GEOPHYSICS, VOL. 81, NO. 1 (JANUARY-FEBRUARY 2016); P. V43–V54, 11 FIGS., 1 TABLE.
10.1190/GEO2015-0182.1

Adaptive multiple subtraction: Unification and comparison of matching filters based on the ℓ_q-norm and statistical independence

Yves-Marie Batany[1], Leonardo Tomazeli Duarte[2], Daniela Donno[3], João Marcos Travassos Romano[4], and Hervé Chauris[3]

ABSTRACT

An adaptive multiple subtraction step is necessary for almost all methods that predict seismic multiple reflected waves. We aim at giving a better understanding of matching filters based on ℓ_q-norms and on statistical independence. We found that the formulation of all of these techniques can be gathered in a mutual framework by introducing a space-time operator, called the *primary enhancer*, acting on the estimated primaries. The differences between the considered matching filters become more intuitive because this operator behaves as a simple amplitude compressor. In this perspective, all the methods tend to uncorrelate the predicted multiples and the enhanced estimated primaries. The study of these matching-filter methods can be narrowed to the study of the primary enhancer operator because it is the only difference. Moreover, we have emphasized the role of using adjacent traces or windowing approaches in terms of statistics, and we show that an adequate windowing strategy may overbear the choice of the objective function. Indeed, our analysis showed that setting a good windowing strategy may be more important than changing the classical least-squares adaptation criterion to other approaches based on ℓ_q-norm minimization or independent component analysis.

INTRODUCTION

Multiple attenuation is crucial for improving the quality of seismic images, especially in marine acquisitions. Several techniques exist to provide a prediction of these multiples such as wavefield extrapolation (Wiggins, 1988, 1999) or surface-related multiple elimination (SRME) (Verschuur et al., 1992).

Unfortunately, none of these prediction-based methods can provide a perfect prediction of the multiples because of phase, wavelet, or space-shift errors (Abma et al., 2005). Therefore, a second step, usually referred to as *adaptive multiple subtraction*, is required to accommodate the prediction to the actual multiples before the subtraction. The most common solutions are based on matching-filter approaches (Verschuur and Berkhout, 1997; Rickett et al., 2001; Guitton and Verschuur, 2004) and on prediction-error filters either in the frequency domain (Spitz, 1999) or in the time domain (Guitton, 2005).

Most adaptive multiple subtraction schemes rely on a linear convolutive model to reshape the predicted multiples. However, they may differ in the following aspects:

- the objective function to be optimized
- the domain to perform the optimization
- the strategy to overcome the nonstationarity of the filter
- the strategy to exploit the space-time coherence of the seismic signal.

Often, nonstationarity and space-time coherence are handled with a common strategy. However, it is important to keep in mind that nonstationarity is a difficulty to overcome, whereas space-time coherence is an asset to capitalize on.

Because of its computational efficiency, the ℓ_2-norm is the most commonly used objective function in the adaptive multiple subtraction. The resulting filter, which is known as a *least-squares* or *Wiener filter*, works under the assumption that primaries and multiples are orthogonal in the considered domain (Verschuur, 2006).

Manuscript received by the Editor 17 March 2015; revised manuscript received 7 July 2015; published online 17 December 2015.
[1]PSL Research University, MINES ParisTech, Centre de Géosciences, Fontainebleau, France and University of Campinas (UNICAMP), School of Electrical and Computing Engineering, Campinas, São Paulo, Brazil. E-mail: yves-marie.batany@mines-paristech.fr.
[2]UNICAMP, School of Applied Sciences, Limeira, São Paulo, Brazil. E-mail: leonardo.duarte@fca.unicamp.br.
[3]PSL Research University, MINES ParisTech, Centre de Géosciences, Fontainebleau, France. E-mail: daniela.donno@mines-paristech.fr; herve.chauris@mines-paristech.fr.
[4]University of Campinas (UNICAMP), School of Electrical and Computing Engineering, Campinas, São Paulo, Brazil. E-mail: romano@decom.fee.unicamp.br.

However, in practice, and in particular in the time-offset domain, this assumption fails and it may lead to an overattenuation of the estimated primaries. For this reason, some works consider some ℓ_1-norm-based filters that seem to overcome the problem by promoting a sparser solution of the estimated primaries (Guitton and Verschuur, 2004). Interestingly, ℓ_q-norms have been considered as a regularization term (Costagliola et al., 2011). Moreover, a Bayesian framework is also investigated by Saab et al. (2007).

More recently, other works propose to use independent component analysis (ICA) (Comon and Jutten, 2010) to separate primaries and multiples. This approach has led to the use of new objective functions associated with methods such as geometric-based ICA (Lu, 2006), FastICA (Kaplan and Innanen, 2008), kurtosis-based methods (Donno, 2011), InfoMax (Liu and Dragoset, 2013), and negentropy maximization (Li and Lu, 2013). The first works (Lu, 2006; Kaplan and Innanen, 2008; Donno, 2011) on ICA-based adaptive multiple subtraction operate in a two-step fashion. They comprise an estimation of the shape of the filter using a classic ℓ_2-norm matching filter or a histogram method to correct for time delay, followed by a more precise adjustment of its amplitude using ICA. More recent works (Liu and Dragoset, 2013; Li and Lu, 2013) propose to directly rely on a convolutive modeling with objective functions based on statistical independence. These last solutions are also proposed for 3D multiple elimination (Li and Lu, 2013).

The domain in which the matching filter is performed is decisive in adaptive multiple subtraction, and a lot of effort has been done to search for domains where primaries and multiples do not overlap. Usually, the adaptive multiple subtraction procedure is carried out in the time-offset domain in which the orthogonal assumption fails. Other domains have been proposed such as the dip-domain (Donno, 2011), wavelet-domain (Ventosa et al., 2012), curvelet-domain (Herrmann et al., 2007, 2008; Donno et al., 2010), Radon domain (Li and Lu, 2014), frequency domain (Spitz, 1999), and adjoint domains (the first derivative along with the Hilbert transform and its first derivative) (Wang, 2003). We only consider the space-time domain, but our conclusions hold for other domains.

After defining a proper objective function and a suitable domain to perform the adaptive multiple subtraction procedure, the last issue to overcome is the nonstationarity of the primaries and the multiples (Guitton, 2005; Fomel, 2009). This means that the statistical features of the data are not steady with respect to the time or the offset, and so neither is the filter we aim to recover. However, the spatial and temporal coherence of the seismic signal prevents from drastic changes, and smooth variations can be assumed. Hence, most of the time, the signal is considered as stationary in a small data window in which a unique filter can be obtained. This operation is then repeated on several overlapping windows to complete the full data length. Finally, we must mention that 1D, 2D, or 3D data windows, and so filters, can be considered (Wang, 2003; Donno, 2011) because the seismic signal is coherent in the full data cube.

We aim to give a better understanding of the matching filters based on ℓ_q-norms and ICA in the context of adaptive multiple subtraction. In particular, we focus on the most common ℓ_1 and ℓ_2-norms and two other ICA-based methods based on independence, namely, information maximization (Liu and Dragoset, 2013) and negentropy maximization (Li and Lu, 2013). By introducing a new function called the *primary enhancer*, we show that all these methods share strong similarities. Basically, they all uncorrelate the estimated multiples with the primaries enhanced by the operator. The differences between the investigated techniques can be explained by introducing different primary enhancer operators.

This article is organized as follows: In the next section, we present the adaptive multiple subtraction problem in formal terms and present different ℓ_q-norm optimization problems. The section "Blind source separation" (BSS) is dedicated to the presentation of BSS problems and ICA in which the matching filters based on statistical independence will be briefly presented. The section "Equivalences and similarities between matching-filter methods" contains our main contribution. It presents some similarities and equivalences between the different matching filters. In this article, an equivalence between two methods means that the objective functions to be optimized are mathematically the same. Then, the section "Comparison of methods" is devoted to providing a statistical description of windowing techniques and then to make a comparison between the methods on 2D real seismic data.

OBJECTIVE FUNCTIONS USED FOR ADAPTIVE MULTIPLES SUBTRACTION

Marine seismic data $\boldsymbol{d}[t,h,s]$, where t, h, and s denote, respectively, the time, offset, and shot position, are corrupted by multiple reflections $\boldsymbol{m}[t,h,s]$ due to the water-free surface and internal layers. Most of the time, these multiples are considered as noise to be removed before the imaging process, and one would like to obtain only primary events $\boldsymbol{p}[t,h,s]$ (Verschuur, 2006). Among the existing techniques providing a prediction $\breve{\boldsymbol{m}}[t,h,s]$ of these multiples (Wiggins, 1988, 1999; Verschuur et al., 1992), none of them are perfect and time shifts or wavelet differences need to be adjusted. To do so, one searches for a filter $\boldsymbol{w}[t,h,s]$ of size $K_t \times K_h \times K_s$ to better match the prediction with the observed data $\boldsymbol{d}[t,h,s]$. Unless otherwise stated, an italic bold lowercase letter $\boldsymbol{d}$ indicates a multidimensional data set, and an italic lowercase letter d indicates an element of it. If subscripts appear, they indicate the dimension of the considered data set. For instance d_t represents an element of a 1D data set and d_{th} an element of a 2D data set.

