GEOPHYSICS, VOL. 80, NO. 6 (NOVEMBER-DECEMBER 2015); P. WD129–WD141, 13 FIGS., 1 TABLE.
10.1190/GEO2015-0066.1

Multidimensional simultaneous random plus erratic noise attenuation and interpolation for seismic data by joint low-rank and sparse inversion

Raphael Sternfels[1], Ghislain Viguier[1], Regis Gondoin[1], and David Le Meur[1]

ABSTRACT

We have developed an efficient convex optimization strategy enabling the simultaneous attenuation of random and erratic noise with interpolation in prestack seismic data. For a particular analysis window, frequency slice spatial data were reorganized into a block Toeplitz matrix with Toeplitz blocks as in Cadzow/singular spectrum analysis methods. The signal and erratic noise were, respectively, modeled as low-rank and sparse components of this matrix, and then a joint low-rank and sparse inversion (JLRSI) enabled us to recover the low-rank signal component from noisy and incomplete data thanks to joint minimization of a nuclear norm term and an L_1-norm term. The convex optimization framework, related to recent developments in the field of compressed sensing, enabled the formulation of a well-posed problem as well as the use of state-of-the-art algorithms. We proposed an alternating directions method of multipliers scheme associated with an efficient singular value thresholding kernel. Numerical results on field data illustrated the effectiveness of the JLRSI approach at recovering missing data and increasing the signal-to-noise ratio.

INTRODUCTION

Various techniques have been proposed to improve the signal-to-noise ratio (S/N) of seismic data by attenuating incoherent noise, including prediction error filtering (Canales, 1984), projection filtering (Soubaras, 1995), and, more recently, rank-reduction filtering, which assumes that the signal component of a matrix or tensor formed from the input data is of low rank. In the last category, we can differentiate eigenimage filtering (Trickett, 2003), Cadzow/singular spectrum analysis (SSA) filtering (Trickett, 2008; Sacchi, 2009), and tensor methods (Kreimer and Sacchi, 2012; Kreimer et al., 2013; Trickett et al., 2013; Da Silva and Herrmann, 2014). These rank-reduction methods have also been extended to robust noise attenuation, dealing with erratic noise as well as data interpolation (Trickett et al., 2010, 2012; Oropeza and Sacchi, 2011; Aravkin et al., 2014; Chen and Sacchi, 2015). Improving the S/N by noise attenuation and missing trace interpolation is key to certain seismic processes because strong noise and suboptimal spatial sampling from acquisition constraints can have an impact on the final migration quality (Trad, 2009). More specific methods addressing the problem of recovery from incomplete data include Fourier-based techniques such as the minimum weighted-norm interpolation (MWNI) (Liu and Sacchi, 2004; Trad, 2009) and the antileakage Fourier transform (ALFT), which handles irregularly sampled data by honoring the real spatial coordinates (Xu et al., 2005, 2010; Poole, 2010), projection onto convex sets (POCS) (Abma and Kabir, 2006), and techniques based on local transforms such as the Radon transform (Trad et al., 2002) or the curvelet transform (Herrmann and Hennenfent, 2008; Hennenfent et al., 2010). Such advances have been fueled by the recent and increasing use in the geophysical community of compressed sensing (Herrmann, 2010; Herrmann et al., 2013; Aravkin et al., 2014) and advanced convex optimization techniques (Pham et al., 2014). Regarding reconstruction methods, the main assumption is that a structured signal is inherently sparse or low-rank in a certain domain, so that limited measurements in the physical domain still hold enough information to accurately recover the signal.

In this paper, we take a new look at the signal model stemming from Cadzow/SSA techniques in which the coherent signal is modeled as a low-rank component of the so-called trajectory matrix, the latter being a Hankel or Toeplitz matrix (or block Hankel/Toeplitz with Hankel/Toeplitz blocks in the case of several spatial dimensions) formed from data in the temporal Fourier domain. We model erratic noise as sparse elements of this matrix: In the temporal

Manuscript received by the Editor 30 January 2015; revised manuscript received 29 June 2015; published online 15 October 2015.
[1]CGG, Massy, France. E-mail: raphael.sternfels@cgg.com; ghislain.viguier@cgg.com; regis.gondoin@cgg.com; david.lemeur@cgg.com.

Fourier domain, erratic noise can be either localized amplitude anomalies or phase distortions (Trickett et al., 2012). Erratic noise needs to be explicitly taken into account because traditional least-squares filtering approaches making use of the L_2-norm such as Cadzow/SSA are highly sensitive to corrupted data with high-amplitude, non-Gaussian noise. We formulate the simultaneous random plus erratic noise attenuation with interpolation problem as a joint low-rank and sparse inversion (JLRSI) convex optimization problem, in which the low-rank signal component is recovered and separated from the sparse erratic noise component thanks to joint minimization of a nuclear norm term and an L_1-norm term, constrained by the fit to the available, incomplete data. The nuclear norm of a matrix is the sum of its singular values, and it can be considered for rank minimization problems as an equivalent of the L_1-norm enhancing sparsity in compressed sensing (Chandrasekaran et al., 2012). This systematic formulation is motivated by recent progress in matrix completion and rank minimization (Candès and Recht, 2008; Candès and Tao, 2009; Recht et al., 2010) and more specifically principal component pursuit or robust principal component analysis problems, i.e., low-rank plus sparse decompositions (Candès et al., 2009; Wright et al., 2009; Chandrasekaran et al., 2011). Our JLRSI formulation can be seen as an extension of the stable principal component pursuit (Zhou et al., 2010) and compressive principal component pursuit (Wright et al., 2013) problems, in which observed data are noisy and incomplete (see Candès and Plan [2009] for matrix completion with noise), and the low-rank component has a Hankel/Toeplitz structure (see Fazel et al. [2013] for rank minimization of such matrices). Some of the aforementioned references include strong theoretical results based on weak assumptions on the input data and its structure, in addition to providing an elegant framework to formulate our problem as a convex semidefinite program (Vandenberghe and Boyd, 1996; Boyd and Vandenberghe, 2004).

We propose to solve the resulting convex optimization problem for each temporal frequency slice with an efficient alternating direction method of multipliers (ADMM) scheme (Wen et al., 2010; Boyd et al., 2011; Tao and Yuan, 2011), which iteratively estimates and recovers the missing entries of the low-rank signal component of the trajectory matrix, while at the same time characterizing erratic noise as a sparse component. The inner updates within an ADMM iteration amount to evaluating so-called proximal operators (Moreau, 1965; Combettes and Pesquet, 2011; Parikh and Boyd, 2013), of which the one associated with the nuclear norm term, singular value thresholding, is the costliest because it involves a singular value decomposition (SVD) of a Hankel/Toeplitz matrix. Because only the leading terms are needed, we make use of an efficient scheme based on fast Toeplitz (or block Toeplitz with Toeplitz blocks [BTTB]) matrix-vector products (Lee, 1986; Strang, 1986; Van Loan, 1992), recently proposed by Korobeynikov (2010) and used for seismic data filtering and reconstruction purposes by Trickett (2003) and Gao et al. (2013).

This article is organized as follows: First, we expose the motivation behind the Cadzow/SSA signal model; second, we formulate our problem as a well-posed convex program; and third, we describe the ADMM algorithm enabling us to solve the latter efficiently. The effectiveness of our approach is subsequently demonstrated on real data examples, and finally Appendices A and B provide additional details concerning the singular value thresholding step.

THEORY

Modeling in the *f-x* domain

We work from the assumption that the signal is predictable in the *f-x* domain. Here, f represents the temporal frequency and x the spatial dimensions, such as shot-receiver coordinates, midpoint coordinates, offset vector components, offset, or azimuth. In our applications, we will work in up to four spatial dimensions, although in theory, the algorithm can be extended to as many spatial dimensions as one deems necessary. For the sake of clarity, in the spirit of the exposition in Sacchi (2009), let us start with the case of one spatial dimension, in which a signal composed of a single dip can be represented as

$$s(t,x) = u(t - px) \tag{1}$$

in the time domain and as

$$S(\omega,x) = U(\omega)e^{i\omega px} \tag{2}$$

in the *f-x* domain. Thus, for regularly sampled data at $x_n = n\Delta x$ for $n \in \mathbb{N}$ and a fixed frequency ω_k,

$$S_n^{(k)} = S(\omega_k,x_k) = U(\omega_k)e^{i\omega_k px_n} = e^{i\omega_k p\Delta x}S_{n-1}^{(k)}. \tag{3}$$

This last recursive relation, showing the signal predictability assumption, is used in various kinds of *f-x* deconvolution and filtering algorithms.

For our purposes, we introduce some useful notations. Working frequency by frequency and dropping the superscript k, let the trajectory matrix be $\mathbf{H}$ as follows for a window of five traces and a frequency slice $\mathbf{S} = (S_1, \ldots, S_5)$:

$$\mathbf{H}(\mathbf{S}) = \begin{pmatrix} S_1 & S_2 & S_3 \\ S_2 & S_3 & S_4 \\ S_3 & S_4 & S_5 \end{pmatrix}. \tag{4}$$

From the predictability assumption, one can see that for a single dipping event, the columns of the trajectory matrix are scaled versions of each other, and thus, $\mathbf{H}(\mathbf{S})$ has rank 1. Hence, a noiseless data set that is the sum of k-plane waves results in a trajectory matrix having at most rank k. This observation is used by Cadzow/SSA filtering (Trickett, 2008; Sacchi, 2009), which attenuates random noise in the recorded data $\mathbf{D}$ by reducing the rank of the trajectory matrix $\mathbf{T}(\mathbf{D})$ via an SVD, discarding the smaller singular values. The resulting matrix is then averaged along its antidiagonals to yield the denoised signal $\mathbf{S}$. This procedure was previously described in the dynamical systems community (Broomhead and King, 1986), in the signal processing community as Cadzow's method (Cadzow, 1988) in which the SVD truncation and averaging process is repeated, and in statistics and climatic time series analysis as SSA (Vautard and Ghil, 1989; Plaut and Vautard, 1993; Ghil et al., 2002; Golyandina and Zhigljavsky, 2013).

