Approximate inversion of the wave-equation Hessian via randomized matrix probing
Pierre-David Letourneau, Stanford University, Laurent Demanet*, Massachusetts Institute of Technology and Henri Calandra, Total SA

SUMMARY

We present a method for approximately inverting the Hessian of full waveform inversion as a dip-dependent and scale-dependent amplitude correction. The terms in the expansion of this correction are determined by least-squares fitting from a handful of applications of the Hessian to random models — a procedure called matrix probing. We show numerical indications that randomness is important for generating a robust preconditioner, i.e., one that works regardless of the model to be corrected. To be successful, matrix probing requires an accurate determination of the nullspace of the Hessian, which we propose to implement as a local dip-dependent mask in curvelet space. Numerical experiments show that the novel preconditioner fits 70% of the inverse Hessian (in Frobenius norm) for the 1-parameter acoustic 2D Marmousi model.

INTRODUCTION

Much effort and progress has been made over the past several years to cast reflection seismic imaging as a waveform inversion problem, and solve it via all-purpose optimization tools. This shift of agenda has generated no correction, many factors come into play to degrade the amplitude resulting from a single migration of the reflectors. The terms in the expansion of this correction are determined by least-squares fitting from a handful of applications of the Hessian to random models — a procedure called matrix probing. We show numerical indications that randomness is important for generating a robust preconditioner, i.e., one that works regardless of the model to be corrected. To be successful, matrix probing requires an accurate determination of the nullspace of the Hessian, which we propose to implement as a local dip-dependent mask in curvelet space. Numerical experiments show that the novel preconditioner fits 70% of the inverse Hessian (in Frobenius norm) for the 1-parameter acoustic 2D Marmousi model.

The difficulty of inverting the Hessian is due to the fact that it is too large to be represented as a matrix, and applying it to a function in model space is in itself a very computationally intensive operation. To leading order, the wave-equation Hessian is the composition $F^*F$, where $F^*$ is the migration operator, and $F$ is the demigration operator (linearized forward modeling). In large-scale industrial applications, a prestack depth migration $F^*$ can take weeks to operationalize on a cluster, since it involves mapping 3D model space (billions of unknowns) to possibly 5D data space (trillions of unknowns). Compressive strategies have recently been proposed to reduce this burden, such as various forms of source encoding or “supershots”, but they don’t change the fact that applying $F$ or $F^*$ is still considered a costly operation.

Accordingly, much of the research effort has so far (rightly) focused on considering inverting the Hessian as a very special kind of preconditioning problem, constrained by the availability of at most a handful of applications of $F$ or $F^*$. An early important contribution is that of (Claerbout and Nichols (1994)), where the illumination is treated as a scalar function, and determined from applying the Hessian to the migrated image. A refinement of this idea is proposed by (Rickett (2003)). (Herrmann (2003); Herrmann et al. (2009)) proposed to approximate the Hessian as a diagonal operation in curvelet space. (Guitton (2004)) proposed a solution based on a “nonstationary convolution” which essentially models illumination as a filter rather than a multiplications. (Symes (2008)) proposed to extend the model for the Hessian by combining the advantages of multiplication and filtering. See also (Bao and Symes (1996)). (Nammour (2008); Nammour and Symes (2009)) extended this model yet again through a dip-dependent scaling. These authors all propose to fit the Hessian or the inverse Hessian from its application to judiciously chosen functions in model space, such as the migrated image — a fitting procedure that has come to be known either as “scaling method” or “matrix probing”.

The contribution of this note is twofold. First, we expand the realization of the inverse Hessian to include additional degrees of freedom that allow to model dip-dependent and scale-dependent scalings in a seamless fashion via pseudodifferential amplitudes, properly discretized via so-called “discrete symbol calculus”. Second, we explain how probing should be extended to the randomized case, why this extension is important for robustness, and why being able to determine the spaces of the Hessian (nullspace and range space) is an important step to properly “color” randomness in model space. While the numerical experiments on the 2D acoustic Marmousi model are on a very modest computational scale, the potential to approximate the whole inverse Hessian in a quantitative manner (as we show, 70% in Frobenius norm) seems to be a first and
Randomized Hessian Preconditioner

should be of interest to the community. The investigation of preconditioning for waveform inversion in larger-scale multi-parameters settings is currently under way and will be reported elsewhere.