We consider the following linear model:

$$\begin{cases} \boldsymbol{d} = \boldsymbol{p} + \boldsymbol{m} \\ \hat{\boldsymbol{m}} = \boldsymbol{w} * \breve{\boldsymbol{m}} \end{cases}, \tag{1}$$

where a breve $\breve{\cdot}$ indicates a prediction, a hat $\hat{\cdot}$ indicates a final estimation, and $*$ denotes the convolution product that can either be 1D, 2D, or 3D, according to the dimension of $\boldsymbol{w}$ and $\breve{\boldsymbol{m}}$. The estimate of the primaries is then given by

$$\hat{\boldsymbol{p}} = \boldsymbol{d} - \hat{\boldsymbol{m}} = \boldsymbol{d} - \boldsymbol{w} * \breve{\boldsymbol{m}}. \tag{2}$$

To find a filter, we need to formulate an optimization problem of the form

$$\text{find } \boldsymbol{w} \quad \text{such that } \phi(\boldsymbol{w}) \text{ is minimum}, \tag{3}$$

where $\phi(\boldsymbol{w})$ is an objective function to be defined. The most common objective functions are based on the ℓ_q-norm, $q \geq 1$, which is defined for a vector $\boldsymbol{x}$ as

$$\|\boldsymbol{x}\|_{\ell_q} = \sqrt[q]{\sum_i |x_i|^q}. \tag{4}$$

From that, the ℓ_2-norm refers to the classical Euclidean distance, whereas the ℓ_1-norm is simply the sum of the absolute values of the vector components. In multiple subtraction, the most commonly adopted approach is based on the ℓ_2-norm such as

$$\phi_{\ell_2} = \|\boldsymbol{d} - \boldsymbol{w} * \check{\boldsymbol{m}}\|_{\ell_2}. \tag{5}$$

This norm is quite convenient from a mathematical point of view because it admits an analytical solution when a linear model is considered (Haykin, 2001). The objective function in equation 5 can be written in matrix form as

$$\phi_{\ell_2} = \|\boldsymbol{d} - \check{\boldsymbol{M}}\boldsymbol{w}\|_{\ell_2}, \tag{6}$$

where $\check{\boldsymbol{M}}$ and $\boldsymbol{w}$ are, respectively, a matrix and a vector constructed such that $\check{\boldsymbol{M}}\boldsymbol{w} = \boldsymbol{w} * \check{\boldsymbol{m}}$. This construction can be done for the 1D, 2D, or 3D convolutional products according to the dimension of $\boldsymbol{w}$ and $\check{\boldsymbol{m}}$. The damped least-squares solution gives

$$\boldsymbol{w}_{\ell_2} = (\check{\boldsymbol{M}}^T\check{\boldsymbol{M}} + \zeta\boldsymbol{I})^{-1}\check{\boldsymbol{M}}^T\boldsymbol{d}, \tag{7}$$

where the term $\zeta\boldsymbol{I}$ regularizes the inversion of $\check{\boldsymbol{M}}^T\check{\boldsymbol{M}}$ if necessary.

However, the ℓ_2-norm filter, also known as a *Wiener* or *least-squares filter*, may lead to overattenuation issues when primaries and multiples overlap. Guitton and Verschuur (2004) analyze the use of the ℓ_1-norm objective function:

$$\phi_{\ell_1} = \|\boldsymbol{d} - \boldsymbol{w} * \check{\boldsymbol{m}}\|_{\ell_1}, \tag{8}$$

and they have shown that it may lead to a sparser estimate of the primaries. Unfortunately, a direct analytic solution does not exist for the ℓ_1-norm. Guitton and Verschuur (2004) propose to use the iterative reweighted least-squares (IRLS) algorithm to approximate the ℓ_1-norm solution by using the objective function:

$$\phi_{\ell_{1/2}} = \|\boldsymbol{F}(\boldsymbol{d} - \boldsymbol{w} * \check{\boldsymbol{m}})\|_{\ell_2}, \tag{9}$$

where $\boldsymbol{F}$ is a diagonal matrix depending on the estimated primaries $\hat{\boldsymbol{p}}$ and iteratively updated with the least-squares solution given by equation 7. By using a specific $\boldsymbol{F}$, they have shown that their method is equivalent to consider the following objective function:

$$\phi_{\ell_{1/2}} = \mathbb{E}\left\{\sqrt{1 + (\hat{p}/\epsilon)^2} - 1\right\}, \tag{10}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator and ϵ is a positive constant. Their analysis suggests to use a constant $\epsilon = \max|\boldsymbol{d}|/100$.

More generally, it is also possible to envisage formulations based on the minimization of a ℓ_q-norm objective function:

$$\phi_{\ell_q} = \|\boldsymbol{d} - \boldsymbol{w} * \check{\boldsymbol{m}}\|_{\ell_q}, \tag{11}$$

with $q \geq 1$. This formulation is adopted, for instance, by Costagliola et al. (2011) for regularization purposes. Pham et al. (2014) also propose a more general framework able to introduce different kinds of norms for the estimated primaries and the filter.

More recently, some authors consider that primaries and multiples can be modeled as statistical independent variables (Kaplan and Innanen, 2008; Donno, 2011; Li and Lu, 2013). As in equation 3, we can write the optimization problem as

$$\text{find } \boldsymbol{w} \quad \text{such that } \hat{\boldsymbol{p}} \text{ and } \check{\boldsymbol{m}} \text{ are independent.} \tag{12}$$

Once again, no analytic solution exists for this problem. However, ICA has emerged as a powerful framework for specifically tackling this problem, and many algorithms can be found (Hyvärinen et al., 2001; Comon and Jutten, 2010). Although all of these methods have in common to try to solve the problem in equation 12, they differ by their objective function. For instance, Donno (2011) uses a kurtosis-based function, Liu and Dragoset (2013) use an information maximization (InfoMax) objective function, and Li and Lu (2013) use a negentropy maximization objective function.

In the section "Equivalences and similarities between matching-filter methods", we will show that for the convolutive model of equation 1, the optimization problems using the ℓ_q-norm or the statistical independence are actually very similar and, therefore, are expected to lead to similar practical results. Before presenting these equivalences, we shall, in the next section, briefly review the BSS problem, the ICA framework for solving BSS, and two ICA-based algorithms, known as information maximization (Liu and Dragoset, 2013) and negentropy maximization (Li and Lu, 2013), that have been used for multiple subtraction.

BLIND SOURCE SEPARATION

Formulation of the blind source separation problem

BSS is a problem in which one aims to recover a set of N sources $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$ from a set of M observed mixtures $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ (Comon and Jutten, 2010). If the mixing process is modeled as a linear and instantaneous system, the observed mixture can be written as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{13}$$

where the matrices $\boldsymbol{S}$ and $\boldsymbol{X}$ contain, respectively, the source and mixed signals, and $\boldsymbol{A}$ is the $M \times N$ mixing matrix. The term *blind* comes from the fact that we assume no information about the mixing process $\boldsymbol{A}$ nor training samples performing supervised learning.

For the determined case, in which $M = N$, the separation can be achieved by finding a demixing matrix $\boldsymbol{W}$ such that, once applied to the observations, it gives an estimate $\hat{\boldsymbol{S}}$ of the sources as

$$\hat{\boldsymbol{S}} = \boldsymbol{W}\boldsymbol{X} = \boldsymbol{W}\boldsymbol{A}\boldsymbol{S} = \boldsymbol{G}\boldsymbol{S}, \tag{14}$$

where $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A}$ represents the global mapping between the true sources $\boldsymbol{S}$ and their estimates $\hat{\boldsymbol{S}}$. It is proved that we can only recover sources up to a scale and a permutation ambiguity (Comon, 1994).

For a convolutive mixing model, equation 13 must be transformed into

$$\boldsymbol{X} = \sum_{l=0}^{L-1} \boldsymbol{A}_l \mathcal{T}_l\{\boldsymbol{S}\}, \tag{15}$$

where L is the size of the filters acting on the sources and $\mathcal{T}_l$ is a time-shifting operator acting on $\boldsymbol{S}$. Analogously to the instantaneous case, separation in the convolutive case can be performed by adjusting a set of separating filters of length K such that the set of retrieved sources is given by

$$\hat{S} = \sum_{k=0}^{K-1} W_k \mathcal{T}_k\{X\}. \tag{16}$$

All the models and strategies discussed in this section for 1D sources can be directly extended to two or three dimensions with filters of the corresponding size.

Solving the blind source separation problem using independent component analysis

Clearly, a BSS problem is ill posed and some a priori information on the desired sources must be added. Historically, the first idea was to consider the assumption of statistically independent sources, which led to ICA (Comon, 1994). However, other information such as sparsity can be considered to achieve the separation (Bofill and Zibulevsky, 2001).

Two sources s_1 and s_2 are said to be statistically independent if their joint probability distribution function (PDF) g'_{s_1,s_2} is the product of their marginal PDF such as

$$g'_{s_1,s_2} = g'_{s_1} \times g'_{s_2}, \tag{17}$$

where the notation $g'(\cdot)$ for the PDF is chosen to make clearer the notations used in the following sections. Therefore, ICA algorithms search for a unique demixing matrix W, in the instantaneous case, or a set of demixing filters W_k, in the convolutive case, that provide independent estimates of the signals to be recovered.

