# Matrix Probing and Simultaneous Sources: A New Approach for Preconditioning the Hessian

Curt Da Silva[1] and Felix J. Herrmann[2]
[1] Dept. of Mathematics [2] Dept. of Earth and Ocean SciencesUniversity of British Columbia Vancouver, BC, Canada

January 18, 2012

**Abstract**

Recent advances based on the mathematical understanding of the Hessian as, under certain conditions, a pseudo-differential operator have resulted in a new preconditioner by L. Demanet et al. Basing their approach on a suitable basis expansion for the Hessian, by suitably 'probing' the Hessian, i.e. applying the Hessian to a small number of randomized model perturbations, one can obtain an approximation to the inverse Hessian in an efficient manner. Building upon this approach, we consider this preconditioner in the context of least-squares migration and Full Waveform Inversion and specifically dimensionality reduction techniques in these domains. By utilizing previous work in simultaneous sources, we are able to develop an efficient least-squares migration scheme which recovers higher quality images and hence higher quality search directions in the context of a Gauss-Newton method for Full Waveform Inversion while simultaneously avoiding inordinate amounts of additional work.

## Introduction

Recent interest in preconditioning the Gauss-Newton Hessian (hereafter referred to simply as the Hessian) in full waveform inversion have yielded a variety of techniques for approximating the inverse Hessian. The crux of a number of these methods relies on exploiting the mathematical structure of the Hessian, understood to act as a pseudo-differential operator in the high-frequency limit under certain conditions on the medium parameter. Recent work by Demanet et al. (2011) outlines an approach to approximating the inverse Hessian by using a suitable, known orthogonal basis expansion for pseudo-differential operators and solving for the unknown coefficients by observing the action of the inverse Hessian on a set of random vectors (this is the so-called technique of matrix probing, see Halko et al. (2011) for further details). By solving a small linear system for the unknown coefficients, one can recover the approximate action of the inverse Hessian on any trial vector, and hence obtain a preconditioner for Gauss-Newton type methods.

This approach to preconditioning the Hessian is attractive for a number of reasons. For one, the size of the models in consideration as well as the highly-ill conditioned nature of the Hessian result in an inability to explicitly compute the matrix representation of the Hessian. As a result, one is limited to using preconditioners that only rely on already expensive Hessian-vector products. Previous approaches to preconditioning the Hessian include, among others, those by Symes (2008) and Herrmann et al. (2009), which involve estimating scaling operations in the spatial and Fourier, and spatial, Fourier, and Curvelet domains, respectively. Compared to scaling-based methods, which typically rely on estimating the scaling coefficients from migrated data, the matrix probing preconditioner instead fits against randomized model perturbations, which avoids biasing the fit against specific characteristics of the particular perturbation used.

In this paper, we explore the efficacy of the aforementioned preconditioner in the context of imaging. In the interest of keeping costs low without sacrificing image fidelty, we explore the effects of introducing dimensionality reduction techniques (in particular, simultaneous sources) into this preconditioning framework.

## Methodology

*Matrix Probing Preconditioner.* The original preconditioning technique described below can be found in Demanet et al. (2011). We consider a representation of $H$ as a *pseudo-differential operator* (of degree 1 and type $(1,0)$)

$$Hm(x) = \int a(x,k)\hat{m}(k)e^{ik\cdot x}\mathrm{d}k$$

which is a valid approximation to the Hessian in the high-frequency limit (due to the body of work by Beylkin (1985), Rakesh (1988), Ten Kroode and Smit (1998), and Stolk (2000)) under the following idealized conditions:

- sufficiently fine sampling of the data in the time and receiver coordinates
- full-aperture acquisition
- a point-impulse source wavelet
- a smooth and generic (in that there are no kinematic exceptions that would compromise migration as a microlocal inverse) background medium

Under these assumptions, the symbol of $H$, $a(x,k)$, exhibits particular decay and smoothness properties that makes it extremely compressible numerically. In particular, work by Demanet and Ying (2011) has shown that one can approximate $a(x,k)$ by

$$a(x,k) = \sum_{\lambda, q_1, q_2} c_{\lambda,m,n} e^{i\lambda \cdot x} e^{im\theta} TL_n(|k|)|k| + O(\varepsilon)$$

where $TL_n(|k|)$ is the $n$th Rational Chebyshev function (refer to Boyd and Boyd (2001) for an extensive treatise on the approximation properties of such functions). Here, the number of terms in the expansion is $O(\varepsilon^{-M})$ for each $M > 0$, regardless of the size of the matrix that realizes the Hessian, i.e. the size of the model space of the problem considered.

