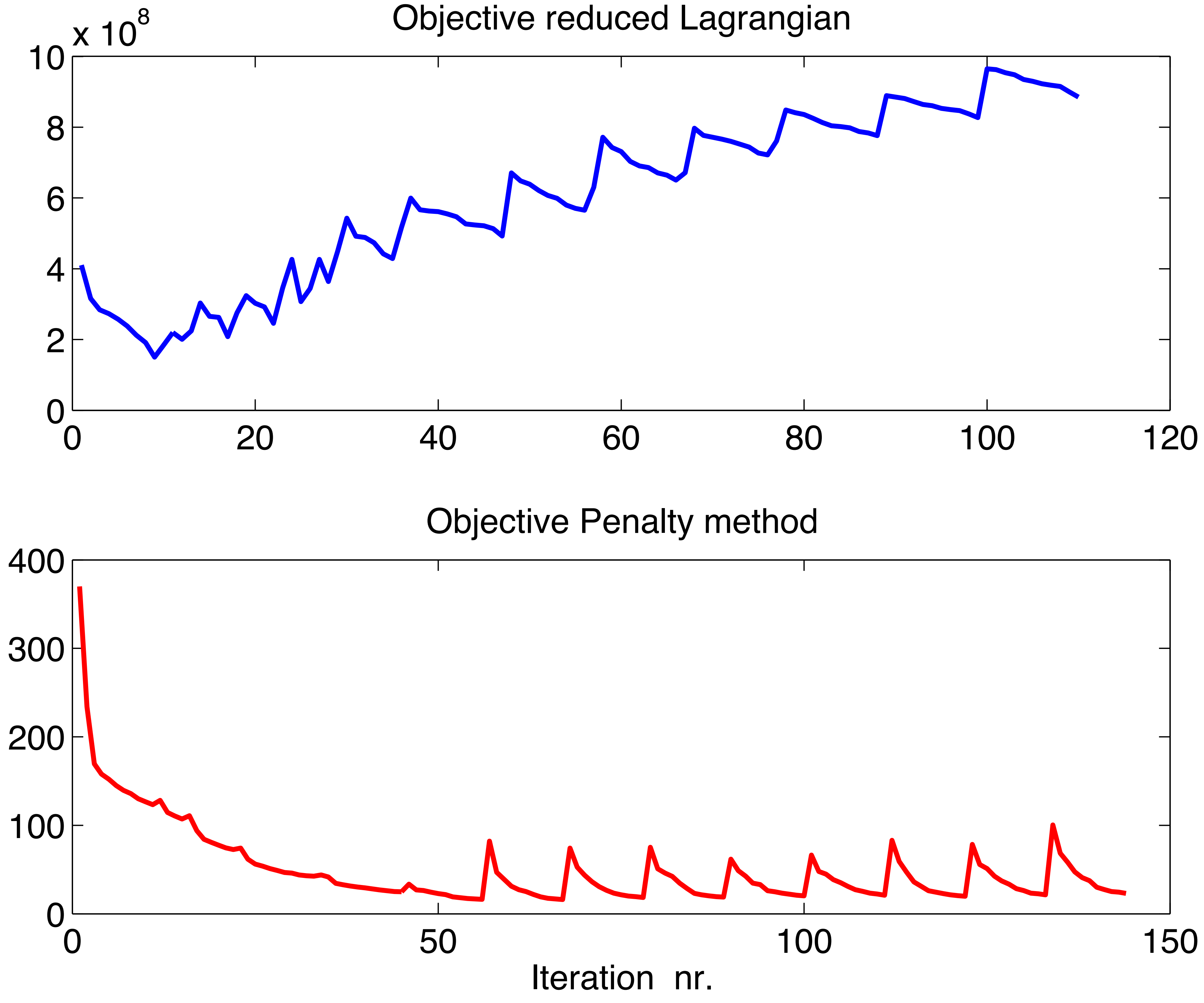
True model, Penalty method result overlay



