Stanford Exploration Project, Report 82, May 11, 2001, pages 1–77

Estimating a pseudounitary operator for velocity-stack inversion

David E. Lumley[1]

ABSTRACT

I estimate a pseudounitary operator for enhancing an iterative conjugate-gradient (CG) inversion of CMP data to derive a best-fitting model in velocity-stack space. The amplitude balance of the operator is approximated with a simple time- and offset-variant operator weighting function. The spectral balance of the operator is tuned with a finite-aperture finite-fold approximation to the continuous integral rho filter. Convergence of the CG inversion and separation of multiple and primary velocity-stack energy is improved by application of the pseudounitary operator weights. Application of the rho filter improves the first few CG iterations, but counterintuitively slows or reverses convergence in subsequent iterations.

INTRODUCTION

Consider two categories of geophysical operators: (1) operators that are derived from fundamental principles of physics and constitutive relations, and (2) operators which map data to model spaces which are convenient for data processing. Physical operators model data from a physically realizable model space. An example of a physical operator is the the acoustic wave equation which models seismic data from a given spatial model of bulk modulus and density. The inverse of this physical operator estimates the earth's acoustic properties given input seismic data. The units of the model space are an important part of the operator. Given a physical operator, we may not have much freedom to modify the operator without violating the underlying physics. In contrast, a processing operator may map data to an unphysical space that happens to be convenient for processing reasons. An example of a processing operator is the hyperbolic velocity-stack operator which maps seismic data to a velocity-stack space. Although this space is unphysical, and the operator cannot be derived from constitutive relations of physics, it is convenient since multiple reflections may be isolated in velocity-stack space, allowing primary reflections to be enhanced in the processed version of the input seismic data (**?**). The units of the processing operator may or may not be important to the processing objective. Given a processing operator, we are completely free to modify the operator to optimize our processing goal, since there are no constitutive relations to satisfy. In this case it may be

[1]**email:** david@sep.stanford.edu

useful to modify the operator to make it more easily invertible (**?**), thus accelerating convergence (**?**) and freeing extra CG iterations to work on another aspect of the inverse problem, e.g., separating primary and multiple reflection velocity peaks in velocity-stack space (**?**). I present a methodology for modifying the hyperbolic velocity-stack operator to make it pseudounitary, and hence readily invertible. I show that improvements in CG convergence are obtained with the pseudounitary operator compared to the unweighted operator. I first review the theory of pseudounitary operators. I then describe how to estimate time and offset-variable weights to amplitude-balance the original unweighted operator. I propose a finite-aperture finite-fold approximation to the continuous integral rho filter to spectrally balance the original operator. Finally, I demonstrate the improved convergence of a CG velocity-stack inversion using the pseudounitary versus original hyperbolic stacking operator.

A PSEUDOUNITARY OPERATOR

Theory

Claerbout (**?**) proposed the concept of a pseudounitary operator. Given a general linear matrix/integral equation

$$\mathbf{y} = \mathbf{A}\mathbf{x}\,, \tag{1}$$

we would like to find a pseudounitary operator $\mathbf{A}$ such that

$$\mathbf{A}'\mathbf{A} \approx \mathbf{I}\,. \tag{2}$$

In this case, a solution $\mathbf{x}$ can be estimated in one gradient step by applying the adjoint of the pseudounitary operator

$$\mathbf{A}'\mathbf{y} = \mathbf{A}'\mathbf{A}\mathbf{x} \quad \approx \quad \mathbf{I}\mathbf{x} = \mathbf{x}\,. \tag{3}$$

Since the gradient $\mathbf{A}'\mathbf{y}$ is the first step in an iterative CG (conjugate gradient) solution for $\mathbf{x}$, finding a good pseudounitary operator $\mathbf{A}$ is likely to speed the convergence of the CG solution for $\mathbf{x}$.

Practice

Figure **??** shows a CMP gather and its semblance velocity scan. Note that the data are rich in multiple reflection energy. The multiples contaminate the primary reflection response and distort AVO amplitude information. This data is the subject of an amplitude-preserved multiple suppression study (**?**). Figure **??** shows the input CMP gather, its linear velocity scan, and its adjoint-modeled CMP gather. The linear velocity scan $\mathbf{v}$ is obtained by applying the velocity scan operator $\mathbf{S}$ to the CMP gather data $\mathbf{c}$:

$$\mathbf{v} \approx \mathbf{S}\mathbf{c}\,. \tag{4}$$

The velocity scan operator $\mathbf{S}$ sums along hyperbolic trajectories in the time domain, uses linear interpolation of input trace data values, and maps only within the non-muted zone of the gather. We would like to satisfy two objectives in this process:

1. The velocity scan operator $\mathbf{S}$ is easily invertible.
2. Primary and multiple reflections are well-separated in the velocity space $\mathbf{v}$.

