Robust estimation of primaries by sparse inversion via one-norm minimization

Tim T.Y. Lin[1] *and Felix J. Herrmann*[1]

ABSTRACT

Even though contemporary methods for the removal of multiple events in seismic data due to a free-surface are built upon reciprocity relationships between wavefields, they are often still implemented as prediction-subtraction processes. The subtraction process does not always compensate for imperfect prediction of multiple events, and itself often leads to distortion of primary events. A recently proposed method called Estimation of Primaries by Sparse Inversion avoids the subtraction process altogether by directly prediction the primary impulse response as a collection of band-limited spikes under sparsity-regulated wavefield inversion approach. Although it can be shown that the correct primary impulse response is obtained through the sparsest possible solution, the Estimation of Primaries by Sparse Inversion algorithm was not designed to seek such a solution, instead depending on a predetermined degree of sparsity as an inversion parameter. This leads to imperfect multiple rejection when the sparsity is overestimated, and problems with recovering late primary events when it is underestimated. In this paper, we propose a new algorithm where we make obtaining the sparsest solution our explicit goal. Our approach remains a gradient-based approach like the original algorithm, but is in turn derived from a new optimization framework based on an extended basis pursuit denoising formulation. We show that the sparsity-minimizing objective of our formulation enables it to operate successfully on a wide variety of synthetic and field marine dataset without excessive tweaking of inversion parameters. We also demonstrate that Robust EPSI produces a more artifact-free impulse response compared to the original algorithm, which has interesting implications for broadband seismic applications. Finally we demonstrate through field data that recovering the primary impulse response under transform domains can significantly improve the recovery of weak primary late arrivals, without appreciable change to the underlying algorithm.

INTRODUCTION

Because multiple removal is often thought of as a prediction-subtraction process, it is easy to overlook that for the past decade or so the seismic industry have gradually shifted towards demultiple methods that are actually solutions to integral equations under various degrees of disguise. The unifying framework of what are commonly called "data-driven" approaches is the Raleigh-Helmholtz reciprocity relationship between the field-measured wavefield data and a hypothetical multiple-free version of it (Fokkema and van den Berg, 1993; Frijlink

---

[1] Seismic Laboratory for Imaging and Modeling, Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, V6T 1Z4, BC, Canada

The University of British Columbia Technical Report. TR-2012-02, March 12, 2012

et al., 2011). This relation can be expressed as wavefield integral equations with suitable boundary conditions, and the demultiple process equates to solving them.

For removal of multiple events due to the free surface, Surface-Related Multiple Elimination (SRME, Verschuur, 1992; Berkhout and Verschuur, 1997; Verschuur and Berkhout, 1997) is one of the most well-known example of the data-driven techniques, and also the most commonly regarded as a prediction-subtraction process. Although its practicality has led to its widespread adoption, the problems with the prediction-subtraction paradigm are well known. Namely, the prediction process is often plagued with undersampling issues, and the adaptive subtraction requires estimating a shaping filter for the predicted multiple events to match its counterpart in the data. The filter estimation process often utilizes a minimum-energy criteria on the primary wavefield that results in corrupted primary signals when primary and multiple events overlap (van Borselen et al., 2003; Guitton and Verschuur, 2004).

To avoid adaptive subtraction altogether, a proposal to directly solve for the multiple-free wavefield from the SRME integral equation (Frijlink et al., 2011) is presented in van Groenestijn and Verschuur (2009a) as a technique called Estimation of Primaries by Sparse Inversion (EPSI). Because the SRME equation treats the surface-free Green's function and the source signature as two separate unknowns, EPSI needs to invert for both at the same time. This results in an inversion that is analogous to blind deconvolution, as opposed to the typical SRME procedure where solution of the two unknowns are implicitly decoupled through the prediction-subtraction process. The main insight of EPSI is that modeling the surface-free Green's function as impulsive spikes is a strong enough prior to overcome the inherent local-minima present in this inversion. EPSI was shown to be a significant improvement over SRME in marine data acquired over shallow water (van Groenestijn and Verschuur, 2009b,c; Baardman et al., 2010; Savels et al., 2010; Toxopeus et al., 2011), and had been adapted to work on passive seismic data (van Groenestijn and Verschuur, 2010a), blended data (van Groenestijn and Verschuur, 2009d; Lin and Herrmann, 2009; van Groenestijn and Verschuur, 2011), and ocean-bottom cable data (van Groenestijn and Ross, 2011). Before the introduction of EPSI, wavefield inversion approaches to multiple removal either required exact knowledge of the source wavelet (van Borselen et al., 1996) or requires high fidelity recording of the direct waves (Ziolkowski et al., 1999; Amundsen, 2001; Majdanski et al., 2011).

In Guitton and Verschuur (2004) and van Groenestijn and Verschuur (2008), it was shown that the minimizer of a sparsity measure on the primary wavefield yields the correct demultipled result during adaptive subtraction. Therefore, it would seem natural for EPSI to seek the sparsest possible solution for the surface-free Green's function. However, as we show in this paper, the EPSI algorithm as proposed in van Groenestijn and Verschuur (2009a) does not automatically seek the sparsest solution, and leaves the burden of determining the appropriate level of sparsity to the practitioner. Furthermore, as we will show, the sparsity control parameters in EPSI do not strongly correlate to the final sparsity level of the solution, so even if the structure of the subsurface is perfectly known (i.e., the number of reflectors) this information cannot be effectively used in the inversion. This shortcoming impacts the effectiveness of the EPSI in terms of its flexibility, reliability, and quality of its solutions, particularly in its ability to recover late events of low amplitude.

The central purpose of this paper is to propose a new approach to EPSI that *does* attempt to seek the sparsest surface-free Green's function in a computationally efficient way.

In doing so we devise a reformulation of EPSI as a modification of the basis pursuit optimization problem (Chen and Donoho, 1994; Chen et al., 2001), and design a custom algorithm to find its solution. This results in a method that is applicable to a wide variety of datasets without readjusting inversion parameters. Also, the fidelity of the recovered surface-free Green's function is markedly improved compared to the original EPSI approach. Because we regard EPSI as a standard optimization problem, this reformulation takes advantage of recent advances in large-scale sparse recovery techniques (specifically that of van den Berg and Friedlander, 2008, 2011) for fast convergence, and is flexible enough to allow numerous extensions without modification to the algorithm, such as finding solutions in arbitrary transform domains and dealing with simultaneously acquired data.

It is important to note beforehand that in this paper we do not regard one important aspect of EPSI, which is the simultaneous reconstruction of missing near-offset data throughout the inversion. Rather, our scope is limited to in-depth analysis of the sparse regularization aspects of EPSI. That said, we do feel that enhancing the robustness of the near-offset reconstruction feature is also an important topic, and that the optimization framework we introduce in this paper will benefit its future investigation.

In the following sections, we first review the central premise of EPSI, and formulate it as an optimization problem to highlight its potential limitations. We then start our reformulation of EPSI by turning it into a basis pursuit problem with multiple unknowns. We then discuss how to use a block-coordinate descent scheme to efficiently solve this extended objective. The particular advantage of our algorithm is that, while exploiting the Pareto curve (van den Berg and Friedlander, 2008; Hennenfent et al., 2008; Daubechies et al., 2008) of basis pursuit problems, it also ensures that the source signature are always estimated with sparse approximations to the surface-free Green's function. We illustrate the improvements of this approach over the original EPSI algorithm on synthetic datasets. Finally, we use field datasets to demonstrate how recovery of primary late-arrivals can further be improved by seeking solutions in alternative wavefield representations.

REVIEW OF THE EPSI ALGORITHM

Our goal in this section is to introduce the methodology of EPSI as presented in van Groenestijn and Verschuur (2009a) and restate the mathematical formulation in a manner that more readily expresses its underlying optimization problem. This not only enables a more concrete notion of the original algorithm's shortcomings as described in the introduction, but also form the basis of our extensions.

We will start with the familiar mathematical language typically used in SRME-related literature, namely the detail-hiding "data matrix" notation similar to Berkhout and Pao (1982), to describe EPSI. The detail-hiding notation describes wavefields as monochromatic slices arranged into a matrix that have columns representing common shot gathers. Uppercase variables denote matrices or linear operators. In order to avoid subscripting by ω we use hatted quantities to represent monochromatic, frequency-dependent variables. An important feature of this notation is that the matrix multiplication of two hatted uppercase data-matrix variables implies multidimensional wavefield convolutions in the physical domain.

We introduce two wavefields pertinent to our discussion: $G(\mathbf{x}_r, \mathbf{x}_s, t)$ represents the desired surface-free Green's function (also called the primary impulse response) due to a dipole source, and may contain internal multiples. $P(\mathbf{x}_r, \mathbf{x}_s, t)$ represents the total upgoing pressure wavefield data, possibly approximated through deghosting (or wavefield decomposition, if the data allows). $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{P}}$ therefore represent their discretized monochromatic slices arranged in a data matrix, such that the monochromatic shot gathers are organized in the columns. We use Q to represent the source properties corresponding to our data P. Throughout the scope of this paper we assume Q to be the same for all shots and traces, such that $\widehat{\mathbf{Q}}$ can simply be modeled by a frequency-dependent scaling $\widehat{\mathbf{Q}} = q(\omega)\mathbf{I}$ with $\mathbf{I}$ the identity matrix. The matrix $\widehat{\mathbf{R}}$ models a surface reflection operator that is assumed to be $-\mathbf{I}$ (i.e., an ideal reflector) for the rest of this paper. The main physical relationship that underlies EPSI is thus expressed as

$$\widehat{\mathbf{P}} = \widehat{\mathbf{G}}(\widehat{\mathbf{Q}} + \widehat{\mathbf{R}}\widehat{\mathbf{P}}). \tag{1}$$

Note that in relation to typical SRME parlance, $\widehat{\mathbf{G}}\widehat{\mathbf{Q}} = \widehat{\mathbf{P}}_0$ is the monochromatic primary wavefield that does not contain multiples due to the free-surface, and $\widehat{\mathbf{Q}}^\dagger\widehat{\mathbf{R}}$ (superscript $\dagger$ denotes pseudoinverse) plays roughly the same role as the free-surface operator (Berkhout and Verschuur, 1997).

On the left-hand side of equation 1 is $\widehat{\mathbf{P}}$ the observed data, while the right-hand side describes how to model this data using the two unknown quantities $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{Q}}$, using $\widehat{\mathbf{P}}$ as essentially an integration kernel. This describes a discrete inhomogenous integral equation. Due to noise, imperfect up-down wavefield decomposition, and inaccuracies in physical assumptions (such as the presence of cross-line dips when assuming single-line survey geometry), the data will not lie entirely in the range of the model. Solving equation 1 for $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{Q}}$ thus requires an inversion that minimizes the energy of the mismatch between the left-hand and right-hand sides over all frequencies. In addition, the combined unknowns $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{Q}}$ exceeds the size of $\widehat{\mathbf{P}}$ and results in an underdetermined system, so the inversion requires additional constraints on the solution. Therefore, we can summarize EPSI as an inversion procedure obtained by choosing a sparsity constraint on G in the *time* domain, motivated by the spiky appearance of the Green's function. Additionally, EPSI ensures that Q is also supported in time within a short interval, to avoid having Q act like a predictive deconvolution filter (i.e., fitting multiple events into Q).