True model, Reduced Lagrangian result overlay

Objective and model error

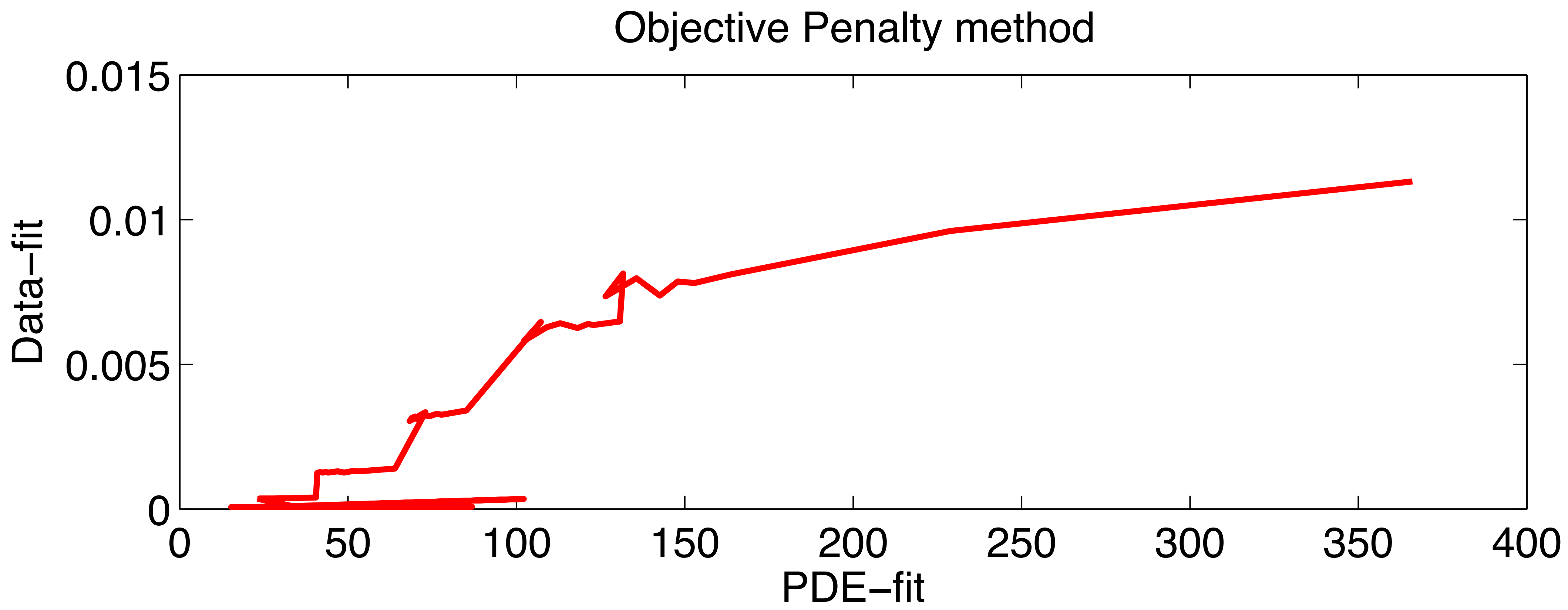


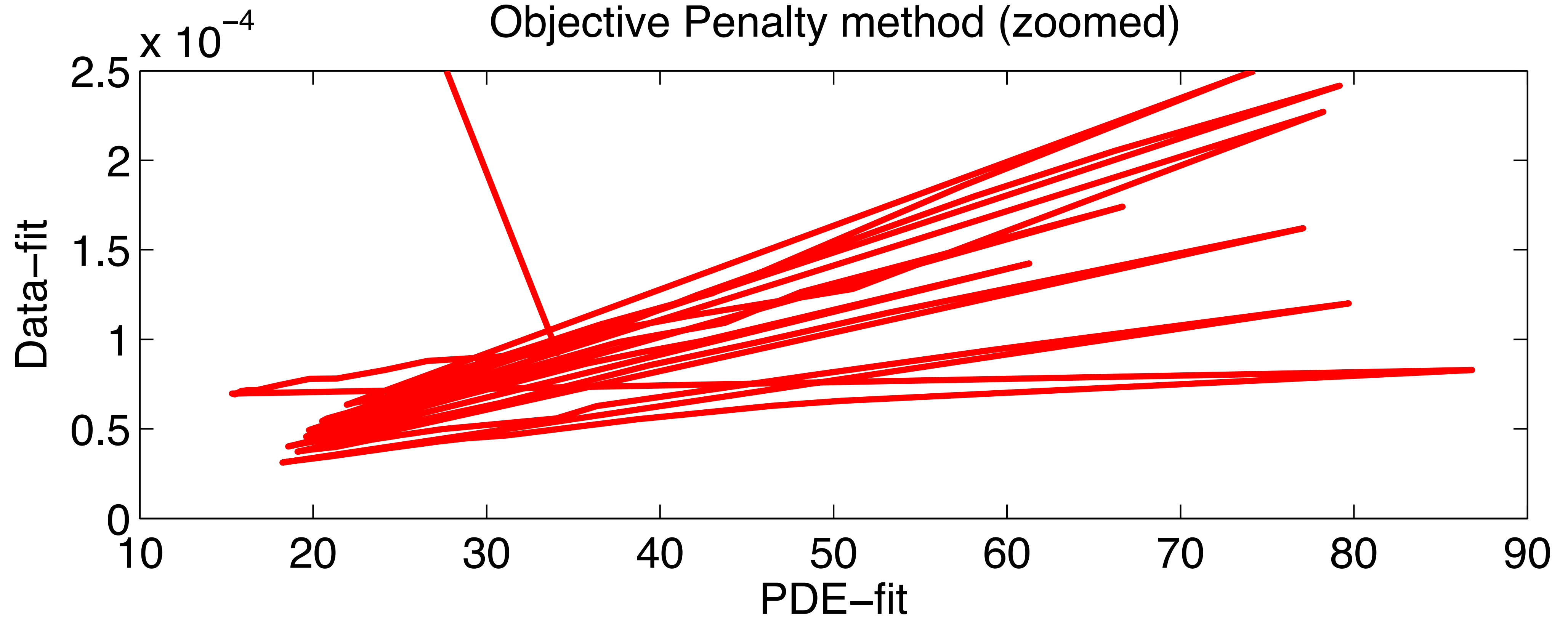


Objective and model error

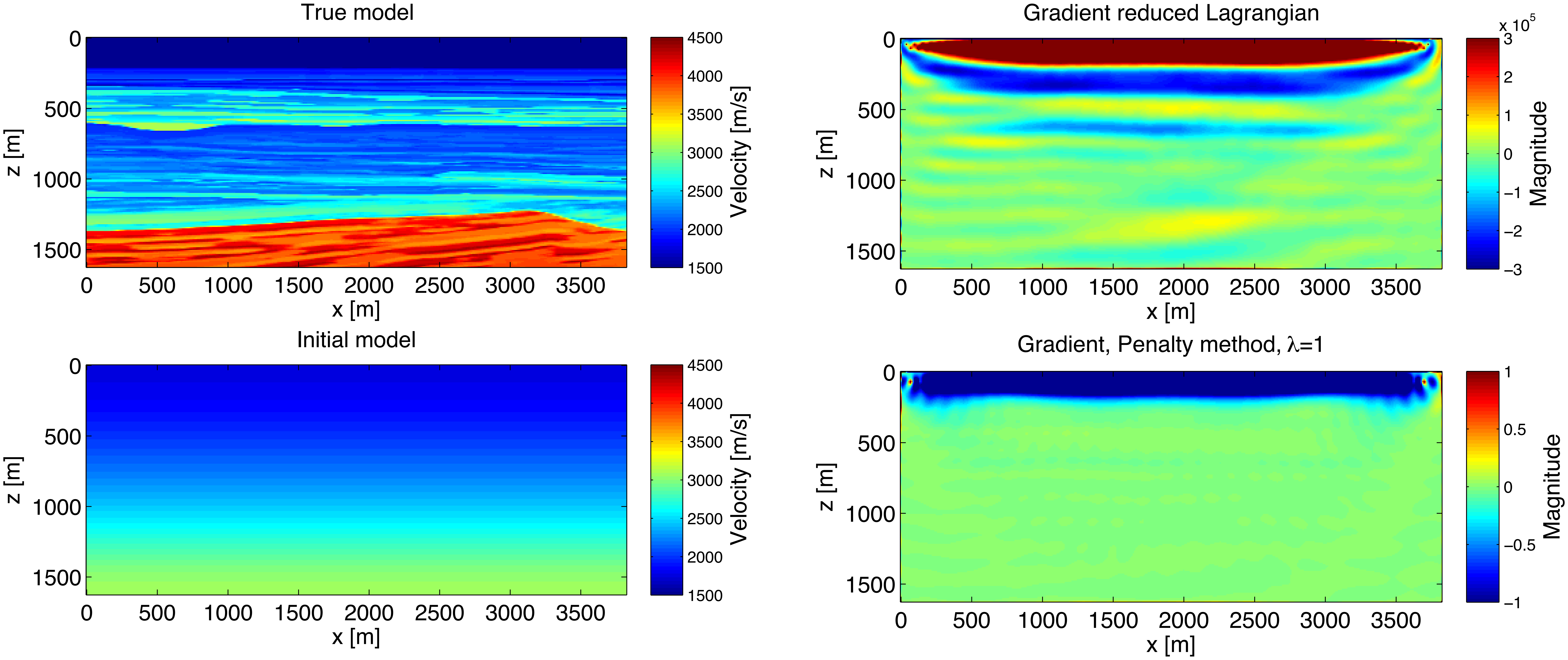
$$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\sum_{kl}\|P\bar{\mathbf{u}}_{kl} \overset{\substack{\text{Data-misfit}\\ \Downarrow}}{-} \mathbf{d}_{kl}\|_2^2 + \frac{\lambda^2}{2}\|H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} \overset{\substack{\text{PDE-misfit}\\ \Downarrow}}{-} \mathbf{q}_{kl}\|_2^2$$

- We can take a look at each part separately:





A look at the first gradient...



