A preconditioner for the wave-equation Hessian via matrix probing

Pierre-David Letourneau, Stanford University, Laurent Demanet, Massachussetts Institute of Technology and Henri Calandra, Total SA*

Summary

In this paper, we present a new way of approximating the inverse of the wave-equation Hessian, also called normal operator, arising in the context of wave-based imaging such as seismic imaging. Our approach is based on the pseudo-differential nature of the operator. Thanks to this property, both the operator and its inverse can be approximated through a linear combination of only a few basis matrices. The coefficients in this expansion are obtained via least-square fitting from a certain number of applications of the normal operator on adequate randomized trial functions built in curvelet space (matrix probing). It can be shown that the least-square system thus obtained is well-conditioned with high probability. The choice of basis matrices is such that the approximation to the operators can be applied with low computational complexity. We give details on how to construct appropriate trial functions and demonstrate the performance of the preconditioner (approximate inverse) through several examples.

Introduction

The major goal of reflection seismology is to recover the physical properties of the subsurface for some region of space. Generally, the only information available to the investigator are surface measurements (seismograms) collected at various locations over some finite time interval. We shall denote these measurements as $d(r, s, t)$ where r and s represent the receiver and source position respectively, and t represents time. In addition to this, we shall assume that the density is constant throughout the medium in the region of interest so that only the speed of sound $m(x)$ needs to be recovered (x is a spatial variable).

A common way to approach this inversion problem is through the least-square functional,

$$J[m] = \frac{1}{2}||d - \mathcal{F}[m]||_2^2.$$

where we seek to minimize the misfit error $J[m]$. Here, $\mathcal{F}$ represents the modeling or imaging operator ; it takes as input a given velocity model $m_0(x)$ and produces an output $d_0(r, s, t)$ in data space. When an acceptable estimate of the background velocity is available, it is mathematically reasonable to assume that $\mathcal{F}$ is a *properly linearized* operator. From a practical point of view, this means that $\mathcal{F}$ can be represented as a matrix. Thus, upon discretizing our region we can write

$$J[m] = \frac{1}{2}||d - Fm||_2^2,$$

where F is now a matrix and m is a vector representing the speed of sound at every point of the discretization. The solution of this system is well-known and given by the normal equations,

$$(F^*F)m = F^*d \quad , \quad m = (F^*F)^{-1}F^*d$$

We refer to the wave-equation Hessian as the normal operator F^*F in this expression. As can been seen, the operator needs to be inverted to reach the solution. This is the problem we shall address throughout the remainder of this paper.

Theory

Several techniques are available to solve such a system. They are commonly separated into two families : direct methods (such as Gauss elimination) and iterative methods (such as Jacobi iterations). For this particular problem however, we need to rule out direct methods. This is because they usually require $O(N^3)$ steps which is prohibitively expensive especially with seismic problems where N, the number of grid points, can be very large. We are therefore left with iterative methods.

Iterative methods generally tend to offer much better performances than direct methods on large problems. However, these performances degrade quickly as the condition number of the matrix, that is the ratio of the largest singular value over the smallest singular value, increases. Now, since the matrix F^*F is generally badly conditioned, some additional step is required for fast convergence, and this comes in the form of a preconditioner.

A preconditioner P is nothing more than an *approximate* inverse. The way it works is as follows : assume we are given a badly conditioned square matrix A and a vector b and need to solve the system $Ax = b$ iteratively. Because of the ill-conditioning, a large number of iterations is likely to be needed. However, by premultiplying both sides of the system by P we get,

$$PAx = Pb$$

where $PA \approx I$ since $P \approx A^{-1}$. This is a much better conditioned system which can now be solved quickly.

We are therefore interested in finding an approximate inverse for the matrix F^*F. At this point, we need to introduce the concept of pseudo-differential operator (ΨDO). A ΨDO L is a linear operator that is usually written in the following form,

$$Lm(x) = \int e^{-2\pi\imath x\xi} a(x,\xi)\hat{m}(\xi)\mathrm{d}\xi$$

where $\hat{m}(\xi)$ represents the Fourier transform of $m(x)$ and the symbol $a(x,\xi)$ satisfies very specific smoothness properties. A ΨDO usually operates on a function as a scaling factor (which depends on the location x) and a filter (which depends on the frequency ξ). To put it in the words of Nammour & Symes ([1]), it is a *dip-dependent* scaling and filtering.

