HAL
archives-ouvertes.fr

A Primal-Dual Proximal Algorithm for Sparse Template-Based Adaptive Filtering: Application to Seismic Multiple Removal

Mai Quyen Pham, Caroline Chaux, Laurent Duval, Jean-Christophe Pesquet

To cite this version:

Mai Quyen Pham, Caroline Chaux, Laurent Duval, Jean-Christophe Pesquet. A Primal-Dual Proximal Algorithm for Sparse Template-Based Adaptive Filtering: Application to Seismic Multiple Removal. IEEE Transactions on Signal Processing, Institute of Electrical and Electronics Engineers, 2014, 62 (16), pp.4256-4269. <10.1109/TSP.2014.2331614>. <hal-00914628>

HAL Id: hal-00914628

https://hal.archives-ouvertes.fr/hal-00914628

Submitted on 5 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Primal-Dual Proximal Algorithm for Sparse Template-Based Adaptive Filtering: Application to Seismic Multiple Removal

Mai Quyen Pham *, Caroline Chaux †, Laurent Duval ‡, and Jean-Christophe Pesquet §,

Abstract

Unveiling meaningful geophysical information from seismic data requires to deal with both random and structured "noises". As their amplitude may be greater than signals of interest (primaries), additional prior information is especially important in performing efficient signal separation. We address here the problem of multiple reflections, caused by wave-field bouncing between layers. Since only approximate models of these phenomena are available, we propose a flexible framework for time-varying adaptive filtering of seismic signals, using sparse representations, based on inaccurate templates. We recast the joint estimation of adaptive filters and primaries in a new convex variational formulation. This approach allows us to incorporate plausible knowledge about noise statistics, data sparsity and slow filter variation in parsimony-promoting wavelet frames. The designed primal-dual algorithm solves a constrained minimization problem that alleviates standard regularization issues in finding hyperparameters. The approach demonstrates significantly good performance in low signal-to-noise ratio conditions, both for simulated and real field data.

1 Introduction

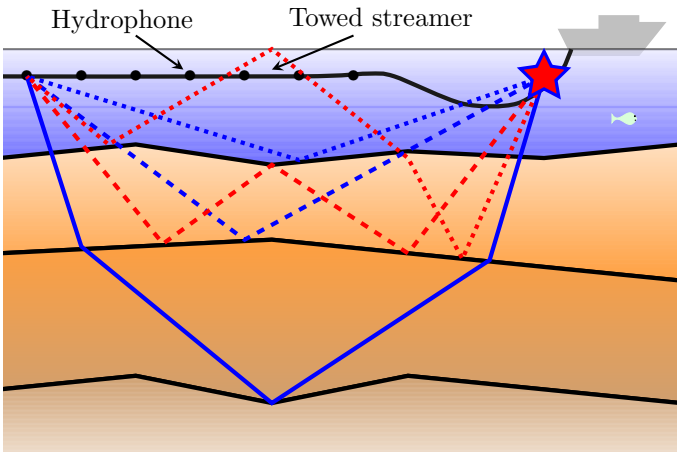


Figure 1: Principles of marine seismic data acquisition and wave propagation. Towed streamer with hydrophones. Reflections on different layers (primaries in blue), and reverberated disturbances (multiple in dotted and dashed red).

Adaptive filtering techniques play a prominent part in signal processing. They cope with time-varying or non-stationary signals and systems. The rationale of these methods is to optimize parameters of variable filters, according to adapted cost functions working on error signals. The appropriate choice of cost functions, that encode a priori information on the system under study, should be balanced with the

*M. Q. Pham is with IFP Energies nouvelles and Université Paris-Est, LIGM UMR-CNRS 8049.
†C. Chaux is with Aix-Marseille Université, LATP UMR CNRS 7353.
‡L. Duval is with IFP Energies nouvelles.
§J.-C. Pesquet is with Université Paris-Est, LIGM UMR CNRS 8049.

tractability of the adaptation. While traditional adaptive algorithms resort to least squares minimization, they may be sensitive to outliers, and may not directly promote simple filters (well-behaved, with concentrated coefficients), especially when the filter length is not well known.

Certain systems, for instance transmission channels, behave parsimoniously. They are modeled by sparse impulse response filters with a few large taps, most of the others being small. Several designs have thus turned toward cost functions promoting filter sparsity [1–3]. Recently, developments around proximity operators [4] with signal processing applications [5] have allowed performance improvements. For instance, [6,7] allow sparsity promotion with ℓ_1 and ℓ_p, $0 < p < 1$, quasi-norms, respectively, via time-varying soft-thresholding operators. Improvements reside in convergence speed acceleration or gains in signal-to-noise ratios (SNRs). These developments are generally performed directly in the signal domain.

Sparsity may additionally be present in signals. Choosing an appropriate transformed domain could, when applied appropriately [8], ease the efficiency of adaptive filters [9–11]. Such transforms include filter banks [12] or redundant wavelets [13]. The usefulness of sparsity-promoting loss functions or shrinkage functions in structured data denoising or deconvolution is well documented [14–16]. Geophysical signal processing [17] is a field where dealing with sparsity, or at least energy concentration, both in the system filter and the data domain, is especially beneficial.

The aim of seismic data analysis is to infer the subsurface structure from seismic wave fields recorded through land or marine acquisitions. In reflection seismology, seismic waves, generated by a close-to-impulsive source, propagate through the subsurface medium. They travel downwards, then upwards, reflected by geological interfaces, convolved by earth filters. They account for the unknown relative distances and velocity contrasts between layers and they are affected by propagation-related distortions. A portion of the wave fields is finally recorded near the surface by arrays of seismometers (geophones or hydrophones). In marine acquisition, hydrophones are towed by kilometer-long streamers.

Signals of interest, named primaries, follow wave paths depicted in dotted, dashed and solid blue in Fig. 1. Although the contributions are generally considered linear, several types of disturbances, structured or more stochastic, affect the relevant information present in seismic data. Since the data recovery problem is under-determined, geophysicists have developed pioneering sparsity-promoting techniques. For instance, robust, ℓ_1-promoted deconvolution or complex wavelet transforms still pervade many areas of signal processing.

We address one of the most severe types of interferences: secondary reflections, named multiples, corresponding to seismic waves bouncing between layers [18], as illustrated with red dotted and dashed lines in Fig. 1. These reverberations share waveform and frequency contents similar to primaries, with longer propagation times. From the standpoint of geological information interpretation, they often hide deeper target reflectors. For instance, the dashed-red multiple path may possess a total travel time comparable with that of the solid-blue primary. Their separation is thus required for accurate subsurface characterization. A geophysics industry standard consists of model-based multiple filtering. One or several realistic templates of a potential multiple are determined off-line, based on primary reflections identified in above layers. For instance, the dashed-red path may be approximately inferred from the dashed-blue, and then adaptively filtered for separation from the solid-blue propagation. Their precise estimation is beyond the scope of this work, we suppose them given by prior seismic processing or modeling. As template modeling is partly inaccurate — in delay, amplitude and frequency — templates should be adapted in a time-varying fashion before being subtracted from the recorded data. Resorting to several templates and weighting them adaptively, depending on the time and space location of seismic traces, helps when highly complicated propagation paths occur. Increasing the number of templates is a growing trend in exploration. Meanwhile, inaccuracies in template modeling, complexity of time-varying adaptation combined with additional stochastic disturbances require additional constraints to obtain geophysically-sound solutions.

We propose a methodology for primary/multiple adaptive separation based on approximate templates. This framework addresses at the same time structured reverberations and a more stochastic part. Namely, let $n \in \{0, \ldots, N-1\}$ denotes the time index for the observed seismic trace z, acquired by a given sensor. We assume, as customary in seismic, an additive model of contributions:

$$z^{(n)} = \overline{y}^{(n)} + \overline{s}^{(n)} + b^{(n)} \,. \tag{1}$$

The unknown signal of interest (primary, in blue) and the sum of undesired, secondary reflected signals (different multiples, in red) are denoted, respectively, by $\overline{y} = (\overline{y}^{(n)})_{0 \leq n < N}$ and $\overline{s} = (\overline{s}^{(n)})_{0 \leq n < N}$. Other unstructured contributions are gathered in the noise term $b = (b^{(n)})_{0 \leq n < N}$. We assume that several approximate templates accounting for multiples are available. As the above problem is undetermined, additional constraints should be devised. We specify sparsity and slow-variation requirements on primaries

and adaptive filters. In Section 2, we analyze related works and specify the novelty of the proposed methodology. Section 3 describes the transformed linear model incorporating the templates with adaptive filtering. In Section 4, we formulate a generic variational form for the problem. Section 5 describes the primal-dual proximal formulation. The performance of the proposed method is assessed in Section 6. We detail the chosen optimization criteria and provide a comparison with different types of frames. The methodology is first evaluated on a realistic synthetic data model, and finally tested and applied to an actual seismic data-set. Conclusions and perspectives are drawn in Section 7. This work improves upon [19] by taking into account several multiple templates. Part of it was briefly presented in [20], by incorporating an additional noise into the generic model, and by introducing alternative norms in multiple selection objective criteria. Here, the approach is extended. In particular, the problem is completely reformulated as a constrained minimization problem, in order to simplify the determination of data-based parameters, as compared with our previous regularized approach involving hyper-parameters.

