A CONSTRAINED-BASED OPTIMIZATION APPROACH FOR SEISMIC DATA RECOVERY PROBLEMS

Mai Quyen Pham[1,3], Caroline Chaux[2], Laurent Duval[1], and Jean-Christophe Pesquet[3]

[1] IFP Energies nouvelles
Direction Technologie
1 et 4 av. de Bois-Préau
92852 Rueil-Malmaison, France
{mai-quyen.pham, laurent.duval}@ifpen.fr

[2] Aix-Marseille Univ.
I2M UMR CNRS 7373
39 rue F. Joliot-Curie,
13453 Marseille, France
caroline.chaux@latp.univ-mrs.fr

[3] Univ. Paris-Est
LIGM UMR-CNRS 8049
5 bd Descartes,
77454 Marne-la-Vallée, France
pesquet@univ-mlv.fr

ABSTRACT

Random and structured noise both affect seismic data, hiding the reflections of interest (primaries) that carry meaningful geophysical interpretation. When the structured noise is composed of multiple reflections, its adaptive cancellation is obtained through time-varying filtering, compensating inaccuracies in given approximate templates. The under-determined problem can then be formulated as a convex optimization one, providing estimates of both filters and primaries. Within this framework, the criterion to be minimized mainly consists of two parts: a data fidelity term and hard constraints modeling a priori information. This formulation may avoid, or at least facilitate, some parameter determination tasks, usually difficult to perform in inverse problems. Not only classical constraints, such as sparsity, are considered here, but also constraints expressed through hyperplanes, onto which the projection is easy to compute. The latter constraints lead to improved performance by further constraining the space of geophysically sound solutions.

Index Terms— Optimization methods, Wavelet transforms, Adaptive filters, Geophysical signal processing, Signal restoration.

1. INTRODUCTION

Adaptive filtering techniques are meant to optimize coefficients of variable filters, according to adapted cost functions working on error signals. Adaptive subtraction [1, 2] is at play in seismic data recovery problems where approximate models are adapted or matched to actual data, throughout adaptive filters. These models are obtained from geophysical modeling, and known a priori. One such situation is the filtering of secondary reflexions, or multiples. Geophysical signals of interest, named primaries, follow wave paths depicted in dotted, dashed and solid blue in Fig. 1. Since the data recovery problem is generally under-determined, geophysicists have developed pioneering sparsity-promoting techniques. For instance, robust, ℓ_1-promoted deconvolution [3] or complex wavelet transforms [4, 5] still pervade many areas of signal processing.
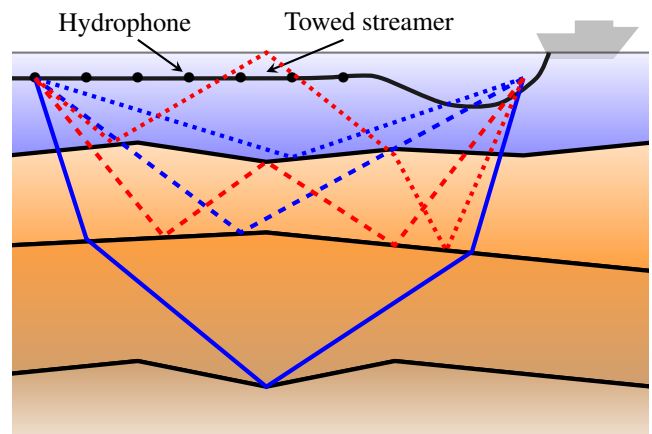

Fig. 1. Principles of marine seismic data acquisition and wave propagation. Towed streamer with hydrophones. Reflections on different layers (primaries in blue), and reverberated disturbances (multiple in dotted and dashed red).

Although the contributions are generally considered linear, several types of disturbances, structured or more stochastic, affect the relevant information present in seismic data. Multiples correspond to seismic waves bouncing between layers [6], 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 bedim 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. We suppose here that one or several approximate templates of potential multiples are determined, off-line, based on primary reflections identified in above layers or wave-propagation modeling. Model-based multiple filtering is similar to adaptive echo cancellation practice (see [7] for details), and is now considered as a geophysics industry standard.