EPSI in optimization form

To facilitate further discussion we now express EPSI as an optimization problem in the form of minimizing an objective, with the unknowns represented by vector quantities. For this we need to introduce the convention of lower-case symbols representing vectorized quantities. For example, $\mathbf{p} = \mathrm{vec}\left(P(\mathbf{x}_r, \mathbf{x}_s, t)\right)$ is the discretized observed data arranged in a vectorized fashion such that the hierarchy of contiguous physical dimensions are, respectively, the time axis, the receiver position axis, and the source position axis. We define the EPSI forward-modeling operator M as the right-hand side of equation 1 acting over all frequencies on $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{Q}}$. Because the constraints we impose are in the time domain, it is more convenient for M to also act on unknowns in the time domain. Formally, M can be written as

$$M(\mathbf{g}, \mathbf{q}) := \mathbf{F}_\mathrm{t}^H \mathrm{Blockdiag}_{\omega_1 \cdots \omega_{nf}} \left((q(\omega)\mathbf{I} - \widehat{\mathbf{P}})^T \otimes \mathbf{I} \right) \mathbf{F}_\mathrm{t}\mathbf{g}, \tag{2}$$

where the block-diagonal elements vary over all frequencies. Vectors $\mathbf{g}$ and $\mathbf{q}$ are respectively the time-domain surface-free Green's function and source signature. We denote adjoints with superscript H and matrix transposes with superscript T. The symbol $\otimes$ defines a Kronecker product that, in this case, simply helps reformulate the matrix-matrix operations in equation 1 into matrix-vector operations. The matrix $\mathbf{F}_{\mathrm{t}}$ represents a Fourier transform along the time axis that also organizes the resulting monochromatic wavefields such that $\mathbf{F}_{\mathrm{t}}\mathbf{g} := [\widehat{\mathbf{g}}_{\omega_1}, \widehat{\mathbf{g}}_{\omega_2}, ..., \widehat{\mathbf{g}}_{\omega_{nf}}]^T$, while the adjoint operation $\mathbf{F}_{\mathrm{t}}^H$ on the left brings the wavefield back to the time domain, sorted in a manner that is consistent with the physical datum.

The definition of the forward-modeling operator $M(\mathbf{g}, \mathbf{q})$ in equation 2 allows us to express the relation in equation 1 as $\mathbf{p} = M(\mathbf{g}, \mathbf{q})$, this time over the whole dataset in physical dimensions instead of monochromatic slices. Note that if we fix either of the arguments of $M(\mathbf{g}, \mathbf{q})$, it becomes a linear operator with respect to the remaining variable; therefore M is a bilinear operator. We denote the linearizations with respect to $\mathbf{g}$ and $\mathbf{q}$ as $\mathbf{M}_{\tilde{g}}$ and $\mathbf{M}_{\tilde{q}}$, and define the partial derivatives $\partial M/\partial \mathbf{q}$ and $\partial M/\partial \mathbf{g}$ evaluated at $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{q}}$ (using the overhead tilde to denote approximations to actual physical entities, or empirically determined quantities):

$$\begin{aligned} \mathbf{M}_{\tilde{q}}\mathbf{g} &:= \left(\frac{\partial M}{\partial \mathbf{g}}\right)_{\tilde{q}} \mathbf{g} = M(\mathbf{g}, \tilde{\mathbf{q}}), \\ \mathbf{M}_{\tilde{g}}\mathbf{q} &:= \left(\frac{\partial M}{\partial \mathbf{q}}\right)_{\tilde{g}} \mathbf{q} = M(\tilde{\mathbf{g}}, \mathbf{q}). \end{aligned} \tag{3}$$

Based on this description, we can use our new notation to express the EPSI procedure as equivalent to solving the following standard-form optimization problem:

$$\underset{\mathbf{g},\, \mathbf{q} \in \Lambda}{\text{minimize}} \, \|\mathbf{p} - M(\mathbf{g}, \mathbf{q})\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_0 \leq \rho, \tag{4}$$

where the ℓ_0-“norm” $\|\mathbf{g}\|_0$ measures the number of non-zero impulsive events in $\mathbf{g}$. Usage of the ℓ_0-“norm” notation instead of the more intuitive name of *sparsity* or *cardinality* is currently standard language in applied mathematics and computer science literature. The parameter ρ controls the sparsity of the solution for the surface-free Green's function. We use Λ to express the set of candidate source signatures $\mathbf{q}$ that satisfy any desired or a-priori constraints we wish to impose, e.g., decay rates in time and frequency, as well as phase properties. In this paper we only impose that $\mathbf{q}$ is constrained in a short time window around $t = 0$, i.e., $q(t) = 0$ for $|t| \geq T$ when q is represented in the time domain. Also, $\mathbf{q}$ is allowed to be anti-causal to account for any possible time shifts in the recorded data. As shown in van Groenestijn and Verschuur (2010b), this model is easily extensible to accommodate more complex source properties by expanding the degree of freedom. It is important to note that, due to the structure of $M(\mathbf{g}, \mathbf{q})$, our final estimate of $\mathbf{q}$ as a process of an inversion will necessarily capture all preprocessing on $\mathbf{p}$ than can be modeled by a filter of the same structure as $\mathbf{q}$. With the assumptions listed above, this means that a estimate of $\mathbf{q}$ will also capture global scalings, spectrum shaping, and various other preprocessing techniques commonly applied to the data.

The term $\widehat{\mathbf{G}}\widehat{\mathbf{Q}}$ in equation 1 (and thus in the forward-modeling operator of the inversion) represents a convolution of two unknown quantities. This bears resemblance to a blind deconvolution problem, which attempts to recover an unknown signal that is observed only through a likewise unknown linear filter. Although blind deconvolution problems are famous for their ill-behavior, in reality we expect our EPSI to be more well posed, as we can not only exploit a sparsity regularizer on $\mathbf{g}$, which comes from our prior assumption of the desired Green's function's spiky appearance, but also we have a second term in the modeling function that only depends on $\mathbf{g}$. Nevertheless, to solve this problems we need algorithms that can deal with the presence of multiple unknowns.

Optimizing over multiple variables

The most straightforward strategy for dealing with multiple unknowns is to simply lump them together into a single aggregate variable, and deal with the new problem using the same standard approaches for single unknowns. However, this approach has immediate difficulties when the different unknowns are strongly coupled and differ in scale (You and Kaveh, 1996), such as in our case of jointly inverting for both a wavefield and its source signature. We also have completely different constraints for $\mathbf{g}$ and $\mathbf{q}$, which exasperates the problem of working on them in unison. Alternatively, we can decrease the objective by iteratively optimizing over $\mathbf{g}$ with $\mathbf{q}$ fixed, then $\mathbf{q}$ with $\mathbf{g}$ fixed. This is an example of the cyclic block-coordinate descent methods (Bezdek et al., 1987).

We apply this approach to solve equation 4 and demonstrate how this subsequently leads to the original EPSI algorithm. Define two new objective functions that are formed by taking the quadratic objective in equation 4 and fixing one of the variables at some value (as discussed above this also linearizes $M(\mathbf{g}, \mathbf{q})$ around that variable), which we write as $f_{\tilde{g}}(\mathbf{q}) := \frac{1}{2}\|\mathbf{p} - \mathbf{M}_{\tilde{g}}\mathbf{q}\|_2^2$ and $f_{\tilde{q}}(\mathbf{g}) := \frac{1}{2}\|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{g}\|_2^2$. Algorithm 1 outlines what a typical cyclic coordinate descent approach would conform to.

Algorithm 1 Cyclic block-coordinate descent for EPSI

```
repeat
    $\tilde{\mathbf{g}}_{k+1} \leftarrow \text{argmin}_{\mathbf{g}}\ f_{\tilde{q}_k}(\mathbf{g})\quad \text{s.t.}\quad \|\mathbf{g}\|_0 \le \rho$
    $\tilde{\mathbf{q}}_{k+1} \leftarrow \text{argmin}_{\mathbf{q}}\ f_{\tilde{g}_{k+1}}(\mathbf{q})\quad \text{s.t.}\quad \mathbf{q} \in \Lambda$
    $k \leftarrow k + 1$
until $\|\mathbf{p} - M(\tilde{\mathbf{g}}_k, \tilde{\mathbf{q}}_k)\|_2$ converges to some criteria
```

Algorithm 1 can also be *inexact* if we relax the two minimization problems and instead simply seek *some* decrease in the objective. The original EPSI algorithm uses just a single gradient update to form the subsequent $\tilde{\mathbf{g}}_{k+1}$ and $\tilde{\mathbf{q}}_{k+1}$, where k is the iteration counter, and therefore is a typical example of such an inexact cyclic block-coordinate descent method.

We now express the original EPSI algorithm in our notation. To begin, a descent direction for $f_{\tilde{q}_0}(\mathbf{g})$ is then obtained though its gradient at $\tilde{\mathbf{g}} = 0$. For this first iteration $k = 1$, if we start with $\tilde{\mathbf{q}}_0 = 0$, this gradient is equivalent to the autocorrelation of $\mathbf{p}$. To ensure $\|\tilde{\mathbf{g}}\|_0 \le \rho$ at completion of the algorithm, this gradient is passed to some sparsifying operator $\mathcal{S}$ that zeroes most of its elements. The resulting sparse gradient is then scaled by a line search and used to form $\tilde{\mathbf{g}}_1$. Next, we form the gradient for $f_{\tilde{g}_1}(\mathbf{q})$ and update $\tilde{\mathbf{q}}_1$ in a straightforward fashion (possibly with an implicit projection to ensure $\tilde{\mathbf{q}} \in \Lambda$). The process then iterates, stopping when a chosen criteria is satisfied, such as an iteration limit or a target misfit. Algorithm 2 provides a summary of this procedure, with α and β denoting scalings that ensure the objective decreases sufficiently (i.e., line-search scalings).

Algorithm 2 Original EPSI algorithm

```
$\tilde{\mathbf{g}}_0$, $\tilde{\mathbf{q}}_0 \leftarrow$ zero vector
$k \leftarrow 0$
repeat
    $\tilde{\mathbf{g}}_{k+1} \leftarrow \tilde{\mathbf{g}}_k + \alpha\mathcal{S}(\nabla f_{\tilde{q}_k}(\mathbf{g}_k))$
    $\tilde{\mathbf{q}}_{k+1} \leftarrow \tilde{\mathbf{q}}_k + \beta\nabla f_{\tilde{g}_{k+1}}(\mathbf{q}_k)$
    Project $\tilde{\mathbf{q}}_{k+1}$ to satisfy $\tilde{\mathbf{q}}_{k+1} \in \Lambda$
    $k \leftarrow k + 1$
until $\|\mathbf{p} - M(\tilde{\mathbf{g}}_k, \tilde{\mathbf{q}}_k)\|_2$ converges to some criteria, or $k = k_{\mathsf{tot}}$
```

Unpredictable nature of sparse updates

The specifics of the sparsifying operator $\mathcal{S}$ greatly influence the behavior of this algorithm. In order to both satisfy the terminating sparsity constraint $\|\tilde{\mathbf{g}}\|_0 \leq \rho$ and decrease the error objective $f_{\tilde{q}_k}(\tilde{\mathbf{g}})$ as much as possible, $\mathcal{S}$ selects and keeps the largest amplitude ρ/k_{tot} elements of the update for $\tilde{\mathbf{g}}$, where k_{tot} is the number of total gradient iterations used for the problem, while the remaining elements are set to zero. Of course k_{tot} is rarely known beforehand, so a practitioner would estimate an empirical per-update sparsity limit $\tilde{\rho}$ in place of ρ/k_{tot}. Furthermore, EPSI subscribes to the argument that seismic reflection events usually appear in all traces, and therefore $\mathcal{S}$ is typically carried out independently on a per trace basis, thus $\tilde{\rho}$ can equivalently be specified by the number of largest-amplitude events kept from each trace.

One of the biggest challenges of the original EPSI algorithm is that it is difficult to justifiably choose a $\tilde{\rho}$ for any non-trivial dataset. Intuitively, this parameter seems to indicate the number of expected reflection events in the Green's function, and would thus correlate to the expected number of subsurface reflectors. However, because the number of gradient updates to recover a particular event with the right amplitude can vary wildly, even in extreme cases where the number of subsurface reflectors are entirely known it is still difficult to choose $\tilde{\rho}$ such that the number of events in our final estimate of $\tilde{\mathbf{g}}$ match our expectations.

Furthermore, a poor choice of $\tilde{\rho}$ can have serious consequences for the output of the algorithm. If $\tilde{\rho}$ is allowed to be too large, then $\tilde{\mathbf{g}}$ may no longer resemble a impulsive Green's function after a few iterations. This will ultimately result in poor multiple removal, and we demonstrate in the first figure of the Example section. In practice, this outcome is prevented by setting $\tilde{\rho}$ to a very small value (typically 2 to 5 events per trace are kept). Additionally, to avoid any energy being put into $\tilde{\mathbf{g}}$ at the location of multiples, a time-window, also empirically chosen, is applied on the updates. This time-window is allowed to grow with the iteration count, again in an arbitrary fashion, as the earlier events are recovered sufficiently accurately to allow for later events to be recovered.