If $\mathbf{S}$ is pseudounitary then a CG routine may rapidly converge on $\mathbf{S}^{-1}$ since $\mathbf{S}'\mathbf{S} \approx \mathbf{I}$. In this case, valuable CG iterations can be retargeted to satisfying constraints on $\mathbf{v}$ that separate multiple and primary velocity peaks.

../david2/./Figs/cmp-semb.ps

Figure 1: A CMP gather and its contoured velocity scan made with a semblance coherency measure and a nonlinear filter to enhance the energy of the velocity peaks. Note that the data are rich in multiple reflection energy.

OPERATOR WEIGHTING

Amplitude balancing

In the velocity-stack operator, I use weighting as a function of offset, operator traveltime, and cosine of the ray angle, such that

$$\mathbf{v} \approx \mathbf{SWc}\,, \tag{5}$$

and

$$W = f^{-d}\,(1+|h|^a)\,(1+t)^b\,\cos^c\theta\,, \tag{6}$$

where $|h|$ is the absolute offset, t is the time along the hyperbolic scan trajectory, θ is the reflection angle, and f is the fold of the data. The values of h and t could be made dimensionless by normalizing by h_{min} and Δt, for example, which would make the estimation of

../david2/./Figs/cmp-v-adj.ps

Figure 2: A CMP gather, its linear velocity scan, and its adjoint-modeled seismograms.

the parameters $\{a, b\}$ dependent only on the dataset and not the units of measurement. I adjust the weighting parameters $\{a, b, c\}$ so that the adjoint-modeled CMP gather is amplitude balanced with respect to the input CMP gather. In other words, the adjoint-modeled CMP gather should have the same amplitude decay in time and offset as the input CMP gather. I found a good amplitude balance can be obtained for this dataset using the parameter values $a = 0.5$, $b = -1$, $c = 0$, where offset is measured in [km] and recording time in [seconds]. The amplitude-balanced result is shown in Figure **??**. Additionally, the absolute scale is approximately correct if the fold weight is used with $d = 0.5$. The optimal operator weights for this dataset and measurement units can be written accordingly as:

$$W \approx \left(\frac{1}{\sqrt{f}}\right)\left(\frac{1+\sqrt{x}}{1+t}\right). \tag{7}$$

In contrast, Figure **??** shows that the adjoint-modeled CMP gather for $a = b = c = 0$ is not amplitude-balanced with the input gather, and also differs by an absolute scale factor of about 3600.

CG convergence

I coded a conjugate gradient solver to use the unweighted and weighted forms of the velocity-stack operators described in this section. The CG algorithm finds a minimum energy velocity scan model which minimizes the least-squares error between the input CMP gather and the adjoint-modeled CMP gather. I used the CG method described by Hestenes and Steifel

(1952). I let the CG algorithm perform 10 iterations and compared the weighted and unweighted operator results. The plot of their respective convergence paths appears in Figure **??**. The pseudounitary operator weighting improves the convergence of a conjugate gradient scheme in the first iteration, but the pseudounitary weighting becomes less effective at higher iterations. Overall, the pseudounitary weighting saves only about 1-2 CG iterations at best for this problem, and after 10 iterations, there is no difference in convergence speed. However, since the operator is modified by the pseudounitary weighting, the estimated velocity scan will necessarily also be different. I have found that pseudounitary weighting gives a slightly more pleasing separation of primary and multiple reflection energy in velocity-stack space than other operator weightings I have tested, including unit weights. The best-fitting adjoint-modeled CMP gather after 10 pseudounitary-weighted CG iterations is compared with the input CMP gather in Figure **??**, and their spectra are compared in Figure **??**.

../david2/./Figs/cmp-w1.ps

Figure 3: The input CMP gather (left) and the weighted adjoint-modeled CMP gather (right). Note that the adjoint is well amplitude balanced with respect to the input.

../david2/./Figs/cmp-w0.ps

Figure 4: The input CMP gather (left) and the unweighted adjoint-modeled CMP gather (right). Note that the adjoint is not amplitude balanced with respect to the input.

../david2/./Figs/iter-w10-ann.ps

Figure 5: Comparison of CG convergence between the unweighted (top curve) and pseudounitary-weighted (lower curve) velocity-stack operator.

../david2/./Figs/cmp-w10.ps

Figure 6: The input CMP gather (left) and the adjoint-modeled CMP gather (right) after 10 iterations of conjugate gradients.

../david2/./Figs/specs-w10.ps

Figure 7: Amplitude spectra of the input CMP gather (left) and the adjoint-modeled CMP gather (right) after 10 iterations of conjugate gradients.