We propose a methodology for primary/multiple adaptive separation based on these approximate templates. It addresses

at the same time structured reverberations and a more stochastic part. Let $n \in \{0, \ldots, N-1\}$ denote the time index for the observed seismic trace z, acquired by a given sensor (here, an hydrophone). We assume, as customary in seismic, a linear 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 $(\overline{s}^{(n)})_{0 \leq n < N}$. Other unstructured contributions are gathered in the noise term $b = (b^{(n)})_{0 \leq n < N}$. We assume that J templates $(r_j^{(n)})_{0 \leq n < N, 0 \leq j < J}$ for the disturbance signal are available, which are related to $(\overline{s}^{(n)})_{0 \leq n < N}$ through an FIR (Finite Impulse Response), possibly non-causal, linear model

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

where $\overline{h}_j^{(n)}$ is an unknown impulse response (P_j tap coefficients) corresponding to template j and time n and where $p' \in \{-P_j + 1, \ldots, 0\}$ ($p' = 0$ 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 not time invariant, although it can be assumed slowly varying in practice.

The purpose of this work is to provide means to identify $\overline{y}$ and $\overline{h}_j$, by imposing hopefully meaningful constraints onto the above system.

2. RELATION TO PRIOR WORK

The separation of primaries and multiples is a classical issue in seismic exploration. Most published solutions, tailored to specific levels of prior knowledge, are very dependent on seismic data-sets. They generally rely on adapted transforms (Radon, Fourier transforms) and some form of least-squares adaptive filtering. Among the vast literature, we refer to [2, 8], for a recent account on adaptive subtraction of multiples, including shortcomings of standard ℓ_2-based methods. With weak primary/multiple decorrelation, poor data stationarity or higher noise levels, traditional methods fail. Due to the parsimonious layering [9] of the subsurface (illustrated in Fig. 1), sparsity promotion suggests the use of sparsifying transforms (e.g. wavelet/curvelet frames [10, 11]), potentially combined with robust norms (approximate ℓ_1 in [12]), quasi-norms or source separation methods [13, 14]. To date, their genericity may be limited by the number of possible penalties to constrain feasible solutions, and the crucial issue of hyperparameter determination in such methods.

In [7, 15], the authors incorporate plausible knowledge via additional metrics. Prior multiple templates are supplemented with Gaussian noise assumptions, wavelet-domain sparsity, smooth variations and energy concentration criteria. Joint estimation of primaries and adaptive filters is performed with a proximal algorithm. To alleviate the hyperparameter estimation issue, we reformulate the previous approach as a constrained minimization problem. This allows us to more easily determine data-based parameters. We focus here on the exploration of various constraint efficiency in wavelet frame subbands for the primary signal. Interestingly, convex sets defined as appropriate hyperplanes can outperform standard ℓ_1-ball constraints.

The paper is organized as follows. In Section 3 we rewrite the observation model and formulate the constrained optimization problem. The definition of the constraint sets and the adopted optimization strategy follows. Section 4 details the simulation results. Finally, conclusions are drawn in Section 5.

3. CONSTRAINED FORMULATION

3.1. Observation model

Model (2) can be written more concisely as

$$\overline{s} = \mathbf{R}\overline{\mathbf{h}} \tag{3}$$

by appropriately defining $\mathbf{R} \in \mathbb{R}^{N \times Q}$, where $Q = NP$ with $P = \sum_{j=0}^{J-1} P_j$ and $\overline{\mathbf{h}} \in \mathbb{R}^Q$ [15]. On the one hand, the matrix $\mathbf{R}$ contains the J templates for every time index n and tap index p. On the other hand, the vector $\overline{\mathbf{h}}$ is similarly defined as the concatenation of all (unknown) time-varying filter impulse responses. With this notation, the observed data z are given by

$$z = \overline{y} + \mathbf{R}\overline{\mathbf{h}} + b. \tag{4}$$

Now, we turn our attention to solving the ill-posed inverse problem of estimating $\overline{y}$ and $\overline{\mathbf{h}}$ from the observation vector z.

