Recent Developments in Preconditioning the FWI Hessian - A Dimensionality Reduction Approach

Curt Da Silva and Felix J. Hermann

SLIM
University of British Columbia

Preconditioning the FWI Hessian

- Work done by L. Demanet et. al - Matrix probing: a randomized preconditioner for the wave-equation Hessian

Motivation

- Traditional least-squares migration is an expensive process
  - Numerous applications of the linearized Born scattering operator and its adjoint

Motivation

- Traditional least-squares migration is an expensive process
  - Each application of the linearized scattering operator requires computation with full data, which is prohibitive cost-wise

Motivation

- We want to both reduce the number of iterations required for our linear solver as well as reducing the amount of data we need to process at each iteration

Inversion

- The problem we effectively try to solve in LS migration is given by

$$\min_{\delta m} \|H[m_0;Q]\delta m - J^*\delta D\|_2^2$$

J - Linearized Born Scattering Operator

$H = J^*J$ - (Gauss-Newton) Hessian

m_0 - (smooth) background model

Q - source matrix

Hessian

- Under certain conditions, the Hessian is known to behave like a *pseudo-differential operator*
- As does its inverse, assuming it exists

Pseudo-differential Operators

- A generalization of differential operators
- Differential operators are polynomials in the Fourier domain (filters with no spatial dependence)

Pseudo-differential Operators

- A generalization of differential operators
- Pseudo-differential operators are polynomial-like in the Fourier domain (act as spatially dependent filters)

Pseudo-differential Operators

- Their particular Fourier-domain behaviour makes them amenable to computation
- Can be stored in an extremely compressed form, efficiently applied to functions

SLIM

Hessian

$$Hf(x) \approx \int a(x,k)\hat{f}(k)e^{2\pi i k\cdot x}dk$$

- $a(x,k)$ is the *symbol* of the pseudo-differential operator

Inverse Hessian

$$H^{-1}f(x) \approx \int b(x,k)\hat{f}(k)e^{2\pi i k\cdot x}dk$$

- Knowledge of $b(x,k)$ gives us knowledge of H^{-1} applied to any vector (which is what we aim for in a preconditioner)

Matrix Probing

- A significant amount of information about the effective range of a matrix can be gleaned simply from observing its action on a set of random vectors

Matrix Probing

- If we can write

$$H^{-1} \approx \sum_i c_i B_i$$

for known operators B_i,
unknown coefficients c_i

SLIM

Matrix Probing

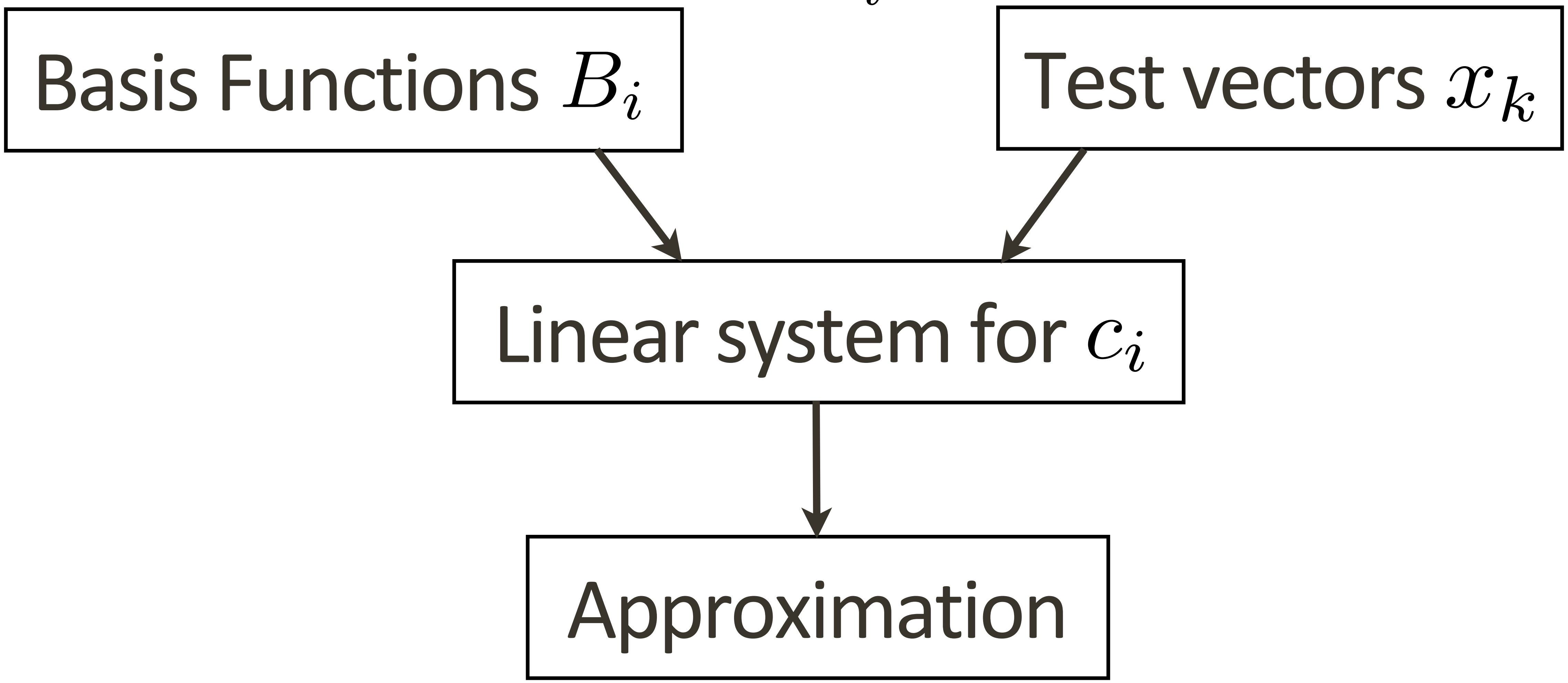
- We can recover the (approximate) action of H^{-1} by forming the linear system

$$H^{-1}x_k = \sum_i c_i B_i x_k$$

(x_k are random test vectors) and solving for c_i

Matrix Probing

$$H^{-1} x_k = \sum_i c_i B_i x_k$$



Basis for PDOs

- Using the basis of operators

$$B_{\lambda,q_1,q_2} = \mathrm{diag}(e^{2\pi i x\cdot\lambda})F^*\mathrm{diag}(e^{iq_1\theta}TL_{q_2}(|k|)|k|)F$$

  results in a fast converging expansion for a large class of PDOs
- F is the Fourier transform

Matrix Probing

$$Hx_k = y_k$$

- x_k are test vectors

SLIM

Matrix Probing

$$H_{approx}^{-1} y_k = x_k$$

- Knowledge of the action of the inverse Hessian can be obtained from observing the behaviour of the forward Hessian

Coefficient Recovery

$$H^{-1} \approx \sum_{(\lambda,q_1,q_2)} c_{\lambda,q_1,q_2} B_{\lambda,q_1,q_2}$$

- B_{λ,q_1,q_2} are known, c_{λ,q_1,q_2} are unknown but theoretically quickly decreasing

SLIM

Summary

- We know from the theory that

$$H^{-1} \approx \sum_{(\lambda, q_1, q_2)} c_{\lambda, q_1, q_2} B_{\lambda, q_1, q_2}$$

- We impose this relationship on a small set of reference vectors in order to solve for c_{λ, q_1, q_2}

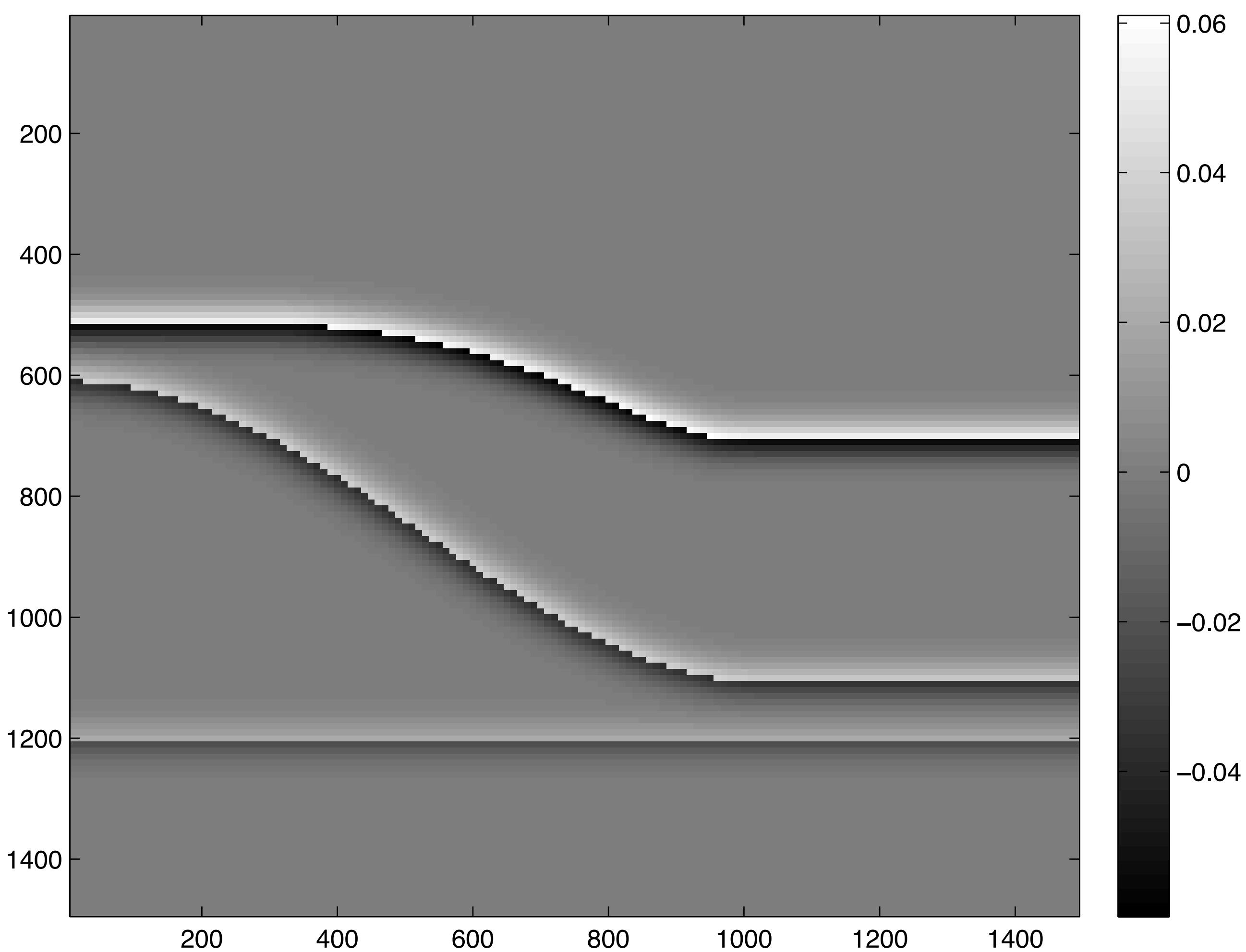
Random Test Vectors

- Various choices of test vectors
- Gaussian noise - too much energy in the nullspace of H
- Gaussian noise to which the Hessian is applied - avoids the nullspace, but expensive to compute (and is apparently terrible)