The trajectory matrix has a Hankel structure; for convenience, from now on we will work with the following "equivalent" Toeplitz matrix:

$$\mathbf{T}(\mathbf{S}) = \begin{pmatrix} S_3 & S_2 & S_1 \\ S_4 & S_3 & S_2 \\ S_5 & S_4 & S_3 \end{pmatrix}. \quad (5)$$

Because the order of matrix rows or columns does not affect the rank, it is strictly a matter of convenience whether we form Hankel or Toeplitz matrices. The matrix $\mathbf{T}(\mathbf{S})$ is square of dimension $(N+1)/2$ for a window of N traces if N is odd, and it is rectangular of size $N/2 \times (N/2+1)$ if N is even.

The extension to several spatial dimensions amounts to building block Toeplitz matrices with Toeplitz blocks (BTTB matrices) of an order corresponding to the additional spatial dimensions. For instance, for two spatial dimensions, we have

$$\mathbf{T}(\mathbf{S}) = \begin{pmatrix} \mathbf{T}_3 & \mathbf{T}_2 & \mathbf{T}_1 \\ \mathbf{T}_4 & \mathbf{T}_3 & \mathbf{T}_2 \\ \mathbf{T}_5 & \mathbf{T}_4 & \mathbf{T}_3 \end{pmatrix}, \quad (6)$$

where each $\mathbf{T}_i$ is a Toeplitz matrix. For additional spatial dimensions, the $\mathbf{T}_i$ matrices would be recursively replaced by BTTB matrices. More details on how to build BTTB matrices can be found in the literature, e.g., in Oropeza and Sacchi (2011) and Gao et al. (2013) in which the term multichannel SSA is used when working with additional dimensions. Adding spatial dimensions allows for a better characterization of the signal coherency along those dimensions, leading to a better separation from incoherent, random noise.

Cadzow/SSA denoising is then performed by forming the Toeplitz or BTTB matrix $\mathbf{T}(\mathbf{D})$ from the input frequency slice $\mathbf{D}$; taking the SVD of the matrix; discarding the smaller singular values to yield $\mathbf{T}_{\mathrm{RR}}(\mathbf{D})$; and finally averaging along the (block) diagonals of the resulting matrix, yielding $\mathbf{S} = \mathcal{T}[\mathbf{T}_{\mathrm{RR}}(\mathbf{D})]$, where $\mathcal{T}[\cdot]$ is the averaging operator. This process should possibly be repeated in the case of Cadzow filtering (Cadzow, 1988). Once all frequency slices have been processed, the original block in the time domain is retrieved via an inverse Fourier transform.

Convex formulation of the problem

Let us start by reformulating the previous problem, i.e., random noise attenuation in the recorded data via Cadzow/SSA filtering. For an assumed level of random noise δ such that $\mathbf{D} = \mathbf{S} + \mathbf{Z}_\delta$, where $\mathbf{Z}_\delta$ represents the additive random noise, we have

$$\text{minimize } \mathrm{rank}(\mathbf{T}(\mathbf{S})) \quad \text{subject to } \|\mathbf{D} - \mathbf{S}\|_2 \le \delta, \quad (7)$$

where $\|\cdot\|_2$ is the entrywise sum of squares norm. Although clearly nonconvex, this specific problem is easily dealt with by using the SVD of the matrix $\mathbf{T}(\mathbf{D})$.

Now, let us assume that recorded data $\mathbf{D}$ are also incomplete and corrupted by erratic noise; in the *f-x* domain, erratic noise amounts to spike bursts or localized phase distortions. We encode the available information by the sampling operator $P_\Omega[\cdot]$, where $(P_\Omega[\mathbf{X}])_i = X_i$ if $i \in \Omega$ and 0 otherwise. The signal model becomes $\mathbf{D} = P_\Omega[\mathbf{S} + \mathbf{E} + \mathbf{Z}_\delta]$, where $\mathbf{E}$ represents erratic noise, and applying the Cadzow/SSA procedure described above directly to $\mathbf{T}(\mathbf{D})$ would yield incorrect results because the SVD is optimal only in the least-squares sense, and thus it is very sensitive to outliers and missing data. Iterative schemes have been previously suggested to extend Cadzow/SSA filtering to either erratic noise attenuation or interpolation (Trickett et al., 2010, 2012; Oropeza and Sacchi, 2011) using POCS-like ideas and robust statistics, although none are in in a theoretical framework enabling simultaneous interpolation and erratic noise attenuation, along with the use of modern convex optimization techniques.

Taking into account the missing data via the sampling operator $P_\Omega[\cdot]$ and the sparsity of the erratic noise component $\mathbf{E}$ as measured by its number of nonzero components $\|\mathbf{E}\|_0$ (the L_0-norm in the compressed sensing literature, which is none other than the cardinality function), the following problem formulation is better adapted for the recovery of the true signal subspace:

$$\begin{aligned} &\text{minimize } \mathrm{rank}\,(\mathbf{T}(\mathbf{S})) + \|\mathbf{E}\|_0 \\ &\text{subject to } \|P_\Omega[\mathbf{D} - \mathbf{S} - \mathbf{E}]\|_2 \le \delta. \end{aligned} \quad (8)$$

This much more difficult problem is intractable in most cases. Indeed, the rank and cardinality functions are nonconvex and generally require algorithms running in exponential time to be minimized. Therefore, we seek a solvable surrogate problem. This can be achieved using a convex relaxation; that is, we "convexify" the cost function as follows to give the JLRSI problem formulation:

$$\begin{aligned} &\text{minimize } \|\mathbf{T}(\mathbf{S})\|_* + \lambda\|\mathbf{E}\|_1 \\ &\text{subject to } \|P_\Omega[\mathbf{D} - \mathbf{S} - \mathbf{E}]\|_2 \le \delta, \end{aligned} \quad (9)$$

where $\|\cdot\|_*$ is the nuclear norm, i.e., the sum of the singular values; $\|\cdot\|_1$ is the entrywise L_1-norm, i.e., the sum of absolute values; and λ is a regularization parameter. This is a convex optimization problem, and although it is nonsmooth, it can be solved efficiently as it will be seen in the next section.

Let us now develop the motivation and the benefits behind this "convexification" before dealing with the description of the proposed ADMM algorithm. The L_1-norm is well known and was used by geophysicists as early as the 1970s and 1980s to deal with erratic data (Claerbout and Muir, 1973; Taylor et al., 1979) and sparse inversion (see Trad et al. [2003] for a review concerning the sparse Radon transform). However, it is only recently with the appearance and expansion of the compressed sensing field (Candès et al., 2006a, 2006b) that a solid mathematical framework emerged: Mathematicians proved the well posedness of such problems, with the L_1-norm acting as a well-behaved proxy for the cardinality function. The nuclear norm heuristic for rank minimization was first introduced in the control community (Fazel et al., 2001), and it was not long before mathematicians unearthed conditions for optimal solutions and recovery for matrix completion problems (Candès and Recht, 2008; Candès and Tao, 2009; Recht et al., 2010). Motivated by this newly developed framework and the shortcomings of classical principal component analysis with outliers, researchers studied the principal component pursuit or robust principal component analysis problem, i.e., the problem of separating a matrix into a sum of a low-rank component and a sparse component (Candès et al., 2009; Wright et al., 2009; Chandrasekaran et al., 2011). Those works include strong theoretical results under various conditions on the matrix structure (for instance, one cannot expect to separate matrices that are low rank and sparse) and have been extended to cases in which available data are limited and/or contaminated with random noise (Zhou et al., 2010; Wright et al., 2013), helped by developments on sparse recovery and matrix completion problems (Candès et al., 2006b; Candès and Plan, 2009). In all those

advances, convex relaxations and mathematical arguments rooted in high-dimensional geometry and functional analysis are essential ingredients, and we let the reader further delve into the provided references to appreciate those achievements and the underlying proofs. We end this interesting detour by mentioning recent developments (Andersson et al., 2014; Condat and Hirabayashi, 2015) in the signal processing community regarding the spectral estimation problem, for which the Cadzow denoising method was originally proposed. In these works, in contrast to Fazel et al. (2013), Hankel/Toeplitz matrices are constrained to be low rank without convex relaxations.

The principal component pursuit problem inspired our JLRSI convex formulation, in which the low-rank signal component takes the form of a trajectory matrix in the vein of the Cadzow/SSA model. We could have directly applied the principal component pursuit theory to the trajectory matrix formed from the input data; although in addition to the added redundancy, it would have been algorithmically less efficient because in the proposed algorithm, the trajectory matrix is actually never formed nor stored in memory. Finally, the following equivalent formulation to equation 9 (see Aybat et al. [2014] for an adaptable proof) will be used in the subsequent sections for convenience purposes:

$$\begin{aligned} &\text{minimize } \|\mathbf{T}(\mathbf{S})\|_* + \lambda\|P_\Omega[\mathbf{E}]\|_1 \\ &\text{subject to } P_\Omega[\mathbf{D}] = \mathbf{S}+\mathbf{E}+\mathbf{Z}, \quad \|\mathbf{Z}\|_2 \le \delta. \end{aligned} \tag{10}$$

Alternating directions method of multipliers algorithm and proximal operators

The JLRSI convex program formulated previously can be solved by specialized techniques for nonlinear programming such as interior-point methods (Boyd and Vandenberghe, 2004; Nocedal and Wright, 2006), but those second-order schemes become prohibitively expensive for trajectory matrices of practical size. Therefore, we suggest the use of a fast first-order scheme, converging quickly to a more modest accuracy compared to second-order schemes, which in our case is sufficient. Such first-order schemes include, among others: Nesterov's method, accelerated proximal gradient , and fast iterative shrinkage-thresholding (Nesterov, 2005; Beck and Teboulle, 2009), of which specific variants have been developed for principal component pursuit problems (Ganesh et al., 2009; Aybat et al., 2014). We suggest the use of the ADMM (Wen et al., 2010; Boyd et al., 2011; Tao and Yuan, 2011) for its simplicity and ease of implementation. ADMM originated in the seventies (Glowinski and Marroco, 1975; Gabay and Mercier, 1976) and has since found numerous applications and has been comprehensively analyzed by many authors, showing close links with independently developed techniques such as Douglas-Rachford splitting (Eckstein and Bertsekas, 1992; Combettes and Pesquet, 2007). Furthermore, it has been recently emphasized that ADMM is a special case of the broader class of primal-dual proximal splitting algorithms: The reader is referred to Komodakis and Pesquet (2014) for a comprehensive review of these methods.