2 Related and proposed work

Primary/multiple separation is a long standing problem in seismic. Published solutions are weakly generic, and often embedded in a more general processing work-flow. Levels of prior knowledge — from the shape of the seismic source to partial geological information — greatly differ depending on data-sets. We refer to [21, 22] for recent accounts on broad processing issues, including shortcomings of standard ℓ_2-based methods. The latter are computationally efficient, yet their performance decreases when traditional assumptions fail (primary/multiple decorrelation, weak linearity or stationarity, high noise levels). We focus here on recent sparsity-related approaches, pertaining to geophysical signal processing. The potentially parsimonious layering of the subsurface (illustrated in Fig. 1) suggests a modeling of primary reflection coefficients with generalized Gaussian or Cauchy distributions [23], having suitable parameters. The sparsity induced on seismic data has influenced deconvolution and multiple subtraction. Progressively, the non-Gaussianity of seismic traces has been emphasized, and contributed to the use of more robust norms [24, 25] for blind separation with independent component analysis (ICA) for the signal of interest. As the true nature of seismic data distribution is still debated, including its stationarity [26], a handful of works have investigated processing in appropriate transformed domains. They may either stationarize [27] or strengthen data sparsity. For instance, [28] applies ICA in a dip-separated domain. In [29], as well as in [30] and subsequent works by the same group, a special focus is laid on separation in the curvelet domain.

Aforementioned works mostly deal with the mitigation of some ℓ_2-norm on residuals, as remnant noise is traditionally considered Gaussian in seismic. They are blended with ℓ_0 objectives, solved through ℓ_1 or hybrid ℓ_1-ℓ_2 approximations [31], resorting for instance to iteratively re-weighted least-squares method. Recently, [32] investigated the use of intermediate ℓ_p-norms, with $p = 1.2$ for instance, accounting for the "super-Gaussian nature of the seismic data due to the interfering fields", in the time domain. Without further insights on precise modeling, a more flexible framework is desirable to adapt to the nature of different seismic data, either in the direct or in a variety of transformed domains.

Data sparsity and noise Gaussianity alone may not be sufficient to solve (1). Additional constraints reduce the set of solutions, hopefully to geologically sounder ones. A first one is the locality of matched filters, traditional in standard multiple filtering. These can be modeled by Finite Impulse Response (FIR) operators. Classical filter support limitations, down to one-tap [22, 29], assorted with ℓ_2 or ℓ_1 criteria, are standard. In other seismic processing fields, [33] has investigated mixed ℓ_p-ℓ_1 loss functions for deconvolution. Recently, in [34, 35] the use of the nuclear norm is promoted for interpolation, combined with a standard ℓ_2-norm penalty. Yet, to the authors' knowledge, no work in multiple removal has endeavored a more systematic study of variational and sparsity constraints on the adaptive filters, in the line of [36]. In this work, we propose a formulation allowing a family of penalties to be applied to the adaptive FIR filters. Since no metric is evidently more natural, such a flexibility is useful to assess different objectives. For instance, one might be interested in either well-preserving primaries, in mild noise cases, or robustly removing the multiples, in high contamination situations. Indeed, when the perturbation is stronger in amplitude than the target signal, geophysicists are interested in uncovering even spoors of potential primaries, obfuscated by noise. As will be seen, the most appropriate norm depends on such contexts. Finally, the propagation medium, as well as the modeled templates, carry continuous variations. With the seismic bandwidth (up to 125 Hz), changes in signals are not as dramatic as in sharp images. Consequently, we expect the adapted filters to exhibit bounded variations from one time index to the next one.

This paper presents for the first time a relatively generic framework for multiple reflection filtering with (i) a noise prior, (ii) sparsity constraints on signal frame coefficients, (iii) slow variation modelling of the adaptive filters, and (iv) concentration metrics on the filters. With the development of recent optimization tools, multiple constraints can now be handled in a convenient manner. Due to the diversity of focus points, paired with data observation, we choose here to decouple effects and to insist on (iv), with respect to different flavors of 1D wavelet bases and frames [37, 38], which appear as natural atoms for sparse descriptions of some physical processes, related to propagation and reflection of signals through media.

The evaluation of the proposed multiple filtering algorithm on seismic data is not straightforward, for two main reasons. First, seismic processing work-flows are neither publicly available for benchmarks and are generally heavily parametrized. Second, quality measures are not easy to devise since visual inspection is of paramount importance in geophysical data processing assessment. We thus compare the proposed approach with a state-of-the-art solution, previously benchmarked against industrial competitors [22].

3 Model description

We assume that multiple templates are modeled at the temporal vicinity of actual disturbances, with standard geophysical assumptions on primaries. The multiple signal possesses a local behavior related to the geological context. Hence, we assume the availability of J templates $(r_j^{(n)})_{0\leq n<N, 0\leq j<J}$, related to $(\overline{s}^{(n)})_{0\leq n<N}$ via a possibly non-causal linear model through a limited support relationship:

$$\overline{s}^{(n)} = \sum_{j=0}^{J-1} \sum_{p=p'}^{p'+P_j-1} \overline{h}_j^{(n)}(p) r_j^{(n-p)} \tag{2}$$

where $\overline{h}_j^{(n)}$ is an unknown finite impulse response (with P_j tap coefficients) associated with template j and time n, and where $p' \in \{-P_j+1, \ldots, 0\}$ is its starting index ($p'=0$ corresponds to the causal case). It must be emphasized that the dependence w.r.t. the time index n of the impulse responses implies that the filtering process is time variant, although it can be assumed slowly varying in practice. Indeed, seismic waveforms evolve gradually with propagation depth, in contrast with steeper variations around contours in natural images. Templates are generated with standard geophysical modeling based on the above primaries. The adaptive FIR assumption is commonly adopted, and applied in partly overlapping, complementary time windows at different scales. The observation that adapted filters are ill-behaved, due to the band-pass nature of seismic data is well known, although rarely documented, motivating the need for filter coefficient control. Defining vectors $\overline{s}$ and $(\overline{h}_j)_{0\leq j<J}$ by:

$$\overline{s} = \left[\overline{s}^{(0)} \quad \cdots \quad \overline{s}^{(N-1)}\right]^\top, \tag{3}$$

$$\begin{aligned}\overline{h}_j = \Big[&\overline{h}_j^{(0)}(p') \ \cdots \ \overline{h}_j^{(0)}(p'+P_j-1) \ \cdots \\ &\cdots \ \overline{h}_j^{(N-1)}(p') \ \cdots \ \overline{h}_j^{(N-1)}(p'+P_j-1)\Big]^\top,\end{aligned} \tag{4}$$

and block diagonal matrices $(R_j)_{0\leq j<J}$ of size $N \times NP_j$:

$$R_j = \begin{bmatrix} R_j^{(0)} & 0 & \ldots & 0 \\ 0 & R_j^{(1)} & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & R_j^{(N-1)} \end{bmatrix}, \tag{5}$$

where $(R_j^{(n)})_{0\leq n\leq N-1}$ are vectors of dimension P_j such that

$$\left[(R_j^{(0)})^\top (R_j^{(1)})^\top \cdots (R_j^{(N-1)})^\top\right]^\top =$$
$$\begin{bmatrix} r_j^{(-p')} & \cdots & r_j^{(0)} & 0 & \cdots & & 0 \\ r_j^{(-p'+1)} & \cdots & & r_j^{(0)} & 0 & \cdots & 0 \\ \vdots & & & & & & \\ r_j^{(N-1)} & r_j^{(N-2)} & & \cdots & & & r_j^{(N-P_j)} \\ 0 & r_j^{(N-1)} & & \cdots & & & r_j^{(N-P_j+1)} \\ \vdots & & & & & & \\ 0 & \cdots & 0 & r_j^{(N-1)} & \cdots & & r_j^{(N-P_j-p')} \end{bmatrix}.$$

Eq. (2) can be expressed more concisely as

$$\overline{s} = \sum_{j=0}^{J-1} R_j \overline{h}_j \,. \tag{6}$$

For more conciseness, one can write $\overline{s} = \mathbf{R}\overline{\mathbf{h}}$ by defining $\mathbf{R} = [R_0 \dots R_{J-1}] \in \mathbb{R}^{N\times Q}$ where $Q = NP$ with $P = \sum_{j=0}^{J-1} P_j$ and $\overline{\mathbf{h}} = [\overline{h}_0^\top \dots \overline{h}_{J-1}^\top]^\top \in \mathbb{R}^Q$.

4 Variational formulation of the problem

4.1 Bayesian framework

We assume that the characteristics of the primary are appropriately described through a prior statistical model in a (possibly redundant) frame of signals, e.g. a wavelet frame [39]. If we denote by $\overline{x}$ the vector of frame coefficients, and $F \in \mathbb{R}^{K\times N}$ designates the associated analysis operator, we have [40]:

$$\overline{x} = F\overline{y}. \tag{7}$$

In addition, we assume that $\overline{y}$ is a realization of a random vector Y, the probability density function (pdf) of which is given by

$$(\forall y \in \mathbb{R}^N) \quad f_Y(y) \propto \exp(-\varphi(Fy)) \tag{8}$$

where $\varphi\colon \mathbb{R}^N \to]-\infty, +\infty]$ is the associated potential. If $\|\cdot\|$ designates the standard Euclidean norm, it can be noticed that, as for every $y \in \mathbb{R}^N$, $\|Fy\| \geq \sqrt{\underline{\nu}}\|y\|$ for some positive constant $\underline{\nu}$ corresponding to the lower bound of the frame F, such a prior distribution is properly defined, provided that $-\varphi$ has a fast enough decay.