3.2. Constrained problem formulation

Our objective here is to propose a variational approach aiming at providing relevant estimates of the primary signal $\overline{y}$ and time-varying filters $\overline{\mathbf{h}}$ related to multiples. To this end, define an objective function composed of two convex terms being related to either $\overline{y}$ or $\overline{\mathbf{h}}$ through functions $\varphi : \mathbb{R}^N \to \,]-\infty, +\infty]$ and $\rho : \mathbb{R}^Q \to \,]-\infty, +\infty]$, respectively. We propose to solve the following constrained minimization problem

$$\underset{y \in \mathbb{R}^N, \mathbf{h} \in \mathbb{R}^Q}{\text{minimize}} \quad \alpha \rho(\mathbf{h}) + (1-\alpha)\varphi(y)$$

$$\text{subject to} \begin{cases} \psi(z - y - \mathbf{R}\mathbf{h}) \leq 1 \\ \mathbf{h} \in C \\ Fy \in D \end{cases} \tag{5}$$

where $\alpha \in [0, 1]$, $F \in \mathbb{R}^{K \times N}$, $K \geq N$, models a (non necessarily tight) frame operator [16], and C and D are nonempty closed convex constraint sets that are defined hereafter.

3.3. Constraint set definitions

We discuss here the different choices that can be adopted for the potential functions as well as the convex sets C and D. These choices reflect some a priori knowledge one may have on the variables to be estimated. The idea in addressing a constrained formulation instead of a regularized formulation is to avoid or, at least to facilitate, hyperparameter estimation. This is detailed later on, in the simulation part.

3.3.1. Coupling constraint: function ψ

The seismic noise b is naturally assumed to be additive white Gaussian with zero-mean and variance σ^2. A natural choice for ψ is thus to take $\psi = \| . \|^2/(N\sigma^2)$. When the noise variance is unknown, it can be easily and accurately estimated by using classical techniques such as the median absolute deviation (MAD) [17] wavelet estimator [18, p. 446].

3.3.2. Hard constraints on filters $\mathbf{h}$: convex set C

As mentioned earlier, the filters are assumed to be time-varying. However, real case study showed that those filters have smooth variations along time index n. To ensure that this a priori characteristic is satisfied for the estimated filters, we propose to introduce the following upper bound on the impulse response variations [15]:

$$\forall (j,p,n), \qquad |h_j^{(n+1)}(p) - h_j^{(n)}(p)| \le \varepsilon_{j,p} \qquad (6)$$

where $\varepsilon_{j,p} \in \,]0, +\infty[$.

3.3.3. Hard constraints on primaries y: convex set D

First of all, the primary signal y is assumed to be well represented onto a wavelet frame [16], whose analysis operator is $F \in \mathbb{R}^{K \times N}$. To further account for the wavelet analysis frame coefficient properties, we propose to split the convex set D as $D_1 \times \cdots \times D_{\mathcal{L}}$. Indeed, the idea here is to construct $\mathcal{L}$ partitions of $\{1, \ldots, K\}$ denoted by $\{\mathbb{K}_\ell \mid \ell \in \{1, \ldots, \mathcal{L}\}\}$ where $\mathcal{L}$ corresponds to the number of subbands and $\mathbb{K}_\ell$ is the ℓ-th subband. In this work, we investigate two kinds of convex sets $(D_\ell)_{\ell \in \{1, \ldots, \mathcal{L}\}}$:

1. The first one is widely used in the literature and consists of defining sets of the form: for every $\ell \in \{1, \ldots, \mathcal{L}\}$, $D_\ell = \{(x_k)_{k \in \mathbb{K}_\ell} \mid \sum_{k \in \mathbb{K}_\ell} \phi_\ell(x_k) \le \eta_\ell\}$, where $\eta_\ell \in \mathbb{R}$, and $\phi_\ell : \mathbb{R}^{|\mathbb{K}_\ell|} \to \,]-\infty, +\infty]$ is a proper lower-semicontinuous convex function. For example, this constraint set definition enables to incorporate sparsity constraints on the wavelet frame coefficients in the optimization problem, by choosing e.g. $\phi_\ell = |\cdot|$.

