Solution of an Acoustic Transmission Inverse Problem by Extended Inversion: Theory

William W. Symes

ABSTRACT

A single-trace transmission inverse problem for the wave equation seeks to determine both the wave velocity in a homogenous acoustic medium and the transient waveform of an isotropic point source. The duration (support) of the source waveform and the source-to-receiver distance are assumed known. A least squares formulation of this problem exhibits the “cycle-skipping” behaviour observed in field scale problems of this type, with many local minima differing greatly from the global minimizer. This behaviour is eliminated by dropping the hard support constraint on the source waveform, replacing it by a soft penalty implemented as a weighted mean-square of the source waveform. For properly chosen weight function, penalizing nonzero values away from $t = 0$, the velocity component of any stationary point differs from the global minimizer of the constrained least-squares formulation by a linear combination of the source waveform support radius and data noise-to-signal ratio. Given an estimate of data noise, the penalty weight can be dynamically adjusted during iterative optimization to maximize predicted data accuracy and closely approximate a support-constrained source waveform.

INTRODUCTION

Inverse problems based on wave equations (acoustic, elastic, Maxwell’s...) are commonplace in geophysics, nondestructive materials testing, medical imaging, and other areas of science and engineering in which wave motion plays an important role. Often these problems are formulated as optimization of mean square (or other scalar measure) functions of misfit between observed and predicted data. In many contexts, such as exploration seismology, the computational size of these problems is very large, and even the prediction (or modeling) of the data requires use of high-performance computing environments. Accordingly, rapidly convergent optimization algorithms based on Newton’s method and its relatives to update model parameters are the numerical methods of choice. These however are local, in the sense that they converge to stationary points of the objective functions, usually near the initial model parameters. For high-frequency (hence high resolution) data regimes, stagnation of local optimization methods at physically irrelevant model vectors is frequently observed. This “cycle-skipping” is one of the main obstacles to widespread successful deployment of optimization-based methods for inverse problems in wave propagation [Gauthier et al., 1986; Virieux and Operto, 2009; Fichtner, 2010; Plessix et al., 2010; Schuster, 2017].

This paper shows how a modification of least-squares data fitting via modeling operator extension overcomes cycle-skipping in a simple example. The problem studied here may be the simplest inverse wave problem that exhibits cycle-skipping - in fact, essentially this
example is routinely used to illustrate cycle-skipping (see for example Virieux and Operto (2009), Figure 7). Its data is a single trace at a point in $\mathbb{R}^3$ ("receiver location") of the causal linear acoustic pressure field, generated by an isotropic radiator acting at another point ("source location") at a known distance from the receiver location. The acoustic material is supposed spatially homogeneous, hence characterized by wave slowness (reciprocal velocity) and density. The object of the inverse problem is to recover the slowness and the time history of the isotropic radiator ("wavelet") from the recorded trace. In order that the data constrain the slowness, the wavelet is presumed to have known compact support. Thus the ("physical") modeling operator has as its domain the Cartesian product of an interval of positive slownesses, and the subspace of $L^2(\mathbb{R})$ consisting of wavelets with the prescribed support. Even for noise-free data (in the range of the modeling operator), the obvious least squares problem formulated in terms of this modeling operator exhibits many local minimizers having nothing to do with the ("target") slowness and wavelet used to create the data - that is, cycle-skipping occurs (Theorem 1).

Dropping the support constraint on the wavelet creates an extension of the modeling map for which any data may be fit exactly, for any choice of slowness. The link between data and slowness is restored by adding a quadratic penalty on non-zero values of the wavelet outside of the physical support interval to the extended mean-square data misfit. The first main result of this paper is that for noise-free data, the slowness component of any stationary point of this extended penalty function differs from the target slowness by a multiple of the target wavelet radius (Theorem 5). That is, cycle-skipping does not occur, and the wavelet support radius determines the resolution of slowness. For noisy data (that is, sum of a noise-free signal trace in the range of the physical modeling operator and a square-integrable noise trace), any stationary point differs from the target slowness by a combination of the wavelet support radius and the noise-to-signal ratio (Theorem 6). Since the wavelet component of the extended domain does not constrain support, stationary points of the extended penalty function do not provide solutions of the original, physical inverse problem, which includes the support constraint. Such solutions can be obtained, with estimates on support radius and data error, simply by truncating the wavelet components of stationary points (Theorem 7).

The inverse problem studied here is far too simple to have any direct use in applications. However it seems worthy of careful study for two reasons. First, its simplicity allows an unusually complete account of its properties. The present paper details its mathematical behaviour, including explicit and sharp error estimates rarely available for nonlinear inverse problems. A companion paper (Symes et al., 2021) describes numerical examples that illustrate the mathematical conclusions.

Second, the analysis reveals several important features shared with similar approaches to other wave inverse problems, and so provides an at least partial pattern for reaching similar conclusions with immediate practical implications. For example, nested optimization (variable projection method, Golub and Pereyra (2003)) based on a model decomposition into inner and outer variables seems to be very important: in some cases, such as the simple problem presented here, the decomposition is obvious (inner variable = wavelet, outer variable = slowness), in others less so (Symes, 2008; Terentyev, 2017). The extended modeling operator must be surjective, or at least have a dense range, for each value of the outer variable, so that data may be fit well even for a poor initial guess of the outer variable. for the extended inversion approach to be successful. The derivative of the extended modeling...
operator with respect to the outer variable is well-approximated by the composition of the extended modeling operator itself and a pseudodifferential operator of order 1 (Symes, 2014; ten Kroode, 2014). The choice of the penalty operator, controlling the extended degrees of freedom, is critical: in order to produce an objective immune from cycle-skipping, this penalty operator must be (pseudo-)differential (Stolk and Symes, 2003). Generally, the extended modeling operator is not differentiable in the outer variable. However these last two facts together imply the differentiability of the reduced objective function from variable projection (Theorem 3, Appendix C). Because of the data-fitting assumption, a straightforward algorithm for scaling the quadratic penalty, based on the Discrepancy Principle, is available - see Fu and Symes (2017); Symes et al. (2021), also Appendix A. This scaling varies dynamically during iterative optimization, in effect changing the objective function sporadically as the iteration converges.

The approach explained here is an example of extended source inversion, see Huang et al. (2019) for a review, and Métivier and Brossier (2020) for a recent variation. Adding degrees of freedom to the coefficients in the wave equation in various ways produces so-called extended model inversion approaches, which actually have a longer history in seismology - see Symes (2008) for details. The need to overcome cycle-skipping has spawned a great number of concepts, of which extension-based inversion is only one. Least-squares inversion succeeds with sufficiently good initial estimates of wave velocity, and the dominant techniques in industrial and academic seismology exploit data other than direct measurements of waves, for example measurement of wave time-of-flight, to supply such initial estimates (Virieux and Operto, 2009; Fichtner, 2010; Schuster, 2017). Use of a data misfit measure other than least squares, such as versions of the Wasserstein metric from optimal transport theory, also shows promise (Métivier et al., 2018; Yang et al., 2018).

The next section gives precise statements of the inverse problem studied here, and its various components. The following sections contain statements and proofs of the main results: the existence of multiple stationary points of the least squares problem, properties of the extended modeling operator and the reduced objective, choice of the penalty operator, the main results mentioned above about stationary points of the reduced objective, and the discrepancy-based algorithm for adjusting the penalty scale. Three appendices justify the discrepancy-based scale adjustment, review the link between constrained support and spectrum provided by the Heisenberg inequality, and expose the abstract structure underlying the differentiability of the reduced objective.

I use the abbreviations $\mathcal{B}(X,Y)$ and $\mathcal{I}(X,Y)$ for the algebra of bounded linear operators from the Hilbert space $X$ to the Hilbert space $Y$, and its subalgebra of invertible operators. All of the operators appearing in this discussion are members of $\mathcal{B}(X,Y)$ for some choice of Hilbert spaces $X$ and $Y$, and superscript $T$ denotes the transpose or adjoint of a such an operator in the sense specified by its domain and range Hilbert structure.
AN ACOUSTIC TRANSMISSION INVERSE PROBLEM

According to linear acoustics, the causal pressure field due to an isotropic point radiator at \( x_s \in \mathbb{R}^3 \) solves the wave equation (Friedlander, 1958):

\[
\left( m^2 \frac{\partial^2 p}{\partial t^2} - \nabla^2 \right) p(x, t) = w(t) \delta(x - x_s) \]

\[
p(x, t) = 0, \ t \ll 0.
\]

(1)

The solution is well-known, see for instance Courant and Hilbert (1962), Chapter VI, section 12, equation 47:

\[
p(x, t) = \frac{1}{4\pi|x - x_s|} w \left( t - m|x - x_s| \right),
\]

(2)

This trace of \( p \) at \( x_r \in \mathbb{R}^3 \) can be viewed as the result of applying an m-dependent linear operator \( F[m] \) to the wavelet \( w \):

\[
F[m]w(t) = p(x_r, t) = \frac{1}{4\pi r} w(t - mr)
\]

(3)

The slowness \( m \) must be positive, as follows from basic acoustics, and in fact reside in a range characteristic of the material model: for crustal rock, a reasonable choice would be \( m_{\text{min}} = 0.125, m_{\text{max}} = 0.6 \) s/km.

Natural choices for domain and range of \( F \) are thus

- \( M = (m_{\text{min}}, m_{\text{max}}), 0 < m_{\text{min}} \leq m_{\text{max}} \);
- \( W = L^2(\mathbb{R}) \);
- \( D = L^2([t_{\text{min}}, t_{\text{max}}]), t_{\text{min}} < t_{\text{max}} \);
- \( F: M \times W \rightarrow D \) as specified in 3.

It is immediately evident from these choices and from the definition 3 that

\[
\text{for } m \in M, F[m] \in \mathcal{B}(W, D), \text{ and } \|F[m]\| = \frac{1}{4\pi r}.
\]

(4)

Note also that \( F[m] \) is surjective for every \( m \in M \).

Remark: In computational practice, \( W \) will have to be replaced by a finite-dimensional subspace of \( L^2(\mathbb{R}) \). Many such choices will implicitly limit the support of \( w \in W \) to a bounded interval, say \( [T_{\text{min}}, T_{\text{max}}] \). To maintain the surjective property, these bounds should be chosen so that \( [t_{\text{min}}, t_{\text{max}}] \subset [T_{\text{min}} + mr, T_{\text{max}} + mr] \) for all \( m \in M \), that is,

\[
t_{\text{min}} \geq T_{\text{min}} + m_{\text{max}} r, \quad t_{\text{max}} \leq T_{\text{max}} + m_{\text{min}} r.
\]

(5)

I will ignore these computational necessities in this work, maintaining the definition \( W = L^2(\mathbb{R}) \).

As mentioned earlier, \( F[m] \) is surjective for every \( m \in M \). Since all possible data lie in the range of \( F[m] \) for any \( m \in M \), some restriction of the domain of \( F \) is necessary in order that fitting the data constrain \( m \). The constraint employed is the specification of a maximum support radius \( \lambda_{\text{max}} > 0 \). Then define for \( \lambda \in (0, \lambda_{\text{max}}] \):
In terms of this infrastructure, the inverse problem studied in this paper may be stated as

**Inverse Problem:** given data \( d \in D \), relative error level \( \epsilon \in [0, 1) \), and support radius \( \lambda \in (0, \lambda_{\text{max}}) \), find \( (m, w) \in M \times W_\lambda \) for which

\[
\| F_\lambda[m]w - d \| \leq \epsilon \| d \|,
\]

(6)

Define the relative mean-square error \( e_\lambda : M \times W_\lambda \times D \to \mathbb{R}^+ \) by

\[
e_\lambda[m, w; d] = \frac{1}{2} \frac{\| F_\lambda[m]w - d \|^2}{\| d \|^2},
\]

(7)

so that inequality (6) is equivalent to

\[
e_\lambda[m, w; d] \leq \frac{1}{2} \epsilon^2,
\]

(8)

Minimization of \( e_\lambda \) is the standard least-squares formulation of the Inverse Problem defined above: if the global minimum value of \( e_\lambda \) is less than \( \frac{1}{2} \epsilon^2 \), then the global minimizer is a solution of the Inverse Problem.

