Seismic wave field restoration using spare representations and quantitative analysis

Mai-Quyen Pham

Université Paris-Est

Lab. Informatique Gaspard Monge

UMR CNRS 8049

THESIS

presented by

Mai Quyen PHAM

15 July 2015, Paris

Seismic wave field restoration using sparse representations and quantitative analysis

---

Reporters:	Jean-Francois Aujol	Prof. Université de Bordeaux
	Mauricio D. Sacchi	Prof. University of Alberta
Examiners:	Jérôme Mars	Prof. Grenoble-INP
	Mai K. Nguyen	Prof. Université de Cergy-Pontoise
PhD supervisor:	Jean-Christophe Pesquet	Prof. Université Paris-Est Marne-la-Vallée
PhD co-supervisors:	Laurent Duval	Dr. IFP Energies nouvelles (IFPEN)
	Caroline Chaux	CNRS researcher, I2M, Aix-Marseille Univ.

Contents

Abstract

This thesis deals with two different problems within the framework of convex and non convex optimization. The first one is an application to multiple removal in seismic data with adaptive filters and the second one is an application to blind deconvolution problem that produces characteristics closest to the Earth layers. More precisely:

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 transforms. 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 seismic data.

In seismic exploration, a seismic signal (e.g. primary signal) is often represented as the results of a convolution between the "seismic wavelet" and the reflectivity series. The second goal of this thesis is to deconvolve them from the seismic signal which is presented in Chapter **??**. The main idea of this work is to use an additional premise that the reflections occur as sparsely restricted, for which a study on the "sparsity measure" is considered. Some well known methods that fall in this category are proposed such as (Sacchi et al., 1994; Sacchi, 1997). We propose a new penalty based on a smooth approximation of the ℓ_1/ℓ_2 function that makes a difficult nonconvex minimization problem. We develop a proximal-based algorithm to solve variational problems involving this function and we derive theoretical convergence results. We demonstrate the effectiveness of our method through a comparison with a recent alternating optimization strategy dealing with the exact ℓ_1/ℓ_2 term.

Résumé

Cette thèse porte sur la restauration de champs d'ondes sismiques perturbés par trois sources de dégradation. Ces sources sont dues à des trajets de propagation complexes, au dispositif d'acquisition, à des sources liées ou non à l'acquisition, et potentiellement présentes simultanément : des réflexions multiples (ou échos), une dégradation de la réponse impulsionnelle attendue (ou flou) et des perturbations plus aléatoires (ou bruits).

Nous avons considéré dans un premier temps le problème des réflexions multiples, réflexions qui se sont réfléchies plusieurs fois sur au moins une interface. Nous nous intéressons ici au filtrage adaptatif de ces réflexions sismiques multiples à partir de modèles approximatifs issus de modélisation sismique. Ce filtrage est réalisé dans un domaine de trames d'ondelettes discrètes, mono- et bidimensionnelles, sous contraintes de parcimonie et de variation lente des filtres adaptatifs. Ceci est intéressant en réflexion sismique, car les méthodes standard peuvent produire des filtres très mal conditionnés, du fait notamment du caractère passe-bande des données sismiques. Dans ce travail, une formulation variationnelle des problèmes de réflexions multiples est proposée. Nous utilisons des algorithmes proximaux, dont la convergence est garantie lorsqu'il s'agit d'optimiser dans un cadre convexe. L'avantage de ces approches est l'utilisation d'une régularisation sophistiquée, permettant de considérer la parcimonie à la fois a) dans le domaine d'ondelettes, b) via des a priori sur les filtres pour lesquels nous avons utilisé différentes fonctions de régularisation (norme ℓ_1, ℓ_2, mixte $\ell_{1,2}$ et nucléaire). Notre méthode vise à étendre et améliorer certains aspects de la méthode proposée par S. Ventosa en collaboration avec CGG en 2012, et testée avec succès sur plusieurs campagnes sismiques. Les résultats que nous avons obtenus démontrent la performance de notre méthode non seulement sur des données synthétiques bruitées mais également sur des données réelles.

Nous nous intéressons ensuite au problème de déconvolution aveugle. En géophysique, un modèle simplifié de la Terre souvent utilisé fait l'hypothèse d'un nombre de couches localement parallèles, chacune avec des propriétés constantes. Mais la vitesse, la densité ou les deux peuvent varier d'une couche à l'autre. L'impédance acoustique est calculée pour chaque couche ; puis les coefficients de réflexion pour une incidence normale sont calculés aux endroits où il y a des changements d'impédance acoustique. Chaque change-

ment d'impédance acoustique opère une modification d'amplitude et de polarisation liée aux coefficients de réflexion. Ainsi, la séquence de réflectivité (réponse impulsionnelle) est convoluée avec la forme d'onde descendante pour donner une trace sismique. Ce problème constitue un contexte de déconvolution aveugle où l'on recherche un signal inconnu, convolué avec une forme d'onde elle aussi inconnue en présence de bruit additif. La déconvolution requiert souvent d'introduire des hypothèses complémentaires sous forme de pénalisation, notamment non convexe. L'ambiguïté d'échelle en déconvolution aveugle suggère l'usage de fonctions de contraste invariantes en échelle. Dans cette thèse, nous proposons un algorithme de minimisation alternée, de type explicite-implicite à métrique variable. Il traite une approximation lisse du rapport ℓ_1/ℓ_2 (SOOT pour "Smoothed One-Over-Two penalty") pour des données réelles signées. Nous étudions les propriétés de convergence de la méthode proposée, basées sur l'inégalité de Kurdyka-Łojasiewicz. Les performances de cette nouvelle approche sont illustrées à travers un exemple en déconvolution aveugle de données sismiques, mais aussi sur des images.

Résumé long

Les méthodes d'exploration basées sur la sismique (Lavergne, 1986) s'appuyent sur la propagation d'ondes générées par un ébranlement du sous-sol. Celles-ci sont réfléchies ou réfractées aux interfaces des couches géologiques. Après une succession de transmissions et de réflexions, une partie de l'énergie de l'ébranlement peut être enregistrée par des capteurs sismiques (géophones ou hydrophones) localisés à proximité de la surface. Ces capteurs font l'acquisition de signaux correspondant aux amplitudes et temps de propagation des ondes considérées. Ces signaux temporels, appelés "traces sismiques" possèdent un spectre de fréquence allant de quelques Hertz à moins de cent Hertz en général. Une acquisition sismique consiste en la collecte de traces sismiques (z) issues de capteurs régulièrement répartis formant des antennes acoustiques en surface ou remorquées par des navires, ce pour un grand nombre d'ébranlements (ou tirs) générés en différentes positions. Par analyse conjointe des traces sismiques, il est possible de déterminer les distances relatives entre les interfaces géologiques et les contrastes de vitesse des ondes s'y propageant, comme illustré dans la Figure 3.1 - p. 34. Ces données sismiques sont affectées de nombreuses distorsions, dues à des trajets de propagation complexes, au dispositif d'acquisition et à des sources externes. En général, ces distorsions ne sont pas aléatoires. Elles consistent en des signaux parasites, souvent structurés et de forte amplitude, qui recouvrent les signaux utiles correspondant aux réflexions intéressantes. L'objet du traitement géophysique (Mars et al., 2004) est l'extraction d'information pertinente à partir de ces importants volumes de données. Du fait de la complexité des champs d'ondes interférents, la géophysique est à l'origine du développement de méthodes de traitement de signal avancées, comme la déconvolution sous contrainte de norme ℓ_1 (Claerbout and Muir, 1973), ou les transformées en ondelettes continues (Morlet et al., 1982).

Cette thèse se divise en deux parties : 1) filtrage adaptatif de réflexions multiples et 2) déconvolution aveugle.

Filtrage adaptatif de réflexions multiples

L'enregistrement sismique est constitué notamment de réflexions primaires et de réflexions multiples qui forment l'objet principal des méthodes développées dans cette thèse. Lorsque

l'onde aborde la première interface, une partie de son énergie est renvoyée vers le haut; une partie de l'énergie est transmise vers bas, l'onde transmise se réfléchit sur la seconde interface et est renvoyée vers le haut; une partie de l'énergie est également transmise vers le bas à travers la seconde interface et les ondes transmises se réfléchissent sur les interfaces suivantes et ainsi de suite jusqu'à la dernière interface susceptible de produire une réflexion sismique. Remarquons que les réflexions primaires se réfléchissent une fois seulement sur les interfaces. Pratiquement, l'enregistrement sismique comporte, outre les réflexions primaires, des réflexions multiples qui se réfléchissent plusieurs fois sur les interfaces.

L'objectif de ce travail est de proposer des méthodes fiables et efficaces pour résoudre le problème de filtrage des réflexions multiples en sismique. Ainsi, nous nous intéresserons à l'emploi de représentations parcimonieuses pour la restauration et l'analyse quantitative de champs d'ondes en sismique.

Nous supposons qu'une trace sismique est composée du primaire d'intérêt y, des multiples s et du bruit b, comme suit:

$$z^{(n_t)} = \overline{s}^{(n_t)} + \overline{y}^{(n_t)} + b^{(n_t)}$$

où $n_t \in \{0, \ldots, N_t - 1\}$ désigne l'indice temporel et $z = (z^{(n_t)})_{0 \leq n_t < N_t}$ correspond aux données observées (trace sismique) combinant: le primaire $\overline{y} = (\overline{y}^{(n_t)})_{0 \leq n_t < N_t}$ (signal d'intérêt, inconnu) représenté en trait continu et pointillés bleu (Figure 3.1 - p. 34), les multiples $(\overline{s}^{(n_t)})_{0 \leq n_t < N_t}$ (somme des signaux réfléchis non désirés) représentés en pointillés rouge (Figure 3.1 - p. 34) et le bruit $b = (b^{(n_t)})_{0 \leq n_t < N_t}$ qui est supposé être une réalisation de bruit additif.

Le problème posé est d'estimer y a partir de z. Les multiples et le bruit étant inconnus, il est nécessaire de faire des hypothèses supplémentaires. Pour cela, nous utilisons des modèles approximatifs des multiples (fournis par des méthodes sismiques non détaillées ici). Par des considérations géophysiques, on sait que ces modèles sont imprécis en amplitude, en fréquence et en localisation temporelle. Pour compenser ces imprécisions, nous effectuons un filtrage adaptif sur les signaux servant de modèles pour les faire "ressembler" le plus possible aux multiples réels, et ensuite les soustraire.

L'idée s'inspire de méthodes de filtrage des multiples en sismique classique. Cela veut dire que nous chercherons une transformée adaptée à notre signal, tout en utilisant un formalisme analogue à celui de la déconvolution.

Nous utiliserons des algorithmes proximaux, dont la convergence est garantie lorsqu'il s'agit d'optimiser dans un cadre convexe (minimiser des critères convexes). Le terme de fidélité sera choisi judicieusement à partir d'un critère adapté aux caratéristiques du bruit. De plus, nous utiliserons des termes de régularisation basés sur la parcimonie des coefficients du signal dans le domaine fréquentiel ou temps-échelle (base, trame...), et du filtre composé avec un opérateur linéaire. En particulier, nous ferons des hypothèses supplémentaires sur les filtres adaptatifs afin d'assurer un meilleur comportement des solutions.

Un exemple modélisé de données sismiques est illustré dans la Figure 1 - p. xiii.

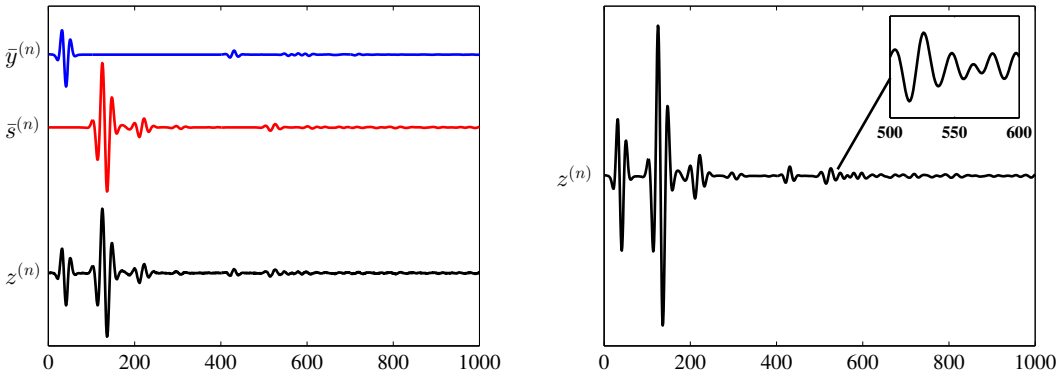


Figure 1: *À gauche : de haut en bas, le signal d'intérêt (y inconnu) en bleu, le multiple (s inconnu) en rouge et le signal observé (z connu) en noir; à droite: zoom sur une zone d'intérêt de z.*

La figure de gauche montre une trace sismique (z, en bas), le signal d'intérêt étant le primaire (y, en haut), auquel s'ajoute le multiple (s, au milieu). On ne connaît ici que z (le signal observé) et on voudrait obtenir y (le primaire d'intérêt). Nous agrandissons une zone d'intérêt (figure de droite) pour le signal observé. Elle met en évidence le mélange des signaux primaires et des multiples. Il n'est donc pas facile d'enlever les multiples dans cette zone. Dans la littérature, on trouve deux grandes classes de méthodes.

Une première approche représente les données dans un domaine transformé avec un chevauchement minimal entre les primaires et multiples. On effectue un filtrage ou une sélection dans le domaine transformé, qui permet d'atténuer des multiples ou de sélectionner des primaires d'intérêt. Les coefficients filtrés sont alors reconvertis dans le domaine temporel ou spatial, en utilisant une transformée inverse. Des outils standard comprennent la sommation combinée avec NMO *(Normal Move Out)*, les filtrages homomorphiques (Buttkus, 1975), $f-k$ (Wu and Wang, 2011), et $\tau-p$ (Nuzzo and Quarta, 2004), ainsi que différents avatars (parabolique, hyperbolique) de la transformée de Radon (Hampson, 1986; Trad et al., 2003; Nowak and Imhof, 2006).

Une seconde approche inclut des filtres de prédiction et leurs variantes (Taner, 1980; Abma et al., 2005; Spitz et al., 2009), aussi appelées filtres de mise en forme, ou adaptés dans la littérature en traitement du signal (Ristow and Kosbahn, 1979) ou de déconvolution prédictive (Taner et al., 1995). Incidemment, le filtrage adapté des multiples peut être vu comme une forme de déconvolution informée. Récemment, des techniques basées sur la modélisation, telle que le SRME *(surface-related multiple elimination)*, ont démontré d'excellentes performances, ce qui permet de guider par les données l'élimination de multiples (Verschuur and Berkhout, 1992; Weglein et al., 1997; Lin et al., 2004) à partir d'un modèle basé sur l'équation des ondes (Pica et al., 2005a; Weisser et al., 2006). Ces méthodes, basées sur des modèles de multiples, consistent à prédire, puis à soustraire les modèles de multiples adaptés aux données sismiques.

Le modèle adapté est généralement obtenu par un filtre linéaire de type Wiener (Robinson and Treitel, 1967) qui minimise l'énergie de la différence entre le signal observé et le modèle. Bien que les primaires et les multiples soient générés à partir de la même source et ne soient pas complètement décorrélés, ou orthogonaux, les solutions basées sur l'erreur quadratique moyenne ou Least Squares Estimation (LSE) sont très utilisées en raison de leur simplicité de calcul. Pour réduire la distorsion des primaires, ces filtres sont généralement appliqués en deux étapes en pratique. Les filtres longs compensent globalement l'amplitude, la forme d'onde et le décalage temporel, puis des filtres plus courts corrigent des écarts variables en temps.

Par expérience, les méthodes hybrides, mélangeant les deux approches de prédiction et de transformation sont généralement plus performantes. Les améliorations peuvent résider dans différentes stratégies. L'utilisation des critères plus robustes que les méthodes basées sur l'erreur quadratique moyenne, comme ceux basés sur la norme ℓ_1 (Guitton and Verschuur, 2004), ou les méthodes basées sur des décompositions en composantes indépendantes (Kaplan and Innanen, 2008) sont des alternatives efficaces. Les méthodes basées sur l'estimation simultanée d'ensembles de filtres présentent une autre option, par exemple, en adaptation multi-canal (Wang, 2003). Une représentation appropriée des données peut en outre améliorer la soustraction adaptative ((Taner, 1980) avec une transformation radiale, (Berkhout and Verschuur, 2006) avec une transformation focale). Une nouvelle tendance met l'accent sur des approches multi-échelles ou en ondelettes (Jacques et al., 2011), ce qui pourrait mieux promouvoir la parcimonie dans l'exploitation de légères différences entre les spectres et les vitesses apparentes des primaires et des multiples. (Pokrovskaia and Wombell, 2004) et (Ahmed, 2007) ont utilisé principalement la transformée en ondelettes discrète. (Herrmann and Verschuur, 2005), (Donno et al., 2010), (Neelamani et al., 2010) ont utilisé la transformée en *curvelet* que l'on considère parfois comme une transformée de Radon multi-échelles locale (de Hoop et al., 2009).

Une méthode a récemment été proposée par S. Ventosa en collaboration avec CGG (Ventosa et al., 2012). Elle minimise l'erreur quadratique moyenne entre des coefficients du signal observé et ceux de la combinaison des modèles, dans un domaine de trame d'ondelettes complexes et continues. Cette méthode de soustraction utilise la redondance fournie par la trame d'ondelettes complexes dans une conception des filtres rapides à calculer. Elle revient à faire un filtre adaptatif non stationnaire dans le domaine des ondelettes complexes. Notre méthode vise à étendre et améliorer certains aspects de cette approche, notamment par:

1. l'utilisation de transformations discrètes plutôt que continues, permettant d'apporter des gains de calculs (temps, mémoire),

2. l'ajout de contraintes sur les filtres adaptatifs assurant un meilleur comportement des solutions,

3. la prise en compte des informations liées au bruit, permettant de mieux enlever les perturbations incohérentes dans le signal,

4. la prise en compte des informations sur les filtres en utilisant différentes fonctions

de régularisation (normes ℓ_1, ℓ_2, $\ell_{1,2}$ et nucléaire),

5. l'extension de la méthode au 2D en utilisant des décompositions M-bandes en arbre dual 2D avec différents filtres d'ondelettes suivant le temps et l'espace (Chaux et al., 2006, 2007b), et le rajout de contraintes sur les filtres adaptatifs (non seulement sur le temps mais aussi sur l'espace),

6. les expériences sur les données synthétiques ainsi que les données réelles permettant de mettre en évidence la performance de notre méthode par rapport aux méthodes de l'état de l'art, par exemple, la méthode de (Ventosa et al., 2012).

Déconvolution aveugle

Un modèle simplifié de la Terre souvent utilisé fait l'hypothèse d'un nombre de couches localement parallèles, chacune avec des propriétés constantes, mais la vitesse, la densité ou les deux peuvent varier d'une couche à l'autre. L'impédance acoustique est calculée pour chaque couche; puis les coefficients de réflexion pour une incidence normale sont calculés pour les endroits où il y a des changements d'impédance acoustique. Chaque changement d'impédance acoustique reflète une ondelette réduite en amplitude et polarité en fonction du coefficient de réflexion, et la somme de ceux-ci est la trace de sortie. Ainsi, la séquence de réflectivité est convoluée avec la forme d'onde descendante (en fait, avec l'ondelette) pour donner une trace sismique synthétique. Un exemple d'un sismogramme synthétique est représenté sur la Figure 1.6 - p. 7. Une information géologique précieuse est contenue dans la véritable amplitude de l'événement de réflexion, qui peut être récupérée à partir d'enregistrements de terrain convenablement calibrés. Toute variation latérale de réflexion en amplitude est due au changement dans la lithologie latérale de la couche de roche ou de la teneur en liquide des pores. Dans la pratique, l'interprétation structurale sismique est dirigée vers la création de cartes de structure de la sous-surface de la sismologie par réflexion. Cependant, la convolution avec une ondelette altère souvent l'information nécessaire à une interprétation stratigraphique complète des données. Ainsi, la restauration de l'amplitude réelle dans les sections sismiques revêt une importance grandissante. Dans sa forme la plus simple, un sismogramme synthétique $y(t)$ peut être considéré comme la convolution de la fonction de source supposée $h(t)$ (ondelette sismique) avec une fonction de réflectivité $x(t)$ qui représente le contraste d'impédance acoustique dans le modèle en couches:

$$y(t) = x(t) * h(t) + w(t) \tag{1}$$

Avec ce problème, on se trouve dans un contexte de déconvolution aveugle où l'on recherche un signal inconnu (la réflectivité du sous-sol), convolué avec une forme d'onde elle aussi inconnue (ondelette sismique) en présence de bruit additif. Elle requiert souvent d'introduire des hypothèses complémentaires sous forme de pénalisation, notamment non convexe. L'ambiguité d'échelle en déconvolution aveugle suggère l'usage de fonctions de contraste invariantes en échelle. La fonction correspondant au rapport de

normes ℓ_1 et ℓ_2 présente de bonnes propriétés d'estimation de la parcimonie de signaux ou d'images. Quelques travaux existent pour minimiser un critère composite faisant intervenir cette fonction, mais très peu d'entre eux offrent des garanties de convergence. Dans ce manuscrit, nous proposons un algorithme de minimisation alternée, de type explicite-implicite à métrique variable. Il traite une approximation lisse du rapport ℓ_1/ℓ_2 (SOOT pour "Smoothed One-Over-Two penalty") pour des données réelles signées. Nous étudierons les propriétés de convergence de la méthode proposée, basées sur l'inégalité de Kurdyka-Łojasiewicz. Les performances de cette nouvelle approche sont illustrées à travers un exemple en déconvolution aveugle de données sismiques, mais aussi sur des images.

Remerciements

Je tiens tout d'abord à exprimer ma gratitude à mon directeur de thèse Jean-Christophe Pesquet et à mes deux encadrants Caroline Chaux et Laurent Duval. Je les remercie pour leur confiance, ainsi que pour leur soutien permanent tout le long de ma thèse. Malgré leurs emploi du temps sans doute très chargé, ils se sont toujours montrés extrêmement disponible et m'ont guidé avec beaucoup d'enthousiasme, de compréhension et de professionnalisme.

Je remercie Monsieur Jean-François Aujol et Monsieur Mauricio D. Sacchi pour avoir accepté de rapporter cette thèse, pour les remarques constructives et pour l'intérêt qu'ils ont porté à mes travaux. Je tiens également à remercier Madame Mai K. Nguyen pour avoir accepté de faire partie de mon jury. Je suis également très honorée que Monsieur Jérôme Mars ait accepté de présider ce jury. Cette thèse a été effectuée au sein du Département Contrôle, Signal et Systèmes de IFPEN et de l'Équipe Signal et Communications du Laboratoire d'informatique Gaspard Monge. J'ai été très bien accueillie et ai pu exercer mes travaux de recherche dans de excellentes conditions. Je remercie tous les membres de ces deux sites pour leur aide, pour les discussions que nous avons eues et pour leurs conseils.

Je tiens à remercier NGO Hoai Diem Phuc qui, plus qu'une amie, est une soeur adorable, sympathique et sincère avec qui j'ai passé de bons moments. J'adresse également une pensée particulière à tous mes amis vietnamiens, que je n'ai pas pu citer ici. Mes remerciements les plus sincères vont à vous qui avez partagé mes joies ainsi que mes difficultés dans ma vie en France.

Enfin, j'ai toujours pu compter sur l'affection et le soutien de ma famille au Vietnam. Je pense à mes parents et à mon frère, ma belle soeur, ma nièce, que je ne remercierai jamais assez.

I dedicate this thesis to my beloved parents Thuy Mai and Duc Chinh.