Each symbol $e^{i\lambda \cdot x} e^{iq_1 \theta} TL_{q_2}(|k|)|k|$ in turn gives rise to a pseudo-differential operator with that symbol (a so-called *elementary operator* in Demanet et al. (2011)), denoted $B_i$. Thus, one can write (disregarding the $O(\varepsilon)$ error and writing $i = (\lambda, q_1, q_2)$)

$$H = \sum_i b_i B_i$$

Due to Shubin (1978), one also has that $H^{-1}$ is a pseudo-differential operator of degree $-1$ and so we can write (ignoring the $O(\varepsilon)$ error as before) $H^{-1} = \sum_i c_i B_i$ using the same elementary operators $B_i$.

In order to estimate the coefficients $c_i$, we employ a variant of the *matrix probing* technique employed in Halko et al. (2011). One of the main ideas with matrix probing is that one can reliably recover a large portion of the range-space of a matrix by observing its action on a sequence of random vectors (which is especially useful with a rank-deficient matrix, where few random vectors are necessary). To that end, by applying $H$ to a randomized test vector $y$, we have that $x = Hy \leftrightarrow y = H^{-1}x$ so that the action of $H^{-1}$ on the vector $x$ is available. By choosing an appropriate sequence of $p$ randomized test vectors $y_k$ and denoting $x_k = Hy_k$, we can form the following linear system

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

which we can subsequently solve for the unknowns $c_i$. In less mathematical terms, we generate $p$ random velocity perturbations, which we subsequently demigrate and migrate (i.e. to which we apply the Hessian). We then fit the proposed expansion to the data generated by this process and use that linear combination of operators in order to precondition the Hessian.

Violating the aforementioned conditions on the sampling, aperture, wavelet, and medium (and in particular, limiting the aquisition aperture and the source wavelet frequencies) results in locations which are "kinematically invisible" to the Hessian, which creates phase-space regions where the symbol $a(x,k)$ takes on small values. These "shadow zones" gives rise to a subspace of the model space where $H$ produces small values, i.e. the numerical version of its nullspace.

To that end, the choice of test functions in this scheme is vital to acquire an accurate fit for $H^{-1}$. Test functions $y_k$ which contain too much energy in the nullspace of $H$ will skew the above fit, as writing $y_k = H^{-1}x_k$ is no longer valid since components in the nullspace of $H$ do not lie in the rangespace of $H^{-1}$. There are a number of available choices for test vectors described in the paper which take this observation into account that we will not detail here.

*Simultaneous Sources.* Despite the search direction improvement in the full waveform inversion framework as a result of using the aforementioned preconditioner, the size of the problems considered in practice is still dauntingly large. In evaluating $H$, one must solve three Helmholtz systems $F[m]u = q$ for each source $q$, where $F[m]$ is the discretized Helmholtz operator and $u$ is the wavefield. The number of sources required to adequately illuminate the body of interest can be prohibitively high, and PDE solves are typically the bottleneck in any Full Waveform Inversion environment. We resort to dimensionality reduction techniques in order to reduce the computational costs while simultaneously maintaining a high fidelity in the computed solution.

To whit, we employ the technique of *simultaneous sources* (detailed in, among other places, Beasley (2008)) in order to further reduce the total number of PDE solves required, and thus the overall computational time needed. The simultaneous sources technique exploits the fact that in the Helmholtz equation $H[m,k]u_s = q_s$ where $H[m,k]$ is a discretization of the Helmholtz operator $k^2 + \nabla^2$, for each source $q_s$, the wavefield $u_s$ is a linear function of $q_s$. Therefore, we can replace $N$ sources with $n \ll N$ sources generated as random superpositions of $q_s$, i.e. $\tilde{q}_i = \sum_{s=1}^{N} w_{is} q_s$ for $i = 1, \ldots, n$ and $w_{is}$ are randomly chosen weights, effectively replacing the $N \times k$ PDE solves in $F[m,k]u_s = q_s$ by $n \times k$ simultaneous source PDE solves. As a result, the cost of applying the migration, demigration, and Hessian operators are effectively reduced by a factor of $\dfrac{N}{n}$ in this framework. By employing this method, we introduce noisy crosstalk between previously distinct source signatures which is coloured Gaussian noise. In the context of the image estimation process, this crosstalk is in the range of migration operator and so by applying the preconditioner to the migrated image, the spectrum of the noise is significantly whitened.

## Examples

We test the proposed method on a simple synthetic clay example, modelling the complete dataset for 51 sources spaced evenly on the surface of a 152 $\times$ 152 grid and 37 frequencies from 1.5 Hz to 19.5 Hz. Using 10 simultaneous sources, we generate data using the linearized Born-scattering operator. We use the aforementioned preconditioner (using 5 Gaussian noise test vectors) in the context of a preconditioned conjugate gradient method, run for 5 iterations, to invert for the model perturbation. For comparison, we compare the result of running LSQR for 10 iterations (which results in the same number of PDE solves in both cases and is numerically equivalent to running CG on the Hessian).