On the other hand, since a small value of $\tilde{\rho}$ would only allow a small number of events in $\tilde{\mathbf{g}}$ to be updated at each gradient iteration, the practitioner is forced to aggressively ramp the number of iterations in order to recover late low-amplitude events that may be paramount to imaging reservoirs. In the effort to lower our error objective as quickly as possible, large-amplitude early events tend to dominate this $\tilde{\rho}$-event allowance of each update at the early stages of the algorithm, and the weaker events are only updated at higher iteration counts. Because each gradient update for $\tilde{\mathbf{g}}$ is computationally comparable to applying a focal transform on the wavefield (i.e., physical wavefield correlations, see van Groenestijn and Verschuur, 2009a, for detailed expressions), the additional gradient costs required are substantial. Thus a typical iteration-constrained application of EPSI, these weaker events may therefore not even be present in the solution. Further exasperating this issue is the trace-independent nature of $\mathcal{S}$. Because reflection events are not in general recovered at the same rate across all traces, $\mathcal{S}$ does not in turn guarantee that each update will affect the same events across all traces. This tends to affect the late low-amplitude events the most, resulting in reflections that appear abruptly on some traces but not on adjacent ones. See Figure 1 for an illustration of this issue, and notice the discontinuous appearance of events past 2 s.

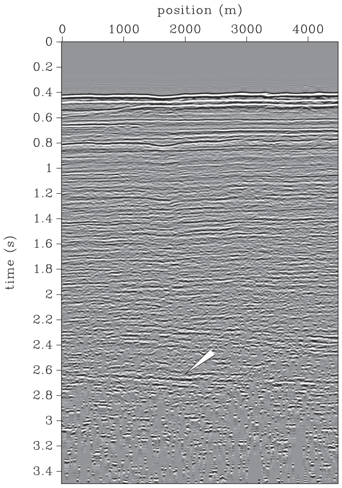


Figure 1: An example of the event continuity issue that plagues direct primary estimates of EPSI. Figure shown is a 200 m offset panel of the primary wavefield directly produced by EPSI, with late events amplified by automatic gain control. Strong sections of late events around 3 s are recovered, but the events appear broken and discontinuous. Automatic gain control was used for the plot.

ROBUST REFORMULATION OF EPSI

The aforementioned complications due to reliance on $\mathcal{S}$ and sparsity-per-update as an optimization parameter are avoided if we instead place the sparsity measure of $\tilde{\mathbf{g}}$ in the objective, which translates to seeking the sparsest solution for the surface-free Green's function. This results in a more robust and flexible formulation of EPSI, but comes with a necessary increase in complexity of the algorithm. Our main contribution is to demonstrate that this can be accomplished in a computationally efficient manner by treating EPSI as an extended basis pursuit (Chen et al., 2001) problem for $\mathbf{g}$ that also simultaneously solves for the source signature $\mathbf{q}$. The final algorithm is outlined in Algorithm 3, and in the following sections we will discuss the measures necessary to arrive at this new approach from the original EPSI formulation.

From sparse regularization to basis pursuit

To avoid having to specify the sparsity level directly, we begin by interchanging the role of the sparsity measure $\|\mathbf{g}\|_0$ with the observation-prediction misfit in equation 4. This results in a new problem

$$\underset{\mathbf{g},\,\mathbf{q}\in\Lambda}{\text{minimize}}\,\|\mathbf{g}\|_0 \quad \text{subject to} \quad \|\mathbf{p}-M(\mathbf{g},\mathbf{q})\|_2 \leq \sigma. \tag{5}$$

To minimize $\|\mathbf{g}\|_0$ means that we seek the smallest number of primary events that explains the recorded wavefield (up to a misfit of σ). This is closer to the arguments for sparse regularization of the primary Green's function laid out in van Groenestijn and Verschuur (2008), which precedes the formal introduction of EPSI. There, a sparsity measure on the primary wavefield is shown to be a better *objective function* compared to the energy measure for the purpose of resolving overlapping primary and multiple events in adaptive subtraction for a purely impulsive source. Compared to equation 4, the parameter in this inversion problem is moved to a more physically intuitive energy of the misfit between the recorded and modeled wavefield. An immediate benefit is that the smallest recoverable event is now decoupled from the way we impose sparsity on $\tilde{\mathbf{g}}$, and now purely depends on the noise-floor σ of the data. This framework also allows a number of common data metrics, such as the relative noise-level in the dataset, to help form a general estimate of σ.

To adapt the block-coordinate descent scheme described in Algorithm 1 to this new inversion problem, we simply replace the two subproblems in Algorithm 1 with the following problems:

$$\tilde{\mathbf{g}}_{k+1} \leftarrow \underset{\mathbf{g}}{\text{argmin}}\,\|\mathbf{g}\|_0 \quad \text{subject to} \quad f_{\tilde{q}_k}(\mathbf{g}) \leq \sigma \tag{6a}$$

$$\tilde{\mathbf{q}}_{k+1} \leftarrow \underset{\mathbf{q}}{\text{argmin}}\,f_{\tilde{g}_{k+1}}(\mathbf{q}) \quad \text{subject to} \quad \mathbf{q}\in\Lambda, \tag{6b}$$

where the objective function in the first subproblem (6a) has been replaced so that the minimization is over $\|\mathbf{g}\|_0$. Note that the subproblem (6b) related to updating the estimate of the source wavelet $\tilde{\mathbf{q}}_{k+1}$ remain untouched compared to Algorithm 1.

The fatal drawback of working with the new optimization problem of equation 5 is that it is a *combinatorial* problem, due wholly to difficulties with the $\|\mathbf{g}\|_0$ objective; in fact, there exist no deterministic methods to solve problems of this type apart from a search through all the possible solutions. It is possible, however, to obtain approximate solutions through various heuristics. Resigning ourselves to simply guessing at the optimal sparsity level leads to the original EPSI problem in equation 4, along with all the drawbacks just discussed in the last section. Instead, we take the approach of posing a basis pursuit denoising (BPDN, Chen et al., 2001; van den Berg and Friedlander, 2008) problem over $\mathbf{g}$. BPDN is a well-studied convex problem (therefore the only local minimum is the global minimum) which remains computationally tractable to solve exactly after sufficient iterations, even for large-scale applications. We will refer to this approach as a convexification of the EPSI problem with respect to $\mathbf{g}$. It allows us to retain the notion of "seeking the sparsest solution" without having to solve a combinatorial problem, and thereby permitting large-scale seismic applications.

It is well-known that, for most large linear systems with redundant solutions, the minimum ℓ_1-nom solution to the system is a close approximation of its sparsest solution (Donoho, 2006). The accuracy of this approximation depends on the sparsest possible (minimum ℓ_0-norm) solution to the system; the sparser it is in relation to the total size of the system, the better the minimum ℓ_1-norm approximation would be. For the purposes of EPSI, we generally expect $\|\mathbf{g}\|_0 \ll n$, where $n := n_t \times n_r \times n_s$ is the total number of time samples in the entire seismic record. Therefore, provided that we have somehow obtained an accurate estimate of $\mathbf{q}$, we can expect the following problem to recover the desired surface-free Green's function in the noise-free case:

$$\underset{\mathbf{g}}{\text{minimize}}\ \|\mathbf{g}\|_1 \quad \text{subject to} \quad \mathbf{p} = \mathbf{M}_q\mathbf{g}. \tag{7}$$

Problems of this form are usually called basis pursuit (BP). It turns out that the relationship of the BP solution to the *sparsest* solution also remains stable under the presence of noise in $\mathbf{p}$ (Fuchs, 2005; Candès et al., 2006b; Donoho, 2006; Tropp, 2006), provided that the equality constraint is relaxed to $\|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{g}\|_2 \leq \sigma$, which results in the aforementioned BPDN approach. Because of the non-differentiable nature of the ℓ_1-norm objective, BPDN problems are inherently more difficult to solve than differentiable problems such as least-squares. Although BPDN is still amenable to many classical optimization approaches, challenges still exist in bringing convergence rate and scalability to the same order as venerable least-squares solvers. Propitiously, computationally efficient solvers for BPDN is currently a very active area of research due in part to its critical role in the fields of compressive sensing and large-scale machine learning.

Our reformulation is not complete without considering how the BPDN approach for $\mathbf{g}$ fits into the cyclic block-coordinate descent method for the general EPSI problem. Specifically, the main concern is with balancing the relative evolution of $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{q}}$, which is important for the practicality of the algorithm (see page 136 of Biggs, 1998, and related discussion in our section titled "Initializing the algorithm"). Recall that EPSI forms estimations of $\mathbf{q}$ by matching the data against a spiky impulse response. In that regard the original EPSI described in Algorithm 2 is immensely flexible in terms of when $\tilde{\mathbf{q}}$ can be updated, because $\tilde{\mathbf{g}}$ is only changed with sparse updates. This behaviour is unfortunately no longer a given if we are to adopt the BPDN approach, as many BPDN solvers only form sparse solutions upon convergence. However, in the early stages of the algorithm when $\tilde{\mathbf{q}}$ is expected to be inaccurate, solving an expensive BPDN problem for a similarly rough estimate of $\mathbf{g}$ is wasteful in terms of computation resources. To resolve this issue we need develop a concrete notion of sparse approximations to a BPDN solution, which can be obtained by analyzing the structure of the Pareto trade-off curve of the BPDN problem, as we discuss in the next section.

Intermediate estimates on the Pareto curve

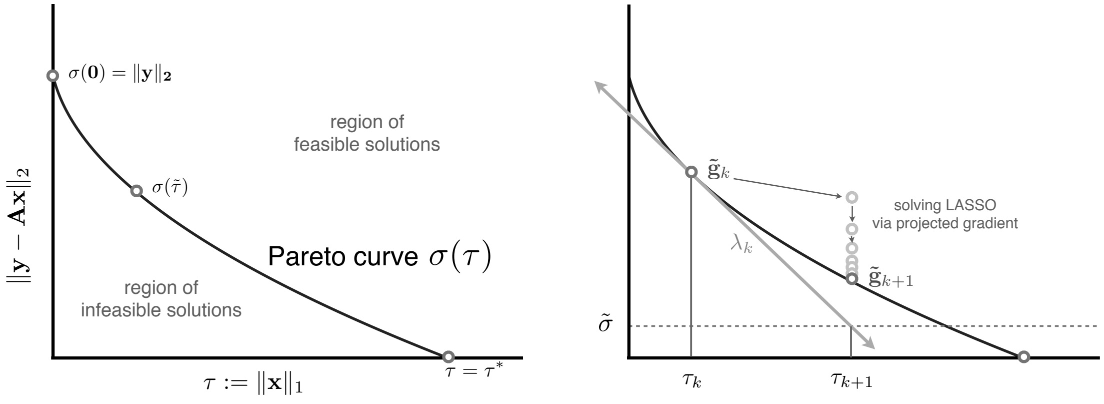
The Pareto curve of a BPDN problem characterizes its *optimal* solutions; it relates any choice of a target misfit σ to its corresponding optimal ℓ_1-norm objective. Because the BPDN problem is convex, this relation is unique, i.e., for any given observation vector $\mathbf{y}$ in the range of a linear system $\mathbf{A}$, the solution to the sparse recovery problem $\min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \tilde{\sigma}$, with $\tilde{\sigma}$ a predetermined target misfit, will have a unique minimum ℓ_1-norm of $\tilde{\tau}$. For reasons that will become evident, we express the Pareto curve as a function of τ, defined as

$$\sigma(\tau) := \underset{\mathbf{x}}{\text{minimize}}\ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau. \tag{8}$$

Note that $\sigma(\tau)$ takes on the value of the optimal objective, and not the solution vector. See Figure 2(a) for an illustration of the Pareto curve in a 2d plane of possible solutions characterized with $\|\mathbf{x}\|_1$ on the abscissae and $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2$ on the vertical axis.