- Chapter 1 -

Introduction

1.1 Seismic data and primaries

Seismic data acquisition involves generating a seismic energy pulse, such as a compressed air gun (marine seismic acquisition), or a seismic vibrator or explosive source (land seismic acquisition) at discrete surface locations. The resulting energy is reflected back to the surface from interfaces where rock properties change. By recording this reflected energy at an array of hydrophones (marine seismic acquisition) or geophones (land seismic acquisition) placed on the ground surface, seismic data can be processed to produce an image of underground geological structures and a range of attributes that can be used to infer the physical rock properties.

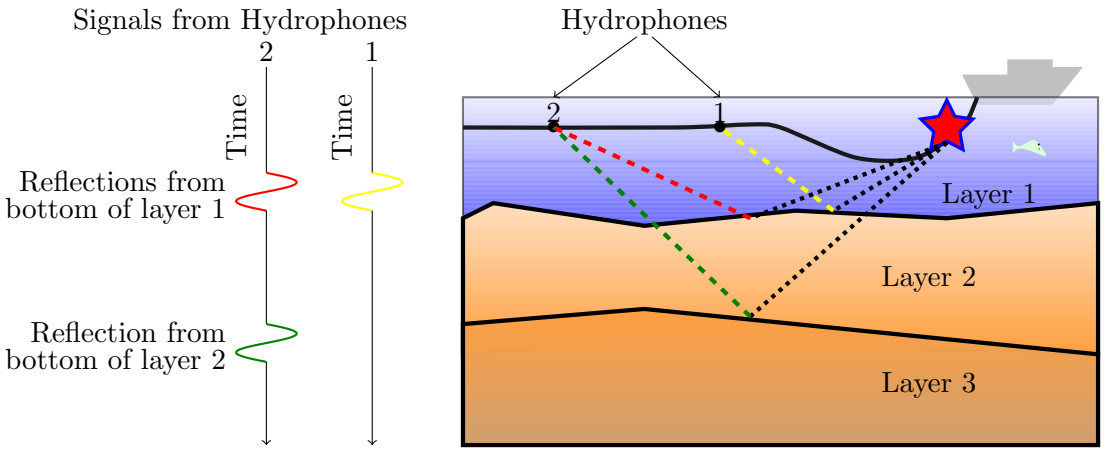


Figure 1.1: *The seismic reflection method.*

Figure 1.1 - p. 1 depicts some acoustic waves and their travel paths for two hydrophones. An acoustic impedance contrast exists between layers 1 (water) and 2 (bedrock), and between layers 2 and 3. At the boundary between layers 1 and 2, part of

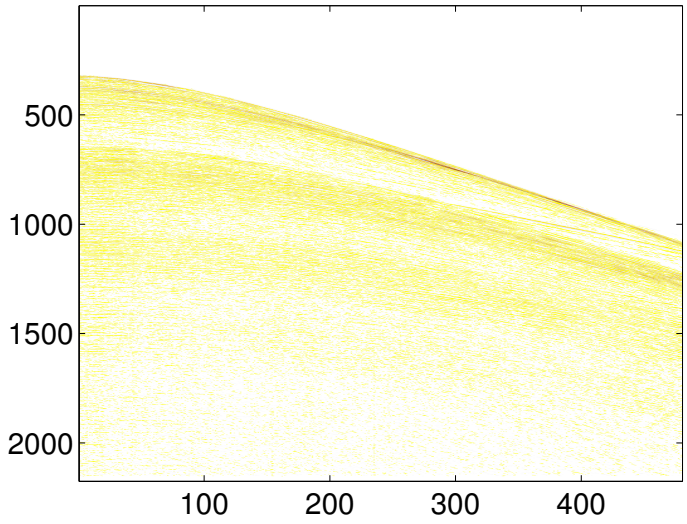


Figure 1.2: *Portion of common shot gather.*

the seismic energy is reflected and part is transmitted and continues through layer 2 until it reaches the impedance contrast between layers 2 and 3. At this interface, part of the wave is again reflected and part is transmitted and continues into layer 3 (not shown). The two simplified traces on the left hand side of the layer picture show the signals arriving at the two hydrophones. The electric current values recorded at the geophone or hydrophone are usually digitized with a sampling period of 1 to 4 milliseconds, typically before being transmitted to the recorder. Theoretically, the value recorded on the tape will either be proportional to the particle velocity or the pressure variation. Normally however, the coefficient of proportionality is disregarded and the seismic "amplitudes" recorded on the tape are relative.

Figure 1.2 - p. 2 illustrates a seismic image that is collected with five hundred of seismic traces. These seismic traces must be analyzed so as to yield contour maps of the subsurface sedimentary structure. These maps are used in order to make decisions about the locations of where to drill exploratory oil and gas wells. There are two important variables in seismic prospecting: time of reflected events (t) and velocity (v). For simplicity, we assume that velocity is a constant v, so that the relationship between arrival time and offset reads (hyperbolic equation)

$$t_x = \sqrt{t_0^2 + \left(\frac{x}{v}\right)^2} \tag{1.1}$$

where

- t_x is the actual reflection time of the seismic event due to *normal moveout correction* (NMO) effects, corresponding to the vertical axis in Figure 1.2 - p. 2,

- t_0 is the zero offset reflection time of the seismic event,
- x is the actual source-receiver offset distance, corresponding to the horizontal axis in Figure 1.2 - p. 2,
- v is the NMO velocity for this reflection event.

The physical rock properties can be determined from knowledge of these variables. The distances to subsurface reflectors are calculated from observed travel times and computed velocities. Seismic waves propagate with the velocity of sound in rock, and so the propagation velocity generally depends on chemical composition and local geology. More precisely, it depends on the elastic constant of rocks and their density. These elastic constants and densities are used to characterize rocks, for instance porosity, texture etc. Generally the velocity increases with depth, although occasionally there may be layers in which a decrease in velocity occurs. For a given surface point, the velocity plotted as a function of depth is called the velocity function. This velocity function varies from one location to another, a given velocity function cannot be assumed to be valid for an entire prospect. As a result, the velocity function must be frequently corrected from place to place over the area of exploration. Each recorded trace is a time series made up of reflected events together with various interfering waves and noise. The "desired reflected" events are the primary reflections who

- are the first reflections of the source wavelet from an interface,
- carry useful information about the velocities and thicknesses of subsurface layers.

On the contrary, seismic noise is considered as everything in the record that we did not want to record ("undesired signal"). There are two types of seismic noise:

- Coherent noise which can be followed across at least a few traces. It includes surface waves (ground roll), direct waves, refractions, diffractions, and multiples.
- Incoherent noise which is apparently random on all traces. It includes noise generated by near-surface scatterers, rain, humans, and interference coherent noise.

Multiple reflections are an important type of undesired waves. Occasionally, some noises can reveal useful for seismic interpretation. However, we consider them in the traditional signal/noise separation context. The importance of the problem they raise will be made more explicit in the next sections.

1.2 Seismic multiples

Seismic multiples are events that have undergone more than one reflection from the same interface. Multiple reflections fall into one of two basic categories (see Figure 1.3 - p. 4) based on their paths: short-path (period) multiples and long-path (period) multiples.

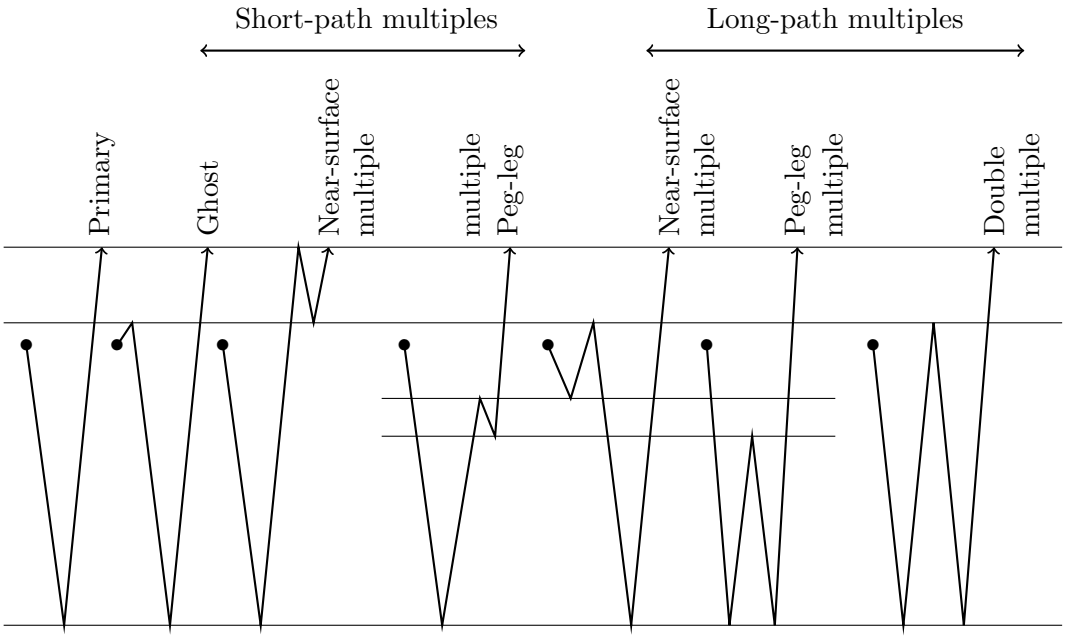


Figure 1.3: *Types of multiple reflections from (Sheriff, 1973).*

- A long-path multiple is generated far from the primary so that it is totally separated from the primary reflection from the same reflector. Therefore, it appears as a separate event on the seismic record.
- A short-path multiple is generated close to the primary so that it interferes with the primary reflection from the same reflector. Its effect is to lower the vertical resolution. The most important short-path multiples are of two main types:
  - Peg-leg multiples resulting from the addition to a primary reflection of energy reflected from both the top and bottom of a thin bed, either on the way to or on the way back from the principal reflecting horizon.
  - Ghosts where part of the energy leaving the source travels upward and is reflected downward either at the base of the low-velocity layer or at the surface.

Figure 1.4 - p. 5 comes from offshore Labrador (an Atlantic-type divergent margin), which, locally, overlies late Cretaceous rift-type basins. This line is influenced by various multiples which have higher amplitude then their corresponding primaries. One is visible and can be easily accounted for by three successive rebounds within the water layer, and at one second on the left of the section, there are several other multiples which are generated by the strong dipping horizon. Probably, they are secondary rebounds of this primary event: (i) within the water layer, (ii) downward on the water bottom or (iii) a combination of both. (Source: http://homepage.ufp.pt/biblioteca/Seismic/Pages/Page10.htm). Figure 1.5 - p. 6 shows several multiples induced by the sea level. A strong reflection, culminating at approximately two seconds, appears in the middle of the section. Because

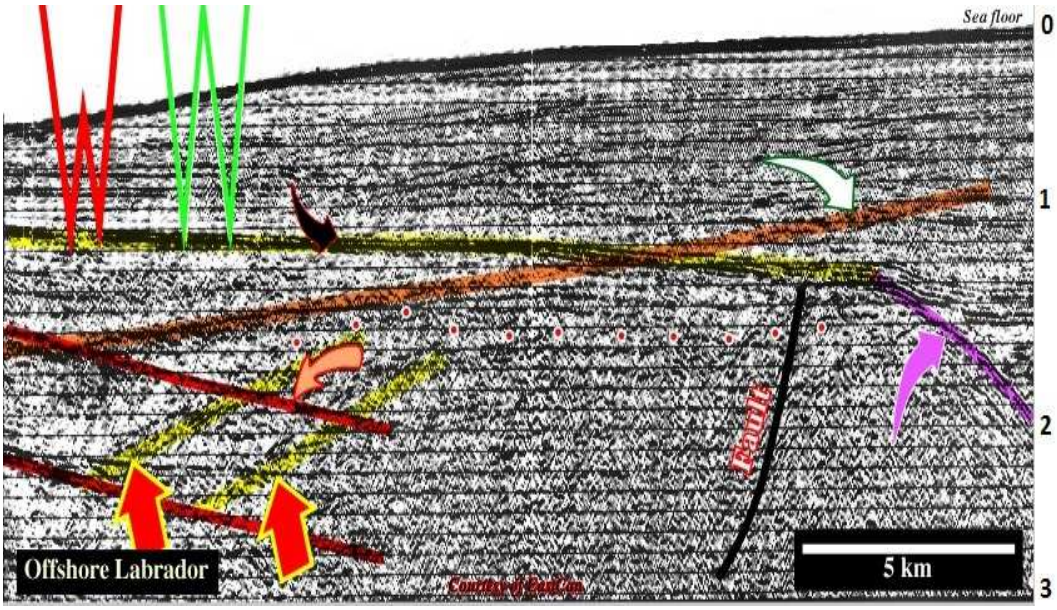


Figure 1.4: *This old seismic line comes from offshore Labrador (an Atlantic-type divergent margin), which, locally, overlies late Cretaceous rift-type basins. (Source: http://homepage.ufp.pt/biblioteca/Seismic/Pages/Page10.htm).*

of the high acoustic impedance contrast at the water-bottom, 1st and 2nd order multiples are produced within the water layer (ringing/ reverberations). The time difference between the primary and its first multiple is everywhere equal to the time through the water layer (Source: http://homepage.ufp.pt/biblioteca/Seismic/Pages/Page10.htm).

We refer to the comprehensive survey in Verschuur (Verschuur, 2006) for a detailed description of seismic multiple removal problem: the classification of multiple reflections and their characteristics, the methods and techniques that have been developed for these problems.

1.3 Seismic deconvolution

An often used simplified Earth model assumes the underground to be composed of a number of locally parallel layers, each with constant properties, but the velocity or density or both may differ from layer to layer. The acoustic impedance is calculated for each layer; then reflection coefficients for normal incidence are calculated for places where the acoustic impedance changes. Each acoustic impedance change reflects a wavelet scaled in amplitude and polarity according to the reflection coefficient, and the sum of these is the output trace. Thus, the reflectivity sequence is convolved with the downgoing waveform (actually, with the seismic wavelet) to give a synthetic seismic trace. An example of a synthetic seismogram is shown in Figure 1.6 - p. 7. The true amplitude of

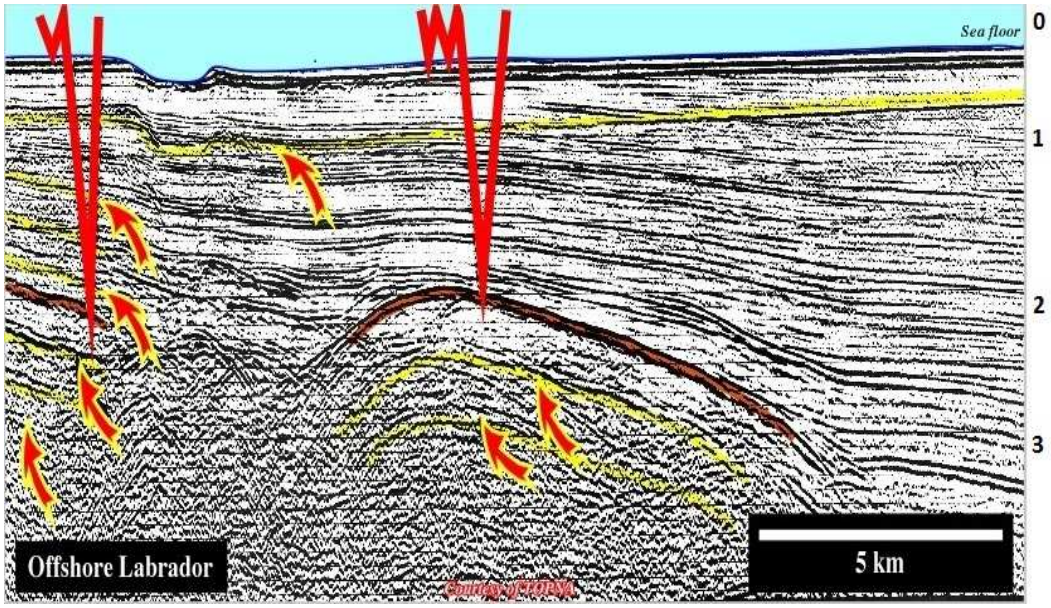


Figure 1.5: *Marine section from offshore Labrador shows several multiples induced by the sea level.(Source: http://homepage.ufp.pt/biblioteca/Seismic/Pages/Page10.htm).*

the reflection event contains most valuable geological information that can be retrieved from suitably calibrated field recordings. Any lateral variation of reflection amplitude is due to lateral change in the lithology of the rock layer or in its pore fluid content. In practice, structural seismic interpretation is directed toward the creation of structural maps of the subsurface from the reflection seismology. However, the convolution with a wavelet sometimes suppresses the information that is essential to a full stratigraphic interpretation of the data. That is why, true-amplitude seismic sections are becoming increasingly important.

In its simplest form, a synthetic seismogram $y : \mathbb{R} \to \mathbb{R}$ may be considered as the convolution of the assumed source function $h : \mathbb{R} \to \mathbb{R}$ (wavelet) with a reflectivity function $x : \mathbb{R} \to \mathbb{R}$ representing the acoustic impedance contrasts in the layered model:

$$(\forall t \in \mathbb{R}) \quad y(t) = (x * h)(t) \tag{1.2}$$

Actually, the convolutional model always contains random noise, thus (1.2) should be written as

$$(\forall t \in \mathbb{R}) \quad y(t) = (x * h)(t) + w(t) \tag{1.3}$$

where w is the noise component.
The recovery of the reflectivity constitutes a problem of main interest that will be addressed in chapter 6.
Figure 1.7 - p. 8 illustrates enlarged portions from 25 m to 30 m of GPR data (Source:

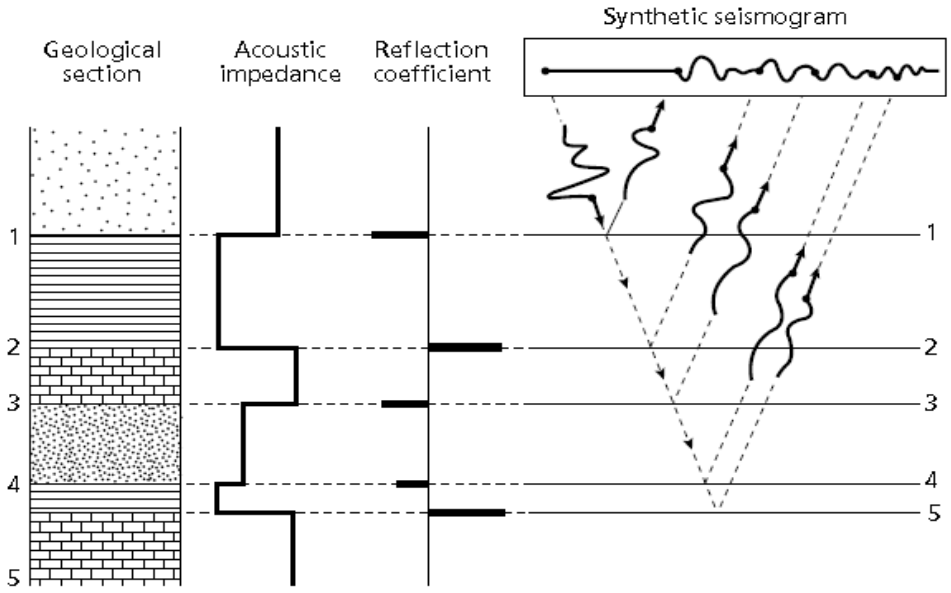


Figure 1.6: *Reflection seismograms showing primary reflection only, reprinted by permission of Birkhauser Verlag From Al-Sadi, 1980.*

http : //www.kgs.ku.edu/Geophysics/OFR/2004/OFR04_41). Comparing the deconvolved data (Figure 1.7b - p. 8) with the photomosaic (Figure 1.7c - p. 8), we can see the direct relationship between them. Subtle changes identified on the quarry face are also identifiable on the enlarged portions of the deconvolved data. The oval reflection portion in Figure 1.7b - p. 8 is an apparent lateral effect from a portion of the limestone layer showing relief away from the quarry wall at 27 m and 0.3 m depth in Figure 1.7c - p. 8. Additionally, the deconvolved data correctly show undulations of Bed B and the smaller scale stratal complexities that occur between Bed • and Bed E. It is also worth pointing out that a slight dip in Bed F identified on the quarry wall (oval on the left side of Figure 1.7c - p. 8) is identified as a dipping reflector on the deconvolved data (oval on the left side of the 25 m mark on Figure 1.7b - p. 8). We observe that all of the features discussed above are determined on the deconvolved data Figure 1.7b - p. 8 and none of them are related to the non-deconvolved data in Figure 1.7a - p. 8.

Contribution

The contribution of this thesis is developed along two axes. Firstly, we contribute to the seismic multiple removal problem. After the primaries reflections are recovered, we contribute secondly to deconvolution problem that produces characteristics closest to the Earth layers. More precisely, the specific contributions are as follows:

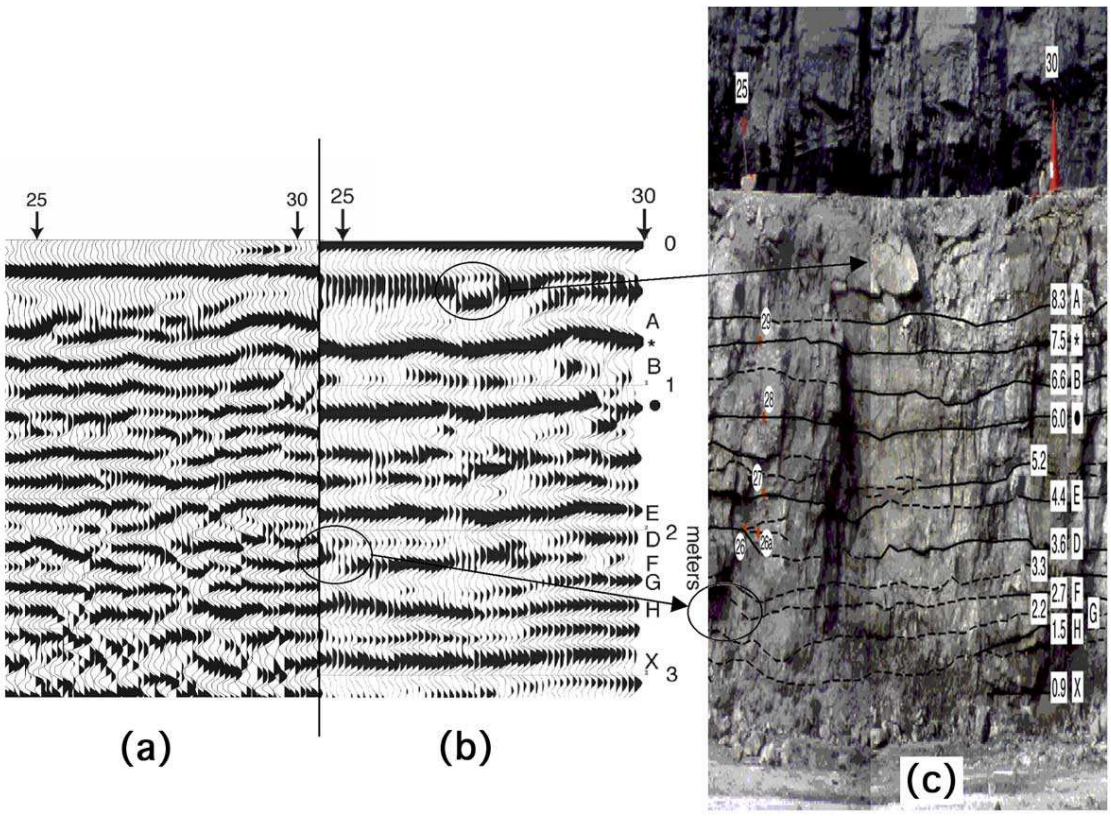


Figure 1.7: *An enlarged portion from the 25 to 30 meter locations on the bench. (a) Non-deconvolved GPR data, (b) deconvolved GPR data, and (c) the photo of the quarry face with an exaggerated vertical scale to match that of the GPR data. (Source:* $http://www.kgs.ku.edu/Geophysics/OFR/2004/OFR04\_41$*.)*

Seismic multiple removal

- We propose a generic methodology to impose sparsity and regularity properties through constrained adaptive filtering in a frame transformed domain. This method exploits side information from approximate disturbance templates.

