Source synthesis for waveform inversion

William Symes, Rice University*

ABSTRACT

Seismologists have long used extended sources, both physical and synthetic, to enhance the effectiveness of data acquisition and processing in various ways. Recently, extended synthetic sources have been suggested as a means to substantially reduce the computational complexity of waveform inversion. Most proposed synthetic sources for waveform inversion are random in some way. In contrast, I suggest a deterministic choice of extended source, one that maximizes the difference between the data predicted by the current inversion iterate and the data synthesized from the target point source records. Optimal source synthesis in this sense has proven effective in several non-seismic inverse problems. I will describe the concept, and present some preliminary evidence concerning its performance in inversion of reflection seismograms.

INTRODUCTION

Construction of data from extended sources by superposition of data from impulsive localized sources, either physically in the field or synthetically in data processing, has a long history in reflection seismology. Simultaneous recording of signals from more than one shot can reduce acquisition costs and enhance source sampling and aperture (Beasley et al., 1998; Berkhout, 2008; Blaquiere et al., 2009). Migration with simultaneous sources can reduce the cost of shot-record imaging by the number of sources taken together (Romero et al., 2000; Verschuur and Berkhout, 2009). Plane wave migration is a special case of migration with extended sources, with an even longer history (Treitel et al., 1982).

Recent intense interest in full waveform inversion has led several groups to propose the use of simultaneous or extended synthetic sources in that context. For example, Krebs et al. (2009) propose synthesis of all shots in a survey (or at least many of them, see the caveat below) by filtered stack, with a different filter for each shot. In common with many other recent source synthesis approaches, Krebs et al. (2009) choose the synthesis filters randomly. They find that random ± 1 length-one filters, that is, weighted stack with no dephasing, work as well as more complicated random filters with phase delays. They also note that convergence of the iterative time-domain least-squares method used in their work is much improved when the weights are chosen anew for each iteration. Synthesis of all shots in this fashion dramatically decreases the inversion workload, essentially by the number of shots combined in the synthesis.

This note proposes an alternative, deterministic method for source synthesis in waveform inversion. In common with Krebs et al. (2009), the method presented here updates the synthesis filters at each model update. Also in common with Krebs et al. (2009), the synthesis filters may have length 1, that is, parametrize a simple weighted stack, although more complex filters with phase delays could also be employed. In contrast to other synthesis

methods, however, this one is entirely deterministic. I will explain the concept, and present some preliminary evidence that it might have some promise.

The principle underlying the deterministic choice of synthesis filter is simple: amongst all admissible synthesis filters, choose the one which makes the predicted data at the current model differ *as much as possible* from the corresponding synthesized target data. Since the difference of the two sets of traces is a linear function of the filter coefficients, this maximization problem is equivalent to finding eigenvector with largest eigenvalue of a certain operator, identified below. Various numerical methods can be employed to approximate this eigenvector, the simplest being the power method.

The author learned this idea through the work on D. Isaacson on electrical impedance tomography, a biomedical inverse problem similar to resistivity imaging (Isaacson, 1987). Essentially the same idea has been used to good effect in ultrasonic and microwave array processing, synthetic aperture radar, underwater acoustics, and other imaging technologies (Borcea et al., 2007; Fink and Prada, 2001). In biomedical EIT, the current-voltage relation amongst a set of electrodes is measured. This relation depends on the conductivity interior to the body to which the electrodes were attached, but the dependence is very weak. Only a few current patterns produce voltage responses capable of distinguishing interior conductivity distributions, therefore it is extremely important to find these optimal patterns. The mathematical setup is the same: the voltage response is linear in the applied currents at the various electrodes, and the best pattern can be found - without knowing the internal conductivity structure - by solving an eigenvalue problem.

This deterministic source synthesis method shares with its random cousins a potential drawback: its formulation requires fixed spread acquisition geometry, at least in principle. While this restriction can be relaxed to some extent, it appears to constitute an important limitation.

The theory section immediately following this introduction describes a simple theoretical framework for the optimal source synthesis algorithm. A few examples of source synthesis follow, based on the Marmousi model (Bourgeois et al., 1991). I will end with some comments on the work remaining to fully evaluate this approach, prospects for relaxing the fixed spread limitation, and a perspective on its meaning for waveform inversion.