As a starting point, the variable splitting performed in equation 10 for our problem leads to the following augmented Lagrangian formulation:

$$\begin{aligned} L(\mathbf{S},\mathbf{E},\mathbf{Z},\mathbf{M}) = {} & \|\mathbf{T}(\mathbf{S})\|_* + \lambda\|P_\Omega[\mathbf{E}]\|_1 \\ & + \langle \mathbf{M}, P_\Omega[\mathbf{D}] - \mathbf{S} - \mathbf{E} - \mathbf{Z}\rangle \\ & + \frac{1}{2\mu}\|P_\Omega[\mathbf{D}] - \mathbf{S} - \mathbf{E} - \mathbf{Z}\|_2^2 + \iota_{\{\|\mathbf{Z}\|_2 \le \delta\}}(\mathbf{Z}), \end{aligned} \tag{11}$$

where $\mathbf{M}$ is the dual-variable or Lagrange multiplier, μ is the parameter associated with the added quadratic penalty, and $\iota_{\mathcal{C}}(\cdot)$ is the characteristic function of the set $\mathcal{C}$: $\iota_{\mathcal{C}}(c) = 0$ if $c \in \mathcal{C}$ and $+\infty$ if $c \notin \mathcal{C}$. The added quadratic penalty term in equation 11 (justifying the augmented Lagrangian name) aims at "robustifying" the cost function and solution procedure (Nocedal and Wright, 2006; Boyd et al., 2011). Using this augmented Lagrangian, the kth iteration of the method of multipliers consists of the following:

$$(\mathbf{S}^{k+1},\mathbf{E}^{k+1},\mathbf{Z}^{k+1}) = \underset{\mathbf{S},\mathbf{E},\mathbf{Z}}{\operatorname{argmin}} L(\mathbf{S},\mathbf{E},\mathbf{Z},\mathbf{M}^k) \tag{12}$$

and

$$\mathbf{M}^{k+1} = \mathbf{M}^k + \frac{1}{\mu}(P_\Omega[\mathbf{D}] - \mathbf{S}^{k+1} - \mathbf{E}^{k+1} - \mathbf{Z}^{k+1}). \tag{13}$$

In contrast with the method of multipliers, the first subproblem is hard to solve efficiently because of the nonseparability of the objective function, which is due to the additional quadratic penalty term. ADMM enables us to tackle this difficulty by splitting and updating the variables one at a time before updating the multiplier. We formulate here the kth ADMM iteration:

$$\mathbf{Z}^{k+1} = \underset{\mathbf{Z}}{\operatorname{argmin}} L(\mathbf{S}^k,\mathbf{E}^k,\mathbf{Z},\mathbf{M}^k), \tag{14}$$

$$\mathbf{E}^{k+1} = \underset{\mathbf{E}}{\operatorname{argmin}} L(\mathbf{S}^k,\mathbf{E},\mathbf{Z}^{k+1},\mathbf{M}^k), \tag{15}$$

$$\mathbf{S}^{k+1} = \underset{\mathbf{S}}{\operatorname{argmin}} L(\mathbf{S},\mathbf{E}^{k+1},\mathbf{Z}^{k+1},\mathbf{M}^k), \tag{16}$$

and

$$\mathbf{M}^{k+1} = \mathbf{M}^k + \frac{1}{\mu}(P_\Omega[\mathbf{D}] - \mathbf{S}^{k+1} - \mathbf{E}^{k+1} - \mathbf{Z}^{k+1}). \tag{17}$$

Unlike the two-variables case, there is currently no strict proof of convergence for ADMM when the number of updated variables is three or higher, although there has been some recent effort in this direction (Chen et al., 2014; Sun et al., 2014). Despite this absence of a guarantee, the extension works well in some applications, e.g., in Tao and Yuan (2011), and satisfactory behavior was observed in our examples. The entire kth ADMM iteration can finally be rewritten in the following form, after merging the relevant multiplier term with the quadratic penalty term:

$$\begin{aligned} \mathbf{Z}^{k+1} = {} & \underset{\mathbf{Z}}{\operatorname{argmin}}\, \iota_{\{\|\mathbf{Z}\|_2 \le \delta\}}(\mathbf{Z}) \\ & + \frac{1}{2\mu}\|P_\Omega[\mathbf{D}] - \mathbf{S}^k - \mathbf{E}^k - \mathbf{Z} + \mu\mathbf{M}^k\|_2^2, \end{aligned} \tag{18}$$

$$\mathbf{E}^{k+1} = \underset{\mathbf{E}}{\operatorname{argmin}} \|P_\Omega[\mathbf{E}]\|_1 + \frac{1}{2\mu\lambda}\|P_\Omega[\mathbf{D}] - \mathbf{S}^k - \mathbf{E} - \mathbf{Z}^{k+1} + \mu\mathbf{M}^k\|_2^2, \quad (19)$$

$$\mathbf{S}^{k+1} = \underset{\mathbf{S}}{\operatorname{argmin}} \|\mathbf{T}(\mathbf{S})\|_* + \frac{1}{2\mu}\|P_\Omega[\mathbf{D}] - \mathbf{S} - \mathbf{E}^{k+1} - \mathbf{Z}^{k+1} + \mu\mathbf{M}^k\|_2^2, \quad (20)$$

and

$$\mathbf{M}^{k+1} = \mathbf{M}^k + \frac{1}{\mu}(P_\Omega[\mathbf{D}] - \mathbf{S}^{k+1} - \mathbf{E}^{k+1} - \mathbf{Z}^{k+1}). \quad (21)$$

The subproblems for $\mathbf{Z}$, $\mathbf{E}$, and $\mathbf{S}$ involve the so-called proximal operators (Moreau, 1965; Combettes and Pesquet, 2011; Parikh and Boyd, 2013); solving one subproblem is equivalent to evaluating the proximal operator of a certain function. The proximal operator of a function f is defined by

$$\operatorname{prox}_f(v) = \underset{x}{\operatorname{argmin}} f(x) + \frac{1}{2}\|x - v\|_2^2. \quad (22)$$

Proximal operators enjoy various useful properties and interpretations (Parikh and Boyd, 2013), a noteworthy one being that a minimizer of a function is a fixed point of its proximal operator. Those connections enabled the development and analysis of several convex optimization methods, including ADMM. In this new light, one can rewrite the subproblems for $\mathbf{Z}$, $\mathbf{E}$, and $\mathbf{S}$ as follows:

$$\mathbf{Z}^{k+1} = \operatorname{prox}_{\mu\iota_{\{\|\cdot\|_2\le\delta\}}}(P_\Omega[\mathbf{D}] - \mathbf{S}^k - \mathbf{E}^k + \mu\mathbf{M}^k), \quad (23)$$

$$\mathbf{E}^{k+1} = \operatorname{prox}_{\mu\lambda\|P_\Omega[\cdot]\|_1}(P_\Omega[\mathbf{D}] - \mathbf{S}^k - \mathbf{Z}^{k+1} + \mu\mathbf{M}^k), \quad (24)$$

and

$$\mathbf{S}^{k+1} = \operatorname{prox}_{\mu\|\mathbf{T}(\cdot)\|_*}(P_\Omega[\mathbf{D}] - \mathbf{E}^{k+1} - \mathbf{Z}^{k+1} + \mu\mathbf{M}^k). \quad (25)$$

Although the functions on which the proximal operator is evaluated are nonsmooth, we fortunately have simple semianalytical formulas that can be derived from the properties of proximal operators (Combettes and Pesquet, 2011; Parikh and Boyd, 2013). Concerning the characteristic function $\iota_{\{\|\cdot\|_2\le\delta\}}$ of an Euclidean norm ball, the proximal operator reduces to the projection operator:

$$\operatorname{prox}_{\iota_{\{\|\cdot\|_2\le\delta\}}}(v) = \Pi_{\{\|\cdot\|_2\le\delta\}}(v) = \underset{\|x\|_2\le\delta}{\operatorname{argmin}} \|x - v\|_2 = \frac{\min(\|v\|_2, \delta)}{\|v\|_2} v. \quad (26)$$

For the entrywise L_1-norm, the operation applied is soft thresholding as

$$(\operatorname{prox}_{\alpha\|\cdot\|_1}(\mathbf{A}))_i = \frac{A_i}{|A_i|}\max(|A_i| - \alpha, 0), \quad (27)$$

and for the nuclear norm, we apply the singular value thresholding operator as

$$\operatorname{prox}_{\beta\|\cdot\|_*}(\mathbf{B}) = \sum_{i=1}^{n} \max(\sigma_i - \beta, 0) u_i v_i^T, \quad (28)$$

where $\mathbf{B} = \sum_{i=1}^n \sigma_i u_i v_i^T$ is the SVD of $\mathbf{B}$. Using the properties of proximal operators in conjunction with the previous formulas, the resulting ADMM iterations can be written as follows:

$$\text{Initialization}: \mathbf{Z}^0 = 0, \mathbf{E}^0 = 0, \mathbf{S}^0 = 0, \mathbf{M}^0 = \mathbf{D}/\|\mathbf{D}\|_2.$$

$$\mathbf{Z}^{k+1} = \frac{\min(\|\mathbf{V}\|_2, \delta)}{\|\mathbf{V}\|_2}\mathbf{V}, \quad \text{with } \mathbf{V} = P_\Omega[\mathbf{D}] - \mathbf{S}^k - \mathbf{E}^k + \mu\mathbf{M}^k; \quad (29)$$

$$(\mathbf{E}^{k+1})_i = \begin{cases} \frac{A_i}{|A_i|}\max(|A_i| - \mu\lambda, 0) & \text{if } i \in \Omega \\ A_i & \text{otherwise,} \\ \text{with } \mathbf{A} = P_\Omega[\mathbf{D}] - \mathbf{S}^k - \mathbf{Z}^{k+1} + \mu\mathbf{M}^k; \end{cases} \quad (30)$$

$$\mathbf{S}^{k+1} = \mathcal{T}\left[\sum_r \max(\sigma_r - \mu, 0) u_r v_r^T\right], \quad \text{where } \mathbf{T}(P_\Omega[\mathbf{D}] - \mathbf{E}^{k+1} - \mathbf{Z}^{k+1} + \mu\mathbf{M}^k) = \sum_r \sigma_r u_r v_r^T \text{ is an SVD;} \quad (31)$$

and

$$\mathbf{M}^{k+1} = \mathbf{M}^k + \frac{1}{\mu}(P_\Omega[\mathbf{D}] - \mathbf{S}^{k+1} - \mathbf{E}^{k+1} - \mathbf{Z}^{k+1}). \quad (32)$$

Iterations are stopped when $\|\mathbf{S}^{k+1} - \mathbf{S}^k\|_2 + \|\mathbf{E}^{k+1} - \mathbf{E}^k\|_2 \le \varepsilon$ for user-defined ε or when a prescribed maximum number of iterations is reached. We recall that $\mathbf{T}(\cdot)$ is the operator tranforming a frequency slice into a Toeplitz (or BTTB) matrix and $\mathcal{T}[\cdot]$ is its adjoint operator, averaging the (block) diagonals of a Toeplitz (BTTB) matrix.