- We use the proximal framework that permits different strategies for sparse modeling, additive noise removal, and adaptive filter design under appropriate regularity and amplitude coefficient concentration constraints.

- We use 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.

- We focus on the adaptation to the anisotropy of two-dimensional seismic data by using shift-invariant tight frames, and directional dual-tree M-band wavelets.
- The approach demonstrates significantly good performance in low signal-to-noise (SNR) ratio conditions, both for simulated and real field data.

Blind deconvolution

- We propose a new penalty based on a smooth approximation to the ℓ_1/ℓ_2 function.
- We develop a proximal-based algorithm to solve variational problems involving this function and we derive theoretical convergence results.
- We demonstrate the effectiveness of our method through a comparison with a recent alternating optimization strategy dealing with the exact ℓ_1/ℓ_2 term.
- Experiments are performed on seismic and image blind deconvolution.

Publications

Journal papers:

1) M.-Q. Pham, L. Duval, C. Chaux and J.-C. Pesquet, "A Primal-Dual Proximal Algorithm for Sparse Template-Based Adaptive Filtering: Application to Seismic Multiple Removal", IEEE Transactions on Signal Processing, vol. 62, No. 16, Aug. 2014, pp. 4256–4269.

2) A. Repetti, M.-Q. Pham, L. Duval, E. Chouzenoux and J.-C. Pesquet, "Euclid in a Taxicab: Sparse Blind Deconvolution with Smoothed ℓ_1/ℓ_2 Regularization", IEEE Signal Processing Letters, vol. 22, No. 5, May 2015, pp. 539–543.

Patent:

3) C. Chaux, L. Duval, J.-C. Pesquet, M.-Q. Pham, "Procédé de traitement de réflexions multiples", France, 12/02.299. Aug. 24, 2012.

Invited conference paper:

4) M.-Q. Pham, C. Chaux, L. Duval and J.-C. Pesquet, "A constrained-based optimization approach for seismic data recovery", Special session on "SS5 - Seismic Signal Processing", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP, May 4-9, 2014), Florence, Italy, pp. 2377–2381.

Regular conference papers:

5) M.-Q. Pham, C. Chaux, L. Duval et J.-C. Pesquet, "Seismic multiple removal with a primal-dual proximal algorithm", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP, May 26-31, 2013), Vancouver, Canada, pp. 2257–2261.

6) M.-Q. Pham, C. Chaux, L. Duval et J.-C. Pesquet, "Filtrage de multiples sismiques par ondelettes et optimisation convexe", (GRETSI, September 3-6, 2013), Brest, France.

7) M.-Q. Pham, C. Chaux, L. Duval et J.-C. Pesquet, "Sparse adaptive template matching and filtering for 2D seismic images with Dual-tree wavelets and proximal methods", IEEE International Conference on Image Processing (ICIP, September 27-30, 2015), Québec, Canada.

Talks:

1) Mai Quyen Pham, "Représentations parcimonieuses pour la restauration et l'analyse quantitative de champs d'ondes en sismiques", IFPEN, France, April 09, 2012.

2) Mai Quyen Pham, "Proximal methods for multiple removal in seismic data", I2M Marseille, France, March 28, 2014.

3) Mai Quyen Pham, "Multiple removal in seismic data", Imaging and Modeling Unit, Institut Pasteur, France, November 07, 2014.

4) Mai Quyen Pham, "Sparse adaptive template matching and filtering for 2D seismic images with dual-tree Wavelets and proximal methods", Créatis, INSA Lyon, France, December 12, 2014.

Organization of the Thesis

This thesis is divided into 7 chapters. Following this introduction, the second chapter is devoted to introduce the reader to the topic of inverse problems and some mathematical tools, i.e. variational methods applied to continuous models. We briefly discuss line search, majorize-minimize and proximal optimization methods. Then, the subsequent Chapters 3-6 address seismic inverse problems. More precisely, in Chapter 3 we develop an optimization framework applied to the seismic multiple removal problem. This method exploits side information from the knowledge of one template and in the non–noisy case. In Chapter 4, we extend the seismic multiple removal problem to the noisy case and exploit several templates. In Chapter 5, we develop the framework of seismic multiple removal for two-dimensional data. Chapter 6 provides an insight with ℓ_1/ℓ_2 sparsity measure to blind deconvolution applied to both seismic data and image processing. Finally, Chapter 7 is a summary of this thesis, including the conclusions and perspectives which can be drawn from this research work. Concluding this thesis is the bibliography.

- Chapter 2 -

Background

2.1 Inverse problems

2.1.1 What are inverse problems?

Solving an inverse problem is often required when recovering information about a physical object from its measurements acquired by a given system. In a direct problem we start from the causes and end up with the results, in an inverse problem we start with partial knowledge of the causes and the results and infer more about the causes. For example, in geophysics in order to predict an earthquake, one can estimate the vibrations produced by an earthquake as well as its propagation by observing the structure of the earth at different points around the globe. Figure 2.1 - p. 12 illustrates an earthquake propagating through the planet. The inverse vision of this is to use earthquake measurements to determine the Earth structure.

Inverse problems appear in signal processing, image analysis, computer vision, and many other applications. In this thesis, we are particularly interested in solving inverse problems related to deconvolution and multiple removal in seismic data and in image processing. For both applications, we can formulate the inverse problem as

$$y = Hx + b \tag{2.1}$$

where $H : \mathbb{R}^N \to \mathbb{R}^M$ is a bounded linear operator, $N \in \mathbb{N}^*$ and $M \in \mathbb{N}^*$ are the sizes of signal x and y, respectively. $b \in \mathbb{R}^M$ is the additive noise perturbation. In many cases, the inverse problem is *ill-posed*, which is the opposite of a well-posed problem. Let us recall the definition of a well-posed problem proposed in (Hadamard, 1923),

Definition 2.1.1 (Hadamard, Figure 2.2 - p. 13) *A problem is called* well-posed *in the sense of Hadamard if*

1. *there exists a solution to the problem (existence),*

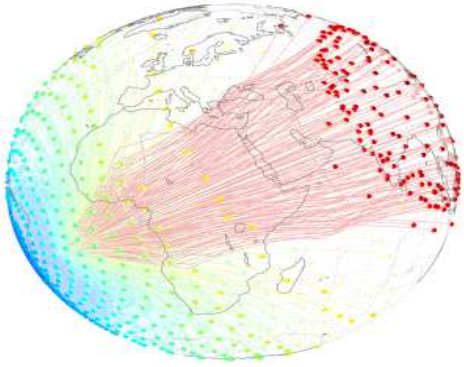


Figure 2.1: *Waves from an earthquake propagating through the planet (Source: http://www.bath.ac.uk/elec-eng/invert/inverse.html).*

2. *there is at most one solution to the problem (uniqueness),*

3. *the solution depends continuously on the data (stability).*

By contrast, a problem is ill-posed when at least one of the properties of a well-posed problem is violated. Hadamard gave an example of an ill-posed problem named the Cauchy problem for Laplace equation.

Example 2.1.1 (Cauchy problem for Laplace equation) (Kabanikhin, 2008) Let $u = u(x, y)$ be a solution to the following problem:

$$\begin{cases} \Delta u(x,y) &= 0 \ , \quad y \in \mathbb{R}, \quad x > 0, \\ u(0,y) &= f(y), \quad y \in \mathbb{R}, \\ \frac{\partial u}{\partial x}(0,y) &= 0 \ , \quad y \in \mathbb{R}. \end{cases} \tag{2.2}$$

Let f be chosen as follows

$$f(y) = u(0,y) = \frac{1}{n}\sin(ny).$$

The solution to Problem (2.2) is given by

$$u(x,y) = \frac{1}{n}\sin(ny)(e^{nx} + e^{-nx}). \tag{2.3}$$

For any fixed $x > 0$ and a sufficiently large n, the value of the solution (2.3) can be as large as desired, while $f(y) \to 0$ as $n \to \infty$. Therefore, small variations in y may lead to arbitrary large variations in the solution, which means that Problem (2.2) is ill-posed.

Nowadays, research about *well-posed* and *ill-posed* inverse problems remains a major issue in many applications. Let us present some approaches for solving these problems, starting with basic ideas, toward more sophisticated ones.

Figure 2.2: *Jacques Hadamard.*

2.1.2 Variational approach

Least squares yield standard solutions to inverse problems because they lead to simple computations. The first statement of the method appeared in Legendre's *"Nouvelles méthodes pour la détermination des orbites des comètes"*, Paris, 1805 (or new Methods for the Determination of Comet Orbits). In this text, Legendre writes: *"De tous les principes qu'on peut proposer pour cet objet, je pense qu'il n'en est pas de plus général, de plus exact, ni d'une application plus facile que celui dont nous avons fait usage dans les recherches précédentes, et qui consiste à rendre minimum la somme des quarrés des erreurs."*– Legendre (1805) (or of all the principles which can be proposed for that purpose, I think there is none more general, more exact, and more easy of application, that of which we made use in the preceding researches, and which consists of rendering the sum of squares of the errors a minimum.). It can also be reformulated as follows:

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \|Hx-y\|^2. \tag{2.4}$$

In the case of an ill-posed problem, (2.4) provides a misbehaved solution. In order to give preference to a particular solution with desirable properties, an extended method based on the generalized Tikhonov regularization is proposed by A.-N. Tikhonov in (Tikhonov and Arsenin, 1977). By adding a regularization term in this minimization, (2.4) now becomes

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \|Hx-y\|^2+\lambda\|x\|^2 \tag{2.5}$$

where $\lambda \in]0,+\infty[$ is named regularization parameter. The choice of the parameter plays an important role in regularization methods and the way of choosing the regularization parameter has attracted many works (Cucker and Smale, 2002; Renaut et al., 2010). In

order to choose the value of these parameters, we first need to understand them. Bayesian estimators give the relationship between these parameters and signal probabilistic models as described below.
Let assume that x is a realization of random vector X with prior probability density function f_X and y is a realization of random vector Y with conditional likelihood $f_{X|Y=y}$. X and Y are supposed independent. Assuming that a desired property of a solution is known a priori, the solution of the inverse problem is given by using the maximum a posteriori (MAP) estimator:

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\operatorname{argmax}}\ f_{X|Y=y}(x). \tag{2.6}$$

By using Bayes rule, this can be simplified as follows:

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\operatorname{argmax}}\ f_{Y|X=x}(y)\frac{f_X(x)}{f_Y(y)} \tag{2.7}$$

$$\Leftrightarrow \widehat{x} \in \underset{x\in\mathbb{R}^N}{\operatorname{argmax}}\ f_{Y|X=x}(y) f_X(x) \tag{2.8}$$

provided that $f_Y(y) \neq 0$. Due to the monotonicity of the logarithm function this can be rewritten as:

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\operatorname{argmin}}\ \Psi(x) + \Phi(x) \tag{2.9}$$

where

- $\Psi(x) = -\log(f_{Y|X=x}(y))$ is called a data fidelity term. For example, when B is a zero-mean Gaussian noise with a covariance matrix Γ, then B has a probability density function defined by the normal distribution i.e.

$$f_{Y|X=x}(y) = \frac{1}{(2\pi)^{N/2}|\Gamma|^{1/2}} e^{-\frac{1}{2}(y-Hx)^\top \Gamma^{-1}(y-Hx)} \tag{2.10}$$

$$\Rightarrow \Psi(x) = \frac{1}{2}(y-Hx)^\top \Gamma^{-1}(y-Hx) - \log\left(\frac{1}{(2\pi)^{N/2}|\Gamma|^{1/2}}\right). \tag{2.11}$$

  When $\Gamma = \sigma^2\,\mathrm{I}$, without loss of generality, we can choose

$$\Psi(x) = \frac{1}{2\sigma^2}\|y-Hx\|^2 \tag{2.12}$$

  Figure 2.3 - p. 15 shows an original image (left) and this image corrupted by Gaussian noise (right).

- $\Phi(x) = -\log(f_X(x))$ is called a regularization term. When $f_X(x) = e^{-\lambda\|x\|^2}$, then

$$\Phi(x) = \lambda\|x\|^2 \tag{2.13}$$

Figure 2.3: *Original image (left); Image corrupted by Gaussian noise (right)*

Legendre's method or Tikhonov's method are two particular cases of Bayesian estimation without/with *a priori* on the solution, respectively. Usually, the considered criterion to be minimized can take a more general form (than simply quadratic); however, it is often assumed to be convex. Let us recall some definitions about convexity.

Definition 2.1.2 *A subset C of $\mathbb{R}^N$ is said to be* convex *if either $C = \emptyset$ or, whenever we take two points in C, the segment that connects them is entirely contained in C, i.e.,*

$$\forall \lambda \in [0,1],\ \forall (x,y) \in C^2, \quad \lambda x + (1-\lambda) y \in C$$

Figure 2.4a - p. 16 illustrates a convex set. The black line segment joining points x and y lies within the green set and this is true for all points x and y in the green set. Figure 2.4b - p. 16 shows a nonconvex set.

Definition 2.1.3 *Let $C \in \mathbb{R}^N$. The* support function *of C is defined as*

$$\sigma_C : \mathbb{R}^N \to]-\infty, +\infty] : u \mapsto \sup \langle C | u \rangle \,. \tag{2.14}$$

Definition 2.1.4 *Let $f : \mathbb{R}^N \to]-\infty, +\infty]$. The* domain *of f is*

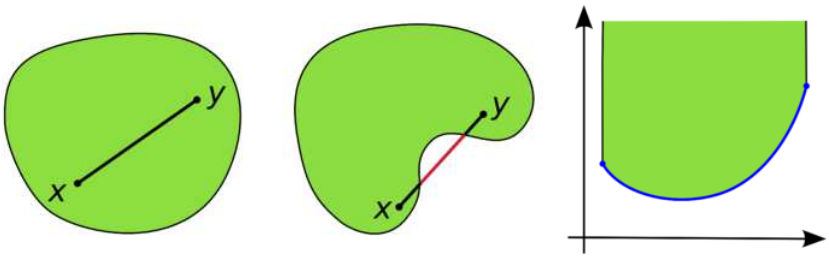
$$\operatorname{dom} f = \left\{ x \in \mathbb{R}^N \mid f(x) < +\infty \right\}, \tag{2.15}$$

the graph *of f is defined as*

$$\operatorname{gra} f = \left\{ (x, \xi) \in \mathbb{R}^N \times \mathbb{R} \mid f(x) = \xi \right\} \tag{2.16}$$

the epigraph *of f is defined as*

$$\operatorname{epi} f = \left\{ (x, \xi) \in \mathbb{R}^N \times \mathbb{R} \mid f(x) \leq \xi \right\}. \tag{2.17}$$



(a) (b) (c)

Figure 2.4: *A convex set (a), a nonconvex set (b) and a convex function*

Definition 2.1.5 *Let* $f : \mathbb{R}^N \to]-\infty, +\infty]$. *The function* f *is* convex *(see Figure 2.4c - p. 16) if its epigraph is a convex subset of* $\mathbb{R}^N \times \mathbb{R}$. *Moreover,* f *is* concave *(or nonconvex) if* $-f$ *is convex.*

Definition 2.1.6 *Let* $f : \mathbb{R}^N \to]-\infty, +\infty]$, *and let* $x \in \mathbb{R}^N$. *Then* f *is* lower semicontinuous *at* x *if, for every sequence* $(x_n)_{n\in\mathbb{N}}$ *in* $\mathbb{R}^N$,

$$x_n \to x \quad \Rightarrow f(x) \leq \liminf f(x_n).$$

Upper semicontinuity *of* f *at* $x \in \mathbb{R}^N$ *holds if* $-f$ *is lower semicontinuous at* x *i.e. if*

$$x_n \to x \quad \Rightarrow f(x) \geq \limsup f(x_n).$$

Definition 2.1.7 $\Gamma_0(\mathbb{R}^N)$ *is the class of lower semi-continuous convex functions* $f : \mathbb{R}^N \to]0, +\infty[$ *such that their domain dom* f *is nonempty.*

In geophysics, the remnant noise is commonly considered to be Gaussian. Thus a quadratic data-fidelity term has been successfully used (i.e. $\Psi = \frac{1}{2\sigma^2}\|\cdot\|^2$, where σ^2 corresponds to the noise variance). We now focus on the data regularization term $\Phi(\cdot)$. In the following, we discuss the sparse representation of data in general and then about seismic data, in particular.

2.2 Sparse representation of signals and images

2.2.1 Introducing sparsity

In signal restoration, the sparsity expressed on data gradient (finite differences) is often used. It is called total variation prior (TV) in image restoration, and it was first used by Rudin, Osher and Fatemi (Rudin et al., 1992). The Hessian based regularization was already discussed in the work by (Geman and Reynolds, 1992; Yunpeng Li, 2008). In

the context of this thesis, we focus in particular on the sparse signal representation in a basis or frame domain. The key idea is to approximate a signal $x \in \mathbb{R}^N$ by a sparse decomposition in a sequence of vectors $\mathcal{F} = \{e_k\}_{k\in\mathbb{L}}$ of waveforms, where $e_k \in \mathbb{R}^N$ or $\mathbb{C}^N$. Thus, we can express x as

$$x = \sum_{k\in\mathbb{L}} c_k e_k \tag{2.18}$$

and the signal is now characterized by fewer significant coefficients c_k. If the set $\mathcal{F}$ does not span $\mathbb{R}^N$ or $\mathbb{C}^N$, then such coefficients do not exist for every x. If $\mathcal{F}$ spans $\mathbb{R}^N$ or $\mathbb{C}^N$ and also is linearly independent, this set forms a *basis* of $\mathbb{R}^N$ or $\mathbb{C}^N$, and the coefficients c_k are uniquely determined by x: they are the components of x relative to this basis. If, however, $\mathcal{F}$ spans $\mathbb{R}^N$ or $\mathbb{C}^N$ but is not linearly independent, this set forms a *frame* of $\mathbb{R}^N$ or $\mathbb{C}^N$ if it satisfies the so-called *frame condition*: There exist two real numbers, $\underline{\nu}$ and $\overline{\nu}$ such that $0 < \underline{\nu} \leq \overline{\nu} < +\infty$ and

$$\underline{\nu}\|x\|^2 \leq \sum_{k\in\mathbb{L}} |\langle x, e_k\rangle|^2 \leq \overline{\nu}\|x\|^2 \quad \text{for all } x \in \mathbb{R}^N \tag{2.19}$$

This means that the constants $\underline{\nu}$ and $\overline{\nu}$ are independently of a specific x: they only depend on the set $\mathcal{F}$. The numbers $\underline{\nu}$ and $\overline{\nu}$ are called lower and upper *frame bounds* respectively. Note that the frame bounds $\underline{\nu}$ and $\overline{\nu}$ are not unique because numbers less than $\underline{\nu}$ and greater than $\overline{\nu}$ are also considered as frame bounds. If $\underline{\nu} = \overline{\nu}$, then a frame is said *tight* and if $\underline{\nu} = \overline{\nu} = 1$, then a frame is either called *normalized* or *Parseval*. The *optimal upper bound* is the infimum of all upper bounds. Likewise, the *optimal lower bound* is the supremum of all lower bounds.

Wavelet bases and frames constitute an important class of bases and frames in general and particularly in this thesis. Thus, let us give a short overview of them. Let $\mathcal{L}^2(\mathbb{R})$ denote the space of all square integrable functions on $\mathbb{R}$ i.e.

$$\text{if } \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty \tag{2.20}$$

then f is square integrable on $\mathbb{R}$ (or $f \in \mathcal{L}^2(\mathbb{R})$). ψ is said to be a *wavelet* if and only if its Fourier transforms (FT) $\hat{\psi}$ satisfies

$$0 < C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty, \tag{2.21}$$

which implies that the wavelet must have a zero mean

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \tag{2.22}$$

and therefore it must be a sort of a *wave*. The dilated-translated wavelets $\psi_{a,b}$ are defined as the following functions

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \tag{2.23}$$

where $b \in \mathbb{R}$ is a translation parameter, whereas $a \in \mathbb{R}_+^*$ is dilation or scale parameter. The factor $\frac{1}{\sqrt{a}}$ is a normalization constant such that the energy i.e. $\|\psi_{a,b}\|_{L_2} = 1$. The *continuous wavelet transform* (CWT) of $x \in \mathcal{L}^2(\mathbb{R})$ is defined as

$$\begin{aligned} W_x(a,b) = \langle x, \psi_{a,b} \rangle &= \int_{-\infty}^{+\infty} x(t)\psi_{a,b}^*(t)dt \\ &= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-b}{a} \right) dt \end{aligned} \tag{2.24}$$

where $\langle\, , \rangle$ is the scalar product of $\mathcal{L}^2(\mathbb{R})$. x can be reconstructed from its wavelet coefficients

$$x(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a^2} W_x(a,b)(t)dadb \tag{2.25}$$

However, in practice a data is represented by a finite number of values, then it is important to consider a discrete version of the CWT. Let consider a signal x of finite energy that can be written as follows

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} c_{m,n} \widetilde{\psi}_{m,n}(t) \tag{2.26}$$

and the coefficients of the expansion are given by

$$c_{m,n} = \int_{-\infty}^{+\infty} x(t) \widetilde{\psi}_{m,n}(t)dt \tag{2.27}$$

where $\widetilde{\psi}_{m,n}$ with $(m,n) \in \mathbb{Z}^2$ are all distillations and translations of a function referred to the as analyzing wavelet $\psi(t)$, and they can be expressed in the form

$$\widetilde{\psi}_{m,n}(t) = \psi_{a^m, na^m b}(t) = a^{-m/2} \psi(a^{-m}t - nb) \tag{2.28}$$

where m and n denote the dilation and translation indices, respectively. Figure 2.5 - p. 19 shows the Ricker (left) and Symlet (right) wavelets with different scalings a. We next show an illustration of sparsity in the wavelet transform domain used to perform signal denoising. In the space $\mathbb{R}^N$, considering a clean signal x, we add some noise to it in order to obtain the noisy signal $y = x + b$, here b is a realization of a Gaussian white noise with variance σ^2 . Figure 2.6 - p. 20 depicts x and its noisy version y. To analyze the performance of wavelet transform, we now compare the estimation reconstructed by a Wiener filter, and by thresholding the wavelet coefficients of the noisy signal.

Signal denoising using a Wiener filter (Mallat, 2009, Chapter 10) can be seen in Figure 2.8-left - p. 22.

