Full waveform inversion by source extension: why it works

William W. Symes, Rice University, Huiyi Chen and Susan E. Minkoff, The University of Texas at Dallas*

SUMMARY

An extremely simple single-trace transmission example shows how an extended source formulation of full waveform inversion can produce an optimization problem without spurious local minima ("cycle skipping"). The data consist of a single trace recorded at a given distance from a point source. The velocity or slowness is presumed homogeneous, and the target source wavelet is presumed quasi-impulsive or focused at zero time lag. The source is extended by permitting energy to spread in time, and the spread is controlled by adding a weighted mean square of the extended source wavelet to the data misfit, to produce the extended inversion objective. The objective function and its gradient can be computed explicitly in this very simple example, and it is easily seen that all local minimizers must be within a wavelength of the correct slowness - that is, cycle-skipping cannot occur. Calculation of the gradient reveals that the minimization avoids cycle-skipping only if the weight operator is differential, a requirement shared with extended source inversion algorithms applicable at field scale.

INTRODUCTION

Full Waveform Inversion (FWI), or estimation of earth structure by model-driven least squares data fitting, is now well-established as a useful tool for probing the earth's subsurface (Virieux and Operto, 2009; Fichtner, 2010). However, so-called "cycle-skipping", the tendency of iterative FWI algorithms to stagnate at suboptimal and geologically uninformative earth models, still impedes its use. Because the computational size of field inversion tasks is very large, only iterative local (descent) minimization of the data misfit function is computationally feasible. However local descent methods avoid suboptimal stagnation only if initial models are already quite close to optimal, in the sense of predicting the arrival times of seismic events to within a small multiple of a dominant wavelength (Gauthier et al., 1986; Plessix et al., 2010).

This paper concerns one of the many ideas that have been advanced to overcome cycle-skipping, namely so-called extended inversion (Symes, 2008; Huang et al., 2019). "Extended" signifies that additional degrees of freedom are provided to the modeling process, in the hope of opening up more effective routes to geologically informative models with acceptable data fit. Since these extra degrees of freedom do not arise from modeling assumptions about the source or receiver acquisition geometry, they should be suppressed in the final solution. Extended inversion methods differ by the choice of additional degrees of freedom, and by choice of penalty applied to eliminate them in the final result.

Many of these extended inversion concepts sound plausible, and appear to work at least to some extent as one might hope from their heuristic justifications. However very few of these approaches have been underwritten by mathematical argument: in essence, they are mostly justified only "in the rear-view mirror", with no assurance that failure is not just around the corner, at the next example. On top of that, some of these approaches, for example those based on the computationally attractive Variable Projection Method ("VPM") of Golub and Pereyra (2003), are cast in such form that the reasons for success are not readily apparent.

This note shows exactly how VPM leads to successful velocity updates for an *extended source inversion* approach to a very simple inverse problem, which asks that a homogeneous velocity field be deduced from one trace at known offset. We put forward this inverse problem and extension-based solution not because there are not simpler ways of answering the question it poses - there certainly are - but because the formal ingredients of waveform-based velocity estimation in this very simple setting are common to many similar extended inversion algorithms, and because in this case every computation can be done analytically, nearly to completion. In particular, it becomes clear why the VPM gradient formula produces a constructive update, with no possibility of stagnation away from the global minimum.

In the following sections, we define the single-trace acoustic model, explain an extended source approach to inverting it, and compare with standard data-fitting FWI both by explicit computations of the gradients and by numerical illustration. We have suppressed some of the mathematical details; these can be found in (Symes, 2020a).

MODELING

Assume small amplitude (linearized) constant-density, acoustic wave propagation and an isotropic point source and receiver. Denote the slowness (reciprocal velocity) by m, which is assumed independent of spatial position. Let $f(t)$ be the time dependence of the point source ("wavelet") at location $\mathbf{x} = \mathbf{x}_s$. Then the pressure trace recorded at location $\mathbf{x} = \mathbf{x}_r$, at distance $r = |\mathbf{x}_r - \mathbf{x}_s|$ from the source position, is given by

$$S[m]f(t) = \frac{1}{4\pi r} f(t - mr). \tag{1}$$

in which $S[m]$ is the operator of convolution with the well-known acoustic 3D Green's function (Courant and Hilbert (1962), Chapter 5). Apart from amplitude scaling, the relation between input wavelet f and output trace $S[m]f$ is an m-dependent time shift. This time shift relation is the basis of many descriptions of the frequency-dependent cycle-skipping phenomenon (for example, Virieux and Operto (2009), Figure 7), so it is unsurprising that an analysis of cycle-skipping can be based on the simple modeling operator described above.