THE RHO FILTER

Theory

It is desirable that the operator's spectral properties are such that adjoint-modeled data have the same spectrum as the input CMP data. This involves the application of a "rho" filter **R** to the operator such that

$$\mathbf{v} \approx \mathbf{SWRc} \ . \tag{8}$$

The infinite-fold, infinite-aperture (continuous integral) form of the rho filter is

$$\mathbf{R} = \rho(\omega) = \sqrt{i\omega} \, . \tag{9}$$

The finite-fold, finite-aperture form of the rho filter may be intuitively guessed as:

$$\rho(\omega) \approx \frac{1}{f^{\alpha}} + (i\omega)^{\frac{1}{2+\beta\frac{h_{min}}{h_{max}}}} \ , \tag{10}$$

where α and β are parameters to be determined. The integral and finite rho filter spectra are sketched in Figure **??**. In the case of infinite aperture but finite fold, the rho filter reduces to

../david2/./Figs/rho-filters.ps

Figure 8: Integral and finite rho filter spectra. The integral form has a notch at zero frequency and overgains high frequencies, compared to the finite-fold finite-aperture rho filter.

the familiar form with the addition of a DC scale component related to the fold:

$$\lim_{h_{max}\to\infty} \{\rho(\omega)\} = \frac{1}{f^{\alpha}} + \sqrt{i\omega} \, . \tag{11}$$

If, in addition to infinite aperture, the fold is also infinite, then the finite rho-filter reduces to the integral rho filter:

$$\lim_{h_{max}\to\infty \; f\to\infty} \{\rho(\omega)\} = \sqrt{i\omega} \, . \tag{12}$$

Practice

I found good parameter values to be $\alpha \approx 0.5$ and $\beta \approx 10$ such that the finite rho filter is approximately

$$\rho(\omega) \approx \frac{1}{\sqrt{f}} + (i\omega)^{1/3} . \tag{13}$$

I also shaped the top half of the spectral band with a taper proportional to $1-(\omega/\omega_{nyq})^2$. The spectra of the input and adjoint-modeled CMP gathers are shown to match well in Figure **??**. The adjoint-modeled CMP gather in Figure **??** is modeled with the optimal finite rho filter for comparison with the input data. In contrast, applying no rho filter leaves the adjoint-modeled data with too much low-frequency energy. Applying the integral rho filter $\sqrt{i\omega}$ cuts too much low frequency and adds too much high-frequency noise, as shown in Figures **??** and **??**.

`../david2/./Figs/specs.ps`

Figure 9: The amplitude spectrum of the input CMP gather (left) and its adjoint-modeled CMP spectrum (right). Instead of the infinite-aperture “rho” filter $(i\omega)^{1/2}$, I used a highcut-tapered $(i\omega)^{1/3}$ filter which seems to shape the aperture-limited spectrum of the adjoint-modeled data better.

CG convergence

I ran the CG algorithm for 10 iterations and compared the rho-filtered and non-rho-filtered weighted operator results. The plot of their respective convergence paths appears in Figure **??**. In the first iteration, the rho filter decreases the misfit error somewhat. Slow convergence with the rho filter was observed independently by Dave Nichols (personal communication).

../david2/./Figs/cmps-r01.ps

Figure 10: Adjoint-modeled CMP gathers with no rho filter (left), and with the integral rho filter (right). No rho filter gives the adjoint-modeled data too much low-frequency energy, whereas the integral rho filter loses some low-frequency energy and overgains high-frequency noise.

../david2/./Figs/specs-r01.ps

Figure 11: Amplitude spectra of adjoint-modeled CMP gathers with no rho filter (left), and with the integral rho filter (right). An additional highcut taper was applied as in the optimal rho filter example.