We now use a thresholding estimator to reconstruct the signal. Let $F \in \mathbb{R}^{K \times N}$ (resp. $F^\top \in \mathbb{R}^{N \times K}$) denotes the frame analysis (resp. synthesis) operator. We restrict here F to correspond to an orthonormal transform, i.e. $F^\top = F^{-1}$. (Donoho et al., 1995) proved that the performance of a thresholding estimator is close to ideal coefficient selections

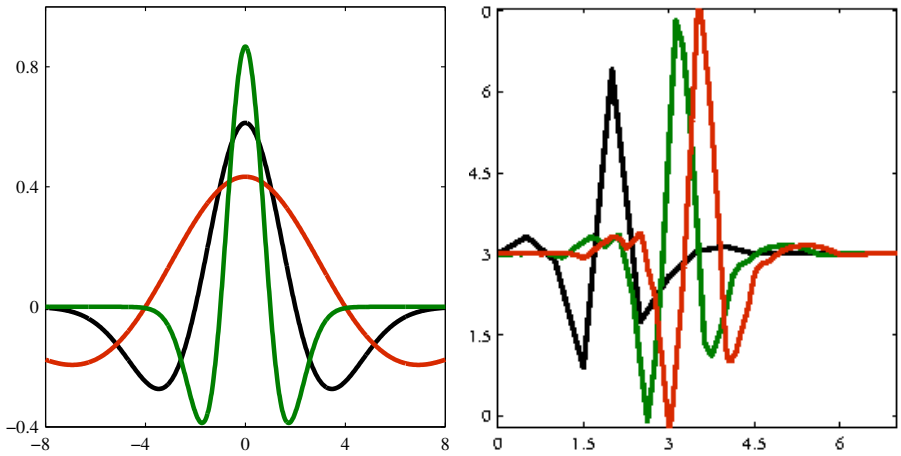


Figure 2.5: *Ricker (left) and Symlet (right) wavelet with different a*

and attenuations. The vector of wavelet coefficients $c \in \mathbb{R}^K$ of the signal $y \in \mathbb{R}^N$ is given by $c = Fy$. Figure 2.7 - p. 21 illustrates the wavelet coefficients of the Piece-Polynomial signal. Symlet wavelets of length 8 over 5 resolution levels are used.

The soft thresholder on $[-T, T]$ for some $T \in]0, +\infty[$ is given by

$$S_T : \mathbb{R} \to \mathbb{R} : \zeta \mapsto \operatorname{sign}(\zeta) \max\{0, |\zeta| - T\}. \tag{2.29}$$

$T \in]0, +\infty[$ is the threshold. An estimate $\hat{x}$ of x can be obtained by applying the soft thresholding to the wavelet coefficients that is $\hat{x} = F^{-1}(S_T(Fy))$.

The result of the thresholding operation on the noisy Piece-Polynomial signal is given in Figure 2.8-right - p. 22. The performance of the wavelet transform is shown by a comparative evaluation of two results (obtained by using either Wiener filter or thresholding the wavelet coefficients). Apparently, in Figure 2.8 - p. 22, a slightly better result is obtained with thresholding the wavelet coefficients. Figure 2.9 - p. 23 shows the comparative results for two dimensional data. The "regularization" parameters of both methods are adjusted so as to maximize the SNR between the original and the reconstructed data.

2.2.2 Evaluations of the sparsity

Let c be a sparse representation of x via a linear operator L, we thus note $c = Lx$. According to (Donoho et al., 1995), an ideal measure of sparsity Φ is the ℓ_0 measure, i.e.

$$\ell_0(c) = \sum_{i \in \mathbb{L}} \delta(c_i) \tag{2.30}$$

$$\text{where} \quad \delta(c_i) = \begin{cases} 0 & \text{if } c_i = 0 \\ 1 & \text{otherwise.} \end{cases}$$

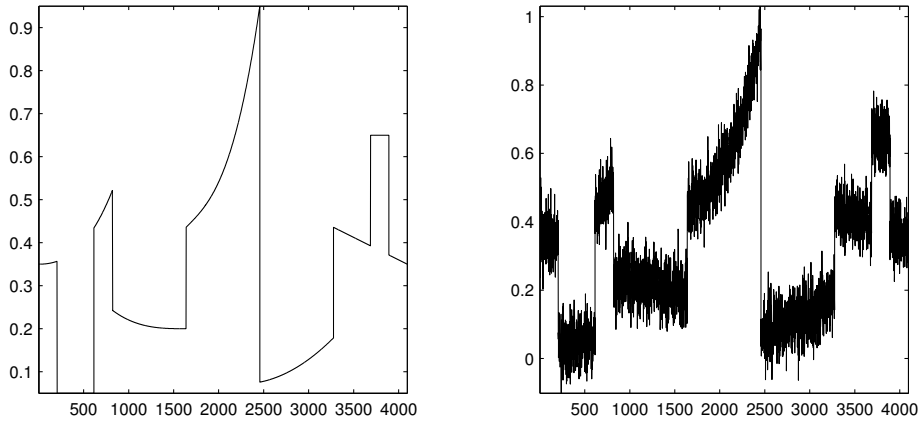


Figure 2.6: *Piece-Polynomial signal (left); Noisy signal (right)* SNR= 17.90dB

However, ℓ_0 is a nonconvex, discrete valued and non-differentiable function. In this case, the problem is not easy to resolve. To circumvent this problem, different measures of sparsity have been proposed in the regularization terms. For instance, quasi-norms are the most widely used measures, where $\ell_p(c) = \left(\sum_{i\in\mathbb{L}} |c_i|^p\right)^{1/p}$, for every $0 < p \leq 1$. (Kowalski, 2009) has investigated mixed $\ell_{p,q}$ loss functions for deconvolution. A detailed discussion about these loss functions is provided in Chapters 3 and 4. In Chapter 6 we propose to use a smooth approximation of the ℓ_1/ℓ_2 ratio function.

As the inverse problem is formulated as a minimization problem, the next question to address is: how to solve minimization problems? In order to answer this question, many approaches have been proposed. Recently, tools for convex optimization, named proximal algorithms, have been proposed. These algorithms are well known for their good convergence properties, and for being able to be parallelized or to be implemented in parallel.

2.3 Algorithms

We consider an optimization problem in the standard form:

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ f(x) \quad \text{subject to} \quad \begin{cases} g_i(x) \leq 0, & \forall i \in \mathcal{E} \\ g_i(x) = 0, & \forall i \in \mathcal{I} \end{cases} \tag{2.31}$$

where f and g_i are scalar-valued functions of the variable x, and $\mathcal{I}$, $\mathcal{E}$ are sets of indices. This problem is a constrained optimization problem if $\mathcal{E} \neq \emptyset$ or $\mathcal{I} \neq \emptyset$, otherwise, it is unconstrained and reads

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ f(x) \tag{2.32}$$

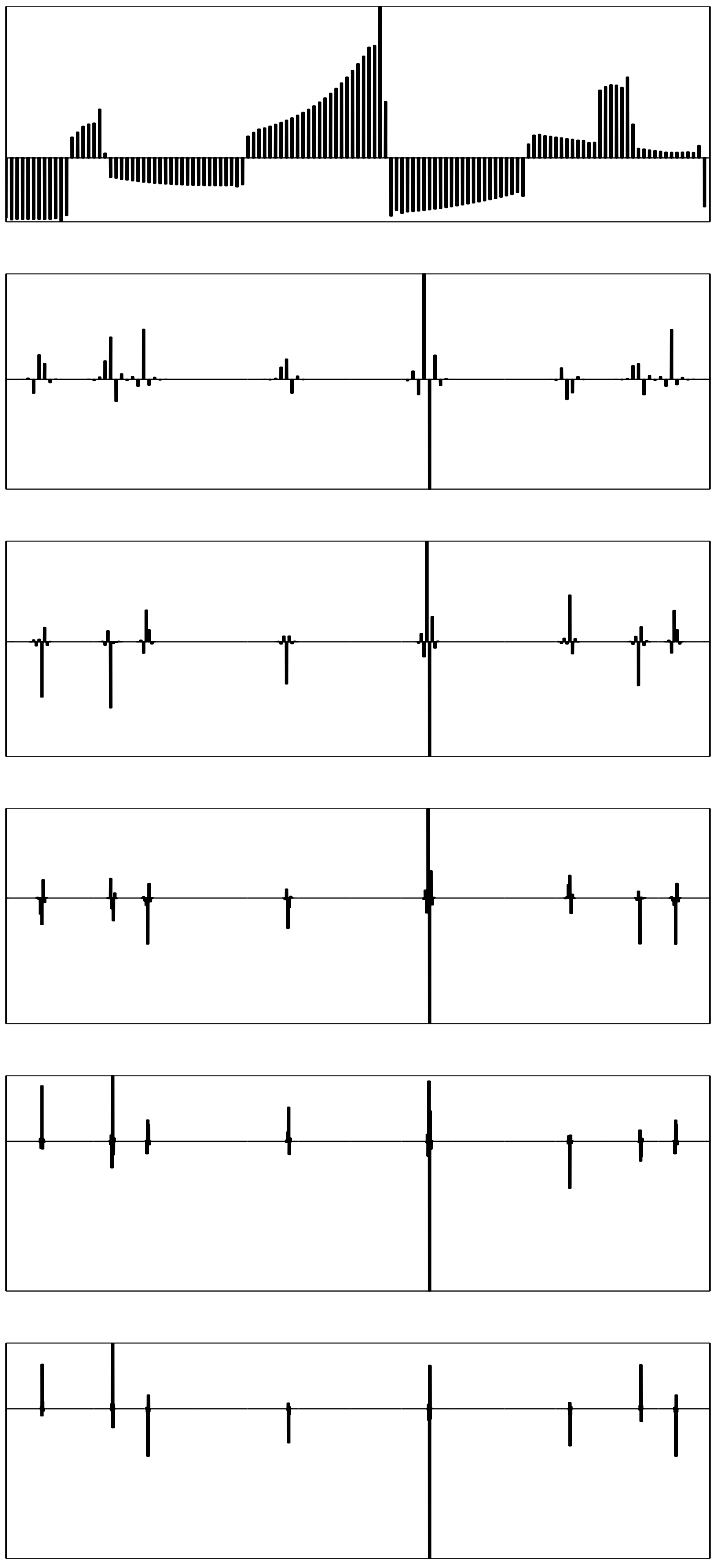


Figure 2.7: *Wavelet coefficients of Piece-Polynomial signal*

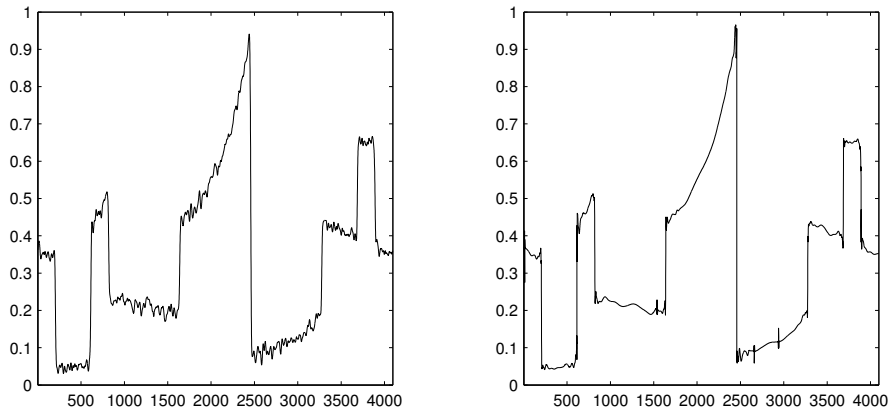


Figure 2.8: *Estimation by using Wiener filter, with* $\lambda = \sigma^2$, SNR= 24.30 dB *(left); estimation with a translation invariant soft thresholding,* SNR= 31.50 dB *(right)*

We define $\left(\forall x \in \mathbb{R}^N\right) \quad C_i = \begin{cases} \{x|\, g_i(x) \leq 0\} & \text{if } i \in \mathcal{E} \\ \{x|\, g_i(x) = 0\} & \text{if } i \in \mathcal{I}. \end{cases}$ We now recall the definition of the indicator function:

Definition 2.3.1 (Indicator function) *Let* C *be a nonempty subset of* $\mathbb{R}^N$*. Its indicator function* ι_C *is defined as* $\left(\forall x \in \mathbb{R}^N\right) \quad \iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$

Thus, we can write (2.31) in the following form

$$\underset{x \in \mathbb{R}^N}{\text{minimize}}\ f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \iota_{C_i}(x) \tag{2.33}$$

Consider a simple example in 2-dimension, we then show a geometrical representation of this optimization problem.

Example 2.3.1

$$\underset{(x_1,x_2)\in\mathbb{R}^2}{\text{minimize}}\ (x_1-2)^2 + (x_2-1)^2 \quad \text{subject to} \begin{cases} x_1^2 - x_2 \leq 0, \\ -3x_1 - 2x_2 \leq 2 \\ (2-\sqrt{3})x_1 + x_2 \leq 2\sqrt{3} \end{cases} \tag{2.34}$$

we can write this problem in the form (2.31) by defining

$$f(x) = (x_1-2)^2 + (x_2-1)^2, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

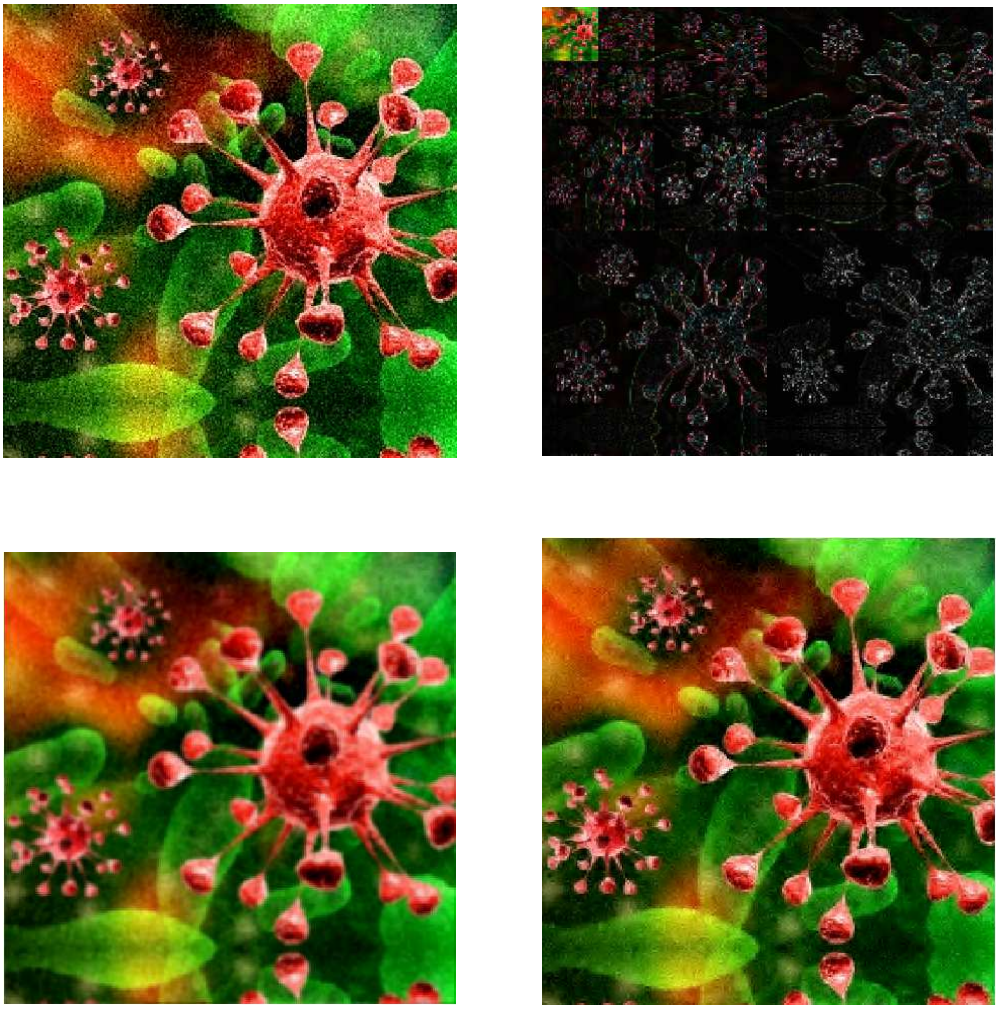


Figure 2.9: *Top: Noisy image* SNR= 14.73 dB *(left); Wavelet coefficients (right), Bottom: Estimation by using Wiener filter, with* $\lambda = 0.7$, SNR= 17.17 dB *(left); Estimation with a translation invariant soft thresholding,* SNR= 18.55 dB *(right)*

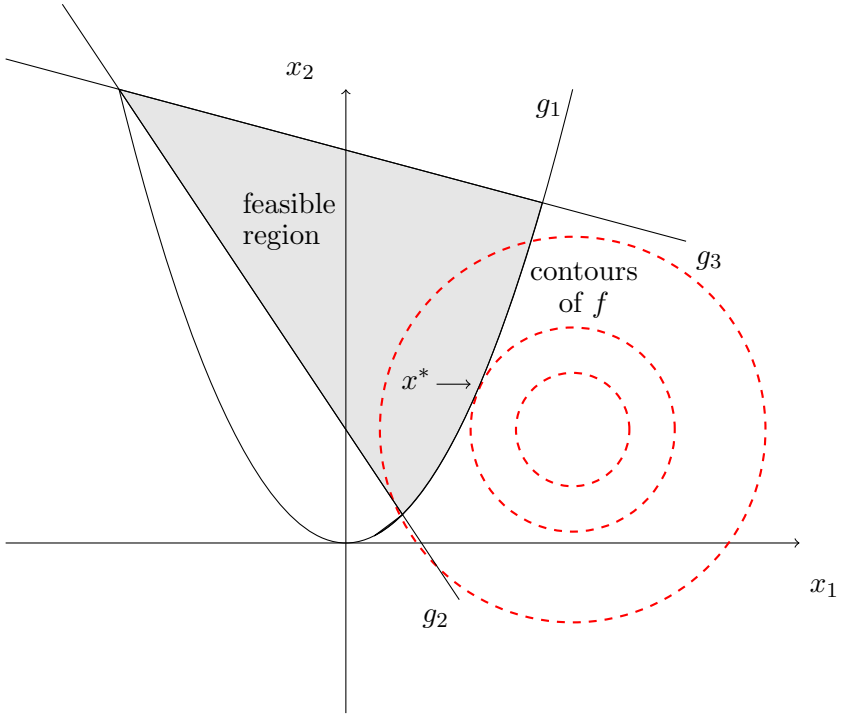


Figure 2.10: *Geometrical representation of* (2.34).

$$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ g_3(x) \end{bmatrix} = \begin{bmatrix} x_1^2 - x_2 \\ -3x_1 - 2x_2 - 2 \\ (2-\sqrt{3})x_1 + x_2 - 2\sqrt{3} \end{bmatrix}, \quad \mathcal{I} = \{1,2,3\}, \quad \mathcal{E} = \emptyset.$$

Figure 2.10 - p. 24 shows the contours of the cost function, i.e., the set of points for which f has a constant value. It also illustrates the feasible region, which is the set of points satisfying all the constraints, and the optimal point x^*, solution to the problem. Note that the "feasible region" of the inequality constraints is shaded.

Optimization algorithms are iterative. More precisely, the strategy used is to move from an initial point to another point which is closer to the solution in using the values of the cost function f and the constraints g. To move from a point to another, we use, for instance, the first order derivatives of these functions and the local information from the current point. In the following, some basic criteria for algorithm evaluation can be considered

- Robustness: for the choices of the initial variables.
- Efficiency: time to converge.

Starting at x_0, optimization algorithms generate a sequence of iterates $\{x_k\}_{k\in\mathbb{N}}$. To move from one iterate x_k to the next, this sequence is often satisfy $f(x_{k+1}) \leq f(x_k)$.

The strategies used for generating it strongly depend on the mathematical properties of the function f. The algorithm stops at the iterate K such that $x_{K+1} \approx x_K$. For instance, there are three fundamental strategies that we will detail in the next sections.

2.3.1 Line search method

Such algorithms choose a direction p_k and search along this direction from the current point x_k for a new iterate with a lower function value. Thus the iteration is given by

$$\forall k \in \mathbb{N}, \quad x_{k+1} = x_k + \alpha_k p_k \tag{2.35}$$

where the positive scalar α_k is called the step-length. The iteration depends on choices of both the direction p_k, ($\|p_k\| = 1$) and the step-length. Using a Taylor's expansion, for any search direction p and step-length parameter α, we have:

$$f(x_k + \alpha p) = f(x_k) + \alpha p^\top \nabla f(x_k) + \frac{1}{2}\alpha^2 p^\top \nabla^2 f(x_k + tp)p, \quad \text{for some } t \in (0, \alpha)$$

f can be reduced along the direction p at x_k when p is the minimum of $p^\top \nabla f(x_k)$. Since $p^\top \nabla f(x_k) = \|p\|\|\nabla f(x_k)\| \cos(\theta) = \|\nabla f(x_k)\| \cos(\theta)$, where θ is the angle between p and $\nabla f(x_k)$, a good direction p is $-\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}$, when $\theta = \pi$. Another search direction method which moves along $p_k = -\nabla f(x_k)$ at every step, is named the steepest descent method (Deift and Zhou, 1993). One of the most famous method is the Newton direction (Nocedal and Wright, 2006) that is derived from the second-order Taylor series. Thus the Newton direction p_k is $-\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$. An extension of this method, called quasi-Newton direction (Nocedal and Wright, 2006), uses an approximation instead of the true Hessian $\nabla^2 f(x_k)$. The search direction of nonlinear conjugate gradient methods (Dai and Yuan, 1999) is $p_k = -\nabla f(x_k) + \beta_k p_{k-1}$, where β_k is a scalar that ensures that p_k and p_{k-1} are collinear. Once we have a good direction, the efficiency of the iterative algorithm depends on the step α, since the new point is given by $x_k + \alpha p_k$. In other word, we need to find α that minimizes the following function:

$$\alpha \in \underset{\alpha \in \mathbb{R}_+}{\operatorname{argmin}} \; f(x_k + \alpha p_k) \tag{2.36}$$

In order to insure the convergence of the algorithm, at each direction p_k we would like to find an exact minimum of (2.36). This leads to the following approximation method.

2.3.2 Majorize-Minimize method (MM)

The main idea of Majorize-Minimize (MM) method is to construct a *majorant function* $m(x_k, x)$ such that m is simpler in seeking a minimizer and very close to the cost function f at a neighborhood of x_k. Thus instead of seeking the minimum of f, we now restrict the search for a minimizer of $m(x_k, x)$ to some region around x_k i.e.

$$x_{k+1} \in \underset{x}{\operatorname{minimize}} \; m(x_k, x) \tag{2.37}$$

where $x = x_k + p$ lies inside the trust region, usually, $\|p\| \leq \delta_k$, $\delta_k > 0$ is called the trust-region radius. The majorant m usually satisfies the following rules

$$\forall x \in \mathbb{R}^N \text{ and } x_k \in \mathbb{R}^N \quad f(x) \leq m(x_k, x) \tag{2.38}$$

and

$$\forall x_k \in \mathbb{R}^N \quad f(x_k) = m(x_k, x_k) \tag{2.39}$$

Indeed (2.38) and (2.39) imply that, for every k

$$f(x_{k+1}) \leq m(x_k, x_{k+1}) \leq m(x_k, x_k) = f(x_k). \tag{2.40}$$

Figure 2.11 - p. 27 illustrates a simple example of MM methods for a one dimensional signal. The majorant is often chosen under a quadratic form

$$m(x_k, x_k + p) = f(x_k) + p^\top \nabla f(x_k) + \frac{1}{2} p^\top M_k p, \tag{2.41}$$

where the matrix M_k is either the Hessian $\nabla^2 f(x_k)$ or an approximation of it. The solution of (2.37) in this case is $p_k = -M_k^{-1} \nabla f(x_k)$. We can see that the Newton and quasi-Newton's methods belong to the class of MM methods in the convex case. The majorant functions can be iteration dependent as in (Jacobson and Fessler, 2007). Many frameworks about the convergence are available in the literature. One global convergence for strictly convex, coercive and differentiable f was proposed in (Lange, 2010, Chapter 10.3). In (Chouzenoux et al., 2011), the authors proposed accelerated subspace optimization methods. At each iteration of such methods, a stepsize vector allowing the best combination of several search directions is computed through multi-dimensional search based on the MM principle. Their method is significantly faster than the Newton methods in terms of computational time before convergence. Concerning the nonconvex case of the cost function, the local convergence was studied in (Chouzenoux et al., 2014) and their convergence speed is discussed in (Chouzenoux et al., 2013). Next we describe a recent strategy for deriving an optimization algorithm relying on proximal minimization.