The idea of looking at the Pareto curve for seismic inversion problems is not new. It was described in Hennenfent et al. (2008) as a way to characterize the various optimization routines for geophysical compressive sensing problems, and in Daubechies et al. (2008) and van den Berg and Friedlander (2008) as the basis of novel solvers for linear ℓ_1 problems. Specifically, van den Berg and Friedlander (2008) proved that the Pareto curve as defined here (as a function of τ) is convex, decreasing, and continuously differentiable for $\tau \in [0, \tau^*]$, where τ^* is the optimal objective of the BP problem $\min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\mathbf{y} = \mathbf{A}\mathbf{x}$. Furthermore, given an optimal solution $\tilde{\mathbf{x}}$ at any point on the Pareto curve, the slope λ of the Pareto curve at that point is given by the closed form expression $\lambda = -\|\mathbf{A}^H\mathbf{r}\|_\infty / \|\mathbf{r}\|_2$, where $\mathbf{r}$ is the residual vector $\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}$, and the ℓ_∞ norm is equal to the largest absolute value amongst all the elements of the vector.

These properties allows us to find the optimal ℓ_1-norm $\tilde{\tau}$ for any BPDN problem, by using Newton's method to solve $\sigma(\tilde{\tau}) - \tilde{\sigma} = 0$ (van den Berg and Friedlander, 2008, 2011). Refer to Figure 2(b) for a depiction of this procedure. Starting with any solution that lies on the Pareto curve $\tilde{\mathbf{x}}_k$, with ℓ_1-norm of τ_k and mismatch σ_k (we overload the iteration counter k to keep track of the Newton's method iterations), we iterate over subsequent guesses of the root using the rule $\tau_{k+1} = \tau_k + \Delta\tau_k$ where $\Delta\tau_k = -(\sigma_k - \tilde{\sigma})/\lambda_k$.



(a) (b)

Figure 2: Illustrations of the Pareto curve. Panel **(a)** shows the Pareto curve in the parameter space of possible solutions. Panel **(b)** depicts one iteration of the root-finding procedure to find the minimum ℓ_1 norm that gives the target energy misfit $\tilde{\sigma}$.

Since obtaining the slope in the Newton root-finding scheme requires knowing an optimal solution $\tilde{\mathbf{x}}$ on the Pareto curve for each of the ℓ_1-norm restrictions τ_k, its success hinges on the ability to quickly solve problems of the form

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tilde{\tau}. \tag{9}$$

This is typically called a Lasso problem, which differentiates itself from BPDN by an interchanged role between the objective and the regularizing metric, resulting in an objective function is actually differentiable (Tibshirani, 1996; Osborne et al., 1999). Because of this differentiability, Lasso problem can be solved much faster than BPDN using projected gradient methods (van den Berg and Friedlander, 2008; Schmidt et al., 2007; Daubechies et al., 2008). Furthermore, the algorithm used to solve each Lasso problem can be initialized with the solution of the previous Lasso problem, which distributes the burden of descending the misfit and ensures that each Lasso problem does not require increasingly larger number of iterations.

The root-finding approach can directly be used to solve the BPDN problem for $\tilde{\mathbf{g}}$. We set $\mathbf{y} = \mathbf{p}$ and $\mathbf{A} = \mathbf{M}_{\tilde{q}}$ to arrive at Lasso problems of the form

$$\underset{\mathbf{g}}{\text{minimize}}\, \|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{g}\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_1 \leq \tilde{\tau} \tag{10}$$

and use the Newton iterations to find the minimum $\tilde{\tau}$ needed for a given target residual. This becomes lines 7-8 in Algorithm 3. In the typical case where noise and preprocessing artifacts will cause the field dataset $\mathbf{p}$ to not be in the range of $\mathbf{M}_{\tilde{q}}$ even for a relatively accurate $\tilde{\mathbf{q}}$, the Pareto curve will not reach $\sigma = 0$, but will instead have an asymptote at the least-squares misfit $\sigma^* = \min \|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{g}\|_2$ (van den Berg and Friedlander, 2011). We therefore need to make sure that our target residual $\tilde{\sigma}$ satisfies $\tilde{\sigma} > \sigma^*$. A pragmatic approach to this condition is to monitor λ^{-1} through the Newton iterations and stop the inversion when its magnitude becomes unreasonably large, which would indicate the asymptote at σ^*.

In practice, we find that the number of projected gradient updates for each Lasso problem is on the order of 10 for tolerances acceptable to geophysical problems. The overall Newton root-finding method for the correct $\tilde{\tau}$ that solves BPDN, when initialized at $\tau_0 = 0$ and a zero solution vector, also tends to converge quickly, typically within 10 Newton steps. See Figure 6.1(c) in van den Berg and Friedlander (2008) for a typical solution path.

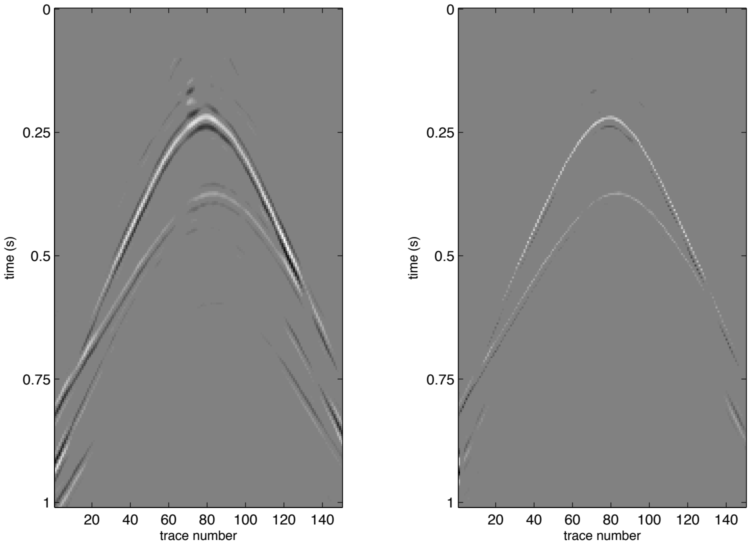
Because the Lasso subproblems are cheaper to solve than the overall BPDN problem, their solutions serve well as the sparse approximations to the full BPDN solution. As mentioned in the last section, this motivated the entire discussion of the Pareto curve. In terms of the overall cyclic block-coordinate descent scheme, solving these Lasso problems play an analogous role to forming the single gradient updates for $\tilde{\mathbf{g}}$ in the original EPSI algorithm – they both yield estimates of $\mathbf{g}$ that are suitable for improving our estimate of the source signature $\mathbf{q}$.

Matching for the source wavelet

As each of the Lasso problems described in equation 10 are solved, we obtain a series of approximations to the primary impulse response $\mathbf{g}$. Since they are on the Pareto curve, any $\tilde{\mathbf{g}}_{k+1}$ also satisfies up to tolerance

$$\underset{\mathbf{g}}{\text{minimize}}\, \|\mathbf{g}\|_1 \quad \text{subject to} \quad \|\mathbf{p} - \mathbf{M}_{\tilde{q}_k}\mathbf{g}\| \leq \sigma(\tau_k), \tag{11}$$

for the corresponding τ_k from the Newton root-finding procedure, and the mapping $\sigma(\tau_k)$ to the corresponding misfit on the Pareto curve. Because this is a BPDN problem, we can justifiably believe that each $\tilde{\mathbf{g}}_k$ are also sparse. Figure 3 illustrates that the Lasso problems lead to sparse estimates of the primary Green's function. The sparse property motivates improving our estimate of the source wavelet around these approximations of $\mathbf{g}$ obtained



(a) (b)

Figure 3: A typical estimate for the Green's function $\tilde{\mathbf{g}}$ at **(a)** the first iteration of a Lasso problem (equation 10), and **(b)** at the final iteration. This Lasso problem was solved via a spectral-projected gradient method (van den Berg and Friedlander, 2008) using 8 gradient updates. This demonstrates that Lasso problems tend to result in sparse solutions.

through the Lasso problems. This is done by solving the second of the set of subproblems stated in equation 6b using the linearized least-squares objectives $f_{\tilde{\mathbf{g}}_k}$ taken at $\tilde{\mathbf{g}}_k$.

The updated wavelet estimate is subsequently used to form a new linearization $\mathbf{M}_{\tilde{q}_k}$ for the next Lasso problem. In theory, this will perturb the underlying Pareto curve, and we can no longer safely expect the original solution $\tilde{\mathbf{g}}_k$ to be approximately optimal after the re-linearization. However, the first-order effect of this perturbation largely contributes to inaccuracies in the computed τ_{k+1}, which mostly governs how quickly the root-finding procedure locate the minimum ℓ_1-norm τ^* for the target $\tilde{\sigma}$. The effects of errors in approximating the Pareto curve solution, and its subsequent effect on the convergence rate of the root-finding procedure, is studied in van den Berg and Friedlander (2008), where the sensitivity of the convergence rate of τ to errors in sampling the Pareto curve is shown to be directly related to the conditioning of each $\mathbf{M}_{\tilde{q}_k}$. Empirically we found that τ_{k+1} as determined before and after the re-linearization rarely differed by more than 2 to 3 percent, especially towards the end of the Newton iterations, and therefore we regard the perturbation of the Pareto curve to be negligible on the quality of the final solutions.

An important caveat needs to be addressed here: Lasso problems place an ℓ_1-norm constraint on the Green's functions, which indirectly limits the amplitude of the primary reflection events in $\tilde{\mathbf{g}}_k$. This is an necessary trade-off unique to the BPDN approach that depends on using the ℓ_1-norm as a heuristic for sparsity. Consequentially, we can expect $\tilde{\mathbf{g}}_k$ to underestimate the amplitudes of the reflection events. To ensure that our estimate of $\mathbf{q}$ have the correct amplitude, we need to appropriately rescale the events in $\tilde{\mathbf{g}}_k$. Under the assumptions of a global source signature for every event in the data, it should suffice to scale the entire $\tilde{\mathbf{g}}_k$ with a single factor s_k that minimizes $\|\mathbf{p} - \mathbf{M}_{\tilde{q}_{k-1}}(s_k\tilde{\mathbf{g}}_k)\|_2$. This is essentially an exact line-search scaling, which is given by

$$s_k = \frac{\tilde{\mathbf{g}}_k^H \mathbf{M}_{\tilde{q}_{k-1}}^H \mathbf{p}}{\|\mathbf{M}_{\tilde{q}_{k-1}} \tilde{\mathbf{g}}_k\|_2^2}. \tag{12}$$

This scaling needs to be restored before the next Lasso problem for $\tilde{\mathbf{g}}_{k+1}$ is formulated, to make sure we return to the neighborhood of the Pareto curve. The wavelet matching and the preceding scaling operation discussed in this section is reflected in line 9-11 of Algorithm 3. Note that this scaling does not perform the "debiasing" that is common in the solution of BPDN problems (for example, Figueiredo et al., 2007). It only attempts to mitigate the issue that Lasso solutions early on in the Newton iterations are obtained with a severely underestimated ℓ_1-norm.

Initializing the algorithm

The matching procedure for the source signature involves solving a vastly overdetermined least-squares problem 6b, as it attempts to match $\mathbf{q}$ to the total data. If a sophisticated algorithm such as Least-Squares QR (Paige and Saunders, 1982) is used, $\tilde{\mathbf{q}}$ will converge extremely quickly, especially compared to the first-order projected gradient methods required for the Lasso problems (cf. equation 10). This leads to a real danger of overfitting $\tilde{\mathbf{q}}$ at the beginning of the algorithm, when $\tilde{\mathbf{g}}$ is expected to be relatively more erroneous, and often tends to result in an erroneous local minima solution. Note that is also a valid concern for the original EPSI algorithm; even though both variables are updated with gradient methods, convergence rate can still differ greatly. In fact, this is a common issue for all blind deconvolution problems.

An effective approach when dealing with local minima is using as much a-priori information as possible to narrow down the appropriate region of the solution space at the outset of the inversion. For example, in the application of blind photographic deblurring, one of the most common strategy is to first identify edges in the image, which allows the algorithm to quickly establish the correct shape of the unknown point spread function by maximizing contrast. We find that an analogous technique is similarly effective for our case. It essentially involves careful identification of the largest impulsive event per trace in $\tilde{\mathbf{g}}$, which is then matched with the data to rapidly obtain a reasonable candidate for $\tilde{\mathbf{q}}_0$.