SETUP

A common high-level formulation of the inversion problem of exploration seismology is the minimization of the least-squares misfit $J[m] = \frac{1}{2} \| d - \mathcal{F}[m] \|^2_2$, where the waveform data $d$ is a function of source position, receiver position, and time; the model $m$ is a function of $x, y, z$ that stands for isotropic wave speed (in this note), or other parameters such as elastic moduli and density; and the (nonlinear) forward modeling operator $\mathcal{F}$ results from simulating wave equations forward in time and sampling the wavefields at the receivers. We denote by $F$ the linearization of $\mathcal{F}$ about a background model velocity, and by $F^*$ the corresponding migration (imaging) operator. A gradient descent step for minimizing $J$ yields a model update of the form $\delta m = \alpha F^* (d - \mathcal{F}[m])$ (for some scalar $\alpha$), while a Gauss-Newton step would give $\delta m = (F^* F)^{-1} F^* (d - \mathcal{F}[m])$. The Gauss-Newton step solves the linearized problem exactly in a least-squares sense, but it is of course much harder to compute than a gradient step. The product $F^* F$ is called normal operator: it is the leading-order approximation to the wave-equation Hessian $H = \frac{\partial^2 J}{\partial m \partial m}$. (A true Newton step would in-
Randomized Hessian Preconditioner

An adequate realization of “local reflectors” in terms of bandlimited wavefronts is the curvelet transform. Curvelets are anisotropic wavelets, well-localized in phase-space (Candes and Donoho (2003a), Candes and Donoho (2003b)). They can be used to filter out the null-space components of any given model as follows,

1. Create a random model \( m \) (white noise);
2. Take a forward fast curvelet transform Candes et al. (2006);
3. Use ray-tracing to remove elements of the null-space (misaligned local reflectors);
4. Apply the inverse fast curvelet transform to get the filtered random model \( m_1 \).

A depiction of the result obtained after applying the above algorithm to white noise is shown in Figure 1. Once such vectors are available, we implement the algorithm introduced earlier (apply \( H \) to \( m_1 \); solve for the coefficients \( c_i \)).

Figure 1: White noise model (top), Typical model in the range space of \( H \): curvelet-masked coloring (bottom)

EXAMPLES

We apply the algorithm presented in the previous section to the single-parameter isotropic acoustic Marmousi benchmark, and compare the performance with the Nammour-Symes deterministic algorithm. Figure 2 shows (from left to right) the original Marmousi model; the result of 200 gradient descent iterations for solving the linearized least-square problem in a regularized fashion (hereby “expensive inversion”); the migrated image \( (F^*d) \); and the solution obtained after applying our approximate inverse with 4 random vectors \( m_{2,k} \), \( k = 1, 2, 3, 4 \) to \( F^*d \).

This suggests that the computational savings over gradient descent are of a factor at most 50. All experiments presented here were carried out with a very smooth background velocity. We found that, for a fixed number of degrees of freedom, the performance mildly degrades as the background becomes rougher.

The migrated image suffers from a lack of illumination at large depths, and a narrow spatial bandlimit. A single application of the preconditioner fixes these problems, and goes a long way toward performing full linearized inversion. Figure 3 shows the relative mean-squared error (MSE) between the “expensive inversion” and the solution obtained after applying the preconditioner. Any MSE below 1 means that the preconditioner is working. \( R_1, R_3 \) and \( R_5 \) refer to the number of random models used to fit the inverse Hessian, namely 1, 3 and 5 respectively. NS1 refers to the deterministic Nammour-Symes algorithm where a single function is used to fit the degrees of freedom (the migrated image). Performance decreases quickly when more vectors belonging to the Krylov subspace of \( H \) (i.e. \( HF^*d, H^2F^*d, \ldots \)) are used to fit the degrees of freedom in the NS algorithm.