On the other hand, to take into account the available information on the unknown filter, it can be assumed that for all $j \in \{0, \dots, J-1\}$, $\overline{h}_j$ is a realization of a random vector H_j. The joint pdf of the filter coefficients can be expressed as

$$(\forall h \in \mathbb{R}^{NP_1} \times \cdots \times \mathbb{R}^{NP_J}) \quad f_{H_0,\dots,H_{J-1}}(h) \propto \exp(-\rho(h)), \tag{9}$$

where $(H_0, \dots, H_{J-1})$ is independent of Y. It is further assumed that the noise vector b is a realization of a random vector B with pdf

$$(\forall b \in \mathbb{R}^N) \qquad f_B(b) \propto \exp(-\psi(b)), \tag{10}$$

where $\psi : \mathbb{R}^N \to]-\infty, +\infty]$, and that B is independent of Y and $H_0, \dots, H_{J-1}$. The posterior distribution of $(Y, H_0, \dots, H_{J-1})$ conditionally to $Z = Y + \sum_{j=0}^{J-1} R_j H_j + B$ is then given by

$$(\forall y \in \mathbb{R}^N)(\forall h \in \mathbb{R}^{NP_1} \times \cdots \times \mathbb{R}^{NP_J})$$
$$f_{Y,H_0,\dots,H_{J-1}|Z=z}(y,h) \propto \exp\left(-\psi\Big(z - y - \sum_{j=0}^{J-1} R_j h_j\Big)\right) f_Y(y) f_{H_0,\dots,H_{J-1}}(h). \tag{11}$$

By resorting to an estimation of $(\overline{y}, \overline{h}_0, \ldots, \overline{h}_{J-1})$ in the sense of the MAP (Maximum A Posteriori), the problem can thus be formulated under the following variational form:

$$\underset{y\in\mathbb{R}^N,\, h\in\mathbb{R}^{NP_1}\times\cdots\times\mathbb{R}^{NP_J}}{\text{minimize}} \ \psi(z - y - \sum_{j=0}^{J-1} R_j h_j) + \varphi(Fy) + \rho(h). \tag{12}$$

4.2 Problem formulation

For simplicity, we propose to adopt uniform priors for Y and $(H_0, \ldots, H_{J-1})$ by choosing for φ and ρ indicator functions of closed convex sets. The associated MAP estimation problem then reduces to the following constrained minimization problem:

$$\underset{Fy\in D,\, \mathbf{h}\in C}{\text{minimize}} \ \Psi(y, \mathbf{h}) \tag{13}$$

where the data fidelity term is defined by function $\Psi\colon (y, \mathbf{h}) \mapsto \psi(z - y - \mathbf{R}\mathbf{h})$, and the a priori information available on the filters and the primary are expressed through hard constraints modeled by nonempty closed convex sets C and D. One of the potential advantages of such a constrained formulation is that it facilitates the choice of the related parameters with respect to the regularized approach which was investigated in some of our previous works [19,20] (this point will be detailed later on). We will now turn our attention to the choice of Ψ, C and D.

4.3 Considered data fidelity term and constraints

4.3.1 Data fidelity term

Function Ψ accounts for the noise statistics. In this work, the noise is assumed to be additive, zero-mean, white and Gaussian. This leads to the quadratic form $\psi = \|\,.\,\|^2$.

4.3.2 A priori information on h

The filters are assumed to be time varying. However, in order to ensure smooth variations along time, we propose to introduce constraint sets

$$C_1 = \Big\{\mathbf{h} \in \mathbb{R}^Q \mid \forall (j,p), \forall n \ \in \Big\{0, \ldots, \Big\lfloor \frac{N}{2} \Big\rfloor - 1\Big\}, \\ |h_j^{(2n+1)}(p) - h_j^{(2n)}(p)| \leq \varepsilon_{j,p}\Big\} \tag{14}$$

$$C_2 = \Big\{\mathbf{h} \in \mathbb{R}^Q \mid \forall (j,p), \forall n \ \in \Big\{1, \ldots, \Big\lfloor \frac{N-1}{2} \Big\rfloor\Big\}, \\ |h_j^{(2n)}(p) - h_j^{(2n-1)}(p)| \leq \varepsilon_{j,p}\Big\}. \tag{15}$$

These constraints prevent strong variations of corresponding coefficients of the impulse response, estimated at two consecutive times. The bounds $\varepsilon_{j,p} \in [0, +\infty[$ may depend on the shape of the expected filter. For example, its dependence on the coefficient index p may enable a larger (resp. smaller) difference for filter coefficients taking larger (resp. smaller) values. Moreover, some additional a priori information can be added directly on the vector of filter coefficients $\mathbf{h}$. This amounts to defining a new convex set C_3 as a lower level set of some lower-semicontinuous convex function $\widetilde{\rho}$, by setting $C_3 = \big\{\mathbf{h} \in \mathbb{R}^Q \mid \widetilde{\rho}(\mathbf{h}) \leq \lambda\big\}$ where $\lambda \in \,]0, +\infty[$. $\widetilde{\rho}\colon \mathbb{R}^Q \to [0, +\infty[$ may correspond to simple norms such as ℓ_1 or ℓ_2-norms but also to more sophisticated ones such as a mixed $\ell_{1,2}$-norm [41]. Hence, the convex set C is defined as $C = C_1 \cap C_2 \cap C_3$. From a computational standpoint (see Section 5), it more efficient to split C in three subsets as described above.

4.3.3 A priori information on y

As mentioned in Section 4.1, we assume that the primary signal is well described through an analysis frame operator $F \in \mathbb{R}^{K\times N}$ [42]. The associated constraint can be split by defining a partition of $\{1,\dots,K\}$ denoted by $\{\mathbb{K}_\ell \mid \ell \in \{1,\dots,\mathcal{L}\}\}$. For example, for wavelet frames, $\mathcal{L}$ may correspond to the number of subbands and $\mathbb{K}_\ell$ is the ℓ-th subband. Then, one can choose $D = D_1 \times \cdots \times D_{\mathcal{L}}$ with $D_\ell = \{(x_k)_{k\in\mathbb{K}_\ell} \mid \sum_{k\in\mathbb{K}_\ell} \widetilde{\varphi}_\ell(x_k) \leq \beta_\ell\}$, where, for every $\ell \in \{1,\dots,\mathcal{L}\}$, $\beta_\ell \in]0,+\infty[$, and $\widetilde{\varphi}_\ell : \mathbb{R}^{\mathrm{card}(\mathbb{K}_\ell)} \to [0,+\infty[$ is a lower-semicontinuous convex function.

5 Primal-Dual proximal algorithm

Our objective is to provide a numerical solution to Problem (13). This amounts to minimizing function Ψ with respect to y and $\mathbf{h}$, the latter variables being constrained to belong to the constraint sets D and C, respectively. These contraints are expressed through linear operators, such as a wavelet frame analysis operator F. For this reason, primal-dual algorithms [43–45], such as the Monotone+Lipschitz Forward-Backward-Forward (M+L FBF) algorithm [46], constitute appropriate choices since they avoid some large-size matrix inversions inherent to other schemes such as the ones proposed in [47, 48]. As mentioned in Section 4.3.1, Ψ is a quadratic function and its gradient is thus Lispchitzian, which allows it to be directly handled in the M+L FBF algorithm. In order to deal with the contraints, projections onto the closed convex sets $(C_m)_{1\leq m\leq 3}$ and D are performed (these projections are described in more details in the next section).

5.1 Gradient and projection computation

From the assumption of additive zero-mean Gaussian noise, we deduce that Ψ is differentiable and

$$\nabla\Psi = 2[\mathrm{I}\ \mathbf{R}]^\top([\mathrm{I}\ \mathbf{R}]\cdot -z). \tag{16}$$

The gradient of Ψ is thus μ-Lipschitzian with

$$\mu = 2|||[\mathrm{I}\ \mathbf{R}]|||^2 \tag{17}$$

where $|||\cdot|||$ denotes the spectral norm. Note that the proposed method could be applied to other functions ψ than a quadratic one, provided that they are Lipschitz differentiable.

Now, we turn our attention to the constraint sets C and D. C models the constraints we set on the filters $\mathbf{h}$, which are split into 3 terms (see Section 4.3). We thus have to project onto each set C_m with $m \in \{1,2,3\}$. The projections onto the two first constraint sets C_1 and C_2 — imposing smooth variations along time of the corresponding tap coefficients — reduce to projections onto a set of hyperslabs of $\mathbb{R}^2$ as illustrated in Fig. 2.

More precisely, the projection onto C_1 (the projection onto C_2 yielding similar expressions) is calculated as follows: let $\mathbf{h} \in \mathbb{R}^Q$ and let $\mathbf{g}_1 = \Pi_{C_1}(\mathbf{h})$; then for every $j \in \{0,\cdots,J-1\}$, $p \in \{p',\cdots,p'+P-1\}$ and $n \in \left\{0,\dots,\left\lfloor\frac{N}{2}\right\rfloor-1\right\}$,

1. if $|h_j^{(2n+1)}(p) - h_j^{(2n)}(p)| < \varepsilon_{j,p}$, then
$$g_{j,1}^{(2n)}(p) = h_j^{(2n)}(p), \qquad g_{j,1}^{(2n+1)}(p) = h_j^{(2n+1)}(p);$$
2. if $h_j^{(2n+1)}(p) - h_j^{(2n)}(p) > \varepsilon_{j,p}$, then
$$g_{j,1}^{(2n)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} - \frac{\varepsilon_{j,p}}{2}$$
$$g_{j,1}^{(2n+1)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} + \frac{\varepsilon_{j,p}}{2};$$
3. if $h_j^{(2n+1)}(p) - h_j^{(2n)}(p) < -\varepsilon_{j,p}$, then
$$g_{j,1}^{(2n)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} + \frac{\varepsilon_{j,p}}{2}$$
$$g_{j,1}^{(2n+1)}(p) = \frac{h_j^{(2n+1)}(p) + h_j^{(2n)}(p)}{2} - \frac{\varepsilon_{j,p}}{2}.$$

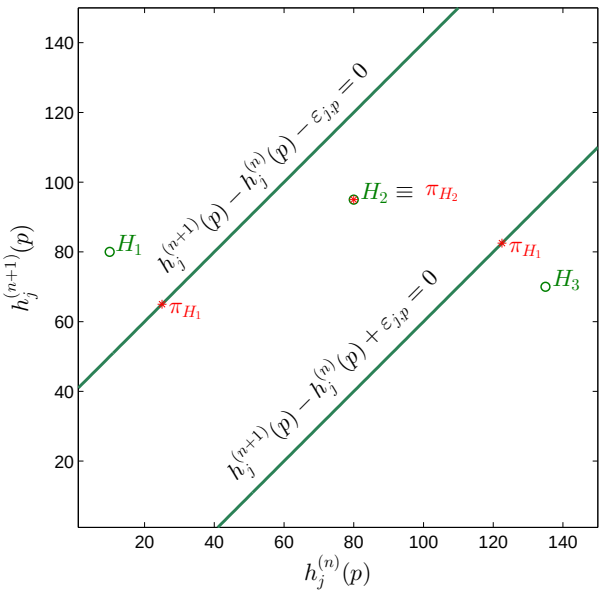


Figure 2: Projection onto C_1/C_2 of points H_1, H_2 and H_3 in $\mathbb{R}^2$.