**Remark:** The constraint \( \epsilon < 1 \) imposed on the target noise level is eliminates the obvious choice \( (m, 0) \), which satisfies the data misfit constraint for any \( m \in M \).

**Remark:** I shall refer to the minimization of \( e_\lambda \) as “Full Waveform Inversion” or “FWI”, as this is the terminology used in the seismology literature to identify this and similar optimization problems.

**Remark:** The support constraint is closely linked to the folk theorem about FWI noted many times in the literature: convergence of a descent method requires that the initial slowness must be known to “within a (fraction of a) wavelength”. The relation is a consequence of Heisenberg’s inequality, and will be reviewed in Appendix B.

The best case for data fitting is clearly the one in which the data can be fit precisely: that is, there exists \( (m_*, w_*) \in M \times W_\lambda \) so that

\[
d = F_\lambda[m_*]w_*.
\]

(9)

Such data \( d \) is noise-free, in the range of the map \( F_\lambda \). For such data a solution of the Problem Statement (6) exists with arbitrarily small \( \epsilon > 0 \).

**FULL WAVEFORM INVERSION**

While \( F \) is surjective, as noted above, it is very far from injective. On the other hand, under a constraint that will be assumed throughout, \( F_\lambda[m] \) is injective for each \( m \in M \) (in fact, \( 4\pi r F_\lambda[m] \) is an isometry):
Proposition 1. Suppose that

\[ [m_{\min} r - \lambda_{\max}, m_{\max} r + \lambda_{\max}] \subset [t_{\min}, t_{\max}]. \]  \(10\)

Then \( F_{\lambda}[m] \in B(W_{\lambda}, D) \) is coercive for every \( m \in M, \lambda \in (0, \lambda_{\max}] \).

Remark: A useful consequence of the condition \(10\) for every \( m \in M, \)

\[ [-\lambda_{\max}, -\lambda_{\max}] \subset [t_{\min} - mr, t_{\max} - mr]. \]  \(11\)

The first main result establishes the existence of large (100 \%) residual local minimizers for the basic FWI objective \( e_\lambda \), even for noise-free data.

Theorem 1. Suppose that \( 0 < \lambda \leq \lambda_{\max}, m_* \in M, w_* \in W_{\lambda}, d = F_{\lambda}[m_*]w_* \) is noise-free data per definition \(9\). Under assumption \(10\) for any \( m \in M \) with \( |m - m_*| r > 2\lambda \),

\[ \min_w e_\lambda[m, w; d] = e_\lambda[m, 0; d] = \frac{1}{2}, \]  \(12\)

and any such \( (m, 0) \) is a local minimizer of \( e_\lambda \) with relative RMS error = 1.0.

Proof. From the definition \(3\)

\[ e_\lambda[m, w; d] = \frac{1}{32\pi^2 r^2 ||d||^2} \int_0^T dt \ |w(t - mr) - w_*(t - m_* r)|^2 \]

Since \( w_* \) vanish for \( |t| > \lambda, F_{\lambda}[m_*]w_* \) vanishes if \( |t - m_* r| > \lambda \) and \( F_{\lambda}[m]w \) vanishes if \( |t - mr| > \lambda \). So if \( |mr - m_* r| = |m - m_*| r > 2\lambda \), then \( |t - mr| + |t - m_* r| \geq |mr - m_* r| > 2\lambda \) so either \( |t - mr| > \lambda \) or \( |t - m_* r| > \lambda \), that is, either \( F_{\lambda}[m]w(t) = 0 \) or \( F_{\lambda}[m_*]w_* = 0 \). Therefore \( F_{\lambda}[m]w \) and \( F_{\lambda}[m_*]w_* \) are orthogonal in the sense of the \( L^2 \) inner product \( \langle \cdot, \cdot \rangle_D \) on \( D \):

\[ |m - m_*| r > 2\lambda \Rightarrow \langle F[m]w, F[m_*]w_* \rangle_D = 0 \]  \(13\)

But \( d = F_{\lambda}[m_*]w_* \), so this is the same as saying that \( d \) is orthogonal to \( F[m]w \). So conclude after a minor manipulation that

\[ |m - m_*| r > 2\lambda \Rightarrow e_\lambda[m, w; d] = \frac{1}{32\pi^2 r^2 ||d||^2} (\|w\|^2 + \|w_*\|^2). \]

\[ \frac{1}{2} \left( \frac{\|w\|^2}{\|w_*\|^2} + 1 \right) \]  \(14\)

That is, for slowness \( m \) in error by more than \( 2\lambda/r \) from the target slowness \( m_* \), the means square error (FWI objective) \( e_\lambda \) is independent of \( m \), and its minimum over \( w \) is attained for \( w = 0 \) \(\Box\)

Therefore local minimizers of \( e_\lambda \) abound, as far as you like from the global minimizer \((m_*, w_*)\). Local exploration of the FWI objective \( e \) gives no useful information whatever about constructive search directions, and descent-based optimization tends to fail if the initial estimate \( m_0 \) is in error by more than \( 2\lambda/r \) (“further than a multiple of a wavelength”, per discussion in the second Appendix). In fact the actual behaviour of FWI iterations is worse (failure if \( m_0 \) is in error by “half a wavelength”), as follows from a more refined analysis of the cycle-skipping local behaviour of \( e_\lambda \) near its global minimizer.
EXTENDED SOURCE INVERSION

The phenomenon explained in the last section can be avoided by reformulating the inverse problem via an extended modeling operator and a soft (penalty) constraint to replace the support requirement. The extension simply amounts to dropping the support constraint, and replacing $W_\lambda$ and $F_\lambda$ by $W$ and $F$.

The hard support constraint implicit in the choice of $W_\lambda$ as domain for the modeling operator is replaced by a soft constraint in the form of a quadratic penalty, with weight operator $A$:

$$Aw(t) = a(t)w(t), \ t \in \mathbb{R}. \quad (15)$$

Explicit choices for $a$ are discussed below.

With these choices, define

$$e[m, w; d] = \frac{1}{2} \| F[m]w - d \|^2 / \| d \|^2; \quad (16)$$
$$g[w; d] = \frac{1}{2} \| Aw \|^2 / \| d \|^2; \quad (17)$$
$$J_\alpha[m, w; d] = e[m, w; d] + \alpha^2 g[w; d]. \quad (18)$$

VARIABLE PROJECTION

The main theoretical device used in the proofs of our main results on extended inversion is Variable Projection reduction of the penalty objective $J_\alpha$ (equation 18) to a function $\tilde{J}_\alpha$ of $m$ alone, by minimization over $w$:

$$\tilde{J}_\alpha[m; d] = \inf_w J_\alpha[m, w; d] \quad (19)$$

A minimizer $w$ on the right-hand side of definition 19 must solve the normal equation

$$(F[m]^T F[m] + \alpha^2 A^TA)w = F[m]^T d. \quad (20)$$

With $A$ of the form 15, $\tilde{J}_\alpha$ is explicitly computable. First observe that apart from amplitude, $F[m]$ is unitary: for $g \in D$,

$$F[m]^T g(t) = \begin{cases} \frac{1}{4\pi r} g(t + mr), & t \in [t_{min} - mr, t_{max} - mr], \\ 0, & \text{else.} \end{cases} \quad (21)$$

so

$$F[m]^T F[m] = \frac{1}{(4\pi r)^2} I_{[t_{min} - mr, t_{max} - mr]} \quad (22)$$

in which $1_S$ denotes multiplication by the characteristic function of a measurable $S \subset \mathbb{R}$.

Therefore the normal equation for the minimizer on the RHS of equation 19 is

$$\left( \frac{1}{(4\pi r)^2} I_{[t_{min} - mr, t_{max} - mr]} + \alpha^2 A^TA \right) w = F[m]^T d. \quad (23)$$
With these choices, the normal equation (23) becomes
\[
\left( \frac{1}{(4\pi r)^2} \mathbf{1}_{[t_{\text{min}}-mr, t_{\text{max}}-mr]} + \alpha^2 \mathbf{a}^2 \right) w = F[m]^T d. \tag{24}
\]

**Proposition 2.** Assume the conditions \(3, 10, 15\). Also assume that \(\lambda \in (0, \lambda_{\text{max}}], \alpha > 0\), and that \(C > 0\) exists so that \(a \in L^\infty(\mathbb{R})\) mentioned in condition \(15\) satisfies condition \(27\). Then

1. the normal operator \(F[m]^T F[m] + \alpha^2 A^T A\) is invertible for any \(m \in M, \alpha > 0\);
2. the solution \(w_\alpha[m; d] \in W\) of the normal equation \(24\) is given by
   \[
w_\alpha[m; d](t) = \left\{ \begin{array}{ll}
\frac{1}{(4\pi r)^2} d(t + mr), & t \in [t_{\text{min}} - mr, t_{\text{max}} - mr]; \\
0, & \text{else;}
\end{array} \right.
\]
   \(\text{supp} F[m]^T d \subset [t_{\text{min}} - mr, t_{\text{max}} - mr].\) \tag{25}
3. if in addition \(d = F[m_*] w_* \in W_\lambda\) is noise-free, as in equation \(9\)
   \[
w_\alpha[m, d](t) = (1 + (4\pi r)^2 \alpha^2 a(t)^2)^{-1} w_* (t + (m - m_*)r). \tag{26}
\]

**Proof.**

1. Note that thanks to \(11\) if \(|t| \leq \lambda \leq \lambda_{\text{max}}\), then \(1_{[t_{\text{min}}-mr, t_{\text{max}}-mr]}(t) = 1\), whereas if \(|t| > \lambda\), then \(a(t) \geq C\), whence
   \[
   \frac{1}{(4\pi r)^2} 1_{[t_{\text{min}}-mr, t_{\text{max}}-mr]} + \alpha^2 a^2 \geq \min \{ (4\pi r)^2, \alpha^2 \min \{1/(4\pi r)^2, C^2 \} \} > 0.
   \]
   Therefore the normal operator is invertible under the stated conditions.

2. From the identity \(21\)
   \[
   \text{supp} F[m]^T d \subset [t_{\text{min}} - mr, t_{\text{max}} - mr].
   \]
   Define \(w_{\text{tmp}}\) to be the right-hand side of equation \(25\). Then from the previous observation and identity \(21\)
   \[
   \text{supp} w_{\text{tmp}} \subset [t_{\text{min}} - mr, t_{\text{max}} - mr].
   \]
   From the identity \(22\) for any \(w \in W\),
   \[
t \in [t_{\text{min}} - mr, t_{\text{max}} - mr] \Rightarrow F[m]^T F[m] w(t) = \frac{1}{(4\pi r)^2} w(t).
   \]
   It follows from this and the previous two observations that \(w_{\text{tmp}}\) solves the normal equation \(20\) and therefore that \(w_\alpha[m; d] = w_{\text{tmp}}\).

3. Follows by inserting the definition \(9\) of \(d\) in \(25\) and rearranging.

\[\Box\]

**Theorem 2.** Assume the condition \(10\) \(C > 0\), and suppose that \(A\) is given by equation \(15\) for \(a \in L^\infty(\mathbb{R})\) satisfying
\[
a \geq 0; a(t) \geq C \text{ for } |t| \geq \lambda_{\text{max}}. \tag{27}
\]
Then
1. the reduced objective $\tilde{J}_\alpha$ is given by

$$\tilde{J}_\alpha[m; d] = J_\alpha[m, w_\alpha[m; d]; d],$$  \hspace{1cm} (28)

in which $w_\alpha[m; d] \in W$ is the unique solution of the normal equation \[20\].

2. The following are equivalent:

i. $(m, w) \in M \times W$ is a local minimizer of $J_\alpha[\cdot, \cdot; d]$, and

ii. $m$ is a local minimizer of $\tilde{J}_\alpha[\cdot; d]$ and $w = w_\alpha[m; d]$.

Proof. These conclusions follow immediately from Proposition 2.