In practice, ICA can be conducted by different approaches, which mainly differ on the choice of the statistical function used to express the independence. Interested readers may refer to Hyvärinen et al. (2001) or Comon and Jutten (2010) for a complete overview of ICA methods. In the sequel, we describe two ICA-based methods that have been recently used for adaptive multiple subtraction, namely, InfoMax (Liu and Dragoset, 2013) and negentropy maximization (Li and Lu, 2013). We focus here on their objective functions.

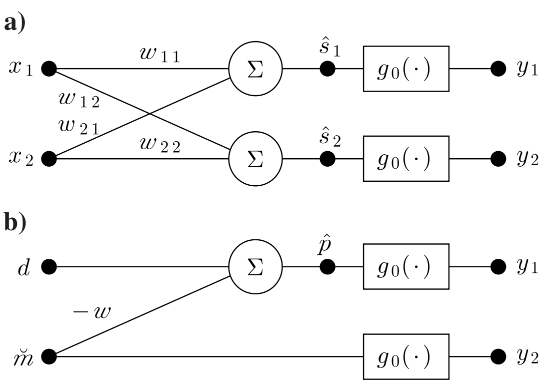

Figure 1. (a) BSS network — neural network of the InfoMax algorithm with two mixtures and two sources, corresponding to the formulation of a BSS problem. (b) Matching filter network — adaptive multiple subtraction (equation 2) described as a neural network. In the image, a line indicates a convolution with the specified filter.

InfoMax

Let us define the mutual information between two random variables x and y as

$$I_{x,y} = H_x + H_y - H_{x,y}, \tag{18}$$

where H_x is the differential entropy of a random variable x following a PDF g'_x and $H_{x,y}$ is the joint entropy of two random variables x and y having a joint PDF $g'_{x,y}$. They are defined as

$$H_x = -\int g'_x \log g'_x dx, \tag{19}$$

and

$$H_{x,y} = -\int g'_{x,y} \log g'_{x,y} dxdy. \tag{20}$$

As proposed by Bell and Sejnowski (1995), maximizing the mutual information $I_{X,Y}$ between the inputs $X = [x_1 \quad x_2]^T$ and the outputs $Y = [y_1 \quad y_2]^T$ of the neural network shown in Figure 1a leads to the recovery of estimated sources $\hat{S} = [\hat{s}_1 \quad \hat{s}_2]^T$ that are statistically independent.

It is important to note that, in the context of BSS, the outputs $y_i = g_0(\hat{s}_i)$ of the neural networks are auxiliary variables used to optimize the statistical independence between the estimated sources $\hat{s}_i$. The functions $g_0(\cdot)$ can be seen as estimates of the cumulative distribution functions (CDF) of the desired sources. Often, logistic functions are used such as the sigmoid function:

$$g_0(s) = \frac{1}{1+\mathrm{e}^{-\lambda s}}, \quad \text{with } g_0{}'(s) = \frac{\lambda \mathrm{e}^{-\lambda s}}{(1+\mathrm{e}^{-\lambda s})^2}, \tag{21}$$

where λ is a shaping parameter. As shown in Figure 2, $g_0{}'(s)$ represents an estimate of the PDF of the desired sources (Cardoso, 1997). Liu and Dragoset (2013) consider the shaping parameter fixed with $\lambda = 1$. In this article, more flexibility is added to take into account a wider range of CDF.

It can be shown (Bell and Sejnowski, 1995) that maximizing the mutual information $I_{X,Y}$ is equivalent to minimizing the following objective function:

$$\phi_{IM} = -\mathbb{E}\{\log(|J_{IM}|)\}, \tag{22}$$

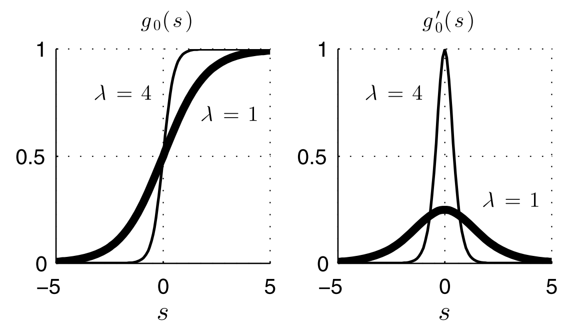

Figure 2. Nonlinear functions $g_0(s)$ are used in an InfoMax network, and their derivatives $g'_0(s)$ with different shaping parameters λ. They respectively represent the estimated CDF and PDF of a desired signal.

where $\log(\cdot)$ denotes the natural logarithm function and J_{IM} is the Jacobian defined for a 2×2 network as

$$J_{IM} = \det \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{bmatrix}. \quad (23)$$

Negentropy maximization

As explained by Comon and Jutten (2010) or Hyvärinen et al. (2001), recovering estimates $\hat{\boldsymbol{S}}$ as far as possible from a Gaussian distribution is a valid method for recovering independent sources. This measure of Gaussianity can be achieved by using the negentropy $Q_{\hat{S}}$ of the estimated sources defined as

$$Q_{\hat{S}} = H_N - H_{\hat{S}}, \quad (24)$$

where H_N is the differential entropy of a random vector following a Gaussian distribution with the same mean vector and covariance matrix as $\hat{\boldsymbol{S}}$.

As proposed by Hyvärinen and Oja (2000), we can approximate the maximization of negentropy by minimizing the following objective function:

$$\phi_Q = -(\mathbb{E}\{g_i(\hat{\boldsymbol{S}}_0)\} - \mathbb{E}\{g_i(\boldsymbol{N}_0)\})^2, \quad (25)$$

where $\hat{\boldsymbol{S}}_0$ has a zero mean vector and unit covariance matrix and $\boldsymbol{N}_0$ is a random standardized Gaussian vector. The function $g_i(\cdot)$ can be chosen within the following set of nonquadratic functions (Li and Lu, 2013):

$$g_1(s) = -\exp\left(-\frac{s^2}{2}\right), \quad (26)$$

$$g_2(s) = \log(\cosh(s)), \quad (27)$$

and

$$g_3(s) = \sqrt{1+s^2} - 1. \quad (28)$$

EQUIVALENCES AND SIMILARITIES BETWEEN MATCHING-FILTER METHODS

In this section, we first show an equivalence between the ℓ_q-norm- and InfoMax-based matching filters, as well as an equivalence between InfoMax- and negentropy-based matching filters. We remind the reader that in this article, an equivalence between two methods means that the objective functions to be optimized are mathematically the same. Then, we introduce an operator that allows showing the similarities between ℓ_2-norm, $\ell_{1/2}$-norm, InfoMax, negentropy, and ℓ_1-norm-based matching filters. In this context, the ℓ_1-norm filter is seen as a limiting case of the others.

Equivalence between InfoMax and ℓ_q-norm

Let us consider the neural network in Figure 1b, which represents the matching filter approach as proposed by the linear convolutive model presented in equations 1 and 2. The difference between this network and the classical formulation of a BSS problem (Figure 1a) is that only one filter $\boldsymbol{w}$ is required to perform the separation for adaptive multiple subtraction. This network could be considered as a special case of the BSS network. Therefore, the objective function of equation 22 also holds for the network of Figure 1b. In this case, the statistical independence is required between the estimated primaries $\hat{\boldsymbol{p}}$ and the predicted multiples $\check{\boldsymbol{m}}$.

For this specific network, the Jacobian given by the equation 23 simplifies and becomes

$$J_{IM} = \frac{\partial y_0}{\partial d}\frac{\partial y_1}{\partial \check{m}} = \frac{\partial y_0}{\partial \hat{p}}\frac{\partial \hat{p}}{\partial d}\frac{\partial y_1}{\partial \check{m}} = g_0'(\hat{p})g_0'(\check{m}). \quad (29)$$

Because the prediction $\check{m}$ does not change in the network, after substituting equation 29 in equation 22, the term $\mathbb{E}\{\log|g_0'(\check{m})|\}$ is constant and the objective function to be minimized becomes

$$\phi_{IM} = -\mathbb{E}\{\log|g_0'(\hat{p})|\}. \quad (30)$$

As discussed in the previous section, the function $g_0(\cdot)$ can be seen as an estimate of the CDF of the desired signals and the function $g_0'(\cdot)$ represents an estimate of their PDF. We can assume that the primaries follow a generalized Gaussian distribution that can either be super- or sub-Gaussian, so that its PDF is given by

$$g_0'(\hat{p}) \propto \exp(-|\hat{p}|^q). \quad (31)$$

The objective function to be minimized in this case becomes

$$\phi_{IM} \propto +\mathbb{E}\{|\hat{p}|^q\}, \quad (32)$$

which is equivalent to the minimization of the ℓ_q-norm of the primaries as in equation 11.

In particular, if we assume that the primaries follow a Laplacian distribution, we can choose $g_0'(\hat{p}) \propto \exp(-|\hat{p}|)$, and so $\phi_{IM} \propto +\mathbb{E}\{|\hat{p}|\}$, which is equivalent to the minimization of the ℓ_1-norm of the primaries as in equation 8. In the same way, if we assume that the primaries follow a Gaussian distribution, we can choose $g_0'(\hat{p}) \propto \exp(-\hat{p}^2)$, and so $\phi_{IM} \propto +\mathbb{E}\{\hat{p}^2\}$, which is equivalent to the minimization of the ℓ_2-norm of the primaries as in equation 5.

