Recent Developments in Preconditioning the FWI Hessian

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

- Gradient-based methods converge *slowly*
- Gauss-Newton type methods require inverting the Hessian
- In seismic applications, the Hessian is large, dense, and poorly conditioned

Motivation

- Solely optimization-based approaches to dealing with the Hessian (or its inverse) are too expensive
- The Hessian is not strictly low rank in the SVD sense

Motivation

- The Hessian has a (somewhat) known mathematical structure that we can exploit

SLIM

Previous Work

- Previous work by Symes, Hermann, others on *scaling-based* methods (diagonal approximations to the Hessian)
- Exploiting the near-diagonal nature of the Hessian acting on curvelets

Overview

- Optimization context
- Pseudo-differential operators
- Matrix probing

Optimization Context

Problem:

$$\min_m \phi(m) = \frac{1}{2}\|F[m] - d\|_2^2$$

where

$$F[m]\big|_{\text{receivers}} = u$$

$$\nabla^2 u + \omega^2 m^2(x) u = f$$

Optimization Context

$$J = \frac{\delta F}{\delta m}$$

$$\nabla_m \phi(m) = J^*(F[m] - d)$$

$$H_{GN} = J^* J$$

Optimization Context

Looking to solve:

$$Hp = -\nabla_m \phi(m)$$

Left Preconditioning

$$PHp = -P\nabla_m \phi(m)$$

Preconditioning

- Ideally:

$$P = H^{-1}$$

- Realistically:
  PH should have *clustered* eigenvalues and/or a lower condition number

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)

SLIM

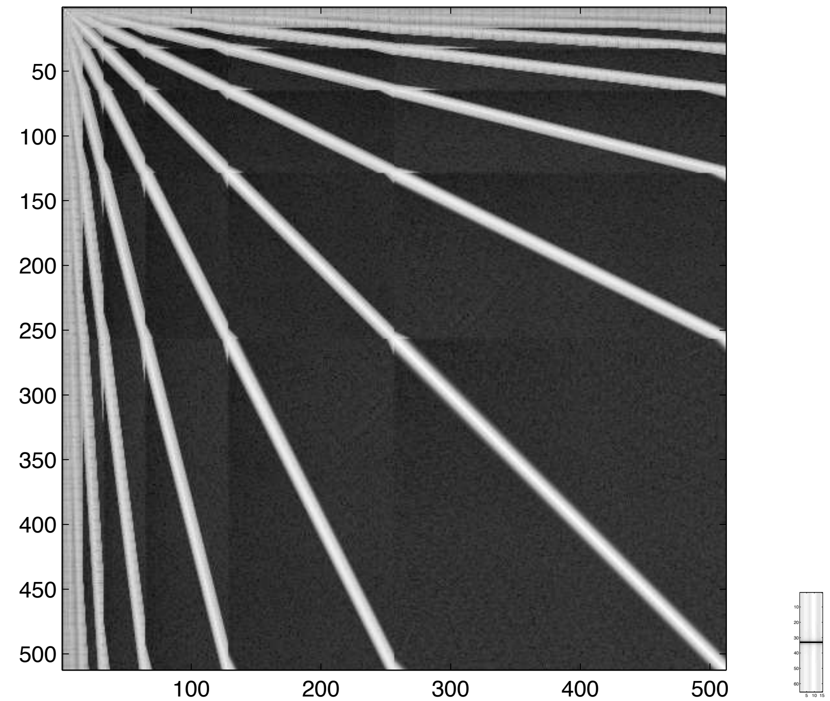
Pseudo-differential Operators

- 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

Pseudo-differential Operators



(a) (b)