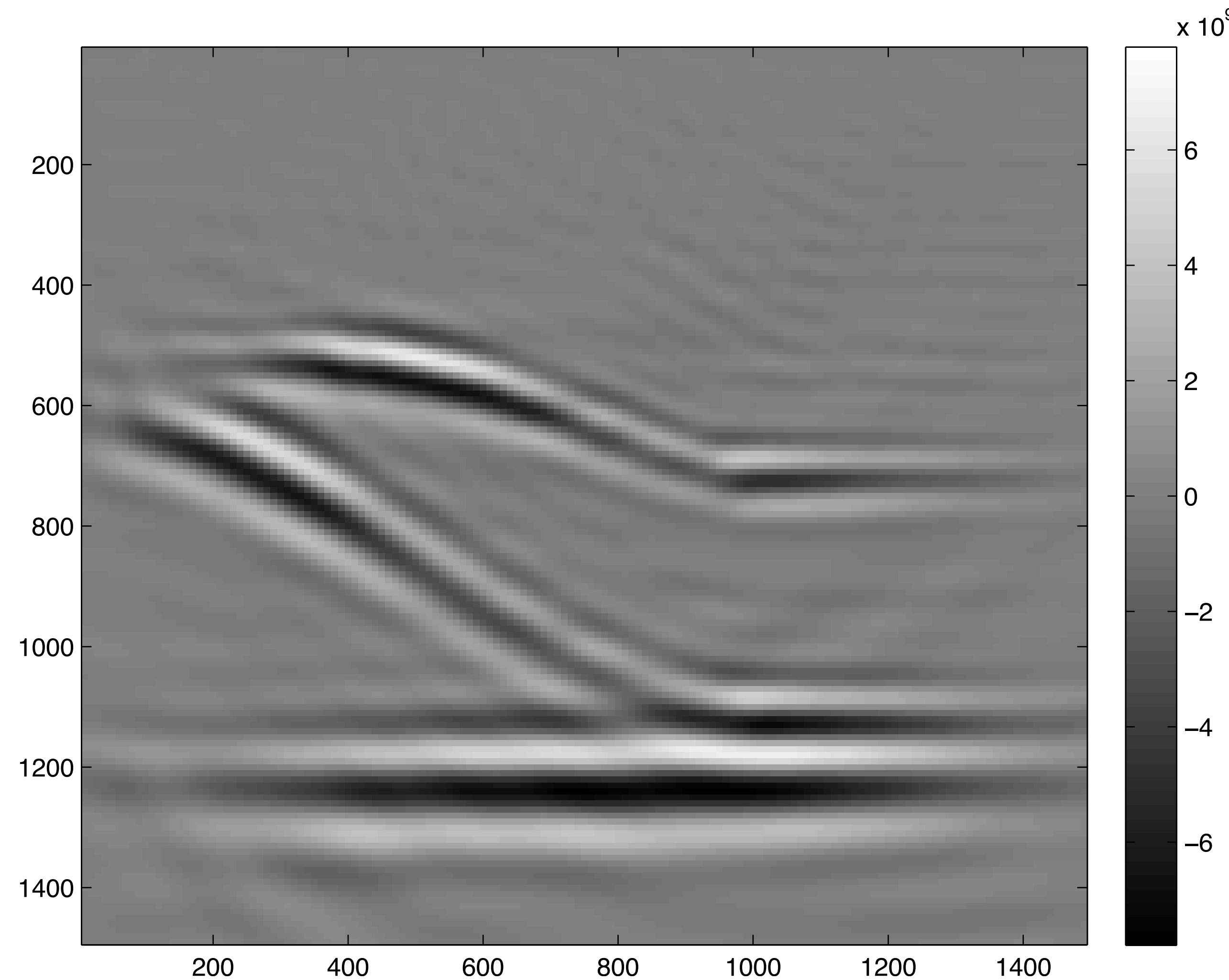
Random Test Vectors

- Various choices of test vectors
- Noise in the Curvelet domain taken with a 1-0 mask
- Curvelets discriminate space-frequency regions where H is effectively zero
- More details in *Matrix Probing*