True model

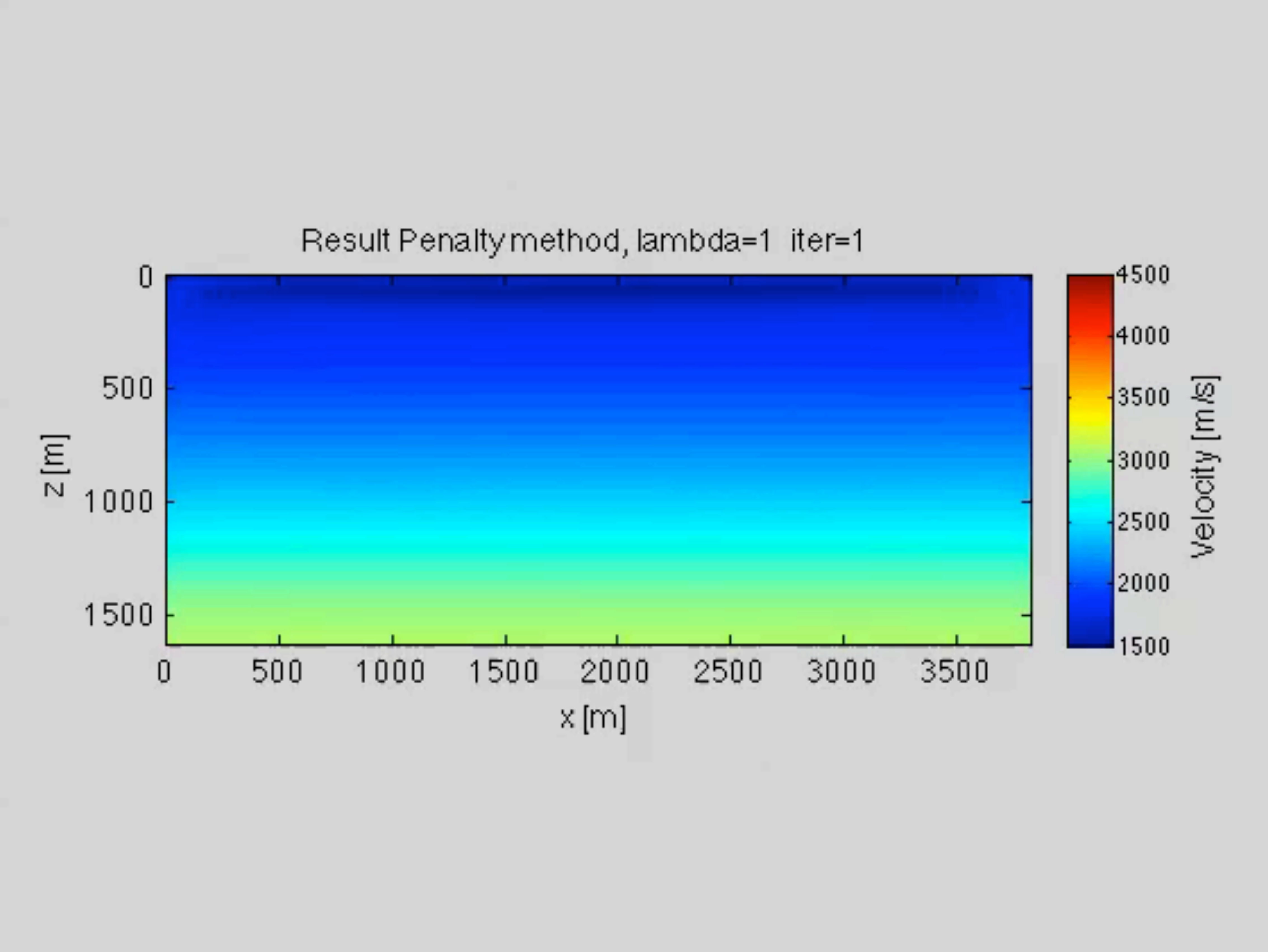
Initial model

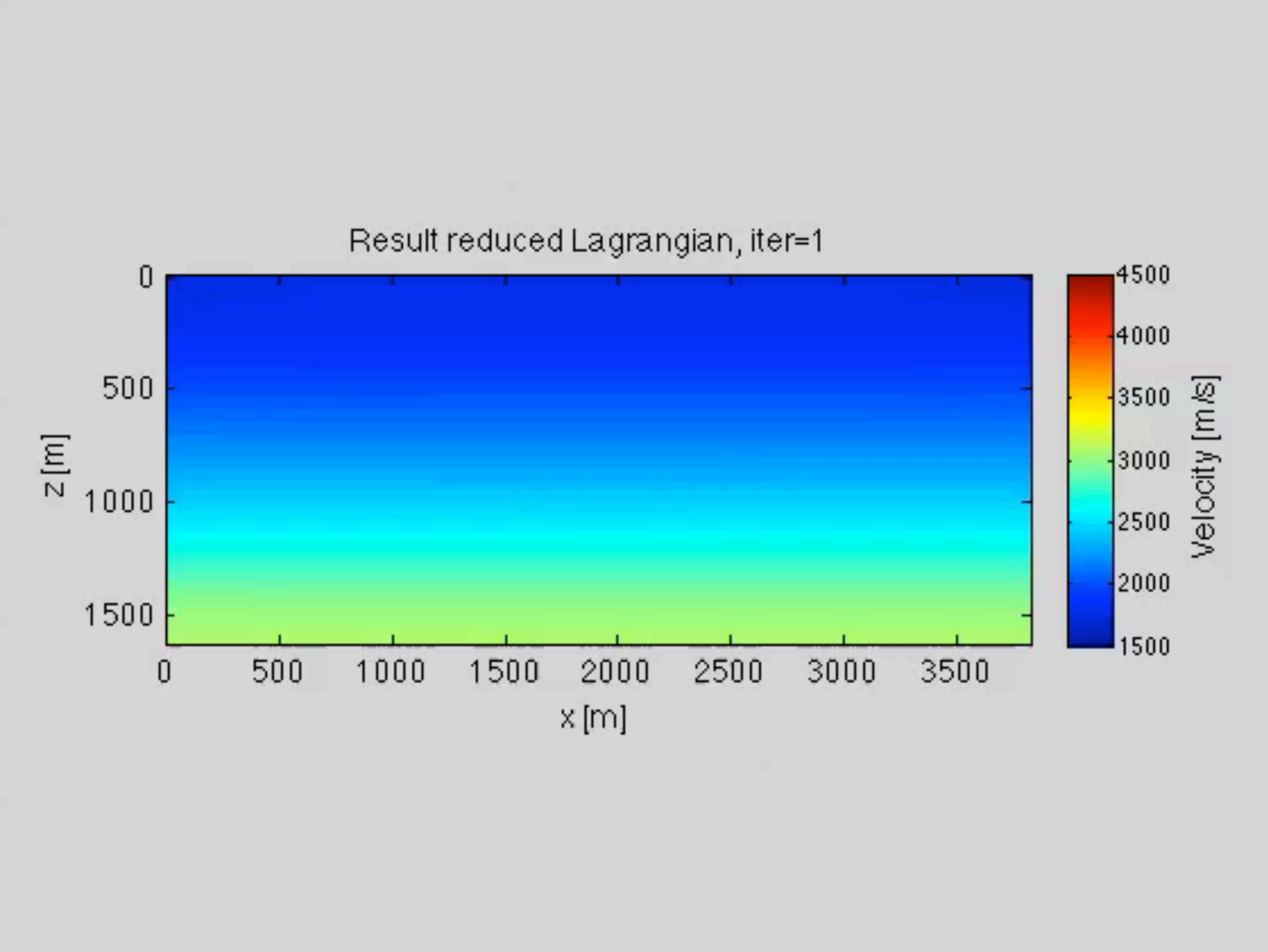
Gradient reduced Lagrangian

Gradient, Penalty method, λ=1

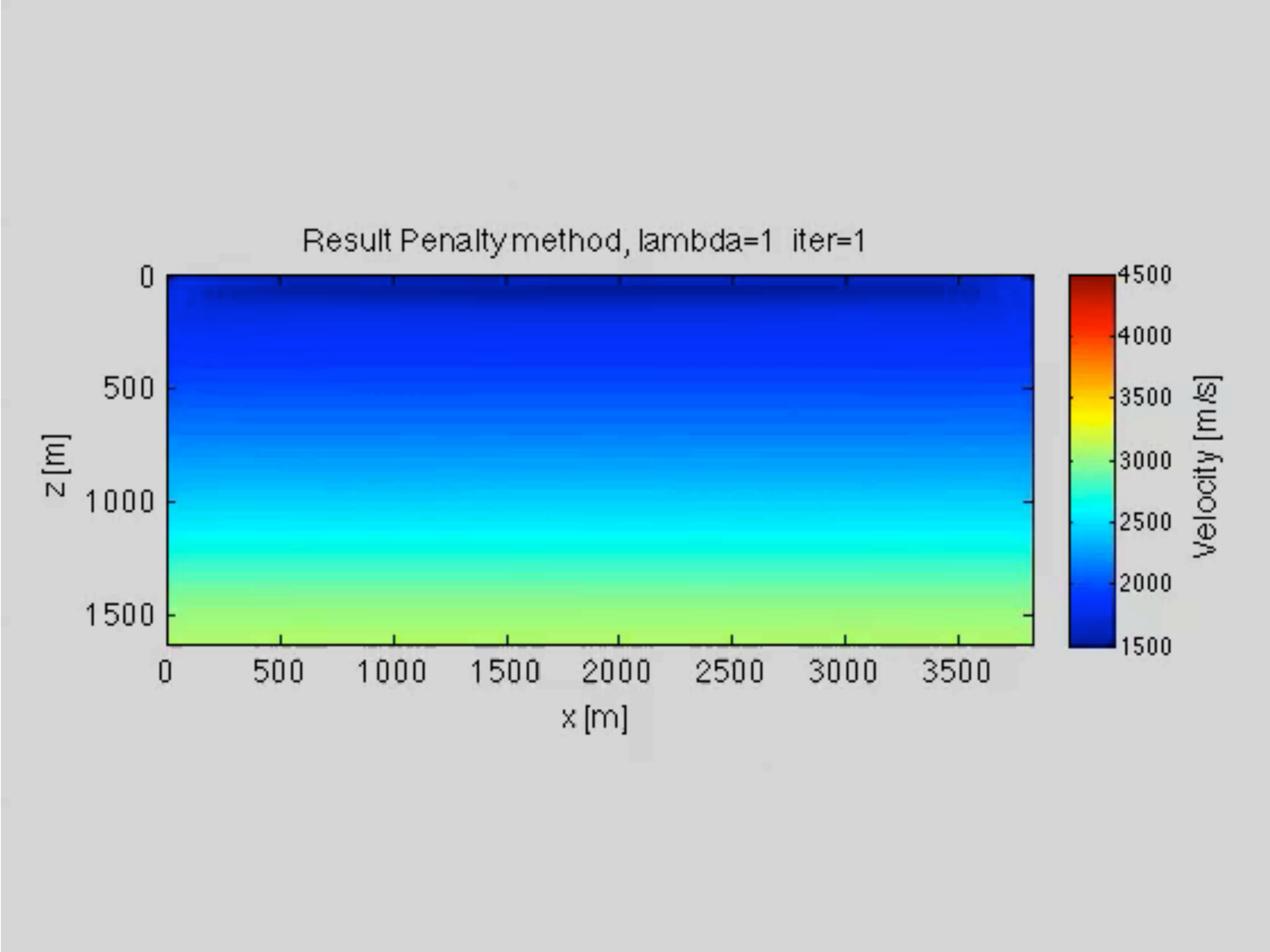
Gradients which will be the first updates, Frequency band: 7 Hz

