Full Waveform Inversion by Source Extension: Why it works

William. W. Symes [1]

ABSTRACT

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, and it is easily seen that all local minimizers must be within a wavelength of the correct slowness. The derivation shows several important features of all similar extended source algorithms. For example, nested optimization, with the source estimation in the inner optimization (variable projection method), is essential. The choice of the weight operator, controlling the extended source degrees of freedom, is critical: the choice presented here is a differential operator, and that property is crucial for production of an objective immune from cycle-skipping.

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 several ideas that have been advanced to overcome cycle-skipping, namely so-called extended inversion (Symes, 2008). "Extended" signifies that addional 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 extended degrees of freedom are not part of the basic physics chosen to model the data acquisition process, they should be suppressed in the eventual 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 a particular extended inversion approach to a very simple inverse problem, which asks that a homogeneous velocity field be deduced from one trace at known offset. I 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.

The extended inversion approach developed here uses a *source extension*, in which source parameters form the additional degrees of freedom. This type of extension presumes that the actual or target source is constrained in some way; the extended source is allowed to violate the constraint. For recent overview of source extension methods, see Huang et al. (2019). Source extension methods have computational complexity approximately the same as that of FWI, a signal advantage over the alterantive medium extension class described for instance in (Symes, 2008).

For the problem considered here, the source model amounts to a wavelet, and the target wavelet is assumed to be non-zero only in a short time interval (an approximate impulse, perhaps as the result of signature deconvolution). The extension consists in permitting energy to spread in time at intermediate iterations of the inversion. A simple penalty for energy spread (second moment of square amplitude) drives the extended source towards a focused source approximately satisfying the assumed constraint. Not all penalties are created equal: the penalty used here has the critical *(pseudo-)differential* attribute necessary for avoidance of cycle-skipping, as will be explained in the Discussion section.

I begin with a quick sketch of constant density acoustics, and describe the single-trace transmission inverse problem. To make the role of data frequency content clear, I introduce a family of noise-free data parametrized by wavelength. For completeness, I show how the standard FWI approach to this problem generates multiple local minima that will be found by any descent method unless the initial estimate predicts travel time from source to receiver with an error on the order of a wavelength. The next section describes the source extension objective, and the reduced objective produced by VPM. As VPM eliminates the extended source, this function depends only on the velocity, just as does the FWI objective. A nearly-explicit calculation of the VPM gradient shows that the only stationary points are “within a wavelength” of the correct velocity, used to build the data: that is, cycle-skipping cannot occur. The paper ends with a discussion of the parallels between the calculations presented here and the structure of other extended inversion methods applicable to field-scale velocity estimation, and the critical role that the differential nature of the extension penalty plays in the success of this and other extension methods.

PRELIMINARIES

Assume small amplitude (linearized) acoustic propagation, constant density, and isotropic point source and receiver. Denote by $m(\mathbf{x})$ the slowness (reciprocal velocity) at spatial position $\mathbf{x}$, $f(t)$ the time dependence of the point source ("wavelet") at location $\mathbf{x} = \mathbf{x}_s$. Then the (excess) pressure field $p(\mathbf{x}, t)$ obeys a scalar wave equation:

$$\begin{aligned} \left(m(\mathbf{x})^2 \frac{\partial^2 p}{\partial t^2} - \nabla^2 \right) p(\mathbf{x}, t) &= f(t)\delta(\mathbf{x} - \mathbf{x}_s) \\ p(\mathbf{x}, t) &= 0, t \ll 0 \end{aligned} \tag{1}$$

Suppose that a single trace is recorded, at distance $r > 0$ from the source position $\mathbf{x}_s$. The dominant information in a single trace is the transient signal time of arrival, constraining only the mean slowness in the region between source and receiver, so assume that the slowness is constant, that is, independent of position $\mathbf{x}$. The pressure field is simply the the source wavelet $f(t)$ convolved with the acoustic Green's function, for which an analytic expression is available in the constant m case (Courant and Hilbert, 1962):

$$p(\mathbf{x}, t) = \frac{1}{4\pi|\mathbf{x} - \mathbf{x}_s|} f\left(t - m|\mathbf{x} - \mathbf{x}_s|\right). \tag{2}$$

The receiver location $\mathbf{x}_r$ lies at distance r from the source location $\mathbf{x}_s$, that is, $|\mathbf{x}_r - \mathbf{x}_s| = r$. The predicted signal at $p(\mathbf{x}_r, t)$ depends nonlinearly on the slowness m and linearly on the source wavelet f. Therefore it is naturally represented as the action of a m-dependent linear operator $S[m]$ on f:

$$S[m]f(t) = p(\mathbf{x}_r, t) = \frac{1}{4\pi r} f\left(t - mr\right). \tag{3}$$

Ignoring amplitude, this map implements a m-dependent time shift. This time shift operator is the basis of many descriptions of the 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, which amounts essentially to a time shift. To make the link with wavelet frequency content manifest, I introduce a family $\{f_\lambda\}$ of wavelets indexed by λ, a parameter having dimensions of time,

$$f_\lambda(t) = \frac{1}{\sqrt{\lambda}} f_1\left(\frac{t}{\lambda}\right). \tag{4}$$

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.