Synthetic Clay Example



Model Perturbation

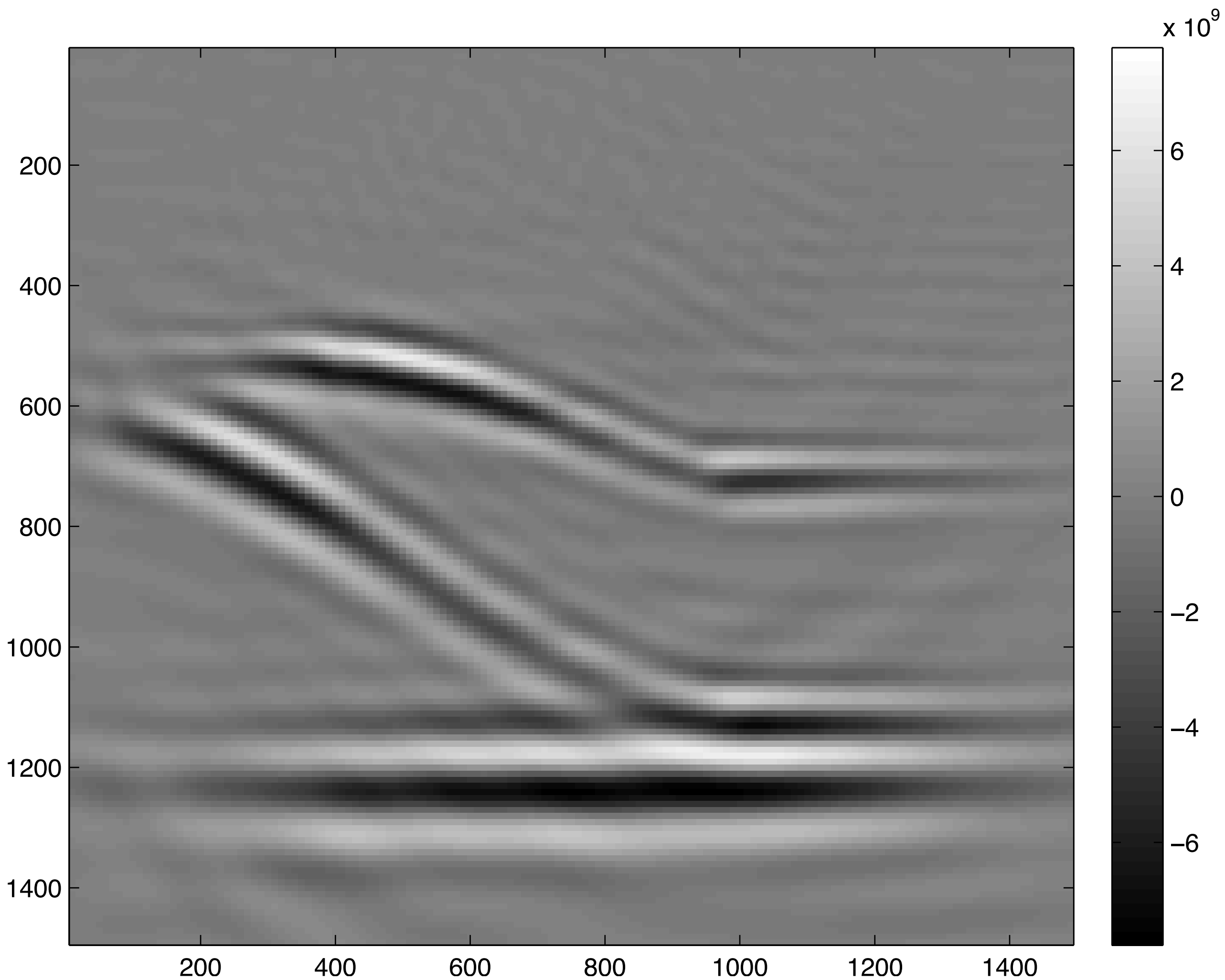


Hessian - Full Sources

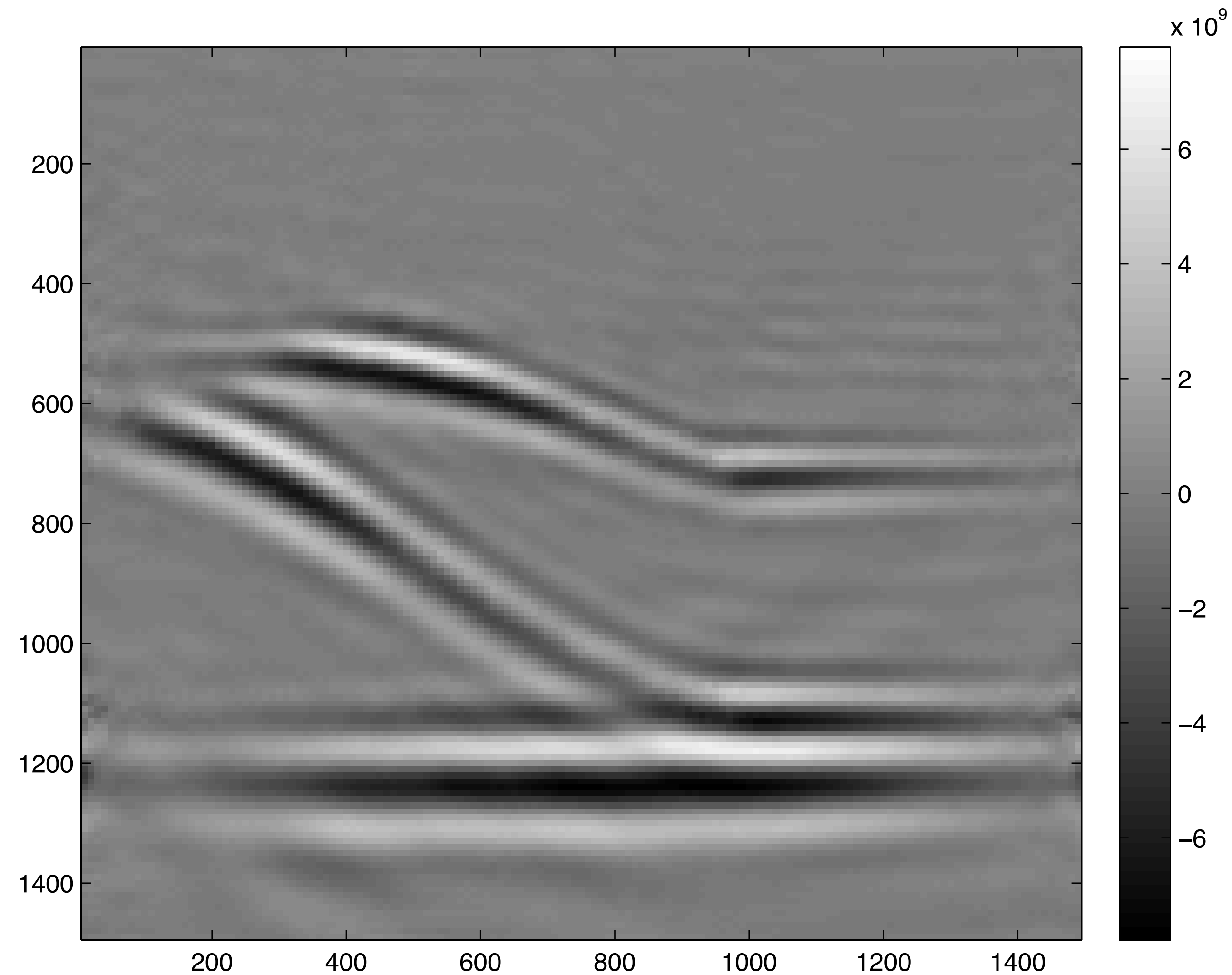
SLIM

Synthetic Clay Example

Using the true perturbation as a test function



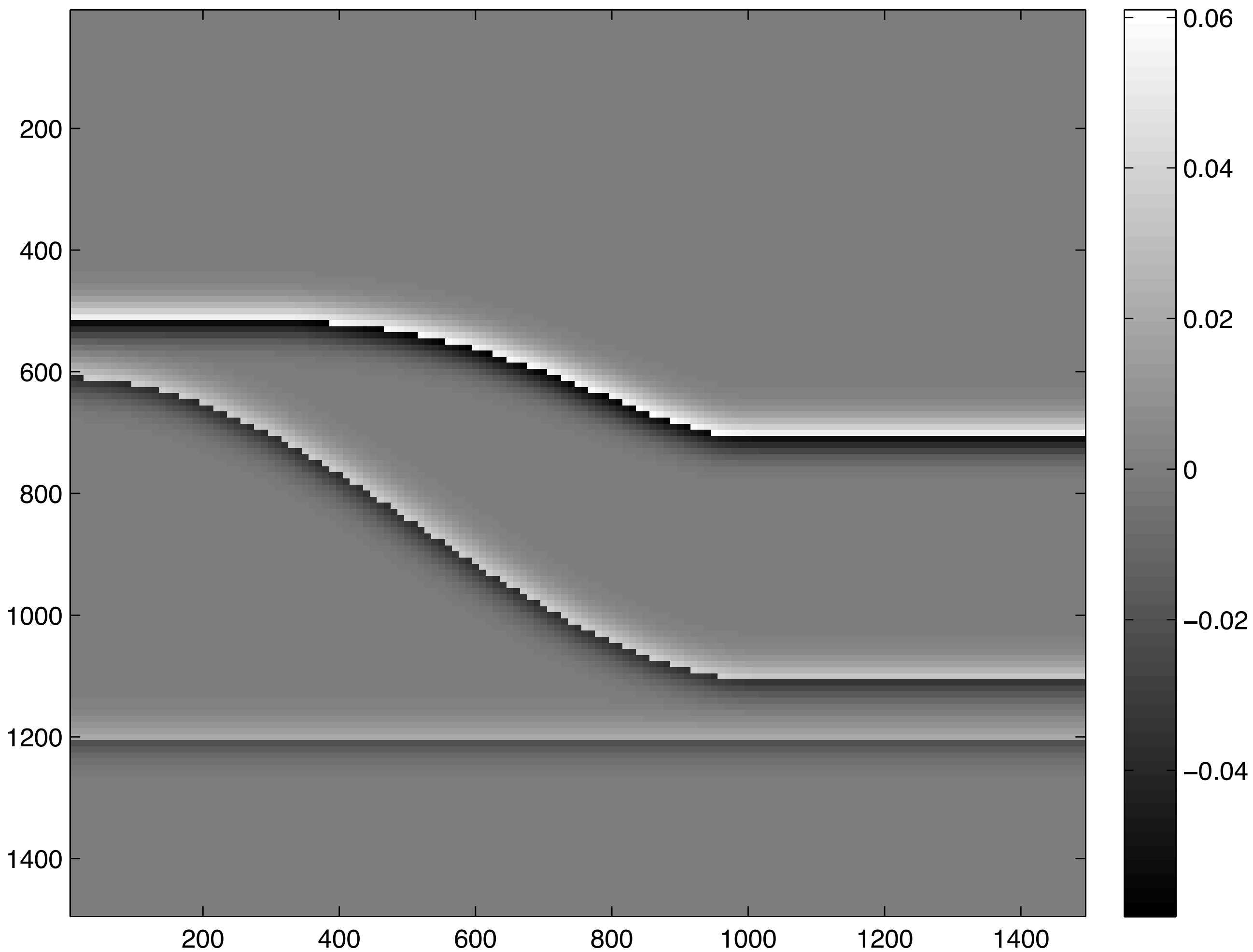
True Hessian



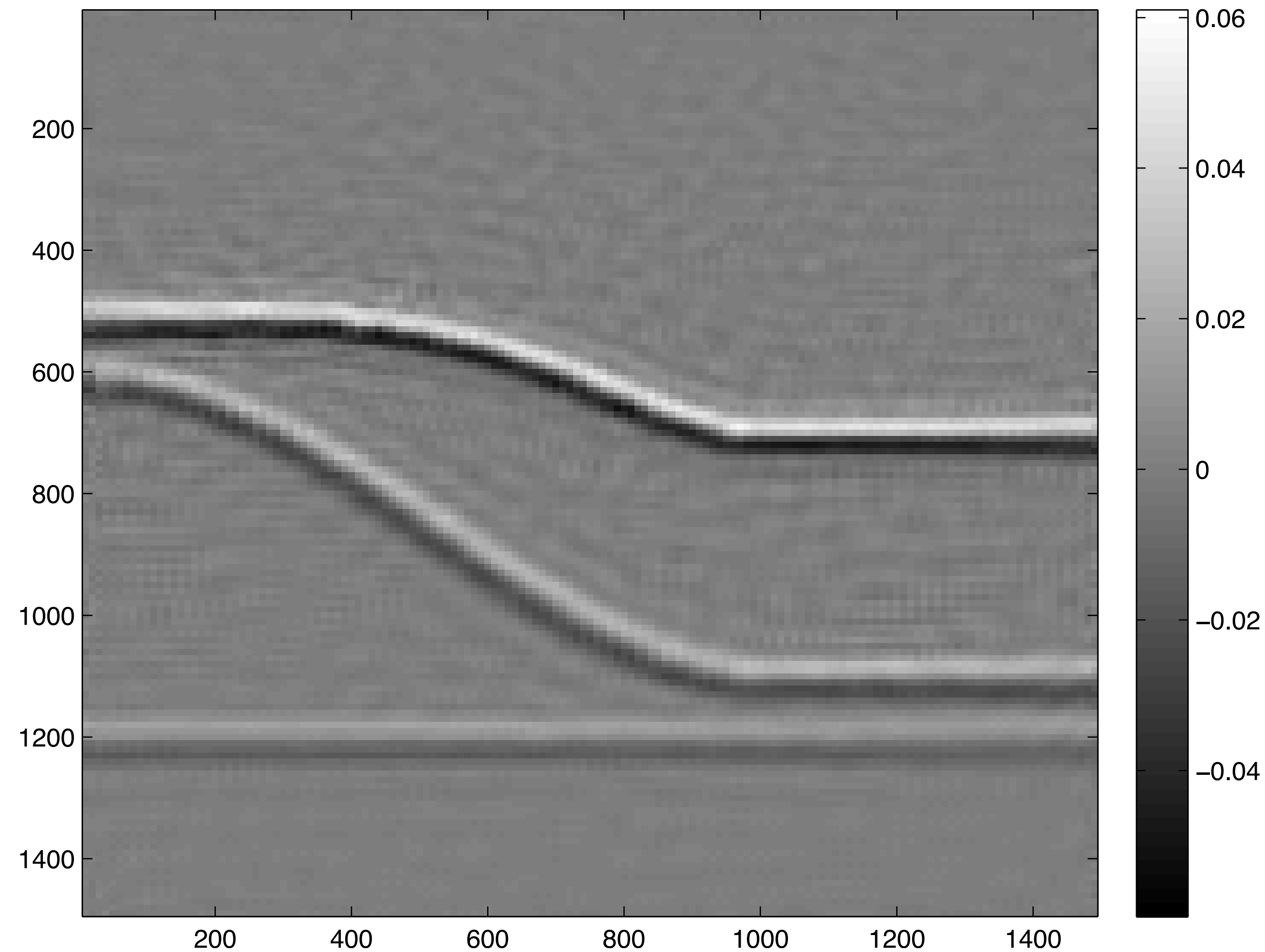
PDO approximation

Synthetic Clay Example

Using the true perturbation as a test function



Model Perturbation



PDO approximation of the inverse Hessian

Simultaneous Sources

- Application of the Hessian with the full number of sources is extremely costly
  - Each application involves a number of PDE solves proportional to the number of sources