This initialization starts with a multidimensional autocorrelation of $\mathbf{p}$ (equivalent to a gradient update for $\mathbf{g}$ when $\tilde{\mathbf{q}} = 0$). From each trace of this autocorrelation, we pick out the single strongest event that coincides with a primary reflection in $\mathbf{p}$, which will typically be the ocean-bottom reflection. This can be achieved in a variety of ways, such as a combination of muting and automatic first break picking, and can be manually verified if needed. We then identify its peak location as the arrival time of an impulsive reflection, and place a single spike at that time in our otherwise empty initial model of the surface-free Green's function. After this is done for each trace, we ensure that the amplitude of these spikes in is in the correct neighborhood, scaling with the exact line-search formula in equation 12, setting $\mathbf{M}_{\tilde{q}_{k-1}}\mathbf{g} = \mathbf{M}_0\mathbf{g} := M(\mathbf{g}, 0)$. A full back-projection can also be done by extending s to a diagonal weighting, allowing each spike to be weighted individually. Physically, this is interpreted as adjusting the amplitudes of every event in the estimate of the surface-free Green's function so that it models the multiple events as much as possible through the operator $M(\tilde{\mathbf{g}}, 0)$.

Finally, setting this methodically calibrated single-event approximation of $\mathbf{g}$ as $\tilde{\mathbf{g}}_0$, we can solve equation 6b using the fastest least-squares solver possible to quickly obtain a good initial source signature estimate $\tilde{\mathbf{q}}_0$ with minimal fear of overfitting. This accomplishes our goal of negating the risk of finding a $\tilde{\mathbf{q}}_0$ so wildly incorrect that it precludes reasonable solutions of $\tilde{\mathbf{g}}$ (even when it satisfies the constraint set Λ that we predetermined for $\mathbf{q}$). Once $\tilde{\mathbf{q}}_0$ is obtained, we can optionally choose to minimize the imprint of the initialization by discarding $\tilde{\mathbf{g}}_0$ and reset it to a zero vector $\tilde{\mathbf{g}}_0 = \mathbf{0}$ before starting the cyclic block-coordinate descent iterations. The steps discussed in this section correspond to lines 2-4 in Algorithm 3.

Summary of the Robust EPSI algorithm

We call our approach the *Robust* Estimation of Primaries by Sparse Inversion. The prefix *robust* is chosen to reflect both a lack of dependence on empirical inversion parameters and the formulation's explicit relation to the expected noise level in the data through σ. Algorithm 3 provides an overview of this method. Compared to the original EPSI algorithm described in Algorithm 2, Robust EPSI retains the general structure of the cyclic block-coordinate descent method to refine both $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{q}}$ in an alternating fashion. In fact, if a gradient-based method is chosen to solve the Lasso problems for $\tilde{\mathbf{g}}$, then Robust EPSI mainly differs algorithmically from the original EPSI in the way that the gradient updates of $\mathbf{g}$ are treated. Yet, the underlying optimization problems being solved in each case are completely different. The original EPSI method seeks *some* sparse model of $\mathbf{g}$ that gives a small prediction error when compared to the measured data, with the sparsity level weakly influenced by empirical parameters. In contrast, Robust EPSI simply demands the *sparsest* model for any given error level.

Algorithm 3 The proposed Robust EPSI algorithm

```
1:  Input: target residual energy $\tilde{\sigma}$, recorded surface seismic data $\mathbf{p}$
2:  $\tilde{\mathbf{g}}_0 \leftarrow$ initialize via single event picking as described in text
3:  $\tilde{\mathbf{q}}_0 \leftarrow \text{argmin}_{\mathbf{q}} \ \|\mathbf{p} - \mathbf{M}_{\tilde{g}_0}\mathbf{q}\|_2^2$ s.t. $\mathbf{q} \in \Lambda$
4:  $\tilde{\mathbf{g_0}} \leftarrow$ reset to zero vector
5:  $k \leftarrow 0, \ \tau_0 \leftarrow 0$
6:  repeat
7:      $\tau_{k+1} \leftarrow$ determine from $\tilde{\sigma}$ and $\tau_k$ using Newton's method on the Pareto curve
8:      $\tilde{\mathbf{g}}_{k+1} \leftarrow \text{argmin}_{\mathbf{g}} \|\mathbf{p} - \mathbf{M}_{\tilde{q}_k}\mathbf{g}\|_2$ s.t. $\|\mathbf{g}\|_1 \leq \tau_{k+1}$
9:      scale $\tilde{\mathbf{g}}_{k+1}$ by $s_{k+1}$ according to (12)
10:     $\tilde{\mathbf{q}}_{k+1} \leftarrow \text{argmin}_{\mathbf{q}} \ \|\mathbf{p} - \mathbf{M}_{\tilde{g}_{k+1}}\mathbf{q}\|_2$ s.t. $\mathbf{q} \in \Lambda$
11:     undo scaling of $\tilde{\mathbf{g}}_{k+1}$ by $s_{k+1}$
12:     $k \leftarrow k + 1$
13: until $\|\mathbf{p} - M(\tilde{\mathbf{g}}_k, \tilde{\mathbf{q}}_k)\|_2 \leq \tilde{\sigma}$, or a predetermined iteration limit is reached
14: Output: estimated primary Green's function $\tilde{\mathbf{g}}_k$, estimated source signature wavelet $\tilde{\mathbf{q}}_k$
```

EXAMPLES ON SYNTHETIC DATA

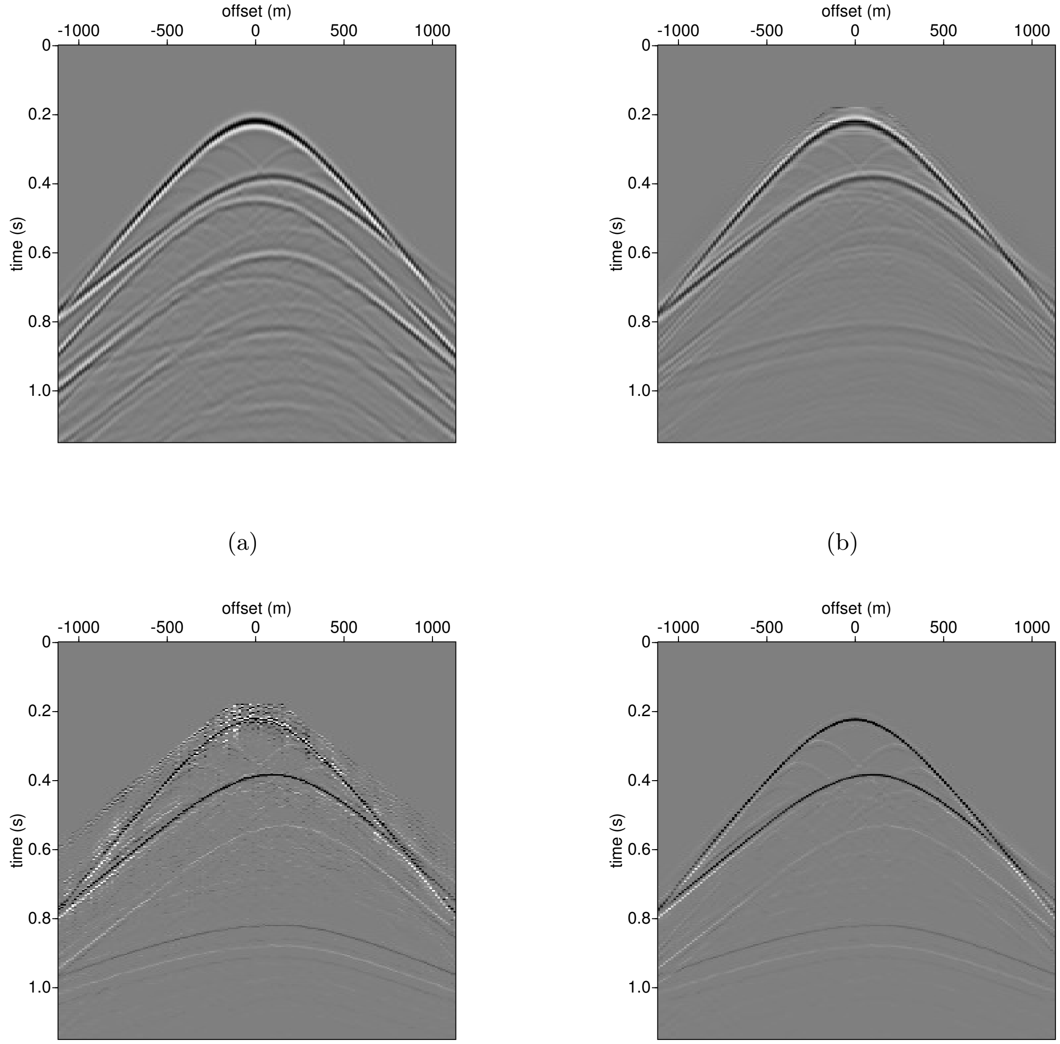
We demonstrate the effectiveness of Robust EPSI by applying it to the same synthetic seismic line as the one used throughout van Groenestijn and Verschuur (2009a). A shot gather from the center of the line is shown in Figure 4(a). The dataset is modeled from a 2D synthetic model that was depicted in Figure 1 of the aforementioned paper. It has a water layer approximately 200 m deep and a laterally-varying salt dome layer that is situated roughly between the depths of 400 m and 800 m. The split-spread pressure data (up to zero-offset) is modeled by a second-order in time, fourth-order in space finite-difference acoustic program using a zero-phase Ricker wavelet centered at 30 Hz as source signature, and is deghosted at the receiver-side. We have a well-sampled source line that exactly coincides with the receiver line, so the multidimensional convolution for multiple prediction is free of aliasing issues. A factor of $\sqrt{\omega}$ is included in $\mathbf{M}_{\tilde{q}}$ to account for the non-impulsive nature of Green's functions in 2D.

From this data we obtained two estimates of the surface-free Green's function according to the original EPSI algorithm as described in Algorithm 2 after 80 iterations. Figure 4(b) depicts the result obtained without the sparsifying operator $\mathcal{S}$ (i.e., kept each gradient update for $\tilde{\mathbf{g}}$ untouched), and serves to illustrate the importance of sparsity in EPSI. Not only was the primary multiple removal relatively ineffectual (e.g., the water-bottom multiple at 0.45 s), but also the result clearly does not resemble reflection events due to an impulsive source. Figure 4(c) depicts the result obtained by setting $\mathcal{S}$ to keep the 4 largest amplitude events per trace for every update on $\tilde{\mathbf{g}}$. This greatly improved the multiple rejection and delivered the impulsive appearance we expect from the surface-free Green's function.

Both of these results required a triangle muting mask on $\tilde{\mathbf{g}}$ throughout the inversion procedure (as part of the global time-window implicitly included in $\mathcal{S}$) to prevent placing events before the first water-bottom reflection event around 0.25 s. However, there is still visible artifacts at the boundaries of the triangle mute. Since these anti-causal events are not in the original data, they indicate update error introduced when $\tilde{\mathbf{q}}$ is severely underestimated in amplitude, most likely at one of the earlier iterations. Because the original EPSI inversion does not attempt to seek the sparsest solution, we find that these errors decay very slowly with the iteration count. Note that we slightly exaggerated the clipping percentile of the colormap, which was kept consistent for all of Figure 4 except for 4(a), so these artifacts appear more visible in the plots.

Meanwhile, Figure 4(d) shows the result produced by Robust EPSI. We set $\tilde{\sigma}$ to 2% of total data energy in anticipation of edge effects, but the program is terminated early at the equivalent of 82 gradient updates of the Green's function. Because we solve the convex Lasso problems to obtain $\tilde{\mathbf{g}}$, we do not need to use a triangle muting mask in this case. Note the absence of the aforementioned anti-causal artifacts, which indicates that we are close to good physical approximations with our $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$. The fidelity of the Green's function obtained from the Robust EPSI procedure is also qualitatively better, with the weaker reflections around 0.8 s and the diffracted events directly below the water-bottom reflection more cleanly defined compared to the original EPSI results. In both programs the wavelet was allowed to live within a ±100 ms time window, and can be seen in Figure 5(a).