I 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

$$d_\lambda = S[m_*]f_\lambda. \tag{5}$$

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

FULL WAVEFORM INVERSION

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

$$J_{\rm FWI}[m] = \frac{1}{2}\|S[m]f_\lambda - d_\lambda\|^2. \tag{6}$$

for all values of $\lambda > 0$.

Written out in detail, this objective function is

$$J_{\rm FWI}[m] = \frac{1}{32\pi^2 r^2}\int dt\ |f_\lambda\,(t-mr) - f_\lambda\,(t-m_* r)|^2$$

Since f_1 vanishes for $|t| > 1$, f_λ vanishes for $|t| > \lambda$, and $S[m]f_\lambda$ vanishes if $|t - mr| > \lambda$. So if $|mr - m_* r| = |m - m_*|r > 2\lambda$, then $|t - mr| + |t - m_* r| \ge |mr - m_* r| > 2\lambda$ so either $|t - mr| > \lambda$ or $|t - m_* r| > \lambda$, that is, either $S[m]f_\lambda(t) = 0$ or $S[m_*]f_\lambda(t) = 0$. Therefore $S[m]f_\lambda$ and $S[m_*]f_\lambda$ are orthogonal in the sense of the L^2 inner product:

$$|m - m_*|r > 2\lambda \ \Rightarrow\ \langle S[m]f_\lambda, S[m_*]f_\lambda\rangle = \int dt\, S[m]f_\lambda(t)S[m_*]f_\lambda(t) = 0 \tag{7}$$

But $d_\lambda = S[m_*]f_\lambda$, so this is the same as saying that d_λ is orthogonal to $S[m]f_\lambda$. So conclude that

$$|m - m_*|r > 2\lambda \ \Rightarrow\ J_{\rm FWI}[m] = \frac{1}{16\pi^2 r^2}\|f_1\|^2. \tag{8}$$

using the previously observed independence of $\|f_\lambda\|$ from λ.

That is, for slowness m in error by more than $2\lambda/r$ from the target slowness m_*, the FWI objective $J_{\rm FWI}$ is perfectly flat: all nearby values of m are local minima. Therefore the local exploration of the objective gives no useful information whatever about constructive search directions towards the global minimizer $m = m_*$, for which the objective value is of course $= 0$, if the initial estimate of m in error by an amount greater than a fixed multiple of λ. This is precisely the behaviour of FWI noted many times in the literature: the model must be known to "within a wavelength".

EXTENDED SOURCE INVERSION

As mentioned in the introduction, the modeling operator introduced in the last section, $m \mapsto S[m]f_\lambda$, 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). With an annihilator A, to be chosen below, the penalty form of extended inversion is: minimize over $\{m, f\}$

$$J_{\mathrm{ESI}}[m,f] = \frac{1}{2}(\|S[m]f - d_\lambda\|^2 + \alpha^2\|Af\|^2). \tag{9}$$

The choice of penalty weight α has a profound influence on the character of this optimization problem. For the moment, I will mandate only that $\alpha > 0$. A detailed discussion of methods for setting α is beyond the scope of this brief paper. In the Discussion section, I mention an effective method for choosing α to optimize convergence of iterative minimization, that applies (and has been applied) to much larger-scale inversion problems with features similar to those of this simple example.