adapted from Demanet '08

a) $O(N)$ storage in the wavelet basis,

b) $O(\log N)$ storage in the basis described in *Discrete Symbol Calculus*

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

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

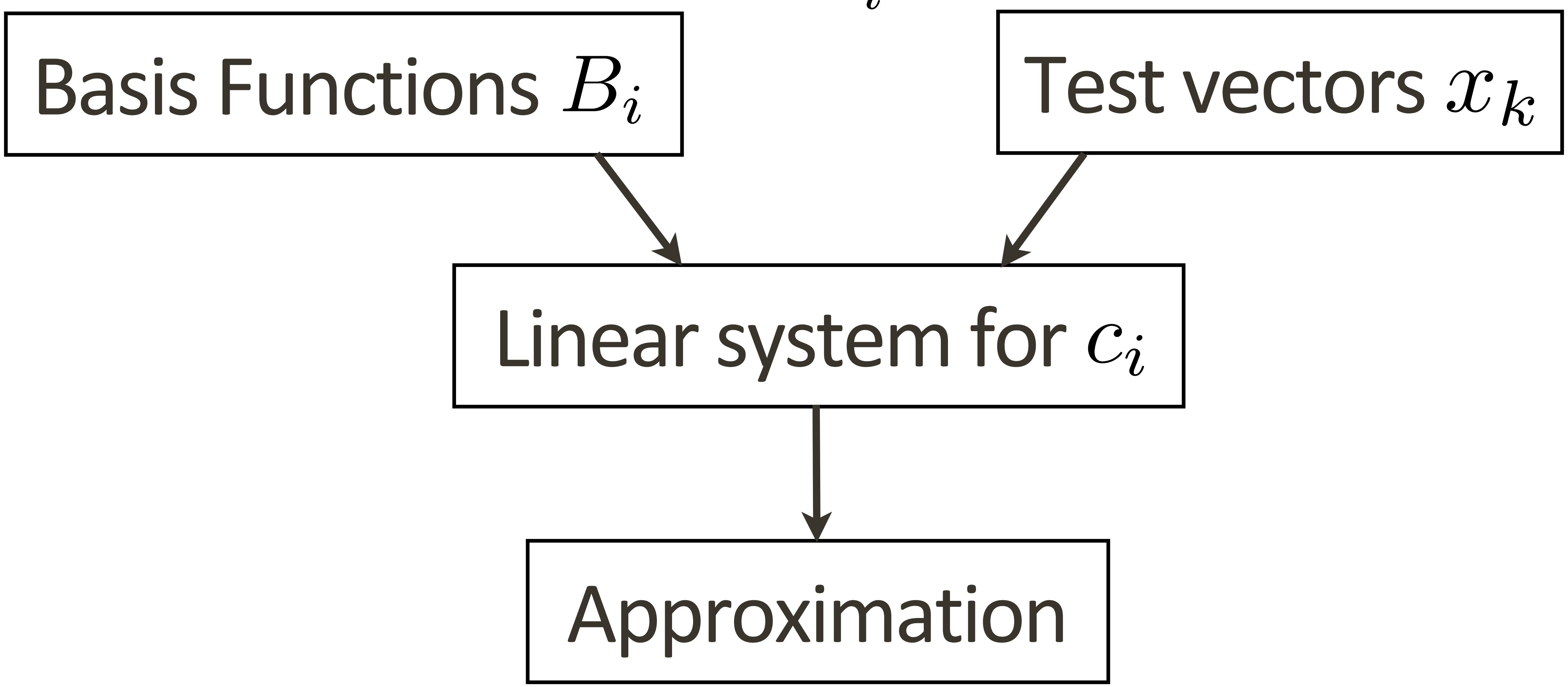
for known operators B_i,
unknown coefficients c_i

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

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



Matrix Probing

$$Hx_k = y_k$$

- x_k are test vectors

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

SVD Approach

$$H^{-1} \approx \sum_{i=1}^{p} \lambda_i u_i v_i^T$$

- u_i, v_i are vectors, so $u_i v_i^T$ are rank-1 matrices
- unfortunately both H and H^{-1} have high-rank in this basis