To make wavelet frequency content manifest, introduce a family $\{f_\lambda\}$ of wavelets indexed by λ, a parameter having dimen-

sions of time,

$$f_\lambda(t) = \frac{1}{\sqrt{\lambda}} f_1\left(\frac{t}{\lambda}\right). \tag{2}$$

The argument s of the "mother wavelet" f_1 is nondimensional. The only constraints placed on f_1 are that (i) $f_1(s) = 0$ for $|s| \geq 1$, and (ii) f_1 has positive mean-square, that is, does not vanish identically. Note that the scaling is such that the mean-square

$$\|f_\lambda\|^2 = \int dt\, |f_\lambda(t)|^2$$

is independent of λ.

We shall refer to λ as "wavelength": if f_1 has a dominant period of oscillation, then so does f_λ, and it is proportional to λ.

To this family of wavelets and a choice of target slowness m_* corresponds a family of noise-free data

$$d_\lambda = S[m_*] f_\lambda. \tag{3}$$

This family of data in turn defines a family of inverse problems, to which we now turn.

FWI

The preceding section provided all of the raw ingredients to define full waveform inversion for estimation of m from a single trace. It is only m that is to be determined: the $\lambda-$dependent family of wavelets $\{f_\lambda\}$ is regarded as known, along with the data family $\{d_\lambda\}$. The aim is to chose m to minimize

$$J_{\mathrm{FWI}}[m] = \frac{1}{2}\|S[m] f_\lambda - d_\lambda\|^2$$

$$= \frac{1}{32\pi^2 r^2} \int dt\, |f_\lambda(t - mr) - f_\lambda(t - m_* r)|^2. \tag{4}$$

The essence of the cycle-skip problem is clearly visible in the integral expression 4. Since $f_\lambda(t) = 0$ if $|t| > \lambda$, the shifted copies of f_λ under the integral sign do not overlap (that is, are not non-zero at the same times) if $|m - m_*| > 2\lambda/r$, so

$$|m - m_*| r > 2\lambda \;\Rightarrow\; J_{\mathrm{FWI}}[m] = \frac{1}{16\pi^2 r^2}\|f_1\|^2. \tag{5}$$

That is, if m differs from the optimal m_* by more than a multiple $(2/r)$ of wavelength (λ), then m is a local minimizer of J_{FWI} - that is, we have found a continuum of spurious local minimizers, not just a few.

For example, suppose that the mother wavelet f_1 is a 1 Hz (peak frequency) zero phase Ricker wavelet (which has negligible values for $|t| > 1$), and that the source-receiver distance $r = 1$ km. Figure 1 shows plots of J_{FWI} for peak frequencies $\lambda = 20, 40$ Hz and offset $r = 1$ km. The structure described above is clearly visible, along with a couple of local minima even closer to the global minimum at $m = m_*$.

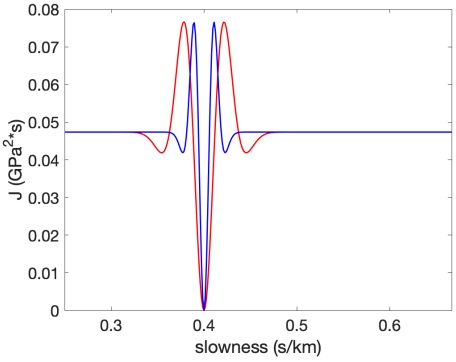

Figure 1: The FWI objective function plotted as a function of slowness for data from two sources: a 40 Hz Ricker (blue curve) and a 20 Hz Ricker (red curve).