Because $\mathbf{Z}, \mathbf{E}, \mathbf{M}$ updates require $O(N)$ floating point operations and storage for a window of N traces, the computational cost per iteration is dominated by the $\mathbf{S}$ update, which involves an SVD. In the appendices, we will describe how this step can be carried out efficiently in $O(N \log N)$ floating point operations with $O(N)$ storage thanks to the low-rank requirements and the Toeplitz (BTTB) matrix structure, enabling fast matrix-vector products (Lee, 1986; Strang, 1986; Van Loan, 1992; Korobeynikov, 2010; Gao et al., 2013) to be used in a robust randomized SVD scheme (Halko et al., 2011; Martinsson et al., 2011), without explicitly forming nor storing in memory Toeplitz (BTTB) matrices.

NUMERICAL EXAMPLES

In the following numerical examples, in the spirit of Candès et al. (2009), we used $\lambda = \sqrt{N/|\Omega|n_{\max}}$ and $\mu = 4\|P_\Omega[\mathbf{D}]\|_1/|\Omega|$, where in the analysis window, $|\Omega|$ is the number of observed traces, N is

the total number of traces including the ones to be interpolated, $n_{\max}$ is the maximum extent of the spatial dimensions, and **D** is the input frequency slice. The number of relevant singular values to be calculated by the scheme described in the appendix was determined adaptively by examining the smallest computed singular values at each iteration, and testing showed that a maximum of a few tens of ADMM iterations were sufficient to reach the convergence.

We first apply the proposed JLRSI method to a land data set arranged in 36 common offset-vector (COV) volumes (Cary, 1999; Vermeer, 2000) corrected for normal moveout (NMO), with inline and crossline increments of 15 m. In this example, we used a total of four spatial dimensions: inline, crossline, and both components of the offset vector. The processing window extent was 600 ms in the temporal direction, 30 traces in the inline and crossline directions, and 4 traces in both binned offset vector component directions. Figure 1 shows the original central COV volume and the stack of the original 36 COV volumes, with prestack data exhibiting a poor S/N. Then, we randomly decimated these data to keep 25% of all original traces in the input prestack volume, and we randomly added high-amplitude spikes acting as additional erratic noise. The corrupted and decimated central COV volume is shown before and after application of the JLRSI reconstruction process in Figure 2; one can discern horizontal events in the reconstructed section in contrast to the original section. The difference between the original and reconstructed central COV volumes is shown in Figure 3. Figure 4 shows the stack of the 36 corrupted and decimated COV volumes before and after JLRSI recovery. Finally, Figure 5 shows the stack of the reconstructed COV volumes compared with the original stack, their difference demonstrating that JLRSI provides effective noise attenuation and reconstruction while still preserving the character of the data, with minimum signal leakage.

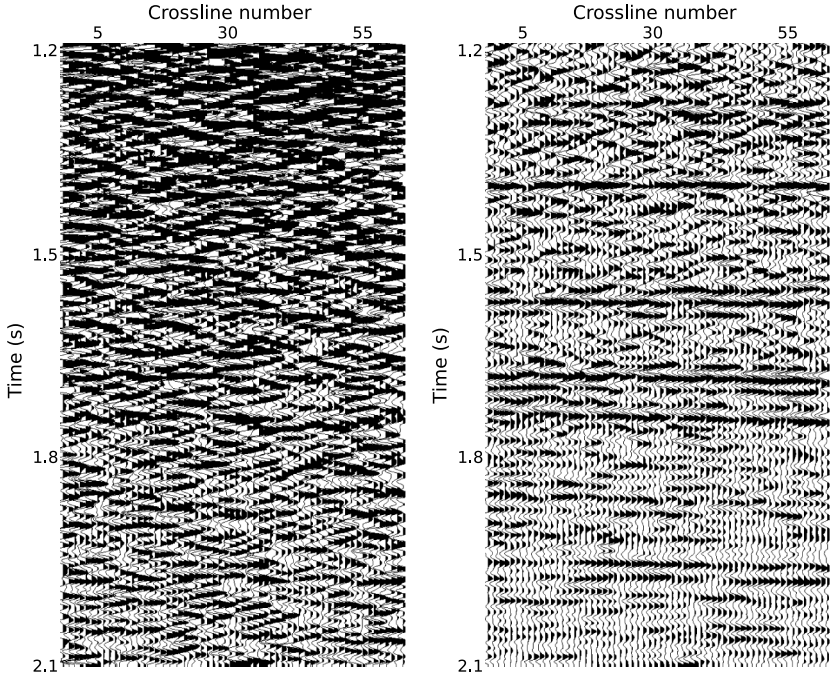

Figure 1. (a) Original central COV volume. The S/N is very low. (b) Stack of the original undecimated, uncorrupted COV volumes.

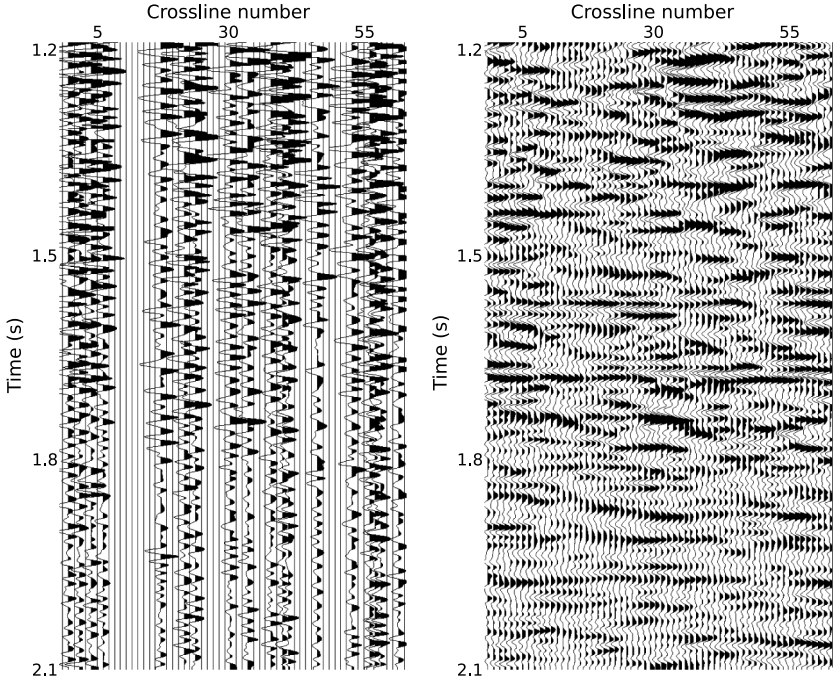

Figure 2. (a) Central COV volume, with 25% traces remaining and random high-amplitude spikes added. (b) Decimated central COV volume with added noise spikes after application of the suggested JLRSI process. The S/N is improved.

Our second example deals with a deep offshore marine data set from Angola, of which we use offset classes with NMO correction. The inline and crossline increments are 25 and 12.5 m, respectively. Here, we use three spatial dimensions: the inline number, crossline number, and offset class number. The processing window temporal size was 500 ms, and the spatial window extent was 20 traces in the inline and crossline directions and four traces in the offset class direction. Figures 6 and 7 show a prestack and a stacked inline section of the original data. Figures 8 and 9 show the prestack inline section, first after additional decimation to reach approximately 75% missing traces in the considered volume, then after JLRSI reconstruction starting from the former decimated data. The difference between the reconstructed and original prestack inline sections is shown in Figure 10. The stack corresponding to the same inline section after decimation is shown in Figure 11, in which one can easily note the effects of data decimation and high-amplitude erratic noise. Figure 12

shows the result of JLRSI reconstruction after stacking, and the difference between the reconstructed and original stacks is shown in Figure 13. Interpolated and denoised traces are consistent with the original volume, and diffractions appear to be correctly reconstructed. Less-aggressive noise attenuation can be obtained by lowering the noise threshold, at the expense of an additional computational cost because more singular values would be needed to reconstruct the seismic volume.

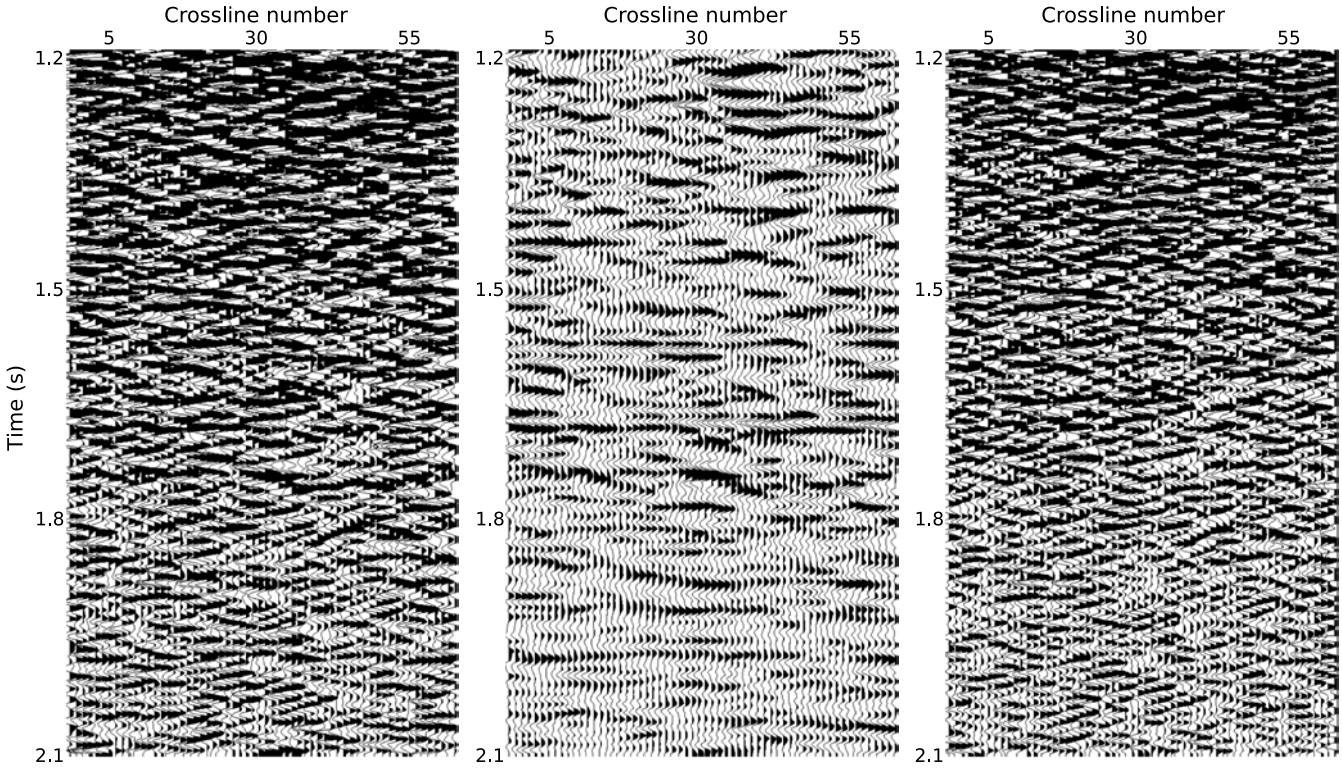

Figure 3. (a) Original undecimated and uncorrupted central COV volume. (b) Decimated and corrupted central COV volume after JLRSI reconstruction. (c) Difference (a and b).