While minimization of J_{ESI} might be tackled directly - by alternately minimizations between m and f, or by computing updates for m and f simultaneously - such joint mimization performs poorly, as Huang (2016) has shown. The reason for this poor performance is that J_{ESI} has dramatically different sensitivity to m versus f, especially for high frequency f, as the reader will see below. 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_{ESI} being quadratic in f to solve for f given m, thus producing a reduced objective of m alone:

$$J_{\mathrm{VPM}}[m] = \min_f J_{\mathrm{ESI}}[m,f] = J_{\mathrm{ESI}}[m, f[m]], \tag{10}$$

where $f[m]$ is the minimizer of J_{ESI} over f for given m. For the problem considered here, J_{VPM} is explicitly computable. First observe that apart from amplitude, $S[m]$ is unitary:

$$S[m]^T g(t) = \frac{1}{4\pi r} g\,(t + mr) \tag{11}$$

so

$$S[m]^T S[m] f(t) = \frac{1}{(4\pi r)^2} f(t). \tag{12}$$

Therefore the normal equation for the minimizer on the RHS of equation 10 is

$$\left(\frac{1}{(4\pi r)^2} I + \alpha^2 A^T A\right) f[m] = S[m]^T d_\lambda. \tag{13}$$

At this point I have to come clean about the actual choice of A. Recall that A is called an "annihilator", as it vanishes when applied to the target source field, at least in an idealized limit. Thus the choice of A depends on modeling assumptions, not fundamental physics, as is characteristic of extended source inversion. 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 I choose A to penalize energy away from $t = 0$:

$$Af(t) = tf(t). \tag{14}$$

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).

With this choice of A, the normal operator on the LHS of 13 is simply multiplication by a positive function of time, and can be inverted by inspection. Using the identity 11 for the adjoint operator, obtain

$$f[m](t) = \left(\frac{1}{(4\pi r)^2} + \alpha^2 t^2\right)^{-1} \frac{1}{4\pi r} d_\lambda\,(t + mr)$$

$$= \left(1 + (4\pi r)^2 \alpha^2 t^2\right)^{-1} f_\lambda\,(t + (m - m_*)r) \tag{15}$$

thanks to the definition of the data d_λ (equation 5). After a little algebra,

$$(S[m]f[m] - d)(t) = \frac{1}{4\pi r}[(1 + (4\pi r)^2 \alpha^2 (t - mr)^2)^{-1} - 1] f_\lambda(t - m_* r) \tag{16}$$

and

$$Af[m] = t\left(1 + (4\pi r)^2 \alpha^2 t^2\right)^{-1} f_\lambda\,(t + (m - m_*)r)\,. \tag{17}$$

So

$$J_{\mathrm{VPM}}[m] = \frac{1}{2}\int dt\, \left(\frac{1}{(4\pi r)^2}[(1 + (4\pi r)^2 \alpha^2 (t + (m_* - m)r)^2)^{-1} - 1]^2 + \right.$$

$$\left. \alpha^2 (t + (m_* - m)r)^2 \left(1 + (4\pi r)^2 \alpha^2 (t + (m_* - m)r)^2\right)^{-2}\right) |f_\lambda(t)|^2$$

$$= \frac{1}{2(4\pi r)^2}\int dt\,[1 - (1 + (4\pi r)^2 \alpha^2 (t + (m_* - m)r)^2)^{-1}]|f_\lambda(t)|^2 \tag{18}$$

The gradient of J_{VPM} can be extracted by elementary means from the identity 18, but instead we will derive it in using some important features that this inverse problem shares with other inverse wave problems with more complex physics. To begin with, the gradient of a VPM objective of the form 10 is given by the formula

$$\nabla J_{\mathrm{VPM}}[m] = (DS[m]f[m])^*(S[m]f[m] - d). \tag{19}$$

This easily derived result is in some sense the main content of (Golub and Pereyra, 2003). In this formula, $DS[m]f$ is the derivative of the modeling operator $S[m]f$ with respect to m, that is, with f held fixed. The adjoint $(DS[m]f)^*$ is the adjoint of the map from model space to data space:

$$(DS[m]f) : \delta m \mapsto (DS[m]\delta m)f$$

and is NOT the same as the adjoint denoted by $S[m]^T$, which is the source-space-to-data-space adjoint. That is, this adjoint make this relation true, for all $m, \delta m, f$, and d:

$$\delta m \cdot (DS[m]f)^* d = \int dt\,[(DS[m]\delta m)f](t) d(t) \tag{20}$$

On the left side is the dot product in model space - since model space is just 1D in this example, that's just the numerical product. On the right is the dot product in data space, in the idealized continuum limit.

Note that the VPM gradient formula is remarkable in two ways. First, it is exactly the same as the FWI gradient, that is, the gradient of J_{FWI} if you happen to insert the solution $f[m]$ of the normal equation for f. Second, it does not mention the annihilator A at all: its impact is locked up in $f[m]$.