Remark: If $J_\alpha[\cdot, \cdot; d]$ and $\tilde{J}_\alpha[\cdot; d]$ were differentiable, then “local minimizer” in the conclusion of the preceding theorem could be replaced by “stationary point”. However, for the problem addressed in this paper, $J_\alpha[\cdot, \cdot; d]$ is not differentiable without added smoothness constraints on $w$, whereas $\tilde{J}_\alpha[\cdot; d]$ is differentiable for proper choice of penalty operator $A$.

This conclusion follows from properties of the modeling operator $F$ shared with many other inverse problems in wave propagation, as explained in the third appendix. Here, I derive it from explicit expressions for $\tilde{J}_\alpha$ and its components.

Proposition 3. Assume the hypotheses of Proposition \[2\]. Then

$$e[m, w_\alpha[m; d]; d] = \frac{1}{2\|d\|^2} \int_{t_{\text{min}}}^{t_{\text{max}}} dt \left(4\pi r a(t - mr)\right)^4 (1 + (4\pi r a(t - mr))^2)^{-2} d(t)^2 \hspace{1cm} (29)$$

$$p[m, w_\alpha[m; d]; d] = \frac{1}{2\|d\|^2} \int_{t_{\text{min}}}^{t_{\text{max}}} dt \left(4\pi r a(t - mr)\right)^2 (1 + (4\pi r a(t - mr))^2)^{-2} d(t)^2 \hspace{1cm} (30)$$

$$\tilde{J}_\alpha[m; d] = \frac{1}{2\|d\|^2} \int_{t_{\text{min}}}^{t_{\text{max}}} dt \left(4\pi r a(t - mr)\right)^2 (1 + (4\pi r a(t - mr))^2)^{-1} d(t)^2. \hspace{1cm} (31)$$

Proof. From equation \[25\]

$$F[m]w_\alpha[m; d](t) = \frac{1}{4\pi r} \left(\frac{1}{(4\pi r)^2 + \alpha^2 a^2(t - mr)}\right)^{-1} \frac{1}{4\pi r} d(t),$$

so

$$(F[m]w_\alpha[m; d] - d)(t) = (1 + (4\pi r a(t - mr))^2)^{-1} - 1) d(t)$$

$$= -(4\pi r a(t - mr))^2 (1 + (4\pi r a(t - mr))^2)^{-1} d(t).$$

Half the integral of the square of this data residual is $e[m, w_\alpha[m; d], d]$, which proves identity \[20\].

To compute $p[m, w_\alpha[m; d], d]$, note that

$$Aw_\alpha[m; d](t) = a(t) \left(\frac{1}{(4\pi r)^2 + \alpha^2 a^2(t)}\right)^{-1} \frac{1}{4\pi r} d(t + mr)$$
for \( t \in [t_{\min} - mr, t_{\max} - mr] \), so squaring, integrating, and changing integration variables \( t \mapsto t - mr \) gives the result \[30\]

That the VPM objective \( \tilde{J}_\alpha \) is given by \[31\] follows from equations \[22\], \[19\] \[29\], and \[30\]. \qed

**Theorem 3.** Suppose that in addition to the hypotheses of Theorem 2, \( a \in W_{1,\infty}^{1}(\mathbb{R}) \), then \( \tilde{J}_\alpha[;d] \in C^{1}(M) \).

**Proof.** Suppose first that \( a \in C^{1}(\mathbb{R}) \). Differentiation under the integral sign yields the expression for its derivative:

\[
\frac{d}{dt} \tilde{J}_\alpha[m;d] = -\frac{(4\pi \rho \alpha)^2}{\|d\|^2} \int_{t_{\min}}^{t_{\max}} dt \left( \frac{da}{dt} \right) (t - mr)(1 + (4\pi \rho \alpha(t - mr))^2)^{-2}d(t)^2.
\]

(32)

For \( a \in W_{loc}^{1,\infty}(\mathbb{R}) \) a limiting argument shows that the same expression gives the derivative of \( \tilde{J}_\alpha \).

It will be useful to record expressions for the various components of \( \tilde{J}_\alpha \) when the data is noise-free, that is, the context of Proposition 2, item 3.

**Corollary 1.** Assume the hypotheses of Proposition 2, item 3. Then noting that \( \|d\| = \|w_*\|/(4\pi r) \)

\[
e[m, w_\alpha[m,d];d] = \frac{\alpha^4}{2\|w_*\|^2} \int dt a(t - (m - m_*)) \left( 1 + (4\pi r)^2 \alpha^2 a(t - (m - m_*)) \right)^{-2} w_*(t)^2.
\]

(33)

\[
p[m, w_\alpha[m,d];d] = \frac{(4\pi \rho \alpha)^2}{2\|w_*\|^2} \int dt \frac{a(t - (m - m_*))}{(1 + (4\pi r)^2 \alpha^2 a(t - (m - m_*))^2)^2} w_*(t)^2.
\]

(34)

so

\[
\tilde{J}_\alpha[m;d] = \frac{(4\pi \rho \alpha)^2}{2\|w_*\|^2} \int dt a(t - (m - m_*)) \left( 1 + (4\pi r)^2 \alpha^2 a(t - (m - m_*))^2 \right)^{-1} w_*(t)^2.
\]

(35)

Finally, if \( a \in W_{loc}^{1,1}(\mathbb{R}) \), then \( \tilde{J}_\alpha[;d] \) is differentiable, and

\[
\frac{d}{dm} \tilde{J}_\alpha[m;d] = -\frac{r(4\pi \rho \alpha)^2}{\|w_*\|^2} \int dt \frac{a da}{dt} (t - (m - m_*)) \left( 1 + (4\pi r)^2 \alpha^2 a(t - (m - m_*))^2 \right)^{-2} w_*(t)^2.
\]

(36)

**CHOICE OF PENALTY OPERATOR**

I examine two choices for \( A \). For each choice, I ask first whether local minimizers of the resulting VPM objective \( \tilde{J}_\alpha[;d] \) occur far from a slowness \( m_* \).

A penalty operator \( A \) of which \( W_\lambda \) is the null space would be a natural choice. Such operators have come to be called “annihilators”, since they map all members of the constraint subspace \( W_\lambda \) to zero. A simple example is

\[
A = E_H^\lambda = I - E_\lambda, \text{ where } \quad E_\lambda w(t) = 1_{[-\lambda,\lambda]}(t)w(t).
\]

(37)
That is, $E_{\lambda}$ is the orthogonal projector onto $W_{\lambda}$, and $E_{\lambda}^c$ is the orthogonal projector onto its orthocomplement, an operator of the form $a = 1 - 1_{[-\lambda,\lambda]}$.

**Theorem 4.** Suppose that

1. $\lambda \in (0, \lambda_{\text{max}}]$;
2. $m_s \in M, w_s \in W_{\lambda}$, and $d = F[m_s]w_s$ (noise-free data);
3. $A = E_{\lambda}^c$, that is, $a = 1 - 1_{[-\lambda,\lambda]}$ in the definition [15].

Then if $|m - m_s| > 2\lambda/r$,

$$J_\alpha[m; d] = \frac{(4\pi r\alpha)^2}{2(1 + (4\pi r\alpha)^2)}.$$ 

**Proof.** of Theorem 4

This is clear from the definition $a = 1 - 1_{[-\lambda,\lambda]}$ and equation [35] as the supports of $w_s$ and $1_{[-\lambda,\lambda]}(t - (m - m_s))$ are disjoint for the range of $m$ identified in the theorem.

**Remark.** One might have thought that $A = E_{\lambda}^c$ would be a better choice of annihilator, as for noise-free data, the solution set defined by the problem statement [6] is the same as the set of global minimizers of $J_\alpha$ in this case. However, for this choice of annihilator, $tJ_\alpha$ exhibits the same feature as the mean square error $e$, namely a continuum of local minimizers at any distance from the global minimizer $m_s$ greater than a multiple of $\lambda$. Therefore the extended inversion with this choice of annihilator is no more amenable to local optimization than is FWI.

A second possible penalty operator penalizes energy away from $t = 0$: choose $\tau > 0$ and set

$$a(t) = \min(|t|, \tau).$$

Note that $a \geq 0$ and $a \in L^\infty(\mathbb{R}) \cap W^{1,1}_{\text{loc}}(\mathbb{R})$. The cutoff $\tau$ will be chosen large enough to be effectively inactive: specifically, hindsight suggests

$$\tau = \max\{|t_{\text{min}} - m_{\min}r|, |t_{\text{min}} - m_{\max}r|, |t_{\max} - m_{\min}r|, |t_{\max} - m_{\max}r|\}.$$  

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].

**STATIONARY POINTS**

**Proposition 4.** Suppose that

1. $m_s \in M$;
2. $0 < \mu \leq \lambda$, and $w_s \in W_{\mu}$;
3. $d_s = F[m_s]w_s$;

**Remark.** One might have thought that $A = E_{\lambda}^c$ would be a better choice of annihilator, as for noise-free data, the solution set defined by the problem statement [6] is the same as the set of global minimizers of $J_\alpha$ in this case. However, for this choice of annihilator, $tJ_\alpha$ exhibits the same feature as the mean square error $e$, namely a continuum of local minimizers at any distance from the global minimizer $m_s$ greater than a multiple of $\lambda$. Therefore the extended inversion with this choice of annihilator is no more amenable to local optimization than is FWI.

A second possible penalty operator penalizes energy away from $t = 0$: choose $\tau > 0$ and set

$$a(t) = \min(|t|, \tau).$$

Note that $a \geq 0$ and $a \in L^\infty(\mathbb{R}) \cap W^{1,1}_{\text{loc}}(\mathbb{R})$. The cutoff $\tau$ will be chosen large enough to be effectively inactive: specifically, hindsight suggests

$$\tau = \max\{|t_{\text{min}} - m_{\min}r|, |t_{\text{min}} - m_{\max}r|, |t_{\max} - m_{\min}r|, |t_{\max} - m_{\max}r|\}.$$  

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].
4. \(a(t) = \min\{|t|, \tau\}\) in the definition \[15\] with \(\tau\) given by equation \[39\] and
5. \(\alpha > 0\).

Then for any \(m \in M\),

\[
|m - m_*|r > \lambda \Rightarrow \left| \frac{d}{dm} \tilde{J}_\alpha[m; d_*] \right| > \frac{r(4\pi r \alpha)^2 (\lambda - \mu)}{(1 + (4\pi r \alpha)^2 (\lambda + \mu)^2)^2}
\] (40)

Proof. As observed before, supp \(w_\alpha[m; d_*] \subset [t_{\min} - mr, t_{\max} - mr] \subset [-\tau, \tau]\), with \(\tau\) defined in \[39\]. Therefore, \(a(t) = |t|, aa'(t) = t\) in the support of the integrand on the RHS of equation \[36\], which therefore becomes (after change of integration variable)

\[
\frac{d}{dm} \tilde{J}_\alpha[m; d_*] = -\frac{r(4\pi r \alpha)^2}{\|w_*\|^2} \int dt (1 + (4\pi r)^2 \alpha^2 t^2)^{-2} w_*(t + (m - m_*)r)^2.
\] (41)

Recall that \(w_*(t + (m - m_*)r)\) vanishes if \(|t + (m - m_*)r| > \lambda\). Therefore the integral on the RHS of equation \[41\] can be re-written

\[
= -\frac{r(4\pi r \alpha)^2}{\|w_*\|^2} \int_{-(m-m_*)r-\lambda}^{-(m-m_*)r+\lambda} dt (1 + (4\pi r)^2 \alpha^2 t^2)^{-2} w_*(t + (m - m_*)r)^2
\]

Suppose that \(\mu \leq \lambda\) and \(w_* \in W_\mu\). If \(m > m_* + \lambda/r\), then \(t + (m - m_*)r \in \text{supp } w_*\) implies \(-\mu - \lambda < t < \mu - \lambda < 0\), so

\[
t(1 + (4\pi r)^2 \alpha^2 t^2)^{-2} < (\mu - \lambda)(1 + (4\pi r)^2 \alpha^2 (\mu + \lambda)^2)^{-2} < 0
\]

in the support of the integrand in equation \[41\]. Arguing similarly for \(m < m_* - \lambda/r\), obtain a similar inequality, implying the conclusion \[40\].