Simultaneous Sources

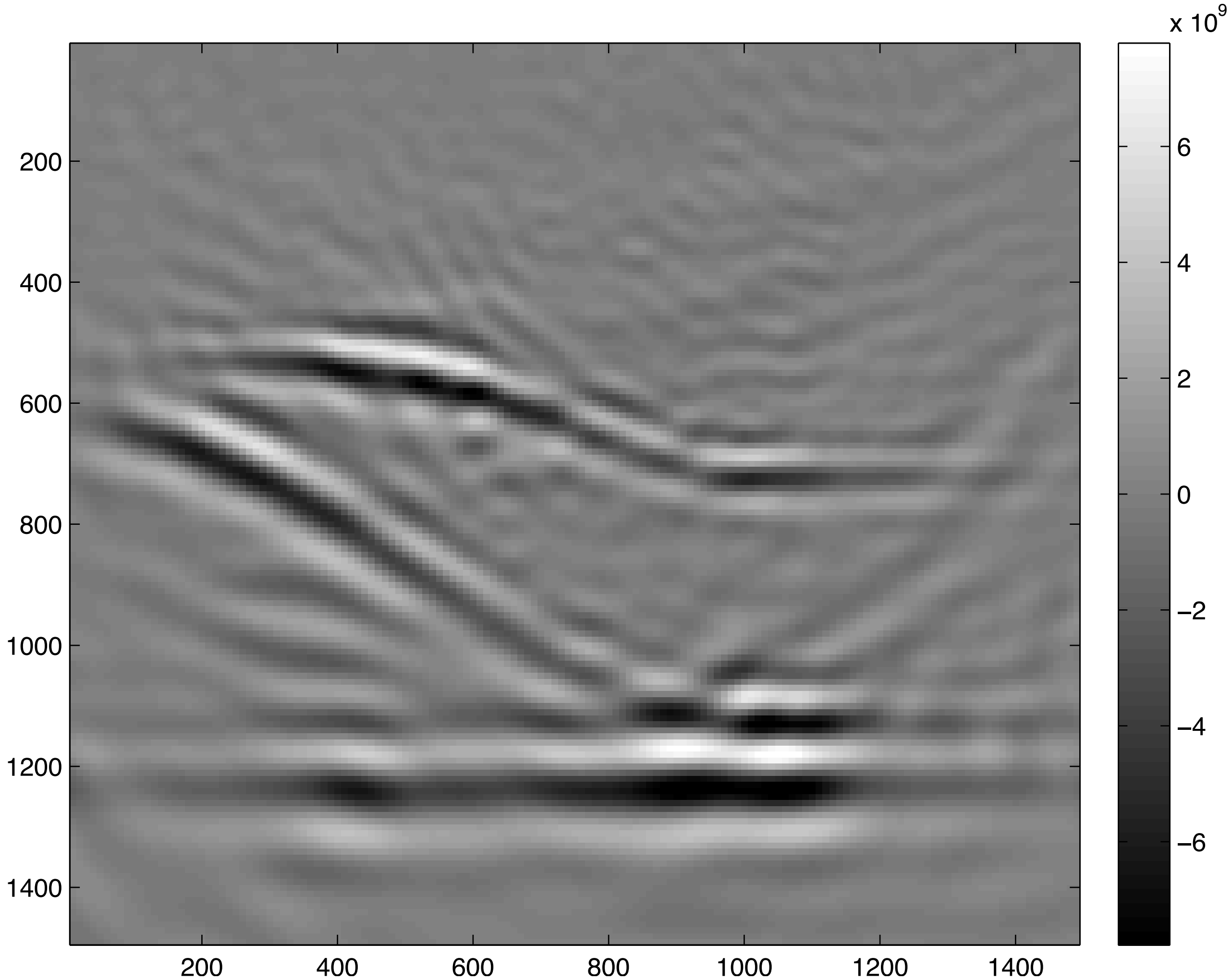
- We apply the Hessian using simultaneous sources, which are randomly weighted superpositions of the original source functions
  - We achieve a Hessian speedup proportional to the ratio of simultaneous to full sources

Simultaneous Sources

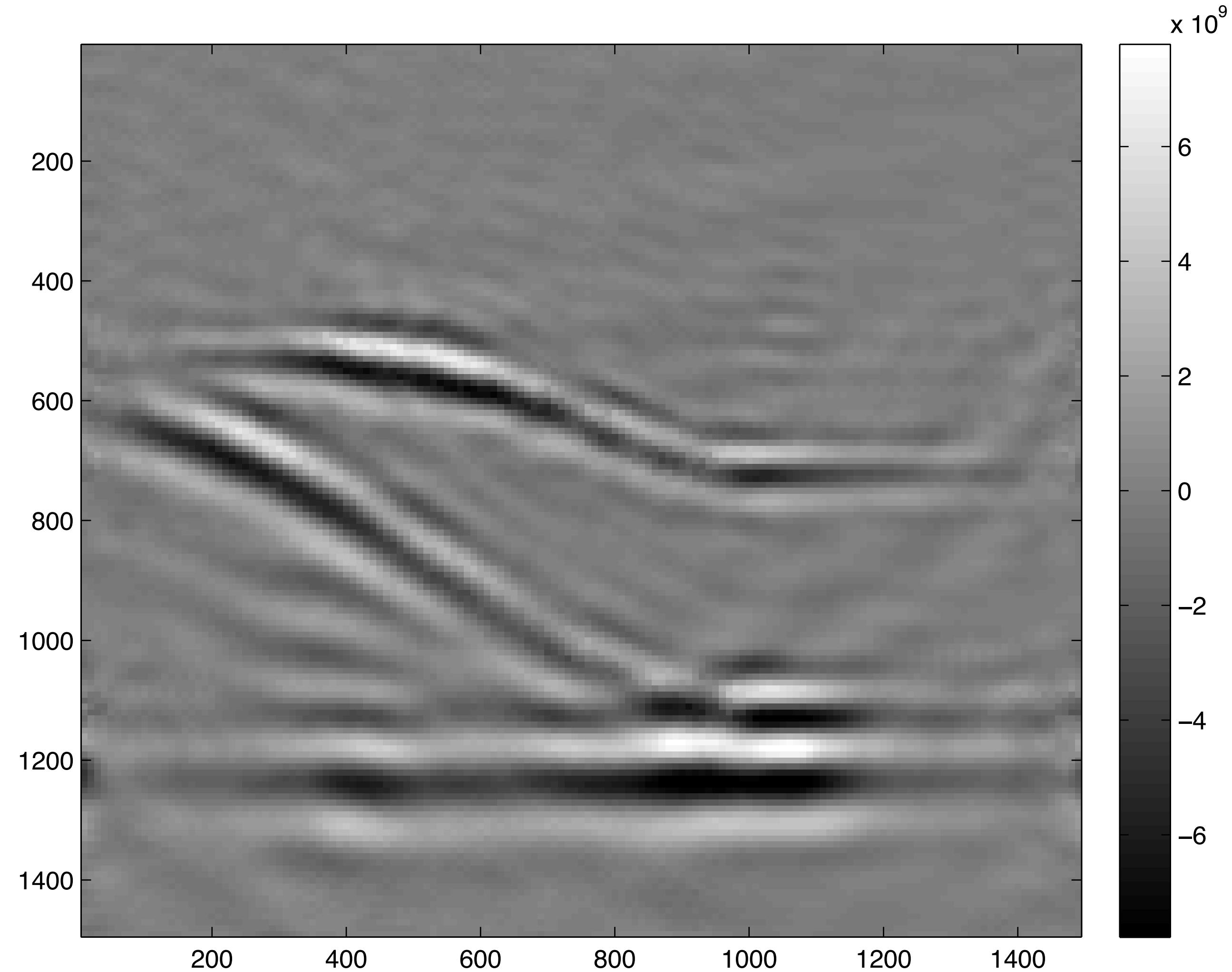
- We can still apply the aforementioned preconditioner in this context to invert for the model perturbation
  - The Hessian behaves roughly as a pseudo-differential operator, but now with a noisy symbol