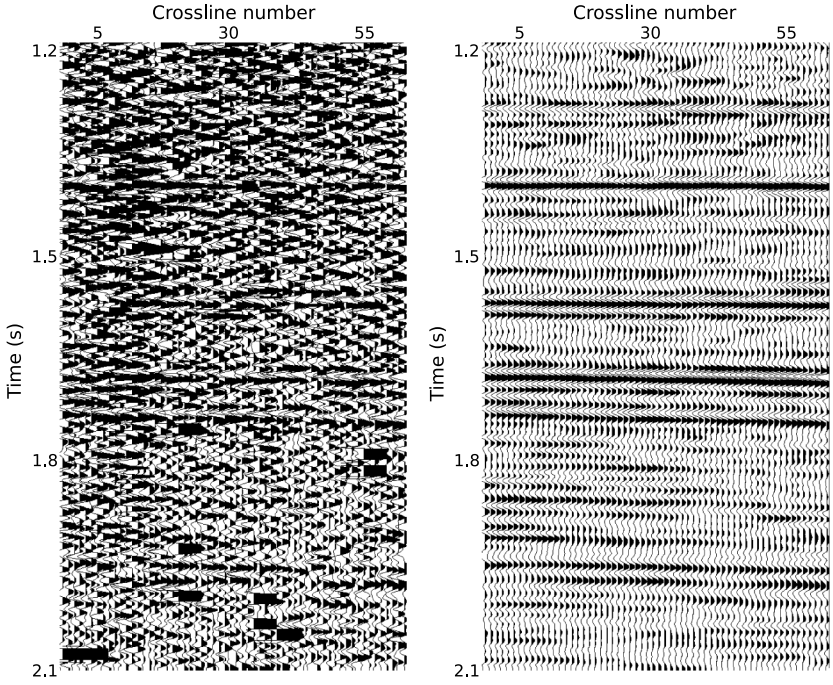

Figure 4. (a) Stack of the 36 COV volumes with 25% prestack traces remaining and randomly added high-amplitude spikes. (b) Stack after simultaneous random and erratic noise attenuation with interpolation of missing traces using JLRSI.

DISCUSSION

As in other algorithms for seismic data interpolation and denoising, there are variable parameters that need to be adjusted to the problem at hand. First of all, the temporal and spatial windowing must be adapted to the bin spacing and offset class interval. Another parameter of importance is the number of kept singular values, which depends on the window size. Although the rank of the signal

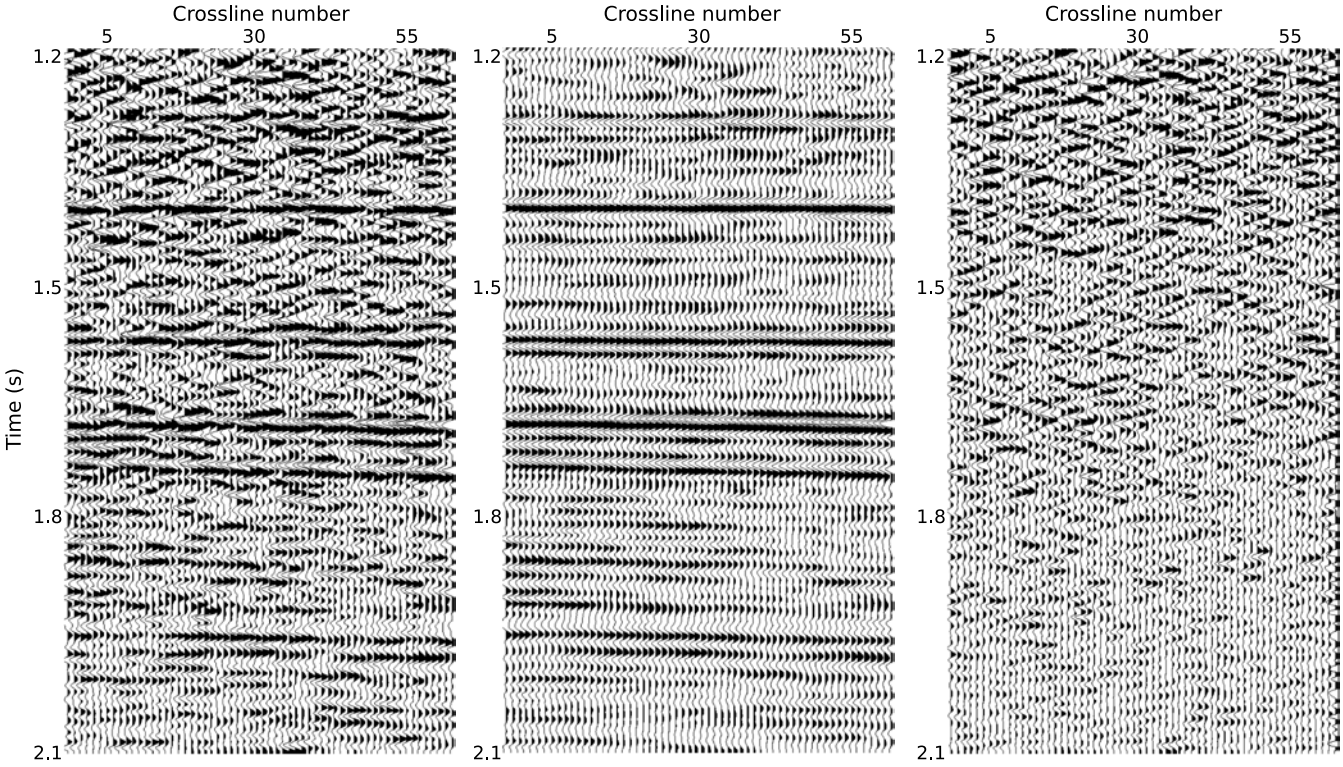

Figure 5. (a) Stack of the original undecimated, uncorrupted COV volumes. (b) Stack of the decimated and corrupted COV volumes after JLRSI reconstruction. (c) Difference (a and b).

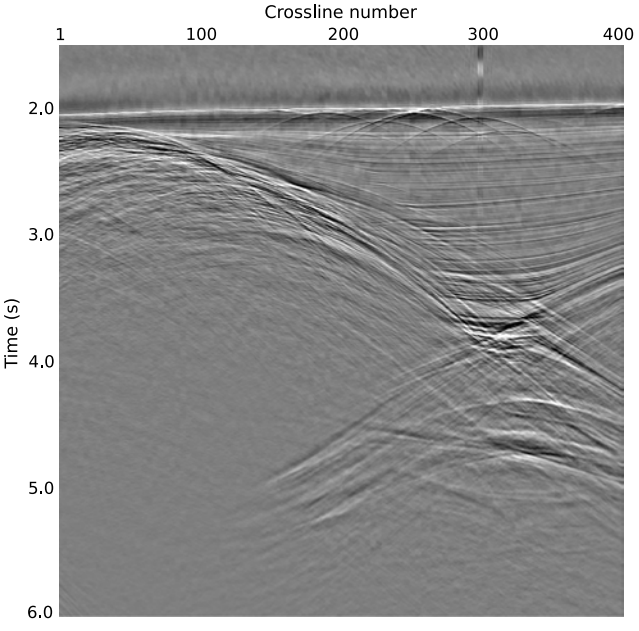

Figure 6. Original inline section of a prestack offset class. Erratic traces are noticeable.

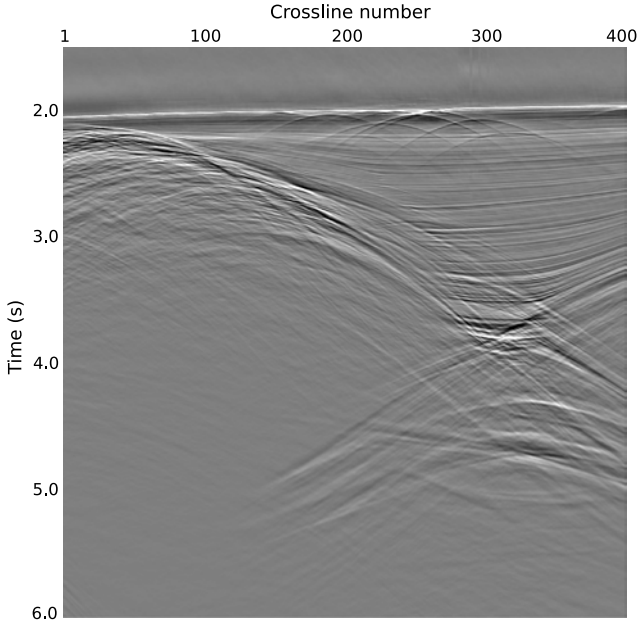

Figure 7. Stacked inline section of undecimated original data.

component is determined adaptively for each frequency via the singular value soft-thresholding operator, the computational cost is proportional to the number of calculated singular values in the randomized SVD scheme. If too many singular values are computed, the computational cost suffers; conversely, if not enough of them are computed, the algorithm might converge toward a local minimum because convexity would be lost and the processing will seem overly aggressive.