C_3 introduces a priori information on the filter vector $\mathbf{h}$ through the lower-semicontinuous convex function $\widetilde{\rho}$. This function can be chosen separable w.r.t. $j \in \{0, \ldots, J-1\}$ in the sense that

$$\widetilde{\rho}(\mathbf{h}) = \sum_{j=0}^{J-1} \widetilde{\rho}_j(h_j). \tag{18}$$

This term can be seen as a concentration measure for the filter tap amplitude. Subsequently, we consider three possible choices for $\widetilde{\rho}_j$:

1. ℓ_1-norm:

$$\widetilde{\rho}_j(h_j) = \|h_j\|_{\ell_1} = \sum_{n=0}^{N-1} \sum_{p=p'}^{p'+P_j-1} |h_j^{(n)}(p)|. \tag{19}$$

   This choice requires to perform projections onto an ℓ_1-ball. This can be achieved by using the iterative procedure proposed in [49], which yields the projection in a finite number of iterations.

2. squared ℓ_2-norm:

$$\widetilde{\rho}_j(h_j) = \|h_j\|_{\ell_2}^2 = \sum_{n=0}^{N-1} \sum_{p=p'}^{p'+P_j-1} |h_j^{(n)}(p)|^2. \tag{20}$$

   In this case, the projection is straightforward.

3. mixed $\ell_{1,2}$-norm:

$$\widetilde{\rho}_j(h_j) = \|h_j\|_{\ell_{1,2}} = \sum_{n=0}^{N-1} \left(\sum_{p=p'}^{p'+P_j-1} |h_j^{(n)}(p)|^2 \right)^{1/2}. \tag{21}$$

   Then, we can use an algorithm similar to [49] computing the projection onto an ℓ_1 ball.

Finally, as mentioned earlier in Section 4.3, the prior information on the primary y is expressed through the frame analysis operator F by splitting the constraint into individual subband constraints. In order to promote sparsity of the coefficients, the potential function employed for the ℓ-th subband with $\ell \in \{1, \ldots, \mathcal{L}\}$ can be chosen equal to $\widetilde{\varphi}_\ell = |\cdot|$. For computing the resulting projection Π_{D_ℓ} onto an ℓ_1-ball, we can again employ the iterative procedure proposed in [49].

5.2 M+LFBF algorithm

The primal-dual approach chosen to solve the minimization problem (13) is detailed in Algorithm 1. It alternates the computations of the gradient of Ψ, and of the projections onto $(C_m)_{1\leq m\leq 3}$ and $(D_\ell)_{1\leq \ell\leq \mathcal{L}}$.

The choice of the step size is crucial for the convergence speed and it has to be chosen carefully. First, the norm of each linear operator involved in the criterion or at least an upper bound of it must be available. In our case, we have:

$$|||[\mathrm{I}\ \mathbf{R}]||| \leq \sqrt{1 + |||R_0|||^2 + \cdots + |||R_{J-1}|||^2} \tag{22}$$

where $|||R_j||| = \max_{n\in\{0,\ldots,N-1\}} \|R_j^{(n)}\|$ for every $j \in \{0,\ldots,J-1\}$. Secondly, the step size $\gamma^{[i]}$ at each iteration i must be chosen so as to satisfy the following rule: let μ be the Lipschitz constant defined in (17), let $\beta = \mu + \sqrt{|||F|||^2 + 3}$ and let $\epsilon \in]0, \frac{1}{\beta+1}[$, then $\gamma^{[i]} \in [\epsilon, \frac{1-\epsilon}{\beta}]$. $|||F|||^2$ can be easily evaluated. Indeed, in the case of a tight frame, it is equal to the frame constant and, otherwise, it can be computed by an iterative approach [50, Algorithm 4]. It is important to emphasize that the convergence of this algorithm to an optimal solution to Problem (13) is guaranteed by [46, Theorem 4.2]. In practice, the higher the norms of F and $(R_j)_{0\leq j\leq J-1}$, the slower the convergence of the algorithm. In order to circumvent this difficulty, one can resort to a preconditioned version of the algorithm [51]. However, this was not found to be useful in our experiments.

Algorithm 1 Primal-dual algo. M+LFBF to solve (13)

Let $\gamma^{[i]} \in [\epsilon, \frac{1-\epsilon}{\beta}]$

Let $\begin{bmatrix} y^{[0]} \\ \mathbf{h}^{[0]} \end{bmatrix} \in \mathbb{R}^{N+Q},\ v^{[0]} \in \mathbb{R}^K,\ \left(u_m^{[0]}\right)_{m\in\{1,2,3\}} \in (\mathbb{R}^Q)^3\ \ s_2^{[0]} \in \mathbb{R}^K,\ w_1^{[0]} \in \mathbb{R}^K$

for $i = 0, 1, \ldots$ **do**

Gradient computation

$$\begin{bmatrix} s_1^{[i]} \\ t_1^{[i]} \end{bmatrix} = \begin{bmatrix} y^{[i]} \\ \mathbf{h}^{[i]} \end{bmatrix} - \gamma^{[i]} \left(\nabla\Psi\Big(\begin{bmatrix} y^{[i]} \\ \mathbf{h}^{[i]} \end{bmatrix} \Big) + \begin{bmatrix} F^* v^{[i]} \\ \sum_{m=1}^3 u_m^{[i]} \end{bmatrix} \right)$$

Projection computation

$x_1^{[i]} = F y^{[i]}$

for $\ell = 1 : \mathcal{L}$ **do**

$$\left(s_2^{[i]}(k)\right)_{k\in\mathbb{K}_\ell} = \left(v^{[i]}(k) + \gamma^{[i]} x_1^{[i]}(k)\right)_{k\in\mathbb{K}_\ell}$$

$$\left(w_1^{[i]}(k)\right)_{k\in\mathbb{K}_\ell} = \left(s_2^{[i]}(k)\right)_{k\in\mathbb{K}_\ell} - \gamma^{[i]} \Pi_{D_\ell} \left(\frac{\left(s_2^{[i]}(k)\right)_{k\in\mathbb{K}_\ell}}{\gamma^{[i]}} \right)$$

end for

for $m = 1 : 3$ **do**

$t_{2,m}^{[i]} = u_m^{[i]} + \gamma^{[i]} \mathbf{h}^{[i]}$

$w_{2,m}^{[i]} = t_{2,m}^{[i]} - \gamma^{[i]} \Pi_{C_m} \left(\frac{t_{2,m}^{[i]}}{\gamma^{[i]}} \right)$

end for

Averaging

$x_2^{[i]} = F s_1^{[i]}$

for $\ell = 1 : \mathcal{L}$ **do**

$$\left(q_1^{[i]}(k)\right)_{k\in\mathbb{K}_\ell} = \left(w_1^{[i]}(k) + \gamma^{[i]} x_2^{[i]}(k)\right)_{k\in\mathbb{K}_\ell}$$

$$\left(v^{[i+1]}(k)\right)_{k\in\mathbb{K}_\ell} = \left(v^{[i]}(k) - s_2^{[i]}(k) + q_1^{[i]}(k)\right)_{k\in\mathbb{K}_\ell}$$

end for

for $m = 1 : 3$ **do**

$q_{2,m}^{[i]} = w_{2,m}^{[i]} + \gamma^{[i]} t_1^{[i]}$

$u_m^{[i+1]} = u_m^{[i]} - t_2^{[i]} + q_{2,m}^{[i]}$

end for

Update

$$\begin{bmatrix} y^{[i+1]} \\ \mathbf{h}^{[i+1]} \end{bmatrix} = \begin{bmatrix} y^{[i]} \\ \mathbf{h}^{[i]} \end{bmatrix} - \gamma^{[i]} \left(\nabla\Psi\Big(\begin{bmatrix} s_1^{[i]} \\ t_1^{[i]} \end{bmatrix} \Big) + \begin{bmatrix} F^* w_1^{[i]} \\ \sum_{m=1}^3 w_{2,m}^{[i]} \end{bmatrix} \right)$$

end for

6 Results

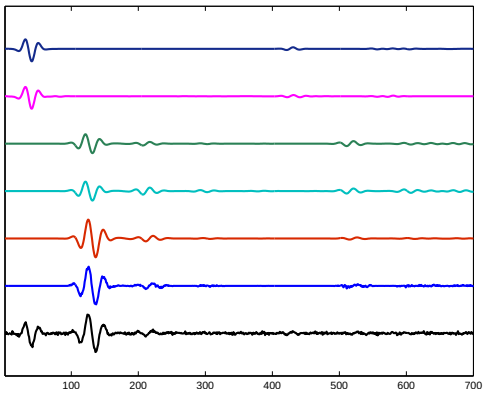
6.1 Evaluation methodology

We consider either synthetic or real data for our evaluations. The first ones are evaluated both qualitatively and quantitatively, and the choice for the sparsity norm for the wavelet coefficients is discussed. We consider a real seismic trace with denoised primaries y. We then reconstruct a synthetic multiple as follows. Two multiple templates ($J = 2$) r_0 and r_1 are independently convolved with time-varying filters and summed up to yield a known, realistic, synthetic secondary reflection signal s. The primaries are then corrupted by s and an additive Gaussian noise. The j-th time-varying filter is built upon averaging filters with length P_j, such that, $\forall p \in \{p', \ldots, p' + P_j - 1\}$, $\overline{h}_j^{(n)}(p) = \eta_j^{(n)}/P_j$ (cf. Eq. (2)). The time-varying filters are thus unambiguously defined, at a given time n, by the constants $\eta_j^{(n)}$. Uniform filters are chosen for their poor frequency selectivity behavior and notches in the frequency domain. Such artifacts for instance happen in marine seismic acquisition.