EXTENDED FWI

The known-source modeling operator $m \mapsto S[m] f_\lambda$, may be extended simply by including the source wavelet as one of the model parameters: that is, the model vector becomes (m, f), and the modeling operator, $(m, f) \mapsto S[m] f$. The reader will have no trouble seeing that the data misfit using this extended modeling operator can always be made to vanish entirely by proper choice of wavelet f, unless f must satisfy some additional constraints. Huang et al. (2019) describe a plethora of possible constraints for this and similar source extensions. Many (but not all) take the form of a quadratic penalty, that is, the mean square of Af, A being a suitable operator, commonly dubbed an *annihilator*: in many examples the ideal output for A applied to a source obeying the target constraints is the zero vector (Symes, 2008). Thus the penalty form of extended inversion is: given data d, minimize over $\{m, f\}$

$$J^\alpha_{\mathrm{ESI}}[m, f; d] = \frac{1}{2}(\|S[m] f - d\|^2 + \alpha^2 \|Af\|^2). \tag{6}$$

The choices of the annihilator A and the weight α are crucial to the performance of extended inversion. We mention some possibilities for the choice of α in the discussion section. The choice of A depends on modeling assumptions, not fundamental physics, as is characteristic of extended source inversion. We have assumed that the data arises from a source wavelet active only over a short time interval - as $\lambda \to 0$, the target source wavelet f_λ focuses at time $t = 0$, in the sense that it vanishes for $|t| > \lambda$. Therefore we choose A to penalize energy away from $t = 0$:

$$Af(t) = t f(t). \tag{7}$$

This particular annihilator has been employed in earlier papers on extended source inversion (Plessix et al., 2000; Luo and Sava, 2011; Warner and Guasch, 2014; Huang and Symes, 2015; Warner and Guasch, 2016; Huang et al., 2017). Amongst many operators that could be used to penalize energy spread, this choice has a very important quality: it is *differential* of order zero. We will explain the importance of this characteristic in the Discussion section.

Gradient-based minimization of J^α_{ESI} performs poorly, as Huang (2016) has shown, because of dramatically different sensitiv-

ity to m versus f. Instead, a nested approach, in which f is eliminated in an inner optimization, generally gives far better numerical performance. This Variable Projection Method (VPM) (Golub and Pereyra, 2003) takes advantage of J^α_{ESI} being quadratic in f to solve for f given m, thus producing a reduced objective function of m alone:

$$J^\alpha_{\mathrm{VPM}}[m;d] = \min_f J^\alpha_{\mathrm{ESI}}[m,f;d] = J^\alpha_{\mathrm{ESI}}[m,f[m;d];d], \quad (8)$$

where $f[m;d]$ is the minimizer of $J^\alpha_{\mathrm{ESI}}[m,f;d]$ over f for given m. As is well-known, $f[m;d]$ solves the *normal equation*

$$(S[m]^T S[m] + \alpha^2 A^T A) f = S[m]^T d \quad (9)$$

For the problem considered here, $f[m;d]$ is explicitly computable. Straightforward computations (Symes, 2020a) show that

$$S[m]^T d(t) = \frac{d(t+mr)}{4\pi r}, \quad (10)$$

$$S[m]^T S[m] f(t) = \frac{f(t)}{(4\pi r)^2}, \quad (11)$$

and $A^T A f(t) = t^2 f(t)$, so taking $d = d_\lambda$ as in equation 3,

$$f[m;d_\lambda] = \left(1 + (4\pi r)^2 \alpha^2 t^2\right)^{-1} f_\lambda\,(t + (m-m_*)r) \quad (12)$$

Explicit expressions for J^α_{VPM} and its gradient can be extracted by elementary means from the identity 12. Instead, we sketch a computation of the gradient that works for other extended inversion methods, with appropriate modifications. To begin with, the gradient of a VPM objective of the form 8 is given by the formula

$$\nabla J^\alpha_{\mathrm{VPM}}[m] = (D(S[m]f)_{f=f[m,d]})^*(S[m]f[m;d] - d). \quad (13)$$

This easily derived result is in some sense the main content of Golub and Pereyra (2003). In this formula, $D(S[m]f)$ is the derivative of the modeling operator $S[m]f$ with respect to m, and the superscript * signifies the adjoint of the linear map from model space to data space: $\delta m \mapsto D(S[m]f)\delta m$. The key to unlocking the meaning of the VPM gradient formula for this and similar problems is a remarkable *factorization identity*:

$$D(S[m]f)\delta m = S[m](Q[m]\delta m) f(t), \quad (14)$$

$$(Q[m]\delta m) f = -r\delta m \frac{df}{dt}. \quad (15)$$

That is, $Q[m]\delta m$ is a skew-adjoint operator depending linearly on δm. The relation defined by equations 14, 15 is a simple calculus identity, but similar factorizations have been established for much more complicated extended models (Symes, 2014; ten Kroode, 2014; Symes, 2015).