Concerning the input data, the best results are obtained with NMO-corrected data (Trad, 2009). For wide-azimuth land data, performing 5D simultaneous interpolation and denoising using JLRSI on COV volumes works well in practice (note that 5D interpolation is done in four spatial dimensions). Instead of both offset vector components, one can use the offset magnitude and azimuth dimensions in addition to the inline and crossline directions. When dealing with highly irregular data sets, the proposed algorithm does not

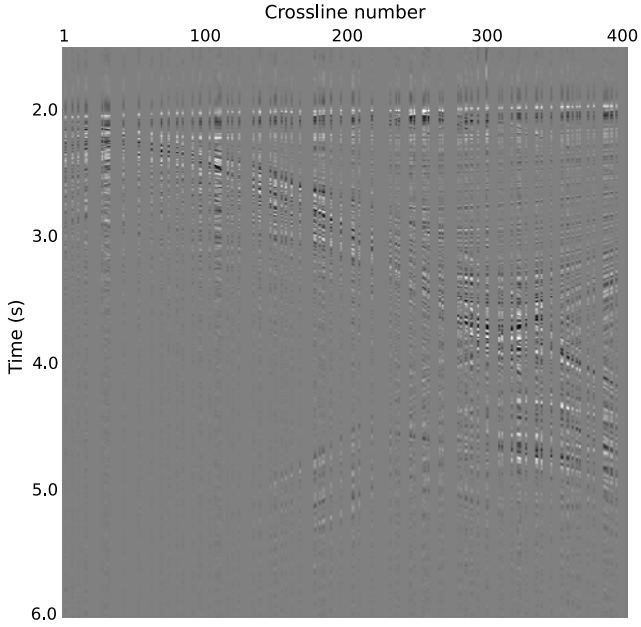

Figure 8. Original prestack inline section after decimation to reach approximately 75% missing traces.

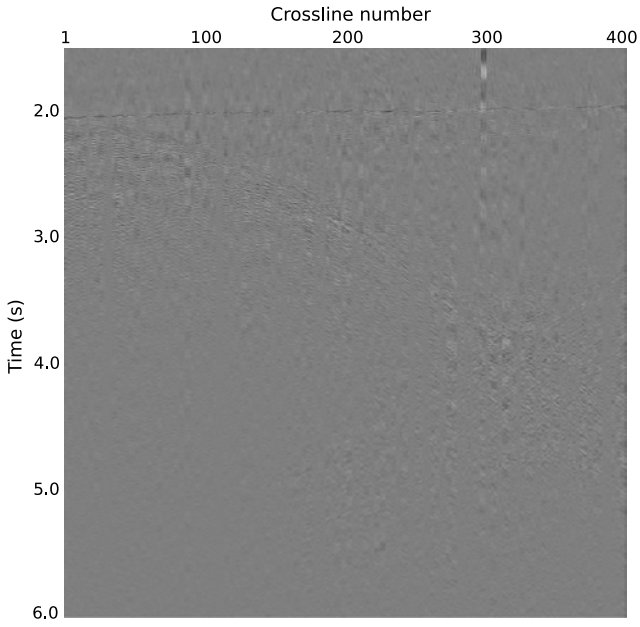

Figure 10. Difference between Figure 9 (after reconstruction, prestack) and Figure 6 (original data, prestack).

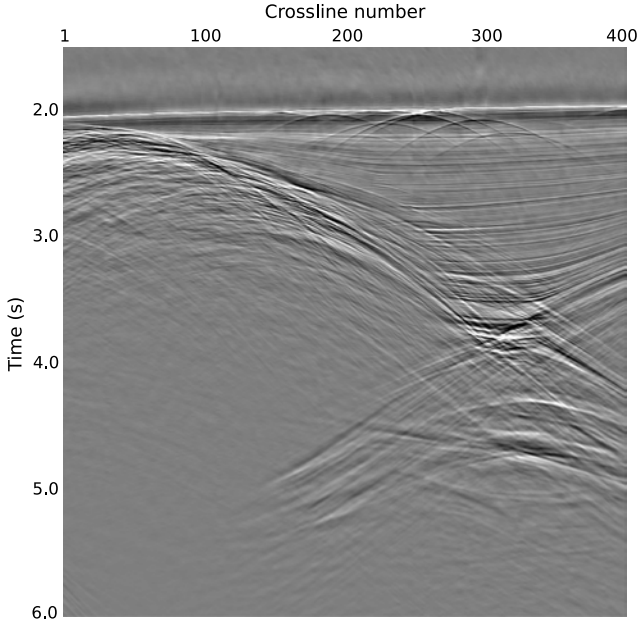

Figure 9. Inline section from Figure 8 after application of the JLRSI reconstruction process.

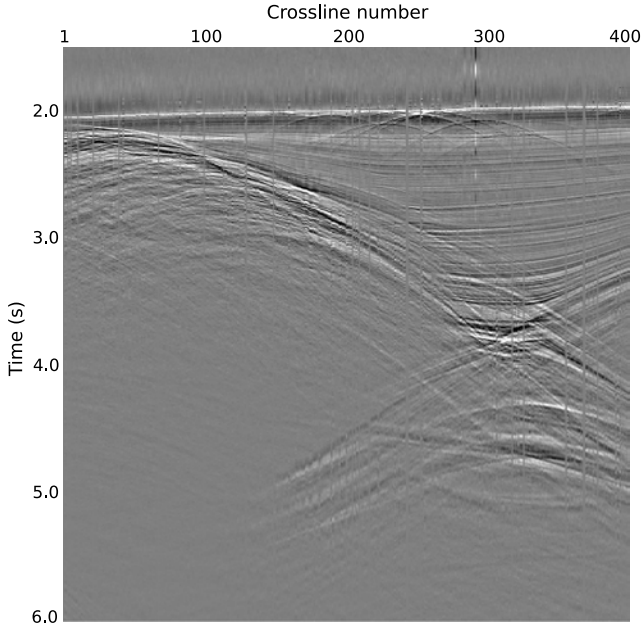

Figure 11. Stacked inline section of original data. Erratic noise and sampling artifacts due to data decimation are noticeable.

have the advantage of methods based on the irregular Fourier transform such as the ALFT (Xu et al., 2005, 2010; Poole, 2010), which honors the exact trace coordinates, at the expense of the additional computational burden of computing the irregular Fourier transform for which the fast Fourier transform (FFT) cannot be directly used. As an alternative, following Trad (2009) in the case of MWNI, we

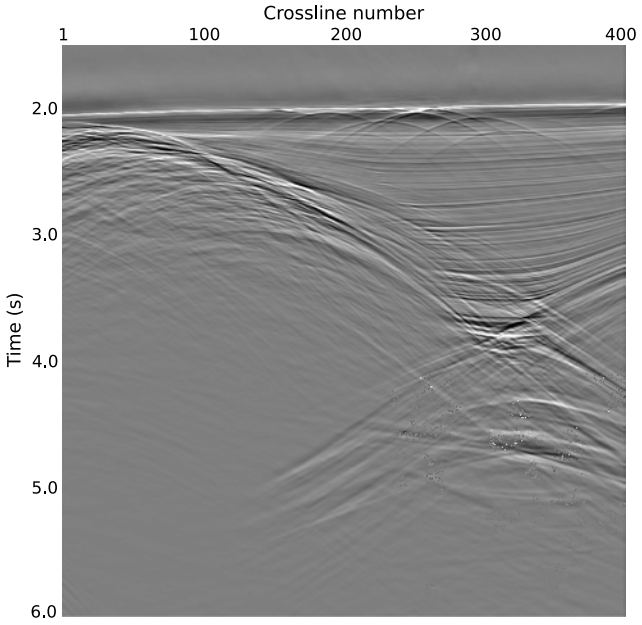

Figure 12. Stacked inline section after application of the JLRSI reconstruction process. The S/N is improved, and the erratic noise and sampling artifacts have been eliminated.

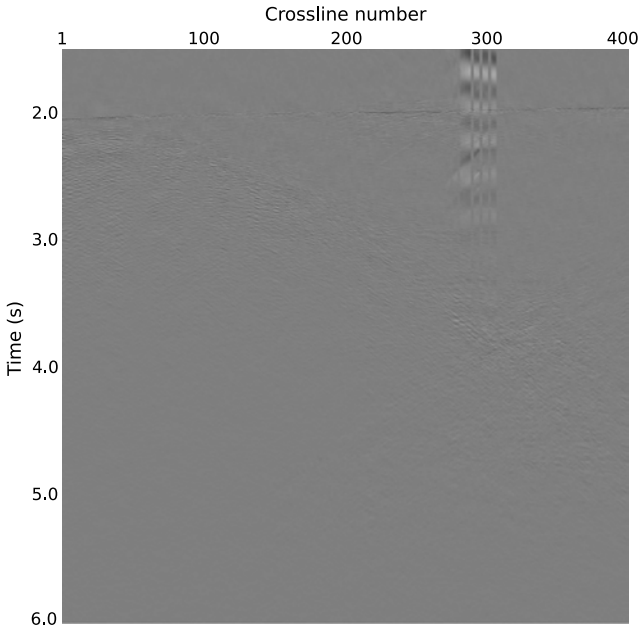

Figure 13. Difference between Figure 12 (after reconstruction, stack) and Figure 7 (original data, stack).

can bin the spatial positions into a finer grid to minimize binning errors and then apply JLRSI on this fine grid. Sparsity of the resulting Toeplitz matrices would increase by a large factor, and this is where compressed sensing takes its whole meaning: Optimal reconstruction is still possible. On the other hand, the proposed method is not currently able to exactly reconstruct strongly aliased data when the original sampling is regular, e.g., for interpolation of regularly missing traces. For this purpose, a frequency extrapolation scheme in the spirit of Spitz (1991) and Naghizadeh and Sacchi (2013) can be used.

A last point is the flexibility of the convex framework: Additional constraints can be added at little to no cost, as long as their corresponding projection or proximal operator can be evaluated without difficulty.

CONCLUSIONS

We have formulated the multidimensional prestack seismic data interpolation with random and erratic noise attenuation problem as a JLRSI convex program. The signal is characterized for each temporal frequency slice in the *f-x* domain as a low-rank component of a Toeplitz or BTTB matrix. We used the signal model pioneered in Cadzow/SSA filtering methods to benefit from the signal predictability property in the *f-x* domain, and a powerful convex optimization framework enables us to formulate a well-posed problem to reconstruct the low-rank signal component from noisy and incomplete data. A joint minimization of the nuclear norm of the signal component and the L_1-norm of the erratic noise component lets us separate out this erratic noise, avoiding the sensitivity to outliers that plagues classical approaches.