Basis for PDOs

- Using the basis

$$B_{\lambda,q_1,q_2} = \int e^{2\pi i x\cdot\lambda} e^{2\pi i q_1 \theta} T L_{q_2}(|k|)|k| dk$$

results in a fast converging expansion for a large class of PDOs

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

Random Test Vectors

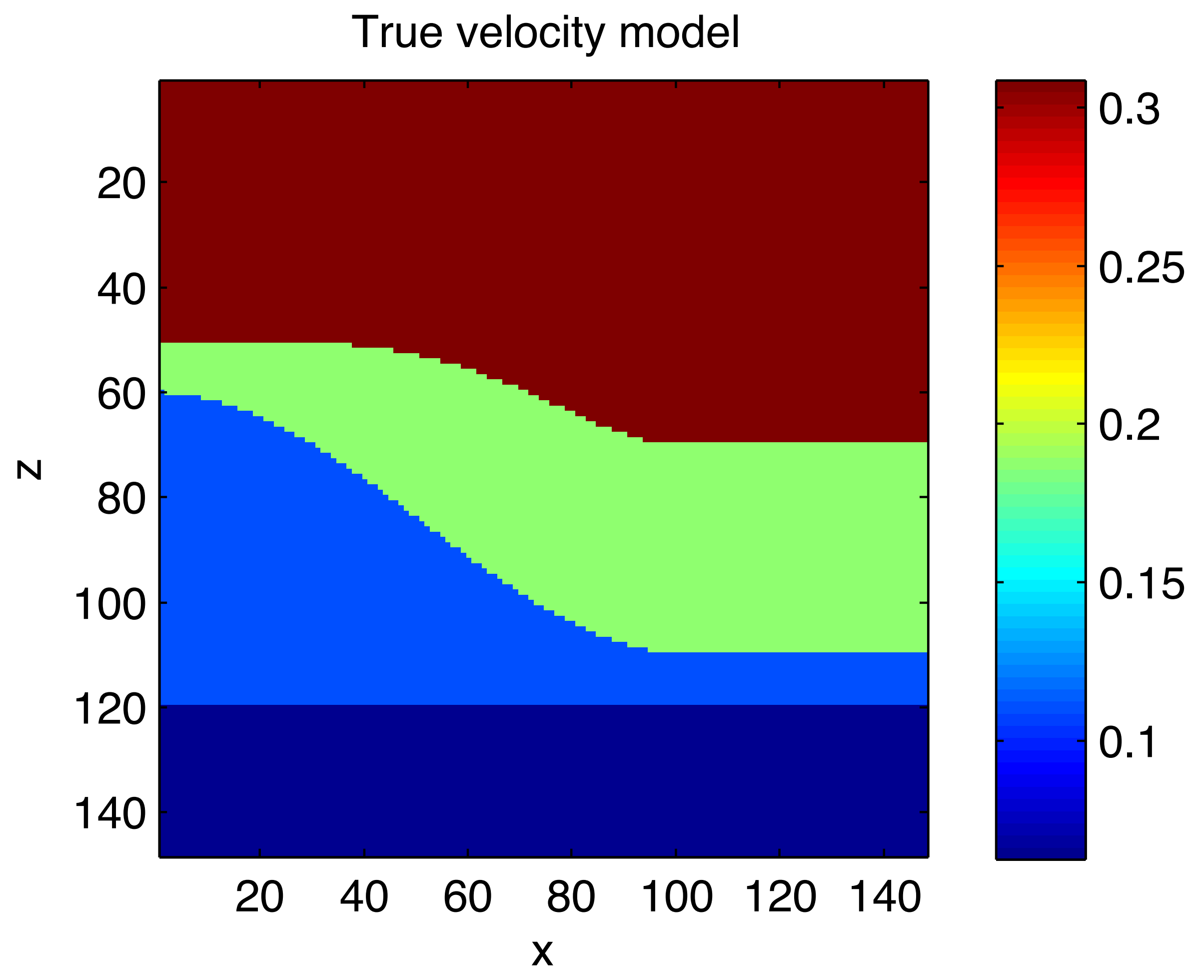
- Various choices of test vectors
- Gaussian noise - too much energy in the nullspace of H
- Gaussian noise to which the Hessian is applied - avoids the nullspace, but expensive to compute