There is a fair amount of literature pointing to the fact that, under fairly general assumptions, the wave-equation Hessian F^*F and its inverse $(F^*F)^{-1}$ belong to the family of pseudo-differential operators ([2], [3], [4], [5]). This is why we introduce the concept here. In addition, thanks to the smoothness of the symbol $a(x,\xi)$, such operators can very easily be compressed ([6]) i.e. we can write

$$F^*F \approx \sum_{n=1}^{p} \alpha_n C_n \quad , \quad (F^*F)^{-1} \approx \sum_{n=1}^{p} \beta_n C_n$$

where $\{C_n\}$ is an adequate set of basis matrices and p is small. The approximation error generally decays very fast with p.

Therefore, in order to find an approximation to $(F^*F)^{-1}$, it is sufficient to seek the coefficients $\{\beta_n\}_{n=1}^{p}$ in the above expression. For this purpose, we chose to use matrix probing. Matrix probing proceeds as follows :

1. Choose an appropriate vector : v
2. Apply the normal operator : $(\mathrm{F}^*\mathrm{F})v$
3. Apply the approximate inverse : $\sum_{n=1}^{p} \beta_n \left[C_n(\mathrm{F}^*\mathrm{F})v\right]$
4. Solve the linear system for the β_n's in the least-squares sense.

In short, what matrix probing does is to find coefficients $\{\beta_n\}$ such that $\sum_{n=1}^{p} \beta_n C_n$ best represents the inverse of F^*F on the vector v *solely*.

Up to this point, the algorithm is very similar in flavour to that of Nammour & Symes ([7], [1]). However, whereas the latter authors use the migrated image (F^*d) as the vector v, we propose to use multiple *random* vectors. There are good reasons why we believe this is an improvement. Among others can be found *generalizability* ; as we shall show, our method not only provide a good preconditioner but recovers an approximate inverse that is *uniformly good* among all vectors in the range of F^*F.

At this stage however, there are obvious issues with the method. First, there is no guarantee that a good approximation of the inverse on a single vector v (or a finite set of vectors $\{v_k\}_{k=1}^{K}$ for that matter) represents a good approximation of the inverse on *any other* vector. Secondly, if the operator F^*F possesses a non-trivial null-space and v happens to belong to this null-space, it is hopeless to try to recover v since $(F^*F)v \equiv 0$.

To alleviate the first issue, we make use of multiple *random* vectors v. In fact, we claim that the use of random vectors is the key to recovering an approximation to the *whole matrix* $(F^*F)^{-1}$. It would be much more difficult to carry through the argument were the vectors deterministic. Numerical experiments confirm this observation. In addition, it has been shown by ([8]) that, under certain conditions on the basis matrices, the use of random vectors leads to a well-conditioned linear system with high probability (step 4 in the above algorithm). Finally, applying the approximation can be done in very low complexity ([9]).

For the second problem, we need to characterize the null-space. For this purpose, the references mentioned above are once again useful since they provide explicit algebraic equations characterizing the null-space. From a physical standpoint, what these equations state is that an element belongs to the range space (the orthogonal complement of the null-space) of F^*F if it takes the form of a *small reflector* for which there exist a ray going from a given source to a given receiver and such that it is reflected in a specular manner on the reflector. The rays are assumed to travel in the background medium in accordance with the laws of geometrical optics.

The adequate mathematical object to express such a concept comes in the form of curvelets. In short, curvelets are anisotropic wavelets ; they are functions that are well-localized in phase-space and segregate between angular regions of the frequency domain ([10],[11]). In other words, they are the mathematical equivalent of the small reflectors introduced above. They can be used to filter out the null-space components of any given vector as follows,

1. Create a random vector v (white noise or Rademacher).
2. Proceed to a forward fast curvelet transform ([12]).
3. Use ray-tracing (ODE45, phase flow ([13]), etc.) to remove elements of the null-space.
4. Apply the inverse fast curvelet transform.