6.2 Qualitative results on simulated data

We choose here two filter families with lengths $P_0 = 10$ and $P_1 = 14$. The two filters evolve complementary in time, emulating a bi-modal multiple mixture at two different depths. They are combined with the two templates in a multiple signal depicted in red at the fifth row from the top of Fig. 3.This figure also displays the other signals of interest, known and unknown, when $\sigma = 0.08$.

We aim at recovering weak primary signals, potentially hidden under both multiple and random perturbations, from the observed signal z in black at the last row. We focus on the rightmost part of the plots, between indices 350 and 700. The primary events we are interested in are located in Fig. 3 (first signal on top) between indices 400-500 and 540-600, respectively. These primary events are mixed with multiples and random noise. A close-up on indices 350-700 is provided in Fig. 5. The first interesting primary event (400-500) is mainly affected by the random noise component. It serves as a witness for the quality of signal/random noise separation, as it is relatively insulated. The second one is disturbed by both noise and a multiple signal, relatively higher in amplitude than the first primary event. Consequently, its recovery severely probes the efficiency of the proposed algorithm.



(a)

Figure 3: Considered simulated seismic signals with noise level $\sigma = 0.08$. From top to bottom: original (unknown) signal y, estimated $\hat{y}$, first template r_0, second template r_1, multiple (unknown) s, estimated $\hat{s}$, and observed signal z.

For the proposed method, we choose the following initial settings. An undecimated wavelet frame transform with 8-length Symmlet filters is performed on $\mathcal{L} = 4$ resolution levels. The loss functions ψ and $(\widetilde{\varphi}_\ell)_{1 \leq \ell \leq \mathcal{L}}$ in (13) are chosen as $\psi = \| \, . \, \|^2$ and $\widetilde{\varphi}_\ell = | \, . \, |$. The latter is based on a selection of power laws (namely, 1, 4/3, 3/2, 2, 3, and 4) for which closed-form proximal operators exist [52, p. 1356]. The best matching power for the chosen wavelet tight frame yields the taxicab metric or ℓ_1-norm, as illustrated in Fig. 4.

The constraints C_1 and C_2 are chosen according to (14) and (15), with $\varepsilon_{1,p} = 0.1$ and $\varepsilon_{2,p} = 0.07$ for every p. The bounds of the constraints are calculated empirically on ideal signals. For real signals, we

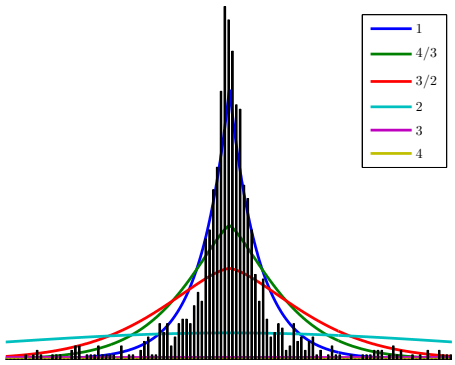


Figure 4: Generalized Gaussian modeling of seismic data wavelet frame decomposition with different power laws.

propose to infer those constants from other methods. In practice, alternative cruder filtering or restoration algorithms indeed exist, with the same purpose. They often are less involved and accurate, and potentially faster. They are run for instance on a small subset of representative real data. Thus, we obtain a first set of solutions, here separating primaries and multiples. Approximate constraints, required in the proposed method, are then computed (in a relatively fast manner) on approximate versions of unknown clean signals. Such a procedure yields coarse bound estimates, upon which the proposed algorithm is run. Although approximate, they are expected to be easier to estimate than regularization hyper-parameters. We hereafter use a fast first-pass separation of signals using [22]. Finally, $\widetilde{\rho}$ is chosen as the $\ell_{1,2}$-norm.

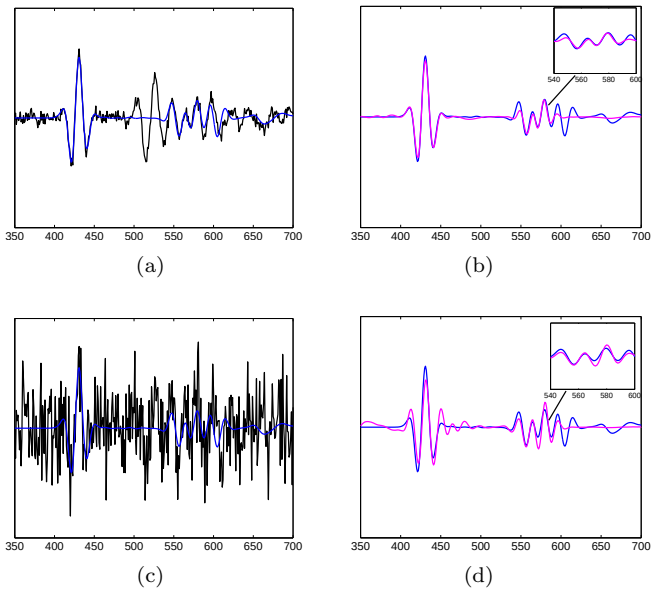


Figure 5: Close-up with $\sigma = 0.01$ (top) and $\sigma = 0.08$ (bottom) and $\widetilde{\rho}$ is the $\ell_{1,2}$-norm; (a) and (c): input data z (black), primary y (blue); (b) and (d): output separated primary $\hat{y}$ (pink) and primary y (blue).

Fig. 5 provides close-ups of delimited areas with weak primaries. With a low noise level (Fig. 5-(a)), the multiple echo at indices 500 to 550 is faithfully removed, as well as the random noise. The first strong primary (indices 400-450) is well recovered. The second one has its first four periods matched correctly. More interesting is the stronger noise condition in Fig. 5-(c). The multiple signal is still correctly removed and the incoherent noise is drastically reduced. Although with a noticeable amplitude distortion and some ringing effect, both primaries are visually recovered. As stated in Section 2, the ability to restore — albeit imperfectly — spurs of strongly hidden primaries (cf. Fig. 5-(c), between indices 540 and 600) is of paramount interest for seismic exploration in greater depths.

6.3 Quantitative results on simulated data

These first qualitative simulation results are complemented with more extensive tests on different settings of wavelet choices, levels, redundancy and adaptive filter norms, to limit the risk of unique parameter set bias effects. We test three different dyadic wavelets (Haar, Daubechies and Symmlet with filter length 8), either in orthogonal basis or shift-invariant tight frame mode, with 3 or 4 decomposition levels, consistent with seismic data bandwidth. These data decomposition settings are tested again for four noise levels and three different choices of concentration metrics for the adaptive filters. For each choice in this parameter set, 100 different noisy realizations are processed. Each of these experiments is represented by its empirical average and standard deviation. As we have seen before, the restoration of primaries or the cancellation of multiples could be jointly pursued. We thus report in Tables 1 and 2 the average SNR, for the clean modeled primary and multiple, with respect to their restored counterpart, respectively. Column headers b and f denote basis and frame results, whose averages are loosely denoted by μ_b and μ_f. To improve reading, numbers in bold (for Table 2 as well) indicate the best result for a given decomposition level $\mathcal{L}$. Numbers in italics denote the best SNRs obtained, irrespective of the number of wavelet levels. The standard deviation tables are combined with Tables 1 and 2 in Table 3 as a "significance index" of the outcome.

We first exemplify results in the leftmost part of Table 1 (Haar wavelet), for the first two rows. For the ℓ_1-norm, we observe a primary restoration improvement of 2.2 dB ($21.3-19.1$, with standard deviations of 0.16 dB for frames and 0.26 dB for bases) for 3 wavelet levels. For 4 levels, we obtain 1.2 dB with standard deviations of 0.18 dB and 0.22 dB. Intuitively, the SNR improvement appears to be significant, relatively to the dispersion. We shall detail this aspect later on. Yet, such a sensible variation of about 1 dB, with only one further wavelet decomposition level, further justifies the need for the given multi-parameter analysis.

In most cases, for frames, the best results are obtained with 4 levels. When this is not the case, the difference in performance generally lies within the dispersion. This assertion cannot be stated with bases, possibly due to shift variance effects. The frame-based SNR for primaries is always greater, or equal to that of the basis one, putting statistical significance aside for the moment. Namely, looking at summary statistics, the minimum, median, mean and maximum improvements for frames over bases are 0.5 dB, 1.8 dB, 2 dB and 4.2 dB respectively. Looking at numbers in bold, we see that a frame with the ℓ_2 loss function is the clear winner in absolute SNR, for every wavelet choice and noise level.