Synthetic Clay Example

Using the true perturbation as a test function



True Hessian - Simultaneous Sources

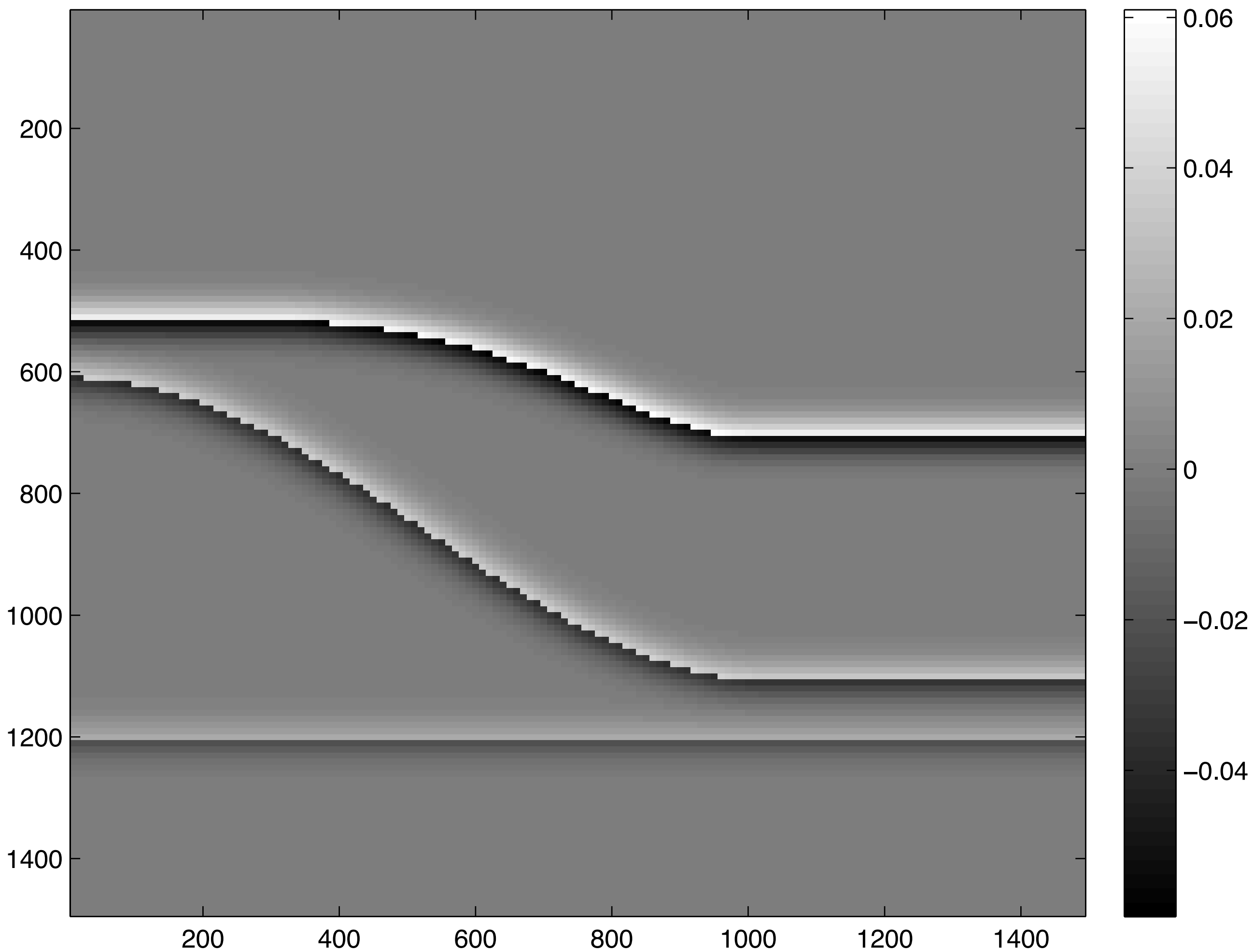


PDO Approximation

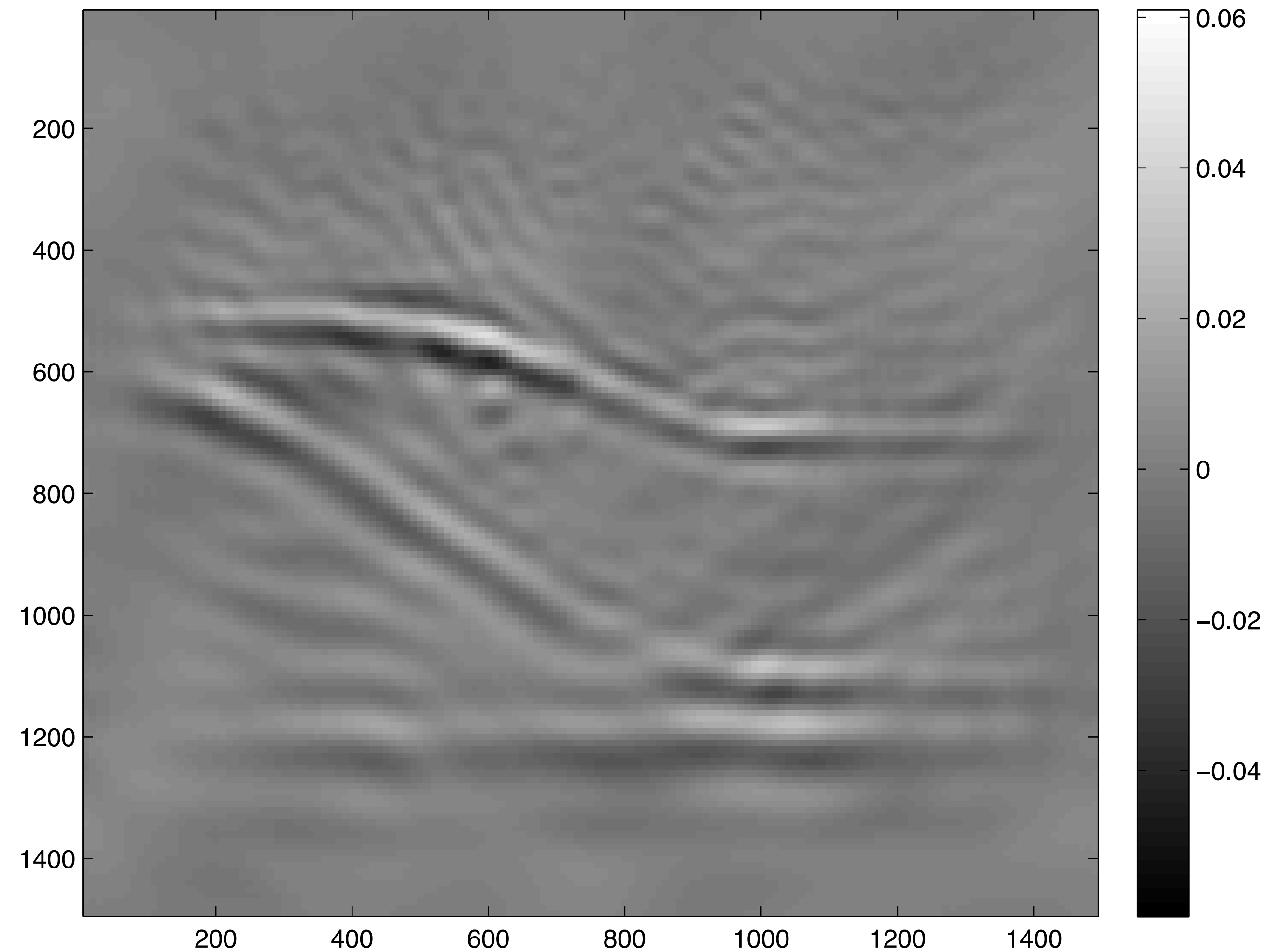
Realistic Approach

- As you can see from the inverse Hessian approximation, this "inverse crime" approach is useful as an illustration that these pseudo-differential relationships actually exist
- Not particularly useful in practice