Equivalence between InfoMax and negentropy maximization

In the case of adaptive multiple subtraction, Li and Lu (2013) show that the objective function in equation 25 can be written as

$$\phi_Q = +\mathbb{E}\{g_i(\hat{p}/\sigma_{\hat{p}})\}, \quad (33)$$

where $\sigma_{\hat{p}}^2$ is the variance of the estimated primaries. As they point out, the use of the function $g_3(\cdot)$ in adaptive multiple subtraction leads to a formulation identical to the IRLS algorithm described in equation 10. Therefore, in the following, we will focus on the first two nonquadratic functions g_1 and g_2 by keeping in mind that for a zero-mean signal, the normalization by $\sigma_{\hat{p}}$ is equivalent to a normalization by the ℓ_2-norm of the estimated primaries.

In particular, if the nonquadratic function g_2 is used, we can write it as

$$\phi_Q = -\mathbb{E}\left\{\log\frac{1}{\cosh(\hat{p}/\sigma_{\hat{p}})}\right\}, \quad (34)$$

Table 1. Primary enhancer operators.

Matching filter	Primary enhancer operator $\bar{g}(\hat{p})$
ℓ_2-norm	$\propto \hat{p}$
Hybrid ℓ_1/ℓ_2-norm	$\propto \hat{p}(1+\hat{p}^2/\epsilon^2)^{-1/2}$
InfoMax	$\propto -g_0''(\hat{p})/g_0'(\hat{p})$
Negentropy	$\propto g_i'(\hat{p})$
ℓ_1-norm	$\propto \mathrm{sign}(\hat{p})$

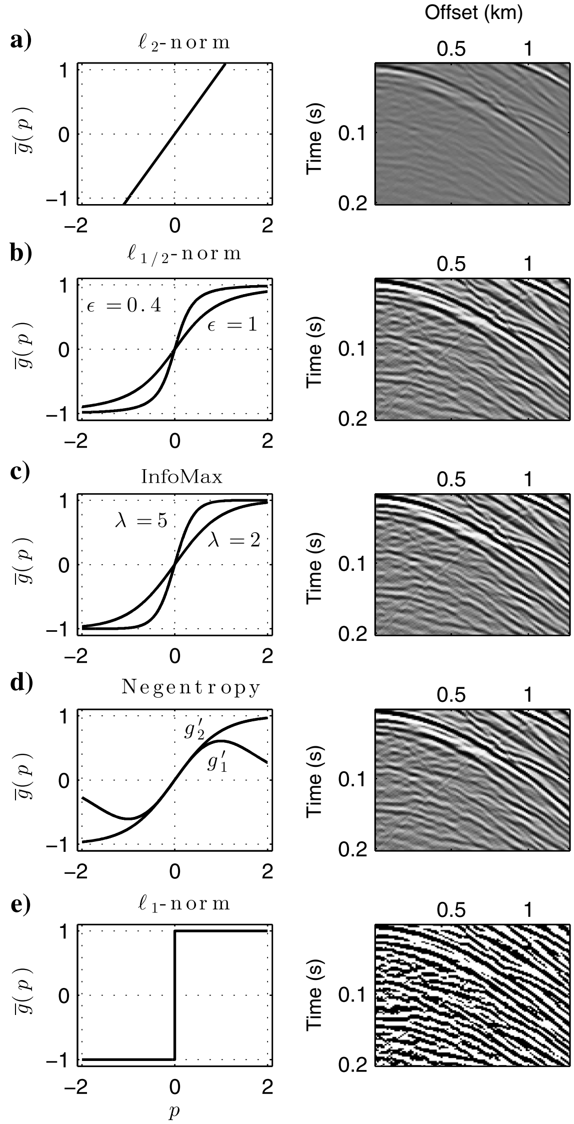

Figure 3. The primary enhancer operators $\bar{g}(\cdot)$, analyzed in equations 39 to 43, are applied on the same small window of seismic data, supposedly the estimated primaries. The mean and the variance of the data have been, respectively, normalized to zero and one. A scaling factor is applied for the operator of the hybrid $\ell_{1/2}$ method and the InfoMax method to bound their value range between -1 and 1.

that is equivalent to the objective function of the InfoMax matching filter in equation 30 with $g_0'(s) = 1/\cosh(s)$. Also, if the nonquadratic function g_1 is used, we have

$$\phi_Q = -\mathbb{E}\left\{\exp\left(-\frac{\hat{p}^2}{2\sigma_{\hat{p}}^2}\right)\right\}, \tag{35}$$

which is equivalent to the objective function of the InfoMax matching filter with $g_0'(s) = \exp\,[\exp\,(-s^2/2)]$. Interestingly, it can also be seen as using the InfoMax objective function after removing the log operator, such that $\phi = -\mathbb{E}\{g_0{}'(\hat{p})\}$ with a Gaussian a priori distribution g_0'.

Uniformization of the objective functions

The first derivative stationary condition applied to a given objective function imposes that we have

$$\nabla\phi = \mathbf{0} \tag{36}$$

at the solution, where $\nabla\phi$ is the gradient of the objective function with respect to each filter coefficient such as

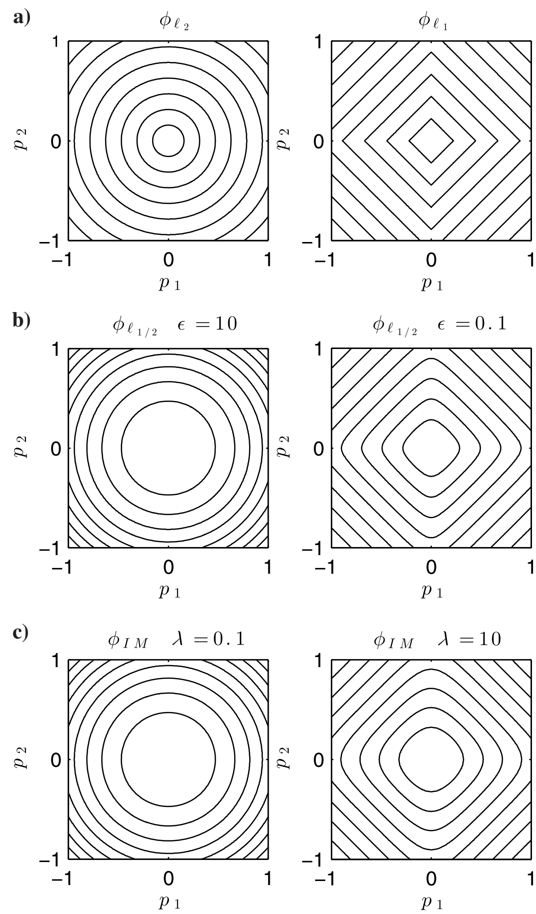

Figure 4. Contour plots in $\mathbb{R}^2$ of: (a) the ℓ_2-norm and the ℓ_1-norm objective functions, (b) the hybrid $\ell_{1/2}$-norm objective function with $\epsilon = 10$ and 0.1, and (c) the InfoMax-based objective function with $\lambda = 0.1$ and 10.

$$\nabla\phi = \left[\frac{\partial\phi}{\partial w_0} \quad , \ldots , \quad \frac{\partial\phi}{\partial w_k} \quad , \ldots , \quad \frac{\partial\phi}{\partial w_{K-1}}\right]^T. \tag{37}$$

Equation 36 is equivalent to a set of K equations,

$$\frac{\partial\phi}{\partial w_k} = \frac{\partial\phi}{\partial\hat{p}_t}\frac{\partial\hat{p}_t}{\partial w_k} = -\frac{\partial\phi}{\partial\hat{p}_t}\check{m}_{t-k} = 0, \quad \text{for } k = 0, \ldots, K-1, \tag{38}$$

from which it is clear that we can restrain our study to the first derivative of the objective function with respect to the estimated primaries. We obtain for all the objective functions considered in this paper (equations 5, 10, 30, 25, and 8, respectively):

$$\frac{\partial\phi_{\ell_2}}{\partial w_k} = -\sum_t \hat{p}_t\check{m}_{t-k} = 0, \tag{39}$$

$$\frac{\partial\phi_{\ell_{1/2}}}{\partial w_k} = -\sum_t \frac{\hat{p}_t}{\epsilon^2}\left(1 + \frac{\hat{p}_t^2}{\epsilon^2}\right)^{-1/2}\check{m}_{t-k} = 0, \tag{40}$$

$$\frac{\partial\phi_{IM}}{\partial w_k} = -\sum_t -\frac{g_0''(\hat{p}_t)}{g_0'(\hat{p}_t)}\check{m}_{t-k} = 0, \tag{41}$$

$$\frac{\partial\phi_Q}{\partial w_k} = -\sum_t g_i{}'(\hat{p}_t)\check{m}_{t-k} = 0, \tag{42}$$