The key to unlocking the meaning of the VPM gradient formula for this and similar problems is a remarkable feature of the derivative operator $DS[m]$: from the definition 3,

$$(DS[m]\delta m)f(t) = -\frac{\delta m}{4\pi}\frac{df}{dt}(t - mr) = S[m](Q[m]\delta m)f(t), \tag{21}$$

where

$$(Q[m]\delta m)f = -r\delta m\frac{df}{dt}. \tag{22}$$

That is, $Q[m]\delta m$ is a skew-adjoint operator depending linearly on δm - more on this below.

A calculation, detailed in Appendix A, yields an expression for the gradient 19 in terms of Q:

$$\delta m \cdot \nabla J_{\rm VPM}[m] = \frac{1}{2}\alpha^2 \int dt\, f[m](t)[(Q[m]\delta m), A^T A]f[m](t) \tag{23}$$

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 in 23. The structure displayed in this expression is common to many other extended inversion methods. The extreme simplicity of the factorization 21, 22 and the gradient expression 23 is modified for more complex inversion problems by asymptotically negligible corrections (Symes, 2014; ten Kroode, 2014; Symes, 2015).

Remember that $A^T A$ amounts to multiplying by t^2, and Q is the scaled time derivative (equation 22), so

$$[(Q[m]\delta m), A^T A] = -2r\delta m t \tag{24}$$

Insert this identity into equation 23 to obtain

$$\delta m \cdot \nabla J_{\rm VPM}[m] = -r\delta m\alpha^2 \int dt\, t f[m]^2(t) \tag{25}$$

Combine this identity with the formula 15 for the solution of the inner problem, and divide out the common factor δm, to obtain

$$\nabla J_{\rm VPM}[m] = -r\alpha^2 \int dt\, t \left(\frac{1}{(4\pi r)^2} + \alpha^2 t^2\right)^{-2} \frac{1}{(4\pi r)^2} f_\lambda\left(t + (m - m_*)r\right)^2 \tag{26}$$

Recall that $f_1(s)$ vanishes if $|s| \geq 1$, so $f_\lambda(t + (m - m_*)r)$ vanishes if $|t + (m - m_*)r| > \lambda$. Therefore the integral on the RHS of equation 26 can be re-written

$$= -r\alpha^2 \int_{-(m-m_*)r-\lambda}^{-(m-m_*)r+\lambda} dt\, \frac{t}{(1 + (4\pi r)^2\alpha^2 t^2)^2} f_\lambda\left(t + (m - m_*)r\right)^2$$

If $m > m_* + \lambda/r$, then the entire interval of integration is a proper subset of the negative half-axis. Consequently the first factor in the integrand satisfies

$$\frac{t}{1 + (4\pi r)^2\alpha^2 t^2} \leq -(m - m_*)r + \lambda < 0$$

over the interval of integration. Consequently, the integral is negative (since f_λ has a positive mean square) and so the gradient is positive. Similar reasoning applies to the case $m < m_* - \lambda/r$.

To summarize,

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

That is, $J_{\rm VPM}$ *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.

On the other hand, careful examination of equation 26 shows that $J_{\rm VPM}$ is not convex: there is an inflection point $O(1/\alpha)$ from the global minimizer m_*, and in fact as $\alpha \to \infty$ for fixed λ, the $J_{\rm VPM}$ approximates $J_{\rm FWI}$.

DISCUSSION

There remain several important points to be made.

First, the formal computations centering on the operator Q (equations 21 through 23) depend only on the relation 21 and the skew-symmetry of Q, which hold with minor modifications for other more complex waveform inversion problems. These other more complex problems do not submit to such a simple treatment as is shown in the equations following 23, but for example it is possible in some cases to use a relation analogous to 21 to extract a relation between extended waveform inversion and traveltime tomography, via analysis of the Hessian at a zero-residual global minimizer. See for instance ten Kroode (2014); Symes (2014, 2015). Up to that point the reasoning is quite general, and central to the understanding gained so far of extended inversion methods.

Another issue mentioned in the last section is that the penalty weight α must be chosen somehow, and its choice influences heavily the convergence rate of iterative optimization algorithms applied to $J_{\rm VPM}$ (or $J_{\rm ESI}$, though (as also mentioned above) it is a poor candidate for local optimization). Two solutions to the weight assignment problem has emerged in the last few years, both quite effective. One applies directly to VPM problem, via use of the Discrepancy Principle (Morozov, 1984). This algorithm controls the size of the data misfit term in $J_{\rm ESI}$ to lie in an interval chosen to contain the assumed level of data noise, by adjusting α sporadically as the optimization proceeds. While this approach requires an estimate of data misfit to be achieved at the optimal model, it is very effective in maintaining convergence (Fu and Symes, 2017). Another approach modifies the VPM problem by means of the Augmented Lagrangian method (Nocedal and Wright, 1999). This reformulation appears to suppress much of the sensitivity to α of VPM optimization, and has been successfully used in extended inversion (Aghamiry et al., 2019).