A straightforward calculation, detailed in (Symes, 2020a) yields an expression for the gradient 13 in terms of Q:

$$\delta m \cdot \nabla J^\alpha_{\mathrm{VPM}}[m] =$$

$$\frac{\alpha^2}{2} \int dt\, f[m;d](t)([(Q[m]\delta m), A^T A] f[m;d])(t) \quad (16)$$

Here, the symbol $[L,M]$ denotes the *commutator* of the operators L and M: $[L,M] = LM - ML$. Note that the annihilator A is explicitly present on the right-hand side of equation 16, whereas its role is hidden in the expression 13. Formulae similar to 16 hold for many other extended inversion methods.

In the present case, $A^T A$ amounts to multiplying by t^2, and Q is the scaled time derivative (equation 15). Together with the expression 12 for $f[m;d]$, these observations imply that

$$\nabla J^\alpha_{\mathrm{VPM}}[m] = -r\alpha^2 \int dt \frac{t(f_\lambda\,(t+(m-m_*)r)^2)}{\left(1+(4\pi r)^2\alpha^2 t^2\right)} \quad (17)$$

Since $f_\lambda(t) = 0$ if $|t| > \lambda$, it is easy to see that

- if $m > m_* + \lambda/r$, then $\nabla J^\alpha_{\mathrm{VPM}}[m] > 0$, and
- if $m < m_* - \lambda/r$, then $\nabla J^\alpha_{\mathrm{VPM}}[m] < 0$.

That is, J^α_{VPM} *has no local minima further than* $O(\lambda)$ *from the global minimum:* the gradient has the correct sign and slowness updates computed from it will be constructive, unless the slowness estimate is already "within a wavelength" of being correct.

Figure 2 shows the behavior described above for the same two examples described earlier for FWI, and presented in Figure 1: that is, data from 20 and 40 Hz Ricker wavelets (f_{20} and f_{40}), and source-receiver distance = 1 km. Note that these two curves are very similar, in dramatic contrast to the behavior shown in Figure 1. The annihilator output Af_{40} is considerably smaller than Af_{20}, which explains the sag towards zero of the minimum value for the higher frequency.

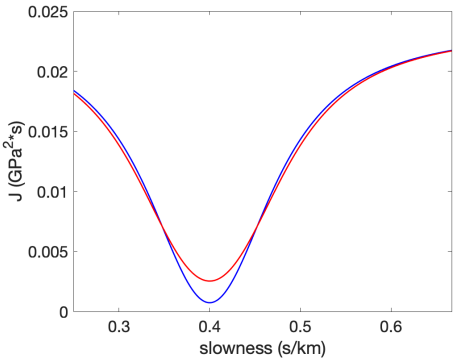

Figure 2: The VPM objective function plotted as a function of slowness for data from two sources: a 40 Hz Ricker (blue curve) and a 20 Hz Ricker (red curve).

The parameter α is set = 1 in this example. Since α is dimensional, this choice is truly arbitrary. The next section mentions an approach to rational choice of α.

DISCUSSION

We emphasize that factorizations similar to relation 14, and the resultant recasting of the VPM gradient similar to equation 16, are common features of extended inversion. Deeper analysis shows that this relation depends on approximate invertibility of the extended modeling operator (obvious for the simple problem considered here) and generic characteristics

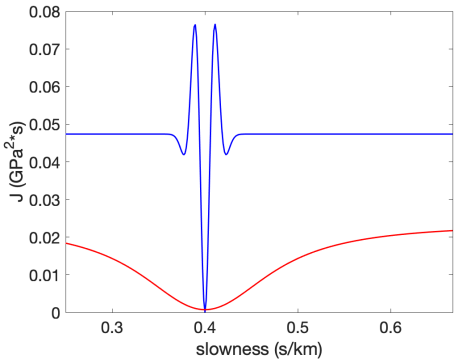

Figure 3: The FWI (blue curve) and VPM (red curve) objective functions plotted as a function of slowness for data from a 40 Hz Ricker wavelet.

of wave propagation. While the implications of equation 16 were easy to understand for the very simple setting of this paper, in other more complex and physically realistic settings the analogous result reveals that extended inversion is equivalent to some form of travel-time tomography (ten Kroode, 2014; Symes, 2014, 2015; Huang and Symes, 2015).