Theorem 5. Suppose that

1. \(m_* \in M\);
2. \(0 < \lambda\), and \(w_* \in W_\lambda\);
3. \(d_* = F[m_*]w_*\);
4. \(a(t) = \min\{|t|, \tau\}\) in the definition \[15\] with \(\tau\) given by equation \[39\];
5. \(\alpha > 0\); and
6. \(m \in M\) is a stationary point of \(\tilde{J}_\alpha[\cdot; d_*]\).

Then \(|m - m_*| < \lambda/r\).

Proof. Follows directly from Proposition \[4\] by taking \(\mu = \lambda\).
how does one use this reduced penalty minimization to produce a solution of the inverse problem as in problem statement 6?

how does one answer the same question for noisy data?

The next result answers the first question, in the case of noise-free data:

**Proposition 5.** Suppose that \( a \) is given by definition 38, \( \mu \in (0, \lambda_{\text{max}}] \), \( d \) is given by 3 and \( w_* \in W_\mu \), and \( m_\infty \) is a stationary point of \( \bar{J}_\alpha[\cdot; d] \). Then \((m_\infty, w_\alpha[m_\infty; d])\) is a solution of the inverse problem 4 for any \( \lambda \geq 2 \mu \) and

\[
\varepsilon \geq \frac{(8\pi \tau \mu \alpha)^2}{1 + (8\pi \tau \mu \alpha)^2},
\]

(42)

**Proof.** From the assumption \( w_* \in W_\mu \) and Theorem 3a, \( |(m_\infty - m_*)r| \leq \mu \). From the identity 26, \( \text{supp} w_\alpha[m_\infty; d] \subset [(m_\infty - m_*)r - \mu, (m_\infty - m_*)r + \mu] \subset [-2\mu, 2\mu] \). Because of the support limitation, \( a(t) = |t| \) in the interval of integration appearing in 33 so

\[
e[m_\infty, w_\alpha[m_\infty, d]; d] = 8\pi^2 r^2 \alpha^4 \int_{-\mu}^{\mu} dt \frac{|t - (m_\infty - m_*)r|^4}{(1 + (4\pi r)^2 \alpha^2 |t - (m_\infty - m_*)r|^2)^2} w_\alpha(t)^2
\]

\[
= \frac{1}{2} (4\pi r \alpha)^4 \int_{-\mu}^{\mu} dt \frac{|t - (m_\infty - m_*)r|^4}{(1 + (4\pi r)^2 \alpha^2 |t - (m_\infty - m_*)r|^2)^2} w_\alpha(t)^2
\]

\[
\leq \frac{1}{2} \|d\|^2 \frac{(8\pi \tau \mu \alpha)^4}{(1 + (8\pi \tau \mu \alpha)^2)^2}.
\]

The inequality 42 can be interpreted as a bound on \( \alpha \), given \( \varepsilon \) and \( \lambda \), for a stationary point of \( \bar{J}_\alpha \) to yield a solution of the inverse problem: one obtains a solution, provided that \( \alpha \) is sufficiently small. On the other hand, it is clear that \( \alpha \) cannot be too large if stationary points of \( \bar{J}_\alpha \) are to yield solutions: the integrand in 33 is increasing in \( \alpha \) for every \( t \) and \( m \), and the multiplier

\[
t \mapsto (4\pi \alpha(t - (m_\infty - m_*)r))^4 (1 + (4\pi r)^2 \alpha^2 |t - (m_\infty - m_*)r|^2)^{-2}
\]

tends monotonically to 1 as \( \alpha \to \infty \), uniformly on the complement of any open interval containing \( t = (m_\infty - m_*)r \). Therefore

\[
\lim_{\alpha \to \infty} e[m_\infty, w_\alpha[m_\infty; d]; d] = \frac{1}{2} \frac{1}{(4\pi r)^2} \|w_*\|^2 = \frac{1}{2} \|d\|^2.
\]

(43)

Consequently, there exists \( \alpha_{\text{max}}(\varepsilon, \lambda, d) \) so that

\[
e[m_\infty, w_\alpha[m_\infty; d]; d] \leq \frac{1}{2} \varepsilon^2 \|d\|^2 \Rightarrow \alpha \leq \alpha_{\text{max}}(\varepsilon, \lambda, d).
\]

The existence of this limiting penalty weight has been inferred indirectly; Appendix A describes a constructive algorithm for its approximation.
I turn now to the second issue identified above, the effect of noise. Suppose that the data trace $d$ takes the form

$$d = F[m_*]w_* + n = d_* + n,$$  \hspace{1cm} (44)

with $m_* \in M, w_* \in W_\mu$, $0 < \mu < \lambda$, and noise trace $n \in D$. Since no support assumptions can be made about $n$, equation 25 implies that $w_\alpha[m; d] \notin W_\lambda$ for any values of $\alpha$ and $\lambda$. Therefore minimization of $\tilde{J}_\alpha$ cannot by itself yield a solution of the inverse problem as defined in the problem statement 4. In this section, I explain how a solution may nonetheless be constructed from a stationary point of $\tilde{J}_\alpha$.

First, examine the effect of additive noise on the estimation of the slowness $m$. In expressing the result, use the dimensionless relative data error

$$\eta = \frac{\|n\|}{\|d_*\|}. \hspace{1cm} (45)$$

**Proposition 6.** Assume the hypotheses of Proposition 4, and that $d$ is given by definition 44. Suppose that $m \in M$ is a stationary point of $\tilde{J}_\alpha[\cdot; d]$, and that

$$\eta(1 + \eta) \leq \frac{16}{3\sqrt{3}} \frac{4\pi r \alpha (\lambda - \mu)}{(1 + (4\pi r \alpha (\lambda + \mu))^2)^2} \hspace{1cm} (46)$$

Then

$$|m - m_*| \leq \frac{\lambda}{r}. \hspace{1cm} (47)$$

**Proof.** From equation 32, $d\tilde{J}_\alpha/dm$ is the value of a quadratic form in $d$ with (indefinite) symmetric operator $B = \text{multiplication by}$

$$b(t; m, \alpha) = -\frac{(4\pi r \alpha)^2 (t - mr)}{(1 + (4\pi r \alpha (t - mr))^2)^2}$$

Therefore

$$\left| \frac{d}{dm} \tilde{J}_\alpha[m; d] - \frac{d}{dm} \tilde{J}_\alpha[m; d_*] \right| = \left| \langle (d + d_*), B(d - d_*) \rangle \right| \leq \max_{t \in \mathbb{R}} |b(t; m, \alpha)|\eta(1 + \eta)\|d_*\|^2 \hspace{1cm} (48)$$

A straightforward calculation shows that

$$\max_{t \in \mathbb{R}} b(t; m, \alpha) = \frac{3\sqrt{3}}{16} 4\pi r \alpha.$$  

For a stationary point $m$ of $\tilde{J}_\alpha[\cdot; d]$, the inequality 48 implies

$$\left| \frac{d}{dm} \tilde{J}_\alpha[m; d_*] \right| \leq \frac{3\sqrt{3}}{16} 4\pi r \alpha \eta(1 + \eta)\|d_*\|^2$$

On the other hand, the conclusion 40 of Proposition 4 implies that if also

$$\frac{3\sqrt{3}}{16} 4\pi r \alpha \eta(1 + \eta)\|d_*\|^2 \leq (4\pi r \alpha)^2 \frac{\lambda - \mu}{(1 + (4\pi r)^2 \alpha^2 (\lambda + \mu)^2)}\|d_*\|^2$$

then $|m - m_*| \leq \lambda/r$. Rearranging, obtain the conclusion. \qed
Corollary 2. Assume the hypotheses of Theorem 6 in particular that \( m \) is a stationary point of \( \tilde{J}_\alpha[\cdot; d] \), \( d = d_\star + n \). Then

\[
|m - m_\star| \leq \frac{\mu}{r} + \frac{\eta}{\alpha} \left( \frac{3\sqrt{3}(1 + \eta)}{64\pi r^2} \left( 1 + (8\pi r \alpha \lambda_{\text{max}})^2 \right)^2 \right)
\]

(49)

Proof. Assume that \( \lambda \) is chosen to obtain equality in the condition 46, substitute the bound \( 2\lambda_{\text{max}} \) for \( \lambda + \mu \) in the denominator, solve for \( \lambda \) and substitute in inequality 47.

Remark: This bound suggests that error in the estimate of \( m \) due to data noise is a decreasing function of \( \alpha \), at least for small \( \alpha \). This result is intuitively appealing, and is supported by numerical evidence. A similar but stronger bound will be demonstrated in the next section.

Theorem 6. Assume that

1. \( \alpha, \mu > 0 \),
2. \( m_\star \in M, w_\star \in W_\mu \),
3. \( d_\star = F[m_\star]w_\star \),
4. \( n \in D \) and \( d = d_\star + n \).

Set \( \eta = \|n\|/\|d_\star\| \). Assume that \( \eta \) satisfies inequality

\[
\eta < \frac{\sqrt{3} - 1}{2},
\]

(50)

and that \( m \) is a stationary point of \( \tilde{J}_\alpha[\cdot; d] \). Then

\[
|m - m_\star| \leq \left( 1 + \frac{2\eta(1 + \eta)}{1 - \eta(1 + \eta)} \right) \frac{\mu}{r}.
\]

(51)

Proof. of Theorem 6 Write \( \lambda = (1 + \delta)\mu \), and \( x = 4\pi r \alpha \mu \). Then the right-hand side of equation 46 may be written as

\[
\frac{16}{3\sqrt{3}} \frac{4\pi r \alpha (\lambda - \mu)}{(1 + (4\pi r \alpha (\lambda + \mu))^2)^2} = D \frac{x}{(1 + C^2 x^2)^2},
\]

(52)

where

\[
D = \frac{16}{3\sqrt{3}} \delta, \quad C = 2 + \delta.
\]

The positive stationary point of the quantity on the right-hand side of 52 is a maximum, and occurs at \( x = 1/(\sqrt{3}C) \), that is

\[
4\pi r \alpha \mu = \frac{1}{\sqrt{3}(2 + \delta)}.
\]

Thus

\[
1 + C^2 x^2 = \frac{4}{3}
\]
hence the maximum value is
\[ \frac{D3\sqrt{3}}{16C} = \frac{\delta}{2+\delta}. \]
This maximum value must be larger than the left hand side of inequality \[46\] that is,
\[ \eta(1+\eta) \leq \frac{\delta}{2+\delta}, \]
in order that there be any solutions at all, but the right hand side is less than 1. This observation establishes the necessity of hypothesis \[50\] of the theorem. Solving this above inequality for \(\delta\) and unwinding the definitions, one finds that the right-hand side of inequality \[46\] is bounded by \[51\] so appeal to Proposition \[6\] finishes the proof. \(\square\)

**Remark.** That is, with the choice of penalty multiplier a given in equation \[38\] and support radius \(\mu\) of the “noise-free” wavelet \(w_\ast\), \(J_\alpha\) has no local minima with slownesses further than \((1+O(\eta))\mu/\gamma\) from the slowness used to generate the data.

**Remark.** The estimate \(|m - m_\ast|/r < \mu(1+O(\eta))\) for local minima of \(J_\alpha\) is sharp: it is possible to choose \(w_\ast \in W_\mu\) so that \(\mu - |m - m_\ast|/r\) is as small as you like. In particular, the “exact” or “true” slowness \(m_\ast\) is not necessarily the only slowness component of a local minimizer, or even the slowness component of any local minimizer, and in particular is not (necessarily) the slowness component of a global minimizer of \(J_\alpha\).

**Remark.** No similar bound could hold for much larger noise levels than specified in condition \[50\] the right-hand side of which is a bit larger than 0.6. For example, if the noise is the predicted data for the same wavelet \(w_\ast\) with a substantially different slowness \(m_\beta\), that is, \(n = F[m_\beta]w_\ast\), then a simple symmetry argument shows that if there is a local minimizer of \(J_\alpha[m, \cdot; d_\ast + n]\) with slowness near \(m_\ast\), there must also be a minimizer with slowness near \(m_\beta\), so that the difference with \(m_\ast\) is not constrained at all by the assumed support radius of \(w_\ast\). So for this example with 100% noise, no bound of the type given by conclusion 2 could possibly hold. The companion paper (Symes et al., 2021) illustrates this phenomenon numerically.