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.

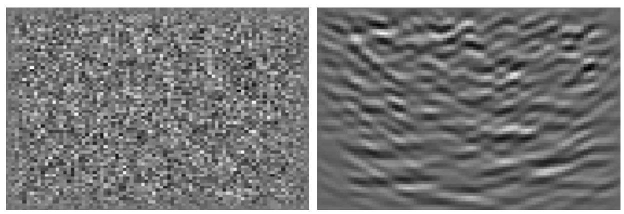

Figure 1: White noise(left), Result of applying a curvelet mask (right)

Examples

In this section, we present results obtained by applying the algorithm presented in the previous section to the classical Marmousi benchmark, and compare the performance with the Nammour-Symes deterministic algorithm.

First, we present results of a more qualitative nature ; Figure shows (from left to right) the original Marmousi model, the true solution of the least-square problem, the migrated image(F^*d only) and the solution obtained after applying our approximate inverse with 4 random vectors $\{v_k\}_{k=1}^4$ to F^*d. All experiments presented here were carried out with a uniform background velocity. Nonuniform background was also investigated and we found that, for a fixed number of parameters, the performance tend to decay as the background gets less and less smooth.

As can be seen, the migrated image already exhibits the high-frequency content of the solution i.e. the scattering surfaces can already be seen. However, it suffers from a lack of of illumination, especially for scatterers deep in the subsurface. On the other hand, we notice that an application of the approximate inverse attenuates this problem significantly. The image thus obtained is much closer to the solution than originally which is a compelling evidence that the preconditioner works.

Figure 3 shows the relative error between the true solution (Figure 2) and the solution obtained after applying the preconditioner (a MSE below 1 means the preconditioner is working). $R1$, $R3$ and $R5$ refer to the number of random

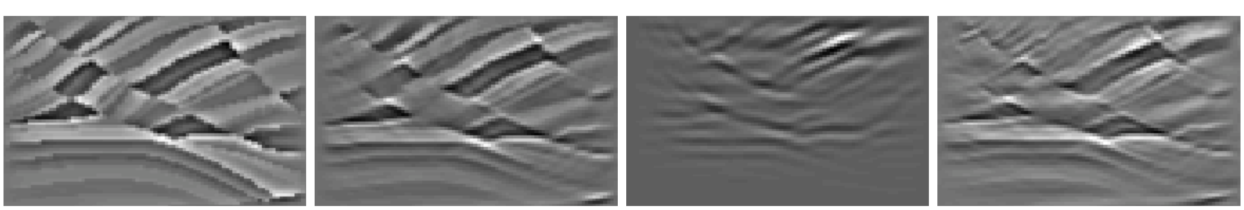

Figure 2: In order from left to right : Marmousi model, true solution, migrated image, application of preconditioner to migrated image

verctors used to fit the data, that is 1, 3 and 5 respectively. NS1 refers to the deterministic Nammour-Symes algorithm where a single vector is used to fit the parameters (the migrated image). For the latter, we only show the case where a single vector is used namely because it is the best case. Indeed, performances diminish when more vectors belonging to the Krylov subspace of F^*F (for instance $((F^*F)F^*d$, $(F^*F)^2F^*d$, ...) are used to fit the parameters.