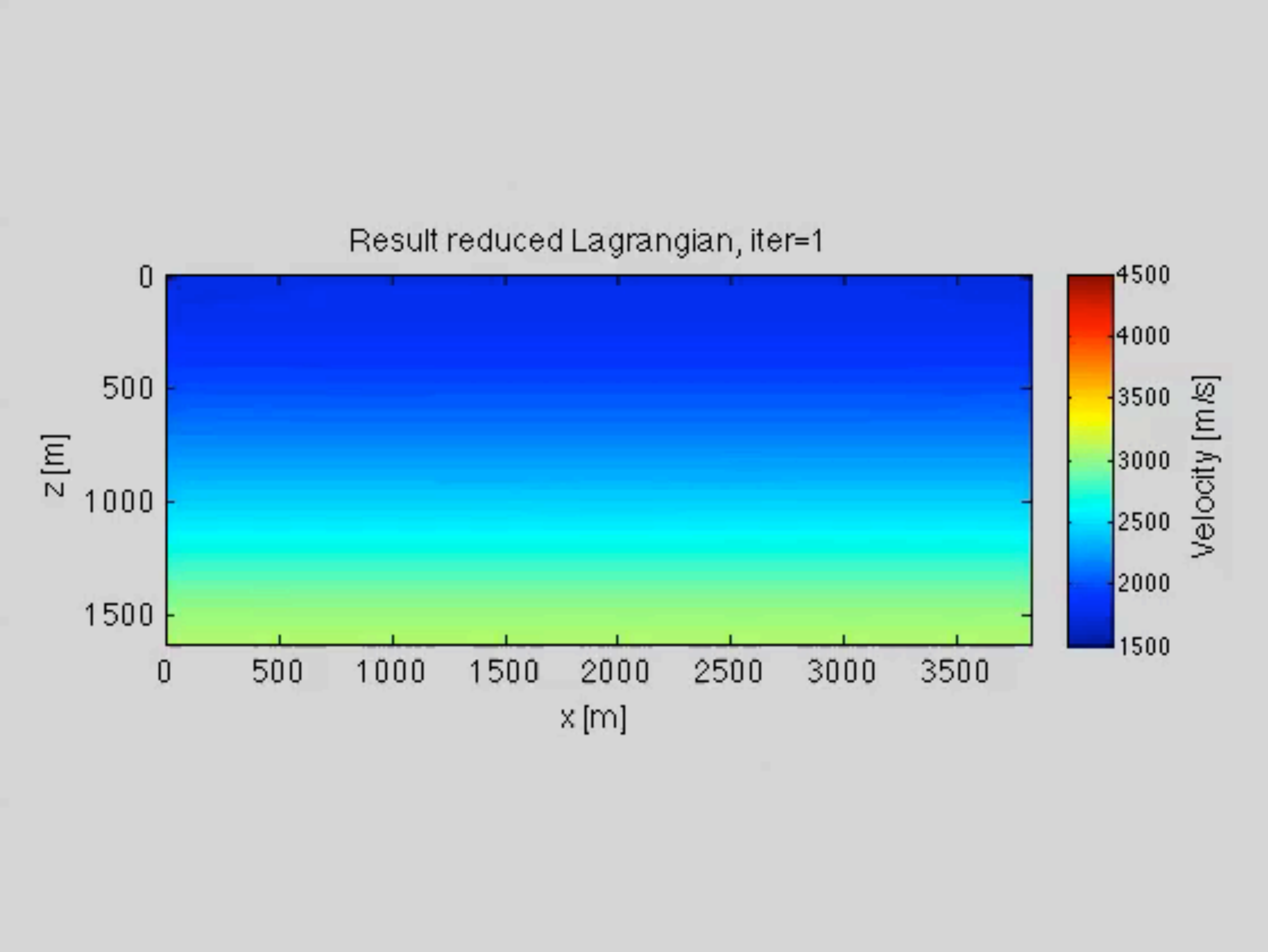
Model estimate at every iteration



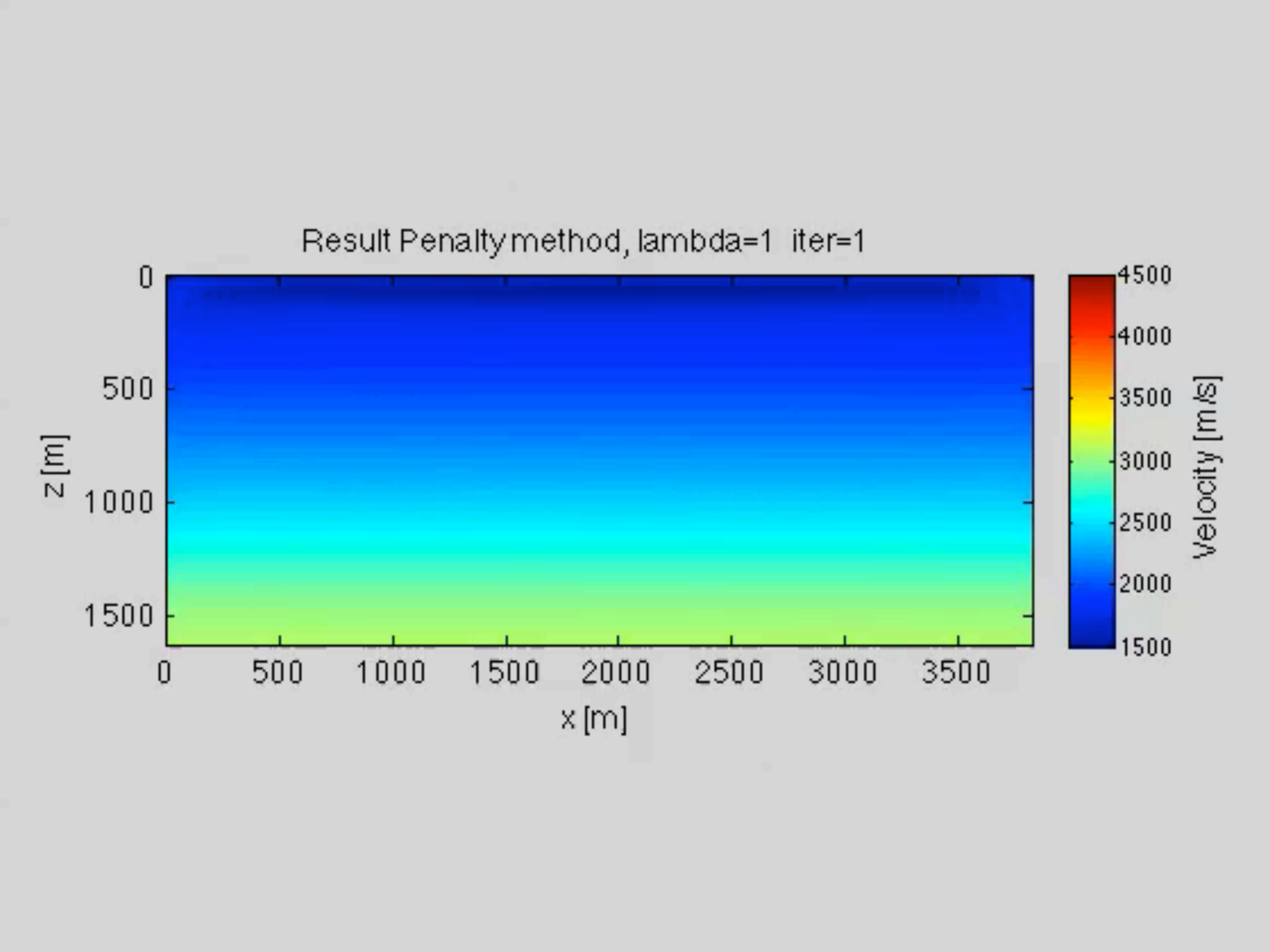


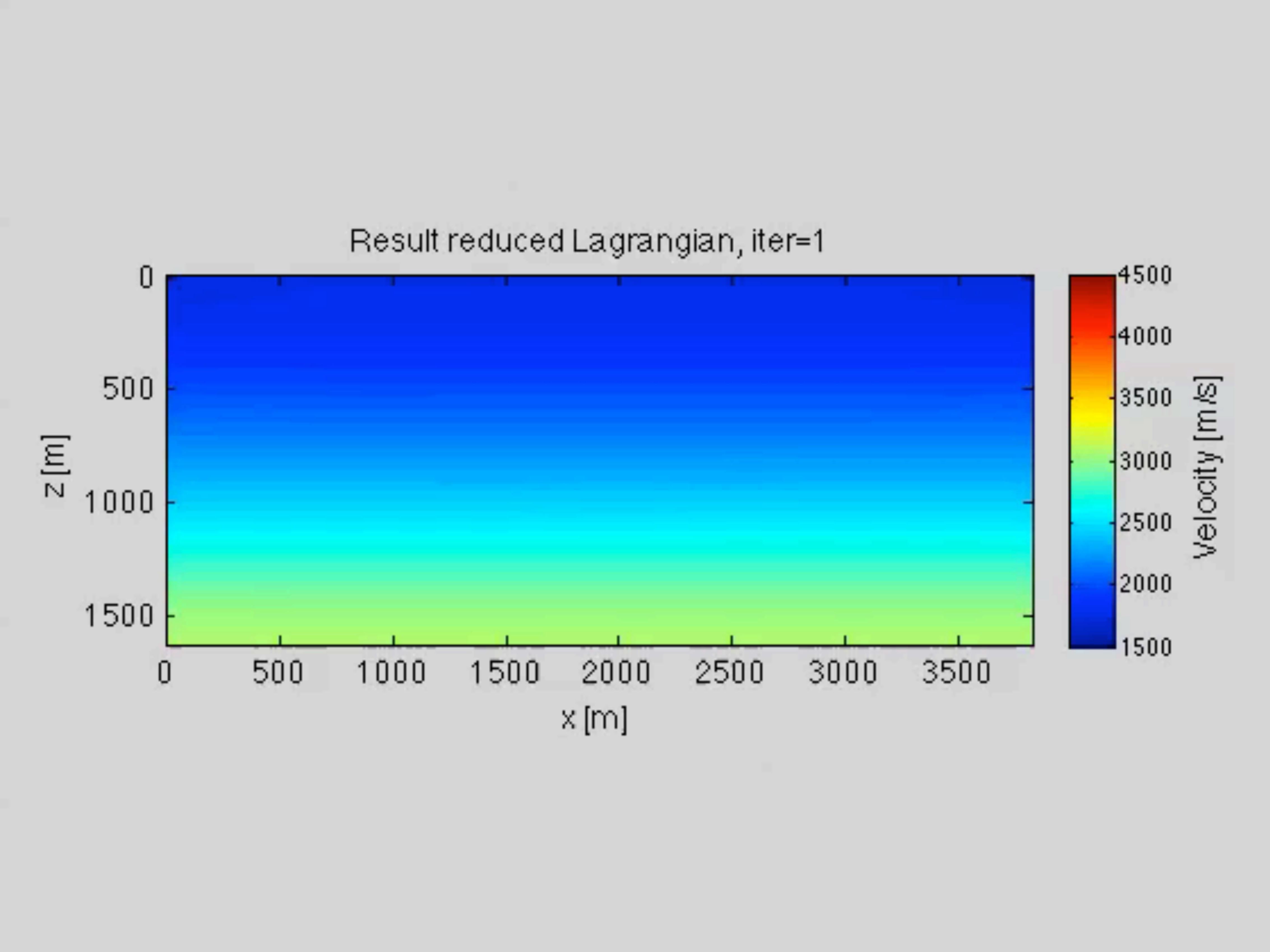
Model estimate at every iteration





Model estimate at every iteration





Observations about waveform inversion

- Penalty method performs much better for hard problems (starting at 7 Hz)
- Penalty method performs a bit better than the reduced Lagrangian method for not so hard problems.

- Penalty method is about 4 times faster (200x400 grid) than the reduced Lagrangian approach, because
  - 1 least-squares problem vs 2 PDE-system-solves per evaluation of the objective & gradient
  - smaller number of line-search steps in l-BFGS (just observed)

Imaging (RTM & ...)

Goal: find the singularities (interfaces) in the medium

Given: smooth function (background velocity), observed data

General recipe for imaging:

1: Formulate an objective functional (least-squares, penalty, ...)

2: Compute the gradient w.r.t. medium parameters (=image)

Penalty-method:

$$\nabla_{\mathbf{m}}\bar{\phi}_{\lambda} = \sum_{kl} \lambda^{2} G_{kl}(\mathbf{m}, \bar{\mathbf{u}}_{kl})^{*} \big(H_{k}(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\big)$$

Reduced Lagrangian:

$$\nabla_{\mathbf{m}}\phi_{\mathrm{red}} = \sum_{kl} G_{kl}^{*} H_{k}(\mathbf{m})^{-*}(-P^{*}(P\mathbf{u}_{kl} - \mathbf{d}_{kl}))$$

Imaging (RTM & ...)