2.3.3 Parallel-Proximal method

In the previous methods, the iteration point is often computed by using the gradient of a differentiable function e.g. the gradient methods $x_{k+1} = x_k - \lambda_k \nabla f(x_{k+1})$. When f is non differentiable then the problem (2.31) is said nonsmooth. In this case, the gradient can be replaced by the subdifferential.

Definition 2.3.2 *The subdifferential (Rockafellar and Wets, 2005) of a function $f \in \Gamma_0(\mathbb{R}^N)$ at $x \in \mathbb{R}^N$, denoted by $\partial f(x)$, is defined as:*

$$\partial f(x) := \left\{ x^{'} \in \mathbb{R}^N \mid \forall x^{''} \in \mathbb{R}^N,\ f(x) + x^{'\top}(x^{''} - x) \leq f(x^{''}) \right\} \tag{2.42}$$

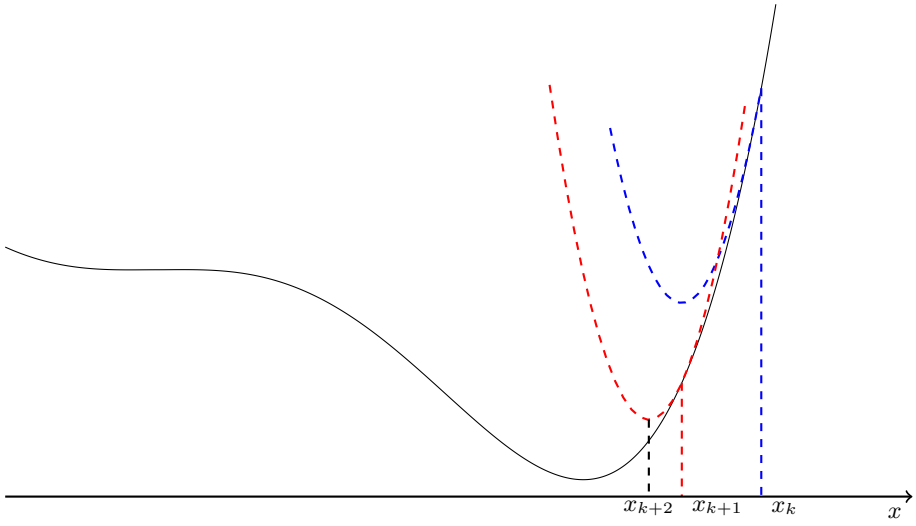


Figure 2.11: *Illustration of the MM methods.*

which implies that

$$\begin{aligned} & x_{k+1} = x_k - \lambda_k t_k^{'}, \quad t_k^{'} \in \partial f(x_{k+1}) \\ \Leftrightarrow \quad & 0 \in x_{k+1} - x_k + \lambda_k \partial f(x_{k+1}) \\ \Leftrightarrow \quad & 0 \in \partial \left(\frac{1}{2} \| \cdot - x_k \|^2 + \lambda_k f(\cdot) \right) (x_{k+1}) \end{aligned} \tag{2.43}$$

(2.43) has a unique solution in the convex case (due the Fermat's rule) named $\text{prox}_{\lambda_k f}(x_k)$.

Definition 2.3.3 *The proximity operator (Rockafellar, 1976) of a function $f \in \Gamma_0(\mathbb{R}^N)$ at $x \in \mathbb{R}^N$, denoted by $\text{prox}_f : \mathbb{R}^N \to \mathbb{R}^N$, is defined as:*

$$\forall p \in \mathbb{R}^N, \ \text{prox}_f(p) := \underset{x \in \mathbb{R}^N}{\text{minimize}} \ \frac{1}{2} \|x - p\|^2 + f(x) \tag{2.44}$$

Example 2.3.2 (ι_C function) If C is a closed non empty and convex subset of $\mathbb{R}^N$

$$\begin{aligned} \text{prox}_{\iota_C}(x) &= \underset{y \in \mathbb{R}^N}{\text{argmin}} \ \iota_C(y) + \frac{1}{2} \|x - y\|^2 \\ &= \underbrace{\underset{y \in C}{\text{argmin}} \ \|x - y\|^2}_{\Pi_C(x): \text{projection operator onto } C} \end{aligned} \tag{2.45}$$

Example 2.3.3 (ℓ_1-norm)

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\lambda|\cdot|}(x) = \underbrace{\text{sign}(x) \max\left(|x| - \lambda, 0\right)}_{\text{soft thresholding}}$$

Example 2.3.4 (ℓ_2^2-norm)

$$(\forall x \in \mathbb{R}^N) \quad \operatorname{prox}_{\lambda\|\cdot\|^2}(x) = \underbrace{\frac{1}{1+2\lambda}}_{\text{Wiener-like filter}} x \quad .$$

The proximity operator is non-expansive, i.e. $\|\operatorname{prox}_f(x) - \operatorname{prox}_f(x')\| \leq \|x - x'\|$ for every $(x, x') \in \mathbb{R}^N \times \mathbb{R}^N$. We can see that i) proximity operator is a generalization of orthogonal projections and ii) if we define $g(x_k, x) = \frac{1}{2}\|x - x_k\|^2 + f(x)$, thus $g(x_k, x)$ can be seen as a majorant function of f at x_k. The proximal algorithms are especially useful for solving large-size optimization problems, when the cost function f is the sum of M functions $(f_m)_{1\leq m\leq M}$. We first discuss the case $M = 2$ and f_1, f_2 are in $\Gamma_0(\mathbb{R}^N)$. The forward-backward algorithm has been proposed when f_2 is differentiable with a Lipschitz continuous gradient. In (Chen and Rockafellar, 1997) the authors used the subdifferential leading to the following update rule $x_{k+1} = x_k - \gamma_k(t_k' + \nabla f_2(x_k)), \quad t_k' \in \partial f_1(x_{k+1})$. In (Combettes and Wajs, 2005) the update rule has been reformulated using the proximity operator which results in $x_{k+1} = x_k + \lambda_k(\operatorname{prox}_{\gamma_k f_1}(x_n - \gamma_k \nabla f_2(x_k)) - x_k)$. This class of algorithms (Nesterov, 2013; Bioucas-Dias and Figueiredo, 2007; Beck and Teboulle, 2009) focuses on the convergence speed, while (Tseng, 2000; Raguet et al., 2013), proposed an extension to more general cases. However if both functions to be minimized are nondifferentiable, then the Douglas-Rachford algorithms proposed in (Lions and Mercier, 1979; Eckstein and Bertsekas, 1992; Combettes and Pesquet, 2007a) may be used. In the case when $M > 2$ one can resort to the Parallel Proximal Algorithm (PPXA) proposed in (Combettes and Pesquet, 2008) that minimizes a sum of a finite number of functions in $\Gamma_0(\mathbb{R}^N)$ or its generalized version (PPXA+) in (Pesquet and Pustelnik, 2012) that minimizes a sum of a finite number of functions in $\Gamma_0(\mathbb{R}^N)$ composed with linear operators. We use PPXA+ in Chapter 3. PPXA+ requires to be able to invert some linear operator (consisting of a sum of linear operators), which may not be straightforward to compute. To tackle this problem, we should also mention the Primal-dual techniques proposed in (Combettes and Pesquet, 2012; Vũ, 2013; Komodakis and Pesquet, 2014; Alghamdi et al., 2014; Combettes et al., 2014). These algorithms are designed to jointly solve the primal and dual formulations of an optimization problem by invoking the Fenchel duality theorem.

Definition 2.3.4 *The Fenchel-Rockafellar duality (Bauschke and Combettes, 2011, Chapter 15) theorem states that under some technical conditions for any convex functions $f_1 : \mathbb{R}^N \to \mathbb{R}$ and $f_2 : \mathcal{V} \to \mathbb{R}$:*

$$\underbrace{\inf_{x\in\mathbb{R}^N} \{f_1(x) + f_2(Vx)\}}_{\textit{Primal problem}} = \underbrace{\sup_{v\in\mathcal{V}} \{-f_1^*(V^*v) - f_2^*(-v)\}}_{\textit{Fenchel-Rockafellar dual problem}} \tag{2.46}$$

where $f^* : \mathbb{R}^N \to \mathbb{R}$ is the convex-conjugate of $f : \mathbb{R}^N \to \mathbb{R}$ (Bauschke and Combettes, 2011, Chapter 13) i.e.

$$\forall u \in \mathbb{R}^N, \quad f^*(u) = \sup_{x\in\mathbb{R}^N} (\langle x, u\rangle - f(x)) \tag{2.47}$$

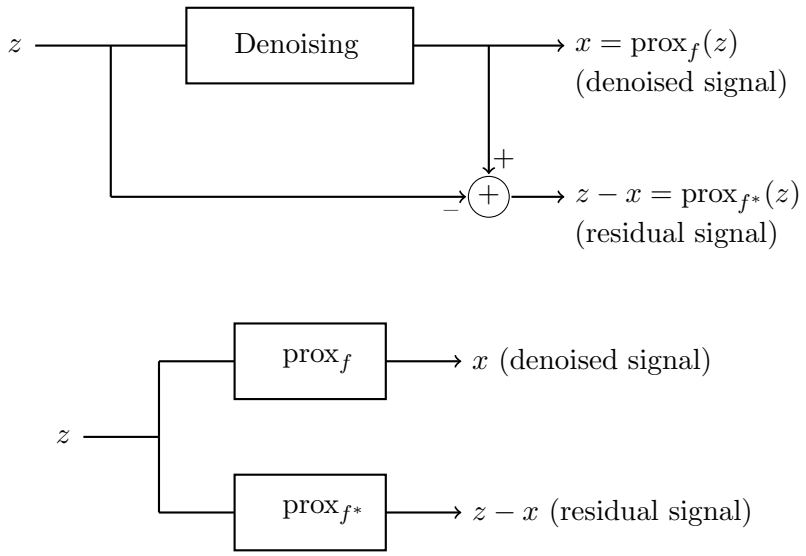


Figure 2.12: *Moreau's decomposition schema.*

$V : \mathbb{R}^N \to \mathcal{V}$ is a linear operator and V^* is its adjoint. In such settings, primal and dual variables refer to x and v, respectively.

(Bauschke and Combettes, 2011, Chapter 14, Theorem 14.3) states that for any convex function $f : \mathbb{R}^N \to \mathbb{R}$ and for any $\gamma \in \mathbb{R}_+^*$, then

$$\mathrm{Id} = \mathrm{prox}_{\gamma f} + \gamma \, \mathrm{prox}_{f^*/\gamma}(\cdot/\gamma) \,. \tag{2.48}$$

We give one example of a denoising problem to illustrate this theorem.

Example 2.3.5 (Denoising) We consider the following denoising problem

$$z = x + b$$

where z corresponds to the observed data, x is the original signal and the noise b which is assumed to be additive zero-mean white Gaussian. We can reformulate the problem as an optimization problem

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \; f(x) + \frac{1}{2}\|z - x\|^2$$

where regularization term $f \in \Gamma_0(\mathbb{R}^N)$ serves to enforce prior knowledge about the solution. We know that this problem has a unique solution $x = \mathrm{prox}_f(z)$, then z can be decomposed as in Figure 2.12 - p. 29.

The algorithms based on primal-dual techniques are also used in this thesis. In general, proximity methods are known for their good convergence properties. Most of the results

are established for convex functions. However, some results in the non convex case can be found in (Kaplan and Tichatschke, 1998; Bolte et al., 2007; Attouch et al., 2013; Attouch and Bolte, 2010; Sra, 2011; Chouzenoux et al., 2014, 2013). In particular, we use the algorithm in (Chouzenoux et al., 2013) to solve the problem of blind deconvolution in Chapter 6.

Property	$f(x)$	$\text{prox}_f(x)$
i, translation	$\varphi(x-z),\ z \in \mathbb{R}^N$	$z + \text{prox}_\varphi(x-z)$
ii, scaling	$\varphi(x/\gamma),\ \gamma \in \mathbb{R} \setminus 0$	$\gamma\, \text{prox}_{\varphi/\gamma^2}(x/\gamma)$
iii, reflection	$\varphi(-x)$	$-\text{prox}_\varphi(-x)$
iv, quadratic function	$\frac{1}{2}\gamma\|Lx-y\|^2$ $L \in \mathbb{R}^{M\times N}, \gamma > 0, y \in \mathbb{R}^M$	$(\mathrm{I}+\gamma L^\top L)^{-1}(x+\gamma L^\top y)$
v, quadratic perturbation	$\varphi(x) + \alpha\|x\|^2/2 + u^\top x + \gamma$ $u \in \mathbb{R}^N, \alpha \geq 0, \gamma \in \mathbb{R}$	$\text{prox}_{\varphi/(\alpha+1)}\left((x-u)/(\alpha+1)\right)$
vi, semi-orthogonal linear transform	$\varphi(Lx)$ $L \in \mathbb{R}^{M\times N}, LL^\top = \nu\,\mathrm{I}, \nu > 0$	$x + \nu^{-1}L^\top\left(\text{prox}_{\nu\varphi}(Lx) - Lx\right)$
vii, indicator function	$\iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases}$	$\Pi_C(x)$

Table 2.1: *Properties of proximity operators (Combettes and Pesquet, 2011):* $\varphi \in \Gamma_0(\mathbb{R}^N)$*;* $C \subset \mathbb{R}^N$ *is nonempty, closed, and convex;* $x \in \mathbb{R}^N$*.*

$\phi(x)$	$\text{prox}_\phi(x)$
i, $\iota_{[\underline{\omega},\overline{\omega}]}(x)$	$\Pi_{[\underline{\omega},\overline{\omega}]}(x)$
ii, $\omega\|x\|$	$\text{sign}(x)\max\{\|x\|-\omega, 0\}$
iii, $\omega\|x\|^{4/3}$	$x + \frac{4\omega}{32^{1/3}}\left((\beta-x)^{1/3} - (\beta+x)^{1/3}\right)$, where $\beta = \sqrt{x^2 + 256\omega^3/729}$
iv, $\omega\|x\|^{3/2}$	$x + 9\omega^2\,\text{sign}(x)\left(1 - \sqrt{1+16\|x\|/(9\omega^2)}\right)/8$
v, $\omega\|x\|^2$	$x/(1+2\omega)$
vi, $\omega\|x\|^3$	$\text{sign}(x)\left(\sqrt{1+12\omega\|x\|}-1\right)/(6\omega)$
vii, $\omega\|x\|^4$	$\left(\frac{\beta+x}{8\omega}\right)^{1/3} - \left(\frac{\beta-x}{8\omega}\right)^{1/3}$, where $\beta = \sqrt{x^2 + 1/(27\omega)}$
viii, $\max\{\|x\|-\omega, 0\}$	$\begin{cases} x & \text{if } \|x\| < \omega \\ \text{sign}(x)\omega & \text{if}\omega \leq \|x\| \leq 2\omega \\ \text{sign}(x)(\|x\|-\omega) & \text{if}\|x\| > 2\omega \end{cases}$

Table 2.2: *Proximity operators of* $\phi \in \Gamma_0(\mathbb{R}^N)$*;* $\alpha \in \mathbb{R}$*,* $\omega > 0$*,* $\underline{\omega} < \overline{\omega}$ *(Combettes and Pesquet, 2011).*

Table 2.1 - p. 30 shows many additional properties of the proximity operator. One will find in Table 2.2 - p. 30 closed-form expressions of the proximity operators of various functions belonging to $\Gamma_0(\mathbb{R})$. These two tables contain some important properties that

we will use in the next chapters of the thesis.

2.4 Conclusion

In this chapter, we first provided an introduction to the topic of inverse problems and several associated mathematical tools. Indeed, we introduced the concept of sparsity and sparse representation, and gave some useful algorithms to solve such inverse problems, possibly favoring sparsity. In the next chapters, we will focus on two inverse problems of particular interest in seismic data processing: multiple removal and blind deconvolution.

- Chapter 3 -

Multiple removal with a single template

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 can be obtained through time-varying filtering, compensating inaccuracies in given approximate templates. We propose here a new variational framework based on a Maximum A Posteriori estimation procedure. More precisely, the problem of multiple removal is formulated as a minimization problem involving time-varying filters, assuming that a single disturbance signal template is available and the target signal is sparse in some decomposition e.g. an orthonormal basis or a frame. We show that estimating these multiples is equivalent to estimating filters and we propose to employ recently proposed convex optimization procedures based on proximity operators to solve the problem. In this chapter we consider the problem for the case where the structured noise is only composed of one type of multiple reflections and without additional noise. This work was initiated during the internship of D. Gragnaniello (summer 2011).

3.1 Introduction

As already detailed in the introduction, 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. Seismic waves 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.

We address here one of the most severe types of interferences: secondary reflections, named multiples, corresponding to seismic waves bouncing between layers (Essenreiter

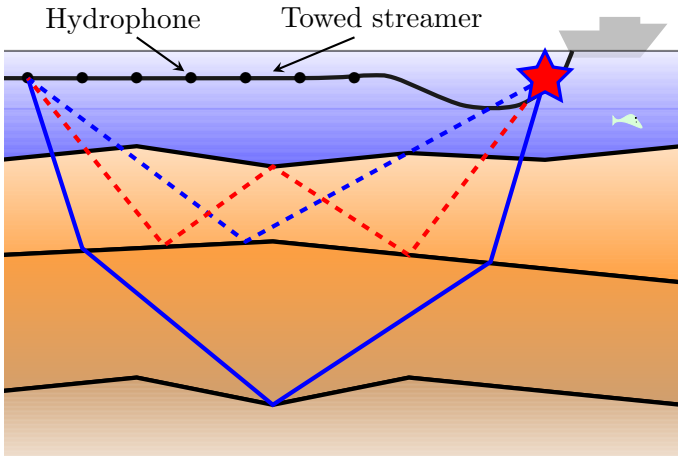


Figure 3.1: *Principles of marine seismic data acquisition and wave propagation. Towed streamer with hydrophones. Reflections on different layers (primaries in blue), and reverberated disturbance (multiple in dashed red).*

et al., 1998), as illustrated with red dashed lines in Figure 3.1 - p. 34. 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 realistic template of a potential multiple is 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 — template should be adapted in a time-varying fashion before being subtracted from the recorded data. Since the data recovery problem is under-determined, geophysicists have developed pioneering sparsity-promoting techniques. Recently, several works in geophysics have revisited the use of non-quadratic criteria, including Huber function (Guitton and Symes, 2003) or $\ell_p (p \in [1,2])$-norms (Costagliola et al., 2011), due to the alleged non-gaussianity of seismic data (Walden, 1985). However, robust, ℓ_1-promoted deconvolution (Claerbout and Muir, 1973) or sparseness measure ℓ_1/ℓ_2 (Zibulevsky and Pearlmutter, 2001; Hoyer, 2004; Hurley and Rickard, 2009; Barak et al., 2014; Gray, 1978; Repetti et al., 2015), or complex wavelet transforms (Morlet, 1975) still pervade many areas of signal processing.

The proposed method is based on a single approximate template, in the absence of noise. This framework addresses at the same time structured reverberations. Namely, let $n_t \in \{0, \dots, N_t - 1\}$ 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:

$$z^{(n_t)} = \overline{y}^{(n_t)} + \overline{s}^{(n_t)} \,. \tag{3.1}$$

The unknown signal of interest (primary, in blue) and the sum of undesired, secondary reflected signals (multiples, in red) are denoted, respectively, by $\overline{y} = (\overline{y}^{(n_t)})_{0 \leq n_t < N_t}$ and $\overline{s} = (\overline{s}^{(n_t)})_{0 \leq n_t < N_t}$. We assume that an approximate template accounting for multiples is available.

This chapter presents a relatively novel framework for multiple reflection filtering with (*i*) data fidelity term taking into account the statistical properties of the basis or frame coefficients, (*ii*) prior information on the filter, and (*iii*) slow variation modelling of the adaptive filter.

In Section 3.2, we analyze related works and specify the novelty of the proposed methodology. Section 3.3 describes the transformed linear model incorporating the template with adaptive filtering. In Section 3.4, we formulate a generic variational form for the problem. We discuss some proximal algorithms for solving our problem in Section 3.5. The slowly-varying property of the filter is to be taken into account and its performance is assessed in Section 3.6. The methodology is evaluated on a realistic synthetic data model. We draw some conclusions in section 3.7.

3.2 Related and proposed 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. Several common processing routines transform the data into a new domain in which they minimize overlap between primaries and multiples. Transformed domain filter subsequently attenuate multiples or select primaries of interest. The filtered data is finally mapped back to data domain, using an appropriate inverse transform. Many standard transform methods are used as homomorphic filtering (Buttkus, 1975), $f-k$ (Wu and Wang, 2011), and $\tau - p$ (Nuzzo and Quarta, 2004) or alternative breeds — parabolic, hyperbolic – of the Radon transform (Hampson, 1986; Trad et al., 2003; Nowak and Imhof, 2006). See more details for these methods in (Verschuur, 2006). Among the vast literature, we refer to (Weglein et al., 2011; Ventosa et al., 2012) 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 Figure 3.1 - p. 34) suggests a modeling of primary reflection coefficients with generalized Gaussian or Cauchy distributions (Walden, 1985; Walden and Hosken, 1986), 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 (Kaplan

and Innanen, 2008; Lu and Liu, 2009; Liu and Dragoset, 2012; Duarte et al., 2012; Takahata et al., 2012; da Costa Filho, 2013) 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 (Bois and Hémon, 1963), a handful of works have investigated processing in appropriate transformed domains. They may either stationarize (Krim and Pesquet, 1995) or strengthen data sparsity. For instance, (Donno, 2011) applies ICA in a dip-separated domain. In (Neelamani et al., 2010), as well as in (Herrmann and Verschuur, 2004) and subsequent works by the same group, a special focus is laid on separation in the curvelet domain.

Data sparsity alone may not be sufficient to solve (3.1). Additional constraints reduce the set of solutions, hopefully to geologically sounder ones. A first one is the locality of matched filter, traditional in standard multiple filtering. These can be modeled by Finite Impulse Response (FIR) operators. Classical filter support limitations, down to one-tap (Neelamani et al., 2010; Ventosa et al., 2012), assorted with ℓ_2 or ℓ_1 criteria, are standard. Yet, no work in multiple removal has endeavored a more systematic study of variational and sparsity constraints on the adaptive filter, in the line of (O'Brien et al., 1994). In this work, we propose a formulation allowing a family of penalties to be applied to the adaptive FIR filter. In standard restoration (Pesquet et al., 2009), knowledge about the degradation kernel is often required. It is replaced here by the knowledge of a template, the degradation kernel being estimated. The resulting algorithm recursively estimates an FIR filter. The propagation medium, as well as the modeled template, carries 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 filter to exhibit bounded variations from one time index to the next one. This hypothesis is consistent with wave propagation assumptions. More precisely, the estimation problem is formulated as a convex variational problem involving a non-smooth cost function.