However, after a few iterations, the convergence with the rho filter plateaus, and then counterintuitively *diverges*. According to CG theory, a CG iteration should never diverge if the operator is linear and the gradient is nonzero. What has gone wrong? Figure **??** shows a convergence comparison for powers of $(i\omega)$ in equal increments from zero to $1/3$. Note that the divergence phenomenon occurs at later iterations for lower powers of $(i\omega)$. I obtained a similar response regardless of whether the rho-filter operator was implemented in the time or frequency domains, using both conjugate gradients and the steepest descent method. I tried double precision computation of the gradient dot-products, and computing the residual fresh (instead of merely updating it) within each iteration, but still observed this divergence behavior. The complete rho-filtered, weighted hyperbolic velocity-stack operator passes the dot-product test to within six significant digits. The best-fitting adjoint-modeled CMP gather after 10 rho-filtered CG iterations is compared with the input CMP gather in Figure **??**, and their spectra are compared in Figure **??**. Note the large gain of high-frequency noise energy in the data spectrum, which would have been more severe had I not been using the high-cut taper as described previously. Clearly, cascaded application of a rho-filter not only slows CG convergence, but apparently has the potential to make conjugate-gradient iterative solvers highly unstable. My hypothesis is that a strong nonlinearity is introduced into the operator by numerical noise. Round-off noise injects small energy into the high-frequency part of the spectrum. Normally, this noise grows linearly with each iteration and doesn't become a problem until the gradient nears zero. However, the rho-filter gains this high-frequency numerical noise at each iteration, and can significantly magnify its energy after a few iterations. An operator which introduces significant energy into a passband that initially contained zero energy is a nonlinear operator. The conjugate gradient technique may diverge in this nonlinear situation. If this hypothesis is correct, then spectral preconditioning that gains high-frequency numerical noise too quickly can lead to an unstable and divergent CG solution.

../david2/./Figs/iter-r10-ann.ps

Figure 12: Comparison of CG convergence between the weighted non-rho-filtered (lower curve) and rho-filtered (top curve) velocity-stack operator. Note that the rho-filtered operator diverges!

../david2/./Figs/iter-rn.ps

Figure 13: Comparison of CG convergence for rho-filtered operators. The curves range in powers of $(i\omega)$, starting from zero (lower) to $1/3$ (top) in equal increments of powers of $(i\omega)$.

../david2/./Figs/cmp-r10.ps

Figure 14: The input CMP gather (left) and the rho-filtered adjoint-modeled CMP gather (right) after 10 iterations of conjugate gradients.

../david2/./Figs/specs-r10.ps

Figure 15: Amplitude spectra of the input CMP gather (left) and the rho-filtered adjoint-modeled CMP gather (right) after 10 iterations of conjugate gradients. Note the unstable gain in high-frequencies, except where artificially damped by the highcut taper.

DISCUSSION

Based on this study, the best procedure for conjugate gradient least-squares velocity-stack inversion using a pseudounitary operator is as follows. Offset and time-variant operator weighting is helpful to speed convergence when applied at each CG iteration. Application of a rho filter can help the first few iterations, but can slow convergence, or worse cause divergence, in subsequent iterations. I recommend applying offset- and time-variant pseudounitary weighting to get good multiple and primary moveout velocity separation, but no rho filter during CG iterations for the best result. Figures **??** and **??** show the best results of L_2 velocity-stack inversion in my pseudounitary tests, and are directly comparable to the L_1 and L_2 results of Nichols (**?**). Note the primary and multiple reflection trends are fairly well separated in velocity space, and the residual error is small. I used ten CG iterations and was able to fit 94% of the input data energy, as shown in the residual plot of Figure **??**. Further iterations invert too many small eigenvalues of the operator and tend to smear the velocity stack and modeled CMP data. All operations for the 10 CG iterations, optional rho filtering, and forward modeling of the best fitting CMP gather take less than 60 CPU seconds per CMP location on an HP 750 workstation in Fortran 77. This approach provides a good candidate for true-amplitude multiple suppression as discussed by (**?**).

../david2/./Figs/cmp-vscan.ps

Figure 16: The input CMP gather (left), and the least-squares velocity scan inversion (right). Note the primary and multiple velocity trends are fairly well separated.

../david2/./Figs/adj-res.ps

Figure 17: The adjoint-modeled CMP gather (left) obtained from the L_2 velocity scan inversion, and the residual data plotted at the same scale. The modeled data fit 94% of the energy of the input CMP gather.

CONCLUSIONS

I have discussed a methodology for making a time-domain hyperbolic velocity-stack operator be pseudounitary. This involved offset and time-variable operator weighting, and the application of a finite-aperture finite-fold rho filter. Conjugate gradient tests showed that the operator weighting increased CG convergence in the first few iterations, but had less impact after several subsequent iterations. However, pseudounitary weighting gave a pleasing separation of primary and multiple reflection energy in velocity-stack model space. Application of a rho filter is a good spectral preconditioner to the first gradient step, but can slow convergence and even cause divergence after a few iterations. I hypothesized that rho-filter divergence was caused by nonlinear cascaded gaining of high-frequency numerical noise from iteration to iteration. If correct, the implication is that spectral preconditioners which gain high-frequencies too fast may result in divergent and unstable CG solutions.

ACKNOWLEDGMENTS

I am grateful for many productive discussions with Dave Nichols and Jon Claerbout on this topic.

REFERENCES

Hestenes, M. R., and Steifel, E., 1952, Methods of conjugate gradients for solving linear systems: J. Res. Nat. Bur. Stand., **49**, 409–436.