The results for multiple estimation, given in Table 2, are more contrasted. Frames and bases yield more similar performance, especially for high Gaussian noise levels. The best overall results (bold) are given by $\ell_{1,2}$ (high noise) and ℓ_2-norms (low noise).

The above "absolute" results allow the observation of global trends. In practice, we consider more important to obtain consistent results, taking into account SNR dispersion. Assuming denoised realizations follow a Gaussian distribution, we now study the difference Gaussian pdf between frames and bases. Its mean is $\mu_{f/b} = \mu_f - \mu_b$ and its variance $\sigma^2_{f/b}$ is $\sigma^2_b + \sigma^2_f$. Table 3 reports the normalized difference significance index $t_{f/b} = \mu_{f/b}/\sigma_{f/b}$, reminiscent of the Student's test.

We deem the difference in distribution between bases and frames significant only if $|t_{f/b}| > 1$. When significant, the index is emphasized in italics or bold, the latter denoting the most significant among the three concentration measures. Since the minimum, median, mean and maximum indices $t_{f/b}$ are 1.3, 3.4, 3.5 and 7.7, we consider the improvement of frames over bases significant for all the tested parameters for primaries (Table 3-left). Interestingly, whereas the ℓ_2-norm gives the best average gain, the $\ell_{1,2}$ and the ℓ_1-norms yield sensibly more significance at lower noises. Choosing the $\ell_{1,2}$ or ℓ_1-norm is consequently more interesting in practice, as we desire more consistent results under unknown noise variations in observed signals. At higher noise levels, the normalized difference index $t_{f/b}$ (between values 3 and 4) is very close for all parameters and wavelets. Thus, the significance of the different filter concentration measures is not fundamentally different. Based on the previous observations, the choice of $\ell_{1,2}$ or ℓ_1-norms would yield more consistent results in all cases.

The rightmost and lower part of Table 3 concerns the restoration of multiples. Although of weaker importance in practice, we notice that most $t_{f/b}$ values vary between -1 and 1. This indicates that bases and frames perform similarly, since there is no significant performance difference. This phenomenon can be explained by the fact that frames or bases, through the wavelet transform sparsity prior, impact primaries rather than multiples. In few cases though, we observe, at low noise levels, some significance in frame outcomes over basis results, obtained by the $\ell_{1,2}$ and ℓ_1-norms again.

Globally, the restoration of both primaries and multiples benefits from the choice of a Daubechies or Symmlet wavelet frame. The best performance — in terms of statistical significance — is offered by

sparsity-promoting ℓ_1 and $\ell_{1,2}$-norms, either at lower and higher noise level.

		mean(SNR_y)																	
		Haar						Daubechies						Symmlet					
$\widetilde{\rho}$		ℓ_1		ℓ_2		$\ell_{1,2}$		ℓ_1		ℓ_2		$\ell_{1,2}$		ℓ_1		ℓ_2		$\ell_{1,2}$	
σ	$\mathcal{L}$	b	f	b	f	b	f	b	f	b	f	b	f	b	f	b	f	b	f
0.01	3	19.1	21.3	20.2	**21.9**	18.9	21.3	*20.9*	21.8	21.1	**22.3**	20.1	21.8	19.9	21.8	21.4	**22.5**	19.2	21.9
	4	*20.6*	*21.8*	*21.6*	***22.4***	*20.3*	*21.8*	20.8	*22.0*	*22.1*	***22.7***	*21.1*	*22.1*	*20.3*	*22.0*	*22.3*	***22.8***	*20.4*	*22.1*
0.02	3	19.2	20.3	20.1	**21.1**	18.6	20.1	*20.6*	21.3	*21.2*	**22.1**	*20.4*	21.2	*20.1*	21.5	21.4	**22.3**	*19.5*	21.3
	4	*19.8*	*20.8*	*20.5*	***21.3***	*19.5*	*20.6*	20.1	*21.6*	21.0	***22.2***	20.0	*21.5*	20.0	*21.6*	*21.5*	***22.4***	19.2	*21.6*
0.04	3	16.5	18.2	16.9	**18.5**	16.2	18.0	*18.0*	20.1	18.3	**20.6**	*17.9*	19.9	*17.9*	20.2	*18.4*	***20.8***	*17.6*	20.0
	4	*16.7*	*18.5*	*16.9*	***18.8***	*16.5*	*18.3*	18.0	*20.2*	*18.3*	***20.7***	17.7	*20.0*	17.9	*20.3*	18.21	**20.7**	17.5	*20.1*
0.08	3	12.3	14.7	12.5	**14.9**	12.2	14.6	*14.3*	17.4	*14.4*	**17.7**	*14.2*	17.3	*14.1*	*17.6*	*14.3*	***17.9***	*14.0*	*17.5*
	4	*12.4*	*15.0*	*12.5*	***15.2***	*12.3*	*14.9*	14.2	*17.5*	14.4	***17.7***	14.1	*17.3*	13.9	17.5	14.0	**17.7**	13.2	17.3

Table 1: SNR, averaged over 100 noise realizations for the estimations of y.

		mean(SNR_s)																	
		Haar						Daubechies						Symmlet					
$\widetilde{\rho}$		ℓ_1		ℓ_2		$\ell_{1,2}$		ℓ_1		ℓ_2		$\ell_{1,2}$		ℓ_1		ℓ_2		$\ell_{1,2}$	
σ	$\mathcal{L}$	b	f	b	f	b	f	b	f	b	f	b	f	b	f	b	f	b	f
0.01	3	26.3	27.0	27.4	**28.4**	26.2	28.0	*27.6*	28.0	28.2	**28.3**	26.9	28.0	26.7	28.0	27.9	**28.4**	26.5	28.0
	4	*27.6*	*28.3*	*28.4*	***28.7***	*27.4*	*28.4*	27.6	*28.1*	***28.6***	*28.5*	*27.7*	*28.1*	*27.1*	*28.1*	*28.5*	***28.6***	*27.1*	*28.2*
0.02	3	25.3	25.6	**25.7**	**25.7**	25.0	25.5	***25.8***	25.5	*25.8*	25.6	***25.8***	25.5	*25.3*	25.5	**25.8**	25.6	25.1	25.5
	4	*25.8*	*25.7*	***26.0***	*25.8*	*25.7*	*25.8*	25.5	*25.6*	25.7	***25.6***	25.5	***25.6***	25.3	*25.6*	***25.9***	*25.6*	*25.3*	*25.6*
0.04	3	22.1	22.1	21.8	21.8	22.2	**22.3**	*22.3*	22.1	20.1	*21.8*	***22.5***	*22.3*	*22.1*	22.1	21.8	21.8	*22.2*	***22.3***
	4	*22.3*	*22.2*	*21.9*	*21.8*	***22.4***	*22.3*	22.2	*22.1*	*21.8*	*21.8*	**22.4**	22.2	21.2	*22.1*	*21.8*	*21.8*	*22.2*	**22.3**
0.08	3	18.4	18.4	17.7	*17.8*	18.6	***18.6***	*18.5*	18.4	*17.8*	*17.8*	***18.7***	*18.6*	18.4	18.4	*17.8*	*17.8*	***18.6***	***18.6***
	4	*18.4*	*18.4*	*17.8*	*17.8*	***18.6***	*18.6*	18.5	*18.4*	*17.8*	17.7	**18.7**	18.6	*18.4*	*18.4*	*17.8*	17.8	**18.6**	18.6

Table 2: SNR, averaged over 100 noise realizations for the estimations of s.

		$t_{f/b}$ for primaries									$t_{f/b}$ for multiples								
		Haar			Daubechies			Symmlet			Haar			Daubechies			Symmlet		
$\widetilde{\rho}$		ℓ_1	ℓ_2	$\ell_{1,2}$	ℓ_1	ℓ_2	$\ell_{1,2}$	ℓ_1	ℓ_2	$\ell_{1,2}$	ℓ_1	ℓ_2	$\ell_{1,2}$	ℓ_1	ℓ_2	$\ell_{1,2}$	ℓ_1	ℓ_2	$\ell_{1,2}$
σ	$\mathcal{L}$	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b	f/b
0.01	3	*7.3*	*4.3*	**7.7**	*2.6*	*3.0*	**4.9**	*5.6*	*2.6*	**7.2**	*2.7*	*3.2*	**7.5**	*1.3*	0.3	**3.5**	*5.3*	*1.5*	**6.0**
	4	*4.2*	*2.8*	**5.1**	**4.2**	*1.9*	*3.3*	**5.3**	*1.3*	*5.1*	*2.9*	*1.1*	**4.2**	**2.5**	-0.2	*1.6*	**4.4**	0.1	**4.4**
0.02	3	*3.0*	*2.7*	**3.4**	*1.5*	**2.1**	*1.6*	*2.8*	*2.1*	**3.1**	0.7	0.0	*1.6*	*-1.1*	-0.8	-0.9	0.5	-0.5	**1.5**
	4	**2.6**	*2.3*	*2.5*	**3.0**	*2.9*	*2.7*	*3.0*	*1.7*	**4.3**	-0.1	-0.6	0.1	0.3	-0.3	0.2	0.8	-0.7	**1.2**
0.04	3	*3.4*	**3.5**	*3.2*	*3.0*	**3.9**	*2.7*	*3.2*	**3.8**	*3.2*	0.0	0.0	0.1	-0.6	**4.7**	-0.6	-0.2	-0.1	0.1
	4	*3.5*	**3.7**	*3.3*	*3.2*	**3.8**	*2.8*	*3.3*	**3.7**	*3.3*	-0.3	-0.1	-0.2	-0.3	-0.1	-0.3	**2.6**	-0.1	0.1
0.08	3	**3.5**	**3.5**	**3.5**	*3.5*	**3.9**	*3.3*	*3.8*	**4.2**	*3.7*	0.0	0.0	0.1	-0.2	0.0	-0.1	-0.1	0.0	0.0
	4	**3.8**	**3.8**	*3.7*	*3.4*	**3.6**	*3.2*	*3.8*	*4.1*	**4.2**	0.0	0.0	0.0	-0.1	0.0	-0.2	-0.1	0.0	0.0

Table 3: Normalized difference significance index.