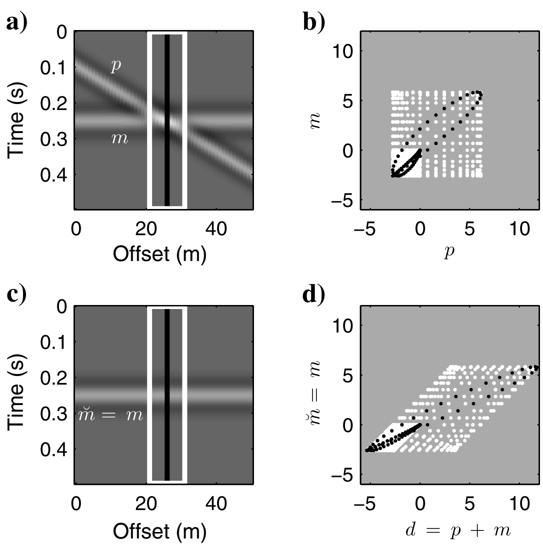

Figure 5. Synthetic toy example of two crossing events. (a) The synthetic data set containing one primary and one multiple that are overlapping at traces 20 to 30. (b) Scatterplot of the primary and the multiple at trace 25 only (in black) and at traces 20 to 30 (in white). (c) The prediction of the multiple that is equal, in this example, to the true multiple. (d) Scatterplot of the data and the prediction of the multiple at trace 25 only (in black) and at traces 20 to 30 (in white).

and

$$\frac{\partial\phi_{\ell_1}}{\partial w_k} = -\sum_t \mathrm{sign}(\hat{p}_t)\check{m}_{t-k} = 0. \tag{43}$$

We propose to unify these conditions by introducing the operator $\bar{g}(\cdot)$, that we call the primary enhancer, which allows writing the first-derivative condition as

$$\nabla\boldsymbol{\phi} = -\bar{\mathrm{g}}(\hat{\boldsymbol{p}})\star\check{\boldsymbol{m}} = \mathbf{0}, \tag{44}$$

where $\star$ denotes the crosscorrelation product. The primary enhancer operators for all of the objective functions analyzed in this paper are presented in Table 1 and are shown in Figure 3.

In particular, if a sigmoid function is used in the InfoMax matching filter, we have $\bar{g}(\hat{p}) = -\lambda(1 - 2g_0(\hat{p}))$. For the negentropy maximization matching filter, we get $\bar{g}(\hat{p}) = \hat{p}\,\exp(-\hat{p}/2)$ if $g_1(\cdot)$ is used and $\bar{g}(\hat{p}) = \tanh(\hat{p})$ if $g_2(\cdot)$ is used.

Figure 3 shows the analyzed primary enhancer operators and their application on a small seismic data window. In this context, the result of the ℓ_1-norm matching filter can be seen as the limit of the hybrid ℓ_1/ℓ_2-norm matching filter when $\epsilon \to 0$ or also as the limit of the InfoMax matching filter when $\lambda \to \infty$. Between those extreme values, the InfoMax and the hybrid ℓ_1/ℓ_2-norm primary enhancer operators share strong similarities because they provide a smooth transition from the ℓ_1-norm to the ℓ_2-norm solution. From our observations, they are the most similar when a relation

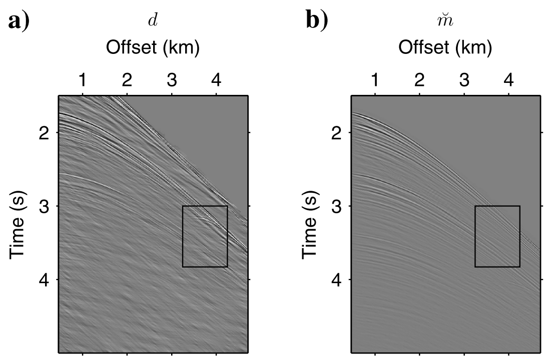

Figure 6. (a) Input common shot gather from real marine data set and (b) SRME-predicted multiples.

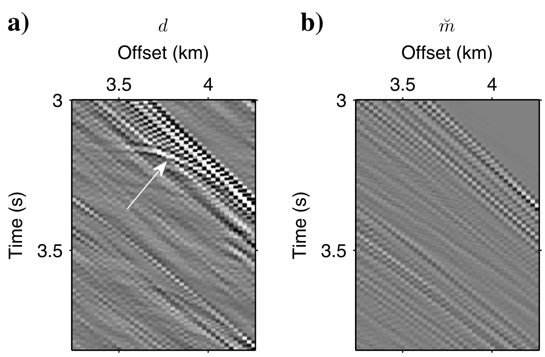

Figure 7. Magnification of the black rectangles of Figure 6. The legends are the same. A primary event is indicated by a white arrow.

$\lambda = 1/\epsilon$ is kept. Figure 4 also shows how the InfoMax and the hybrid $\ell_{1/2}$-norm objective functions make a smooth transition between the ℓ_2-norm and the ℓ_1-norm objective function depending on the value of their shaping parameter.

It is already known that the least-squares filter aims at canceling the crosscorrelation between the estimated primaries and the predicted multiples in a vicinity defined by the dimensions of the filter. Equation 44 unifies the methods analyzed in this paper in a same fashion. They can all be seen as canceling the crosscorrelation between the enhanced primaries and the predicted multiples. As we indicated in the two previous subsections, there exist equivalences between certain methods if specific shaping parameters or nonlinear functions are used. In those cases, the enhanced operators are equal. Otherwise, the methods are similar and their practical differences will be discussed in the next section.

COMPARISON OF METHODS

Considerations about space-time coherence and data windowing

Because the seismic signal is not stationary either in space or in time, windowing strategies are usually used. The space-time coherence of the seismic signal can be used to smooth the variations in the filter and avoid drastic changes in space and time.

Until now, we mainly consider the 1D-1D strategy (1D filter, 1D data window) for which one wants to recover a single 1D matching filter of length K_t for a segment of seismic trace of length F_t. However, this 1D-1D strategy does not avoid a drastic change in space, but only in time. That is the reason why any matching filter based on the ℓ_q-norm or independence applied trace by trace with the 1D-1D strategy may lead to overattenuation of the primaries if they do overlap with multiple events.

To overcome the overattenuation problem, the 1D-2D strategy (1D filter, 2D data window) uses adjacent traces to find a 1D filter. The result of using adjacent traces in terms of statistical diversity is shown in Figure 5 in the case of crossing events. In this toy example, the prediction of the multiples is equal to the true multiples. Figure 5b and 5d, respectively, shows the scatterplot of the primary versus the multiple (Figure 5a) and the data versus the predicted multiple (Figure 5c) at a single offset (in black) and in a small window (in white). We see that if a single trace containing the crossing event (in black) is considered, the primary and the multiple are highly correlated, and so they are highly statistically dependent. Hence, any strategy trying to make them uncorrelated (least-squares) or independent (e.g., InfoMax) will systematically fail. In other words, overattenuation will systematically happen with the 1D-1D strategy. However, when adjacent traces are used (in white), the primaries and the multiples became statistically independent events and a strategy forcing the independence may work. We

