Compressive sensing, sparse transforms and deblending

Daniel Trad

ABSTRACT

Data acquisition is by far the most expensive and problematic part of seismic methods. In particular, 3-dimensional data surveys always have deficient sampling in at least 2 of the 4 spatial dimensions. As a consequence, geophysicists make extensive efforts in the mitigation of sampling problems. These efforts usually involve two directions: data interpolation and simultaneous acquisition. Interpolation is intended to create new seismic traces from the acquired samples by using sparse transformations. Simultaneous acquisition, also known as blending, attempts to mitigate the sampling problem by acquiring more data without increasing the acquisition cost. Simultaneous acquisition is a very cost-effective approach that reduces the cost of seismic information in both marine and land settings. Its main difficulty is the processing of the resulting seismic data, which requires shot separation or deblending, very early in the signal processing chain.

In the last few years, the two approaches have been merged in geophysics with the name of Compressive Sensing (CS). CS refers to an approach developed in the field of mathematics, which permits to obtain information with less sampling by relying on the combination of irregular sampling and sparseness to extract information from sparsely sampled data. This name is a bit unfortunate, because in reality CS, as used in seismic, is a combination of sparse transforms plus deblending, and both technologies have been performed in geophysics much before Compressive Sensing existed. Nonetheless, the name has now stuck in geophysics, and involves acquiring data in a random fashion, using simultaneous sources, and performing deblending and denoising right at the beginning of processing by using sparse transforms.

In this paper, I will discuss the relationships between CS and sparse transforms, showing that both are just the same approach with a different name. Then, I will discuss one particular approach for deblending based on migration/demigration as the transform operator. Finally, I will consider the merge of 5D interpolation with LSMIG as a single approach for deblending.

INTRODUCTION

Compressive sensing Candès et al. (2006) is a relatively new idea in signal processing that exploits three main concepts, sparse representation, irregular sampling and sparse inversion (typically ℓ_1). These concepts were quite revolutionary for the signal processing community where data are usually acquired in regular grids. Signal processing uses the fundamental principle that if sampling honours the Nyquist condition, it is guaranteed from Shannon's theorem to recover the exact values of any sample in between. This principle ruled signal processing for many years, and still it does, because given any electronic recording equipment it is relatively easy to sample with Nyquist distance. However, the data explosion we see in modern society has brought the need to reduce the size of images, and therefore many compression techniques have been developed to solve that issue. What the CS revolution brought was an alternative where sampling can be done much more sparsely to reduce the amount of data required to guarantee a proper reconstruction.

This new idea opened the door to new possibilities for geophysics, in particular for surface seismic. Acquisition designers now have a new target: to create economical sparse acquisitions with irregular sampling that provides the same or more information than more expensive traditional techniques. Although this new idea brings a new useful framework, one has to question how much of the components are truly new to geophysicists. Does not this concept sound familiar? It certainly should. Geophysicists have been applying similar ideas about signal reconstruction for four decades (sparse transforms) and two decades (deblending). The most obvious resemblance between CS, as is used in mathematics, and sparse transforms, as used in geophysics, is multidimensional interpolation.

Multidimensional interpolation (Sacchi et al., 1998; Duijndam et al., 1999), in particular in the form of Five-Dimensional interpolation (Liu and Sacchi, 2004; Trad, 2009), is a popular method to pre-condition seismic data for migration. In the form of sparse transforms, it is rooted in a vast literature in geophysics (Thorson and Claerbout, 1985; Sacchi and Ulrych, 1995, 1996)). Although many different varieties have been developed with impressive results, Fourier interpolation is still the most popular, followed quite closely by low-rank techniques. This is due partly to its computational efficiency, which permits one to work with large multidimensional windows, and partly because of the similarity between Fourier impulses and seismic plane waves. This similarity allows us to use sparse inversion to filter out sampling artifacts. A key component of this process is the subdivision of seismic data into groups or windows where the plane wave approximation works well. Although the algorithm has no constraints on what type of space coordinates should it use, the most common approach is to use midpoints, offsets, and azimuths (Trad, 2009). This choice is related to the way multidimensional sampling works. Although two adjacent samples along any particular dimension can be far apart, the multidimensional distance between samples is usually small. In particular, the sampling interval is smaller along these processing coordinates than along acquisition coordinates. This permits sampling high wavenumbers which would not be possible in one dimension by using the Nyquist theorem. Another key aspect of these chosen dimensions is that they are always irregular for 3D data. So we can see that one apparent difference between CS and multidimensional interpolation by sparse transforms is that geophysicists considered irregular sampling as an opportunity that was already present, instead of a new way to sample data. Another difference is that geophysicists did not try to mathematically guarantee that a particular acquisition design could lead to a perfect reconstruction.

Around the same time when multidimensional interpolation became very popular to mitigate the sampling problem, a different approach was born to improve sampling. This approach was oriented towards getting more data at a lower cost. The most effective way to do this was by simultaneous shooting (Beasley et al., 1998; Mahdad et al., 2011; Vermeer, 2009; Abma and Yan, 2009; Abma et al., 2015), which also relies on sparse transforms and multidimensional sampling as a way to separate the simultaneous shots. Perhaps, because of similarity on principles between interpolation and deblending, it did not take long before both approaches where combined. In principle, if interpolation can be used to generate new data from existing shots and receivers, it should be possible also to separate simultaneous shots by creating new data from the existing blended acquisition. The main challenge, however, is that the process of generating new data moves towards the beginning of the processing sequence. Interpolation, therefore, becomes a double process of denoising+regularization. Although is usually not effective in standard sparse acquisitions, CS and deblending proponents argue that it works well when more (blended) data are available. This is the big challenge of this new philosophy, a challenge that requires much testing and effort, since involves not one operation but several changes in standard dataflows.

The goal of this report is to take a close view to compressive sensing and sparse inversion as done in geophysics and analyze the different components of both approaches to understand what is new and what isn't. Also, I discuss one particular approach for deblending based on the idea of physical transformations, that is decomposition of the data in basis functions that have physical meaning as opposed to pure mathematical transforms. Considering migration as the adjoint of the more fundamental process of modeling, migration/demigration is the most obvious choice of this kind of transforms.

COMPRESSIVE SENSING AND SPARSE INVERSION

Following Candès and Wakin (2008), CS relies on two principles very common in signal processing and in particular in geophysics:

- Sparsity- "information rate of a continuous time signal may be much smaller than suggested by its bandwidth";
- Incoherence- "objects having a sparse representation in one domain may be spread out in another domain".

The key ingredients in CS according to Herrmann and Hennenfent (2008) are the following:

- A sparse representation of the target image or wavefield in some transform domain (i.e. Fourier, Radon, wavelet, or curvelet transforms);
- Advantageous sampling in the measurement domain that adequately characterizes the sparse wavefield components in the transform domain;
- Convex optimization algorithms for recovering a sparse representation of the seismic wave-field from the observed data.