Preconditioned Image

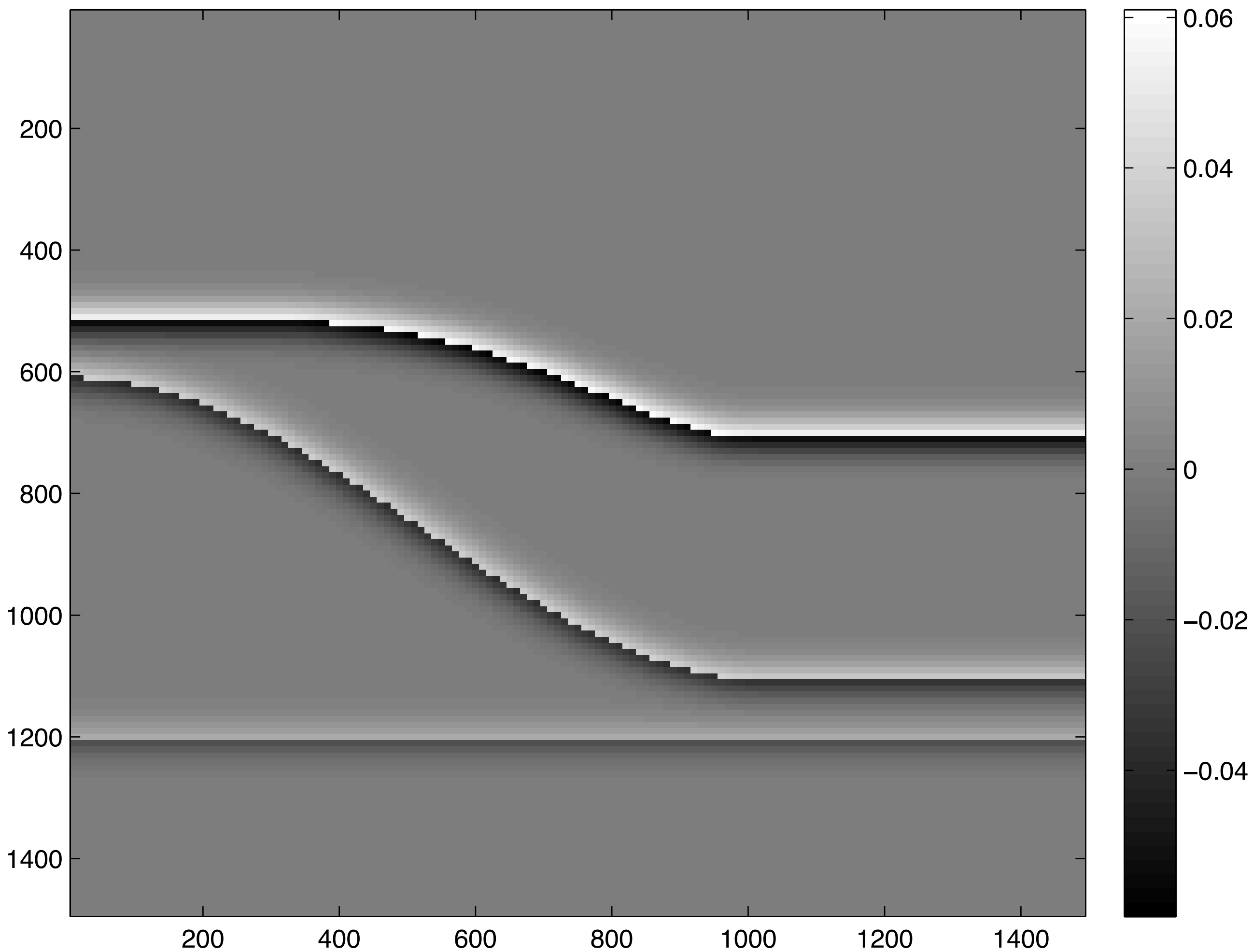


Model Perturbation

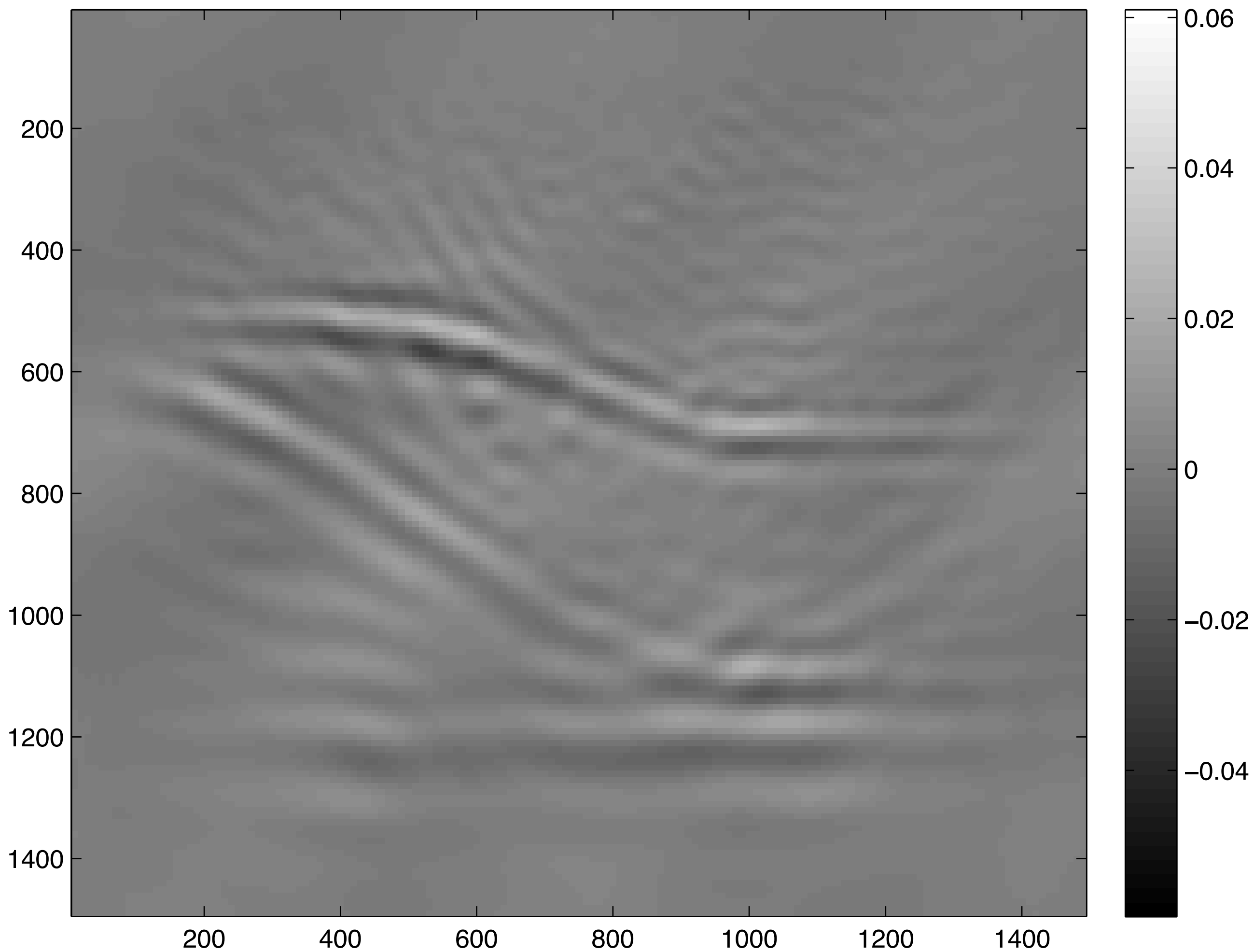


PDO approximation of the inverse Hessian - noisy test functions

Preconditioned Image

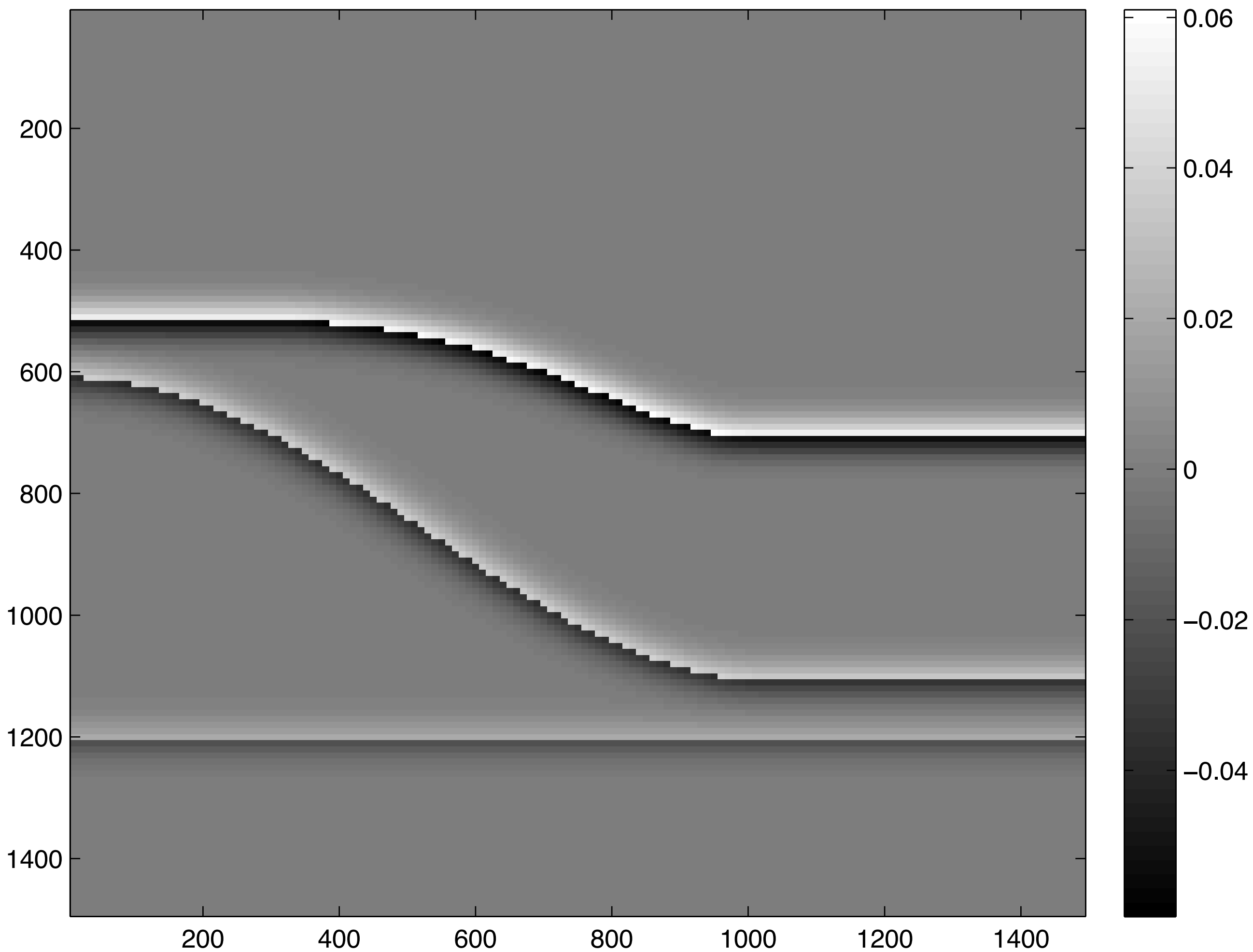


Model Perturbation

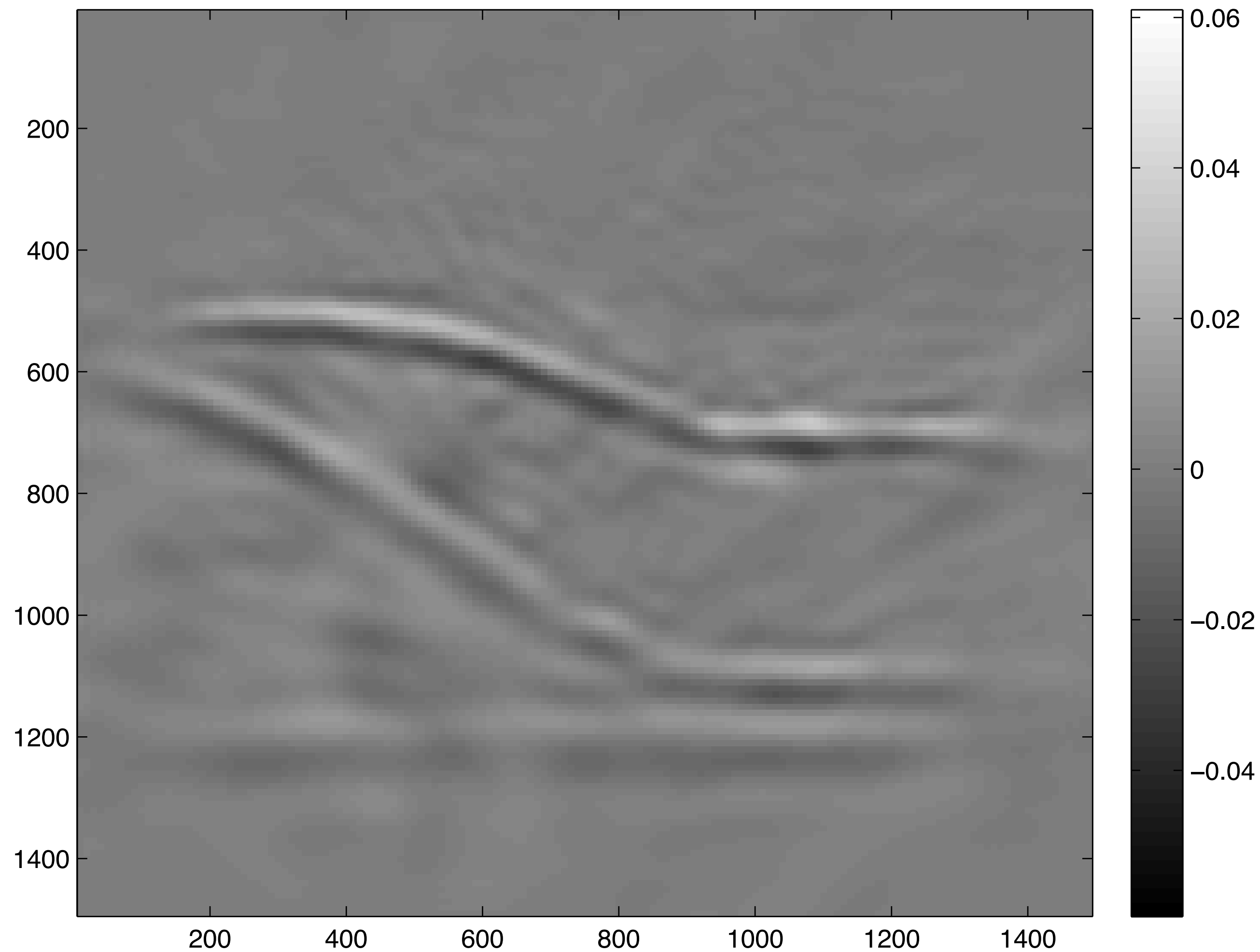


PDO approximation of the inverse Hessian - noisy curvelet test functions