3.3 Model description

We assume that a multiple template is 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 one template $(r^{(n_t)})_{0\leq n_t<N_t}$, related to $(\overline{s}^{(n_t)})_{0\leq n_t<N_t}$ via a possibly non-causal linear model through a limited support relationship:

$$\overline{s}^{(n_t)} = \sum_{p=p'}^{p'+P-1} \overline{h}^{(n_t)}(p) r^{(n_t-p)} \tag{3.2}$$

where $\overline{h}^{(n_t)}$ is an unknown finite impulse response (with P tap coefficients) associated with time n_t, and where $p' \in \{-P+1, \ldots, 0\}$ is its starting index ($p' = 0$ corresponds to the causal case). It must be emphasized that the dependence with respect to the time index n_t of the impulse responses implies that the filtering process is time variant,

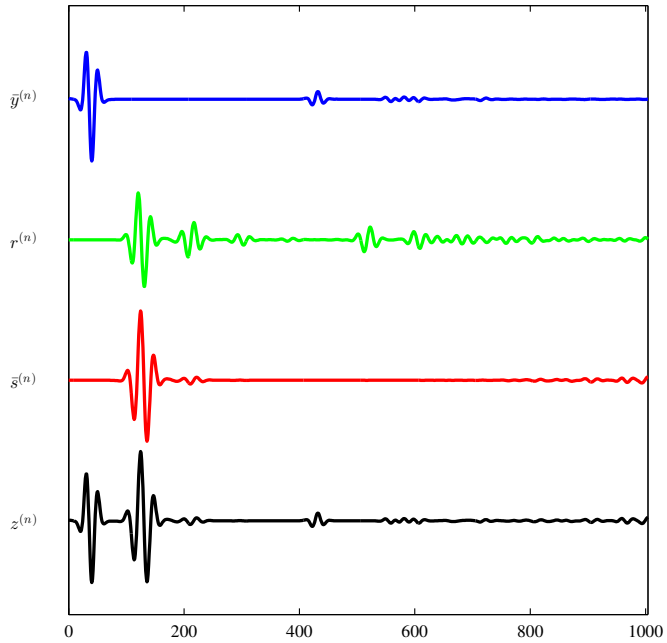


Figure 3.2: *Considered simulated seismic signals. From top to bottom: primary (unknown) $\overline{y}$, template r, multiple (unknown) $\overline{s}$ and observed signal z.*

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. A template is 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 filter might be ill-behaved, due to the band-pass nature of seismic data is well known, although rarely documented, motivating the need for filter coefficient control. Figure 3.2 - p. 37 shows an example of synthetic data. We define vectors $\overline{s}$ and $\overline{h}$ by:

$$\overline{s} = \left[\overline{s}^{(0)} \quad \cdots \quad \overline{s}^{(N_t-1)}\right]^\top, \tag{3.3}$$

$$\overline{h} = \left[\overline{h}^{(0)}(p') \cdots \overline{h}^{(0)}(p'+P-1) \cdots \overline{h}^{(N_t-1)}(p') \cdots \overline{h}^{(N_t-1)}(p'+P-1)\right]^\top, \tag{3.4}$$

and

$$R = \begin{bmatrix} R^{(0)} & 0 & \dots & 0 \\ 0 & R^{(1)} & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & R^{(N_t-1)} \end{bmatrix}, \tag{3.5}$$

where $(R^{(n_t)})_{0\leq n_t\leq N_t-1}$ are vectors of dimension P such that

$$\begin{bmatrix} R^{(0)} \\ R^{(1)} \\ \vdots \\ R^{(N_t-1)} \end{bmatrix} = \begin{bmatrix} r^{(-p')} & \cdots & r^{(0)} & 0 & \cdots & & 0 \\ r^{(-p'+1)} & \cdots & & r^{(0)} & 0 & \cdots & 0 \\ \vdots & & & & & & \\ r^{(N_t-1)} & r^{(N_t-2)} & & \cdots & & & r^{(N_t-P)} \\ 0 & r^{(N_t-1)} & & \cdots & & & r^{(N_t-P+1)} \\ \vdots & & & & & & \\ 0 & \cdots & 0 & r^{(N_t-1)} & \cdots & & r^{(N_t-P-p')} \end{bmatrix}. \tag{3.6}$$

One can note that the matrix R is a block diagonal matrix and that the concatenation of its block diagonal elements is a Toeplitz matrix of size $N_t \times P$. With this formulation, the problem of providing an estimate $\widehat{y}$ of the primary turns out to be equivalent to computing an estimate $\widehat{h}$ of the impulse response and to deduce $\widehat{y} = z - R\widehat{h}$.

3.4 Proposed variational approach

Let us now show how the problem can be addressed from a Bayesian perspective. We assume that the characteristics of the primary are appropriately described through a prior statistical model in a different representation, e.g. a wavelet basis/frame (Mallat, 2009). For example, if we denote by x the vector of frame coefficients and $F \in \mathbb{R}^{K \times N_t}$ designates the associated analysis operator, we have (Chaux et al., 2007a)

$$x = Fy. \tag{3.7}$$

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

$$(\forall y \in \mathbb{R}^{N_t}) \quad f_Y(y) \propto \exp(-\varphi(Fy)) \tag{3.8}$$

where $\varphi : \mathbb{R}^{N_t} \to]-\infty, +\infty]$ is the associated potential, assumed to be such that

$$(\forall x \in \mathbb{R}^{K}) \quad \varphi(x) \geq \alpha \|x\|^{2\theta} \tag{3.9}$$

where $\alpha \in]0, +\infty[$ and $\theta \in]0, +\infty[$. This allows us to deduce that

$$(\forall y \in \mathbb{R}^{N_t}) \quad \varphi(Fy) \geq \alpha \nu^{\theta} \|y\|^{2\theta} \tag{3.10}$$

where $\nu \in]0, +\infty[$ is the lower frame bound. This in turn implies that f_Y as defined in (3.8) is a valid pdf.

For simplicity, φ can be chosen to be separable, which corresponds to an independence assumption on the coefficients when the frame reduces to a basis:

$$(\forall x = (x_k)_{1 \leq k \leq K} \in \mathbb{R}^{K}) \quad \varphi(x) = \sum_{k=1}^{K} \varphi_k(x_k). \tag{3.11}$$

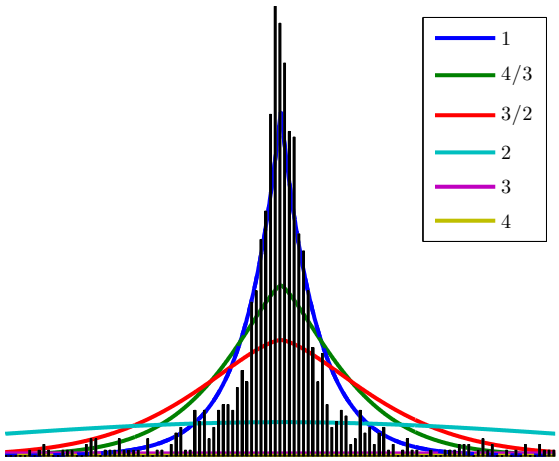


Figure 3.3: *Generalized Gaussian modeling of seismic data wavelet basis decomposition with different power laws.*

where, for every $k \in \{1, \cdots, N_t\}$, $\varphi_k : \mathbb{R} \to]-\infty, +\infty]$. In order to promote the sparsity of the decomposition, a standard choice for this potential is $\varphi_k = \kappa_k |\cdot|$ where $\kappa_k > 0$. (This choice satisfies (3.9) with $\alpha = \min_{1 \leq k \leq K} \kappa_k$ and $\theta = 1/2$).

On the other hand, to take into account the available information on the unknown filter, especially its regular variations along the time dimension, it can be assumed that $\overline{h}$ is a realization of a random vector H, the pdf of which is expressed as

$$(\forall h \in \mathbb{R}^{N_t P}) \quad f_H(h) \propto \exp(-\rho(h)), \tag{3.12}$$

and which is independent of y. By resorting to an estimation of h in the sense of the Maximum A Posteriori (MAP), the problem can be formulated as

$$\underset{h \in \mathbb{R}^{N_t P}}{\text{minimize}}\ \varphi(F(z - Rh)) + \rho(h). \tag{3.13}$$

In this approach, φ represents some data fidelity term taking into account the statistical properties of the basis coefficients and ρ models prior informations that are available on h.

3.5 Filter estimation

3.5.1 Algorithm

The objective here is to perform the minimization of a criterion composed of only two functions:

$$\underset{h \in \mathbb{R}^{N_t P}}{\text{minimize}}\ \Phi(h) + \rho(h) \tag{3.14}$$

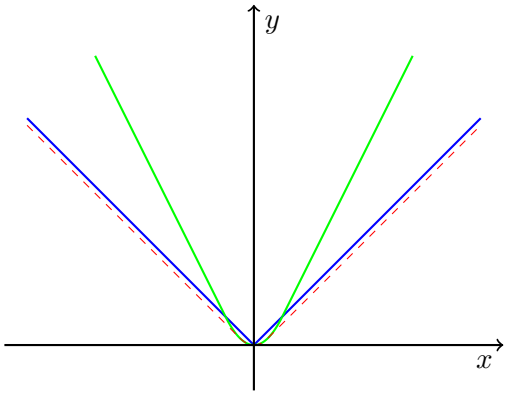


Figure 3.4: *ℓ_1-norm (in blue), Huber function with $\chi = 8$ and $\tau = 1$ (in green) and Huber function with $\chi = 8$ and $\tau = \frac{1}{2}$ (in dashed red).*

where, $\Phi = \varphi(Fz - \Omega\cdot)$, and $\Omega = FR$. The panel of available methods appears to be very large and we have to take into account the linear operator R, in this case matrix R verify $R^\top R = \mathrm{Diag}(\|R^{(0)}\|^2, \ldots, \|R^{(N-1)}\|^2)$. As mentioned in the background (see Chapter 2), the Douglas-Rachford (DR) algorithm (Combettes and Pesquet, 2007a) can be used, however this algorithm requires to be able to compute the proximity operators of Φ and ρ. Function Φ contains the linear operator Ω which drastically reduces the choice one can make for φ e.g. when φ is a ℓ_2-norm and F is a wavelet orthogonal basis, then DR can be applied by using iv, in Table 2.1 - p. 30. However, when φ is a ℓ_2-norm and F is a wavelet orthogonal basis matrix then

$$\varphi(F(z - Rh)) = \|F(z - Rh)\|^2 = (z - Rh)^\top F^\top F(z - Rh) = \|z - Rh\|^2. \tag{3.15}$$

This implies that an orthonormal wavelet basis is uneffective in this case.

Algorithm 1 Forward-Backward algorithm

Set $h^{[0]} \in \mathbb{R}^{N_t P}$
for $i = 0, 1, \ldots$ **do**
  $\widetilde{h}^{[i]} = h^{[i]} - \gamma^{[i]} \nabla\Phi(h^{[i]})$
  $h^{[i+1]} = h^{[i]} + \lambda^{[i]}(\mathrm{prox}_{\gamma^{[i]}\rho}(\widetilde{h}^{[i]}) - h^{[i]})$
end for

We now turn our attention to the Forward-Backward (FB) splitting algorithm. We assume that φ has a μ-Lipschitz continuous gradient, then the function Φ also has a β-Lipschitz continuous gradient with $\beta = \mu\|\Omega\|^2$. This algorithm implies to be able to compute the gradient of Φ and the proximity operator of ρ. At each iteration i, the step-size $\gamma^{[i]}$ and the relaxation parameter $\lambda^{[i]}$ must be chosen so as to satisfy the following rule: let $\epsilon \in]0, \min\{1, 1/\beta\}[$, $\gamma^{[i]} \in [\epsilon, 2/\beta - \epsilon]$, and $\lambda^{[i]} \in [\epsilon, 1]$. We choose to use a wavelet orthogonal basis with 8-length Symlet filters, performed on 4 resolution levels.

Figure 3.3 - p. 39 illustrates a generalized Gaussian modeling of seismic data wavelet basis decomposition with different power laws. The best matching power for the chosen wavelet basis yields the taxicab metric or ℓ_1-norm. However, we assumed here that Φ has a μ-Lipschitz continuous gradient. For this reason we choose the Huber function as follows:

$$\varphi_k(x) = \begin{cases} \chi x^2 & \text{if } |x| \leq \frac{\tau}{\sqrt{2\chi}} \\ \tau\sqrt{2\chi}|x| - \frac{\tau^2}{2} & \text{otherwise} \end{cases} \tag{3.16}$$

where $\chi > 0$ and $\tau > 0$. Then we have $\varphi(x) = \sum_{k=1}^{K} \kappa_k \varphi_k(x_k)$, which is a function that well approximates the ℓ_1-norm and which has a Lipschitz continuous gradient, $\mu = 2\chi$ being the Lipschitz constant of φ_k, with its gradient given by

$$\nabla\varphi_k(x) = \begin{cases} 2\chi x & \text{if } |x| \leq \frac{\tau}{\sqrt{2\chi}} \\ \tau\sqrt{2\chi}\,\text{sign}(x) & \text{otherwise.} \end{cases} \tag{3.17}$$

Figure 3.4 - p. 40 illustrates the ℓ_1-norm (in blue) and the Huber functions with different parameters (in green and red). The red line is very close to the ℓ_1-norm.
Indeed, (3.13) can be written as

$$\underset{h \in \mathbb{R}^{N_t P}}{\text{minimize}}\ \Phi(\Omega h - Fz) + \rho(h) \tag{3.18}$$

where Φ is a function defined such that $\Phi(x) = \varphi(-x)$. For solving monotone inclusions involving a mixture of sums, linear compositions, we choose to employ Monotone + Skew Forward Backward Forward (M+SFBF) algorithm (Briceños Arias and Combettes, 2011). With this algorithm the ℓ_1-norm can be used without any approximation to the φ function.

Algorithm 2 M+SFBF algorithm

Set $\gamma_n \in [\epsilon, \frac{1-\epsilon}{\beta}]$. Set $h^{[0]} \in \mathbb{R}^{N_t P}, u^{[0]} \in \mathbb{R}^K$.
for $i = 0, 1, \ldots$ **do**
$\widetilde{v}^{[i]} = h^{[i]} - \gamma^{[i]}\Omega^* v^{[i]}$
$\widetilde{h}^{[i]} = v^{[i]} + \gamma^{[i]}\Omega h^{[i]}$
$w_1^{[i]} = \text{prox}_{\gamma^{[i]}\rho}(\widetilde{v}^{[i]})$
$w_2^{[i]} = \text{prox}_{\gamma^{[i]}\Phi^*}(\widetilde{h} - \gamma^{[i]} Fz)$
$q_1^{[i]} = w_1^{[i]} - \gamma^{[i]}\Omega^* w_2^{[i]}$
$q_2^{[i]} = w_2^{[i]} + \gamma^{[i]}\Omega w_1^{[i]}$
$h^{[i+1]} = h^{[i]} - \widetilde{v}^{[i]} + q_1^{[i]}$
$v^{[i+1]} = v^{[i]} - \widetilde{h}^{[i]} + q_2^{[i]}$
end for

We have

$$\begin{aligned}\operatorname{prox}_{\gamma^{[i]}\Phi^*}(\widetilde{h}-\gamma^{[i]}Fz) &= (\widetilde{h}-\gamma^{[i]}Fz)-\gamma^{[i]}\operatorname{prox}_{\frac{\Phi}{\gamma^{[i]}}}(\frac{\widetilde{h}}{\gamma^{[i]}}-Fz) \quad \text{using Tab. 2.1, (iii)}\\ &= (\widetilde{h}-\gamma^{[i]}Fz)-\gamma^{[i]}\operatorname{prox}_{\frac{\varphi(-\cdot)}{\gamma^{[i]}}}(\frac{\widetilde{h}}{\gamma^{[i]}}-Fz)\\ &= (\widetilde{h}-\gamma^{[i]}Fz)+\gamma^{[i]}\operatorname{prox}_{\frac{\varphi}{\gamma^{[i]}}}(Fz-\frac{\widetilde{h}}{\gamma^{[i]}}) \quad \text{using Tab. 2.1, (i).}\end{aligned}$$

Furthermore, we can still use the Huber function for φ_k with this algorithm. In this case, the proximity operator of the Huber function (3.16) reads:

$$\operatorname{prox}_{\varphi_k}(x)=\begin{cases}\frac{x}{2\chi+1} & \text{if } |x|\leq\frac{\tau(2\chi+1)}{\sqrt{2\chi}}\\ x-\tau\sqrt{2\chi}\operatorname{sign}(x) & \text{otherwise.}\end{cases} \tag{3.19}$$

The convergence of the algorithm to an optimal solution to Problem (3.18) is guaranteed by (Briceños Arias and Combettes, 2011, Theorem 3.1). According to this theorem, the step size $\gamma^{[i]}$ at each iteration i must be chosen so as to satisfy the following rule: let $\epsilon\in]0,\frac{1}{\|\Omega\|+1}[$, then $\left(\gamma^{[i]}\right)_{i\in\mathbb{N}}\in[\epsilon,\frac{1-\epsilon}{\|\Omega\|}]$.

3.5.2 Results

In this chapter, the considered data $y=(y^{(n_t)})_{0\leq n_t<N_t}$ and multiple template $r=(r^{(n_t)})_{0\leq n_t<N_t}$ were generated from actual seismic data primaries and multiples. We generated the observed data $z=(z^{(n_t)})_{0\leq n_t<N_t}$ according to the model given in (3.1). In the following, we choose the length of signal $N_t=1024$, $P=10$ tap coefficients and $p'=0$. We use a Symlet wavelet decomposition with filter length 8 over 4 resolution levels. The convex optimization algorithm evaluates the value of $\hat{h}$ that solves problem (3.13).

We choose $\varphi=\sum_{k=1}^{K}\kappa_k\varphi_k$ where for every $k\in\{1,\dots,K\}$, φ_k is the Huber function recalled in (3.16) with $\chi=8$ and $\tau=\frac{1}{2}$ and $\kappa_k=0.8$ for FB algorithm and $\varphi_k=|\cdot|$ and $\kappa_k=0.9$ for M+SFBF algorithm. For both algorithms, function ρ is chosen equal to $(1-\kappa_k)\ell_2^2$. The algorithms are launched on 10000 iterations and may stop earlier at iteration i if the filter becomes stable (i.e. $\|h^{[i+1]}-h^{[i]}\|<\sqrt{N_tP}\times10^{-6}$).

The result of these algorithms are depicted in Figure 3.5 - p. 43 and Figure 3.6 - p. 43. We focus on the leftmost part of the plots, between indices 1 and 500. While Figure 3.6 - p. 43 shows the comparisons of: on the left; primary $\overline{y}$ (in blue) and observed data z; on the right; primary $\overline{y}$ (in blue) and estimated signal $\widehat{y}_1$ (in magenta) by using FB and estimated signal $\widehat{y}_2$ (in cyan) by using M+SFBF. We observe firstly that FB and M+SFBF algorithms give very close results, secondly the primary y is well retrieved. However between indices 80 and 200, the multiple energy is higher in this part, the estimated signal is not compatible with the original, and the result obtained by M+SFBF algorithm is better than the one by FB algorithm. In the next section we propose to add constraints for time-varying adaptive filtering of seismic signal.

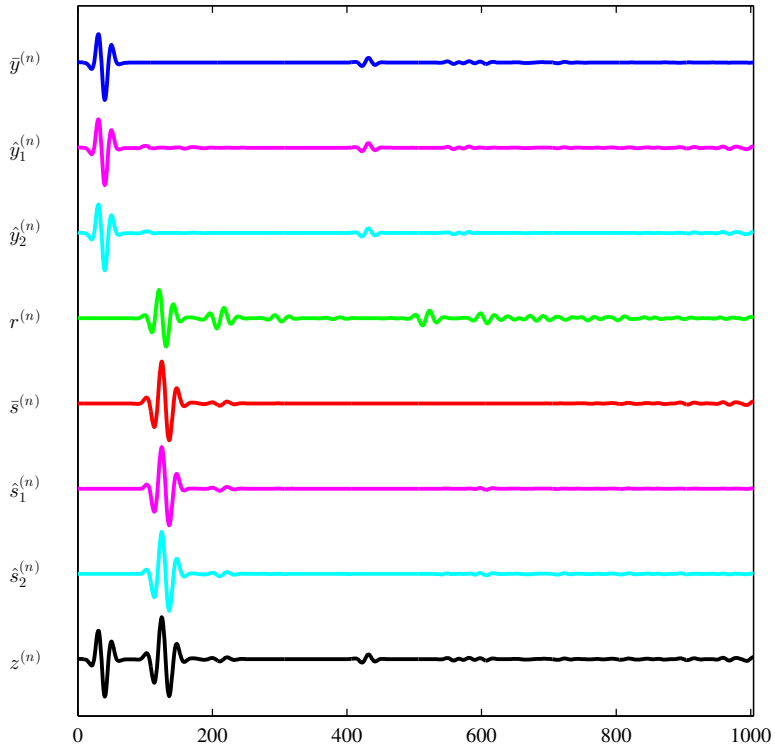


Figure 3.5: *Results by using* FB *and* M+SFBF *algorithms. From up to down: original signal* $\overline{y}$*, estimated signal* $\widehat{y}_1$ *by* FB*, estimated signal* $\widehat{y}_2$ *by* M+SFBF*, model* r*, multiple* $\overline{s}$*, estimated multiples* $\widehat{s}_1$ *by* FB*, estimated multiple* $\widehat{s}_2$ *by* M+SFBF *and observed signal* z*.*

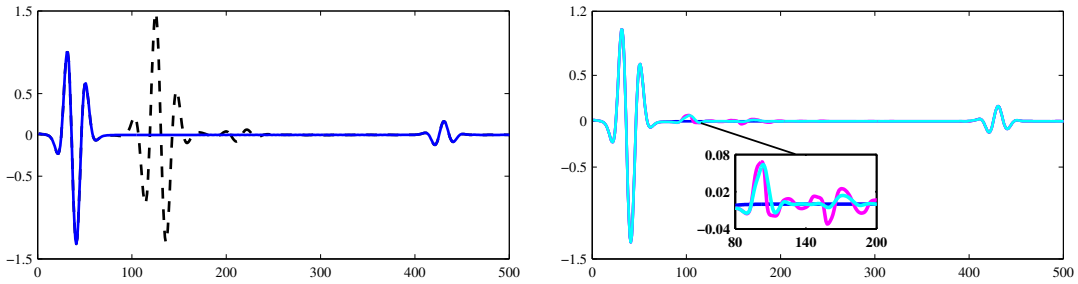


Figure 3.6: *Cropped version of the results by using* FB *algorithm. Left: original signal* $\overline{y}$ *(in blue) and observed signal* z *(in black); Right: original signal* $\overline{y}$ *(in blue) and estimated signal* $\widehat{y}_1$ *(in magenta) and estimated signal* $\widehat{y}_2$ *(in cyan) by* M+SFBF*.*

3.6 Filter estimation adding constraints

The problem is underdetermined because we should estimate an amount of coefficients of filter h much greater than the length of signal. We know that the choice of a good prior (or equivalently, a good regularization term) is the key for solving most ill-posed inverse problems. In this case from wave propagation properties, the filter is here assumed to vary along the time index n_t. However, as mentioned earlier, the filter variations are usually slow. In order to take into account this prior knowledge, the following bounded variation constraint can be introduced

$$(\forall(n_t,p)) \quad |h^{(n_t+1)}(p) - h^{(n_t)}(p)| \leq \epsilon_p. \tag{3.20}$$

This inequality links corresponding coefficients of the impulse response which are estimated between two consecutive time samples (in other words, we impose consecutive filter to take close values). The parameter ϵ_p may depend on the shape of the expected filter. For example, its dependence on the coefficient index p may enable a larger (reps. smaller) difference for filter coefficients taking larger (reps. smaller) values. The associated closed convex set is defined as

$$C = \{h \in \mathbb{R}^{N_t P} \mid \forall(n_t,p) \quad |h^{(n_t+1)}(p) - h^{(n_t)}(p)| \leq \epsilon_p\}. \tag{3.21}$$

We subsequently assume that the pdf f_H is compactly supported on C, so yielding the following criterion to be minimized:

$$\underset{h\in\mathbb{R}^{N_t P}}{\text{minimize}}\ \varphi(F(z - Rh)) + \widetilde{\rho}(h) + \iota_C(h) \tag{3.22}$$

where $\widetilde{\rho} : \mathbb{R}^{N_t P} \to]-\infty, +\infty]$. For computational issues, the convex set C can be expressed as the intersection of two convex subsets C_1 and C_2:

$$C_1 = \left\{h \in \mathbb{R}^{N_t P} \mid \forall p, \forall n_t \in \left\{0, \ldots, \left\lfloor \frac{N_t}{2} \right\rfloor - 1\right\}, |h^{(2n_t+1)}(p) - h^{(2n_t)}(p)| \leq \varepsilon_p\right\} \tag{3.23}$$

$$C_2 = \left\{h \in \mathbb{R}^{N_t P} \mid \forall p, \forall n_t \in \left\{1, \ldots, \left\lfloor \frac{N_t - 1}{2} \right\rfloor\right\}, |h^{(2n_t)}(p) - h^{(2n_t-1)}(p)| \leq \varepsilon_p\right\}. \tag{3.24}$$