The exact meaning of each of these conditions requires a lengthy explanation and I refer the interested reader to the sources, for example Candes et al. (2006). Instead, in this report, I offer a simpler explanation using the language of operators and inversion. Given a mapping from a regular data set $\mathbf{x}$ to a sampled data set $\mathbf{y}$ through an extraction operator $\mathbf{\Phi}$:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

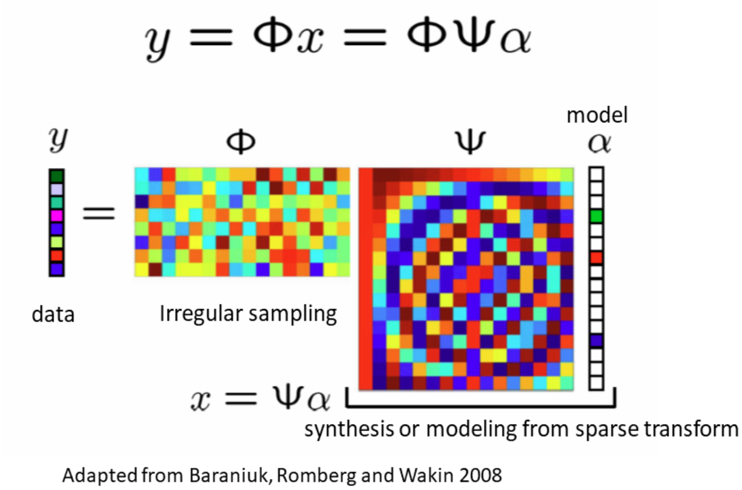

FIG. 1. Compressive sensing modeling

and a synthesis operator $\mathbf{\Psi}$, with adjoint operator $\mathbf{\Psi^H}$, such that:

$$\mathbf{x} = \mathbf{\Psi}\alpha, \tag{2}$$

compressing sampling can recover the full data set $\mathbf{y}$ with maximum error σ by the following optimization (Candès and Wakin, 2008):

$$\text{minimize } \|\mathbf{\Psi^H}_m\mathbf{x}\|_1 \text{ subject to } \|\mathbf{\Phi x} - \mathbf{y}\|_2 \leq \sigma \tag{3}$$

Using signal processing language, $\mathbf{\Phi}$ represents a random sampling operator, $\mathbf{\Psi^H}$ is sparsifying transform with $\mathbf{\Psi}$ being its inverse transform, and x are the data sampled in a uniform way which is not measured but calculated by sparse inversion (ℓ_1 inversion) for the model that matches the data in the least-squares sense with error σ. These equations define an operator to map regular data to sampled data and defines the solution to be one that has a minimum size in an ℓ_1 sense (sum of absolute values) and matches the acquired data when the data exist. A visualization of the main concepts behind CS is shown in Figure 1. Here α represents a sparse model that can predict the data y accurately if a set of basis functions $\mathbf{\Psi}$ that resembles the data is used. The sampling matrix $\mathbf{\Phi}$ is required to destroy organized noise (aliasing). Compressive sensing uses mathematical conditions on these matrices to provide convergence theorems for idealized scenarios.

Sparse least squares inversion

The general philosophy for sparse least squares inversion is explained in many places in the geophysical literature, (Claerbout, 1992; Sacchi and Ulrych, 1996; Trad et al., 2003). An early paper that encapsulates the essence of the sparse inversion with the generality of CS is Harlan et al. (1984). The two key concepts to describe sparse transforms are, just like in CS:

- A transformation that maps a spread out signal in one domain to a dense signal in another domain.

- A sparse inversion (ℓ_1) or algorithm that, given a data set and a sparse operator, finds the model that explains the data using the minimum possible number of elements.

Just as before I used symbols common in CS, here I will use symbols common in sparse transforms. Given a transformation of some model $\mathbf{m}$ to some data $\mathbf{d}$ through some mathematical operator $\mathbf{L}$,

$$\mathbf{d} = \mathbf{L}\mathbf{m}, \tag{4}$$

in sparse inversion we find the model by inverting the operator in a least squares sense with a constraint that the number of coefficients in the model should be the minimum possible. This is the key difference with standard inversion where the minimum energy constraint (ℓ_2) enforces the coefficients to change smoothly. In ℓ_1 we penalize less the individual coefficient sizes than their existence. In a mathematical language we say

$$\text{minimize } \|\mathbf{m}\|_1 \text{ subject to } \|\mathbf{L}\mathbf{m} - \mathbf{d}\|_2 \leq \sigma$$

This is the most common description for sparse inversion. It is quite similar to the CS formulation in Eq. (3), but we need a few more steps to make clear that it is in fact the same. With that purpose, it is useful to look at the sparse transform formulation most commonly used in 5D interpolation (Liu and Sacchi, 2004). First of all, we do not use the ℓ_1 because is not as efficient to implement as the ℓ_2. Instead, we will use iterative re-weighted least squares with weight functions such that:

$$\|\mathbf{W}_\mathbf{m}\mathbf{m}\|_2 = \|\mathbf{m}\|_1 \tag{5}$$

which amounts just to a change of variables (Trad et al., 2003), or using Bayes's theorem to a particular choice of prior information for the model (Sacchi and Ulrych, 1996). Second we use a multidimensional Fourier transformation $\mathbf{L}$ to map seismic plane waves, which are non-sparse in the *time-space* ($t - x$) domain, into Fourier coefficients in the *frequency-wavenumber* ($\omega - k_x$) domain. This transform is a choice not a definition of the method. Because we assume that the coefficients are sparse but the data are not, we need to minimize the norm of the coefficients, not the data. Given an unknown regular data set $\mathbf{x}$, we will use the adjoint operator $\mathbf{L}^H$ to map these regular data to the Fourier coefficients

$$\mathbf{m} = \mathbf{L}^\mathbf{H}\mathbf{x} \tag{6}$$

Notice that, with this new definition we need to update equation 4 to

$$\mathbf{x} = \mathbf{L}\mathbf{m}, \tag{7}$$

where we have made a distinction between sampled data $\mathbf{d}$ and modeled data $\mathbf{x}$. We will make explicit that $\mathbf{L}$ contains a sampling operator $\mathbf{T}$ if we want to produce data only at the locations where samples exist:

$$\mathbf{d} = \mathbf{T}\mathbf{L}\mathbf{m} \tag{8}$$

If $\mathbf{L}$ where implemented as a generic discrete Fourier transform, the sampling operator would be unnecessary. If we use a Fast Fourier transform (FFT), $\mathbf{T}$ is required because the FFT produces regular data. After these definitions, equation 5 becomes the Minimum Weighted Norm Interpolation method described in Liu and Sacchi (2004):

$$\text{minimize } \|\mathbf{W}_\mathbf{m}\mathbf{L}^\mathbf{H}\mathbf{x}\|_2 \text{ subject to } \|\mathbf{T}\mathbf{x} - \mathbf{d}\|_2 \leq \sigma \tag{9}$$

Comparing equation 9 and 3, we can see that $\mathbf{L} = \mathbf{\Psi}$, $\mathbf{T} = \mathbf{\Phi}$, $\mathbf{m} = \alpha$, $\mathbf{d} = \mathbf{y}$. There are two differences: First, $\mathbf{W_m}$ does not appear in the CS formulation because it is used here to implement the ℓ_1 norm as a weighted ℓ_2. Second, $\mathbf{\Psi}$ in CS always implies a random sampling (as shown in Figure 1). We will discuss that in the next section.