**Remark.** Note that \(\alpha\) plays no role in the conclusions of this theorem. It is only required that \(\alpha > 0\).

**Remark:** I emphasize that Theorem \[6\] states *sufficient* conditions for a bound on the slowness error \(|m - m_\ast|\) in terms of the relative data noise level \(\eta\), giving an additional “fudge factor” beyond the support size \(\mu\) of the noise-free wavelet \(w_\ast\) for an interval within which the slowness error is guaranteed to lie.

Conclusion 1 in Theorem \[6\] constrains the range of noise level to which these results apply to a bit more than 60%. That is, the bound given by conclusion 2 is useful only for small noise. In the limit as \(\eta \to 0\), conclusion 2 becomes \(\lambda/\mu \gtrsim 1 + 2\eta\), that is, the “fudge factor” beyond the noise-free bound is approximately twice the noise level.

On the other hand, stronger bounds than given by Theorem \[6\] are possible, given additional constraints on the noise \(n\). A natural example is uniformly distributed random noise, filtered to have the same spectrum as the source. The expression \[32\] implies that the interaction of noise \(n\) and signal \(d_\ast\) in the derivative of \(\tilde{J}_\alpha\) is local, so that the coefficient of \(\eta\) on the left-hand side of inequality \[46\] is effectively much less that 1, resulting in a larger...
range of allowable $\eta$. While I will not formulate such a result, one of the numerical examples in the companion paper (Symes et al., 2021) suggests its feasibility.

Unless the data is noise-free, there is no reason to suppose that the estimated wavelet $w_\alpha[m;d]$ (Theorem 2) will lie in $W_\lambda$, unless the support of the noise $n$ is not restricted.

In order to construct a solution of the inverse problem 6, project $w_\alpha[m;d]$ onto $W_\lambda$. For sufficiently large $\lambda, \epsilon$, the result is a solution of the inverse problem:

**Theorem 7.** Assume the hypotheses of Theorem 6 and that inequality 50 holds, and $\mu \in (0, \lambda_{\text{max}}]$. Then the pair

$$(m, 1_{[-\lambda, \lambda]} w_\alpha[m,d])$$

solves the inverse problem as stated in 6 if

$$\left(2 + \frac{2\eta(1 + \eta)}{1 - \eta(1 + \eta)}\right) \mu \leq \lambda \leq \lambda_{\text{max}}, \quad (53)$$

and

$$\epsilon \geq \frac{(8\pi r\alpha \lambda)^2}{1 + (8\pi r\alpha \lambda)^2} + \eta. \quad (54)$$

**Proof.** of Theorem 7. From Theorem 2,

$$w_\alpha[m;d](t) = (1 + (4\pi r\alpha t)^2)^{-1} (w_\ast(t + (m - m_\ast)) + 4\pi r n(t + mr))$$

$$= w_\alpha[m;d_\ast] + (1 + (4\pi r\alpha t)^2)^{-1} 4\pi r n(t + mr)$$

From Theorem 6,

$$|m - m_\ast| \leq \left(1 + \frac{2\eta(1 + \eta)}{1 - \eta(1 + \eta)}\right) \frac{\mu}{r}$$

which from assumption 53 is

$$\leq \left(2 + \frac{2\eta(1 + \eta)}{1 - \eta(1 + \eta)}\right) \frac{\mu}{r} - \frac{\mu}{r} \leq \left(\frac{\lambda}{\mu} - 1\right) \frac{\mu}{r}$$

That is,

$$|m - m_\ast| r \leq \lambda - \mu.$$
\[-(1 - 1_{[-\lambda + mr, \lambda + mr]}n(t))\]

From [33] and the bound on \(m - m_*\),

\[
\|F[m]E_\lambda w_\alpha[m, d_*] - d_*\|^2 = (4\pi r \alpha)^4 \int_{-\lambda + (m - m_*)r}^{\lambda + (m - m_*)r} dt (t - (m - m_*)r)^4 \\
\times (1 + (4\pi r \alpha)^2(t - (m - m_*)r)^2)^{-2} d_* (t)^2 \\
\leq (4\pi r \alpha)^4 \|d_*\|^2
\]

Similarly, the norm squared of the sum of the last two terms is

\[
\leq (4\pi r \alpha)^2 \|1_{[-\lambda + mr, \lambda + mr]} n\|^2 + \|1 - 1_{[-\lambda + mr, \lambda + mr]} n\|^2
\]

Without additional hypotheses to outlaw the accumulation of \(n\) near \(t = mr\), all that can be said is that this is

\[
\leq \max\{(4\pi r \alpha)^2, 1\} \|n\|^2
\]

Putting this all together,

\[
\|F[m]E_\lambda w_\alpha[m, d] - d\| \leq (4\pi r \alpha)^2 \|d_*\| + \max\{4\pi r \alpha \lambda, 1\} \|n\| \\
\leq ((4\pi r \alpha \lambda)^2 + \max\{4\pi r \alpha \lambda, 1\} \eta) \|d_*\|.
\]

If the right-hand side is to be less than \(\|d_*\|\) as required by the definition [6] of the inverse problem, then necessarily \(4\pi r \alpha \lambda < 1\), so the right hand side in the preceding inequality is bounded by the right hand side of assumption [54] of the theorem. Therefore this assumption implies that the relative residual is \(\leq \epsilon\). 

**Remark:** Note that the sufficient condition [53] for \(\lambda\) is independent of \(\alpha\). It follows that for any choice of \(\lambda\) consistent with this bound, \((m, 1_{[-\lambda, \lambda]} w_\alpha[m, d])\) is a solution of the inverse problem for any \(\epsilon > 0\) provided that \(\alpha\) is chosen sufficiently small \((O(\sqrt{\epsilon}))\).

**SELECTION OF \(\alpha\)**

So far in this story, the selection of the penalty weight \(\alpha\) has played a relatively minor role. In practical calculation, on the other hand, selection of \(\alpha\) strongly influences the convergence of iterative methods and the quality of the results.

Concentration of the wavelet near \(t = 0\) clearly favors larger \(\alpha\): as will be established in Appendix A, the reduced penalty term \(g[m, w_\alpha[m, d]; d]\) is a decreasing function of \(\alpha\) for any \(m\). As \(g\) measures the dispersion of the wavelet away from \(t = 0\), larger \(\alpha\) is to be preferred, all else being equal.

However all else is not equal: the choice of \(\alpha\) affects to VPM reduced objective \(\tilde{J}_\alpha[m; d]\) and therefore the estimation of the nonlinear variable \(m\) as well. To express this connection, observe that for any \(d \in D, \lambda \in (0, \lambda_{\text{max}})\) and \(m \in M\), the orthogonal projection of \(d\) onto the range \(F[m]W_\lambda\) is \((1_{[mr-\lambda, mr+\lambda]}d)(t + mr)\), whence the distance from \(d\) to the range of \(F\) restricted to \(W_\lambda\) is

\[
\inf_{w \in W_\lambda} \|F[m]w - d\| = \|(1 - 1_{[mr-\lambda, mr+\lambda]}d)\|.
\]
Denote by \( r\lambda[m; d] \) this distance normalized by \( \|d\| \):

\[
r\lambda[m; d] = \frac{\|(1 - \mathbf{1}_{[mr - \lambda, mr + \lambda]}d)\|}{\|d\|}.
\]

Denote by \( v\lambda[m; d] \) the pre-image under \( F \) of the projection of \( d \) onto \( F[m]W\lambda \):

\[
v\lambda[m; d](t) = 4\pi r 1_{[-\lambda, \lambda]}(t)d(t + mr).
\]

Then using notation introduced earlier in this section for the residual norm,

\[
r\lambda[m; d] = r[m, v\lambda[m; d]; d]
\]

**Theorem 8.** Suppose \( d \in D, \alpha > 0, 0 < \lambda \leq \lambda_{\text{max}}, m, m_\ast \in M, \) and condition 10 holds. Also assume that

\[
\gamma = \frac{r[m, w\alpha[m; d]; d]}{\sqrt{1 - r\lambda[m_\ast; d]^2}} < 1.
\]

Then

\[
|m - m_\ast| \leq \lambda/r + \frac{1}{4\pi r^2 \alpha} \left( \frac{\gamma}{1 - \gamma} \right)^{1/2}
\]

**Proof.** From the definition, \( 0 \leq r\lambda[m; d] < 1 \) for all \( m \in M \), so \( \gamma \) is well-defined.

Since condition 10 is assumed to hold, the conclusions of Proposition 3 are available, in particular, equation 29:

\[
e[m, w\alpha[m; d]; d] = \frac{1}{2} r[m, w\alpha[m; d]; d]^2 = \frac{1}{2\|d\|^2} \int dt H(4\pi r\alpha|t - mr|)d(t)^2
\]

in which

\[
H(x) = \left( \frac{x^2}{1 + x^2} \right)^2.
\]

Note that \( 0 \leq H \leq 1 \) and \( H \) is strictly increasing on \( \mathbb{R}^+ \).

Either \( |m - m_\ast|r > \lambda \), or not. In the latter case, the conclusion 56 follows trivially. In the former case,

\[
r[m, w\alpha[m; d]; d]^2 = \frac{1}{\|d\|^2} \int dt H(4\pi r\alpha|t - mr|)d(t)^2
\]

\[
\geq \frac{1}{\|d\|^2} \int dt H(4\pi r\alpha|t - mr|)(1_{[m_r - \lambda, mr + \lambda]}d(t))^2
\]

\[
= \frac{1}{\|d\|^2} \int dt H(4\pi r\alpha|t - (m - m_\ast)r|)(1_{[-\lambda, \lambda]}d(t + m_\ast r))^2
\]

\[
\geq \frac{H(4\pi r\alpha(|m - m_\ast|r - \lambda))}{\|d\|^2} \int dt (1_{[-\lambda, \lambda]}d(t + m_\ast r))^2
\]

\[
= \frac{H(4\pi r\alpha(|m - m_\ast|r - \lambda))}{\|d\|^2} \int dt (1_{[m_r - \lambda, mr + \lambda]}d(t))^2
\]

\[
= H(4\pi r\alpha(|m - m_\ast|r - \lambda))(1 - r\lambda[m_\ast; d]^2)
\]
20

whence from the definition of \( \gamma \)

\[
H(4\pi r\alpha(|m - m_*|r - \lambda)) \leq \gamma. \tag{57}
\]

H is bijective: \( \mathbb{R}^+ \to [0, 1) \). Solving inequality 57 for \( 4\pi r\alpha(|m - m_*|r - \lambda) \) and rearranging, obtain the conclusion.

**Remark:** The parameters \( \alpha \) and \( \lambda \) play completely independent roles in this result. Given \( d \in D, m_* \in M, \) and \( \lambda \in (\lambda_{\text{max}}, \lambda) \), \( r_\lambda[m_*;d] \) is the lower bound on relative RMS data error with slowness \( m_* \), attained at \( w = v_\lambda[m;d] \). Select \( r_* > 0 \) so that \( r_*^2 < 1 - r_\lambda[m_*;d]^2 \), and suppose the for each \( \alpha > 0 \), it is possible to find \( m_\alpha \in M \) so that \( r[m_\alpha, w_\alpha[m_\alpha;d];d] = r_* \). Then from 56

\[
|m_\alpha - m_*| \leq \lambda/r + O(1/\alpha).
\]

That is, this bound suggests that amongst \( \alpha > 0 \) that permit attainment (via proper choice of \( m \)) of a prescribed value for the RMS data error at the minimizer of \( J_\alpha[m, \cdot; d] \), then the bound 56 on the error between this \( m \) and the “target” slowness \( m_* \) is smaller if \( \alpha \) is larger. This observation motivates the Discrepancy Algorithm described in Appendix A.

**REFERENCES**

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

Duistermaat, J. J., 1996, Fourier integral operators: Birkhauser.

Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.

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

Folland, G. B., 2007, Real analysis: Wiley. (2nd edition).