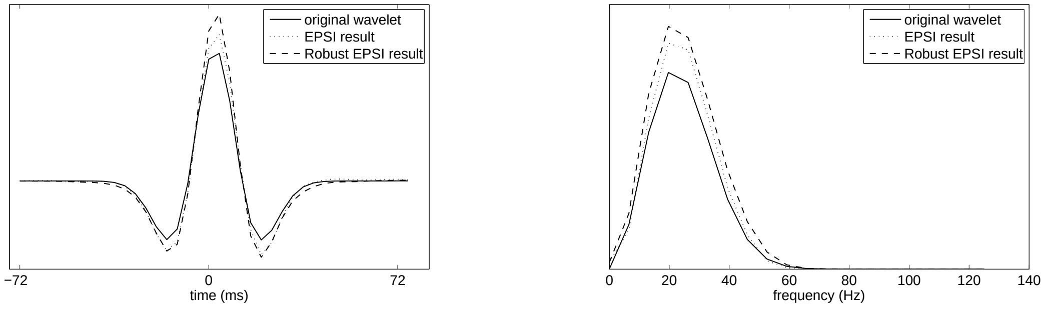
Figure 6 depicts the objective function $\|\mathbf{p} - M(\tilde{\mathbf{g}}, \tilde{\mathbf{q}})\|_2$ as a function of gradient updates on $\tilde{\mathbf{g}}$ (which is the main burden of the computation). Although the data-space convergence rate and final residual are similar in both cases, the actual solutions (cf. Figures 4(c) and 4(d)) greatly differ. The projected gradient methods used to solve the set of Lasso problems in Robust EPSI converged very quickly, typically only requiring 5 to 20 projected gradient updates for each Lasso problem. The root-finding procedure updated τ a total of 6 times throughout this procedure.



(a) (b)

(c) (d)

Figure 4: Comparison of the surface-free Green's function obtained via different formulations of EPSI. **(a)** The input data simulated from a 2D marine model. **(b)** shows the result produced by the original EPSI outlined in Algorithm 2 without using the sparsifying operator $\mathcal{S}$ on updates to $\tilde{\mathbf{g}}$. When $\mathcal{S}$ is employed on all updates to $\tilde{\mathbf{g}}$, keeping the 4 largest events per trace per update, the result obtained is shown in **(c)**. The result produced by Robust EPSI outlined in Algorithm 3 is shown in **(d)**, using a roughly equivalent number of gradient updates compared to the two other solutions. A comparison between **(b)** and **(c)** shows that sparsity regularization is important in producing correct estimations of the surface-free Green's function. The solution produced by Robust EPSI is cleaner and more free of artifacts compared to that of the original EPSI with use of the sparsifying operator.



(a) (b)

Figure 5: A comparison between the source-signature estimates produced by the original EPSI and the Robust EPSI approaches as **(a)** time signals, and **(b)** amplitude spectra. The original wavelet used, plotted in solid line, is a Ricker wavelet centered at 30 Hz. Note that both methods produced slight overestimates of the wavelet amplitude. However, when we forced both algorithms to directly use the input wavelet, the resulting Green's function contained various surface multiple events. This may be due to errors introduced in the data-simulation process, or the limitations of imperfect deghosting in approximating upgoing wavefields. Nevertheless, this serves to demonstrate that the correct source signature for the purpose of rejecting multiple events is not necessarily the true physical source signature.

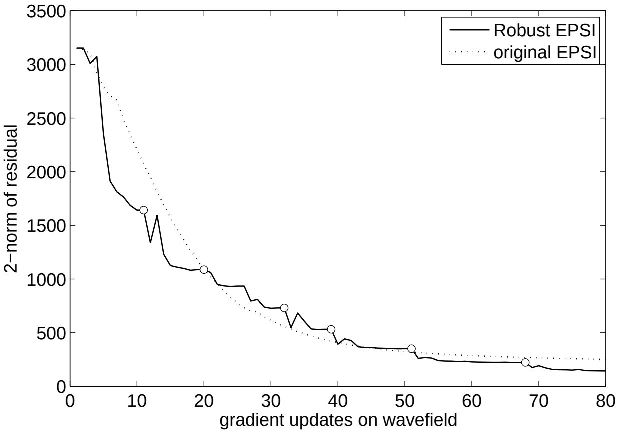
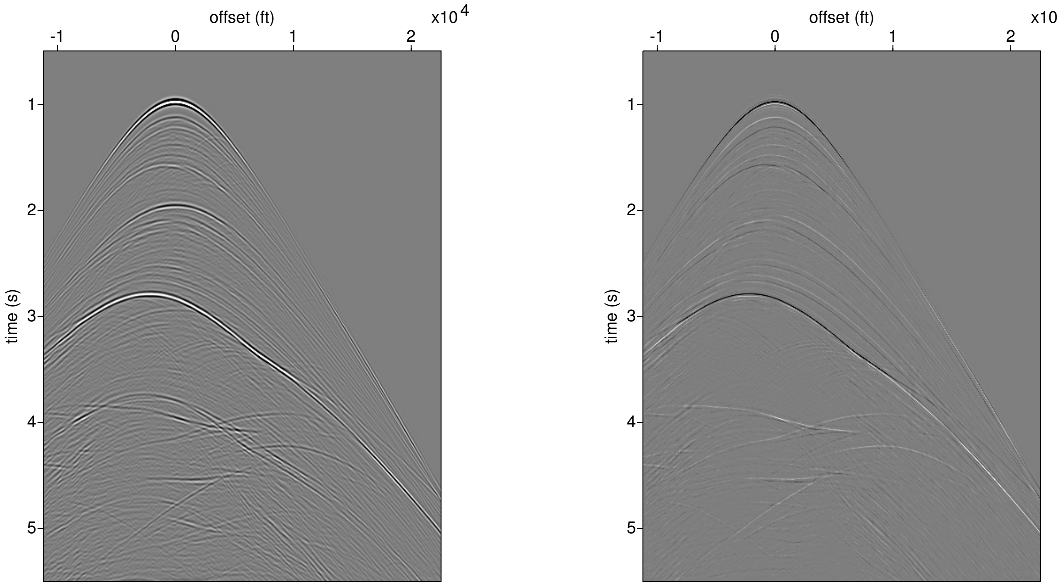


Figure 6: Comparison of the convergence rates between the original EPSI algorithm and Robust EPSI. Circle dots for the Robust EPSI convergence curve indicate where the solution have converged to the Pareto curve and a new Lasso problem was formed. Target residual for Robust EPSI was $\sigma = 64$. Due to properties of the spectral projected-gradient solver we used for the Lasso problems, the objective for Robust EPSI is not monotonically decreasing.

To further demonstrate the resiliency of the Robust EPSI algorithm, we also used it to process the Pluto 1.5 dataset publicly released by the SMAART JV consortium, which was produced using fourth-order finite difference P-SV modeling code with a Ricker wavelet source centered at 15 Hz. Not only did we keep the exact same parameters used in the previous example, we also directly used the pressure recording at the surface, foregoing any preprocessing such as deghosting or up-down wavefield decomposition. Nevertheless, the inversion process remained stable and produced a Green's function with clean impulsive appearance while successfully rejecting multiples. See Figure 7 for a snapshot of the results. Without any change in the parameters used compared to the previous example, and despite the large difference between the two datasets, Robust EPSI was able to produce a clean estimate of the surface-free Green's function while successfully rejecting the main surface multiples at 2 s, 3.8 s, and 4.8 s.



(a) (b)

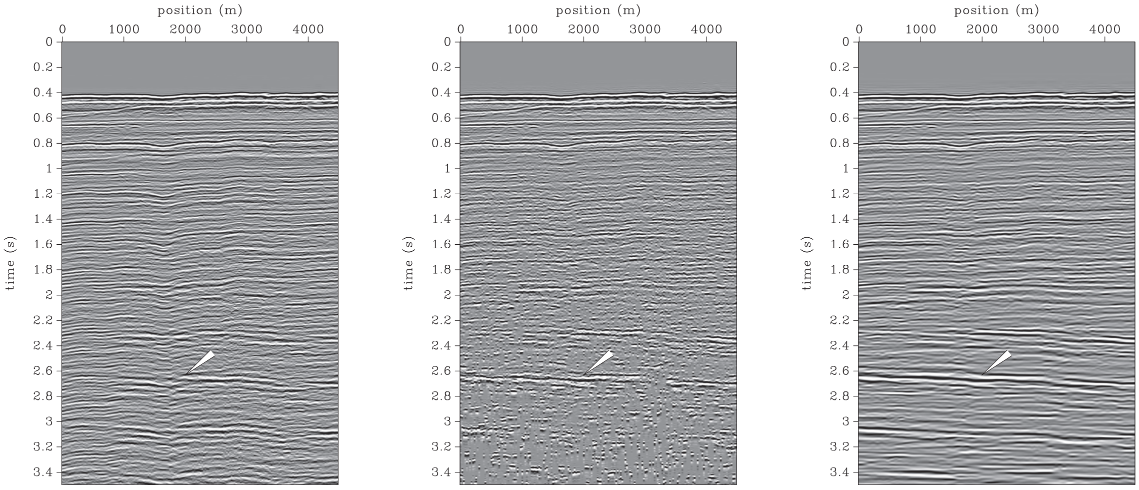
Figure 7: Shot gather of Pluto 1.5 data is shown in **(a)**, while the surface-free Green's function produced by Robust EPSI after 80 gradient updates on the Green's function is shown in **(b)**. Parameters for the Robust EPSI algorithm remain unchanged from the ones used to produce Figure 4(d) yet managed to produce a clean solution for the surface-free Green's function without the surface multiples at 2 s, 3.8 s, and 4.8 s. This demonstrates the robustness of Algorithm 3 in adapting to different input datasets without empirical parameter selection.

INCORPORATING SPARSIFYING TRANSFORMS

One of the advantages of the Robust EPSI algorithm is the ease of seeking the primary Green's function $\tilde{\mathbf{g}}$ in terms of coefficients of transform domains. This is because the sparsity of our solution comes from solving Lasso problems, and the algorithms we utilize are agnostic to the linear forward operator. If the original EPSI approach is used, $\mathcal{S}$ might need to be redesigned with the specific transform in mind. For Robust EPSI, we are free to compound $\mathbf{M}_{\tilde{g}}$ with any linear synthesis transform without changing the underlying algorithm.

Depending on the specific transform chosen, this can be leveraged for several benefits. As discussed in Herrmann (2010), a representation for $\tilde{\mathbf{g}}$ that permits fast decay of the non-linear approximation error (i.e., a sparsifying representation) will enhance the effectiveness of ℓ_1-norm minimization, which in turn strengthens our BPDN approach. Also, a representation that is composed of atoms that decay smoothly in physical space can be exploited to mitigate spatial undersampling issues in $\mathbf{p}$. For example, $\tilde{\mathbf{g}}$ can be sought as coefficients in the hyperbolic Radon domain, analogous to the approach used in van Dedem and Verschuur (2005). Furthermore, we will show that these two properties can be leveraged to improve the recovery of weak late arrivals that can prove challenging even to the Robust EPSI formulation.

Recall Figure 1 which presents an example of the lack of wavefront continuity in $\tilde{\mathbf{g}}$ using a field dataset collected from the North Sea (shown in Figure 8(a)). This dataset is known to be problematic for 2D SRME due to strong out-of-plane scattering of the reflected events and imperfect interpolation of the near-offset traces, and should be similarly challenging for EPSI since they operate under the same physical assumptions. Figure 1 reflected the original EPSI result that is terminated after 85 gradient updates on $\tilde{\mathbf{g}}$. Note that while the strong early events are reconstructed smoothly, the much weaker late events (after 2 s) exhibit severe wavefront discontinuities.



(a) (b) (c)

Figure 8: Common-offset plot (200 m) of **(a)** the North Sea data used to obtain the REPSI result shown in Figure 1, **(b)** the primary wavefield obtained by Robust EPSI in the physical domain, and **(c)** the Robust EPSI result obtained under a Curvelet-Wavelet representation. Compared to the result in Figure 1, the Robust EPSI algorithm was able to recover the late primary event under the arrow in a more continuous fashion, but is still not satisfactory without stacking. The result obtained under the Curvelet-Wavelet representation shows a significantly improved recovery for this event. Automatic gain control was used for the plots.