- In case the Helmholtz equation is discretized as:

$$[\nabla^2 + \omega^2 \mathrm{diag}(\mathbf{m})]\mathbf{u} = \mathbf{q} \rightarrow [L + \omega^2 \mathrm{diag}(\mathbf{m})]\mathbf{u} = \mathbf{q}$$

- The imaging operation can be explicitly written as:

Reduced Lagrangian: $\nabla_{\mathbf{m}} \phi_{\mathrm{red}} = \sum \omega^2 \mathrm{diag}(\mathbf{u})^* \mathbf{v}$

with back prop. data residue $\mathbf{v} = H(\mathbf{m})^{-*}(P^*(\mathbf{d} - P\mathbf{u}))$

Penalty-method: $\nabla_{\mathbf{m}} \bar{\phi}_\lambda = \lambda^2 \sum \omega^2 \mathrm{diag}(\bar{\mathbf{u}})^* \delta \bar{\mathbf{u}}$

with PDE-residual $\delta \bar{\mathbf{u}} = H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}$

Imaging

- 150 sources and receivers near the surface
- 30 frequencies from 4 to 50 Hz
- Ricker waveform with 30Hz peak frequency

Penalty-method imaging

For every choice of λ in

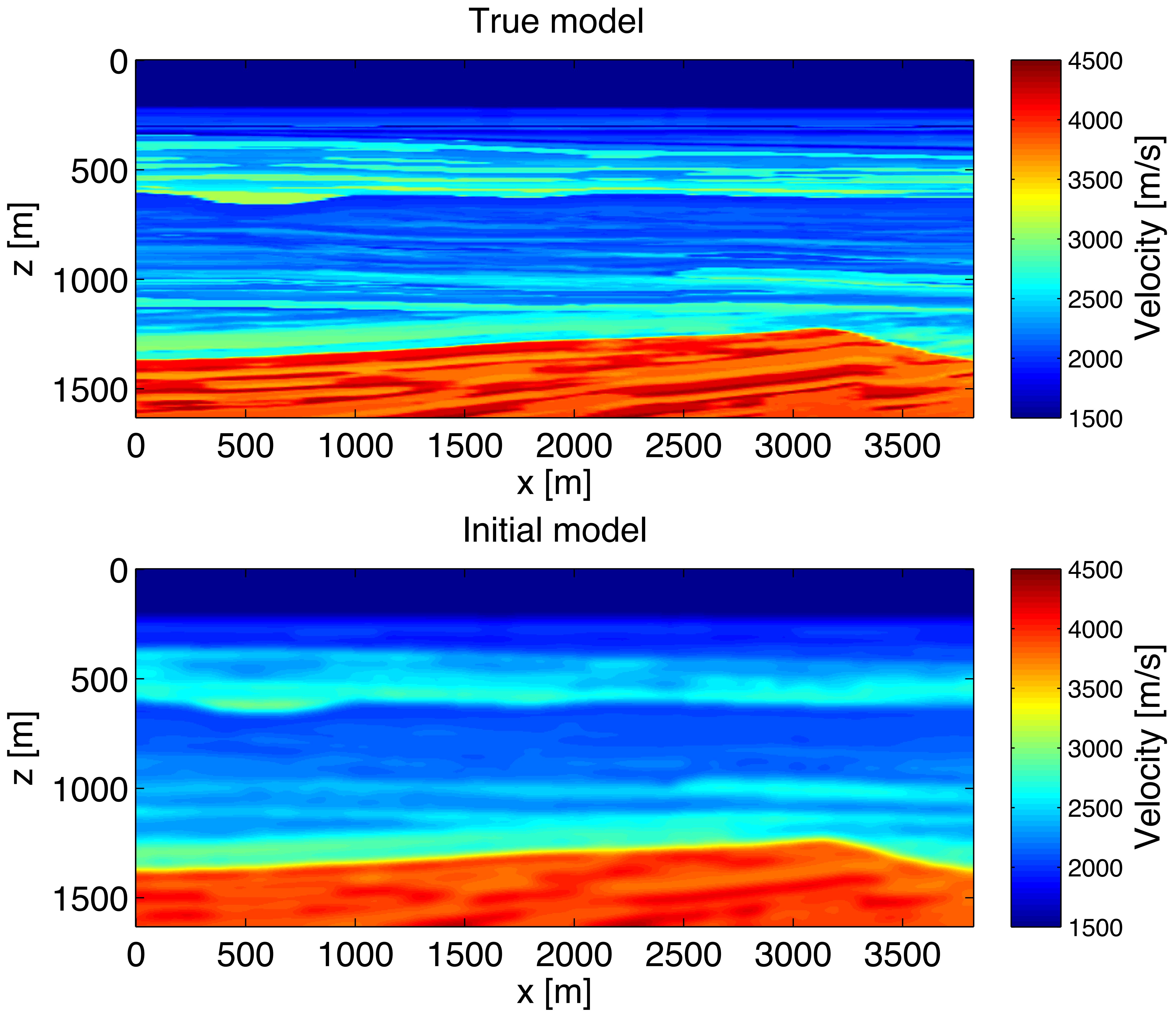
$$\nabla_{\mathbf{m}}\bar{\phi}_\lambda = \sum_{kl} \lambda^2 G_{kl}(\mathbf{m}, \bar{\mathbf{u}}_{kl})^* \left(H_k(\mathbf{m})\bar{\mathbf{u}}_{kl} - \mathbf{q}_{kl}\right)$$

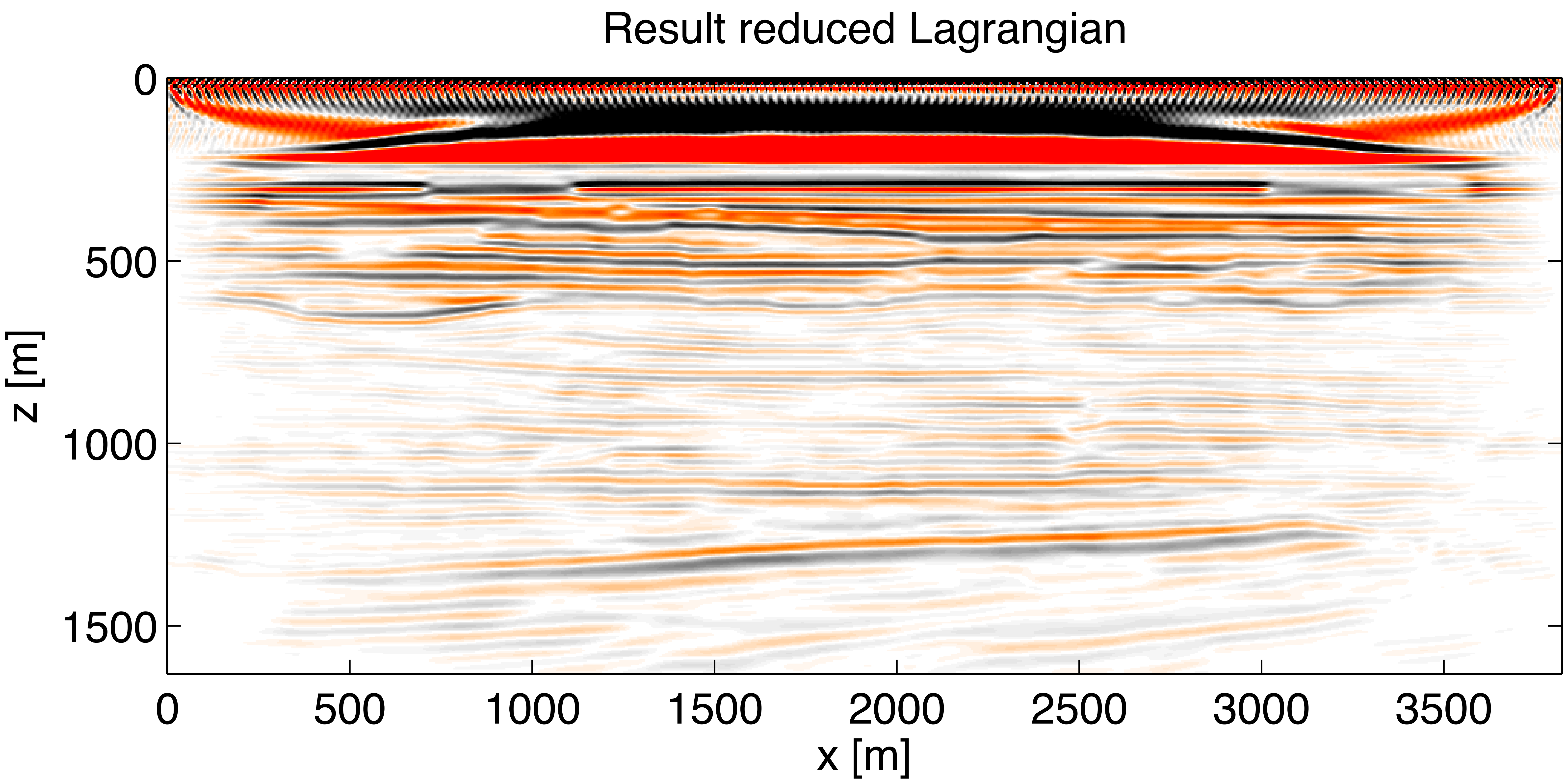
a different image is generated ($\bar{\mathbf{u}}_{kl}$ depends on λ)