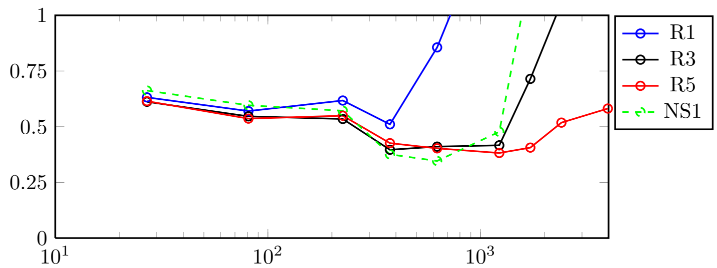

Figure 3: Relative MSE vs number of parameters for the Marmousi model

On this particular example, it seems that it would not be much advantageous to use our algorithm instead of the NS algorithm since both exhibit similar performance at their best. However, as was mentioned earlier we have reasons to believe that our approximation is more robust ; thanks to the randomness of the trial vectors, we are really recovering an approximate to the *full* inverse. That is, given *any* other vector of the form $w = F^*Fv$, the application of our approximate inverse to w shall recover a good approximation of v. This is hard to guarantee with deterministic vectors. This claim is bolstered through numerical experiments that are presented in Figure 4. In this particular example, we generated three random vectors (different from the original trial functions) and applied the NS scheme and our algorithm to try to recover each one. We present the averaged MSE as a function of the number of trial vectors and the number of parameters.

Conclusions

In conslusion, we have presented a way of obtaining a preconditioner for the wave equation Hessian based on ideas of randomized testing, pseudo-differential symbols, and phase-space localization. Numerical experiments show that the proposed solution belongs to a class of effective preconditioners. The precomputation only requires applying the wave equation Hessian once, or a small number of times. Fitting the inverse Hessian involves solving a small least-squares problem, of size p-by-p, where p is ordinarily much smaller than n and the Hessian is n-by-n.

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.

Aknowledgements

PDL and LD were supported by a grant from Total SA, and a grant from the National Science Foundation.

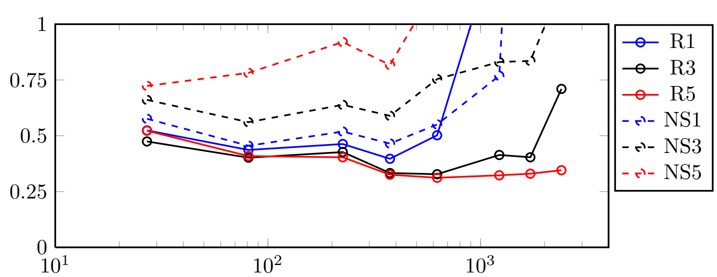

Figure 4: Generalization error (Relative MSE vs number of parameters

References

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

[2] G. Beylkin, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform, J. Math. Phys. 26 (1985) 99–108.

[3] Rakesh, A linearized inverse problem for the wave equation, Comm. PDE 13 (1988) 53–601.

[4] W. W. Symes, Mathematics of reflection seismology (1995).

[5] A. ten Kroode, D. J. Smit, A. R. Verdel, A microlocal analysis of migration, Wave Motion 24 (1998) 149–172.

[6] L. Demanet, L. Ying, Fast wave computation via Fourier integral operators, preprint (2010).

[7] R. Nammour, Approximate inverse scattering using pseudodifferential scaling, Master's thesis, Rice University (October 2008).

[8] J. Chiu, L. Demanet, Matrix probing and its conditioning, preprint (2011).

[9] G. Bao, W. W. Symes, Computation of pseudo-differential operators, SIAM J. Sci. Comput. 17 (1996) 418–429.

[10] E. J. Candes, D. L. Donoho, Continuous curvelet transform: I. Resolution of the wavefront set, Appl. Comput. Harmon. Anal. 19 (2003) 162–197.

[11] E. J. Candes, L. Demanet, D. L. Donoho, L. Ying, Fast discrete curvelet transforms, Multiscale Model. Simul. 5 (2006) 861–899.

[12] L. Ying, E. J. Candes, The phase-flow method, Journal of Computational Physics 220 (2006) 184–215.

[13] A. Guitton, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators, Geophysics 69 (2004) 1017–1024.

[14] J. E. Rickett, Illumination-based normalization for wave-equation depth migration, Geophysics 68 (2003) 1371–1379.

[15] J. Claerbout, D. Nichols, Spectral preconditioning, Tech. rep., Stanford University (1994).

[16] W. W. Symes, Approximate linearized inversion by optimal scaling of prestack depth migration, Geophysics 73 (2008) 23–35.

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

[18] F. Herrmann, Multi-fractional Splines: application to seismic imaging, Vol. 5207, Proc. SPIE Wavelets X conf., 2003.