6.4 Comparative evaluation: synthetic data

The above objective and subjective results are completed by a comparative evaluation with the empirical algorithm proposed in [22]. It is based on adaptive filtering on sliding windows in a complex continuous wavelet domain. The chosen complex Morlet wavelet is very efficient at concentrating seismic data energy. One-tap Wiener-like (unary) complex filters are adaptively estimated from the complex-valued discretized wavelet transform. This algorithm was successfully tested against industry standards. It is quantitatively faster and does not allow complicated constraints on signal sparsity or filter regularity. Fig. 6 represents synthetic 2D seismic data, constructed similarly as the previous 1D traces in Section 6.2, with a noise level of $\sigma = 0.01$. Vertical traces are stacked laterally. From left to right, the synthetic traces drift away from the seismic source. The bended hyperbolas correspond to primaries. The flatter one, below, mimics

a multiple event. Here, $P_0 = 6$, $P_1 = 6$, and constraints C_1 and C_2 are chosen according to (14) and (15), where $\varepsilon_{1,p} = 0.1$ and $\varepsilon_{2,p} = 0.1$ for every p. Apparently, slightly better primary preservation is obtained with the proposed method, for a very simple synthetic data set. This phenomenon is observed at the crossing between primaries and multiples. The proposed method also effectively gets rid of more incoherent noise.

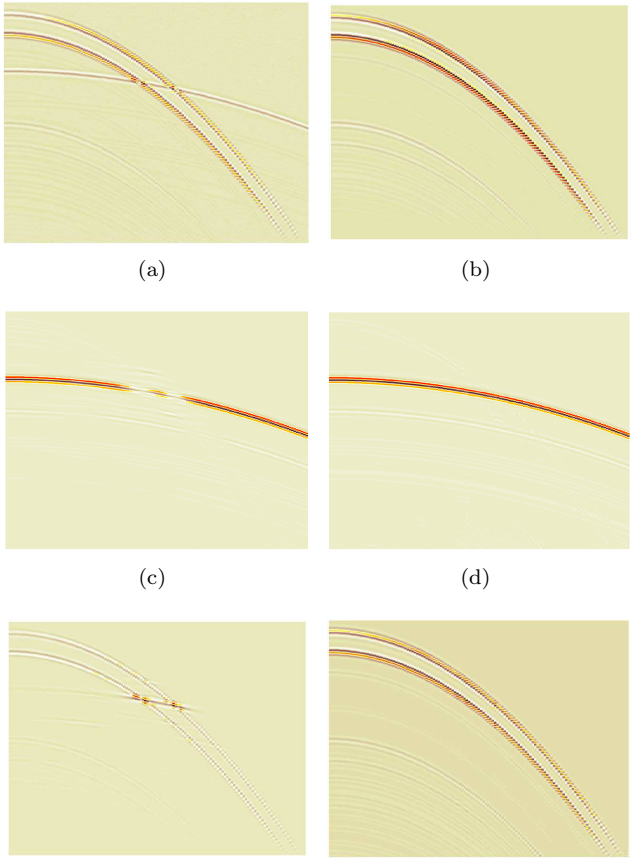


Figure 6: Data composed by three events ($\sigma = 0.01$), one primary and two multiples (a); multiples composed by two estimated events (b); output separated primaries with [22] (c) our method (d); output adapted multiples with [22] (e) our method (f).

6.5 Comparative evaluation: real data

The previous example is a little bit idealistic. We finally compare our algorithm with [22] on a portion of a real seismic data set, displayed in Fig. 7. The main objective is to uncover a potential primary, masked by strong multiple events that mostly contribute to the observed signal energy. It appears partially as an horizontal stripe in the middle of the figure, on the left side. Geophysically speaking, it should be re-linked with one of the slopped stripes on the left side. In this case, we use three multiple templates, not represented. Fig. 8 compares the results obtained with [22] and our algorithm. Since interpreting seismic data is not straightforward, we focus on three different zones indicated in red, related to noise, primaries and multiples. Both methods are able to successfully retrieve a weak primary below the multiple level. First, the red square indicates that the level of random noise looks improved with our method. Second, the oval depicts a better recovery of the primary, that has gained in continuity at the apex. Last, the junction is made with one of the left side slopped events. On Fig. 8-(a), arrows indicates remnants of multiples, parallel to the primary, that have been incorrectly subtracted. Although the random noise is apparently highly heteroskedastic, those observations are consistent with previously stated results, and expected by the management of additive noise in our algorithm.

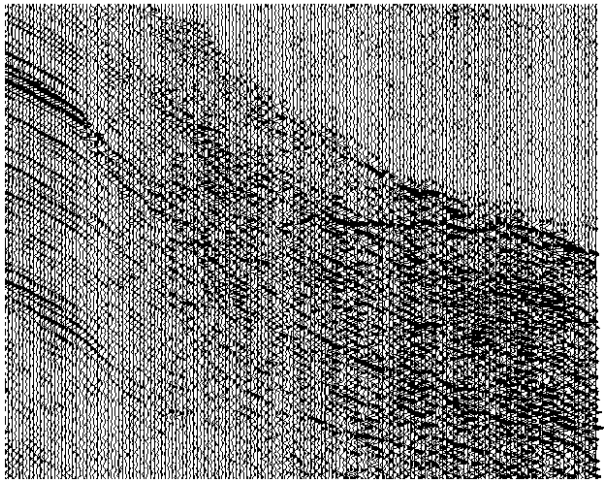


Figure 7: Portion of a receiver gather: recorded data.

7 Conclusions

We have proposed a generic methodology to impose sparsity and regularity properties through constrained adaptive filtering in a transformed domain. This method exploits side information from approximate disturbance templates. The employed proximal framework permits different strategies for sparse modeling, additive noise removal, and adaptive filter design under appropriate regularity and amplitude concentration constraints. The proposed approach is evaluated on seismic data using different orthogonal wavelet bases and tight frames, and various sparsity measures for wavelet coefficients. The standard sparsity-prone ℓ_1-norm is usefully complemented by alternative concentration measures, such as ℓ_2 or $\ell_{1,2}$-norms, which seem better suited to design adaptive filters. The potential of the proposed method is demonstrated on both synthetic and real seismic data, in comparison with a reference multiple filtering algorithm. Its performance is interesting for instance in recovering weak signals buried under both strong random and structured noise.

In our future work, two-dimensional directional multiscale approaches [42] may provide sparser representations for seismic data. Second, sparsity priors could be enforced on multiple signals as well, with a need for more automation in optimal choices on loss functions in the proximal formulation, potentially by using other measures than ℓ_1 [53–55].

References

[1] D. L. Duttweiler. Proportionate normalized least-mean-squares adaptation in echo cancelers. *IEEE Trans. Speech Audio Process.*, 8(5):508–518, 2000.

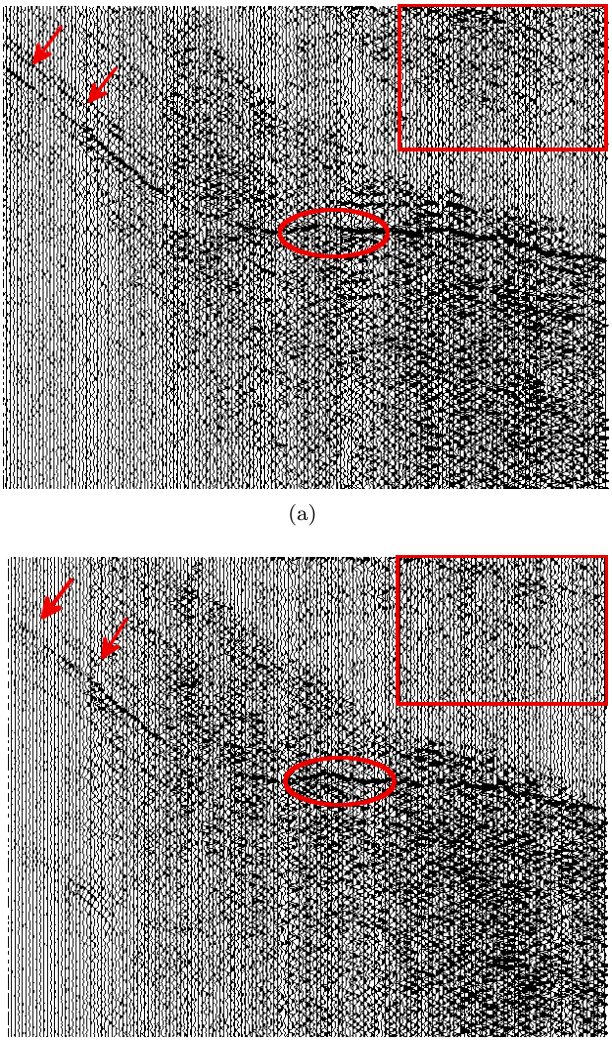
[2] R. K. Martin, W. A. Sethares, R. C. Williamson, and C. R. Johnson, Jr. Exploiting sparsity in adaptive filters. *IEEE Trans. Signal Process.*, 50(8):1883–1894, 2002.