Preconditioned Image



Model Perturbation



PDO approximation of the inverse Hessian - scaled image test function

SLIM

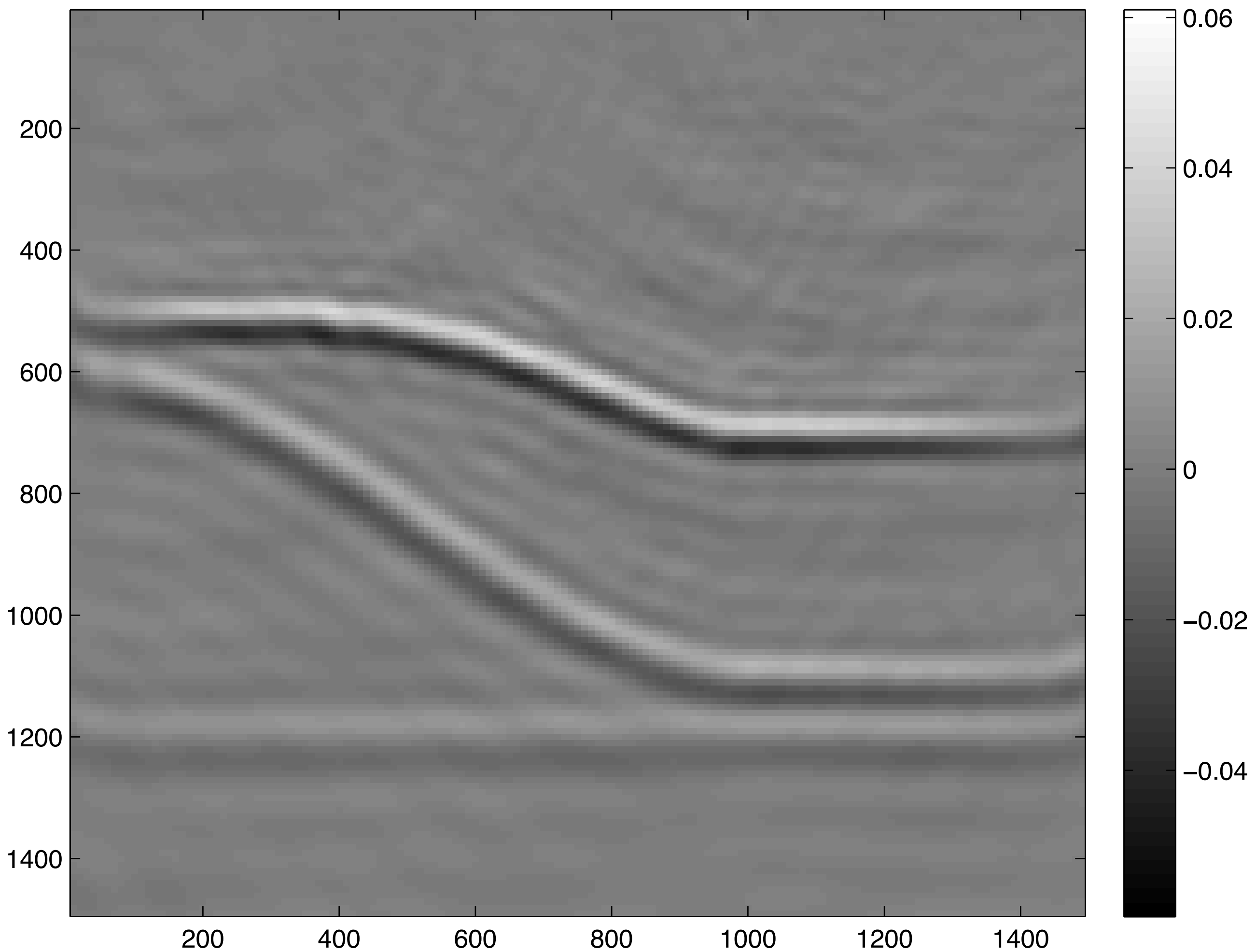
Challenges

- The resulting operator we obtain isn’t symmetric positive semidefinite (unlike the Hessian itself) and the preconditioned system causes CG to terminate early