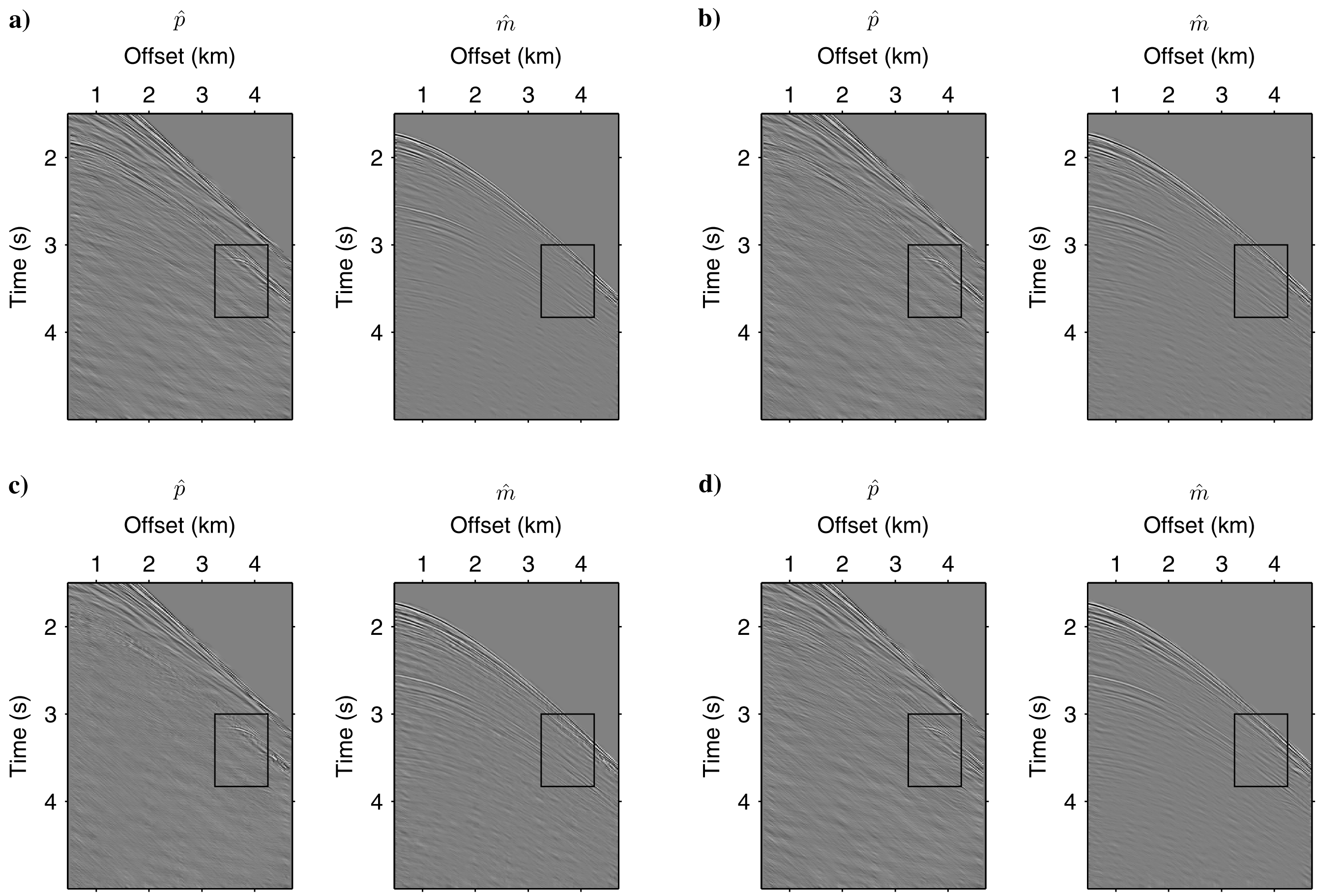

Figure 8. Primaries and multiples estimated by the ℓ_2-norm with a 1D-2D strategy (five adjacent traces) of (a) $F_t = 200$ ms and $K_t = 40$ ms, (b) $F_t = 200$ ms and $K_t = 80$ ms, (c) $F_t = 200$ ms and $K_t = 160$ ms, and (d) $F_t = 400$ ms and $K_t = 160$ ms.

emphasize here again that considering primaries and multiples as independent events does not help if they overlap. It is the use of adjacent traces that help overcoming the overattenuation problem in adaptive multiple subtraction.

The 2D-2D strategy (2D filter, 2D data window) explicitly uses the coherence of the seismic signals in space and time. It is expressed as finding a 2D matching filter of size $K_t \times K_h$, over a data window of size $F_t \times F_h$. In this case, the convolutional product is defined in two dimensions. Most of the time, we have $K_t < F_t$ and $K_h < F_h$ to solve a well-posed problem. A 3D strategy can also be considered using the shot number as the third dimension; the convolutional product will be defined this time on three dimensions.

Results on real data set

We compare the results of matching filter methods on a 2D real marine data set. The common shot gather is presented in Figures 6a and 7a and the 2D SRME prediction in Figures 6b and 7b. The spatial sampling is 25 m, and the time sampling is 2 ms. The minimum and maximum offsets are 225 and 4700 m, respectively. The 2D SRME prediction is realized with a 600-m aperture around the source and receivers to reduce aliasing. Some primary events surrounded by multiple events are clearly identifiable. A global time shift correction of 40 ms is preapplied on the prediction, but no spatial correction is necessary. Hence, the matching filter we are seeking should mainly compensate for the surface operator due to the autoconvolutions of the data during the prediction process. In a first test, we use the ℓ_2-norm matching filter with different windowing strategies. In a second test, we use the same windowing strategy with different objective functions.

The negentropy maximization matching filter is performed with the nonlinear function g_1 and the IRLS algorithm proposed by Li and Lu (2013). The hybrid $\ell_{1/2}$-norm (IRLS) has one parameter ϵ that is estimated for each window by the relation proposed by Guitton and Verschuur (2004). Finally, our formulation of the InfoMax matching filter needs the estimation of the parameter λ that is related to the prior CDF of the primaries we tend to estimate. We propose here to use a fixed value of λ, but an adaptive scheme could also be used to take better into account the nonstationarity of the signal. Because one assumes that the multiples should be removed from the signal, the primaries should have a more spiky PDF compared to the data. First, an optimum parameter λ_d is determined to fit +the data, and then the value for the primaries is overevaluated by $\lambda \approx 5\lambda_d$.

The hybrid $\ell_{1/2}$-norm and negentropy methods are implemented by using the IRLS algorithm. If an identity matrix is chosen for the initialization of the matrix of weights $\boldsymbol{F}$ in equation 9, they have the advantage to give the ℓ_2-norm solution at the first iteration (Guitton and Verschuur, 2004). A gradient method is used for the InfoMax method (Liu and Dragoset, 2013), and it has the advantage of actually computing the nonlinear correlation between the estimated primaries and the predicted multiples for the gradient update rule.

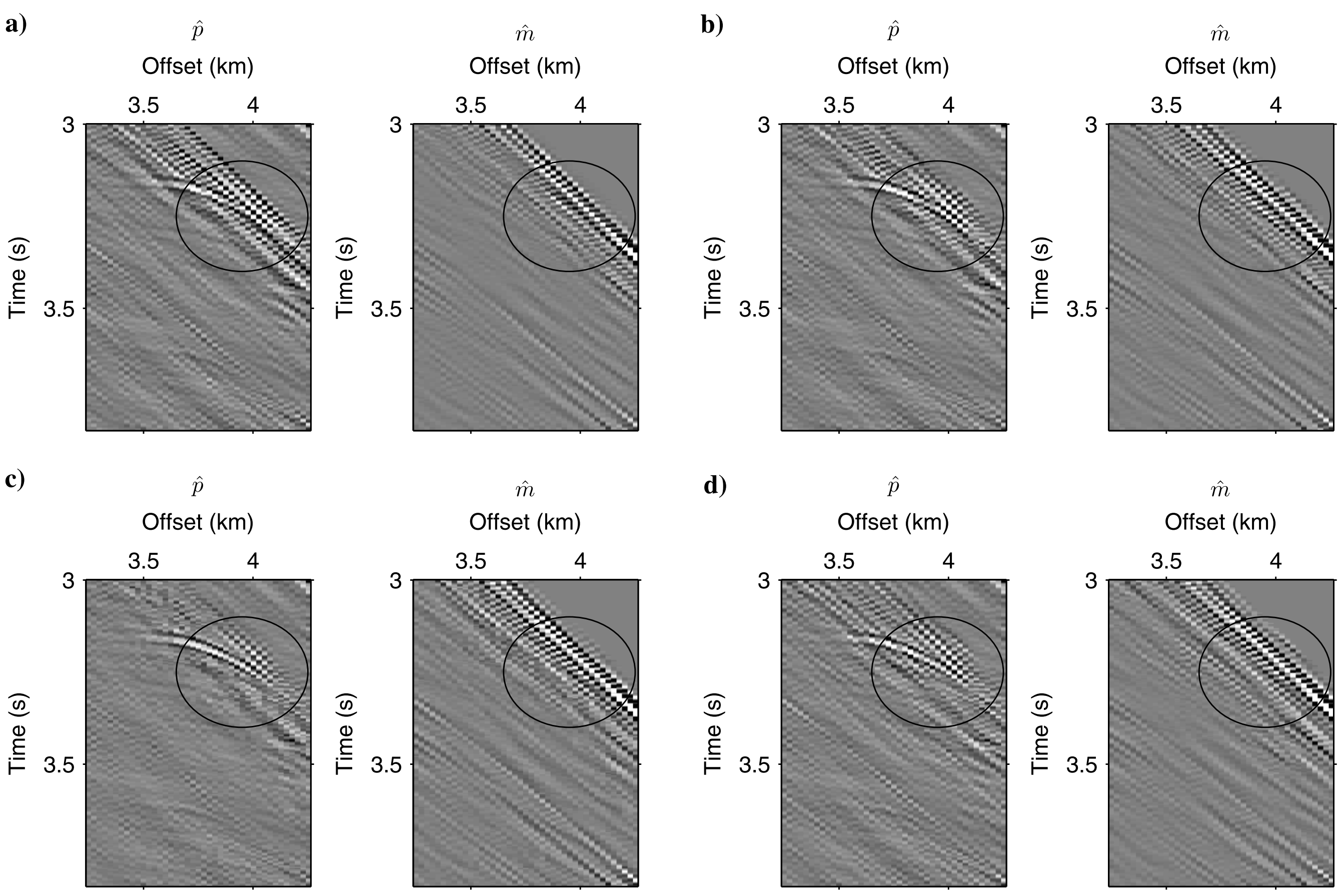

Figure 9. Magnification of the black rectangles of Figure 8. The legends are the same. The main differences are indicated by an ellipse.

However, InfoMax is generally more time consuming compared to the IRLS methods for a small matrix $\breve{\boldsymbol{M}}$.

The results of the ℓ_2-norm objective function with four windowing strategies are shown in Figures 8 and 9. A 50% window overlapping strategy is used for all tests. In all the tests, five adjacent traces are used to compute a 1D filter. As expected, the increase in the temporal length of the filter leads to more attenuation of the multiples, with an eventual overattenuation of the primaries.

The results of the four objective functions described in this article with the same windowing strategy are shown in Figures 10 and 11. In all the tests, five adjacent traces are used to compute a 1D filter. In this example, fewer differences are visible, which is consistent with the previous theoretical analysis showing the similarities between the analyzed objective functions. The observation of fewer dissimilarities between the primaries estimated by different objective functions is also valid for other windowing strategies. This suggests that the windowing strategy has more impact on the results than the choice of an objective function. In this context, Liu and Kostov (2015) recently focus on a criterion to find a proper filter size for a given data set.