A more intuitive way to explain the concept of sparse inversion is represented in Figures 2, 3 and 4. The problem of finding a model that explains (predicts) the data is underconditioned when the data are incomplete. Only adding additional information we can distinguish which model is correct. In a Bayesian framework (Tarantola, 2005), the additional information can be represented in terms of the probability density function for the model. If we can use a transformation whose basis functions resemble the data, then we can expect that few coefficients can suffice to represent the model. The widely ℓ_2 norm used for inversion assumes that coefficient amplitudes in the transform domain change smoothly (Claerbout, 1992). This leads to a smeared model (low-resolution transform). The ℓ_1 norm, on the other hand, favors few coefficients in the transform. This leads to the sparse transforms concept that forms the basis of CS, as properly referenced by Candès and Wakin (2008).

Irregular sampling in CS vs sparse transforms

One of the arguments in favor of CS as a new paradigm is the advocacy of irregular sampling. That aliasing can be attenuated by irregular sampling has been known and used from a very long time. For example Trad and Ulrych (1999) describe, as an extension of the Fourier case, how aliasing can be attenuated in the Radon transform by irregular sampling. However, the role of irregular sampling seems key for CS, which encourage the use of irregular sampling for shots and receivers, but only secondary for sparse transforms. Was this idea missing from geophysics? It was not. Sparse transforms did not ask acquisition designers to create surveys with irregular sampling because for most seismic applications these coordinates are already irregular.

Probably the difference on the emphasis of irregular sampling comes from the fact the CS was created for two-dimensional regular data like photographs. Seismic data are not regular. Furthermore, the redundancy of five-dimensional sampling in seismic data provides advantages not present in image processing: we can choose to work in the dimensions where data are the most irregular. Acquisition coordinates are often set in a regular grid for practical reasons although usually, they have significant irregularities as well (Figures 5 a and b). Processing coordinates, midpoints, offsets, and azimuths, are not regular ((Figures 5 c and d). They come from transforming Cartesian grid to polar coordinates, so they can never be regular unless the acquisition coordinates are set in a very special pattern. If offset x/y coordinates are used instead, they are also irregular except if the line spacing is as fine as group spacing (in that case, we don't need interpolation but larger computer disks). In addition, while acquisition coordinates can have sampling interval of the line intervals (hundreds of meters), midpoint coordinates have dimensions of bin sizes (ten meters or less).

In which conditions then should we need irregular shot and receiver layouts? If interpolation has to be done in acquisition coordinates for example. This happens when midpoints

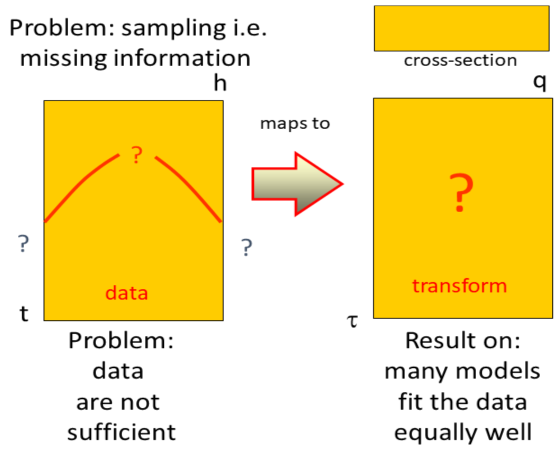

FIG. 2. Inversion uncertainty because of missing data

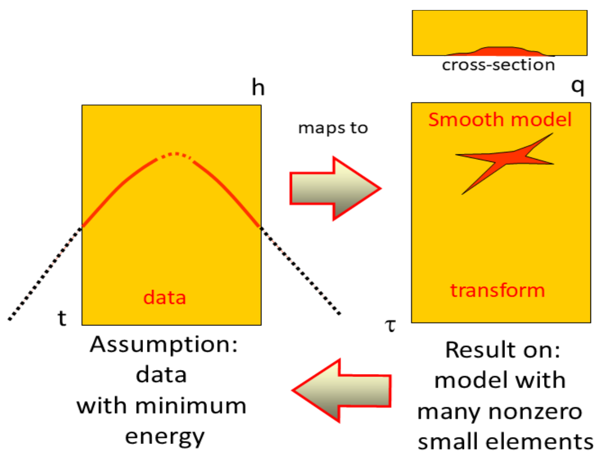

FIG. 3. Standard transforms principle: ℓ_2 norm inversion.

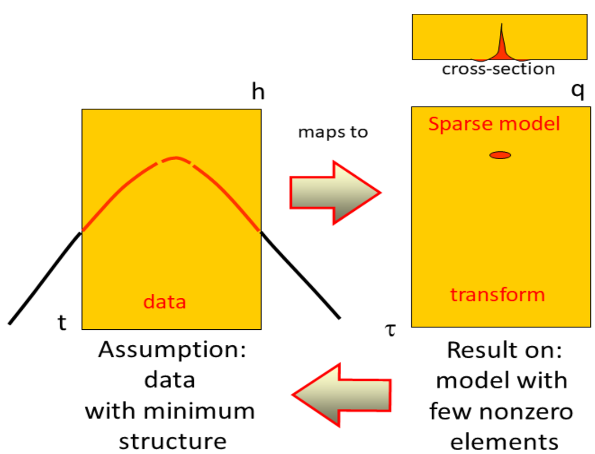

FIG. 4. Sparse transforms principle: ℓ_1 norm inversion.

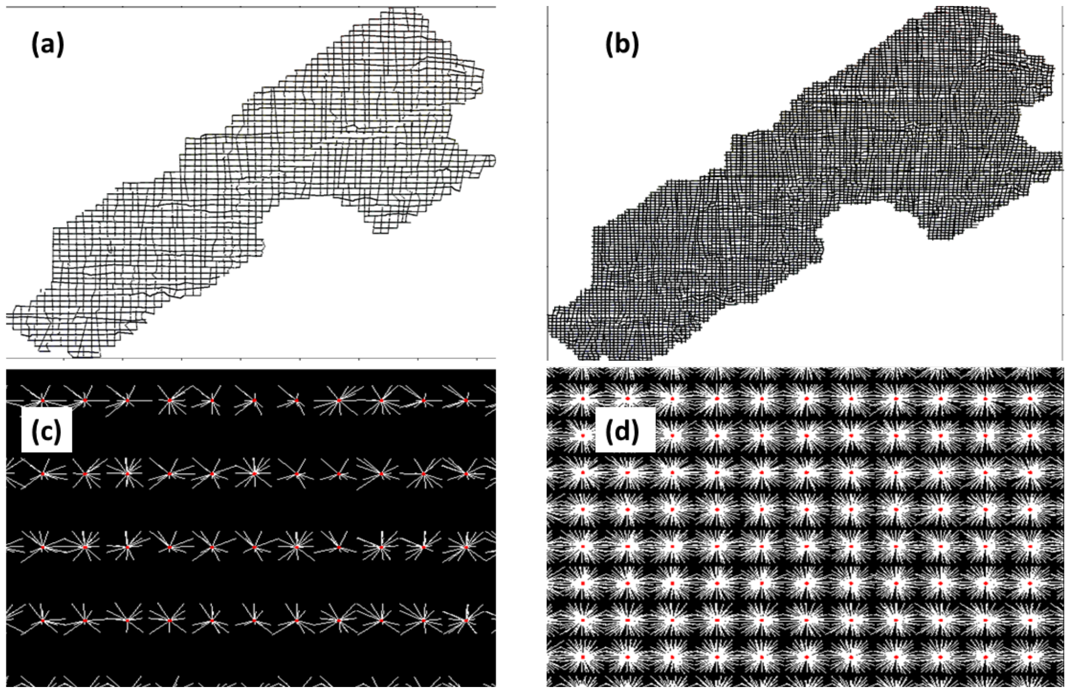

FIG. 5. A real 3D geometry plotted in acquisition coordinates a) before interpolation b) after interpolation. The same geometry plotted in interpolation coordinates c) before interpolation d) after interpolation

are time variant like it happens in converted wave data or irregular topography. However, it is not common to do so. The reason is that the large distance between shot and receiver lines will make interpolation very difficult. Another situation where that can be justified is for example for marine streamers, where data are not truly 5D but 4D.