The proposed solution to the formulated JLRSI convex program resides in an ADMM algorithm, in which the singular value thresholding step is efficiently carried out thanks to fast Toeplitz/BTTB matrix-vector products and a robust randomized SVD scheme. Furthermore, the suggested procedure evaluates automatically the optimal rank of the recovered signal component for each frequency slice, in addition to converging quickly to a satisfactory solution. The effectiveness of our approach at recovering the signal from incomplete and corrupted data was finally demonstrated on real-data examples.

Finally, we would like to mention that many of the techniques used in this paper are based on recent progress in the field of compressed sensing, providing us with elegant frameworks and theoretical properties on top of state-of-the-art algorithms; we anticipate an increasing use of such techniques in exploration geophysics.

ACKNOWLEDGMENTS

We thank CGG for the permission to publish this work. Special thanks go to D. Trad, D. Carotti, G. Poole, R. Taylor, T. Hertweck, the editors, and the four reviewers for their constructive help and suggestions.

APPENDIX A

RANDOMIZED SINGULAR VALUE DECOMPOSITION ALGORITHM

We saw in the previous sections that an SVD is required to recover the low-rank component of the data. In fact, due to the thresh-

olding operator, only the first few largest singular values are needed. If a fast matrix-vector product is available and if the number of relevant singular values of an $m \times n$ matrix is much less than $\min(m, n)$, then it is much more efficient to compute them via alternative methods other than a full decomposition; e.g., Krylov subspace methods based on the Lanczos iteration (Golub and Kahan, 1965; Golub and Van Loan, 2013) such as in Trickett (2003), Korobeynikov (2010), Gao et al. (2013), or randomized methods on which we will focus, already used in Oropeza and Sacchi (2011), although without the advantages provided by fast matrix-vector products. Such randomized methods use random projections to uncover a lower dimensional subspace capturing most of the action of a matrix and are competitive in speed and accuracy with Krylov methods while being conceptually simple, robust, and easily parallelizable on multiprocessor architectures (Halko et al., 2011). Here, we make use of Algorithm 1, described and analyzed in Halko et al. (2011) and Martinsson et al. (2011).

This algorithm requires $(2q+2)(k+p)\mathcal{C}_{\mathbf{A}} + O(k^2(m+n))$ floating point operations, where $\mathcal{C}_{\mathbf{A}}$ is the number of operations required by the application of $\mathbf{A}$ or $\mathbf{A}^*$. Krylov methods are in the same range, compared to a full SVD factorization that requires $O(mn^2)$ operations (Golub and Van Loan, 2013), recalling that k is taken much smaller than m and n. Back to our problem, the Toeplitz or BTTB structures of the input matrix enables further gains on $\mathcal{C}_{\mathbf{A}}$, which is the subject of the next section.

Algorithm 1

Input: an $m \times n$ matrix $\mathbf{A}$ of which we wish to compute the rank-k SVD approximation.

1) Generate an $n \times (k+p)$ test matrix $\mathbf{\Phi}$ with randomly sampled Gaussian entries, where p is a small oversampling factor of the order of k.

2) Compute $\mathbf{Y} = (\mathbf{A}\mathbf{A}^*)^q\mathbf{A}\mathbf{\Phi}$, with q a small integer (usually 1 to 3) and potential orthonormalization steps between applications of $\mathbf{A}$ and $\mathbf{A}^*$ to alleviate round-off errors.

3) Compute a matrix $\mathbf{Q}$ whose columns are an orthonormal basis of the range of $\mathbf{Y}$, for instance, using a QR decomposition.

4) Compute $\mathbf{B} = \mathbf{A}^*\mathbf{Q}$ and its SVD $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\tilde{\mathbf{V}}^*$.

5) Compute $\mathbf{V} = \mathbf{Q}\tilde{\mathbf{V}}$; then $\mathbf{A} \approx \mathbf{V}\mathbf{\Sigma}\mathbf{U}$ is the sought approximate decomposition.

APPENDIX B

FAST TOEPLITZ MATRIX-VECTOR PRODUCT

In this appendix, we will deal with the following Toeplitz matrix:

$$\mathbf{T} = \begin{pmatrix} a_3 & a_2 & a_1 \\ a_4 & a_3 & a_2 \\ a_5 & a_4 & a_3 \end{pmatrix}. \tag{B-1}$$

The Hermitian transpose of a Toeplitz matrix is also a Toeplitz matrix; thus, the following development equally applies to $\mathbf{T}^*$. Given a Toeplitz matrix $\mathbf{T}$, we can embed it into a circulant matrix $\mathbf{C}$ as follows:

$$\mathbf{C} = \begin{pmatrix} a_3 & a_2 & a_1 & 0 & 0 & 0 & a_5 & a_4 \\ a_4 & a_3 & a_2 & a_1 & 0 & 0 & 0 & a_5 \\ a_5 & a_4 & a_3 & a_2 & a_1 & 0 & 0 & 0 \\ 0 & a_5 & a_4 & a_3 & a_2 & a_1 & 0 & 0 \\ 0 & 0 & a_5 & a_4 & a_3 & a_2 & a_1 & 0 \\ 0 & 0 & 0 & a_5 & a_4 & a_3 & a_2 & a_1 \\ a_1 & 0 & 0 & 0 & a_5 & a_4 & a_3 & a_2 \\ a_2 & a_1 & 0 & 0 & 0 & a_5 & a_4 & a_3 \end{pmatrix}. \tag{B-2}$$

Now, if one takes $\tilde{\mathbf{x}} = (x_1, x_2, x_3, 0, 0, 0, 0, 0)$, then the vector made from the first three components of $\mathbf{C}\tilde{\mathbf{x}}$ is equal to $\mathbf{T}\mathbf{x}$ where $\mathbf{x} = (x_1, x_2, x_3)$. We can finally recognize that the application of the matrix $\mathbf{C}$ is a circular convolution by the vector $(a_3, a_4, a_5, 0, 0, 0, a_1, a_2)^T$; thus, a Toeplitz matrix-vector product can be done in $O(n \log n)$ floating point operations thanks to the FFT (Strang, 1986; Van Loan, 1992; Korobeynikov, 2010) with n being the number of distinct entries in the Toeplitz matrix, without having to explicitly form (and thus store in memory) any of the matrices $\mathbf{C}$ or $\mathbf{T}$. Similarly, diagonal averaging can be performed in $O(n \log n)$ floating point operations via the FFT as described in Korobeynikov (2010): Indeed, for each singular component, the summation over the diagonals is an acyclical convolution of the left and right singular vectors.

This procedure can directly be generalized to BTTB matrices, making use of higher dimensional convolutions and FFTs. The reader is referred to Lee (1986) and Gao et al. (2013) for additional details concerning fast BTTB matrix-vector products.

REFERENCES

Abma, R., and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, no. 6, 91–97, doi: 10.1190/1.2356088.

Andersson, F., M. Carlsson, J.-Y. Tourneret, and H. Wendt, 2014, A new frequency estimation method for equally and unequally spaced data: IEEE Transactions on Signal Processing, **62**, 5761–5774, doi: 10.1109/TSP.2014.2358961.

Aravkin, A., R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, 2014, Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation: SIAM Journal on Scientific Computing, **36**, 237–266, doi: 10.1137/130919210.

Aybat, N. S., D. Goldfarb, and S. Ma, 2014, Efficient algorithms for robust and stable principal component pursuit problems: Computational Optimization and Applications, **58**, 1–29, doi: 10.1007/s10589-013-9613-0.

Beck, A., and M. Teboulle, 2009, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: SIAM Journal on Imaging Sciences, **2**, 183–202, doi: 10.1137/080716542.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2011, Distributed optimization and statistical learning via the alternating direction method of multipliers: Foundations and Trends in Machine Learning, **3**, 1–122, doi: 10.1561/2200000016.

Boyd, S., and L. Vandenberghe, 2004, Convex optimization: Cambridge University Press.

Broomhead, D., and G. P. King, 1986, Extracting qualitative dynamics from experimental data: Physica D, **20**, 217–236, doi: 10.1016/0167-2789(86)90031-X.

Cadzow, J. A., 1988, Signal enhancement — A composite property mapping algorithm: IEEE Transactions on Acoustics Speech and Signal Processing, **36**, 49–62, doi: 10.1109/29.1488.

Canales, L. L., 1984, Random noise reduction: 54th Annual International Meeting, SEG, Expanded Abstracts, 525–527.

Candès, E. J., X. Li, Y. Ma, and J. Wright, 2009, Robust principal component analysis?: Journal of ACM, **58**, 1–37, doi: 10.1145/1970392.1970395.

Candès, E. J., and Y. Plan, 2009, Matrix completion with noise: Proceedings of the IEEE, **98**, 925–936, doi: 10.1109/JPROC.2009.2035722.

Candès, E. J., and B. Recht, 2008, Exact matrix completion via convex optimization: Foundations Of Computational Mathematics, **9**, 717–772, doi: 10.1007/s10208-009-9045-5.

Candès, E. J., J. Romberg, and T. Tao, 2006a, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on Information Theory, **52**, 489–509, doi: 10.1109/TIT.2005.862083.

Candès, E. J., J. K. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, 1207–1223, doi: 10.1002/cpa.20124.

Candès, E. J., and T. Tao, 2009, The power of convex relaxation: Near-optimal matrix completion: IEEE Transactions on Information Theory, **56**, 2053–2080, doi: 10.1109/TIT.2010.2044061.

Cary, P. W., 1999, Common-offset-vector gathers: An alternative to cross-spreads for wide-azimuth 3D surveys: 69th Annual International Meeting, SEG, Expanded Abstracts, 1496–1499.

Chandrasekaran, V., B. Recht, P. A. Parrilo, and A. S. Willsky, 2012, The convex geometry of linear inverse problems: Foundations of Computational Mathematics, **12**, 805–849, doi: 10.1007/s10208-012-9135-7.

Chandrasekaran, V., S. Sanghavi, P. Parrilo, and A. Willsky, 2011, Rank-sparsity incoherence for matrix decomposition: SIAM Journal on Optimization, **21**, 572–596, doi: 10.1137/090761793.

Chen, C., B. He, Y. Ye, and X. Yuan, 2014, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent: Mathematical Programming Series A, doi: 10.1007/s10107-014-0826-5.