DISCUSSION

Fundamentally, adaptive multiple subtraction is an underdetermined problem that consists of the joint recovery of a filter $\boldsymbol{w} \in \mathcal{W}$ and the primary signal $\hat{\boldsymbol{p}} \in \mathcal{P}$ from the data $\boldsymbol{d}$. Hence, for the problem to become determined in this form, it always misses a few equations equal to the size of the filter, at least. By setting $\hat{\boldsymbol{p}} \approx \boldsymbol{0}$, the problem can become virtually overdetermined and $\boldsymbol{w}$ can be estimated with an outnumber of linear constraints. Because the primaries (what we could see as the "noise" in the Wiener filtering) are not zeros (indeed, it is the signal), an objective function must weight the contribution of each constraint to be able to specify which solution is the best and unique estimate $\hat{\boldsymbol{p}}$.

Most of adaptive multiple subtraction schemes consider ℓ_q-norm matching filters for which it is assumed that the estimated primaries have minimum energy in the ℓ_q-norm sense (equation 4). However, the desired geophysical solution may not coincide with the optimized solution by ℓ_q-norms. In particular, ℓ_q-norm objective functions have their minimum at $\hat{\boldsymbol{p}} = \boldsymbol{0}$, leading to overattenuation problems if the outnumbered constraints in $\mathcal{P}$ are actually passing by this solution. To overcome this inherent problem, some authors recently propose to use objective functions based on the statistical independence of primaries and multiples. However, as we have shown in this article, there is an equivalence between them and ℓ_q-norm objective functions, if the right nonlinear function (or parameter) is chosen to approximate independence (via InfoMax or negentropy maximization). Hence, independence-based objective functions share the same issue as ℓ_q-norm objective functions because their minimum is obtained for $\hat{\boldsymbol{p}} = \boldsymbol{0}$.

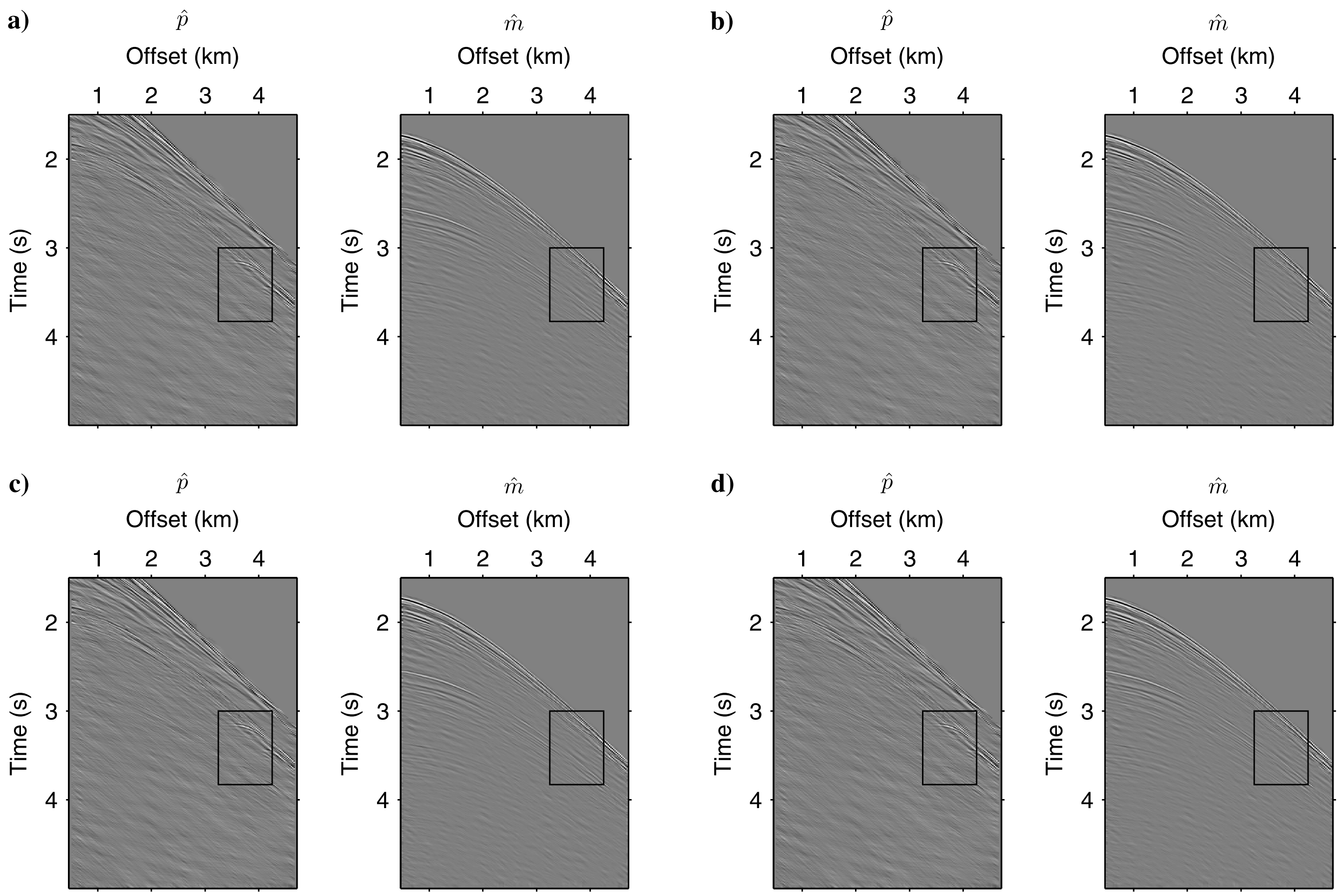

Figure 10. Primaries and multiples estimated with a 1D-2D strategy (five adjacent traces) of $F_t = 200$ ms and $K_t = 80$ ms by (a) the ℓ_2-norm, (b) the $\ell_{1/2}$-norm (IRLS), (c) InfoMax ($\lambda = 400$), and (d) negentropy maximization (g_1).

From a statistical point of view, the least-squares solution (ℓ_2-norm) assumes that primaries and multiple are uncorrelated, and we must remind ourselves that correlation is a measure of linear statistical dependence. When primaries and multiples overlap, they are actually correlated, and so, dependent. Hence, considering primaries and multiples as independent events is not a better strategy if they do overlap. In fact, as demonstrated in this article, it is the use of adjacent traces that increases the statistical diversity of primaries and multiples in a given window, thus allowing to overcome the overattenuation problem as we pointed out in Figure 5. Moreover, we have shown that forcing the independence between the predicted multiples and the estimated primaries can be seen as a nonlinear decorrelation between the same predicted multiples and the estimated primaries enhanced by a chosen operator. This operator has to be chosen to respect an a priori information about the PDF of the desired primaries.

All the methods analyzed in this article can be seen as adding a priori information about the statistical distribution of the primaries, so that the underdetermined adaptive multiple subtraction problem can be virtually overdetermined. If a sigmoid function is used, the InfoMax method becomes really similar to the hybrid $\ell_{1/2}$-norm method because these methods make a smooth transition between the ℓ_2 and the ℓ_1-norm solutions that, respectively, assume a Gaussian and a Laplacian distribution. Other nonlinear functions could be used in the InfoMax network, such as an asymmetric distribution. Unfortunately, the true distribution is not known, and its estimation is a difficult task. Hence, parametric methods, such as InfoMax and the hybrid $\ell_{1/2}$-norm, may be challenging in practice at choosing the appropriate parameter. On the other hand, nonparametric methods, such as ℓ_q-norm or negentropy methods, are easier to use and easier to interpret but less flexible.

If the ℓ_1 and the ℓ_2 solutions are close in the parameter space $\mathcal{W}$, all of the methods are expected to give similar results. It is well known that a subtle balance exists between the use of a short filter underestimating the noise and the use of a long filter overestimating it. The use of shorter filters may lead to more significant differences between two methods such as ℓ_1 and ℓ_2-norms, and we notice the same behavior with the analyzed methods on synthetic examples. However, on our real data set, the statistical diversity of the windows and the need of longer filters to well attenuate the noise seem to bring closer ℓ_1 and ℓ_2-norm solutions, leading to fewer significant differences between the analyzed methods.