THEORY

For the sake of concreteness, seismic modeling in this paper will be acoustic. Linear acoustics relates the pressure p and velocity $\mathbf{v}$ fields by

$$\frac{1}{\kappa(\mathbf{x})} \frac{\partial}{\partial t} p(\mathbf{x}, t; \mathbf{x}_s) = -\nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s) + s(\mathbf{x}, t; \mathbf{x}_s), \quad (1)$$

$$\rho(\mathbf{x}) \frac{\partial}{\partial t} \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s) = -\nabla p(\mathbf{x}, t; \mathbf{x}_s). \quad (2)$$

The bulk modulus κ and material density ρ are functions of position $\mathbf{x}$. The right-hand side s in the first equation represents the source of acoustic energy as a constitutive law defect. The parameter $\mathbf{x}_s$ represents source location, and $s(\mathbf{x}, t; \mathbf{x}_s)$ is presumed to be localized at $\mathbf{x} = \mathbf{x}_s$. In the numerical experiments to be reported below, I used an approximation to the isotropic point radiator: $s(\mathbf{x}, t; \mathbf{x}_s) = w(t)\delta(\mathbf{x} - \mathbf{x}_s)$, in which source wavelet $w(t)$ is a

suitable signature. Since the right-hand side in equation 1 depends on $\mathbf{x}_s$, so do the solution fields p , $\mathbf{v}$.

Define the point-source predicted data is simply the collection of time traces $\{p(\mathbf{x}_r, t; \mathbf{x}_s)\}$ as the receiver locations $\mathbf{x}_r$ ranges over the receivers active for the source at $\mathbf{x}_s$. From now on assume a *fixed spread*: that is, the *same* receiver locations, hence data traces, are available for all sources.

Denote by $f(t; \mathbf{x}_s)$ a vector of filters, one for each source position $\mathbf{x}_s$. Since the equations 1, 2 are a linear, the synthesized data

$$p_f(\mathbf{x}_r, t) = \sum_{\mathbf{x}_s} \int d\tau f(t - \tau; \mathbf{x}_s) p(\mathbf{x}_r, \tau; \mathbf{x}_s)$$

is the data obtained by solving the wave equations 1, 2 with the right-hand side of 1 replaced by the extended source

$$s_f(\mathbf{x}, t) = \sum_{\mathbf{x}_s} \int d\tau f(t - \tau; \mathbf{x}_s) s(\mathbf{x}, \tau; \mathbf{x}_s) \quad (3)$$

To emphasize: the data p_f is the result of solving the wave equations for *one* right-hand side s_f .

Assuming the point source suite $\{s(\mathbf{x}, t; \mathbf{x}_s)\}$ known and fixed, the synthesized data p_f is linear in f , and depends on the model vector $m = (\kappa, \rho)$. To emphasize the linear dependence on f , write

$$p_f = \Lambda[m]f \quad (4)$$

and so define the linear operator $\Lambda[m]$.

Given a set of target data traces $\{d(\mathbf{x}_r, t; \mathbf{x}_s)\}$ for the same source and receiver positions, supposed to be responses to the point source suite $\{s(\mathbf{x}, t; \mathbf{x}_s)\}$, it's also possible to synthesize the data that would have been acquired with source s_f :

$$d_f(\mathbf{x}_r, t) = \Lambda^d f(\mathbf{x}_r, t) = \sum_{\mathbf{x}_s} \int d\tau f(t - \tau; \mathbf{x}_s) d(\mathbf{x}_r, \tau; \mathbf{x}_s). \quad (5)$$

If the model m predicts point source data close to the target data d , then the same must be true of the extended source data: that is, if

$$d(\mathbf{x}_r, t; \mathbf{x}_s) \simeq p(\mathbf{x}_r, t; \mathbf{x}_s), \text{ all } \mathbf{x}_r, t, \mathbf{x}_s$$

then

$$\Lambda^d f \simeq \Lambda[m]f$$

for any filter vector f , in which the last approximate inequality is to be interpreted as meaning that the two sides differ by a small residual relative to the size of the filter vector f .