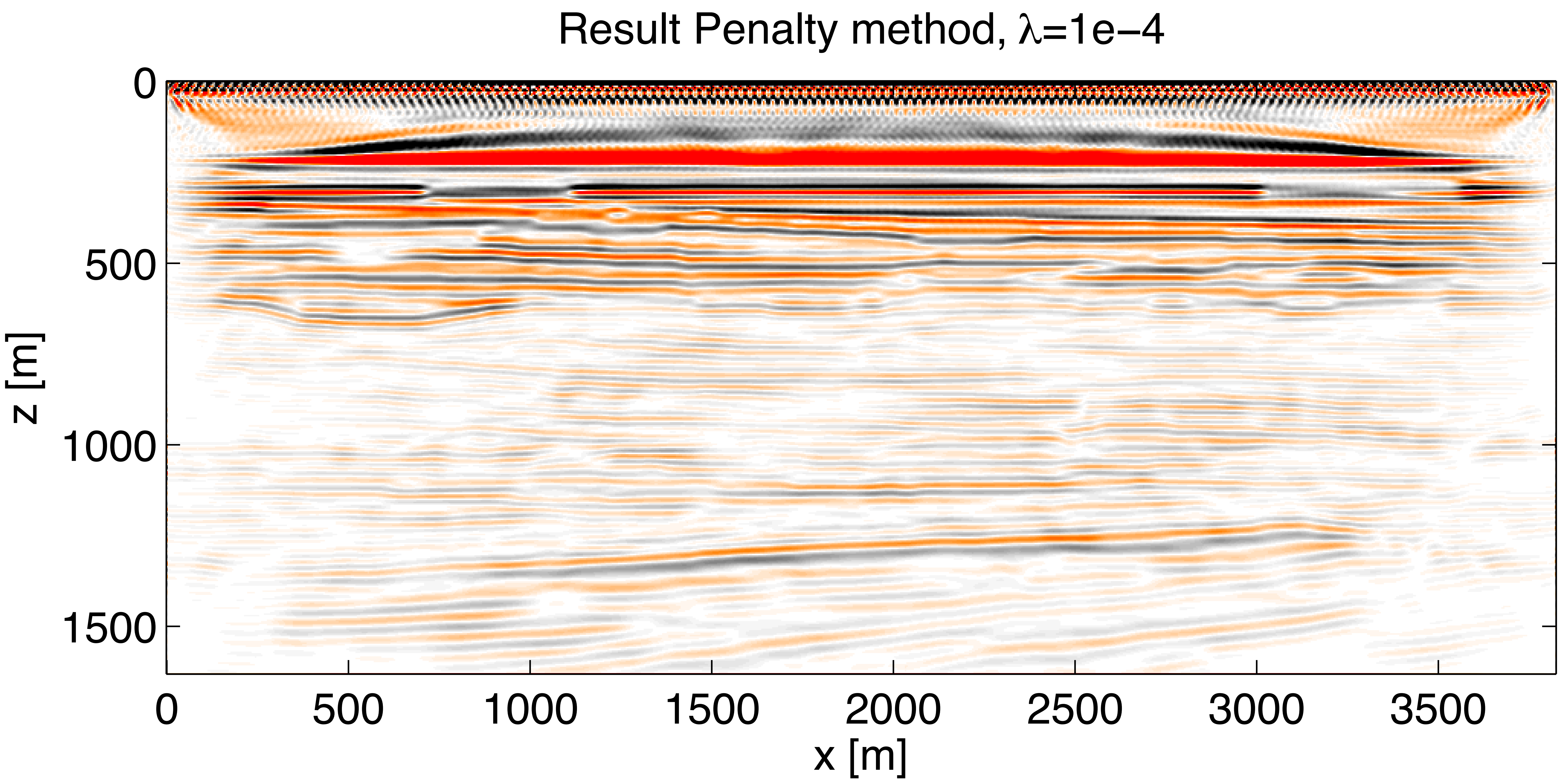
Next few slides show the image generated for various choices of λ(using the same data and background model)

Example 2a: uses a quite accurate background velocity and no noise.