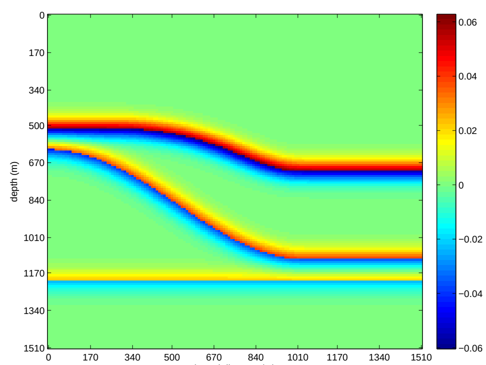


***Figure 1*** *Reference model perturbation.*

As we can see from the figures below, amplitudes along the reflector in the model perturbation are better restored via the preconditioning method compared to the vanilla LSQR algorithm, at the expense of increased noise due to the use of simultaneous sources. The inverted image can conceivably be denoised through some sort of sparsity promoting recovery algorithm such as SPGL1, a topic on which we will report at a future date.

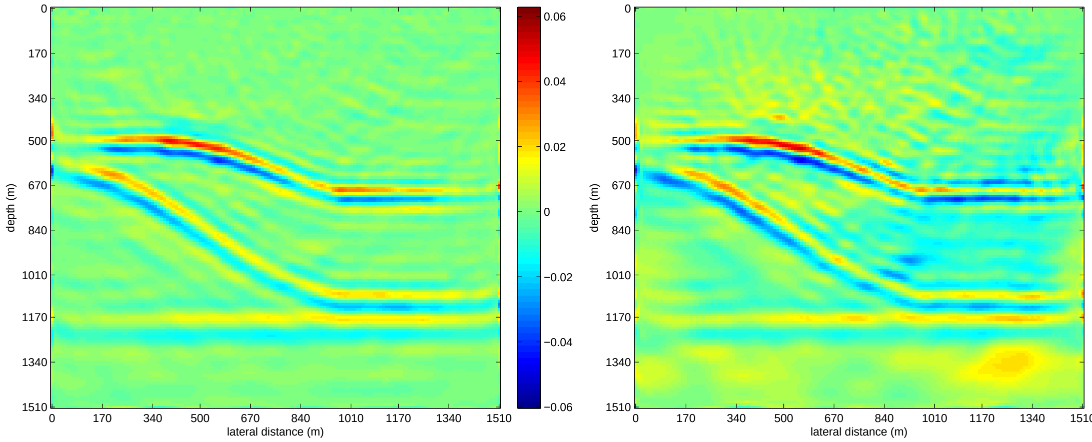


***Figure 2** Comparison between amplitude restored images.* Left: *LSQR with 10 iterations.* Right: *PCG with 5 iterations + Probing preconditioner.*

## Conclusions

Techniques for reducing the computational effort of least-squares migration continue to be relevant in a world of increasingly large data sets and time constraints for processing such data. The combination of the two techniques outlined above yield higher quality results for least-squares migration than traditional approaches while keeping costs manageable.

## Acknowledgements

We would like to thank all of our sponsors of the SINBAD project for their continued support. Many thanks to Tristan van Leeuwen for providing an appropriate problem with which to test this method.

## References

Beasley, C.J. [2008] A new look at marine simultaneous sources. *The Leading Edge*, **27**, 914–917.

Beylkin, G. [1985] Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics.*

Boyd, J.P. and Boyd, J.P. [2001] *Chebyshev and Fourier spectral methods*. Dover Pubns.

Demanet, L., Létourneau, P.D., Boumal, N., Calandra, H., Chiu, J. and Snelson, S. [2011] Matrix probing: A randomized preconditioner for the wave-equation hessian. *Applied and Computational Harmonic Analysis*, ISSN 1063-5203, doi:10.1016/j.acha.2011.03.006.

Demanet, L. and Ying, L. [2011] Discrete symbol calculus. *SIAM Review*, **53**(1), 71–104, doi:10.1137/080731311.

Halko, N., Martinsson, P.G. and Tropp, J.A. [2011] Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, **53**(2), 217–288, doi:10.1137/090771806.

Herrmann, F., Brown, C., Erlangga, Y. and Moghaddam, P. [2009] Curvelet-based migration preconditioning and scaling. *Geophysics*, **74**(4), A41.

Rakesh [1988] A Linearised inverse problem for the wave equation. *Communications in Partial Differential Equations*, **13**(5), 573–601.

Shubin, M.A. [1978] Almost periodic functions and partial differential operators. *Russian Mathematical Surveys*, 1+, doi:10.1070/RM1978v033n02ABEH002303.

Stolk, C. [2000] Microlocal analysis of a seismic linearized inverse problem. *Wave Motion*, **32**(3), 267–290.

Symes, W.W. [2008] Approximate linearized inversion by optimal scaling of prestack depth migration. *Geophysics*, **73**(2), R23–R35, doi:10.1190/1.2836323.

Ten Kroode, A. and Smit, D. [1998] ScienceDirect - Wave Motion : A microlocal analysis of migration. *Wave Motion*.