What sets apart randomized matrix probing is that the approximation is robust: thanks to the randomness of the training models, we are recovering an approximation to the full inverse. Numerical experiments confirm this claim in Figure 4. In this particular example, we generated several random trial models (different from the original training functions) and applied the NS scheme and our algorithm to recover each one. We present the averaged MSE as a function of the number of training functions and the number of degrees of freedom in the preconditioner. In the limit of a large number of trial models, this average MSE converges to the Frobenius (a.k.a. Hilbert-Schmidt norm), hence our claim that the inverse Hessian is approximated with a 30% relative error.

CONCLUSIONS

We have presented a new design for a preconditioner for the wave-equation Hessian based on ideas of randomized testing, pseudo-differential symbols, and phase-space localization. The proposed solution is effective both visually and quantitatively (error bounds). The precomputation requires applying the Hessian once, or a handful of times. Fitting the inverse Hessian involves solving a small least-squares problem, of size \( p \times p \), where \( p \) is much smaller than the size of model space. It is anticipated that the techniques developed in this paper will be of particular interest in 3D seismic imaging and with more sophisticated physical models that require identifying a few different parameters (elastic moduli, density). In that setting, properly inverting the Hessian with low complexity algorithms to unscramble the multiple parameters will be particularly desirable.
Figure 2: In order from top to bottom: Reference model; “expensive inversion”; migrated image; preconditioned migrated image.

Figure 3: Relative MSE vs number of degrees of freedom for the Marmousi model. The rank of the Hessian at level $1e-3$ is $\sim 2500$, and correspondingly a good number of degrees of freedom for the preconditioner is in the high hundreds.

Figure 4: Generalization error (Relative MSE vs number of degrees of freedom). Performance increases with the number of training models (randomized strategy, solid lines) but decreases with the dimension of the Krylov subspace (deterministic strategy, dotted lines).

ACKNOWLEDGMENTS

PDL and LD are grateful to Total SA, the National Science Foundation, and the Alfred P. Sloan Foundation for support.
REFERENCES

Bao, G., and W. W. Symes, 1996, Computation of pseudodifferential operators: SIAM J. Sci. Comput., 17, no. 2, 418–429.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform: J. Math. Phys., 26, 99–108.

Candes, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Model. Simul., 5, no. 3, 861–899.

Candes, E. J., and D. L. Donoho, 2003a, Continuous curvelet transform: I. resolution of the wavefront set: Appl. Comput. Harmon. Anal., 19, 162–197.

——–, 2003b, Continuous curvelet transform: II. discretization and frames: Appl. Comput. Harmon. Anal., 19, 198–222.

Chiu, J., and L. Demanet, 2012, Matrix probing and its conditioning: SIAM J. Numerical Analysis, 50, 171–193.

Claerbout, J., and D. Nichols, 1994, Spectral preconditioning: Technical Report 82, Stanford University.

Demanet, L., P. Letourneau, N. Boumal, H. Calandra, and S. Snelson, 2012, Matrix probing: a randomized preconditioner for the wave-equation hessian: Appl. and Comp. Harmon. Anal., 32, 155–168.

Demanet, L., and L. Ying, 2011, Discrete symbol calculus: SIAM Review, 53, 71–104.

Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, 69, 1017–1024.

Herrmann, F., 2003, Multi-fractional splines: application to seismic imaging: Proc. SPIE Wavelets X conf., 5207.

Herrmann, F., C. Brown, Y. Erlangga, and P. Moghaddam, 2009, Multi-fractional splines: application to seismic imaging: Geophysics, 74, 41–46.

Nammour, R., 2008, Approximate inverse scattering using pseudodifferential scaling: Master’s thesis, Rice University.

Nammour, R., and W. W. Symes, 2009, Approximate constant-density acoustic inverse scattering using dip-dependent scaling: SEG 2009 Proceedings.

Rakesh, 1988, A linearized inverse problem for the wave equation: Comm. PDE, 13, no. 5, 53–601.

Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: Geophysics, 68, 1371–1379.

Symes, W., 1995, Mathematics of reflection seismology.

Symes, W. W., 2008, Approximate linearized inversion by optimal scaling of prestack depth migration: Geophysics, 73, 23–35.

ten Kroode, A., D. J. Smit, and A. R. Verdel, 1998, A microlocal analysis of migration: Wave Motion, 24, 149–172.