Note that in each subset, the involved variables are decoupled. Using these two subsets, Criterion (3.22) becomes

$$\underset{h\in\mathbb{R}^{N_t P}}{\text{minimize}}\ \varphi(F(z - Rh)) + \widetilde{\rho}(h) + \iota_{C_1}(h) + \iota_{C_2}(h). \tag{3.25}$$

Concerning the data fidelity term, it can be observed that it is equal to $h \to \Phi(FRh)$ where, $\Phi \triangleq \varphi(Fz - \cdot)$. The regularization term is function $\widetilde{\rho}$. For the constraints

modeled by the closed convex sets C_1 and C_2, the proximity operators of the associated indicator functions are given by the projections onto these sets. These projections reduce to projections onto a set of hyperslabs of $\mathbb{R}^2$. More precisely, the projection onto C_1 is calculated as follows: let $h \in \mathbb{R}^{N_t P}$ and let $g_1 = \Pi_{C_1}(h)$; and $g_2 = \Pi_{C_2}(h)$; then for every $p \in \{p', \ldots, p' + P - 1\}$ and for every $n_t \in \left\{0, \ldots, \left\lfloor \frac{N_t}{2} \right\rfloor - 1\right\}$,

1. if $|h^{(2n_t+1)}(p) - h^{(2n_t)}(p)| < \varepsilon_p$, then

$$\begin{aligned} g_1^{(2n_t)}(p) &= h^{(2n_t)}(p), & g_1^{(2n_t+1)}(p) &= h^{(2n_t+1)}(p); \\ g_2^{(2n_t)}(p) &= h^{(2n_t)}(p), & g_2^{(2n_t-1)}(p) &= h^{(2n_t-1)}(p) \end{aligned}$$

2. if $h^{(2n_t+1)}(p) - h^{(2n_t)}(p) > \varepsilon_p$, then

$$\begin{aligned} g_1^{(2n_t)}(p) &= \frac{h^{(2n_t+1)}(p) + h^{(2n_t)}(p)}{2} - \frac{\varepsilon_p}{2} \\ g_1^{(2n_t+1)}(p) &= \frac{h^{(2n_t+1)}(p) + h^{(2n_t)}(p)}{2} + \frac{\varepsilon_p}{2}; \\ g_2^{(2n_t)}(p) &= \frac{h^{(2n_t)}(p) + h^{(2n_t-1)}(p)}{2} + \frac{\varepsilon_p}{2} \\ g_2^{(2n_t-1)}(p) &= \frac{h^{(2n_t)}(p) + h^{(2n_t-1)}(p)}{2} - \frac{\varepsilon_p}{2} \end{aligned}$$

3. if $h^{(2n_t+1)}(p) - h^{(2n_t)}(p) < -\varepsilon_p$, then

$$\begin{aligned} g_1^{(2n_t)}(p) &= \frac{h^{(2n_t+1)}(p) + h^{(2n_t)}(p)}{2} + \frac{\varepsilon_p}{2} \\ g_1^{(2n_t+1)}(p) &= \frac{h^{(2n_t+1)}(p) + h^{(2n_t)}(p)}{2} - \frac{\varepsilon_p}{2}; \\ g_2^{(2n_t)}(p) &= \frac{h^{(2n_t)}(p) + h^{(2n_t-1)}(p)}{2} - \frac{\varepsilon_p}{2} \\ g_2^{(2n_t-1)}(p) &= \frac{h^{(2n_t)}(p) + h^{(2n_t-1)}(p)}{2} + \frac{\varepsilon_p}{2}. \end{aligned}$$

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 Figure 3.7 - p. 46.

3.6.1 Algorithm

We propose to employ the iterative algorithm in (Pesquet and Pustelnik, 2012) (see Algorithm 3). This algorithm mainly consists of computing alternately proximity operators and projections. It can be noticed that the algorithm requires to compute the inverse of matrix $Q \in \mathbb{R}^{N_t \times N_t}$. If F corresponds to a wavelet orthonormal basis or a tight frame

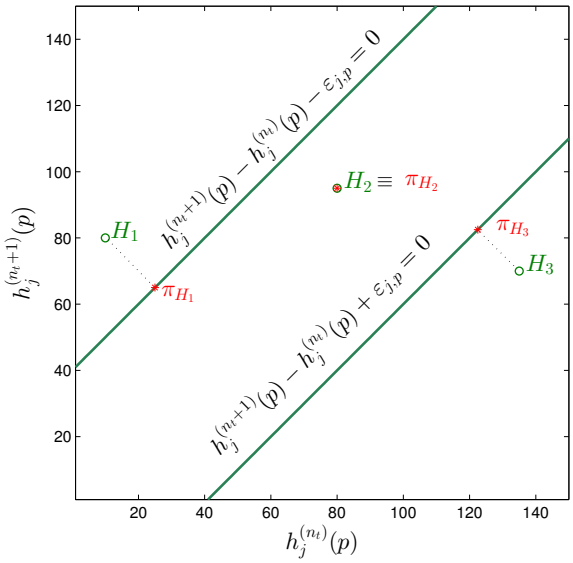


Figure 3.7: *Projection onto C_1/C_2 of points H_1, H_2 and H_3 in $\mathbb{R}^2$.*

Algorithm 3 PPXA+

Set $(o_1, o_2, o_3, o_4) \in]0, +\infty[^4$ and $t_1^{[0]} \in \mathbb{R}^K$, $t_2^{[0]} \in \mathbb{R}^{N_t P}$, $t_3^{[0]} \in \mathbb{R}^{N_t P}$, $t_4^{[0]} \in \mathbb{R}^{N_t P}$
$Q = o_1 \nu R^\top R + (o_2 + o_3 + o_4)\,\mathrm{I}$
$h^{[0]} = Q^{-1}\left(o_1 R^\top F^\top t_1^{[0]} + o_2 t_2^{[0]} + o_3 t_3^{[0]} + o_4 t_4^{[0]}\right)$
for $i = 0, 1, \ldots$ **do**
  $w_1^{[i]} = \mathrm{prox}_{\frac{\Phi}{o_1}}(t_1^{[i]})$
  $w_2^{[i]} = \mathrm{prox}_{\frac{\tilde{\rho}}{o_2}}(t_2^{[i]})$
  $w_3^{[i]} = \Pi_{C_1}(t_3^{[i]})$
  $w_4^{[i]} = \Pi_{C_2}(t_4^{[i]})$
  $c^{[i]} = Q^{-1}\left(o_1 R^\top F^\top w_1^{[i]} + o_2 w_2^{[i]} + o_3 w_3^{[i]} + o_4 w_4^{[i]}\right)$
  $\lambda^{[i]} \in]0, 2[$
  $t_1^{[i+1]} = s_1^{[i]} + \lambda^{[i]}\left(FR(2c^{[i]} - h^{[i]}) - w_1^{[i]}\right)$
  $t_2^{[i+1]} = s_2^{[i]} + \lambda^{[i]}\left(2c^{[i]} - h^{[i]} - w_2^{[i]}\right)$
  $t_3^{[i+1]} = t_3^{[i]} + \lambda^{[i]}\left(2c^{[i]} - h^{[i]} - w_3^{[i]}\right)$
  $t_4^{[i+1]} = t_4^{[i]} + \lambda^{[i]}\left(2c^{[i]} - h^{[i]} - w_4^{[i]}\right)$
  $h^{[i+1]} = h^{[i]} + \lambda^{[i]}\left(c^{[i]} - h^{[i]}\right)$
end for

with constant $\nu \in]0, +\infty[$ ($\nu = 1$ when F is a orthonormal basis), this can be done by applying the Woodbury formula (Golub and Van Loan, 1996)

$$\begin{aligned} Q^{-1} =& \frac{1}{o_2+o_3+o_4}\mathrm{I} - \frac{o_1\nu}{o_2+o_3+o_4}R^\top\left(\mathrm{I}+\frac{o_1\nu}{o_2+o_3+o_4}RR^\top\right)^{-1}R\frac{1}{o_2+o_3+o_4}\mathrm{I} \\ =& \frac{1}{o_2+o_3+o_4}\mathrm{I} - \frac{o_1\nu}{(o_2+o_3+o_4)^2}R^\top\left(\mathrm{I}+\frac{o_1\nu}{o_2+o_3+o_4}RR^\top\right)^{-1}R. \end{aligned}$$

By using the fact that

$$RR^\top = \mathrm{Diag}(\|R^{(0)}\|^2, \ldots, \|R^{(N_t-1)}\|^2) \tag{3.26}$$

we get

$$Q^{-1} = \frac{1}{o_2+o_3+o_4}\left(\mathrm{I} - R^\top DR\right) \tag{3.27}$$

where

$$D^{-1} = \mathrm{Diag}\left(\frac{o_2+o_3+o_4}{o_1\nu} + \|R^{(0)}\|^2, \ldots, \frac{o_2+o_3+o_4}{o_1\nu} + \|R^{(N_t-1)}\|^2\right). \tag{3.28}$$

Concerning the proximity operator of function Φ, it can be computed, by using Table 2.1 - p. 30, (i) and (iii), we get

$$\begin{aligned} \mathrm{prox}_\Phi(x) &= \mathrm{prox}_{\varphi(Fz-\cdot)}(x) \\ &= \mathrm{prox}_{\varphi(-(\cdot - Fz))}(x) \\ &= Fz - \mathrm{prox}_\varphi(Fz - x) \quad \text{using Tab. 2.1, (i) and (iii).} \end{aligned}$$

3.6.2 Results

The inner parameters of PPXA+ have been chosen in an empirical manner: $\lambda^{[i]} = 1.5$, $o_1 = 10000/N_t, o_2 = o_1/P, o_3 = o_4 = 10\,o_2$; the algorithm is initialized by randomly generated N_t positive-valued vectors of size P summing up to one. The stopping criteria for PPXA+ algorithm is similar to the one for FB and M+SFBF algorithms.

The numerical results below have been obtained with $\varphi_k \equiv \kappa_k|\cdot|$ and $\widetilde{\rho} = (1-\kappa_k)\|\cdot\|^2$, $\kappa_k = 0.9$. Symlet wavelets of length 8 over 4 resolution levels are used.

In a first set of experiments, the observations are generated according to (3.1), and we assume that P, p' and $\epsilon = \max_{p'\le p\le p'+P-1, 0\le n_t<N_t-1} |\overline{h}^{(n_t+1)}(p) - \overline{h}^{(n_t)}(p)|$ are known, where $\overline{h}$ denotes the "true" impulse response. We subsequently set $\epsilon_p \equiv \epsilon$ for defining the convex constraints on the filter.

Results obtained considering $P = 10$, $p' = 0$ and $\epsilon = 0.14$ are displayed in Figure 3.8 - p. 48 and Figure 3.9 - p. 48. The signal is satisfactorily recovered. As we can observe in Figure 3.9 - p. 48, the primary is well recovered, in particular, the result is improved between indices 80–200. This shows the performance of time-varying filter constraints.

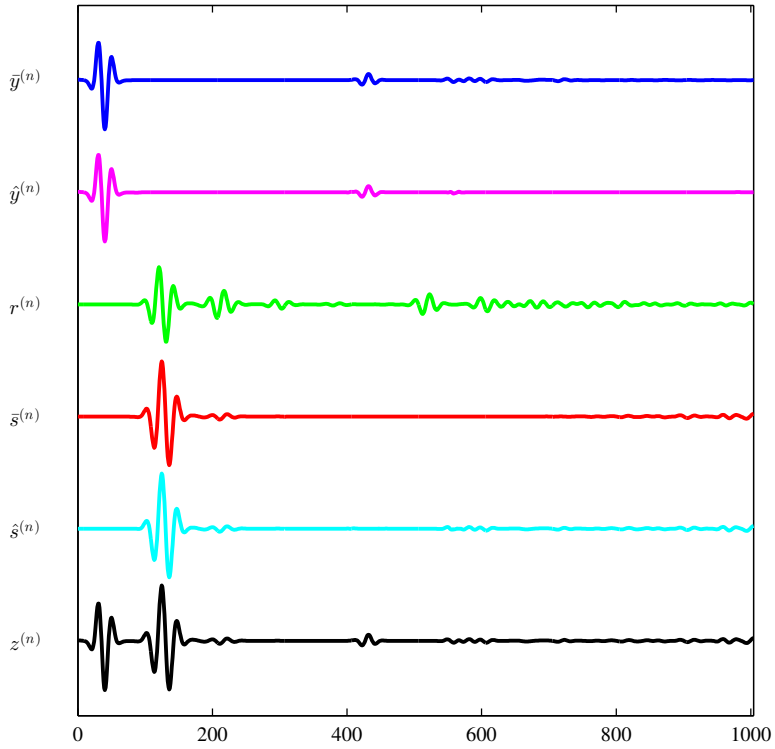


Figure 3.8: *Results by using* PPXA+ *algorithm. From top to bottom: original signal* $\overline{y}$*, estimated signal* $\widehat{y}$*, model* r*, multiple* $\overline{s}$*, estimated multiples* $\widehat{s}$ *and observed signal* z*.*

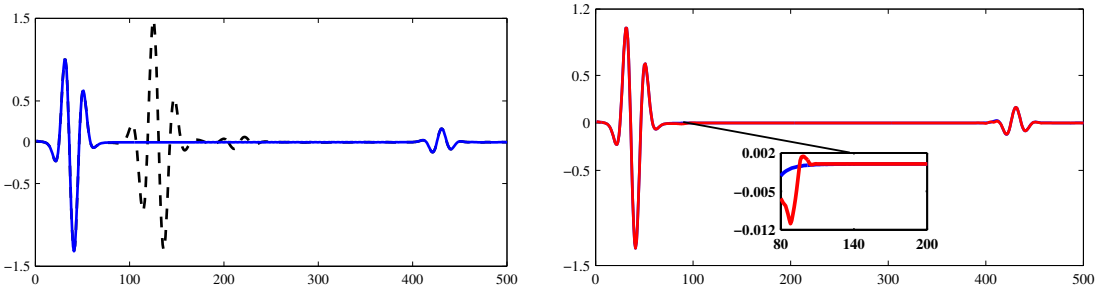


Figure 3.9: *Cropped version of the results by using* PPXA+ *algorithm. Left: original signal* $\overline{y}$ *(in blue) and and observed signal* z *(in black); Right: original signal* $\overline{y}$ *(in blue) and estimated signal* $\widehat{y}$ *(in red).*

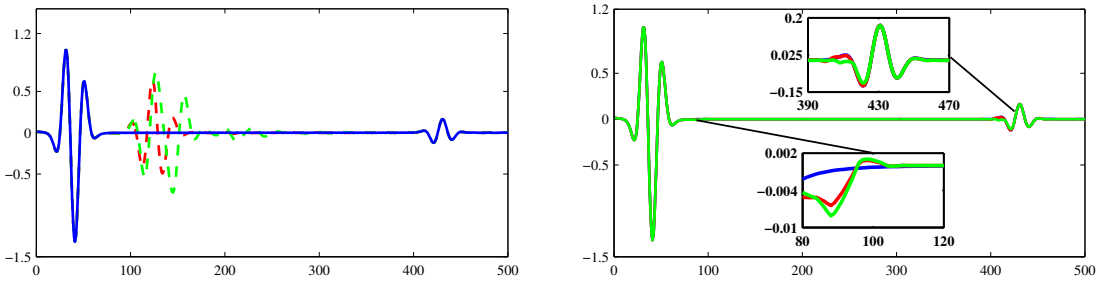


Figure 3.10: *Cropped version of the results by using* PPXA+ *algorithm. Left: original signal $\bar{y}$ (in blue) and observed signal z in red for $P = 6$ and in green for $P = 16$; right: original signal $\overline{y}$ (in blue), estimated signal $\hat{y}$ in red for $P = 6$ and in green for $P = 16$.*

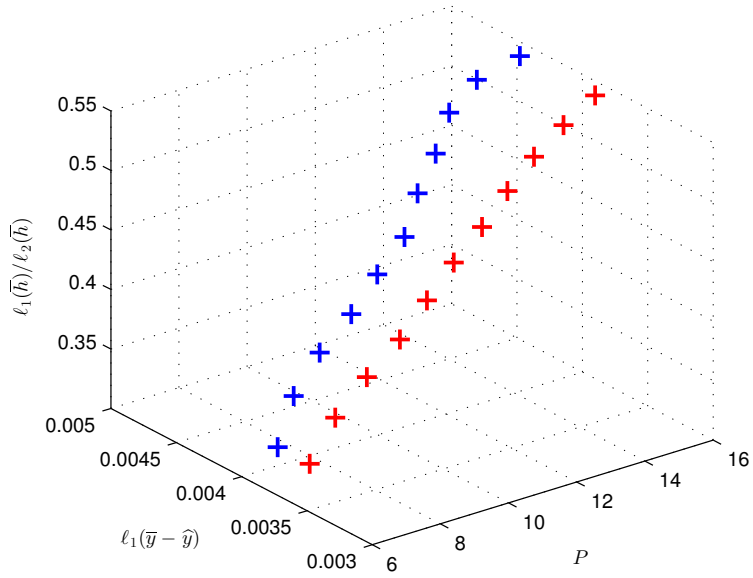


Figure 3.11: *Comparison of the results obtained when varying P tap coefficient from 6 to 16 in using two a priori functions $\widetilde{\rho} = \ell_1$ (in blue) and $\widetilde{\rho} = \ell_2^2$ (in red).*

To emulate actual geophysical configurations, the data and templates obtained from actual seismic surveys have been combined with filter of varying length (from five to twenty taps), starting time (p') and shape (from fully symmetric to asymmetric). Due to amplitude and spectrum variations in actual multiples, an absolute statistical analysis is not straightforward. Instead, we focus on errors of estimated signal and reference signal ($\ell_1(\overline{y}-\hat{y})$) and the $\ell_1/\ell_2(\overline{h})$. Nowadays, the ℓ_1/ℓ_2 criterion is used frequently to measure the sparsity of a signal (Repetti et al., 2015; Krishnan et al., 2011; Benichoux et al., 2013; Zibulevsky and Pearlmutter, 2001; Hoyer, 2004; Hurley and Rickard, 2009; Barak et al., 2014). In Chapter **??**, we propose a method based on a smooth approximation of the ℓ_1/ℓ_2 function. Figure 3.11 - p. 49 shows the results obtained by two different functions ρ (ℓ_1-norm and ℓ_2-norm). We observe that in both cases, the bigger $\ell_1(\overline{y}-\hat{y})$, the bigger $\ell_1/\ell_2(\overline{h})$. In other words, if a filter presents more sparsity, then the estimated signal is better retrieved. Figure 3.10 - p. 49 shows a comparison between the two estimated signals when $P = 6$ (in red) and $P = 16$ (in green) with original signal (in blue). We observe that the signal obtained for $P = 6$ is recovered perfectly, while for $P = 16$, as the filter length is quite large, it is much more difficult to detect. However, the signal is satisfactorily recovered.

3.7 Conclusion

In this chapter, we developed a new variational framework for multiple removal in seismic data under the assumption that a disturbance signal template is available. The proposed algorithm is based on recent advances in the theory of proximal methods. This allows us to estimate FIR filters that vary along the time dimension and the results provided by this approach appear to be very promising. In a first time we have started with some standard algorithms such as FB and M+SFBF for minimizing the sum of two functions and no constraint is considered. In a second time, the slowly-varying property of filter is taken into account. We then proposed to employ PPXA+ algorithm. The results showed the good performance of time varying filter constraint and the ℓ_1-norm penalty on the filter. In the following chapter, we extend this method to the noisy case. Moreover we propose a technique to combine different multiple models together to achieve better results.

- Chapter 4 -

Multiple removal with several templates and noise

In the previous chapter we discussed the problem of seismic multiple removal for only one template in a noise-free case. The focus of this chapter addresses different types of multiples with noise estimation. We recall the problem of jointly estimating the filters and the signal of interest (primary) in a new convex variational formulation, allowing the incorporation of knowledge about the noise statistics. We propose a primal-dual algorithm to solve the derived minimization problem which yields good performance in both simulated and real field seismic data.

4.1 Introduction

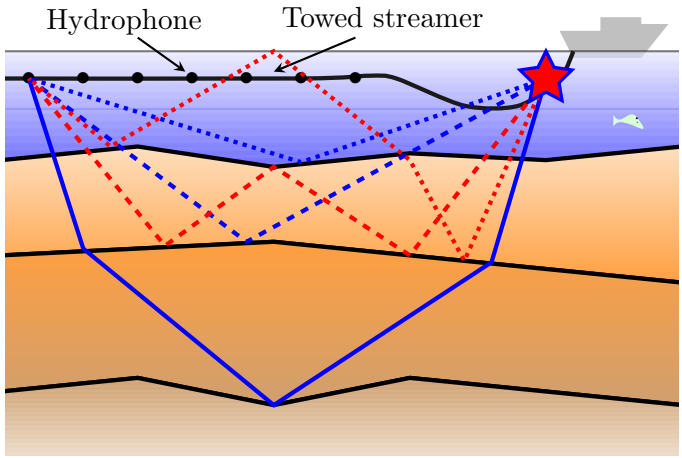


Figure 4.1: *Principles of marine seismic data acquisition and wave propagation. Towed streamer with hydrophones. Primaries (in blue), and two different types of multiple peg-leg (dashed red) and water-bottom (dotted red).*

We now consider the problem for a more general case where the primaries are degraded by both random and structured noises. Consequently, the problem (3.1) becomes

$$z^{(n_t)} = \overline{y}^{(n_t)} + \overline{s}^{(n_t)} + b^{(n_t)} \tag{4.1}$$

where $b = (b^{(n_t)})_{0 \leq n_t < N_t}$ denotes the additional noise. It is well-known that seismic data often contains random noises related to near-surface scatterers. This work mostly deals with the mitigation of some ℓ_2-norm on residuals, as remnant noise is traditionally considered as Gaussian in seismic. As far as sparsity is concerned, ℓ_2-norm penalties are blended with ℓ_0 objectives, solved through ℓ_1 or hybrid ℓ_1-ℓ_2 approximations (Guitton and Verschuur, 2004), resorting for instance to iteratively re-weighted least-squares method. Recently, (Costagliola et al., 2011) 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. In reality, there exist several types of multiples. As illustrated in Figure 4.1 - p. 51, there is not only the peg-leg multiple, but also a water-bottom multiple. In (Mei and Zou, 2010), the authors proposed a multiple subtraction method for two or more multiple models, using matched filters to modify them. Then they minimize the energy between the input data and these modified multiple models. Note that this step is carried out in local time-space windows, and thus the matched filters can vary from one window to another. They showed that the type of weighted subtraction with weights based on the similarity between the multiples and the input data gives the best result when one multiple model is more reliable than the other one. Their experiments were performed only for two multiple models. Recently, we refer to (Ventosa et al., 2012) for a combination of several multiple models. The authors proposed a model-based multiple subtraction that combines complex Morlet wavelet frame with complex unary Wiener filters. Their methods are demonstrated on field data with three different multiple models. Also many papers show the good performance of the multi-model in the multiple removal problem such as (Sanchis and Hanssen, 2011; van Borselen et al., 2012; Liu and Dragoset, 2012; Moni et al., 2012; Mei and Zou, 2010; Neelamani et al., 2008; Pica, 1992; Huo and Wang, 2009). Sometimes, one can consider that there is a single model, however it is turned into a multiple model with the derivatives and the Hilbert transform such as in (Wang, 2003). As problem (4.1) is underdetermined, additional constraints should be devised. In particular, we investigate sparsity issues on primaries and adaptive filters. As for sparse representation on primaries, this can be computed using wavelet transformations such as bases or frames (Pesquet et al., 1996; Fowler, 2005) in which the regularizations are adapted (e.g. the ℓ_1-norm defined on the coefficients). Since each subband has different properties due to the energy distribution, it is reasonable to use the different regularization parameters depending on the content of each subband. In image processing, ℓ_1-norm seems to be a better cost function for wavelet coefficient in subbands and squared ℓ_2-norm is more appropriate on the approximation subband. In (Chaux and Blanc-Féraud, 2012) the authors proposed a method to estimate the regularization hyperparameters per subband based on a Maximum Likelihood (ML) estimator. In this chapter, we propose to define constraints for each subband. This allows us to more easily determine data-based

parameters and to improve the performance of frame decompositions. Concerning the sparsity on adaptative filters, we employ a nuclear norm regularization. This choice is due to the fact that low-rank matrix approximations and sparsity are related. Indeed, a low-rank property well represents matrix sparsity (Zeng et al., 2012; Candès et al., 2011; Chen and Sacchi, 2014; Kreimer et al., 2013). It is well-known that the nuclear norm (sum of singular values) regularization promotes a low rank solution. The singular values of a complex matrix are always positive (Golub and Van Loan, 1996). This implies that the nuclear norm is equivalent to the ℓ_1-norm on singular values. In particular, we demonstrate that the nuclear norm is equivalent to the mixed $\ell_{1,2}$-norm in some cases.