Figure 8(b) shows the Robust EPSI result obtain after an equivalent of 85 gradient updates on $\tilde{\mathbf{g}}$, with $\tilde{\sigma}$ set to 10% of total data energy. Note that although late events, such as the one under the arrow, is slight better recovered under Robust EPSI, the results are still unsuitable for many seismic applications that rely on prestack data. This due to these events falling below the presumed noise-floor in the data, as reflected by our setting of $\tilde{\sigma}$. Figure 8(c) shows the results of the same Robust EPSI algorithm used for Figure 8(b) but seeking the sparse solution in terms of Curvelet coefficients (Candès et al., 2006a) in the source-receiver plane of $\tilde{\mathbf{g}}$. In the time domain, we used a wavelet representation built from a third-order Battle-Lemarié wavelet (Battle, 1987). We chose to use the Curvelet representation for several reasons. It is composed of basis functions of finite energy and approximate local support in the physical domain, so the notion of a sparsest solution still correlates with a minimum number of events in the physical space. Additionally, because individual Curvelets live in thin wedges in the phase space, it is very compactly supported in the wavenumber domain in certain directions, which will align to give significant physical support to wavefront-like structures across a large number of traces. Notice that the late events, especially the ones below 3 s, are much better represented in this result.

To modify the Lasso problems for this extension, the synthesis operator $\mathbf{S}^H$ of the transform domain ($\mathbf{S}^{-1}$ for unitary transforms), which maps the domain coefficients to the physical signal, simply needs to be compounded with $\mathbf{M}_{\tilde{q}}$ in (10) to form a new Lasso problem

$$\underset{\mathbf{x}}{\text{minimize}} \, \|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{S}^H\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau. \tag{13}$$

which is solved using exactly the same algorithm as the one used for (10). This essentially re-parameterizes the unknown Green's function as coefficients in the transform domain. Once the solution coefficients $\tilde{\mathbf{x}}_k$ are obtained, the physics surface-free Green's function can be recovered once again by the synthesis operation $\tilde{\mathbf{g}}_k = \mathbf{S}^H\tilde{\mathbf{x}}_k$. This should be done before subsequently refining $\tilde{\mathbf{q}}$. Both the Lasso algorithm and the Newton iterations do not need to be modified, as long as we redefine $\mathbf{A} = \mathbf{M}_{\tilde{q}}\mathbf{S}^H$ instead of $\mathbf{A} = \mathbf{M}_{\tilde{q}}$ for equations 8 and 9. The scalings s_k also need to be calculated with this overloaded operator. We did not modify the initialization steps in this approach.

In summary, seeking our Green's function in a transform domain requires nothing more than the ability implement the synthesis transform and its adjoint. We note that (13) is a well-known application of the basis pursuit method (Chen et al., 2001), and in most cases to tends to behave well and produce ideally parsimonious coefficient representations in the transform domain.

Example on field data

As a final demonstration of Robust EPSI's capabilities we a look at a seismic line from marine streamers in the Gulf of Suez region. The pre-processing included a 2x receiver trace interpolation, Radon domain near-offset interpolation, reconfiguration to a split-spread geometry via reciprocity. Neither deghosting nor up-down decomposition was performed on the data except for an attempted removal of direct waves. The region of interest is composed of 361 shot records and 361 hydrophone traces at co-located positions. Robust EPSI was performed on this data using both the original variant (Algorithm 3) and the variant that seeks the solution in a transform domain. Following the previous section, we choose the transform domain of a combination of 2D Curvelet frame with 5 scales and 16 angles along the source-receiver (constant time) plane, and the basis associated with the DWT using a third-order Battle-Lemarié wavelet (Battle, 1987) representing time-traces. Notationally we have $\mathbf{S}^H := \mathbf{C}^H \otimes \mathbf{W}^{-1}$, where $\mathbf{C}^H$ is the synthesis operator of the 2D Curvelet frame and $\mathbf{W}^{-1}$ is the inverse DWT.

We show a 250 m offset panel of the data and the Robust EPSI results in Figure 9. With the same number of gradient updates we have a very evident improvement in reconstructing the later events past 1 s by using the transform domain. The direct Robust EPSI result only reproduces the strongest sections of the late events, while the transform domain solution is largely successful in reconstructing them in their entirety. Our transform domain solution also produced a more smoothed appearance, which was also the case for Figure 8. Figure 10(a) shows the resulting estimated global source wavelet in Figure 9(c), where again no constraint is put on the shape besides a time window that coincides with the boundaries of the plot.

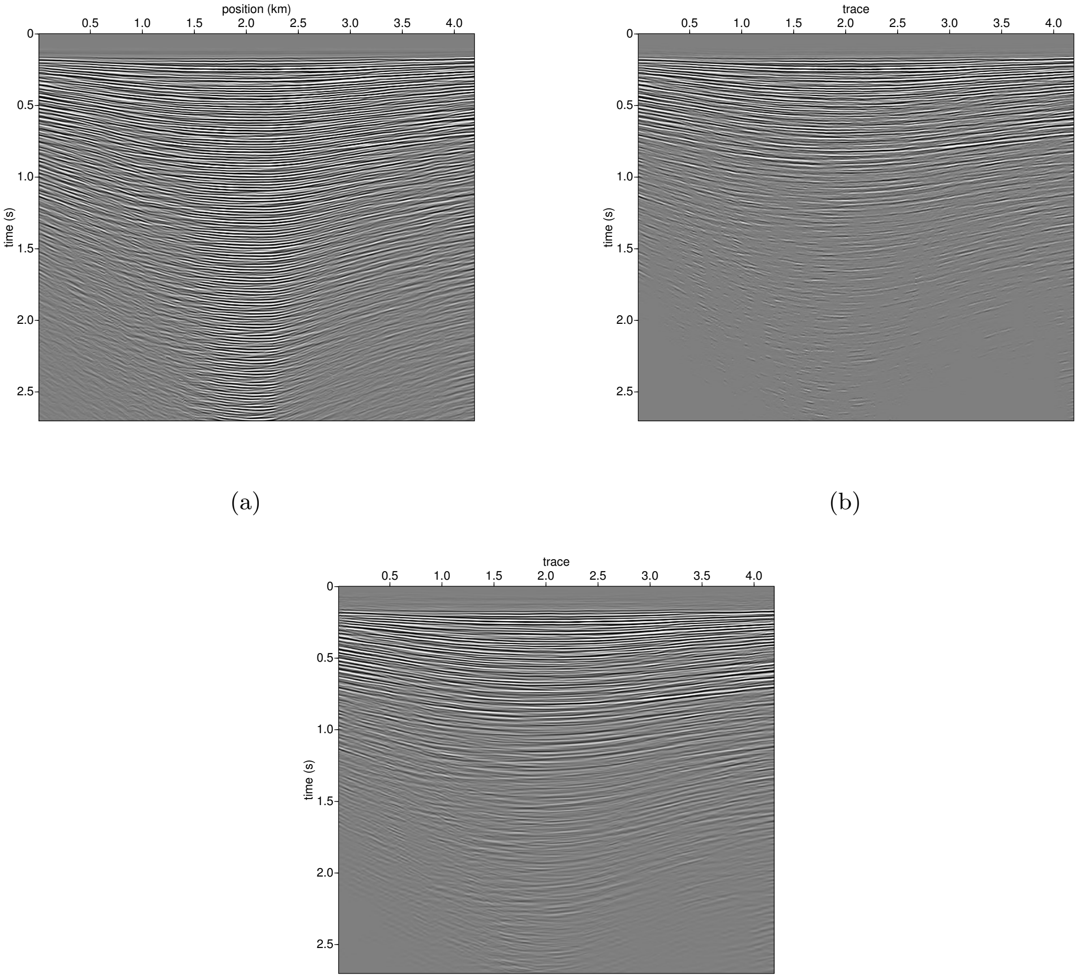
DISCUSSION

Computational considerations

The majority of the cost involved in Robust EPSI is the same as that of the original EPSI program; both consists of gradient updates on both the Green's function and on the source signature. Computing the gradient on the Green's function requires a multidimensional cross-correlation of seismic wavefields, while evaluation of the forward modeling operator requires the multidimensional convolution (identical to the SRME prediction step). These are still considered costly operations, even though for most implementations they are carried out in parallel over frequency.

Nonetheless, the engineering task of making these operations computationally economical directly benefit from the widespread efforts in efficient implementations of SRME. Furthermore, due to the nature of EPSI as a wavefield inversion, it is compelling to explore whether the algorithm is tolerant to using low-rank approximations of $\mathbf{p}$ both as part of the modeling operator M and also as the observable. The data-structure of low-rank wavefield approximations can directly be exploited to reduce the cost of the wavefield convolutions and correlations that underlie the evaluation and gradient steps of EPSI. Provided that the low-rank approximation may be obtained at a low overhead cost compared to the overall cost of EPSI, this will greatly increase the computational tractability of EPSI, especially on large 3D datasets where actual information of the reflection events are overwhelmed by the size of the ambient wavefield dimensions. Exploratory work on this topic had been done in Jumah and Herrmann (2011) with the low-rank approximations built up efficiently via random vector probes into $\mathbf{p}$.

Another avenue to further reduce the computation cost of wavefield operations come from exploiting simultaneous acquisition. It was shown in van Groenestijn and Verschuur (2011) that the physics of EPSI can work directly on simultaneously acquired data, and that information from the multiple events can even aid in the separation of shot records. Because Robust EPSI uses the same ℓ_1-norm minimization that had been previously proposed to directly separate simultaneous seismic data (Herrmann et al., 2009; Wason et al., 2011), there is reason to believe that Robust EPSI can recover its solution under a much denser

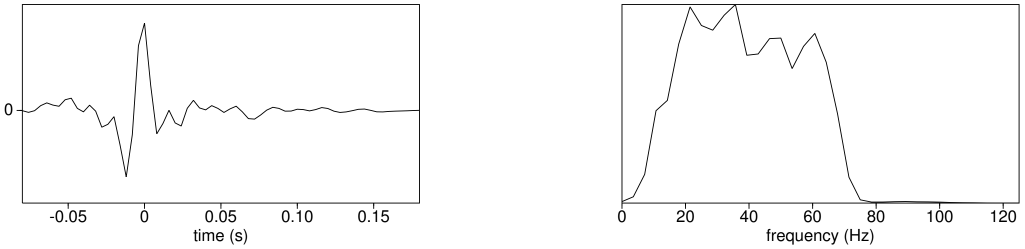


(a)

(b)

(c)

Figure 9: Common-offset plot (250 m) of **(a)** a field dataset obtained in the Gulf of Suez used as an input to produce **(b)** a direct Robust EPSI result, and **(c)** a Robust EPSI result obtained under a Curvelet-Wavelet representation (see text). Although the direct Robust EPSI result demonstrated a strong ability to reject surface multiple events in shallow water-bottom data, reflection events past 1 s are poorly recovered. Using a Curvelet-Wavelet representation, these events are much more clearly recovered.



(a) (b)

Figure 10: Estimate of the source signature $\tilde{\mathbf{q}}$ that was produced in the process of obtain Figure 9(c) as part of Robust EPSI, plotted as **(a)** time signal and **(b)** amplitude spectrum. Note that, as mentioned previously, this wavelet does not necessarily reflect the true physical source signature, and furthermore will capture all preprocessing on the data that be modeled by a global short-time filter.

simultaneous acquisition system compared to the method shown in van Groenestijn and Verschuur (2011), using the combined framework proposed in Lin and Herrmann (2009). This opens the possibility of directly using post-acquisition summing of seismic shot records (sometimes referred to as "shot-encoding") to reduce the effective size of $\mathbf{p}$ for the goal of further speeding up the wavefield convolutions and correlations.

The spectral projected gradient method we used (van den Berg and Friedlander, 2008, section 4.1) to solve the Lasso problems for the Green's function incurs negligible computation overhead compared to the large costs of computing the gradients. It mainly introduces a further fixed memory overhead (two to three times) due to having to store one previous solution of the Green's function and its updates, but is responsible for a major acceleration in convergence rate. Overall, we expect the slight increase in computation costs using Robust EPSI to be more than compensated by the increased convergence rate and the quality of the resulting Green's function, as well as the reduced dependence on parameter search compared to the original EPSI algorithm.