Looking at this slightly differently, m is *not* a satisfactory model to explain the target data if for *any* filter vector f , the predicted extended data $\Lambda[m]f$ differs significantly from that synthesized from the target data, that is, $\Lambda^d f$, relative to some measure of the size of f .

Use the RMS to measure the size of both residual and filter vector:

$$E(f) = \sum_{\mathbf{x}_s} \int dt f(t; \mathbf{x}_s)^2; \quad (6)$$

$$E(\Lambda[m]f - \Lambda^d f) = \sum_{\mathbf{x}_r} \int dt (p_f(\mathbf{x}_r, t) - d_f(\mathbf{x}_r, t))^2. \quad (7)$$

As is well known (Golub and Loan, 1989), for given m (the current model estimate), the maximum value of $E(\Lambda[m]f - \Lambda^d f)/E(f)$ over all f is the largest *singular value* of the operator $\Lambda[m] - \Lambda^d$, and the “worst” f is the corresponding singular vector. Alternatively, the “worst” f is the eigenvector with largest eigenvalue of the symmetric operator

$$N[m, d] = (\Lambda[m] - \Lambda^d)^T (\Lambda[m] - \Lambda^d), \quad (8)$$

in which the superscript T denotes the transpose or adjoint.

These observations suggest an algorithm: for each model update m , find the “worst” filter vector f by estimating the largest eigenvalue of $N[m, d]$. Then use this eigenvector f as source filter, and update m to reduce $E(\Lambda[m]f - \Lambda^d f)$. It should neither be necessary to determine the maximum eigenvalue of N or its eigenvector precisely in the first step, nor to drive $E(\Lambda[m]f - \Lambda^d f)$ to its minimum for fixed f ; substantial steps towards both goals should produce an effective iterative method.

The simplest method for updating an (eigenvector, eigenvalue) pair f, λ is the power method:

$$f^+ \leftarrow N[m, d]f \quad (9)$$

$$f \leftarrow f^+ / \sqrt{E(f^+)} \quad (10)$$

$$\lambda \leftarrow f^T N[m, d]f. \quad (11)$$

The proposed algorithm is:

- repeat until λ falls below tolerance
 1. fix m ; update (f, λ) using rules 9, 10, and 11 several times;
 2. fix f ; update m by reducing $E(\Lambda[m]f - \Lambda^d f)$ using a quasi-Newton or other suitable nonlinear optimization method.

A FEW PRELIMINARY EXAMPLES

This section presents an exploration of the first half of the algorithm just presented, that is, the model m will not be updated. The examples show the ability of the power method to create extended sources that highlight the difference between predicted and target data, at the cost of a few simulations. We use only length-one filters after the fashion of Krebs et al. (2009), that is weighted stacks. The filter vector in that case is simply a function of source position $f(\mathbf{x}_s)$ giving the stack weights.

The example is based on the Marmousi constant-density acoustic model, with fixed spread geometry:

- 96 shots, $\Delta x_s = 80$ m, $x_{s,\min} = 800$ m, $z_s = 4$ m
- 381 receivers (fixed), $\Delta x_r = 20$ m, $x_{r,\min} = 200$ m, $z_r = 8$ m
- $nt = 626$, $dt = 4$ ms
- source wavelet $w(t) = \text{Ricker wavelet}$, peak frequency 25 Hz
- high order finite difference implementation of $\Lambda[m]$
- $\Lambda[m]^T = R\Lambda[m]R$, in which R is the time-reversal operator: $Rd(t) = d(-t)$ for any time series $d(t)$.

The model m used as the current iterate is the homogeneous aqueous half-space, that is, $\kappa = 2.25$ MPa, $\rho = 1.00$ kg/m³. Corresponding data gathers contain only the direct wave.

Figure 2 shows data stacked with uniform weight, that is, $\Lambda[m]f - \Lambda^d f$ with $f(\mathbf{x}_s) \equiv 1$. This is the difference of responses to a normally incident plane wave. Ten iterations of the power method increases the Rayleigh quotient (right-hand side of 11 from 823 to 1413, or nearly a factor of 2. The strongest signal in the data is the near-surface fault-block reflection near $x = 5500$ m, and the power method shifts most of the energy in the source to that region, in order to most effectively reveal the difference between Marmousi and homogeneous models (Figure ??).