However, the current trend proposed by CS includes a new element that makes irregular sampling essential: the need for deblending of simultaneous shots in raw data. Blending techniques are again not a product of CS but have been around from much before (see for example Beasley et al. (1998), but certainly have been a core of CS for seismic applications. Most deblending techniques work better in acquisition coordinates, and that is why CS requires irregular shots/receiver patterns that were not necessary for interpolation.

Is CS a new paradigm in Geophysics?

Although rarely mentioned by geophysicists doing compressive sensing, Candès and Wakin (2008) acknowledge that these concepts come from many decades ago and reference two papers from geophysics Claerbout and Muir (1973) and Santosa and Symes (1986). Considering that these two papers have been and continue to be followed by geophysicists over the years, it is unfortunate that CS advocates choose to ignore this.

However, CS brings to Geophysics a new source of inspiration and mathematical rigor. There are several components that may justify the claim of CS to be more general than sparse transforms, but these components are very often miss-understood, perhaps not by CS researchers but certainly by CS followers. For example, CS is often credited to be more generic than multidimensional interpolation because of using other transformations like curvelets, which have a more compact support than Fourier. It is true that the same researchers that developed CS brought to us many new sparse transforms like curvelets, beamleats, and other mathematical constructs. However, sparse transforms included many other ideas as generic as curvelets, for example, Generalized Radon transforms. In fact, sparse inversion was used with Radon transforms before it was used with Fourier (Thorson and Claerbout (1985); Sacchi and Ulrych (1995)). Furthermore, sparse transforms have often been used with physical basis functions, like interpolation by least squares migration that uses Green functions as its bases functions.

It is fair to say that CS brought a much richer set of theorems to study optimal sampling. Whether those theorems are used in seismic acquisition design is difficult to say, probably not. Much of the tests CS uses are not computationally feasible for standard acquisition design. Do they bring more understanding of the problem? Perhaps yes. Geophysicists have been always very creative but perhaps more interested in practical results than mathematical proofs of convergence. Mathematicians rediscovered with scientific rigor concepts that geophysicists used before. That is a valuable contribution. However, in my opinion, it is unfair to refer to CS is a ‘’new” paradigm. Just an old paradigm with a new name. What I think is fair to say is that CS describes a general framework that uses several common geophysical practices, and propose a new dataflow combining simultaneous shooting, deblending, denoising, and interpolation by using sparse transforms.

Have the goals of interpolation changed?

Original work on seismic data interpolation had different goals depending on the geological complexity where the data were acquired:

- Structured areas: the goal was to reduce sampling intervals to relax anti-alias conditions. Although high fidelity in amplitude was not critical in this case, complexity of seismic events made interpolation quite difficult.
- Stratigraphic plays: the goal was to create new shots and receivers to improve AVO and AVAz inversion results. The challenge in these cases was not the complexity of events but the preservation of the amplitude trends of events both in offset and azimuth directions.

Both of these cases involved interpolation just before migration, when the data had already been cleaned through signal processing. Interpolation was not asked to simultaneously apply denoising and create new samples because that made the problem very poorly conditioned. Simultaneous shooting, however, brings a different problem. Acquisitions are very dense because they have lower cost, but blended data has to be separated into individual shots for velocity analysis, statics, and other standard processing steps. For one side, blended acquisition does not require interpolation because it is already quite dense, but deblending does require regularization, which now involves denoising and interpolation simultaneously.

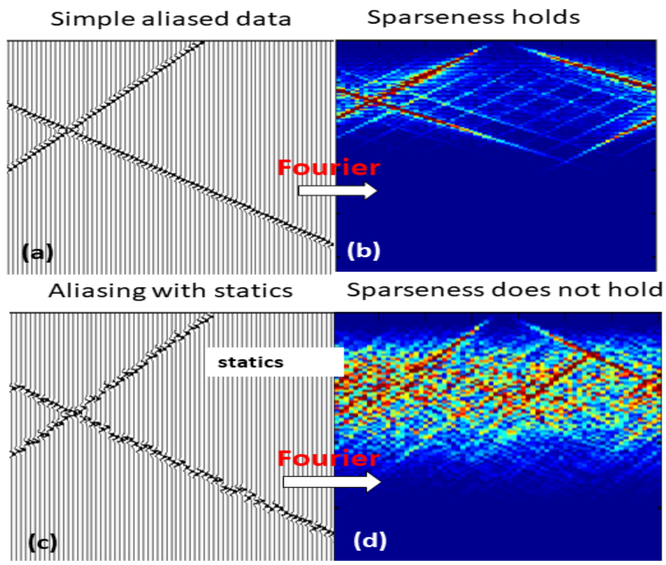

FIG. 6. a) Linear events map to well defined Fourier components in b), where standard techniques for sparse inversion would work very well, even if there is aliasing. c) Static shifts, as commonly presented in raw data, map to mixed Fourier coefficients in d) where sparse inversion would not work.

DEBLENDING

One way by which geophysicists can decrease the acquisition time and therefore greatly reduce the cost of information is by simultaneous shooting (Beasley et al., 1998). Acquiring data in this manner brings, however, significant difficulties because most processing algorithms use a unique source location, either by working on shot gathers or other groups derived from acquisition coordinates (for example offset and common midpoint gathers, inlines, crosslines). Even if migration and Full Waveform Inversion (FWI) can be done without deblending, as there are examples in the literature, we still need to obtain information from the data and apply corrections and denoising, before we can apply simultaneous shot migration and FWI. Therefore, to process blended data, it is necessary to separate at the front the information coming from different shots, which is a process known as deblending. Although there are many deblending tools available, often they use data properties that may not be well defined before processing. For example, methods based on coherence and sparseness may not work in raw data before static corrections are applied as Figure 6 shows.

Deblending tools have strong similarities with regularization and interpolation techniques, but they need to work well at the very beginning of the processing sequence, before any denoising or static corrections are applied.

There are several deblending approaches based on current processing techniques:

- Noise attenuation to eliminate interference: data are replicated as many times as sources are blended, and data sorted in such a way that interference appears incoherent or with different velocities or slopes. Many standard tools like FK and FX filters, Radon transforms, curvelet transforms, and others can be used to remove the interference. Basically, the information coming from the blended shots are treated as noise and removed as such. These techniques usually rely on random dithering or time delay introduced between simultaneous shots during acquisition, but not always.