Random Test Vectors

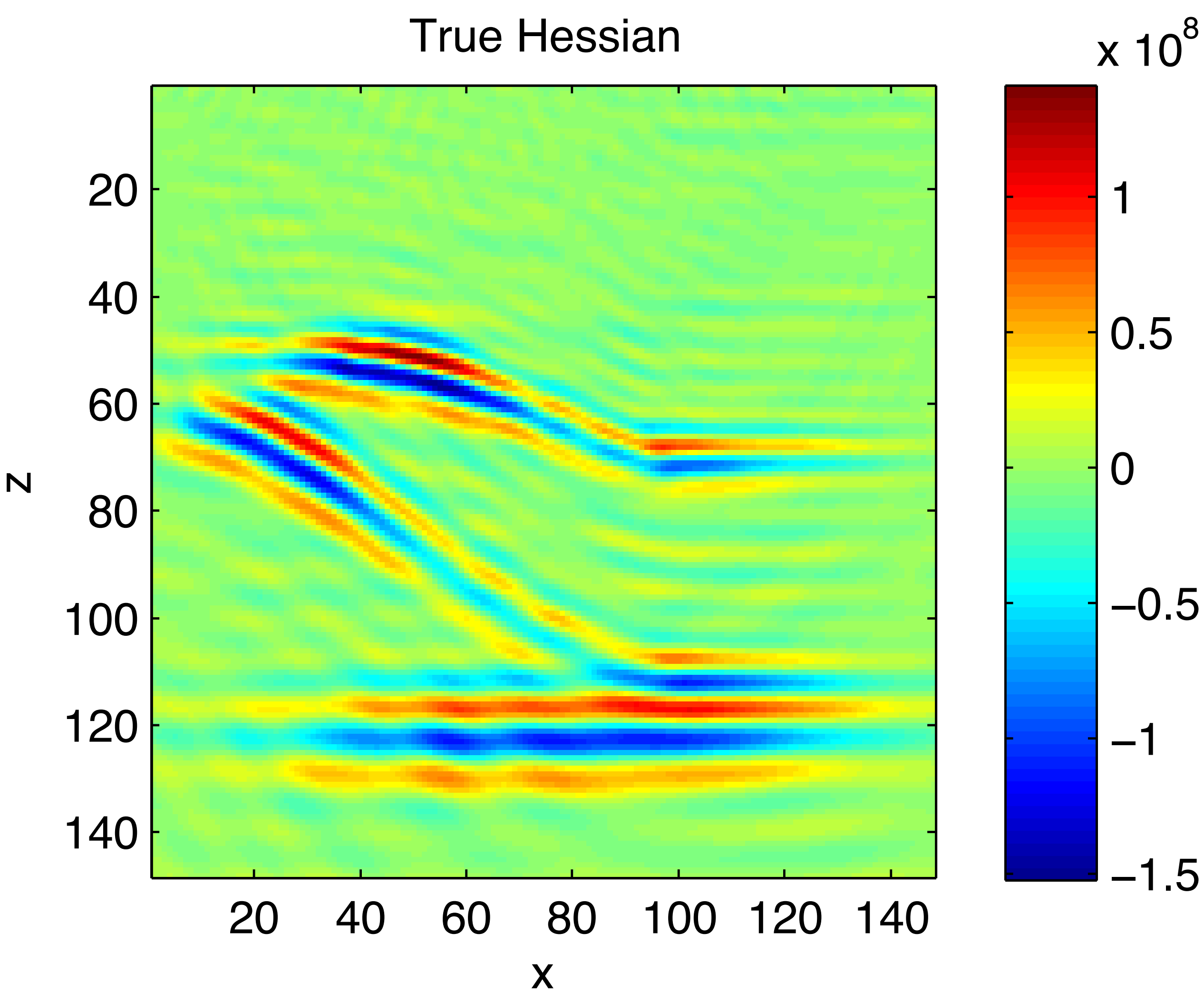
- Various choices of test vectors
- Noise in the Curvelet domain taken with a 1-0 mask
- Curvelets discriminate space-frequency regions where H is effectively zero
- More details in L. *Demanet Et Al. - 2011*