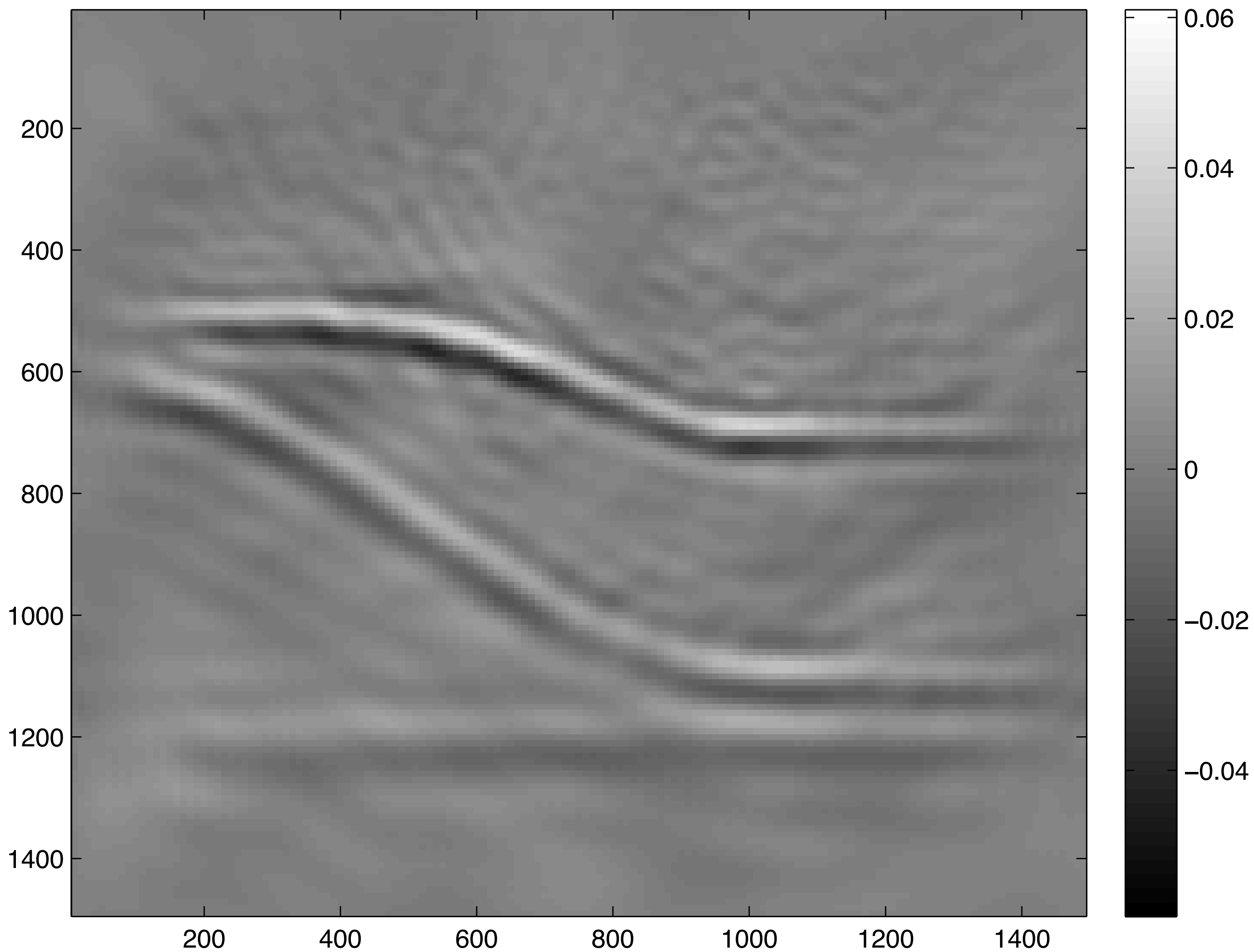
Challenges

- We can still investigate the result of PCG after its early termination

Inversion Results

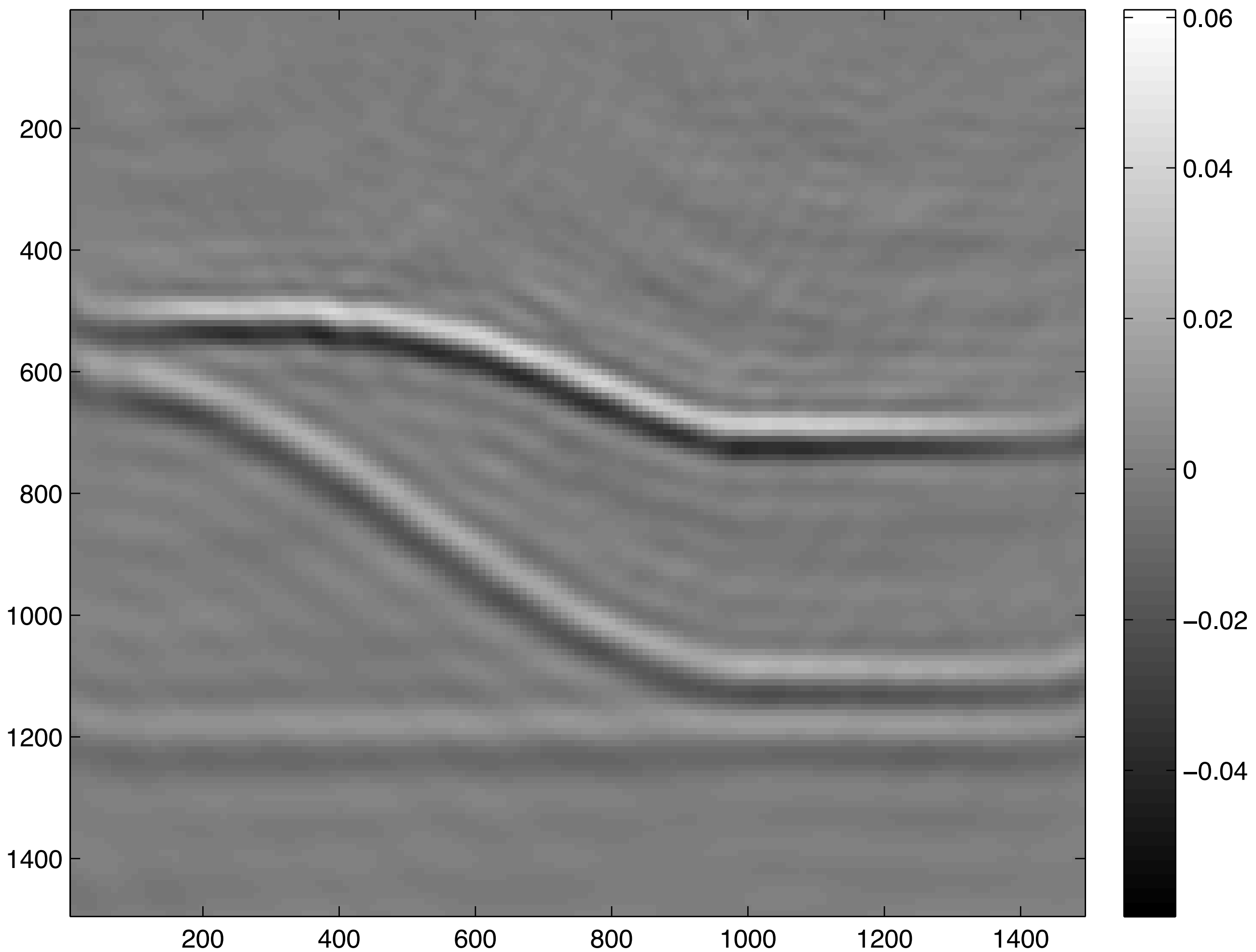


LSQR - 15 iterations

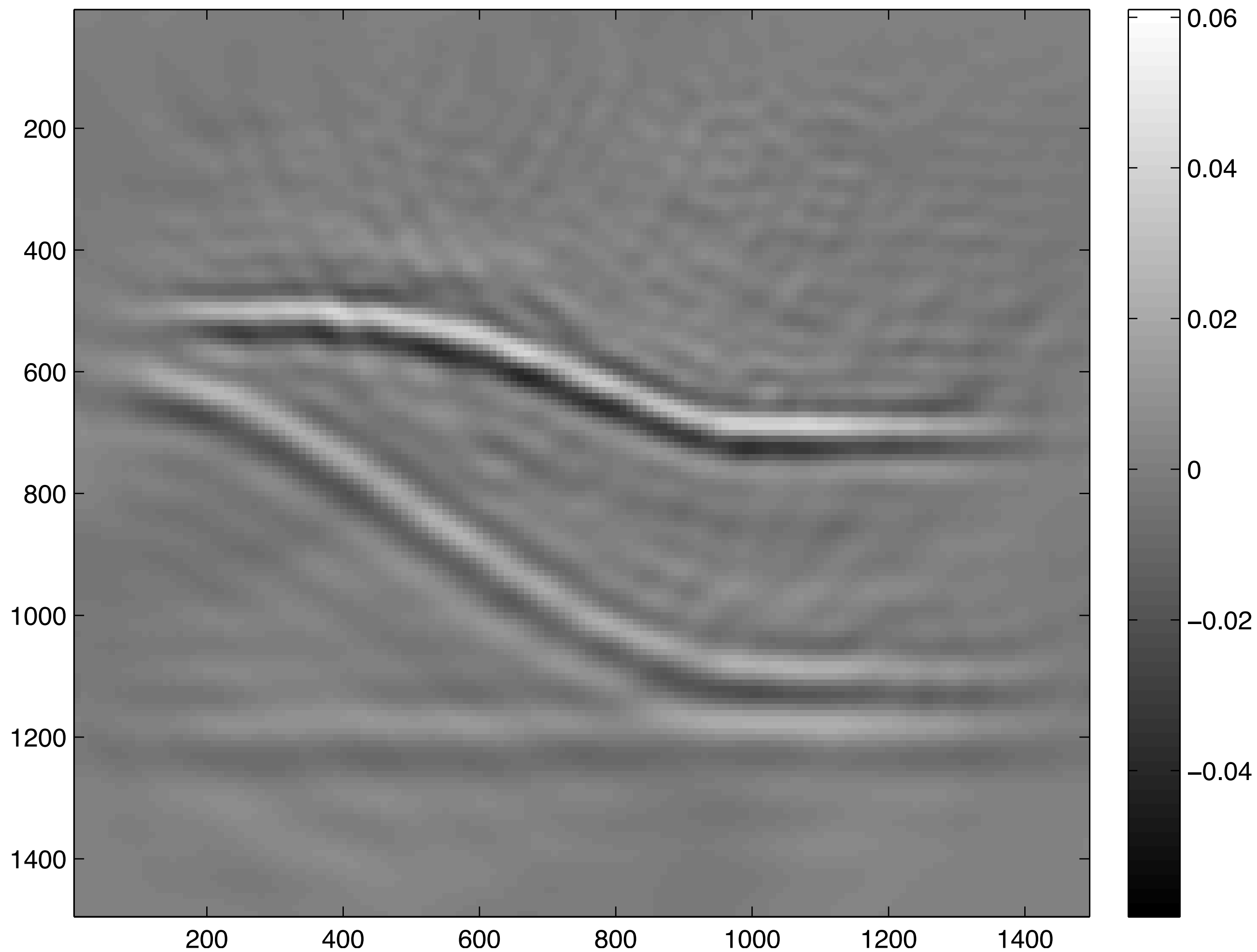


PCG - 2 iterations using noisy test functions

Inversion Results



LSQR - 15 iterations



PCG - 3 iterations using noisy curvelet test functions

Future Work

- Enforce positive, symmetric definiteness of this preconditioner
- Approximate the inverse square root of the Hessian - useful in the context of sparse inversion
- Understand the role of the expansion parameters

Future Work

- Test out this approach on a more realistic model

Conclusion

- Here we have extended the work of L. Demanet et al. to a more realistic scenario than previously considered
- This method improves the convergence of least-squares migration
- Still much work to be done to develop this technique into a practical preconditioner

References

- L Demanet, PD Létourneau, et al. Matrix probing: a randomized preconditioner for the wave- equation Hessian. Applied and Computational Harmonic Analysis, 2011.

- Laurent Demanet and Lexing Ying. Discrete symbol calculus. SIAM Review, 2008

References

- Halko, N. and Martinsson, P. G. and Tropp, J. A. *Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions.* SIAM Review, 2011

SLIM

References

Scaling-based Methods

- R. Nammour and W. Symes, *Approximate constant-density acoustic inverse scattering using dip- dependent scaling,* Proceedings of the 80th SEG Annual International Meeting, 2009
- F. Herrmann, C. Brown, et al., *Curvelet-based migration preconditioning and scaling*, Geophysics 2009