Applications to source signature deconvolution and other future extensions

As Figure 4 showed, the Robust EPSI approach results in a very clean estimate of the Green's function. This opens up the possibility of using Robust EPSI as a way to deconvolve the source signature. This approach will differ from the traditional "spiking" deconvolution in that it additionally exploits information from the multiples to construct the impulse response. Some preliminary work on this topic in terms of spectrum recovery of seismic data can be seen in Lin and Herrmann (2011). When considered together with the wavefield-decomposition application of EPSI and the possible extension of interpolating near-offset direct-waves (van Groenestijn and Verschuur, 2009c), it becomes clear that (Robust) EPSI can play a very important role in preparing seismic data for velocity analysis (and perhaps even full-waveform inversion from reflected events). It is possible that the EPSI modeling operator itself can even act as a preconditioner-like operator during any iterative procedure to obtain seismic velocities, much like the framework introduced in the context of sparsity-promoting least-squares migration in Lin et al. (2010).

Other important existing extensions of EPSI include the ability to simultaneously estimate missing near-offset reflection data and the ability to account for source-arrays. With the Pareto root-finding approach, we believe there are also straightforward adaptations of these extensions. The near-offset traces could, for example, be updated along with the source signature upon reaching the Pareto curve. The complete decoupling of the source signature estimation from the Green's function in Robust EPSI means we can trivially change our assumptions on the source without modification to the algorithm. We look forward to reporting on implementing these extensions as a part of Robust EPSI in the future. Due to the optimization framework that it is derived from and the generic solvers it employ, we believe Robust EPSI presents itself as an even more flexible framework compared to the original EPSI formulation for future extensions of the method.

SUMMARY

Similar to many applications in signal processing and engineering design, the method of Estimation of Primaries by Sparse Inversion (EPSI) uses prior assumption on the structure of the solution to regularize an ill-posed inverse problem, which in this case attempts recover the surface-free Green's function of the earth and the source signature from seismic field acquisitions. Because the Green's function is expected to be impulsive in nature, EPSI uses the sparsity of the solution wavefield as the regularization parameter. It can be shown that for correct recovery of primary events while rejecting surface multiple events, it is desirable to obtain the sparsest possible solution for a given level of expected noise in the data. However, sparsity-minimizing problems in general are difficult and computationally prohibitive for large signal sizes, so the original EPSI approach (cf. Algorithm 2) only attempted to seek a sparse solution through gradients of fixed sparsity, which is an approach that can only guarantee a upper limit to the sparsity of the solution.

By reformulating EPSI under an optimization-based framework, we were able to propose the new Robust EPSI algorithm (cf. Algorithm 3) that retains the notion of solving a sparsity-minimizing problem, but uses a modified BPDN approach to keep the algorithm computationally tractable. To efficiently solve for both the unknown source signature and the surface-free Green's function in conjunction, we exploited the Pareto curve of the basis pursuit denoising problem to find well-defined switching points in the algorithm where improvement in either variable is alternatively sought. To mitigate the underestimation of amplitude for intermediate sparse approximations of the surface-free Green's function when using the Pareto root-finding approach, we proposed a scaling factor that ensures the subsequent source signature approximation will not have an overestimated amplitude as a result. We also proposed a careful initialization step that attempts to further dissuade the solution from entering an undesired local minima. Our goal was to arrive at an algorithm that works well for a large diversity of datasets without excessive tweaking of inversion parameters that is inherent in the original approach.

Experimentally, we showed results of the Robust EPSI algorithm as applied to various synthetic and field datasets. Compared to the original EPSI approach, Robust EPSI was shown to produce a more artifact-free estimate of the Green's function using synthetic data. Robust EPSI was also shown to work on vastly different datasets using near identical parameter settings, demonstrating the resiliency of the basis pursuit denoising approach. Further improvements in the quality of the solution was shown to be possible by seeking the Green's function in a hybrid transform domain composed of a Curvelet transform in spatial directions and a wavelet transform in the time direction, with no appreciable change to the underlying solvers used. This was the first of many extension that were discussed which takes advantage of the very generic optimization framework that Robust EPSI was developed under, and we hope to report on more in the near future.

ACKNOWLEDGEMENTS

We would like to thank Eric Verschuur and G.J.A. van Groenestijn for the genesis of our involvement in EPSI as well as valuable discussions, as well as Aleksandr Aravkin and Tristan van Leeuwen for helpful suggestions in the editing of this paper. Our Lasso solver is derived from SPGℓ_1, which is the work of Michael Friedlander and Ewout van den Berg, who were also responsible for the Sparse Linear Operator Toolbox upon which our implementation of Robust EPSI was based. The Pluto 1.5 dataset was provided by SMAART JV. The salt dome model and the Gulf of Suez dataset was generously provided by Eric Verschuur and G.J.A. van Groenestijn. This work was in part financially supported by the NSERC Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, BGP and WesternGeco.

REFERENCES

Amundsen, L., 2001, Elimination of free-surface related multiples without need of the source wavelet: Geophysics, **66**, 327–341.

Baardman, R. H., D. J. Verschuur, R. G. van Borselen, M. O. Frijlink, and R. F. Hegge, 2010, Estimation of primaries by sparse inversion using dual-sensor data: SEG Technical Program Expanded Abstracts, 3468–3472.

Battle, G., 1987, A block spin construction of ondelettes. Part I: Lemarié functions: Communications in Mathematical Physics, **110**, 601–615.

Berkhout, A. J., and Y. H. Pao, 1982, Seismic Migration - Imaging of Acoustic Energy by Wave Field Extrapolation: Journal of Applied Mechanics, **49**, 682.

Berkhout, A. J., and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, Part I: Theoretical considerations: Geophysics, **62**, 1586–1595.

Bezdek, J. C., R. J. Hathaway, R. E. Howard, C. A. Wilson, and M. P. Windham, 1987, Local convergence analysis of a grouped variable version of coordinate descent: Journal of Optimization Theory and Applications, **54**, 471–477.

Biggs, D. S. C., 1998, Accelerated iterative blind deconvolution: PhD thesis, University of Auckland, Auckland.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, 861–899.

Candès, E. J., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, 1207.

Chen, S. S., and D. L. Donoho, 1994, Basis pursuit: Proceedings of 28th Asilomar Conference on Signals, Systems and Computers, IEEE Comput. Soc. Press, 41–44.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129.

Daubechies, I., M. Fornasier, and I. Loris, 2008, Accelerated projected gradient method for linear inverse problems with sparsity constraints: Journal of Fourier Analysis and Applications, **14**, 764–792.

Donoho, D. L., 2006, For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution: Communications on Pure and Applied Mathematics, **59**, 797–829.

Figueiredo, M. A. T., R. D. Nowak, and S. J. Wright, 2007, Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems: IEEE Journal of Selected Topics in Signal Processing, **1**, 586–597.

Fokkema, J. T., and P. M. van den Berg, 1993, Seismic applications of acoustic reciprocity: Elsevier Science.

Frijlink, M., R. van Borselen, and W. Söllner, 2011, The free surface assumption for marine data-driven demultiple methods: Geophysical Prospecting, **59**, 269–278.

Fuchs, J., 2005, Recovery of Exact Sparse Representations in the Presence of Bounded Noise: IEEE Transactions on Information Theory, **51**, 3601–3608.

Guitton, A., and D. J. Verschuur, 2004, Adaptive subtraction of multiples using the L1-norm: Geophysical Prospecting, **52**, 27–38.

Hennenfent, G., E. van den Berg, M. P. Friedlander, and F. J. Herrmann, 2008, New insights into one-norm solvers from the Pareto curve: Geophysics, **73**, A23.

Herrmann, F. J., 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, WB173–WB187.

Herrmann, F. J., Y. A. Erlangga, and T. T. Y. Lin, 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35.

Jumah, B., and F. J. Herrmann, 2011, Dimensionality-reduced estimation of primaries by sparse inversion: SEG Technical Program Expanded Abstracts, 3520–3525.

Lin, T. T., and F. J. Herrmann, 2009, Unified compressive sensing framework for simultaneous acquisition with primary estimation: SEG Technical Program, SEG, 3113–3117.

Lin, T. T., N. Tu, and F. J. Herrmann, 2010, Sparsity-promoting migration from surface-related multiples: SEG Technical Program Expanded Abstracts, 3333–3337.

Lin, T. T. Y., and F. J. Herrmann, 2011, Robust source signature deconvolution and the estimation of primaries by sparse inversion: SEG Technical Program, 4354–4359.

Majdanski, M., C. Kostov, E. Kragh, I. Moore, M. Thompson, and J. Mispel, 2011, Attenuation of free-surface multiples by up/down deconvolution for marine towed-streamer data: Geophysics, **76**, V129.

Osborne, M. R., B. Presnell, and B. A. Turlach, 1999, On the LASSO and Its dual.

Paige, C. C., and M. A. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **8**, 43–71.

Savels, T., K. de Vos, and J. W. de Maag, 2010, Surface Multiple Attenuation Through Sparse Inversion Attenuation Results for Complex Synthetics and Real Data: Presented at the 72nd EAGE Conference & Exhibition.

Schmidt, M., G. Fung, and R. Rosales, 2007, Fast optimization methods for L1 regularization: A comparative study and two new approaches: ECML '07 Proceedings of the 18th European conference on Machine Learning, Springer Berlin Heidelberg, 286–297.

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: Journal of the Royal Statistical Society Series B Methodological, **58**, 267–288.

Toxopeus, G., R. van Borselen, R. H. Baardman, L. Auer, and E. Odegaard, 2011, Advanced geohazards assessment in shallow water through the estimation of primaries by sparse inversion: SEG Technical Program Expanded Abstracts, 3526–3530.

Tropp, J., 2006, Just relax: convex programming methods for identifying sparse signals in noise: IEEE Transactions on Information Theory, **52**, 1030–1051.

van Borselen, R., G. Fookes, and J. Brittan, 2003, Target-oriented adaptive subtraction in data-driven multiple removal: The Leading Edge, **22**, 340.

van Borselen, R. G., J. T. Fokkema, and P. M. van den Berg, 1996, Removal of surface-related wave phenomena—The marine case: Geophysics, **61**, 202–210.

van Dedem, E. J., and D. J. Verschuur, 2005, 3D surface-related multiple prediction: A sparse inversion approach: Geophysics, **70**, V31.

van den Berg, E., and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.

———, 2011, Sparse optimization with least-squares constraints: SIAM Journal on Optimization, **21**, 1201.

van Groenestijn, G. J. A., and W. Ross, 2011, Primary estimation on OBC data by sparse inversion: SEG Technical Program Expanded Abstracts, 3531–3535.

van Groenestijn, G. J. A., and D. J. Verschuur, 2008, Towards a new approach for primary estimation: SEG Technical Program Expanded Abstracts, **27**, 2487.

———, 2009a, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, A23–A28.

———, 2009b, Estimation of primaries and near-offset reconstruction by sparse inversion: Marine data applications: Geophysics, **74**, R119–R128.

———, 2009c, Estimation of primaries by sparse inversion applied to up/down wavefields: SEG Technical Program Expanded Abstracts, SEG, 3143–3147.

———, 2009d, Estimation of primaries by sparse inversion from blended data: Presented at the 71st EAGE Conference & Exhibition.

———, 2010a, Estimation of primaries by sparse inversion from passive seismic data: Geophysics, **75**, SA61.

———, 2010b, Incorporating the source array into primary estimation: Presented at the 72nd EAGE Conference & Exhibition.

———, 2011, Using surface multiples to estimate primaries by sparse inversion from blended data: Geophysical Prospecting, **59**, 10–23.

Verschuur, D. J., 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166.

Verschuur, D. J., and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, Part II: Practical aspects and examples: Geophysics, **62**, 1596–1611.

Wason, H., F. J. Herrmann, and T. T. Y. Lin, 2011, Sparsity-promoting recovery from simultaneous data: A compressive sensing approach: SEG Technical Program, 6–10.

You, Y. L., and M. Kaveh, 1996, A regularization approach to joint blur identification and image restoration.: IEEE Transactions on Image Processing, **5**, 416–428.

Ziolkowski, A., D. B. Taylor, and R. G. K. Johnston, 1999, Marine seismic wavefield measurement to remove sea-surface multiples: Geophysical Prospecting, **47**, 841–870.