2. The second one is more original and consists of defining hyperplanes: for every $\ell \in \{1, \ldots, \mathcal{L}\}$, $D_\ell = \{(x_k)_{k \in \mathbb{K}_\ell} \mid \sum_{k \in \mathbb{K}_\ell} \phi_\ell((FLz)_k)x_k = \eta_\ell\}$, where $L \in$

$\mathbb{R}^{N \times N}$ is an appropriate linear operator and $\phi_\ell : \mathbb{R} \to \mathbb{R}$. The simplest choice for L is to take the identity operator $L = \mathrm{I}$. An alternative choice, which is reminiscent of Wiener filtering, is

$$L = \lambda_1 \,\mathrm{Diag} \left((1 + \lambda_1 + \lambda_2 \|\mathbf{R}^{(0)}\|^2)^{-1}, \ldots, \right.$$
$$\left. (1 + \lambda_1 + \lambda_2 \|\mathbf{R}^{(N-1)}\|^2)^{-1} \right) \quad (7)$$

where $(\lambda_1, \lambda_2) \in \,]0, +\infty[^2$ and for every $n \in \{0, \ldots, N-1\}$, $\mathbf{R}^{(n)}$ denotes the n-th row of matrix $\mathbf{R}$.

3.4. Optimization strategy

One can note that Problem (5) can be reexpressed as

$$\underset{y \in \mathbb{R}^N, \, \mathbf{h} \in \mathbb{R}^Q}{\mathrm{minimize}} f(y, \mathbf{h}) + \iota_S \left(z - [\mathrm{I} \ \ \mathbf{R}] \begin{bmatrix} y \\ \mathbf{h} \end{bmatrix} \right) + \iota_C(\mathbf{h}) + \iota_D(Fy)$$
$$(8)$$

where $f : \mathbb{R}^N \times \mathbb{R}^Q \to \,]-\infty, +\infty] : (u, v) \mapsto \alpha\rho(v) + (1 - \alpha)\varphi(u)$, $S = \{w \in \mathbb{R}^N \mid \|w\|^2 \le N\sigma^2\}$ and ι_S is the indicator function of the set S defined as

$$\iota_S(u) = \begin{cases} 0 & \text{if } u \in S \\ +\infty & \text{otherwise} \end{cases} \qquad (9)$$

(a similar notation being used for C and D). Such convex optimization problems, involving the sum of 4 convex functions and various linear operators, can be solved in an efficient manner by using primal-dual approaches such as the Monotone+Lipschitz Forward-Backward-Forward (M+L FBF) algorithm [19] as well as the algorithm in [20], which was recently extended in [21, 22]. The functional to be minimized being composed of convex functions as well as indicator functions of convex sets, the algorithm typically requires to compute, in parallel, proximal operators and projections onto the different closed convex sets. Concerning proximity operators, closed-form expressions for a wide class of convex functions can be found in [23]. The projection onto C is explicit and reduces to projections onto hyperslabs (after appropriate splitting). Similarly, when considering affine constraint for convex set D (second case) the projection is explicit. For all the remaining cases, projections onto ℓ_p-ball are performed, some of which can be computed explicitly (e.g. ℓ_2-ball or ℓ_∞-ball) or iteratively (e.g. ℓ_1-ball [24]).