Chen, K., and M. D. Sacchi, 2015, Robust reduced-rank filtering for erratic seismic noise attenuation: Geophysics, **80**, no. 1, V1–V11, doi: 10.1190/geo2014-0116.1.

Claerbout, J. F., and F. Muir, 1973, Robust modeling with erratic data: Geophysics, **38**, 826–844, doi: 10.1190/1.1440378.

Combettes, P., and J.-C. Pesquet, 2007, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery: IEEE Journal of Selected Topics in Signal Processing, **1**, 564–574, doi: 10.1109/JSTSP.2007.910264.

Combettes, P., and J.-C. Pesquet, 2011, Proximal splitting methods in signal processing, *in* H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, eds., Fixed-point algorithms for inverse problems in science and engineering: Springer, 185–212.

Condat, L., and A. Hirabayashi, 2015, Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements: Sampling Theory in Signal and Image Processing, **14**, 17–47.

Da Silva, C., and F. Herrmann, 2014, Low-rank promoting transformations and tensor interpolation — Applications to seismic data denoising: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, Th E104 03.

Eckstein, J., and D. P. Bertsekas, 1992, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators: Mathematical Programming, **55**, 293–318, doi: 10.1007/BF01581204.

Fazel, M., H. Hindi, and S. Boyd, 2001, A rank minimization heuristic with applications to minimum order system approximation: Proceedings of the American Control Conference: IEEE 6, 4734–4739.

Fazel, M., T. Pong, D. Sun, and P. Tseng, 2013, Hankel matrix rank minimization with applications to system identification and realization: SIAM Journal on Matrix Analysis and Applications, **34**, 946–977, doi: 10.1137/110853996.

Gabay, D., and B. Mercier, 1976, A dual algorithm for the solution of nonlinear variational problems via finite element approximation: Computers and Mathematics with Applications, **2**, 17–40, doi: 10.1016/0898-1221(76)90003-1.

Ganesh, A., Z. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma, 2009, Fast algorithms for recovering a corrupted low-rank matrix: International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP): IEEE, 213–216.

Gao, J., M. D. Sacchi, and X. Chen, 2013, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: Geophysics, **78**, no. 1, 21–30, doi: 10.1190/geo2012-0038.1.

Ghil, M., M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, 2002, Advanced spectral methods for climatic time series: Reviews of Geophysics, **40**, 1–41, doi: 10.1029/2000RG000092.

Glowinski, R., and A. Marroco, 1975, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires: Revue française d'automatique, informatique, recherche opérationnelle, **9**, 41–76.

Golub, G. H., and W. Kahan, 1965, Calculating the singular values and pseudo-inverse of a matrix: SIAM Journal on Numerical Analysis Series B, **2**, 205–224, doi: 10.1137/0702016.

Golub, G. H., and C. F. Van Loan, 2013, Matrix computations 4th ed.: Johns Hopkins University Press.

Golyandina, N., and A. Zhigljavsky, 2013, Singular spectrum analysis for time series: Springer.

Halko, N., P. Martinsson, and J. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: SIAM Review, **53**, 217–288, doi: 10.1137/090771806.

Hennenfent, G., L. Fenelon, and F. J. Herrmann, 2010, Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach: Geophysics, **75**, no. 6, 203–210, doi: 10.1190/1.3494032.

Herrmann, F., M. Friedlander, and O. Yilmaz, 2013, Fighting the curse of dimensionality: Compressed sensing in exploration seismology: IEEE Signal Processing Magazine, **29**, 88–100, doi: 10.1109/MSP.2012.2185859.

Herrmann, F. J., 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, no. 6, 173–187, doi: 10.1190/1.3506147.

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysics Journal International, **173**, 233–248, doi: 10.1111/j.1365-246X.2007.03698.x.

Komodakis, N., and J.-C. Pesquet, 2014, Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems: IEEE Signal Processing Magazine, arXiv manuscript, arXiv:1406.5429.

Korobeynikov, A., 2010, Computation- and space-efficient implementation of SSA: Statistics and its Interface, **3**, 357–368, doi: 10.4310/SII.2010.v3.n3.a9.

Kreimer, N., and M. D. Sacchi, 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, no. 3, 113–122, doi: 10.1190/geo2011-0399.1.

Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, no. 6, 273–284, doi: 10.1190/geo2013-0022.1.

Lee, D., 1986, Fast multiplication of a recursive block Toeplitz matrix by a vector and its application: Journal of Complexity, **2**, 295–305, doi: 10.1016/0885-064X(86)90007-5.

Liu, B., and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: Geophysics, **69**, 1560–1568, doi: 10.1190/1.1836829.

Martinsson, P., V. Rokhlin, and M. Tygert, 2011, A randomized algorithm for the decomposition of matrices: Applied and Computational Harmonic Analysis, **30**, 47–68, doi: 10.1016/j.acha.2010.02.003.

Moreau, J., 1965, Proximité et dualité dans un espace hilbertien: Bulletin de la S.M.F., **93**, 273–299.

Naghizadeh, M., and M. Sacchi, 2013, Multidimensional de-aliased Cadzow reconstruction of seismic records: Geophysics, **78**, no. 1, A1–A5, doi: 10.1190/geo2012-0200.1.

Nesterov, Y., 2005, Smooth minimization of non-smooth functions: Mathematical Programming, **103**, 127–152, doi: 10.1007/s10107-004-0552-5.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization 2nd ed.: Springer.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, no. 3, 25–32, doi: 10.1190/1.3552706.

Parikh, N., and S. Boyd, 2013, Proximal algorithms, in foundations and trends in optimization: NOW Publishers.

Pham, M. Q., C. Chaux, L. Duval, and J.-C. Pesquet, 2014, A primal-dual proximal algorithm for sparse template-based adaptive filtering: Application to seismic multiple removal: IEEE Transactions on Signal Processing, **62**, 4256–4269, doi: 10.1109/TSP.2014.2331614.

Plaut, G., and R. Vautard, 1993, Spells of low-frequency oscillations and weather regimes in the northern hemisphere: Journal of the Atmospheric Sciences, **51**, 210–236, doi: 10.1175/1520-0469(1994)051<0210:SOLFOA>2.0.CO;2.

Poole, G., 2010, 5D data reconstruction using the anti-leakage Fourier transform: 72th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Recht, B., M. Fazel, and P. A. Parrilo, 2010, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization: SIAM Review, **52**, 471–501, doi: 10.1137/070697835.

Sacchi, M. D., 2009, FX singular spectrum analysis: 2009 Convention, CSPG CSEG CWLS, Extended Abstracts, 392–395.

Soubaras, R., 1995, Prestack random and impulsive noise attenuation by *f-x* projection filtering: 65th Annual International Meeting, SEG, Expanded Abstracts, 711–714.

Spitz, S., 1991, Seismic trace interpolation in the *f-x* domain: Geophysics, **56**, 785–794, doi: 10.1190/1.1443096.

Strang, G., 1986, A proposal for Toeplitz matrix calculations: Studies in Applied Mathematics, **74**, 171–176.

Sun, D., K. C. Toh, and L. Yang, 2014, A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type of constraints: SIAM Journal on Optimization, **25**, 882–915, doi: 10.1137/140964357.

Tao, M., and X. Yuan, 2011, Recovering low-rank and sparse components of matrices from incomplete and noisy observations: SIAM Journal on Optimization, **21**, 57–81, doi: 10.1137/100781894.

Taylor, H. L., S. C. Banks, and J. F. McCoy, 1979, Deconvolution with the l1 norm: Geophysics, **44**, 39–52, doi: 10.1190/1.1440921.

Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: Geophysics, **74**, no. 6, 123–132, doi: 10.1190/1.3245216.

Trad, D., T. J. Ulrych, and M. D. Sacchi, 2002, Accurate interpolation with high-resolution time-variant Radon transforms: Geophysics, **67**, 644–656, doi: 10.1190/1.1468626.
Trad, D., T. J. Ulrych, and M. D. Sacchi, 2003, Latest views of the sparse Radon transform: Geophysics, **68**, 386–399, doi: 10.1190/1.1543224.
Trickett, S., 2003, F-xy eigenimage noise suppression: Geophysics, **68**, 751–759, doi: 10.1190/1.1567245.
Trickett, S., 2008, F-xy Cadzow noise suppression: 78th Annual International Meeting, SEG, Expanded Abstracts, 2586–2590.
Trickett, S., L. Burroughs, and A. Milton, 2012, Robust rank-reduction filtering for erratic noise: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0129.1.
Trickett, S., L. Burroughs, and A. Milton, 2013, Interpolation using Hankel tensor completion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 3634–3638.
Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: 80th Annual International Meeting, SEG, Expanded Abstracts, 3829–3833.
Vandenberghe, L., and S. Boyd, 1996, Semidefinite programming: SIAM Review, **38**, 49–95, doi: 10.1137/1038003.
Van Loan, C. F., 1992, Computational frameworks for the fast Fourier transform: SIAM.
Vautard, R., and M. Ghil, 1989, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series: Physica D, **35**, 395–424, doi: 10.1016/0167-2789(89)90077-8.
Vermeer, G. J., 2000, Processing with offset-vector-slot gathers: 70th Annual International Meeting, SEG, Expanded Abstracts, 5–8.
Wen, Z., D. Goldfarb, and W. Yin, 2010, Alternating direction augmented Lagrangian methods for semidefinite programming: Mathematical Programming Computation, **2**, 203–230, doi: 10.1007/s12532-010-0017-1.
Wright, J., A. Ganesh, K. Min, and Y. Ma, 2013, Compressive principal component pursuit: Information and Inference, **2**, 32–68, doi: 10.1109/ISIT.2012.6283062.
Wright, J., A. Ganesh, S. Rao, Y. Peng, and Y. Ma, 2009, Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization, *in* Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, eds., Advances in Neural Information Processing Systems: NIPS Foundation 22, 2080–2088.
Xu, S., Y. Zhang, and G. Lambaré, 2010, Antileakage Fourier transform for seismic data regularization in higher dimensions: Geophysics, **75**, no. 6, 113–120, doi: 10.1190/1.3507248.
Xu, S., Y. Zhang, D. Pham, and G. Lambaré, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, no. 4, 87–95, doi: 10.1190/1.1993713.
Zhou, Z., J. Wright, X. Li, E. J. Candès, and Y. Ma, 2010, Stable principal component pursuit: International Symposium on Information Theory (ISIT): IEEE, 1518–1522.