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. The annihilator used here has a property whose importance can be guessed by examining the gradient formula 23. The operator Q is a first order differential operator. The gradient is a quadratic form whose Hessian is the communator $[Q, A^TA]$, and whose argument is $f[m]$. 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 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). For the present operator, $A^T A$ is multiplication by t^2, a very simple pseudodifferential operator. For more discussion of this requirement in the context of annihilators in extended inverseion, 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 behaviour 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 which once again has a non-pseudodifferential Hessian. Not coincidentally, it also exhibits cycle-skipping behaviour. Just as for FWI, 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, 2020).

CONCLUSION

Desipite 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 analysed 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.

APPENDIX A

CALCULATION OF THE VPM GRADIENT

Using equations 21, 22 and 20, and 19,

$$\delta m \cdot \nabla J_{\rm VPM}[m] = \int dt\, [(DS[m]\delta m) f[m]](t)(S[m]f[m] - d)(t)$$

$$= \int dt\, [S[m](Q[m]\delta m) f[m](t)](S[m]f[m]-d)(t) = \int dt\, (Q[m]\delta m) f[m](t) S[m]^T (S[m]f[m]-d)(t)$$

Now invoke the normal equation 13: and replace the last factor:

$$= -\alpha^2 \int dt\,(Q[m]\delta m) f[m](t) A^T A f[m](t) \tag{A-1}$$

Since $Q[m]\delta m$ is skew-symmetric, shift it onto the other factor in this L^2 inner product (why not):

$$= \alpha^2 \int dt\, f[m](t)(Q[m]\delta m) A^T A f[m](t)$$

$$= \alpha^2 \int dt\,(A^T A f[m](t)(Q[m]\delta m) f[m](t) + f[m](t)[(Q[m]\delta m), A^T A] f[m](t)) \tag{A-2}$$

where I swapped Q and $A^T A$ at the cost of introducing a term involving the commutator $[Q, A^T A] = QA^T A - A^T AQ$, and rearranged the first term using the symmetry of $A^T A$. Now notice that this first term is exactly the same as the RHS of equation A-1, except for the minus sign - so subtracting the RHS of A-2 from A-1 and rearranging, get

$$-\alpha^2 \int dt\,(Q[m]\delta m) f[m](t) A^T A f[m](t) = \frac{1}{2}\alpha^2 \int dt\, f[m](t)[(Q[m]\delta m), A^T A] f[m](t)$$

hence

$$\delta m \cdot \nabla J_{\mathrm{VPM}}[m] = \frac{1}{2}\alpha^2 \int dt\, f[m](t)[(Q[m]\delta m), A^T A] f[m](t) \tag{A-3}$$

which is identical to equation 23.

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**, no. 1, R125–R148.

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

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**, no. 5, R287–R298.

Gauthier, O., A. Tarantola, and J. Virieux, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Geophysics, **51**, 1387–1403.

Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26.

Huang, G., R. Nammour, and W. Symes, 2017, Full waveform inversion via matched source extension: Geophysics, **82**, R153–R171.

Huang, G., R. Nammour, W. Symes, and M. Dollizal, 2019, Waveform inversion by source extension: 89th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4761–4765.

Huang, G., and W. W. Symes, 2015, Full waveform inversion via matched source extension: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1320–1325.

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, Expanded Abstracts, Society of Exploration Geophysicists, 2788–2792.

Morozov, V., 1984, Methods for solving incorrectly posed problems: Springer Verlag.

Nocedal, J., and S. Wright, 1999, Numerical Optimization: Springer Verlag.

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, Society of Exploration Geophysicists, 930–934.

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.

Symes, W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, no. 6, 765–790.

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

———, 2015, Algorithmic aspects of extended waveform inversion: 76th Annual International Conference and Exhibition, Expanded Abstracts, European Association of Geoscientists and Engineers, WS05–A01.

———, 2020, Wavefield reconstruction inversion: an example. (arXiv:2003.14181).

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.

———, 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**, no. 6, WCC127–WCC152.

Wang, C., D. Yingst, P. Farmer, and J. Leveille, 2016, Full-waveform inversion with the reconstructed wavefield method: 86th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1237–1241.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1089–1093.

———, 2016, Adaptive waveform inversion: theory: Geophysics, **81**, R429–R445.