- Multidimensional inversion: the blending process is taken as a mixing matrix or mixing operator and some sort of least squares inversion with regularization constraints is used to invert this operator to produce a deblended data set. This process can be combined with interpolation/regularization operators. Similarly to the previous technique, this method exploits some time or space perturbation on the shooting that makes the interference to appear incoherent along some spatial direction.
- Physical transformation: the blended data are transformed into a model with physical (rather than mathematical) meaning. Physical constraints are applied to this model, for example, reflector positions and velocities. The constraints serve as a deblending technique. Data are then predicted from this model into separated shots.

In this paper, I focus on the third approach by using migration-demigration to generate either predicted data directly or processing masks to be used with other denoising techniques, for example, double guided interpolation (Trad, 2014, 2015). However, as we will see later, the three types of techniques are very much connected, and all of them are instances of sparse transforms (Claerbout and Muir, 1973; Stanton and Sacchi, 2013), also known as CS (Candes et al., 2006).

A sparse-transform approach to deblending, denoising and interpolation for raw data

There are many different approaches for deblending of seismic data (see for example Abma et al. (2015)). In this report, I would like to describe an approach based on migration/demigration. Figure 7a shows a two-shot blended gather generated by finite differences using the Marmousi model. Figure 7b shows one of the shots after separation by a migration/demigration RTM. The high quality of this result would not be possible in real data with an inaccurate velocity model, but this example serves as an introduction to the type of methods discussed in this report. Obtaining an accurate image of the earth for blended data has different aspects. First, we need to think about denoising and statics (signal processing). Second, we need to think about velocity analysis. Third, we need to think about migration. In this report, I will discuss whether the physical transforms based on migration/demigration could help on the three tasks.

Usually, we think of migration as a post-processing approach because it requires an accurate velocity model since velocity errors translate into large imaging errors. Data produced by migration/demigration however, does not have the same sensitivity. Kinematic errors in reflectivity created during the migration are compensated by opposite errors in the modeling direction. Dynamic (amplitude) errors, on the other hand, are more persistent, unless we use a data fitting approach (least squares migration or LSMIG). The problem is that LSMIG is very expensive, but we will see later approaches to correct for amplitude errors without iterations. For the moment, let us compare the results of migration/demigration for the two end-members of the LSMIG family: Kirchhoff and RTM. Figures 8a shows a

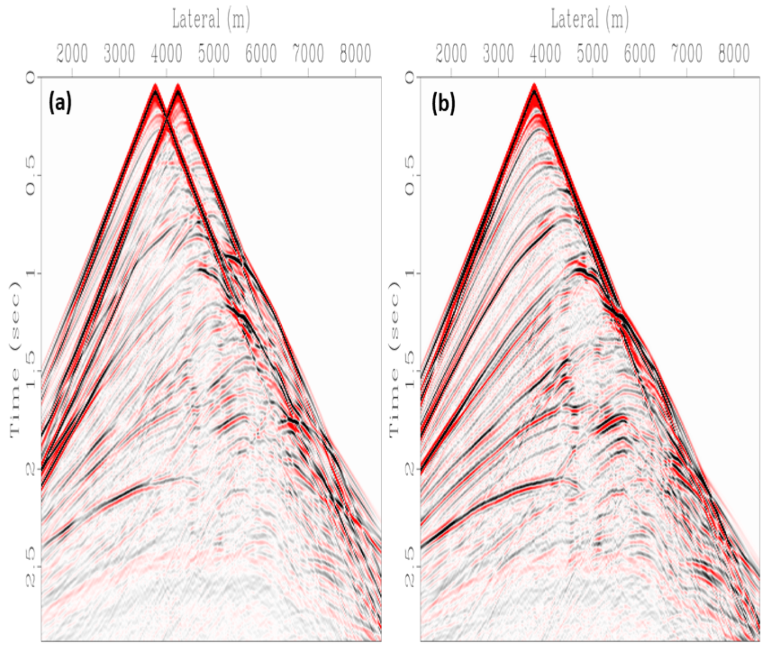

FIG. 7. a) A blended shot produced by finite differences, b) One of the shots predicted by RTM migration/demigration

two-shot blended gather generated by finite differences using the Marmousi model. Figure 8b shows one of the shots in a) after separation by a migration/demigration process using Kirchhoff modeling. This deblending was done by one iteration of a LSMIG algorithm. Better accuracy in amplitudes could be achieved by using more iterations. In this example, I use a smooth velocity model and ray tracing, but simpler structures could be deblended by using time migration instead. Figures 8c shots a three-shots blended gather, and Figure 8d shows one of the shots after separation with Born modeling performed by migration-demigration using a LSRTM algorithm (one iteration). More iterations would lead to better results but with a higher computational cost. In both tests, I use smooth velocity models but otherwise accurate, results would degrade with inaccurate velocity models. Finite-difference methods are more sensitive to velocity errors than ray tracing, although results are more accurate if an exact velocity is used (as in Figure 7b).

Clearly, these examples represent the most accurate end of physical transforms, also the most computationally expensive, where the model is a reflectivity model (in depth or time). The advantage of such transformations is that we can use at full our prior information about the geology of the subsurface. Their disadvantage is that are sensitive to this prior information. However, the idea of migration/demigration can be modified to make it less sensitive to the prior information. Figure 9 shows a deblending sequence using the Apex Shifted Radon Transform (Trad et al., 2012; Ibrahim and Sacchi, 2013; Trad and Siliqi, 2015). The Apex Shifted Radon transform (Trad, 2003) is similar to a LSMIG but uses a modified Stolt mapping to collapse hyperbolic events to their apexes. This transformation is faster than standard Radon transform, but deblending with it works well only on relatively simple scenarios. It is an example of the low-accuracy end of physical transforms, but is fast and relatively robust, since it can use very simple velocity models.

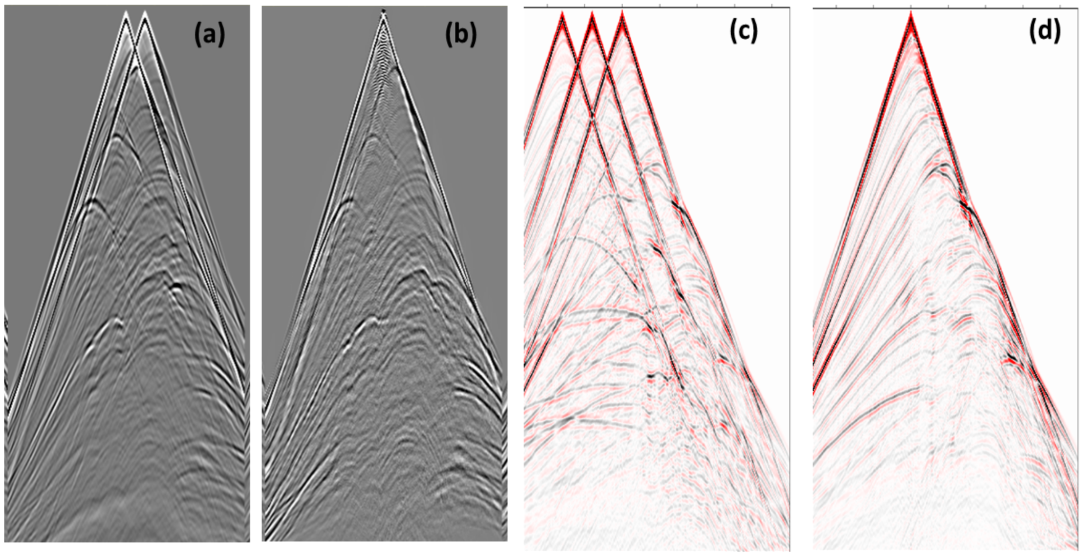

FIG. 8. a) A two-shot gather produced by finite differences; b) deblended shot by Kirchhoff modeling; (c) A three-shot blended gather (finite differences); d) deblended shot by RTM and Born modeling