The penalty weight α has been "along for the ride" in the discussion above. In any practical inversion algorithm, some mechanism must be introduced to choose it. Fu and Symes (2017) describe a dynamic updating rule of α based on the *Discrepancy Principle*, in which the first (mean-square error) term of the extended objective 6 is kept within a range representing the expected error in the data. See the cited reference for an extended discussion and evidence of effectiveness.

Perhaps the most important general message implicit in the example presented here is that the choice of the annihilator A determines whether the extended inversion algorithm achieves global or semiglobal convexity, as is accomplished in the present example. (See Figure 3 for a comparison of the FWI and VPM objective functions as functions of slowness for data generated by a 40 Hz Ricker wavelet.) The annihilator used here has a property whose importance can be guessed by examining the gradient formula 16. The operator Q is a first order differential operator. The gradient is a quadratic form whose Hessian is the commutator $[Q, A^T A]$, and whose argument is $f[m;d]$. In order that the VPM objective be continuous for any model m and finite energy data d, this form should admit any finite-energy source as argument: in technical terms, it should be a bounded (or continuous) operator on the Hilbert space of finite energy traces. If one asks for a bit more, namely that the VPM objective function have derivatives of arbitrary order, then it is not too hard to see that the iterated commutators $[Q, ...[Q, A^T A]...]$ must all be bounded operators. This is a very strong restriction on $A^T A$: this operator must be *pseudodifferential*, that is, a combination of differential operators and powers of the Laplace operator (Taylor (1981), Chapter VIII, Lemma 5.3). For the present example, $A^T A$ is multiplication by t^2, a very simple differential (therefore pseudodifferential) operator. For more discussion of this requirement in the context of annihilators in extended inversion, see Symes (2008), where you can also find references to the technical backstory.

This constraint actually rules out some popular approaches to FWI. To begin with, basic least-squares FWI as formulated in the third section above can be reformulated as a quadratic form whose Hessian turns out not to be pseudodifferential. Therefore the FWI objective is not smooth in data and model jointly, a fact that is linked to the cycle-skipping behavior demonstrated above. More surprising, perhaps, the same turns out to be true for Wavefield Reconstruction Inversion (WRI), an extended source inversion algorithm introduced by van Leeuwen and Herrmann (2013), and further developed by van Leeuwen and Herrmann (2016), Wang et al. (2016), and Aghamiry et al. (2019), amongst others. This approach turns out to be closely linked to basic FWI, and can be formulated as minimization of a similar quadratic form: as in FWI, its Hessian is not pseudodifferential. Not coincidentally, it also exhibits cycle-skipping behavior. In simple cases such as the problem studied in this paper, WRI can be shown explicitly to have local minima far from the global minimum, and a region of attraction for the global minimum on the order of a wavelength in diameter, just as does FWI (Symes, 2020b).

CONCLUSION

Despite its simplicity, the single-trace transmission inversion problem proves typical of many more complex waveform inversion problems. The structure of the derivative is similar in many of these problems, and for the particularly simple one explained here, can be analyzed on paper to the point of showing explicitly why a simple extended source approach to waveform inversion works - that is, generates an objective all of whose local minima are "within a wavelength" of the global minimizer. Otherwise stated, this particular extended source inversion is genuinely immune to cycle-skipping. The simple structure of this problem showcases the importance of the variable projection reduction (elimination of the extended source) and a proper choice of annihilator in the formulation of the basic objective. It also makes clear the central role played by a factorization of the linearized modeling operator, a feature shared with many more complex extended source methods applicable at field scale.

ACKNOWLEDGEMENTS

This research is partially supported by the sponsors of the UT Dallas "3D+4D Seismic FWI" research consortium. WS thanks the sponsors of The Rice Inversion Project for their long-term support.