On the other hand, starting with a source synthesized with random ± 1 weights yields a similar improvement: the Rayleigh quotient is 802 at the start, 1375 at the end, and very similar stack. Once again the near-optimal source created by the power method is quite different than the random uniform distribution, focusing most of its energy on the near-surface fault block signal.

To avoid the focus on the near-surface reflections, which would presumably be accounted for early in a model updating process, one can simply mute the data to a second or so. Similar results obtain: starting with either uniform or random weights, the power method was able to substantially upgrade the ability of the synthetic extended source data to distinguish between the homogeneous medium and target model. As was the case for the unmuted data, the plane-wave weights were actually somewhat more efficient to begin with than the random ± 1 weights.

CONCLUSION

Optimization of the difference energy gives a deterministic method for selecting source synthesis filters. The very preliminary examples presented in the last section limited the filters to length-one weight vectors (Krebs et al., 2009), but suggested that systematic search for good eigenvectors of the operator N may yield sources that ferret out the important differences between models more effectively than do ad hoc choices.

The ultimate effectiveness of this algorithm remains to be seen. The additional degrees of freedom in longer filters may lead to substantially better distinguishability. The filter update must be coupled with a model updating algorithm. The power method is simple but not as efficient as variants of the Lanczos process, for instance. Since every iteration

costs two simulations (albeit in all cases with a *single source!*), faster convergence is highly desirable. Finally, a variety of obvious algorithmic details remain to be pinned down.

The main limitation of this algorithm, shared with many others (Krebs et al., 2009; Verschuur and Berkhout, 2009) is the fixed spread requirement. To some extent one can apply these techniques to subsets of traces in a towed streamer survey, for example, that share a common set of sources, but it remains to be seen precisely how effective these generalizations might be. Ocean bottom node data may be a particular attractive target for algorithms like those sketched here.

ACKNOWLEDGMENTS

I am grateful to the sponsors of The Rice Inversion Project for their long-term support.

Figure 1: Shot gathers for $sx = 800, 1600, 2400, \dots, 8000$ m ./. marmfsdec

REFERENCES

- Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous sources: 65th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 133–135.
- Berkhout, G., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.
- Blaquiere, G., G. Berkhout, and E. Verschuur, 2009, Survey design for blended acquisition: 80th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, XX–XX.
- Borcea, L., G. Papanicolaou, and C. Tsogka, 2007, Optimal waveform design for array imaging: Inverse Problems, **23**, 1973–2020.
- Bourgeois, A., P. Lailly, and R. Versteeg, 1991, The Marmousi model, *in* The Marmousi Experience: IFP/Technip.

Figure 2: Difference of Marmousi, homogeneous medium responses for normally incident plane wave (length-1 filter with uniform weights), Rayleigh quotient = 823. `./nipwstack`

Figure 3: Response to synthetic source resulting from 10 iterations of power method, starting with normal incidence plane wave source, Rayleigh quotient = 1413. `./pow1stk`

Figure 4: Weights computed by 10 iterations of power method, starting with uniform weights (normal incidence plane wave source). `./pow1wts`

Figure 5: Differences of Marmousi, homogeneous medium responses, length-1 filter source with random ± 1 weights, Rayleigh quotient = 802. `./randpm1stk`

Figure 6: Response to synthetic source resulting from 10 iterations of power method, starting with random ± 1 weights, Rayleigh quotient = 1375. `./pow2stk`

Figure 7: Weights computed by 10 iterations of power method, starting with random ± 1 weights. `./pow2wts`

- Fink, M., and C. Prada, 2001, Acoustic time-reversal mirrors: Inverse Problems, **17**, R1–R38.
- Golub, G. H., and C. F. V. Loan, 1989, Matrix computations, second ed.: The John Hopkins University Press.
- Isaacson, D., 1987, Distinguishability of conductivities by electric current computed tomography: IEEE Transactions on Medical Imaging, **MI-5**, 91–95.
- Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, WCC177–WCC188.
- Romero, A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.
- Treitel, S., P. Gutowski, and D. Wagner, 1982, Plane-wave decomposition of seismograms: Geophysics, **47**, 1375–1401.
- Verschuur, D. J., and G. Berkhout, 2009, Target-oriented, least-squares imaging of blended data: 80th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, XX–XX.