CONCLUSION

We have shown that InfoMax, negentropy maximization, and hybrid ℓ_1/ℓ_2-norm-based matching filters share strong similarities. All of these techniques aim at minimizing the crosscorrelation between the predicted multiples and the estimated primaries enhanced by a chosen operator. It is this operator that links all the analyzed filtering techniques. Because correlation is a particular case of linear statistical dependence, the primary and multiple of a crossing event are statistically dependent. Then, forcing their statistical independence does not lead to a better solution. However, the windowing strategy,

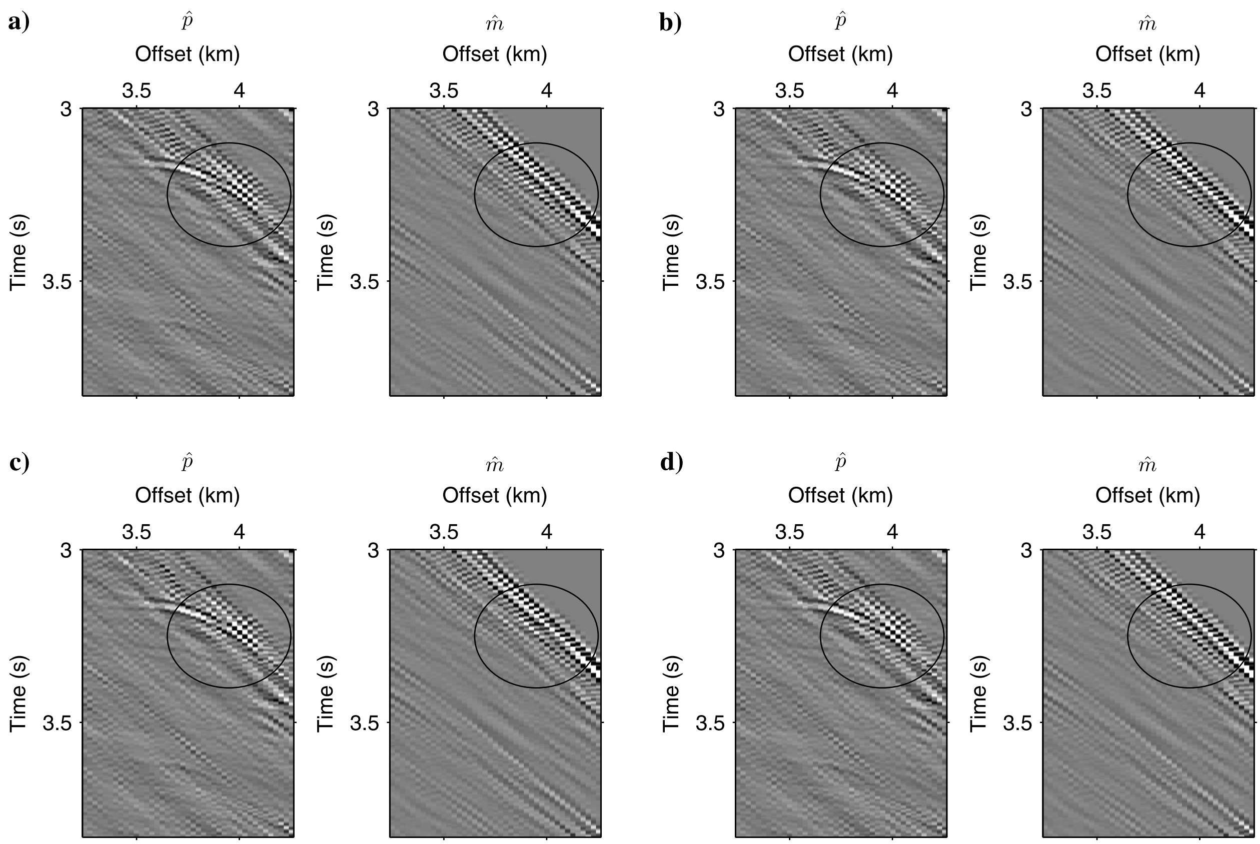

Figure 11. Magnification of the black rectangles of Figure 10. The legends are the same.

increasing the statistical diversity around the crossing event by the use of adjacent traces, is decisive because it actually allows us to model primaries and multiples as independent events.

ACKNOWLEDGMENTS

The authors would like to thank D. Velis, C. Kostov, and an anonymous reviewer for helping to improve this paper during revision. We would like to acknowledge Petrobras for their financial support. The research described in this paper was carried out as part of the Paris Exploration Geophysics Group project (GPX) funded by the French National Research Agency, CGG, Total, and Schlumberger.

REFERENCES

Abma, R., N. Kabir, K. Matson, S. Michell, S. Shaw, and B. McLain, 2005, Comparisons of adaptive subtraction methods for multiple attenuation: The Leading Edge, **24**, 277–280, doi: 10.1190/1.1895312.

Bell, A. J., and T. J. Sejnowski, 1995, An information-maximization approach to blind separation and blind deconvolution: Neural Computation, **7**, 1129–1159, doi: 10.1162/neco.1995.7.6.1129.

Bofill, P., and M. Zibulevsky, 2001, Underdetermined blind source separation using sparse representations: Signal Processing, **81**, 2353–2362, doi: 10.1016/S0165-1684(01)00120-7.

Cardoso, J., 1997, InfoMax and maximum likelihood for blind source separation: IEEE Signal Processing Letters, **4**, 112–114, doi: 10.1109/97.566704.

Comon, P., 1994, Independent component analysis, a new concept?: Signal Processing, **36**, 287–314, doi: 10.1016/0165-1684(94)90029-9.

Comon, P., and C. Jutten, 2010, Handbook of blind source separation: Independent component analysis and applications: Academic Press.

Costagliola, S., P. Mazzucchelli, and N. Bienati, 2011, Hybrid norm adaptive subtraction for multiple removal: 73rd Annual International Conference and Exhibition, EAGE, Extended Abstracts, P231.

Donno, D., 2011, Improving multiple removal using least-squares dip filters and independent component analysis: Geophysics, **76**, no. 5, V91–V104, doi: 10.1190/geo2010-0332.1.

Donno, D., H. Chauris, and M. Noble, 2010, Curvelet-based multiple prediction: Geophysics, **75**, no. 6, WB255–WB263, doi: 10.1190/1.3502663.

Fomel, S., 2009, Adaptive multiple subtraction using regularized nonstationary regression: Geophysics, **74**, no. 1, V25–V33, doi: 10.1190/1.3043447.

Guitton, A., 2005, Multiple attenuation in complex geology with a pattern-based approach: Geophysics, **70**, no. 4, V97–V107, doi: 10.1190/1.1997369.

Guitton, A., and D. J. Verschuur, 2004, Adaptive subtraction of multiples using the L_1-norm: Geophysical Prospecting, **52**, 27–38, doi: 10.1046/j.1365-2478.2004.00401.x.

Haykin, S., 2001, Adaptive filter theory: Prentice Hall.

Herrmann, F. J., U. Böniger, and D. J. Verschuur, 2007, Non-linear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799, doi: 10.1111/j.1365-246X.2007.03360.x.

Herrmann, F. J., D. Wang, and D. J. Verschuur, 2008, Adaptive curvelet-domain primary-multiple separation: Geophysics, **73**, no. 3, A17–A21, doi: 10.1190/1.2904986.

Hyvärinen, A., J. Karhunen, and E. Oja, 2001, Independent component analysis: John Wiley & Sons.

Hyvärinen, A., and E. Oja, 2000, Independent component analysis: Algorithms and applications: Neural Networks, **13**, 411–430, doi: 10.1016/S0893-6080(00)00026-5.

Kaplan, S. T., and K. A. Innanen, 2008, Adaptive separation of free-surface multiples through independent component analysis: Geophysics, **73**, no. 3, V29–V36, doi: 10.1190/1.2890407.

Li, Z.-X., and W.-K. Lu, 2013, Adaptive multiple subtraction based on 3D blind separation of convolved mixtures: Geophysics, **78**, no. 6, V251–V266, doi: 10.1190/geo2012-0455.1.

Li, Z.-X., and W.-K. Lu, 2014, Demultiple strategy combining Radon filtering and Radon domain adaptive multiple subtraction: Journal of Applied Geophysics, **103**, 1–11, doi: 10.1016/j.jappgeo.2014.01.004.

Liu, K.-H., and W. H. Dragoset, 2013, Blind-source separation of seismic signals based on information maximization: Geophysics, **78**, no. 4, V119–V130, doi: 10.1190/geo2012-0136.1.

Liu, K.-H., and C. Kostov, 2015, Multimodel adaptive subtraction with regularized parameter selection via generalized information criterion: Geophysics, **80**, no. 3, V33–V45, doi: 10.1190/geo2014-0474.1.

Lu, W., 2006, Adaptive multiple subtraction using independent component analysis: Geophysics, **71**, no. 5, S179–S184, doi: 10.1190/1.2243682.

Pham, M. Q., L. Duval, C. Chaux, 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.

Rickett, J., A. Guitton, and D. Gratwick, 2001, Adaptive multiple subtraction with non-stationary helical shaping filters, 63rd Annual International Conference and Exhibition, EAGE, Extended Abstracts, P167.

Saab, R., D. Wang, O. Yilmaz, and F. J. Herrmann, 2007, Curvelet-based primary-multiple separation from a Bayesian perspective: 77th Annual International Meeting, SEG, Expanded Abstracts, 2510–2514.

Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58, doi: 10.1190/1.1438154.

Ventosa, S., S. Le Roy, I. Huard, A. Pica, H. Rabeson, P. Ricarte, and L. Duval, 2012, Adaptive multiple subtraction with wavelet-based complex unary Wiener filters: Geophysics, **77**, no. 6, V183–V192, doi: 10.1190/geo2011-0318.1.

Verschuur, D. J., 2006, Seismic multiple removal techniques: Past, present and future: EAGE Publications BV.

Verschuur, D. J., and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion. Part II: Practical aspects and examples: Geophysics, **62**, 1596–1611, doi: 10.1190/1.1444262.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177, doi: 10.1190/1.1443330.

Wang, Y., 2003, Multiple subtraction using an expanded multichannel matching filter: Geophysics, **68**, 346–354, doi: 10.1190/1.1543220.

Wiggins, J. W., 1988, Attenuation of complex water-bottom multiples by wave equation-based prediction and subtraction: Geophysics, **53**, 1527–1539, doi: 10.1190/1.1442434.

Wiggins, J. W., 1999, Multiple attenuation by explicit wave extrapolation to an interpreted horizon: The Leading Edge, **18**, 46–54, doi: 10.1190/1.1438152.