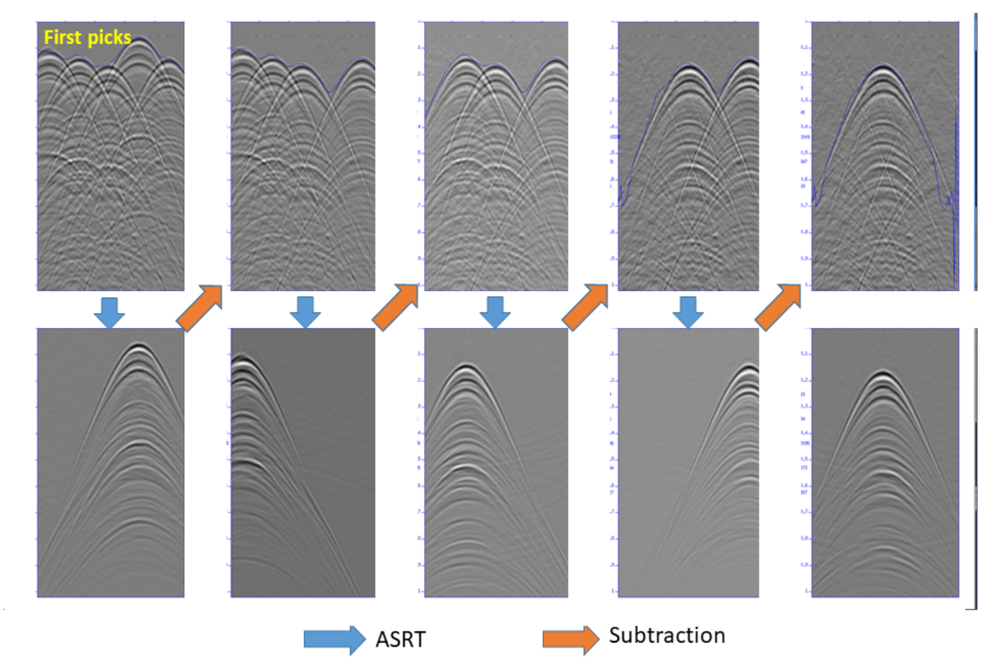

FIG. 9. A five-shot gather deblended by an Apex Shifted Radon Transform sequence

Migration images

It has been shown in the literature and in practice that is possible to get a good image of blended data by direct migration of least squares fitting of simultaneous shots. Similarly, it is possible to perform FWI without deblending, which is often done to diminish computational cost. However, the computational advantage of simultaneous migration is much less important than the economical advantage of cheaper acquisition, since parallelization with computer clusters can be very efficient. In Figure 10 we can see the four most common scenarios. Figure 10a shows a RTM image for Marmousi from 6 shots never blended. Figure 10b shows the RTM for the same 6 shots but this time blended in groups of 3, that is with a computational cost of 2 shots. There is no deblending, but each shot is taken into account during the migration at its correct location, which does not add any additional computation. Some noise is present in the shallow. Figure 10c shows the LSRTM image after 5 iterations, each iteration with a cost of 2 shots (6 shots in groups of 3 again). LSRTM has eliminated some of the noise and increased the resolution somewhat, as expected in LSMIG. However, although the interference noise in the shallow has been attenuated, we can still see it. Figure 10d is the result of the migration from the deblended shots obtained by demigration in individual shots from Figure 10b. The image obtained by using the predicted data from the blended image (Figure 10d) seems to be the least affected by the blended acquisition approach. It is also cheaper to obtain than the LSRTM image (Figure 10c) but it does not show the LSMIG benefits. In summary, one can obtain computational advantages if LSMIG is the goal. Otherwise, simple migration/demigration could be the most cost effective. Alternatively, it is possible to modify the LSRTM dataflow to achieve the deblending internally.

The RTM result in the previous test is very positive because the prediction algorithm (Born modeling) and the data modeling (finite differences) are very similar. Also, the velocity model was accurate. When using an inaccurate or very smooth velocity model, as is always the case for real data, the prediction loses some information. Applying iterations, as in LSRTM fitting, the predictions can be made to match the input better, but the computational cost makes this non-practical. For the case of Kirchhoff modeling, the situation is more difficult than for RTM because the modeling operator for demigration (Kirchhoff) is very different from the mechanism that generates the data (finite differences in this case).

At the time of this report, my implementation of Kirchhoff migration with simultaneous shots is not ready yet. Therefore, the examples in Figure 11 serve mostly just as a quality control (QC) for the Kirchhof deblending. If the image deteriorates significantly, it means that the predicted data have lost information valuable for the migration. Figure 11a shows the Kirchhoff image for 50 individual shots, Figure 11b shows the image for 25 blended shots, although in this case the blended part (25 shots) are only used as interference. Although the deblended shots looked reasonable (Figure 8b), migration from the predicted data (Figure 11c) shows that information is missing from the predictions. Traveltime tables have deficiencies for complex models like Marmousi, so we can see many problems in the image, in particular for the shallow part where rays disappear at critical reflections and traveltime tables are incomplete. This is a very challenging model for Kirchhoff modeling, but considering that the velocity model at the time of deblending is poorly approximated, it is important to consider alternatives. For example, we can use the predicted data not as