DOI:10.1190/segam2020-3424509.1

REFERENCES

Aghamiry, H., A. Gholami, and S. Operto, 2019, Improving full-waveform inversion by wavefield reconstruction with the alternating direction method of multipliers: Geophysics, **84**, R125–R148, doi: https://doi.org/10.1190/geo2018-0093.1.

Courant, R., and D. Hilbert, 1962, Methods of mathematical physics: Wiley-Interscience, II.

Fichtner, A., 2010, Full seismic waveform modelling and inversion: Springer Verlag.

Fu, L., and W. W. Symes, 2017, A discrepancy-based penalty method for extended waveform inversion: Geophysics, **82**, R287–R298, doi: https://doi.org/10.1190/geo2016-0326.1.

Gauthier, O., A. Tarantola, and J. Virieux, 1986, Twodimensional nonlinear inversion of seismic waveforms: Geophysics, **51**, 1387–1403, doi: https://doi.org/10.1190/1.1442188.

Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26, doi: https://doi.org/10.1088/0266-5611/19/2/201.

Huang, G., R. Nammour, W. Symes, and M. Dollizal, 2019, Waveform inversion by source extension: 89th Annual International Meeting, SEG, Expanded Abstracts, 4761–4766, doi: https://doi.org/10.1190/segam2019-3216338.1.

Huang, G., R. Nammour, and W. W. Symes, 2017, Full waveform inversion via matched source extension: Geophysics, **82**, R153–R171, doi: https://doi.org/10.1190/geo2016-0301.1.

Huang, G., and W. W. Symes, 2015, Full waveform inversion via matched source extension: 85th Annual International Meeting, SEG, Expanded Abstracts, 1320–1325, doi: https://doi.org/10.1190/segam2015-5872566.1.

Huang, Y., 2016, Born waveform inversion in shot coordinate domain: PhD thesis, Rice University.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792, doi: https://doi.org/10.1190/1.3627773.

Plessix, R.-E., G. Baeten, J. W. de Maag, M. Klaassen, R. Zhang, and Z. Tao, 2010, Application of acoustic full waveform inversion to a low-frequency large-offset land data set: 81st Annual International Meeting, Expanded Abstracts, doi: https://doi.org/10.1190/1.3513930.

Plessix, R.-E., W. Mulder, and F. ten Kroode, 2000, Automatic cross-well tomography by semblance and differential semblance optimization: theory and gradient computation: Geophysical Prospecting, **48**, 913–935, doi: https://doi.org/10.1046/j.1365-2478.2000.00217.x.

Symes, W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790, doi: https://doi.org/10.1111/j.1365-2478.2008.00698.x.

Symes, W., 2014, Seismic inverse problems: recent developments in theory and practice: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 2–5.

Symes, W., 2015, Algorithmic aspects of extended waveform inversion: 76th Annual International Conference and Exhibition, Expanded Abstracts, EAGE, WS05–A01, doi: https://doi.org/10.3997/2214-4609.201413492.

Symes, W., 2020a, Full waveform inversion by source extension: Why it works. (arXiv:2003.12538).

Symes, W., 2020b, Wavefield reconstruction inversion: an example (arXiv).

Taylor, M., 1981, Pseudodifferential Operators: Princeton University Press.

Ten Kroode, F., 2014, A Lie group associated to seismic velocity estimation: Inverse Problems - from Theory to Application: Proceedings, Institute of Physics, 142–146.

van Leeuwen, T., and F. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667, doi: https://doi.org/10.1093/gji/ggt258.

van Leeuwen, T., and F. Herrmann, 2016, A penalty method for pde-constrained optimization in inverse problems: Inverse Problems, **32**, 1–26.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152, doi: https://doi.org/10.1190/1.3238367.

Wang, C., D. Yingst, Farmer, J. Leveille, 2016, Full-waveform inversion with the reconstructed wavefield method: 86th Annual International Meeting, SEG, Expanded Abstracts, 1237–1241, doi: https://doi.org/10.1190/segam2016-13870317.1.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, doi: https://doi.org/10.1190/segam2014-0371.1.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: theory: Geophysics, **81**, R429–R445, doi: https://doi.org/10.1190/geo2015-0387.1.