Friedlander, F., 1958, Sound pulses: Cambridge University Press.

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.

Hanke, M., 2017, A taste of inverse problems: Society for Industrial and Applied Mathematics.

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

———, 2018a, Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation: Geophysics, 83, R449–R461.

———, 2018b, Volume source-based extended waveform inversion: Geophysics, 83, R369–R387.

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.
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.

Métilvier, L., A. Allain, B. Brossier, Q. Merigot, E. Oudet, and J. Virieux, 2018, Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach: Geophysics, 83, no. 5, R515–R540.

Métilvier, L., and R. Brossier, 2020, A receiver-extension approach to robust full waveform inversion: 90th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 641–645.

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.

Schuster, G., 2017, Seismic inversion: Society of Exploration Geophysicists.

Stolk, C. C., M. V. de Hoop, and W. Symes, 2009, Kinematics of shot-geophone migration: Geophysics, 74, no. 6, WCA18–WCA34.

Stolk, C. C., and W. Symes, 2003, Smooth objective functionals for seismic velocity inversion: Inverse Problems, 19, 73–89.

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

———, 2009, The seismic reflection inverse problem: Inverse Problems, 25, 123008:1–24.

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

Symes, W., H. Chen, and S. Minkoff, 2021, Solution of an acoustic transmission inverse problem by extended inversion. (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.

Terentyev, I., 2017, Nonlinear waveform inversion with surface-oriented extended modeling: PhD thesis, Rice University.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, 74, no. 6, WCC127–WCC152.

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.

Yang, Y., B. Engquist, J. Sun, and B. Hamfeldt, 2018, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: Geophysics, 83, no. 1, R43–R62.

Yosida, K., 1996, Functional analysis: Springer Verlag.
APPENDIX A

PENALTY WEIGHT SELECTION VIA THE DISCREPANCY PRINCIPLE

This appendix describes an algorithm for controlling the penalty weight $\alpha$ based on a version the Discrepancy Principle (Engl et al., 1996; Hanke, 2017; Fu and Symes, 2017). The algorithm solves the problem

given $d \in D$ and $0 < e_- < e_+$, find $m \in M, \alpha \in \mathbb{R}^+$ so that

(i) $m$ is a stationary point of $\tilde{J}_\alpha[\cdot; d]$, and
(ii) $e[m, w_\alpha[m; d]; d] \in (e_-, e_+)$.

I describe an alternating, or coordinate search, algorithm for solution of this problem, combining a local optimization algorithm for updating $m$, and a second algorithm for updating $\alpha$. A first version of this algorithm appeared in Fu and Symes (2017). Note that the mean square error $e$ lies in an open interval $(e_-, e_+)$ at a solution of this problem. Use of an interval, rather than a single target error level, accomplishes two objectives:

- it is consistent with the general lack of precise knowledge of data error in most applications;
- it permits a local optimization algorithm to make several small updates of $m$ before an update of $\alpha$ is required, as is required for good performance of algorithms such as BFGS.

Given an objective function $\Phi : M \rightarrow \mathbb{R}$, a local optimization algorithm generates a map $G[\Phi, \ldots] : M \rightarrow M$, mapping a current estimate of $m$ to an updated estimate. The update rule may depend only on the current estimate, as is the case for steepest descent, the Gauss-Newton method, or Newton’s method, or may also depend on information generated during earlier updates, as for secant-type methods such as BFGS. The notation is intended to allow for this latter possibility. Under “standard conditions” on $\Phi$ and the initial estimate for $m$, and equipped with so-called globalization safeguards to ensure satisfaction of sufficient decrease conditions (see Nocedal and Wright (1999)), iteration of $G$ produces a sequence in $M$ converging to a local minimizer of $\Phi$.

A normal stopping criterion for such an algorithm would be a tolerance test for the norm of $\nabla \Phi$ (and a limit on iteration count, of course). In application to $J_\alpha = e + \alpha^2 g$, add another stopping criterion:

stop if $e \notin [e_-, e_+]$.

Since $e$ is a summand in $\tilde{J}_\alpha$, $e$ will typically decrease along a minimizing sequence, so the expected condition invoking this stopping rule is $e < e_-$. Denote a update rule with this enhancement, applied to $\tilde{J}_\alpha$, by $G[\alpha, e_-, e_+, \ldots]$, still allowing the possibility of information in addition to the current iterate.
The algorithm also requires a rule for updating $\alpha$, defining a map $H : M \times E \rightarrow \mathbb{R}^+$, in which $E = \{(e_-, e_+) \in \mathbb{R}^2 : 0 < e_- < e_+\}$. For any $m \in M$ the rule is required to produce an $\alpha$ for which the error bounds are satisfied: If $(m, e_-, e_+) \in M \times E$ and $\alpha = H(m, e_-, e_+)$, then $e[m, w_\alpha[m, d]; d] \in [e_-, e_+]$. A usable rule is described below (equation A-11).

In outline, the algorithm is as follows:

**Algorithm 1** Scheme for updating $m, \alpha$

1. Choose $m \in M$
2. repeat
3. \[ \alpha \leftarrow H[m, e_-, e_+] \]
4. repeat
5. \[ m \leftarrow G[\alpha, e_-, e_+, \ldots](m) \]
6. until $e[m, w_\alpha[m, d]; d] \notin [e_-, e_+]$ or $\|\nabla \tilde{J}_\alpha[m; d]\|$ sufficiently small, or...
7. until $e[m, w_\alpha[m, d]; d] \in [e_-, e_+]$

Note that after step 3, the error bounds are satisfied, that is, $e[m, w_\alpha[m, d]; d] \in [e_-, e_+]$, but after step 4, that is likely not to be the case: the alteration of $m$ is likely to reduce $e$, as it is a summand in the definition of $\tilde{J}_\alpha$. If the $e$ is reduced below $e_-$, or if an approximate local minimizer is detected, then the condition in step 6 is satisfied, so control returns to step 7. If it is satisfied, the algorithm terminates, else control loops back to step 2, $\alpha$ is updated, and the inner $m$ update loop is entered again. Termination requires that the bounds on $e$ are satisfied (step 7), hence that the $m$ update loop terminates by finding a local min which satisfies these bounds.

The penalty parameter update strategies are based on the following fact about linear combinations of quadratic forms, similar to well-known results from the theory of Tihonov regularization ([Hanke, 2017]):

**Theorem 9.** Suppose that $W, D$ are Hilbert spaces, $F \in \mathcal{B}(W, D)$, $A \in \mathcal{B}(W, W)$, $F^T F + \alpha^2 A^T A > 0$ for any $\alpha \geq 0$ (in particular, $F^T F > 0$), and $A$ is injective. For $d \in D$, define $w : D \times \mathbb{R}^+ \rightarrow W$, $e : D \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$, and $g : D \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ by

\[
\begin{align*}
    w[d, \alpha] &= (F^T F + \alpha^2 A^T A)^{-1} F^T d \\
    e[d, \alpha] &= \frac{1}{2\|d\|^2} \|Fw[d, \alpha] - d\|^2 \\
    g[d, \alpha] &= \frac{1}{2\|d\|^2} \|Aw[d, \alpha]\|^2
\end{align*}
\]

(A-1)

Then for any $d \in D$, $w[d, \cdot]$, $e[d, \cdot]$, and $g[d, \cdot]$ are of class $C^\infty(\mathbb{R}^+)$, and

0. EITHER $d$ is perpendicular to the range of $F$, $e \equiv \frac{1}{2}, g \equiv 0$ for all $\alpha \geq 0$,
1. OR $e$ is positive and strictly increasing, $g$ is positive and strictly decreasing for $\alpha \geq 0$, and
\[
\frac{d e}{d\alpha^2} \leq 2g.
\]

(A-2)

**Remark:** The conditions on $F$ and $A$ in the statement of this theorem might seem unduly restrictive. Both hold for the cartoon problem studied in the body of this paper. In fact,
in most cases studied in the literature, the extended modeling operator represented here by $F$ is not coercive, and must be modified by regularization to meet the conditions of the theorem. Instances of the penalty operator $A$ studied in the literature are in fact usually injective, though this fact is commonly overlooked. See Symes (2009) for some discussion.

**Proof.** Since the normal operator $F^TF + \alpha^2 A^TA$ is boundedly invertible and smooth in $\alpha \geq 0$, $w$ is smooth in $\alpha$, as are $e$ and $g$. Differentiate the definition of $w$ with respect to $\alpha^2$ to obtain

$$ (F^TF + \alpha^2 A^TA) \frac{dw}{d\alpha^2} = -A^T Aw $$

whence

$$ \frac{de}{d\alpha^2} = \left\langle \frac{dw}{d\alpha^2}, F^T(Fw - d) \right\rangle $$

$$ = -\alpha^2 \left\langle \frac{dw}{d\alpha^2}, A^T Aw \right\rangle $$

$$ = \alpha^2 \langle A^T Aw, (F^TF + \alpha^2 A^TA)^{-1} A^T Aw \rangle $$

$$ \geq 0 \quad (A-4) $$

Note that the inequality in equation A-4 is strict if $g[d, \alpha] > 0$ hence $A^T Aw[d, \alpha] \neq 0$, since the normal operator is assumed to be positive definite. Also,

$$ \frac{dg}{d\alpha^2} = -\langle A^T Aw, (F^TF + \alpha^2 A^TA)^{-1} A^T Aw \rangle $$

$$ \leq 0 \quad (A-5) $$

similarly a strict inequality if $g \neq 0$.

It follows that either $g[d, \alpha] > 0$ for all $\alpha \geq 0$, or there exists $\alpha_0 \geq 0$ for which $g[d, \alpha] > 0$ for $0 \leq \alpha < \alpha_0$ and $g[d, \alpha] = 0$ for $\alpha \geq \alpha_0$, hence $Aw[d, \alpha] = 0$ for $\alpha \geq \alpha_0$. Since $A$ is assumed injective, $w[d, \alpha] = 0$ for $\alpha > \alpha_0$. From the definition of $w$, it follows that $F^Td = 0$, that is, $d$ is orthogonal to the range of $F$, and in fact $\alpha_0 = 0$, $g \equiv 0$ and $e \equiv \frac{1}{2}$ for all $\alpha \geq 0$.

In the first case, that is, $F^Td \neq 0$, equations A-4 and A-5 show that increasing $\alpha^2$ implies increasing $e$ while decreasing $g$, and

$$ \langle A^T Aw, (F^TF + \alpha^2 A^TA)^{-1} A^T Aw \rangle $$

$$ = \lim_{\epsilon \to 0} \langle (A^TA + \epsilon^2 I)^{1/2} w, [(A^TA + \epsilon^2 I)^{-1/2} F^TF(A^TA + \epsilon^2 I)^{-1/2} + \alpha^2 I]^{-1} (A^TA + \epsilon^2 I)^{1/2} w \rangle $$

$$ \leq \lim_{\epsilon \to 0} \frac{1}{\alpha^2 \epsilon^2} \langle (A^TA + \epsilon^2 I)w, w \rangle = \frac{2}{\alpha^2 g}. \quad (A-6) $$

which establishes inequality A-2. \hfill \Box

**Proposition 7.** Under the hypotheses of Theorem 7,

$$ \lim_{\alpha \to \infty} w(d, \alpha) = 0, \quad (A-7) $$

$$ \lim_{\alpha \to \infty} e(d, \alpha) = 1/2. \quad (A-8) $$
Proof. Since \( F^T F \) is positive-definite, it has a positive-definite square root, and the definition \( A-1 \) of \( w(d, \alpha) \) is equivalent to
\[
(F^T F)^{1/2} w(d, \alpha) = (I + \alpha^2 (F^T F)^{-1/2} A^T A (F^T F)^{-1/2})^{-1} (F^T F)^{-1/2} F^T d.
\]  
(A-9)

Denote by \( E : \mathbb{R} \rightarrow \mathcal{B}(W, W) \) the resolution of the identity for the self-adjoint bounded operator \((F^T F)^{-1/2} A^T A (F^T F)^{-1/2}\). According to the theorem on spectral representation of functions of self-adjoint operators \( \text{(Yosida (1996), section XI.5, Theorem 1)} \),
\[
\| (F^T F)^{1/2} w(d, \alpha) \|^2 = \int_0^\infty \| F^T F \| \frac{1}{1 + \alpha^2 \lambda} d\| E(\lambda) (F^T F)^{-1/2} F^T d \|^2.
\]  
(A-10)