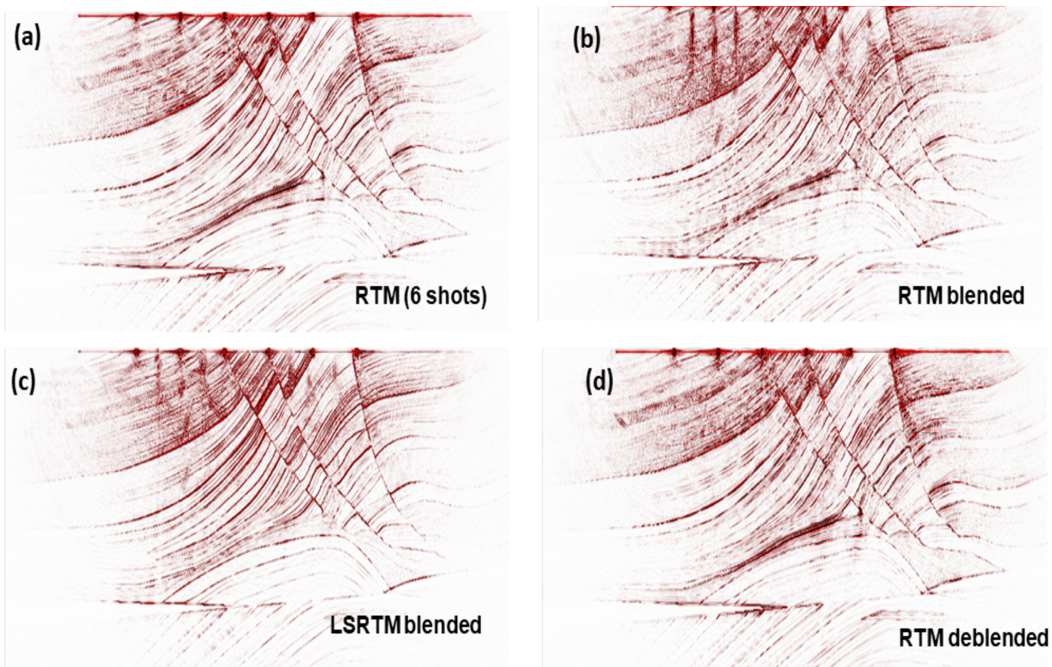

FIG. 10. a) RTM for 6 individual shots, b) RTM for 6 shots blended in groups of 3 (2 blended shots), c) LSRTM for the same 6 shots (groups of 3) with 5 iterations (migration cost of LSRTM for 2 individual shots), d) RTM for deblended shots from b) by migration/demigration of 2 shots with groups of 3).

denoised data but simply as prior information to be used in other algorithms.

Predictions as prior information

An alternative to using predictions from demigration as the processed data, we can use them merely as prior information for other deblending tools. That relieves the strong constraint of getting true amplitude predictions, which usually requires least squares fitting. There are many possible dataflows to follow this approach. Figure 12 shows one possibility, where data are migrated/demigrated in an iterative process like for example conjugate gradients (CG). The estimated model at each iteration can be used to generate predictions in new desired locations, which can be feed to the migration operator together with the input (blended) data. Also, this approach can be used for data regularization and interpolation using physical transforms (Trad, 2015). A problem with this approach is that there are no residuals between original and predicted data (since they have different locations), so a different adjustment of amplitudes is necessary (model updates based on mapping residuals will not work in this case). Two possibilities are using matching filters and migration in the image space.

An example of using predictions as prior information is shown in Figure 13, taken from Trad (2015). The raw input data in Figure 13a, is migrated and demigrated (or generated by a few iterations of LSMIG) to produce the data in Figure 13b. The two data sets are now combined using double guided interpolation Trad (2014). This method permits interpolation and denoising for the raw data shown in Figure 13a. The predictions in Figure 13b

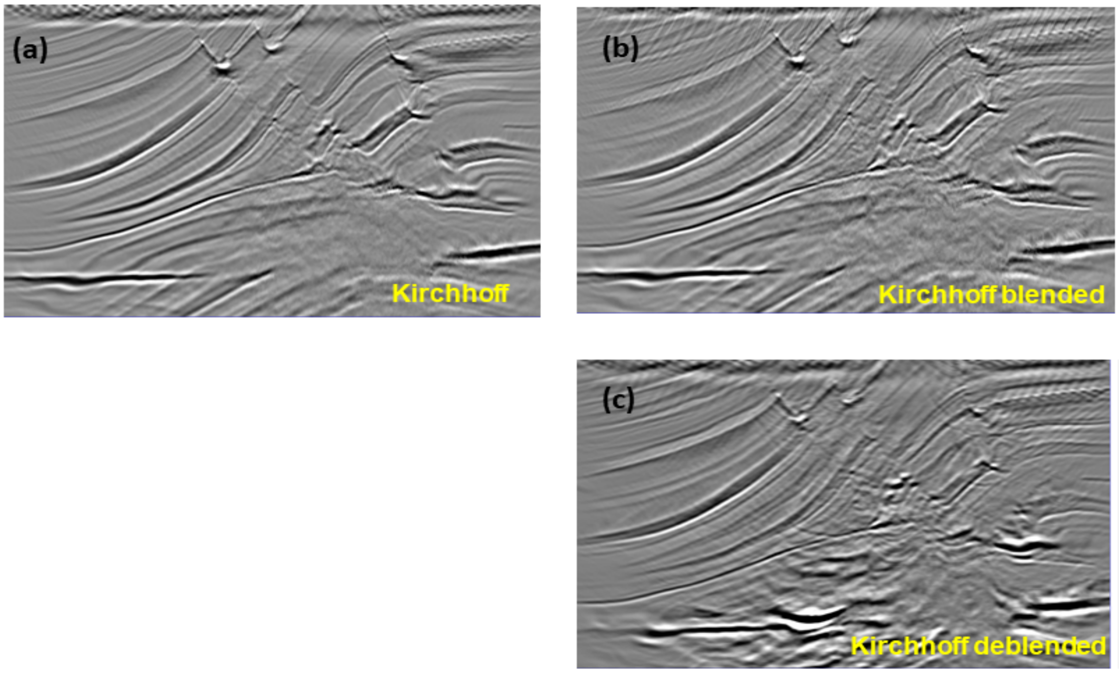

FIG. 11. a) Kirchhoff migration for 50 shots, b) Kirchhoff migration for 25 blended shots with blended interference, c) Kirchhof migration for 25 shots predicted from b) with interference removed

provide a guide for spectral weights and initial model.

CONCLUSIONS

In this paper, I argued that CS, as used in geophysics, is not a new paradigm but a combination of two pre-existing geophysical techniques (sparse transforms and deblending). On the other hand, CS as used in mathematics is the same as sparse transforms and it has been used in geophysics for four decades. Also, I discussed the challenges of deblending and present examples of deblending technology using physical transforms. These transforms, although more difficult to implement and computationally expensive, have the potential for using geophysical prior information in the form of velocity and geological structure.

ACKNOWLEDGMENTS

I thank CMC Research Institutes, Inc (CMC). for access to the seismic data and CREWES sponsors for contributing to this seismic research. I also gratefully acknowledge support from NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13.

REFERENCES

Abma, R., Howe, D., Foster, M., Ahmed, I., Tanis, M., Zhang, Q., Arogunmati, A., and Alexander, G., 2015, Independent simultaneous source acquisition and processing: Geophysics, **80**, No. 6, WD37–WD44.

Abma, R., and Yan, J., 2009, Separating simultaneous sources by inversion, *in* 71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009.