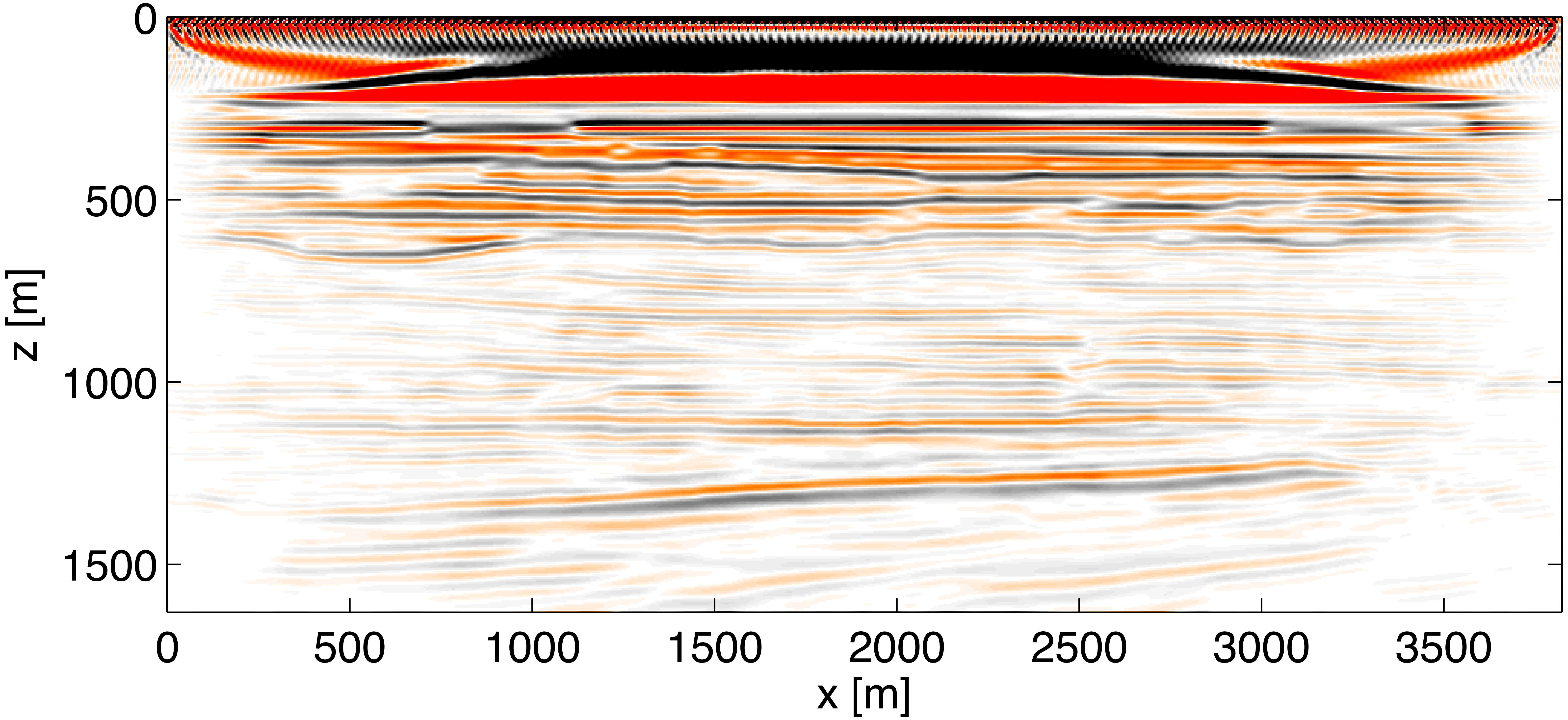
Penalty-method imaging

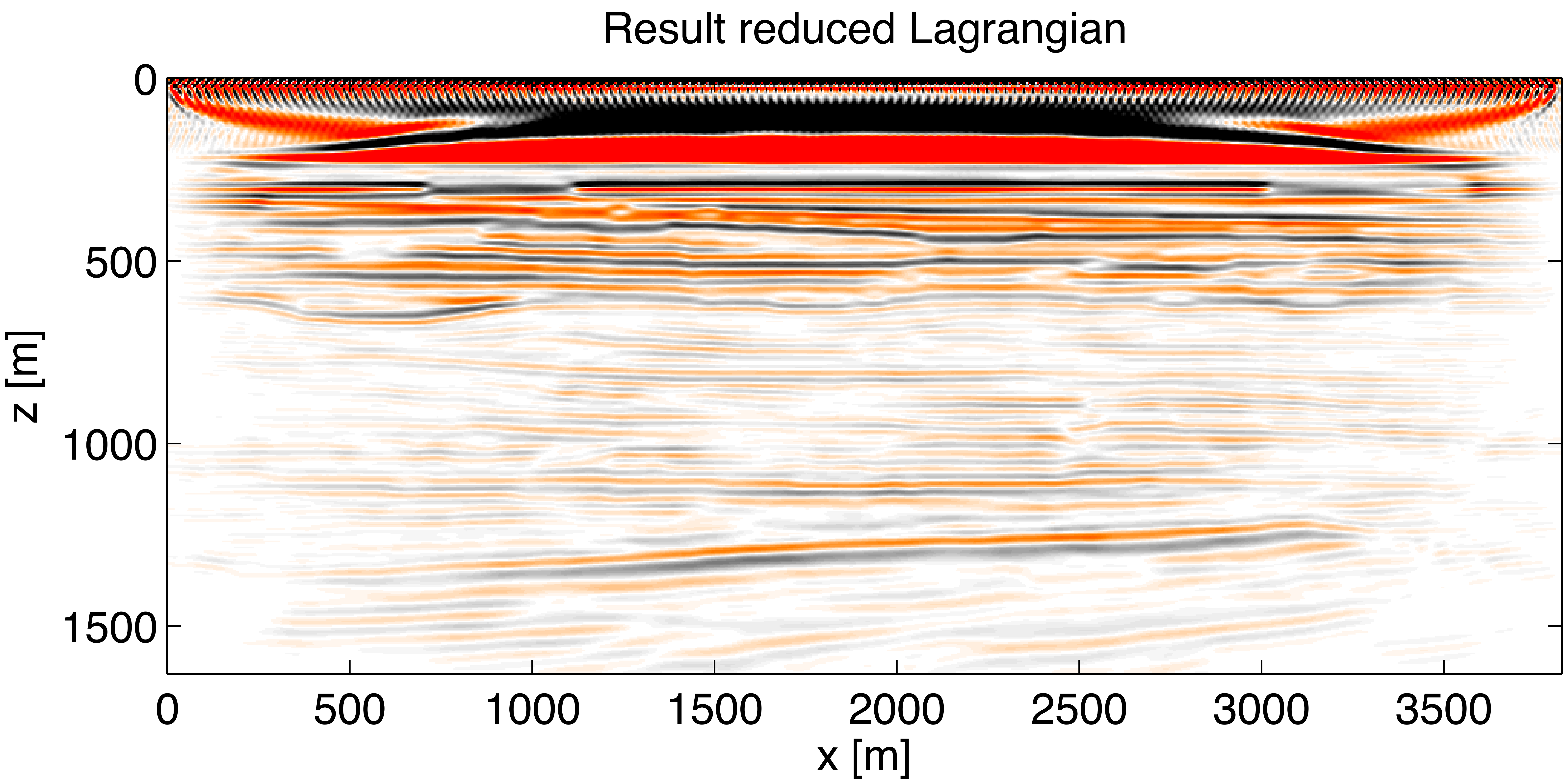



Result reduced Lagrangian
0
500
1000
1500
z [m]
0
500
1000
1500
2000
2500
3000
3500
x [m]


Result Penalty method, λ=1e−4
0
500
1000
1500
z [m]
0
500
1000
1500
2000
2500
3000
3500
x [m]

Result Penalty method, λ=1e4




Result reduced Lagrangian
0
500
1000
1500
z [m]
0
500
1000
1500
2000
2500
3000
3500
x [m]

Penalty-method imaging

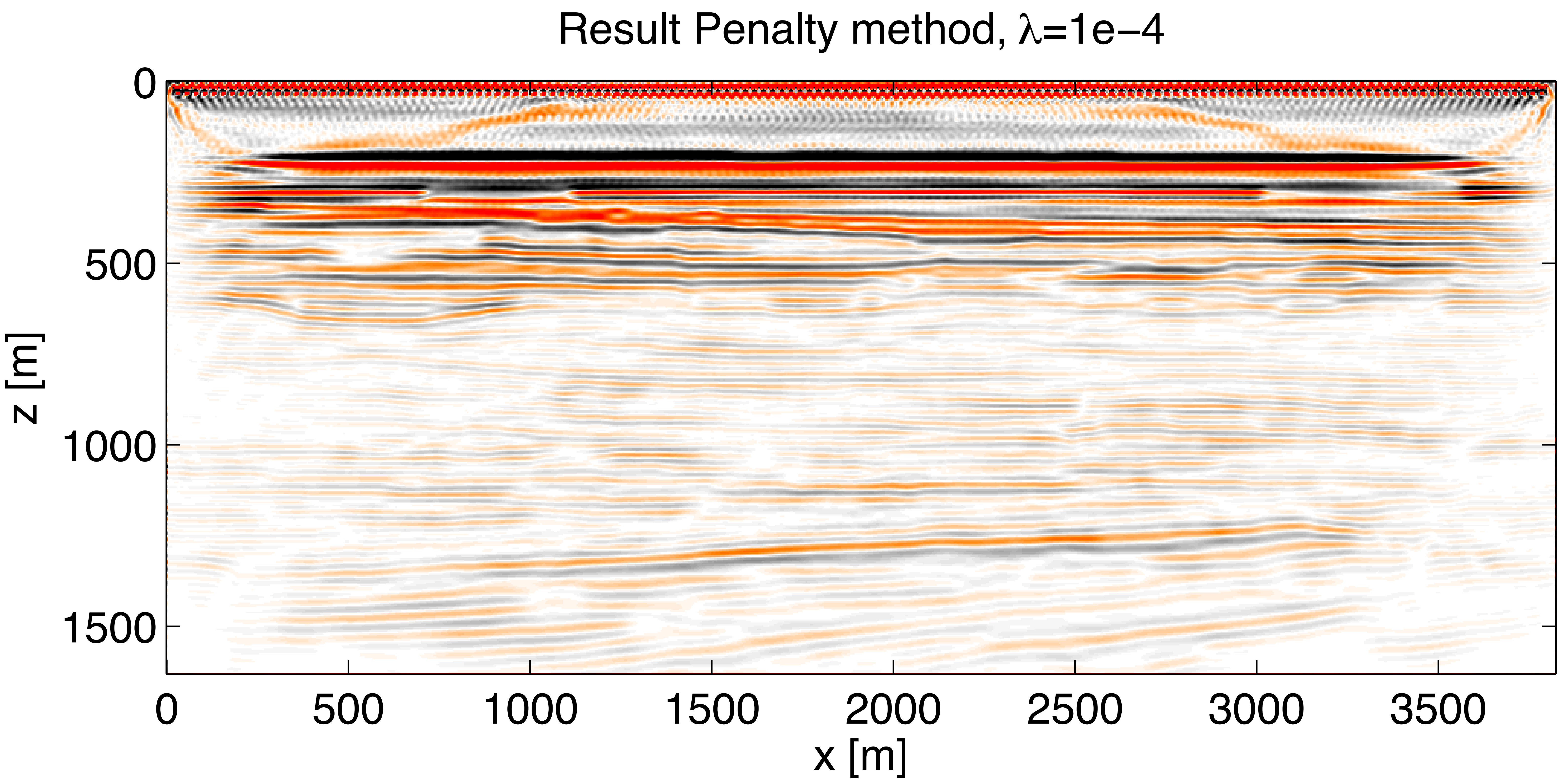
- For sufficiently small λ, the Hessian is diagonal! [van Leeuwen & Herrmann, 2013]

  →Gauss-Newton step at the cost of a gradient computation

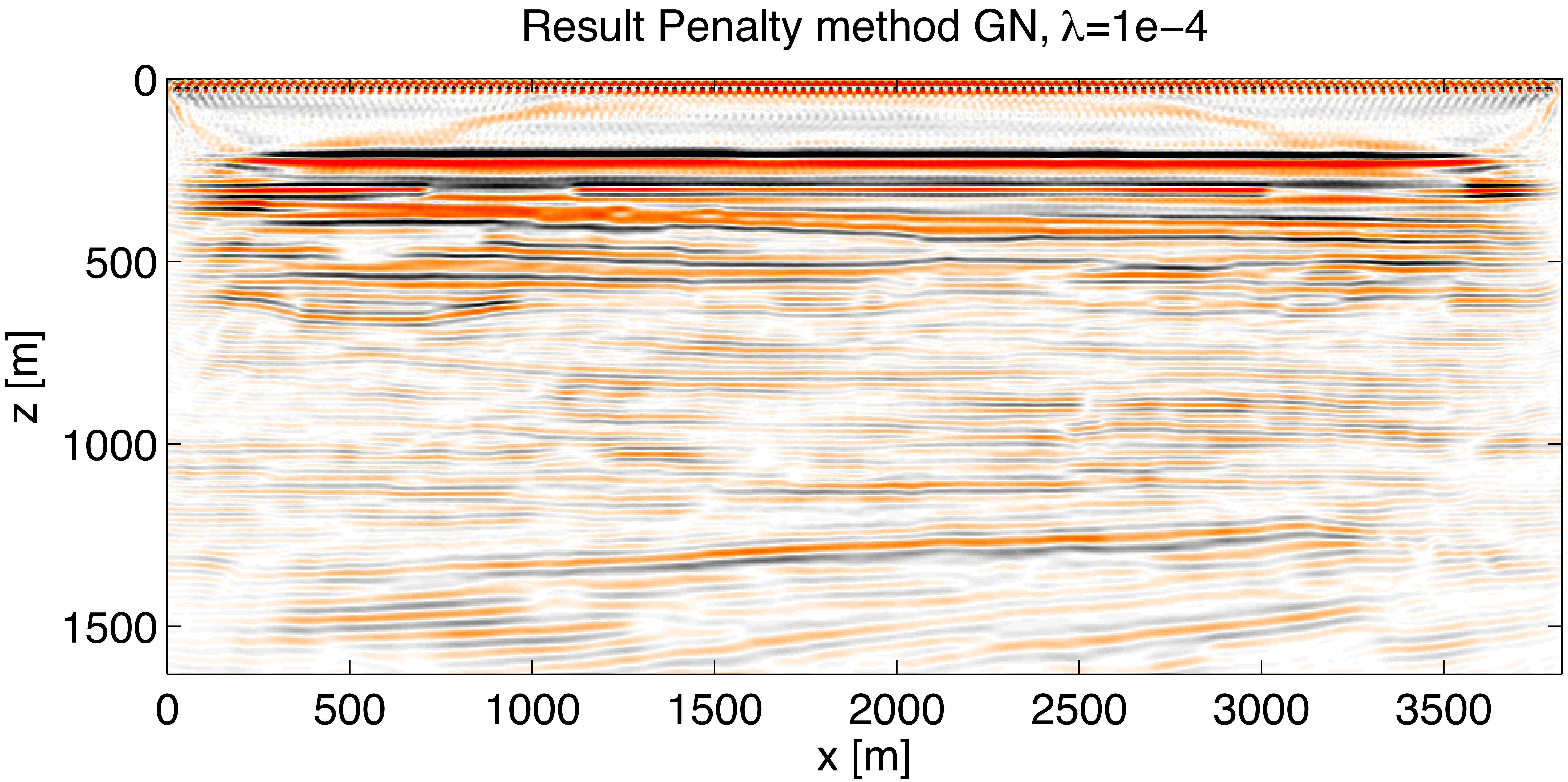
$$H_{\text{pen}} = (\lambda^2 - 1)\sum G^* G = (\lambda^2 - 1)\sum \omega^4 \text{diag}(\mathbf{u})^* \text{diag}(\mathbf{u})$$