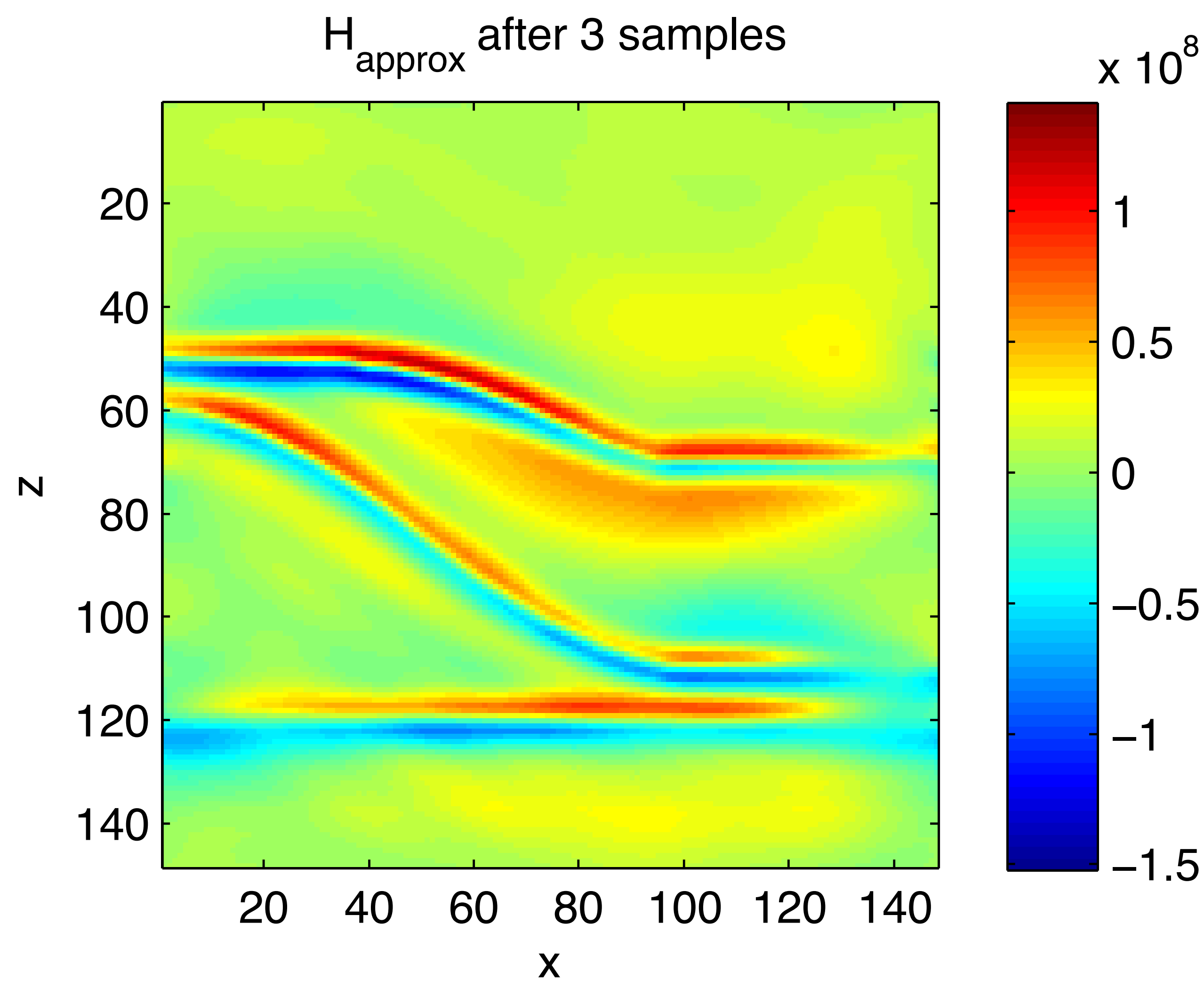
Results

- Synthetic Clay Model
- Forward Hessian probing

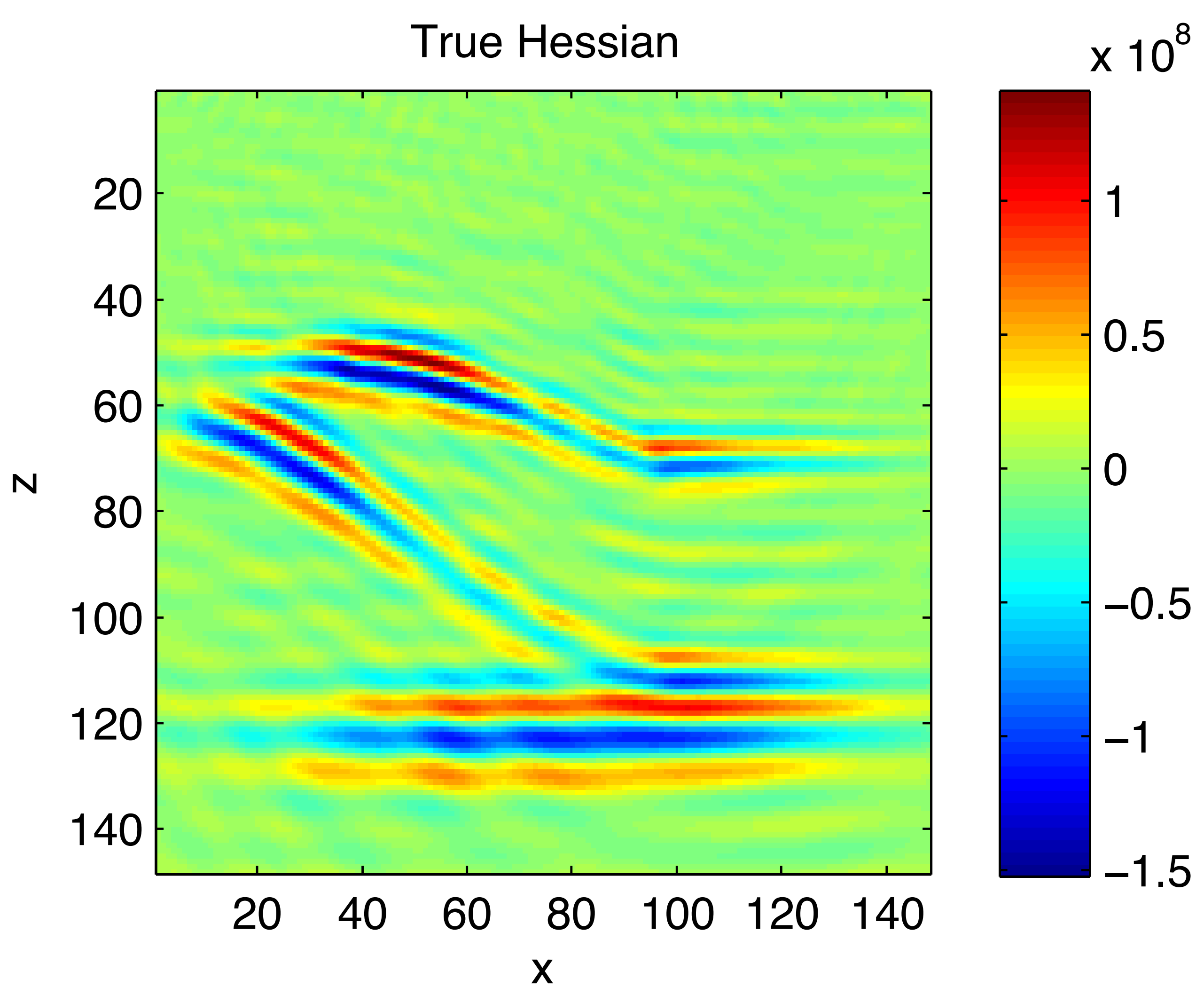


Results

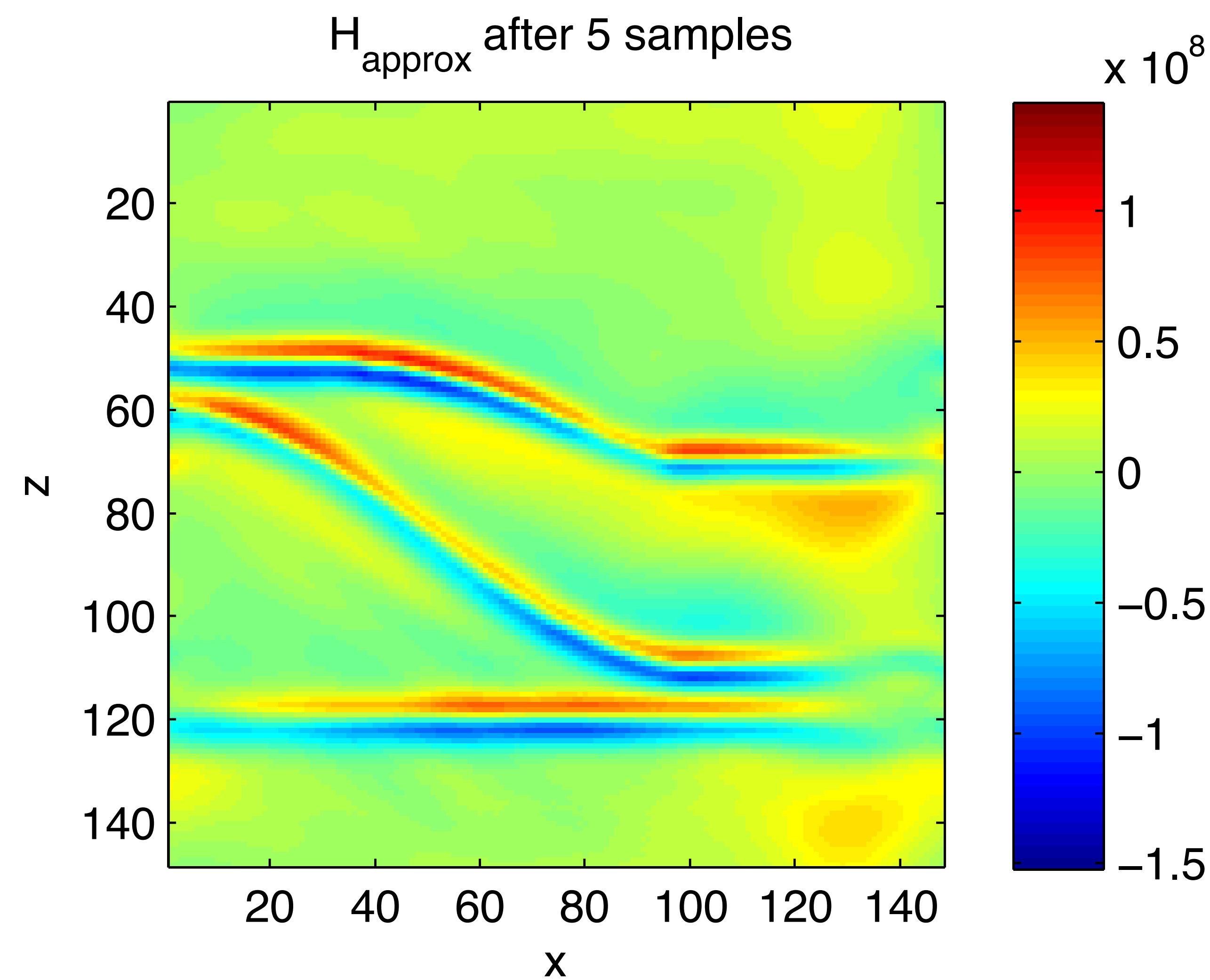

True Hessian
x 10^8
1
0.5
0
−0.5
−1
−1.5
20
40
60
80
100
120
140
z
x


H approx after 3 samples
x 10^8
1
0.5
0
−0.5
−1
−1.5
20
40
60
80
100
120
140
z
x

Results



True Hessian
x 10^8
1
0.5
0
−0.5
−1
−1.5
z
20
40
60
80
100
120
140
x



H approx after 5 samples
x 10^8
1
0.5
0
−0.5
−1
−1.5
z
20
40
60
80
100
120
140
x

Conclusion

- Use existing mathematical knowledge of the structure of the Hessian to obtain information about the inverse Hessian
- Fewer overall optimization iterations as a result

Future Directions

- Dimensionality reduction
- Probing the full Hessian (of which the Gauss-Newton Hessian is a part)
- Adapting this technique to large scale problems

Future Work

- Dealing with low-frequencies
- Incorporating work into a hierarchical low-rank matrix framework

Acknowledgements

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

SLIM

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

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