[3] A. W. H. Khong and P. A. Naylor. Efficient use of sparse adaptive filters. In *Proc. Asilomar Conf. Signal, Syst. Comput.*, pages 1375–1379, 2006.

[4] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. r. Acad. Sci., Paris, Sér. A Math.*, 255:2897–2899, 1962.

[5] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer Verlag, 2011.

[6] Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada. A sparse adaptive filtering using time-varying soft-thresholding techniques. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, pages 3734–3737, Dallas, Texas, USA, Mar. 14-19, 2010.



(a)

(b)

Figure 8: Subtraction results on the first channel gather: Output separated primaries with (a) [22] (b) with our method.

[7] M. Yukawa, Y. Tawara, M. Yamagishi, and I. Yamada. Sparsity-aware adaptive filters based on ℓ_p-norm inspired soft-thresholding technique. In *Proc. Int. Symp. Circuits Syst.*, pages 2749–2752, Seoul, Korea, May 20-23, 2012.

[8] A. Gilloire and M. Vetterli. Adaptive filtering in sub-bands with critical sampling: analysis, experiments, and application to acoustic echo cancellation. *IEEE Trans. Signal Process.*, 40(8):1862–1875, 1992.

[9] B. E. Usevitch and M. T. Orchard. Adaptive filtering using filter banks. *IEEE Trans. Circ. Syst. II*, 43(3):255–265, 1996.

[10] S. Hosur and A. H. Tewfik. Wavelet transform domain adaptive FIR filtering. *IEEE Trans. Signal Process.*, 45(3):617–630, 1997.

[11] M. R. Petraglia and P. B. Batalheiro. Nonuniform subband adaptive filtering with critical sampling. *IEEE Trans. Signal Process.*, 56(2):565–575, 2008.

[12] J. Gauthier, L. Duval, and J.-C. Pesquet. Optimization of synthesis oversampled complex filter banks. *IEEE Trans. Signal Process.*, 57(10):3827–3843, Oct. 2009.

[13] C. Chaux, J.-C. Pesquet, and L. Duval. Noise covariance properties in dual-tree wavelet decompositions. *IEEE Trans. Inform. Theory*, 53(12):4680–4700, Dec. 2007.

[14] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet. A nonlinear Stein based estimator for multichannel image denoising. *IEEE Trans. Signal Process.*, 56(8):3855–3870, Aug. 2008.

[15] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, 57(12):4616–4632, Dec. 2009.

[16] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51(1):34–81, 2009.

[17] E. A. Robinson and S. Treitel. *Geophysical Signal Analysis.* Prentice Hall, Englewood Cliffs, NJ., 1980.

[18] R. Essenreiter, M. Karrenbach, and S. Treitel. Multiple reflection attenuation in seismic data using backpropagation. *IEEE Trans. Signal Process.*, 46(7):2001–2011, Jul. 1998.

[19] D. Gragnaniello, C. Chaux, J.-C. Pesquet, and L. Duval. A convex variational approach for multiple removal in seismic data. In *Proc. Eur. Sig. Image Proc. Conf.*, pages 215–219, Bucharest, Romania, Aug. 27-31, 2012.

[20] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet. Seismic multiple removal with a Primal-Dual proximal algorithm. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, pages 2257–2261, Vancouver, Canada, May 2013.

[21] A. B. Weglein, S.-Y. Hsu, P. Terenghi, X. Li, and R. H. Stolt. Multiple attenuation: Recent advances and the road ahead (2011). *The Leading Edge*, 30(8):864–875, 2011.

[22] S. Ventosa, S. Le Roy, I. Huard, A. Pica, H. Rabeson, P. Ricarte, and L. Duval. Adaptive multiple subtraction with wavelet-based complex unary Wiener filters. *Geophysics*, 77(6):V183–V192, Nov.-Dec. 2012.

[23] A. T. Walden and J. W. J. Hosken. The nature of the non-gaussianity of primary reflection coefficients and its significance for deconvolution. *Geophys. Prospect.*, 34(7):1038–1066, 1986.

[24] L. T. Duarte, E. Z. Nadalin, K. N. Filho, R. A. Zanetti, J. M. T. Romano, and M. Tygel. Seismic wave separation by means of robust principal component analysis. In *Proc. Eur. Sig. Image Proc. Conf.*, pages 1494–1498, Bucharest, Romania, Aug. 2012.

[25] A. K. Takahata, E. Z. Nadalin, R. Ferrari, L. T. Duarte, R. Suyama, R. R. Lopes, J. M. T. Romano, and M. Tygel. Unsupervised processing of geophysical signals: A review of some key aspects of blind deconvolution and blind source separation. *IEEE Signal Process. Mag.*, 29(4):27–35, Jul. 2012.

[26] P. Bois and C. Hémon. Étude statistique de la contribution des multiples aux sismogrammes synthétiques et réels. *Geophys. Prospect.*, 11(3):313–349, 1963.

[27] H. Krim and J.-C. Pesquet. Multiresolution analysis of a class of nonstationary processes. *IEEE Trans. Inform. Theory*, 41:1010–1020, Jul. 1995.

[28] D. Donno. Improving multiple removal using least-squares dip filters and independent component analysis. *Geophysics*, 76(5):V91–V104, 2011.

[29] R. Neelamani, A. Baumstein, and W. S. Ross. Adaptive subtraction using complex-valued curvelet transforms. *Geophysics*, 75(4):V51–V60, 2010.

[30] F. J. Herrmann and D. J. Verschuur. Curvelet imaging and processing: adaptive multiple elimination. In *Proc. CSEG Nat. Conv.* Canadian Soc. Expl. Geophysicists, 2004.

[31] A. Guitton and D. J. Verschuur. Adaptive subtraction of multiples using the l_1-norm. *Geophys. Prospect.*, 52:27–38, 2004.

[32] S. Costagliola, P. Mazzucchelli, and N. Bienati. Hybrid norm adaptive subtraction for multiple removal. In *Proc. EAGE Conf. Tech. Exhib.*, Vienna, Austria, May 23-26, 2011.

[33] A. Gholami and M. D. Sacchi. A fast and automatic sparse deconvolution in the presence of outliers. *IEEE Trans. Geosci. Rem. Sens.*, 50(10):4105–4116, Oct. 2012.

[34] Y. Yang, J. Ma, and S. Osher. Seismic data reconstruction via matrix completion. Technical Report UCLA CAM report 12-14, University of California, Los Angeles, 2012.

[35] N. Kreimer and M. D. Sacchi. Nuclear norm minimization and tensor completion in exploration seismology. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, Vancouver, Canada, May 2013.

[36] M. S. O'Brien, A. N. Sinclair, and S. M. Kramer. Recovery of a sparse spike time series by l_1 norm deconvolution. *IEEE Trans. Signal Process.*, 42(12):3353–3365, 1994.

[37] J.-C. Pesquet, H. Krim, and H. Carfantan. Time-invariant orthonormal wavelet representations. *IEEE Trans. Signal Process.*, 44(8):1964–1970, Aug. 1996.

[38] J. E. Fowler. The redundant discrete wavelet transform and additive noise. *Signal Process. Lett.*, 12(9):629–632, Sep. 2005.

[39] O. Christensen. *Frames and bases. An introductory course.* Birkhäuser, 2008.

[40] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame based inverse problems. *Inverse Probl.*, 23(4):1495–1518, Aug. 2007.

[41] M. Kowalski. Sparse regression using mixed norms. *Appl. Comp. Harm. Analysis*, 27(3):303–324, 2009.

[42] L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12):2699–2730, Dec. 2011.

[43] L. M. Briceños Arias and P. L. Combettes. A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Optim.*, 21(4):1230–1250, Oct. 2011.

[44] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40(1):120–145, May 2011.

[45] L. Condat. A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, pages 1–20, 2012.

[46] P. L. Combettes and J.-C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.*, 20(2):307–330, Jun. 2012.

[47] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, 20(3):681–695, Mar. 2011.

[48] J.-C. Pesquet and N. Pustelnik. A parallel inertial proximal optimization method. *Pacific J. Opt.*, 8(2):273–305, Apr. 2012.

[49] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, Nov. 2008.

[50] L. Chaari, N. Pustelnik, C. Chaux, and J.-C. Pesquet. Solving inverse problems with overcomplete transforms and convex optimization techniques. In *Proc. SPIE, Wavelets*, San Diego, California, USA, Aug. 2-6, 2009.

[51] A. Repetti, É. Chouzenoux, and J.-C. Pesquet. A penalized weighted least squares approach for restoring data corrupted with signal-dependent noise. In *Proc. Eur. Sig. Image Proc. Conf.*, pages 1553–1557, Bucharest, Romania, Aug. 27-31, 2012.

[52] P. L. Combettes and J.-C. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Optim.*, 18(4):1351–1376, 2007.

[53] I. W. Selesnick and İ. Bayram. Sparse signal estimation by maximally sparse convex optimization. *PREPRINT*, Aug. 8, 2013. arXiv:1302.5729.

[54] G. Marjanovic and V. Solo. On exact l_q denoising. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, Vancouver, Canada, May 26-31, 2013.

[55] A. E. Çetin, A. Bozkurt, O. Gunay, Y. H. Habiboglu, K. Kose, I. Onaran, and R. A. Sevimli. Projections onto convex sets (POCS) based optimization by lifting. *PREPRINT*, Jun. 13, 2013. arXiv:1306.2516.