Since \( A^T A \) is assumed injective, so is \( (F^T F)^{-1/2} A^T A (F^T F)^{-1/2} \), whence 0 is either a member of the resolvent set of the latter operator, or of its continuous spectrum. In either case, \( \{0\} \) is a set of measure zero with respect to \( \| E(\lambda) (F^T F)^{-1/2} F^T d \|^2 \), so the integrand \( (1 + \alpha^2 \lambda) \) converges to zero \( \| E(\lambda) (F^T F)^{-1/2} F^T d \|^2 \)-almost everywhere as \( \alpha \to \infty \), and is bounded above by 1 on the spectrum. The Dominated Convergence Theorem implies that \( \| (F^T F)^{1/2} w(d, \alpha) \|^2 \to 0 \) as \( \alpha \to 0 \), which establishes the first conclusion. The second follows from the definition of \( e \) and the continuity of the operators involved in it. \( \square \)

**Proposition 8.** In addition to the hypotheses of Theorem 9, assume that \( F^T d \neq 0 \), and that \( e[d, 0] < e_+ \leq 1/2 \). Then there exists a unique \( \alpha_+ > 0 \) satisfying \( e_+ = e[d, \alpha_+] \), and \( e[d, \alpha] < e_+ \) for \( \alpha \in [0, \alpha_+] \). Define \( \Psi : [0, \alpha_+] \to \mathbb{R} \) by
\[
\Psi(\alpha) = \left( \alpha^2 + \frac{e_+ - e[d, \alpha]}{2g[d, \alpha]} \right)^{1/2}
\]  
(A-11)

Then \( \Psi([0, \alpha_+]) \subset (0, \alpha_+] \) and \( \alpha < \Psi(\alpha) < \alpha_+ \) for \( \alpha \in [0, \alpha_+] \).

**Remark:** Note that \( \Psi(0) \) is well-defined and \( > 0 \).

**Proof.** Existence and uniqueness of \( \alpha_+ \) follows from Theorem 9, Proposition 7 and the assumption that \( e[d, 0] < e_+ \leq 1/2 \). Since \( e \) is strictly increasing under the assumption that \( d \) is not perpendicular to the range of \( F \), \( e[d, \alpha] < e_+ \) for \( \alpha \in [0, \alpha_+] \).

Suppose that \( 0 \leq \alpha < \psi \leq \alpha_+ \). From inequality A-2
\[
e[d, \psi] - e[d, \alpha] \leq \int_{\alpha^2}^{\psi^2} d\tau g[d, \tau]
\]
\[
< 2g[d, \alpha](\psi^2 - \alpha^2)
\]  
(A-12)

since \( g \) is positive and strictly decreasing. From the definition A-11 \( \Psi(\alpha) \geq \alpha \) for \( \alpha \in [0, \alpha_+] \), and \( \Psi(\alpha) > \alpha \) if \( \alpha \in [0, \alpha_+] \). Setting \( \psi = \Psi(\alpha) \), inequality A-12 and definition A-11 imply
\[
e[d, \Psi(\alpha)] - e[d, \alpha] < e_+ - e[d, \alpha]
\]
so
\[
\alpha \in [0, \alpha_+] \Rightarrow e[d, \alpha] < e[d, \Psi(\alpha)] < e_+.
\]  
(A-13)

The last inequality and the strict increase of \( e \) imply that \( \Psi(\alpha) < \alpha_+ \) if \( \alpha \in [0, \alpha_+] \). \( \square \)
These results suggest an algorithm to determine \( \alpha \): select an initial \( \alpha_0 \geq 0 \) for which 
\( e[d, \alpha_0] < e_- \), then set \( \alpha_n = \Psi(\alpha_{n-1}), n = 1, 2, \ldots \), until \( e[d, \alpha_n] > e_- \). Since \( e_- < e_+ \), the preceding result implies that \( e[d, \alpha_n] < e_+ \) also.

That is, iteration of \( \Psi \) produces an increasing sequence in \([0, \alpha_+]\). Two further ingredients are required to view it as an algorithm for production of an \( \alpha \) satisfying the discrepancy principle: a method for selection of an initial \( \alpha_0 \), and assurance that the error \( e \) will eventually exceed the lower bound \( e_- \).

**Theorem 10.** In addition to the hypotheses of Proposition 8, assume that \( e[d, 0] < e_- < e_+ \leq 1/2 \). Define the sequence \( \{\alpha_n : n \in \mathbb{Z}_+\} \) by

\[
\begin{align*}
\bullet & \quad \alpha_0 = 0, \text{ and} \\
\bullet & \quad \alpha_{n+1} = \Psi(\alpha_n), n \in \mathbb{Z}_+.
\end{align*}
\]

Then there exists \( N = N(d, e_-, e_+) \in \mathbb{Z}_+ \) for which

\[
\begin{align*}
\bullet & \quad e[d, \alpha_n] \leq e_- \text{ for } 0 \leq n < N, \text{ and} \\
\bullet & \quad e[d, \alpha_N] \in (e_-, e_+).
\end{align*}
\]

**Proof.** By hypothesis, the first conclusion is satisfied for \( N = 0 \). From the definition A-11 of \( \Psi \) and the conclusions of Proposition 8, \( \{\alpha_n\} \) is strictly increasing and \( \alpha_n \leq \alpha_+ \) for all \( n \geq 0 \), so has a limit point \( \alpha_\infty \in (0, \alpha_+] \). If \( \alpha_\infty < \alpha_+ \), \( e = \Psi(\alpha_\infty) - \alpha_\infty > 0 \). Since \( \alpha \mapsto \Psi(\alpha) - \alpha \) is continuous, there is \( \delta > 0 \) so that if \( |\alpha - \alpha_\infty| < \delta \), then \( \Psi(\alpha) - \alpha > \epsilon/2 \). Since \( \alpha_\infty = \lim_{n \to \infty} \alpha_n \), there is \( m \in \mathbb{Z}_+ \) for which \( \alpha_m \in (\alpha_\infty - \min(\epsilon/2, \delta), \alpha_\infty) \), so that \( \alpha_{m+1} = \Psi(\alpha_m) > \alpha_m + \epsilon/2 > \alpha_\infty \), contradicting the definition of \( \alpha_\infty \) as the limit of the increasing sequence \( \{\alpha_n\} \). Therefore, conclude that \( \alpha_\infty = \alpha_+ \).

According to Proposition 8, \( \alpha_n \to \alpha_+ \) and \( \alpha_n \) increases with \( n \), whence \( e[d, \alpha_n] \to e_+ \) and \( e[d, \alpha_n] \) increases with \( n \). Therefore there must exist a least \( n = N(d, e_-, e_+) \) for which 
\( e[d, \alpha_n] \in (e_-, e_+) \).

**Remark:** The condition \( e(d, 0) \leq e_- \) means that \( \sqrt{2e_-} \) is an upper bound for the distance from \( d \) to the range of \( F \). Recall that \( e_- \) represents an underestimate for data noise level. That is, it is assumed that the data is within data noise level of the range of \( F \). Such an assumption is unlikely to be tenable for the physical (non-extended) modeling operator, unless the nonlinear model parameters (\( m \) in this discussion) are chosen near-optimally. The ability to fit data, even when the parameters represented by \( m \) are greatly in error, is an essential characteristic of successful extended inversion strategies.

**Remark:** The update rule A-11 is convergent to a satisfactory value, as just shown, but in practice is rather slow. Accelerated updates using variants of the secant rule are possible, but an analysis of those possibilities is beyond the scope of this paper.

A satisfactory penalty parameter selection rule can now be defined: in the notation introduced at the beginning of this appendix, define

\[
H[m, e_-, e_+] = \alpha_{N(d, e_-, e_+)}. \tag{A-14}
\]
Choosing $F = F[m]$ in Theorem 10 conclude that $H$, so defined, has the required properties.

This is as far as one can go without more concrete assumptions on $F$ and $A$. If the reduced objective $\tilde{J}_\alpha$ possesses a large (in some sense) domain of convexity for a range of $\alpha$, then the discrepancy algorithm sketched here can efficiently drive $\alpha$ to near its largest feasible value, thus giving the best resolution near the global minimizer. That program has been carried out for the simple cartoon problem of this paper, and partly for some other more prototypical inverse wave problems (Fu and Symes, 2017).

**APPENDIX B**

HEISENBERG, SUPPORT, AND WAVELENGTH

Without further contraints on data or solution, nothing more can be said about bounds on $\alpha$. If the data $d$ (and the target wavelet $w_*$) are assumed to have a square integrable derivative, then a necessary condition follows from the Heisenberg inequality (see for example Folland (2007), p. 255). To formulate this result in its most general form, introduce the Hilbert subspaces $V^0 \subset L^2(\mathbb{R}), V^1 \subset H^1(\mathbb{R})$:

$$V^0 = \{f \in L^2(\mathbb{R}) : Af \in L^2(\mathbb{R})\},$$
$$V^1 = V^0 \cap H^1(\mathbb{R}).$$

(B-1)

$V^0$ is the domain of $A$, and equipped with the graph norm of $A$. A natural norm in $V^1$ is

$$\|f\|^2_{V^1} = \|f\|^2_{V^0} + \|f\|^2_{H^1}.$$

$V^j$ is the completion of $C^\infty_0(\mathbb{R})$ in the corresponding norm, $j=0,1$.

**Proposition 9.** For $w \in V^1$,

$$\|Aw\| \|w'\| \geq \frac{1}{2} \|w\|^2$$

(B-2)

**Proof.** For $w \in C^\infty_0(\mathbb{R})$,

$$\int w^2 = \left| \int dt \langle w(t)^2 \rangle \right| = 2 \int dt tw(t)w'(t) \leq 2\|Aw\| \|w'\|$$

by the Cauchy-Schwarz inequality. Since $C^\infty_0(\mathbb{R})$ is dense in $V^1$, the conclusion follows by continuity.

In the conventional formulation of the Heisenberg inequality, the $L^2$ norm of $w'$ is replaced by its equivalent in terms of the Fourier transform $\hat{w}$. Adopting temporarily the use of dummy variables in the expression of functions, the identity (B-2) turns into the usual form of the Heisenberg inequality: for $w \in V^1$,

$$\|tw(t)\| \|k\hat{w}(k)\| \geq \frac{1}{4\pi} \|w\|^2.$$

(B-3)

Define $k_{\text{RMS}}[w]$, the root mean square estimator of frequency of $w \in V^1$, by

$$k_{\text{RMS}}[w] = \frac{1}{2\pi} \frac{\|w'\|}{\|w\|} = \left( \int dk \frac{\|\hat{w}(k)\|^2}{\|\hat{w}\|^2} k^2 \right)^{1/2}.$$

(B-4)
Then the inequalities $\|B-2\|, B-3$ can be rewritten as

$$\|Aw\| \geq \frac{\|w\|}{4\pi k_{\text{RMS}}[w]}.$$  

(B-5)

For $\lambda > 0$, define

$$W^1_\lambda = W_\lambda \cap H^1(\mathbb{R}).$$  

(B-6)

Note that $W^1_\lambda \subset V^1$ is a closed subspace for any $\lambda > 0$.

**Proposition 10.** For $w \in W^1_\lambda$,\n
$$k_{\text{RMS}}[w] \geq \frac{1}{4\pi \lambda}$$\n
Proof. Follows directly from inequality (B-5) and the obvious bound $\|A|_{W_\lambda}\| \leq \lambda$.

**Remark:** This result is the link mentioned earlier between the support constraint and the well-known frequency-based criteria for success of FWI. The result of Theorem 1 can be rephrased as showing the existence of many stationary points of the mean-square error function for which the travel time is in error by more than $\lambda \geq 1/4\pi k_{\text{RMS}}[w]$. In fact, the usual error criterion mentioned in the literature is that the initial estimate of travel time must be in error by at most “half a wavelength” if FWI is to converge reliably. This is correct in some circumstances, depending on features of the target wavelet $w_*$ of which the arguments in this paper do not take account.