- Image is now computed as: $[\nabla^2_{\mathbf{m}} \bar{\phi}_\lambda]^{-1} [\nabla_{\mathbf{m}} \bar{\phi}_\lambda] = H^{-1}_{\text{pen}} \mathbf{g}$

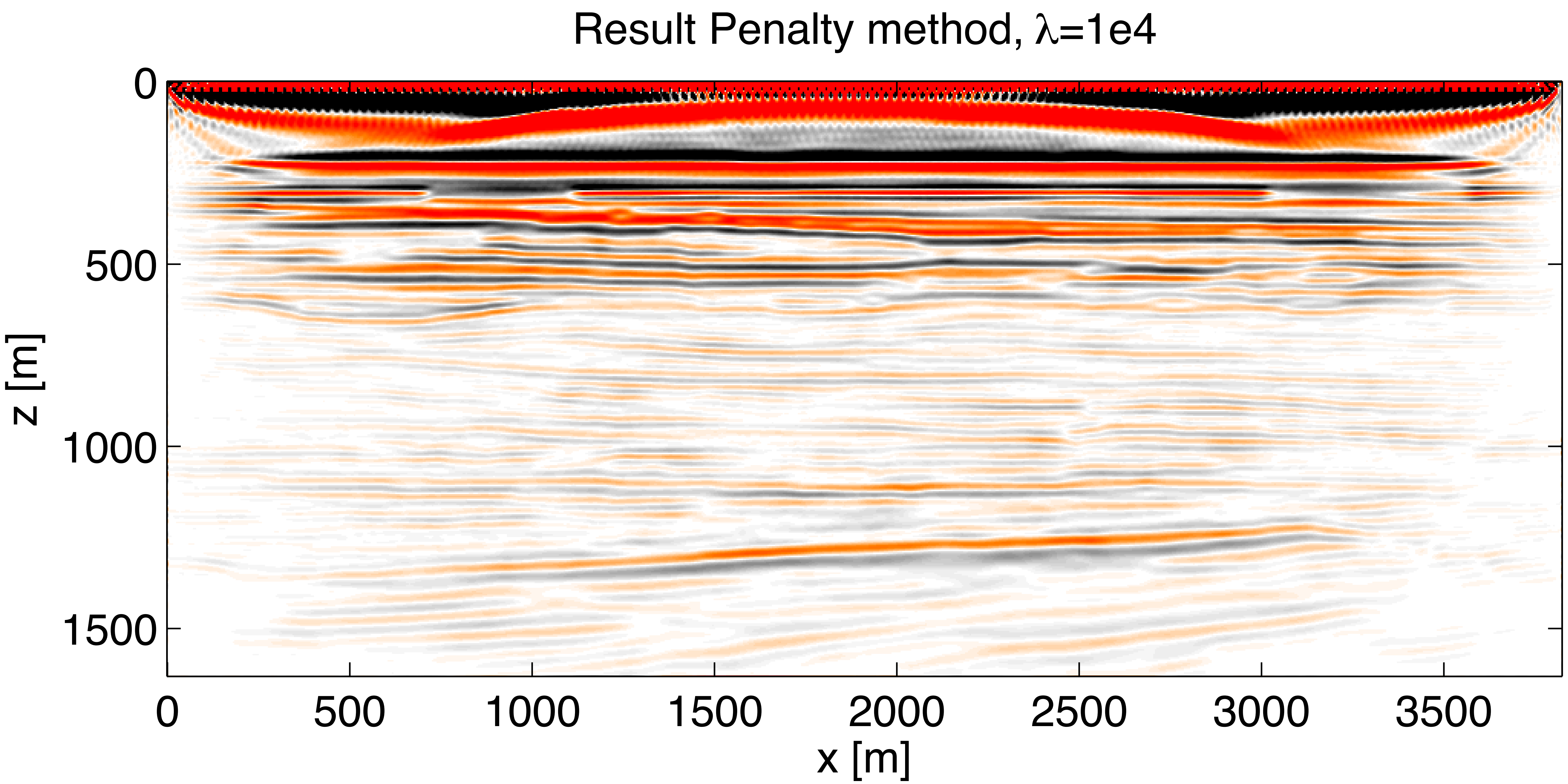
- Next slides: compare Penalty method gradient image vs. Penalty method Gauss-Newton direction



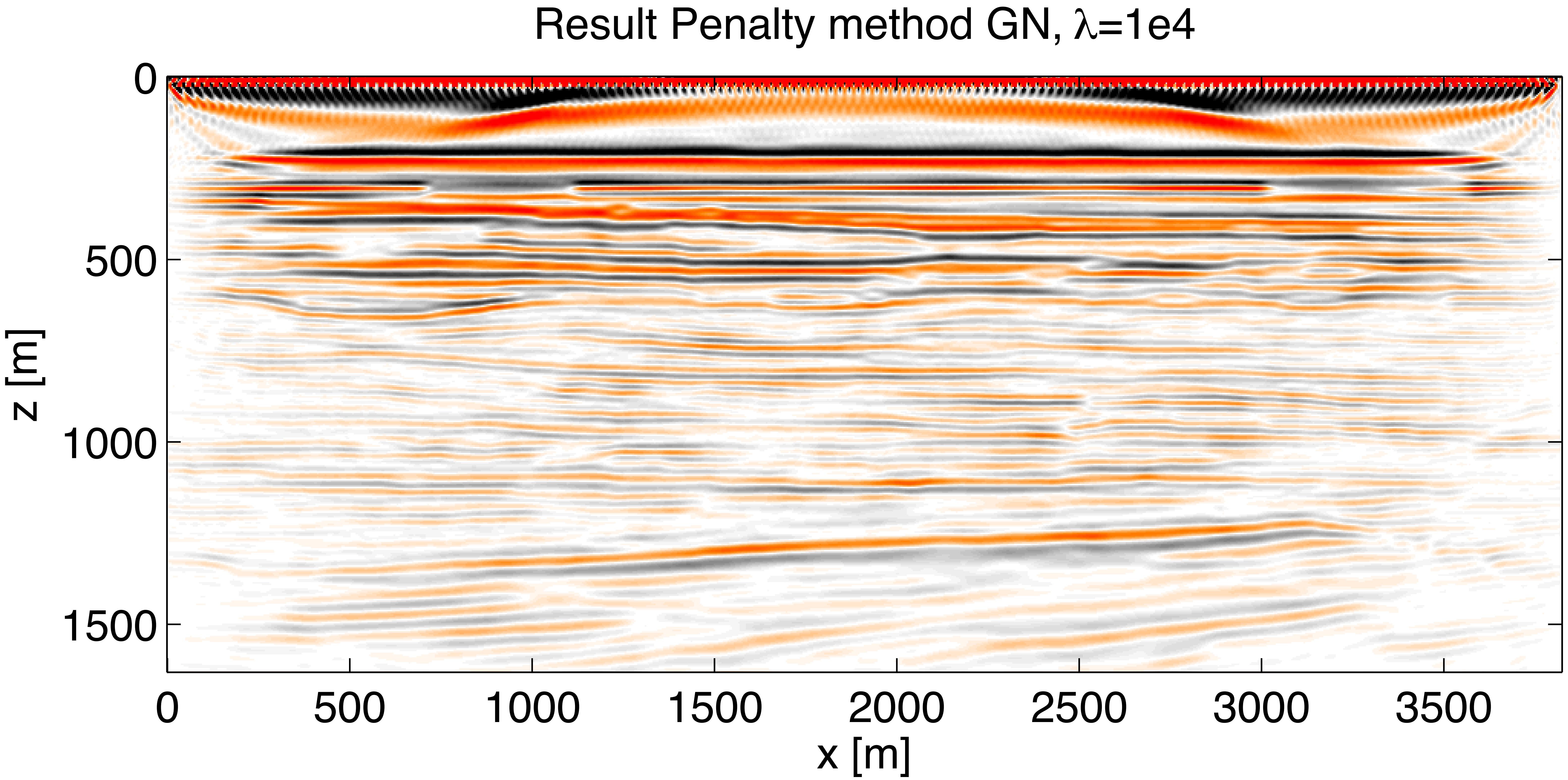
Not so good background model, random noise



Not so good background model, random noise



Not so good background model, random noise



Not so good background model, random noise

Observations about imaging

- The tradeoff parameter λ does not significantly influence the resulting imaging obtained using the Penalty method.

- In the examples, the Penalty method and reduced Lagrangian methods result in similar images.
- Penalty-method Gauss-Newton step may improve the quality further (at no extra cost)

- The Penalty method is about 3 times faster (400x900 grid) than the adjoint-state method. This is the result of 1 least squares problem (SuiteSparseQR) vs 2 linear system solves (UMFpack). Both as implemented in Matlab '\'

Conclusions

Penalty method vs. the Reduced Lagrangian method

- Penalty method:
    - Much better waveform inversion results for some difficult problems
    - Less sensitive to missing low frequencies
    - A bit better waveform inversion results when both converge
    - Results in similar or better quality images
    - Not more sensitive to noise

    - Solution via direct solvers can be more than a factor 2 faster (depending on problem size).

Outlook

- Find most efficient ways to solve the least-squares problem (also using iterative methods), see talk on Wednesday
- Find 'optimal' combination of tradeoff parameter λ and inversion set up. (Penalty method results in this talk may not be the best possible)
- Show how to work with simultaneous sources and limited source-receiver offset

Acknowledgements

The SLIM students & postdocs

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, Total SA, WesternGeco, and Woodside.