4. RESULTS

Simulation tests are performed on synthetic seismic data. From realistic primary signal $\overline{y}$ and templates $\mathbf{R}$ where $J = 2$, we generated observations according to model (4) where appropriate time-varying filters $\overline{\mathbf{h}}$ are used with $P_1 = 6$ and $P_2 = 6$. The primary signal as well as the observations with $\sigma = 0.01$ are represented Fig. 2. The criterion to be minimized is defined by (5) where φ is chosen to be the ℓ_1-norm, and ρ
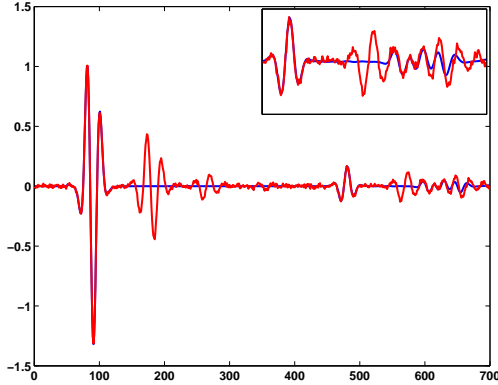
Fig. 2. Observed signal z (red; $\sigma = 0.01$), original y (blue).
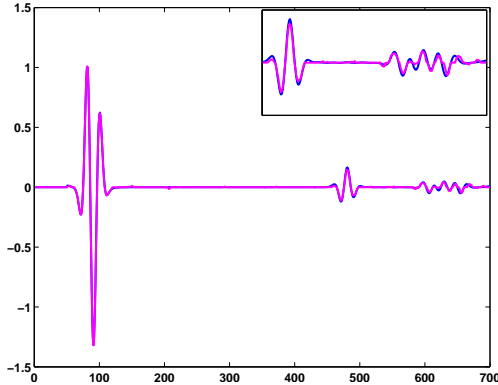

Fig. 3. Estimated signal $\hat{y}$ (magenta), original signal y (blue). D is the intersection of two hyperplanes defined from the identity and the sign functions.

Fig. 4. Estimated multiples $\hat{s}$ (magenta), original multiples s (blue). D is the intersection of two hyperplanes defined from the identity and the sign functions.

σ	0.01			0.04		
ϕ	α	SNR_y	SNR_s	α	SNR_y	SNR_s
0	0.4	23.98	15.79	0.9	15.03	9.60
ℓ_1	0.4	25.98	16.16	0.9	18.19	6.61
ℓ_2	0.6	25.59	16.02	0.8	17.84	9.20
ℓ_∞	0.6	24.48	15.81	0.8	16.24	8.69
I	0.4	26.19	15.81	0.2	**19.74**	8.84
sign	0.3	24.43	14.73	0.1	14.75	4.58
I+sign	0.3	**26.40**	15.56	0.1	18.43	5.94

Table 1. SNR for the estimations of y and s (SNR_y and SNR_s, resp.) in dB considering different convex constraint sets D and two noise levels. Upper table part: "classical constraints" and lower table part: hyperplane contraints.

is the squared ℓ_2-norm (see [7] for more extensive tests on different choices for ρ). Concerning the constraint set definitions, on the one hand, C is defined by (6) where, for every p, $\varepsilon_{1,p} = \varepsilon_{2,p} = 0.17$. On the other hand, D is defined by choosing F to be a (non tight) undecimated wavelet frame with Daubechies wavelets of length 8 and 4 resolution levels. We have considered the two possibilities described in Section 3.3.3 where, in the first case (inequality constraint), $\phi_\ell \equiv \phi$ where ϕ is either the ℓ_1, ℓ_2 or the ℓ_∞-norm. In the second case, $\phi_\ell \equiv \phi$ where ϕ is either the identity or the sign function; furthermore, L is chosen according to (7) where $\lambda_1 = 0.02$ and $\lambda_2 = 0.001$. In this last case, both affine constraints have also been considered jointly (intersection of the two constraint sets).