**Proposition 11.** For $\lambda > 0, w \in W^1_\lambda$,\n
$$\|A^2 w\| \geq \frac{\|w\|}{24\lambda(2\pi k_{\text{RMS}}[w])^3}.$$  

(B-7)

Proof. Since $A$ preserves $W^1_\lambda$, inequality (B-2) implies

$$\|A^2 w\| \geq \frac{\|Aw\|^2}{2\|(Aw)\|^2}. \quad \text{(B-8)}$$

From the definition of $A$, $(Aw)' = w + Aw'$, so for $w \in W^1_\lambda$,

$$\|(Aw)'\| \leq \|w\| + \lambda\|w'\|.$$\n
Applying (B-2) again,

$$\|w\|^2 \leq 2\|w'\|\|Aw\| \leq 2\lambda\|w'\|\|w\|,$$

so obtain the 1D Poincaré inequality: for $w \in W^1_\lambda$,

$$\|w\| \leq 2\lambda\|w'\|.$$\n
Thus

$$\|(Aw)'\| \leq 3\lambda\|w'\|.$$\n
Apply this estimate together with the basic Heisenberg estimate (B-2) to the inequality (B-8) to obtain

$$\|A^2 w\| \geq \frac{\|w\|^4}{24\lambda\|w'\|3}.$$\n
Rearranging and using the definition (B-4) of $k_{\text{RMS}}[w]$, arrive at inequality (B-7).
Theorem 11. Suppose that $a$ is given by definition \( \alpha, \mu \in (0, \lambda_{\text{max}}/2] \), $d$ is given by \( \alpha \) with $w_* \in W^1_\mu$, and that a stationary point $m_\infty$ of $\tilde{J}_\alpha[\cdot; d]$ yields a solution of the inverse problem \( \Theta \) for $\lambda \geq 2\mu$, $\epsilon > 0$. Then

\[
3(4\pi \lambda k_{\text{RMS}}[w_*])^3 \epsilon \geq \frac{(4\pi r \alpha \lambda)^2}{(1 + (4\pi r \alpha \lambda)^2)} \quad (B-9)
\]

Proof. Rearranging the identity \( \Theta \) and observing as in the proof of Theorem 5 that the support of the integrand is contained in \([-2\mu, 2\mu] \subset [\lambda, \lambda] \),

\[
e[m_\infty, w_\alpha[m_\infty, d]; d] = 8\pi^2 r^2 \alpha^4 \int dt t^4 (1 + (4\pi r)^2 \alpha^2 t^2)^{-2} w_*(t + (m_\infty - m_*))^2
\]

\[
\geq \frac{8\pi^2 r^2 \alpha^4}{(1 + (4\pi r)^2 \alpha^2 \lambda^2)^2} \int dt t^4 w_*(t + (m_\infty - m_*))^2
\]

\[
\geq \frac{8\pi^2 r^2 \alpha^4}{(1 + (4\pi r)^2 \alpha^2 \lambda^2)^2} \|A^2 w_*(\cdot + (m_\infty - m_*)r)\|^2. \quad (B-10)
\]

Since $k_{\text{RMS}}[w]$ is invariant under translation of $w \in V^1$, Proposition \( \Theta \) implies that this is

\[
\geq \frac{8\pi^2 r^2 \alpha^4}{(1 + (4\pi r)^2 \alpha^2 \lambda^2)^2} \|w_*\|^2 / (24)^2 \lambda^2 (2\pi k_{\text{RMS}}[w_*])^6
\]

\[
= \frac{1}{2} \frac{(4\pi r \alpha \lambda)^4}{(1 + (4\pi r \alpha \lambda)^2)^2} \frac{\|d\|^2}{(24)^2 (2\pi \lambda k_{\text{RMS}}[w_*])^6}
\]

The pair $(m_\infty, w_\alpha[m_\infty, d])$ is presumed to solve the inverse problem as stated in \( \Theta \) in particular

\[
\epsilon \geq (2e[m_\infty, w_\alpha[m_\infty, d], d])^{1/2}/\|d\|
\]

the inequality \( \Theta \) follows.

\[
\square
\]

Inequality \( \Theta \) couples the dimensionless quantities $\epsilon$, $4\pi r \alpha \lambda$, and $4\pi \lambda k_{\text{RMS}}[w_*]$. Proposition \( \Theta \) implies that the left hand side is $\geq 3\epsilon$. Since the right hand side is $\leq 1$, the inequality implies no limitation on $\alpha$ if the left hand side is $\geq 1$. For small $\epsilon$, the largest permissible $\alpha$ is $O(\sqrt{\epsilon})$. The permissible range of $\alpha$ increases with nondimensionalized RMS frequency $4\pi \lambda k_{\text{RMS}}[w_*]$. Since $k_{\text{RMS}}$ is invariant under translation and scaling of its argument, $k_{\text{RMS}}[w_*] = k_{\text{RMS}}[d]$, that is, the nondimensionalized RMS frequency is an observable property of the data, in the noise-free case at least.

APPENDIX C

ABSTRACT STRUCTURE OF THE GRADIENT

The expression \( \Theta \) for the derivative of the reduced objective $\tilde{J}_\alpha$ is the result of elementary manipulations, based on the explicit expression \( \Theta \) for the modeling operator $F$. In this appendix I give an alternative derivation that generalizes to extended inversion formulations for much less constrained physics. I will point out the additional reasoning necessary to reach
similar conclusions in these more complex instances of extended inversion, as presented for example in Stolk and Symes (2003); Stolk et al. (2009); Symes (2014); ten Kroode (2014); Huang et al. (2017, 2018a,b, 2019).

Recall that $w_\alpha[m;d]$ is the solution of the normal equation $20$. The reduced objective $\tilde{J}_\alpha$ is given by

$$
\tilde{J}_\alpha[m;d] = J_\alpha[m, w_\alpha[m;d]; d] = \frac{1}{2}( \|F[m]w_\alpha[m;d] - d\|^2 + \alpha^2\|Aw[m;d]\|^2 )
$$

(equation 28)

$$
= \frac{1}{2}( \|d\|^2 - \langle d, F[m]w_\alpha[m;d] \rangle ),
$$

(C-1)
after a little algebra.

As mentioned above, $F : M \times W \to D$ is not differentiable. Neither is $w_\alpha : M \times D \to W$, as follows immediately from the identity 25. However $Fw_\alpha : M \times D \to D$ is differentiable, hence so is $\tilde{J}_\alpha : M \times D \to \mathbb{R}$ thanks to the identity (C-1) under the conditions on the multiplier $a$ identified in Theorem 3.

For the remainder of this appendix, assume that $a \in C^\infty(\mathbb{R}) \cap L^\infty(\mathbb{R})$ and $a(t) = t$ for $|t| < \tau$ and $|a(t)| \geq \tau$ for $|t| \geq \tau$, with $\tau$ defined in equation 39.

Proposition 2, item [1], implies that

$$
F[m]w_\alpha[m;d] = F[m](F[m]^TF[m] + \alpha^2A^TA)^{-1}F[m]^Td.
$$

(C-2)
The normal operator and its inverse are multiplication operators, whereas $F$ and $F^T$ are scaled shift operators, inverse to each other except for scale. From this it is easy to see that the operator on the RHS of equation [C-2] is multiplication by a smooth function, with its arguments shifted by $mr$. Such an operator is smooth in $m$, hence so is $Fw_\alpha$.

In the other extended inversion settings mentioned at the beginning of this appendix, $F$ is a microlocally elliptic Fourier Integral Operator (FIO) (Duistermaat, 1996). The canonical relation takes on the role of the shift operator $t \to t - mr$, and has the properties necessary to conclude that the composition of $F$ and $F^T$ in both orders is pseudodifferential, at least in an open conic subset of the cotangent bundle. If $A^TA$ is pseudodifferential, then so is the normal operator. The inverse in the RHS of equation (C-2) must be replaced by a microlocal parametrix, and a smooth error added to the RHS. Thanks to Egorov’s Theorem (Taylor, 1981), a special case of the rules for composing FIOs, the operator on the RHS of equation (C-2) is a pseudodifferential operator whose symbol is algebraic in geometric optics quantities, hence depends smoothly on the coefficients in the wave equation, as does $Fw_\alpha$.

Since $\tilde{J}_\alpha$ and its gradient depend smoothly on their arguments, it is possible to derive an alternate expression assuming that $d \in H^1(\mathbb{R})$, then extend it by continuity to $d \in L^2(\mathbb{R})$. Note that if $w \in H^1(\mathbb{R})$, then differentiable, and

$$
(D(F[m]w)\delta m)(t) = F[m](Q[m]\delta m)w(t),
$$

(C-3)

where

$$
(Q[m]\delta m)w = -r\delta m \frac{dw}{dt}.
$$

(C-4)
That is, $DF[m]δm$ factors into $F[m]$ following $Q[m]δm$, where the latter a skew-adjoint differential operator of order 1, depending linearly on $\delta m$.

Assume that $d \in H^1(\mathbb{R})$. From equation $25$ it follows that $w[m;d] \in H^1(\mathbb{R})$ and moreover that $m \mapsto w_\alpha[m;d]$ is differentiable as a map from $\mathbb{R}^+$ to $L^2(\mathbb{R})$. From equation $28$

$$J_\alpha[m;d] = \frac{1}{2} ∥F[m]w[m;d] - d∥^2 + α^2 ∥Aw[m;d]∥.$$ 

whence $m \mapsto J_\alpha[m;d]$ is differentiable. A standard calculation invoking the normal equation $20$ shows that

$$dJ_\alpha[m;d]δm = \langle DF[m]w_\alpha[m;d], F[m]w_\alpha[m;d] - d \rangle$$

$$= \langle F[m](Q[m]δm)w_\alpha[m;d], F[m]w_\alpha[m;d] - d \rangle$$

$$= \langle (Q[m]δm)w_\alpha[m;d], F[m](F[m]w_\alpha[m;d] - d) \rangle$$

$$= -α^2 \langle (Q[m]δm)w_\alpha[m;d], A^T Aw_\alpha[m;d] \rangle$$

$$= α^2 \langle w_\alpha[m;d], (Q[m]δm)A^T Aw_\alpha[m;d] \rangle$$ (using antisymmetry of $Q$)

$$= α^2 \langle w_\alpha[m;d], [(Q[m]δm), A^T A]w_\alpha[m;d] \rangle + α^2 \langle (Q[m]δm)w_\alpha[m;d], A^T Aw_\alpha[m;d] \rangle$$

Rearranging,

$$dJ_\alpha[m;d]δm = \frac{1}{2} α^2 \langle w_\alpha[m;d], [(Q[m]δm), A^T A]w_\alpha[m;d] \rangle. \quad (C-5)$$

Since $Q$ is a differential operator of order 1, and $A^T A$ an operator of order 0, the commutator has order 0. That is, the RHS of equation $[C-5]$ defines a continuous quadratic form in $d$ with respect to the $L^2$ norm. As shown above, the same is true of the left hand side. Therefore their extensions by continuity to $d \in L^2(\mathbb{R})$ are the same, and the identity $[C-5]$ holds for $d \in L^2(\mathbb{R})$.

Making the identity $[C-5]$ explicit by means of equation $[C-4]$ and $A^T A = I^2$, substituting the expression $25$ for $w_\alpha[m;d]$, and rearranging: one obtains precisely the expression $32$.

For the more general contexts mentioned earlier, a similar derivation is possible. The analog of $Q$ is a pseudodifferential operator of order 1, which is essentially (to leading order) skew-symmetric, and the more general version of the identity $[C-5]$ is approximate, with lower-order (smoother) error terms.

These ideas were used in ten Kroode (2014); Symes (2014) to show that the analog of $J_\alpha$ is tangent to second order to a convex quadratic form, related to the Hessian of an objective formulated in terms of traveltime. This relation explains the ability of the variant of extended inversion studied in the cited references to recover kinematically accurate velocity fields. Of course, that is what is shown in detail in this paper for the very simple model problem studied here.