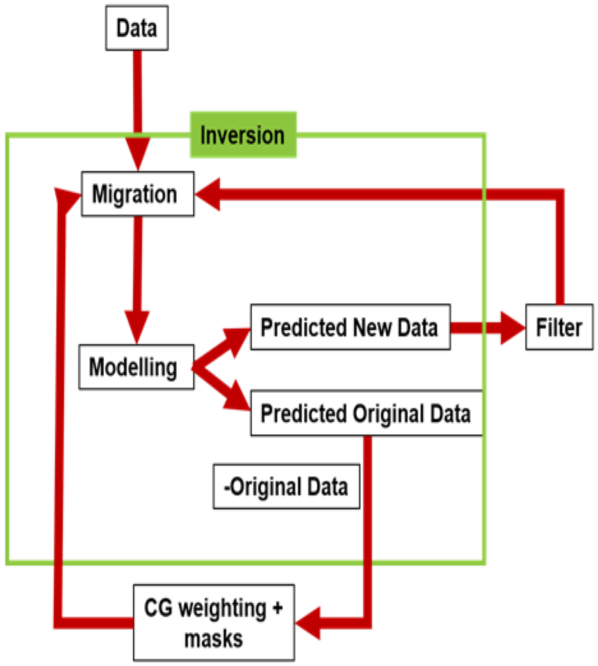

FIG. 12. A general flow for combining predictions from migration model with original data. Filter can include other processes, like 5D regularization or matching filters

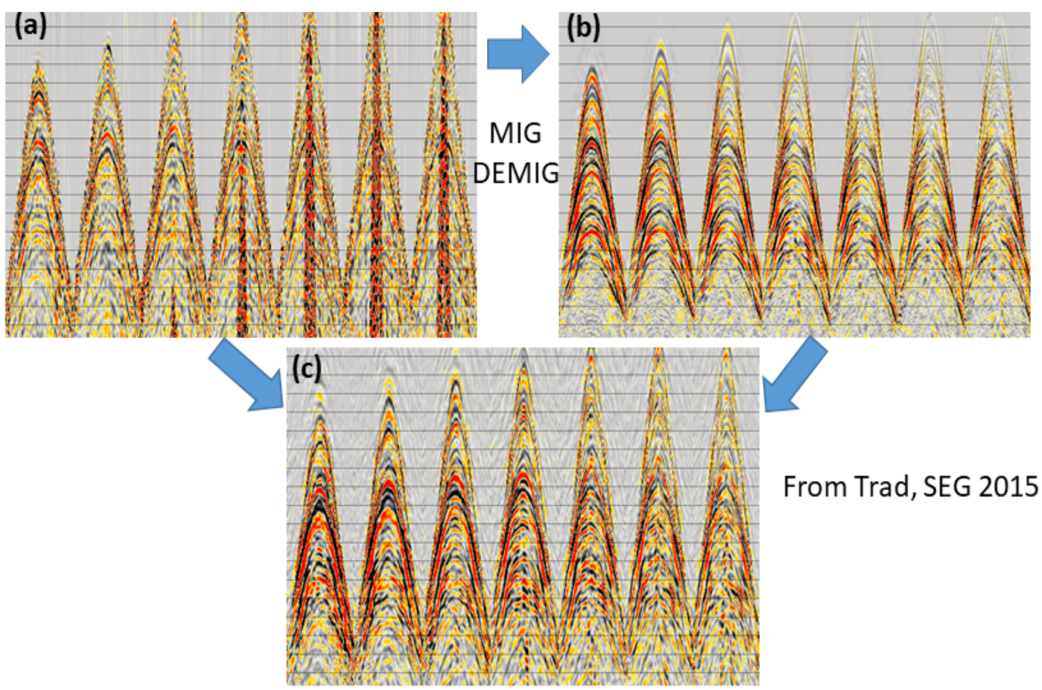

FIG. 13. a) Raw real shots before noise attenuation, b) predicted shots, c) denoised shots obtained by using predictions in b) as an inversion guided for 5D (from Trad (2015)

Beasley, C. J., Chambers, R. E., and Jiang, Z., 1998, A new look at simultaneous sources, *in* SEG Technical Program Expanded Abstracts 1998, Society of Exploration Geophysicists, 133–135.

Candes, E. J., Romberg, J. K., and Tao, T., 2006, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, **59**, No. 8, 1207–1223.

Candès, E. J., and Wakin, M. B., 2008, An introduction to compressive sampling: IEEE signal processing magazine, **25**, No. 2, 21–30.

Candès, E. J. et al., 2006, Compressive sampling, *in* Proceedings of the international congress of mathematicians, vol. 3, Madrid, Spain, 1433–1452.

Claerbout, J., 1992, Earth sounding analysis, Processing versus inversion: Blackwell Scientific Publications, Inc.

Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: Geophysics, **38**, No. 5, 826–844.

Duijndam, A., Schonewille, M., Romijn, R., Koek, E., and Ongkiehong, L., 1999, A general reconstruction scheme for dominant azimuth 3d seismic data, *in* SEG Technical Program Expanded Abstracts 1999, Society of Exploration Geophysicists, 1441–1444.

Harlan, W. S., Claerbout, J. F., and Rocca, F., 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, 1869–1880.

Herrmann, F. J., and Hennenfent, G., 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, No. 1, 233–248.

Ibrahim, A., and Sacchi, M. D., 2013, Simultaneous source separation using a robust radon transform: Geophysics, **79**, No. 1, V1–V11.

Liu, B., and Sacchi, M. D., 2004, Minimum weighted norm interpolation of seismic records: GEOPHYSICS, **69**, No. 6, 1560–1568.

Mahdad, A., Doulgeris, P., and Blacquiere, G., 2011, Separation of blended data by iterative estimation and subtraction of blending interference noise: Geophysics, **76**, No. 3, Q9–Q17.

Sacchi, M. D., and Ulrych, T. J., 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, No. 4, 1169–1177.

Sacchi, M. D., and Ulrych, T. J., 1996, Estimation of the discrete fourier transform, a linear inversion approach: GEOPHYSICS, **61**, No. 4, 1128–1136.

Sacchi, M. D., Ulrych, T. J., and Walker, C. J., 1998, Interpolation and extrapolation using a high-resolution discrete fourier transform: IEEE Transactions on Signal Processing, **46**, No. 1, 31–38.

Santosa, F., and Symes, W. W., 1986, Linear inversion of band-limited reflection seismograms: SIAM Journal on Scientific and Statistical Computing, **7**, No. 4, 1307–1330.

Stanton, A., and Sacchi, M., 2013, All roads lead to rome: Predictability, sparsity, rank and pre-stack seismic data reconstruction: CSEG Recorder, **38**, 1919–1928.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation, vol. 89: siam.

Thorson, J. R., and Claerbout, J. F., 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, **50**, No. 12, 2727–2741.

Trad, D., 2003, Interpolation and multiple attenuation with migration operators: Geophysics, **68**, No. 6, 2043–2054.

Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: GEOPHYSICS, **74**, No. 6, V123–V132.

Trad, D., 2014, Five-dimensional interpolation: New directions and challenges: CSEG Recorder, March, 40–46.

Trad, D., 2015, Least squares kirchhoff depth migration: implementation, challenges, and opportunities, *in* 2015 SEG Annual Meeting, Society of Exploration Geophysicists, 4238–4242.

Trad, D., and Siliqi, R., 2015, Device and method for de-blending simultaneous shooting data with apex shifted radon transform, uS Patent 9,091,788.

Trad, D., Siliqi, R., Poole, G., Boelle, J.-L. et al., 2012, Fast and robust deblending using apex shifted radon transform, *in* 2012 SEG Annual Meeting, Society of Exploration Geophysicists.

Trad, D., Ulrych, T., and Sacchi, M., 2003, Latest views of the sparse radon transform: GEOPHYSICS, **68**, No. 1, 386–399.

Trad, D. O., and Ulrych, T. J., 1999, Radon transform: beyond aliasing with irregular sampling, *in* 6th International Congress of the Brazilian Geophysical Society.

Vermeer, G. J., 2009, Wide-azimuth towed-streamer data acquisition and simultaneous sources: The Leading Edge, **28**, No. 8, 950–958.