Restoration results, using M+L FBF algorithm, for the primary signal in the case when $\sigma = 0.01$, are displayed in Fig. 3. The associated estimated multiples are plotted in Fig. 4. From these two figures, one can note that the multiples are quite well estimated and adequately separated from the primary. The stochastic part is accurately removed, even if some residual noise remains, for instance when the signal is of small amplitude. Table 1 shows the signal-to-noise ratios obtained for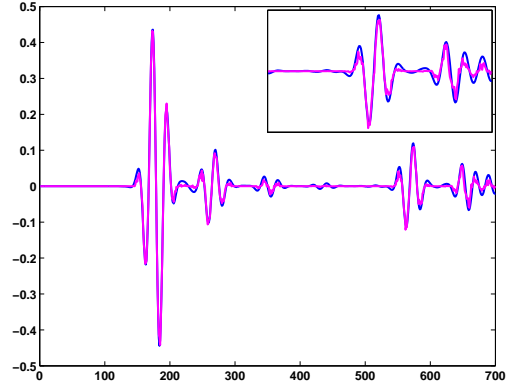 the estimation of y and s. Simulations have been run for different convex sets D and for two noise levels (with standard-deviation $\sigma = 0.01$ and $\sigma = 0.04$). The notation $\phi = 0$ has been used in the case when no constraint is applied to Fy. This allows us to evaluate the gain (up to 1.4 dB) brought by the introduction of prior information on Fy through a constrained formulation.

5. CONCLUSIONS

This paper focuses on the constrained convex formulation of adaptive multiple removal. The proposed approach, based on proximal methods, is quite flexible and allows us to integrate a large panel of hard constraints corresponding to a priori knowledge on the data to be estimated (i.e. primary signal and time-varying filters). A key observation is that some of the related constraint sets can be expressed through hyperplanes, which are not only more convenient to design, but also easier to implement through straightforward projections. Since sparsifying transforms and constraints strongly interact [7], we now study the class of hyperplane constraints of interest as well as their inner parameters, together with the extension to higher dimensions.

6. REFERENCES

[1] S. Liu, "Progress report on SRME with an iterative adaptive subtraction filter," 2009, UTAM Annual Report.

[2] 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*, vol. 77, no. 6, pp. V183–V192, Nov.-Dec. 2012.

[3] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, vol. 38, no. 5, pp. 826–844, Oct. 1973.

[4] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory—part I: Complex signal and scattering in multilayered media," *Geophysics*, vol. 47, no. 2, pp. 203–221, 1982.

[5] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory—part II: Sampling theory and complex waves," *Geophysics*, vol. 47, no. 2, pp. 222–236, 1982.

[6] R. Essenreiter, M. Karrenbach, and S. Treitel, "Multiple reflection attenuation in seismic data using backpropagation," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 2001–2011, Jul. 1998.

[7] M. Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet, "A primal-dual proximal algorithm for sparse template-based adaptive filtering: Application to seismic multiple removal," *IEEE Trans. Signal Process.*, 2014, http://tinyurl.com/proximal-multiple.

[8] J. Song, E. Verschuur, and X. Chen, "Comparing three feedback internal multiple elimination methods," *J. Appl. Geophys.*, vol. 95, no. 0, pp. 66–76, 2013.

[9] 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.*, vol. 34, no. 7, pp. 1038–1066, 1986.

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

[11] R. Neelamani, A. Baumstein, and W. S. Ross, "Adaptive subtraction using complex-valued curvelet transforms," *Geophysics*, vol. 75, no. 4, pp. V51–V60, 2010.

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

[13] D. Donno, "Improving multiple removal using least-squares dip filters and independent component analysis," *Geophysics*, vol. 76, no. 5, pp. V91–V104, 2011.

[14] 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.*, Bucharest, Romania, Aug. 2012, pp. 1494–1498.

[15] 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.*, Vancouver, Canada, May 2013, pp. 2257–2261.

[16] L. Jacques, L. Duval, C. Chaux, and G. Peyré, "A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity," *Signal Process.*, vol. 91, no. 12, pp. 2699–2730, Dec. 2011.

[17] F. R. Hampel, "The influence curve and its role in robust estimation," *J. Am. Stat. Assoc.*, vol. 69, pp. 383–393, 1974.

[18] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[19] 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.*, vol. 20, no. 2, pp. 307–330, Jun. 2012.

[20] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[21] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, 2013.

[22] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Aug. 2013.

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

[24] M. P. Friedlander, "Sparse optimization: applications and algorithms," Proc. SIAM Optimization, May 16–19, 2011, Tutorial.