This chapter extends a more generic framework for multiple reflection filtering with (i) a noise prior, (ii) sparsity constraints (for each of subbands) on signal frame coefficients, (iii) slow variation modeling of the adaptive filters, and (iv) concentration metrics on the filters. Multiple constraints can now be handled in a convenient manner. Due to the diversity of focus points which are 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, which appear as natural atoms for sparse descriptions of some physical processes, related to propagation and reflection of signals through media.

Section 4.2 describes the transformed linear model incorporating the templates with adaptive filtering. In Section 4.3, we formulate a generic variational form for the problem. Section 4.4 describes the primal-dual proximal formulation. The performance of the proposed method is assessed in Section 4.5. 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 then tested and applied to an actual seismic data-set. Actually, 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 a paramount feature in geophysical data processing assessment. We thus compare the proposed approach with a state-of-the-art solution, previously benchmarked against industrial competitors (Ventosa et al., 2012). Conclusions and perspectives are drawn in Section 4.6.

4.2 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 $(r_j^{(n_t)})_{0\leq n_t<N_t, 0\leq j<J}$, related to $(\overline{s}^{(n_t)})_{0\leq n_t<N_t}$ via a possibly non-causal linear model through a limited support relationship:

$$\overline{s}^{(n_t)} = \sum_{j=0}^{J-1} \sum_{p=p'}^{p'+P_j-1} \overline{h}_j^{(n_t)}(p) r_j^{(n_t-p)} \tag{4.2}$$

where $\overline{h}_j^{(n_t)}$ is an unknown finite impulse response (with P_j tap coefficients) associated with template j at time n_t, and where $p' \in \{-P_j+1, \dots, 0\}$ is its starting index ($p'=0$ corresponds to the causal case). Templates are generated with standard geophysical modeling based on the above primaries and wave-equation modeling. 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 $(\overline{h}_j)_{0\le j<J}$ by:

$$\begin{aligned}
\overline{s} &= \left[\overline{s}^{(0)} \quad \dots \quad \overline{s}^{(N_t-1)}\right]^\top, \\
\overline{h}_j &= \left[\overline{h}_j^{(0)}(p') \ \dots \ \overline{h}_j^{(0)}(p'+P_j-1) \ \dots \overline{h}_j^{(N_t-1)}(p') \ \dots \ \overline{h}_j^{(N_t-1)}(p'+P_j-1)\right]^\top,
\end{aligned}$$

and block diagonal matrices $(R_j)_{0\le j<J}$ of size $N_t \times N_t P_j$:

$$R_j = \begin{bmatrix} R_j^{(0)} & 0 & \dots & 0 \\ 0 & R_j^{(1)} & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & R_j^{(N_t-1)} \end{bmatrix},$$

where $(R_j^{(n_t)})_{0\le n_t\le N_t-1}$ are vectors of dimension P_j such that

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

Eq. (4.2) can be expressed more concisely as

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

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

4.3 Proposed variational approach

4.3.1 Bayesian framework

As mentioned in the previous chapter, we assume that $\overline{y}$ is a realization of a random vector Y, the probability density function (pdf) of which is given by

$$(\forall y \in \mathbb{R}^{N_t}) \quad f_Y(y) \propto \exp(-\varphi(Fy)) \tag{4.3}$$

where $\varphi\colon \mathbb{R}^{N_t} \to]-\infty, +\infty]$ is the associated potential, assumed to have a fast enough decay.

To take into account the available information on the unknown filters, it can be assumed that for all $j \in \{0, \ldots, J-1\}$, $\overline{h}_j$ is a realization of a random vector H_j. Let $\mathbb{H} = \mathbb{R}^{N_t P_0} \times \ldots \times \mathbb{R}^{N_t P_{J-1}}$. The joint pdf of the filter coefficients can be expressed as:

$$(\forall h \in \mathbb{H}) \quad f_{H_0,\ldots,H_{J-1}}(h) \propto \exp(-\rho(h)),$$

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

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

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

$$(\forall y \in \mathbb{R}^{N_t})(\forall h \in \mathbb{H})$$
$$f_{Y,H_0,\ldots,H_{J-1}|Z=z}(y,h) \propto \exp\left(-\psi\Big(z - y - \sum_{j=0}^{J-1} R_j h_j\Big)\right) f_Y(y) f_{H_0,\ldots,H_{J-1}}(h).$$

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

$$\underset{y \in \mathbb{R}^{N_t},\, h \in \mathbb{H}}{\text{minimize}}\ \psi(z - y - \sum_{j=0}^{J-1} R_j h_j) + \varphi(Fy) + \rho(h). \tag{4.4}$$

4.3.2 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:

$$\underset{Fy \in D,\, \mathbf{h} \in C}{\text{minimize}}\ \Psi\left(\begin{bmatrix} y \\ \mathbf{h} \end{bmatrix}\right) \tag{4.5}$$

where the data fidelity term is defined by function $\Psi\colon \left(\begin{bmatrix} y \\ \mathbf{h} \end{bmatrix}\right) \mapsto \psi\left(z - [\mathrm{I}\ \mathbf{R}]\begin{bmatrix} y \\ \mathbf{h} \end{bmatrix}\right)$, 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 (Gragnaniello et al., 2012; Pham et al., 2013) (this point will be detailed later on). We will now turn our attention to the choice of Ψ, C and D.

4.3.3 Considered data fidelity term and constraints

4.3.3.1 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 $\psi = \|\,.\,\|^2$.

4.3.3.2 A priori information on adaptive filter

The filters are assumed to be time–varying. Naturally, the hypothesis is difficult to verify in practice. However, in order to ensure smooth variations along time, we propose to introduce constraint sets

$$C_1 = \left\{\mathbf{h} \in \mathbb{R}^Q \mid \forall (j,p), \forall n_t \ \in \left\{0, \ldots, \left\lfloor \frac{N_t}{2} \right\rfloor - 1\right\} |h_j^{(2n_t+1)}(p) - h_j^{(2n_t)}(p)| \leq \varepsilon_{j,p}\right\} \tag{4.6}$$

$$C_2 = \left\{\mathbf{h} \in \mathbb{R}^Q \mid \forall (j,p), \forall n_t \ \in \left\{1, \ldots, \left\lfloor \frac{N_t - 1}{2} \right\rfloor\right\} |h_j^{(2n_t)}(p) - h_j^{(2n_t-1)}(p)| \leq \varepsilon_{j,p}\right\}. \tag{4.7}$$

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 functions, by setting $C_3 = \left\{\mathbf{h} \in \mathbb{R}^Q \mid \ \forall j \in \{0, 1, \ldots, J-1\},\ \widetilde{\rho}_j(h_j) \leq \lambda_j\right\}$ where for every $j \in \{0, 1, \ldots, J-1\}$, $\lambda_j \in\,]0, +\infty[$, $\widetilde{\rho}_j\colon \mathbb{R}^{N_t P_j} \to [0, +\infty[$ 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 (Kowalski, 2009). Hence, the convex set C is defined as $C = C_1 \cap C_2 \cap C_3$. From a computational standpoint (see Section 4.4), it is more efficient to split C into three subsets as described above.

4.3.3.3 A priori information on primary signal

As mentioned in Section 4.3.1, we assume that the primary signal is sparsely described through an analysis frame operator $F \in \mathbb{R}^{K \times N_t}$ (Jacques et al., 2011), which may ease its processing, by increasing the data-domain discrepancy between primaries and multiples. The associated constraint can be split by defining a partition of $\{1, \ldots, K\}$ denoted by $\{\mathbb{K}_l \mid l \in \{1, \ldots, \mathcal{L}\}\}$. For example, for wavelet frames, $\mathcal{L}$ may correspond to the number of subbands and $\mathbb{K}_l$ is the l-th subband. Then, one can choose $D = D_1 \times \ldots \times D_{\mathcal{L}}$ with $D_l = \{(x_k)_{k \in \mathbb{K}_l} \mid \sum_{k \in \mathbb{K}_l} \widetilde{\varphi}_l(x_k) \leq \beta_l\}$, where, for every $l \in \{1, \ldots, \mathcal{L}\}$, $\beta_l \in]0, +\infty[$, and $\widetilde{\varphi}_l : \mathbb{R} \to [0, +\infty[$ is a lower-semicontinuous convex function.

4.4 Primal-Dual proximal algorithm

Our objective is to provide a numerical solution to Problem (4.5). 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 constraints are expressed through linear operators, such as a wavelet frame analysis operator F. For this reason, primal-dual algorithms (Briceños Arias and Combettes, 2011; Chambolle and Pock, 2011; Condat, 2013), such as the Monotone+Lipschitz Forward-Backward-Forward (M+LFBF) algorithm (Combettes and Pesquet, 2012), constitute appropriate choices since they avoid some large-size matrix inversions inherent to other schemes such as the ones proposed in (Afonso et al., 2011; Pesquet and Pustelnik, 2012). As mentioned in Section 4.3.3.1, Ψ is a quadratic function and its gradient is thus Lipschitzian, which permits it to be handled by the M+LFBF algorithm. In order to deal with the constraints, 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).

4.4.1 Gradient and projection computation

From the assumption of additive zero-mean Gaussian noise, we deduce that Ψ is differentiable with a μ-Lipschitzian gradient, i.e. $(\forall \begin{bmatrix} y \\ \mathbf{h} \end{bmatrix} \in \mathbb{R}^{N_t+Q})(\forall \begin{bmatrix} y' \\ \mathbf{h}' \end{bmatrix} \in \mathbb{R}^{N_t+Q})$:

$$\left\| \nabla \Psi \left(\begin{bmatrix} y \\ \mathbf{h} \end{bmatrix} \right) - \nabla \Psi \left(\begin{bmatrix} y' \\ \mathbf{h}' \end{bmatrix} \right) \right\| \leq \mu \left\| \begin{bmatrix} y \\ \mathbf{h} \end{bmatrix} - \begin{bmatrix} y' \\ \mathbf{h}' \end{bmatrix} \right\|$$

and

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

The gradient of Ψ is thus μ-Lipschitzian with

$$\mu = 2 \left| \left| \left| [\mathrm{I}\ \mathbf{R}] \right| \right| \right|^2 \tag{4.8}$$

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

We now 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 three terms (see Section 4.3.3). 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 — impose smooth variations along time of the corresponding tap coefficients.

More precisely, similarly to Chapter 3, 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 $j \in \{0,\dots,J-1\}$, $p \in \{p',\dots,p'+P-1\}$ and $n_t \in \left\{0,\dots,\left\lfloor \frac{N_t}{2} \right\rfloor - 1\right\}$,

1. if $|h_j^{(2n_t+1)}(p) - h_j^{(2n_t)}(p)| < \varepsilon_{j,p}$, then
$$g_{j,1}^{(2n_t)}(p) = h_j^{(2n_t)}(p), \quad g_{j,1}^{(2n_t+1)}(p) = h_j^{(2n_t+1)}(p);$$

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

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

C_3 introduces a priori information on the filter vector $\mathbf{h}$ through the lower-semicontinuous convex functions $\widetilde{\rho}_j$. In this framework, we consider three possible choices for $\widetilde{\rho}_j$:

1. ℓ_1-norm:
$$\widetilde{\rho}_j(h_j) = \|h_j\|_{\ell_1} = \sum_{n_t=0}^{N_t-1} \sum_{p=p'}^{p'+P_j-1} |h_j^{(n_t)}(p)|.$$
This choice requires to perform projections onto an ℓ_1-ball. This can be achieved by using the iterative procedure proposed in (van den Berg and Friedlander, 2008), which yields the projection in a finite number of iterations.

2. squared ℓ_2-norm:
$$\widetilde{\rho}_j(h_j) = \|h_j\|_{\ell_2}^2 = \sum_{n_t=0}^{N_t-1} \sum_{p=p'}^{p'+P_j-1} |h_j^{(n_t)}(p)|^2.$$
In this case, the projection is straightforward.

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

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

Then, we can use an algorithm similar to the one proposed in (van den Berg and Friedlander, 2008) computing the projection onto an $\ell_{1,2}$-ball.

4. nuclear norm $\|\cdot\|_*$:

$$\widetilde{\rho}_j(h_j) = \|L_j h_j\|_*.$$

Following the work in (Grave et al., 2011) on sparse regression, a good choice for L_j may be:

$$\begin{aligned} L_j : \mathbb{R}^{N_t P_j} &\to \mathbb{R}^{N_t \times N_t P_j} \\ h_j &\mapsto R_j \operatorname{Diag}(h_j). \end{aligned}$$

First, let have a look on the adjoint operator L_j^* of L_j. For every $x \in \mathbb{R}^{N_t \times N_t P_j}$, $y \in \mathbb{R}^{N_t P_j}$, we have

$$\begin{aligned} \langle L_j^* x, y \rangle &= \langle x, L_j y \rangle \\ &= \operatorname{Tr}\left(x (L_j y)^\top \right) \\ &= \operatorname{Tr}\left(x \operatorname{Diag}(y) R_j^\top \right) \\ &= \operatorname{Tr}\left(R_j^\top x \operatorname{Diag}(y) \right). \end{aligned}$$

For every $l \in \{1, 2, \ldots, N_t P_j\}$ and for every $k \in \{1, 2, \ldots, N_t P_j\}$, we have

$$\begin{aligned} \big(R_j^\top x \underbrace{\operatorname{Diag}(y)}_{D}\big)_{l,k} &= \sum_{m=1}^{N_t P_j} (R_j^\top x)_{l,m} D_{m,k} \\ &= (R_j^\top x)_{l,k} y_k \\ \Rightarrow \operatorname{Tr}\Big(R_j^\top x \operatorname{Diag}(y)\Big) &= \sum_{k=1}^{N_t P_j} (R_j^\top x)_{k,k} y_k \\ &= \langle \operatorname{diag}(R_j^\top x), y \rangle. \end{aligned}$$

We finally obtain $L_j^* x = \operatorname{diag}(R_j^\top x)$.
We now compute the projection Π_{C_3} onto a "nuclear-ball". Thus, we define $C_3 = C_{3,0} \times \ldots \times C_{3,J-1}$, where, for every $j \in \{0, 1, \ldots, J-1\}$,

$$C_{3,j} = \left\{ X_j \in \mathbb{R}^{N_t \times N_t P_j} \middle| \, \|X_j\|_* \leq \lambda_j \right\}.$$

By definition, we have:

$$\Pi_{C_{3,j}}(X_j) = \left\{ \underset{Y_j \in \mathbb{R}^{N_t \times N_t P_j}}{\operatorname{argmin}} \|Y_j - X_j\| \text{ subject to } \|Y_j\|_* \leq \lambda_j \right\}. \tag{4.9}$$

Ignoring the case $\|X_j\|_* \leq \lambda_j$ which has the trivial solution $Y_j = X_j$, there exists for each λ_j a χ such that

$$\underset{Y_j \in \mathbb{R}^{N_t \times N_t P_j}}{\text{minimize}} \frac{1}{2}\|Y_j - X_j\|^2 + \chi\|Y_j\|_* \tag{4.10}$$

has the same solution as (4.9). The solution of this penalized formulation is $\operatorname{prox}_{\chi\|\cdot\|_*}(X_j)$.

We now turn our attention to the proximity operator of the nuclear norm. The nuclear norm is also known as the trace norm. The nuclear norm of a matrix $X_j \in \mathbb{R}^{N_t \times N_t P_j}$ of rank T_j is defined as the sum of its singular values $(\upsilon_{j,i})_{1 \leq i \leq T_j}$ indexed in a decreasing order

$$\|X_j\|_* = \sum_{i=1}^{T_j} \upsilon_{j,i} \tag{4.11}$$

where the singular values $\upsilon_{j,i}$ are positive. The singular value thresholding operator is the proximity operator associated with the nuclear norm (Cai et al., 2010). More precisely, the proximity operator of the nuclear norm at $X_j \in \mathbb{R}^{N_t \times N_t P_j}$ is

$$\operatorname{prox}_{\chi^{-1}\|\cdot\|_*}(X_j) = U_j \mathcal{D}_{\chi^{-1}}(\Sigma_j) V_j^\top \quad \text{for every } \chi \in]0, +\infty[\tag{4.12}$$

where $\mathcal{D}_{\chi^{-1}}(\Sigma_j) = \operatorname{Diag}\left((\{\upsilon_{j,i} - \chi^{-1}\}_+)_{1 \leq i \leq T_j}\right)$ and U_j, V_j are respectively $N_t \times T_j$ and $N_t P_j \times T_j$ matrices with orthonormal columns i.e. $U_j^\top U_j = \mathrm{I}_{T_j}$ and $V_j^\top V_j = \mathrm{I}_{T_j}$, resulting from the singular value decomposition (SVD) of X_j.

Finally, we can find the value of χ that makes (4.9) and (4.10) equivalent, so that both can be solved using the proximity operator of $\chi\|\cdot\|_*$ at X_j. This comes down to finding χ such that $\|\operatorname{prox}_{\chi^{-1}\|\cdot\|_*}(X_j)\|_* = \lambda_j$ i.e.

$$\|U_j \mathcal{D}_{\chi^{-1}}(\Sigma_j) V_j^\top\|_* = \lambda_j \;\Leftrightarrow\; \sum_{i=1}^{T_j} \{\upsilon_{j,i} - \chi^{-1}\}_+ = \lambda_j. \tag{4.13}$$

Without loss of generality we can define $\upsilon_{j,T_j+1} = 0$ and the function $\phi(\cdot) = \sum_{i=1}^{T_j} \{\upsilon_{j,i} - \cdot\}_+$ which is continuous and monotonically decreasing from $\|X_j\|_*$ to 0 on $[0, \upsilon_{j,i}]$. Therefore, there exists an integer i_0 such that

$$\phi(\upsilon_{j,i_0+1}) \leq \chi^{-1} < \phi(\upsilon_{j,i_0}). \tag{4.14}$$

This implies that there exists one solution χ of (4.13) and $\chi^{-1} = \frac{\lambda_j - \sum_{i=1}^{i_0} v_{j,i}}{i_0}$.

We are now interested in the particular case when $X_j = R_j \operatorname{Diag}(h_j) \in \mathbb{R}^{N_t \times N_t P_j}$. To compute the proximity operator of the nuclear norm at X_j, we need to determine matrices U_j, V_j and $\mathcal{D}_{\chi^{-1}}(\Sigma_j)$ as stated in (4.12). In this case, X_j is a block diagonal matrix having the same form as matrix R_j. More precisely, X_j can be written as follows:

$$X_j = \begin{bmatrix} X_j^{(0)} & 0 & \dots & 0 \\ 0 & X_j^{(1)} & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & X_j^{(N_t-1)} \end{bmatrix}$$

where, for every $n_t \in \{0, \dots, N_t - 1\}$, $X_j^{(n_t)}$ are row vectors of length P_j. We have

$$X_j X_j^\top = \operatorname{Diag}\left(\left[\|X_j^{(0)}\|^2,\ \|X_j^{(1)}\|^2,\ \dots,\ \|X_j^{(N_t-1)}\|^2 \right] \right).$$

On the other hand we have

$$X_j X_j^\top = U_j (\Sigma_j)^2 U_j^\top. \tag{4.15}$$

Let define the set $\boldsymbol{\mathcal{N}}_{T_j}$ as

$$\boldsymbol{\mathcal{N}}_{T_j} = \left\{ n_t \in \{0, \dots, N_t - 1\} \mid\ X_j^{(n_t)} \neq \mathbb{0}_{P_j} \right\}$$

where $\mathbb{0}_{P_j}$ is a row vector of zeros of dimension P_j. Without loss of generality, let assume that the matrix X_j is ordered such that $\boldsymbol{\mathcal{N}}_{T_j} = \{0, 1, \dots, T_j - 1\}$. Consequently, we have

$$\Sigma_j = \operatorname{Diag}\left(\left[\|X_j^{(n_t)}\| \right]_{n_t \in \boldsymbol{\mathcal{N}}_{T_j}} \right) \in \mathbb{R}^{T_j \times T_j} \tag{4.16}$$

and

$$U_j = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{N_j \times T_j}.$$

Finally, $V_j = X_j^\top U_j \Sigma_j^{-1}$.

Note that, when we define $L_j : \mathbb{R}^{N_t P_j} \to \mathbb{R}^{N_t \times N_t P_j} : h_j \mapsto \widetilde{R}_j \operatorname{Diag}(h_j)$ where

$$\widetilde{R}_j = \begin{bmatrix} \mathbb{1}_{P_j} & 0 & \dots & 0 \\ 0 & \mathbb{1}_{P_j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbb{1}_{P_j} \end{bmatrix},$$

and $\mathbb{1}_{P_j}$ is a row vector of ones of dimension P_j, the projection onto the "nuclear-ball" is equivalent to the projection onto the mixed $\ell_{1,2}$-norm.

Finally, as mentioned earlier in Section 4.3.3, the prior information on the primary y is expressed through the frame analysis operator F by splitting constraints into individual subband constraints. In order to promote sparsity of the coefficients, the potential function employed for the l-th subband with $l \in \{1, \dots, \mathcal{L}\}$ can be chosen equal to $\widetilde{\varphi}_l = |\cdot|$. For computing the resulting projection Π_{D_l} onto an ℓ_1-ball, we can again employ the iterative procedure proposed in (van den Berg and Friedlander, 2008).

4.4.2 M+LFBF algorithm

The primal-dual approach chosen to solve the minimization problem (4.5) is detailed in Algorithm 5.5.1. It alternates computations of the gradient of Ψ, and of the projections onto $(C_m)_{1 \le m \le 3}$ and $(D_l)_{1 \le l \le \mathcal{L}}$.

The choice of the step size is crucial for the convergence speed and it has to be selected 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:

$$|||[\mathrm{I}\ \mathbf{R}]||| \le \sqrt{1 + |||R_0|||^2 + \dots + |||R_{J-1}|||^2} \tag{4.17}$$

where $|||R_j||| = \max_{n_t \in \{0,\dots,N_t-1\}} \|R_j^{(n_t)}\|$ for every $j \in \{0, \dots, J-1\}$. 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 (4.8), 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 (Chaâri et al., 2009, Algorithm 4). It is important to emphasize that the convergence of this algorithm to an optimal solution to Problem (4.5) is guaranteed by (Combettes and Pesquet, 2012, Theorem 4.2). In practice, the higher the norms of F and $(R_j)_{0 \le j \le J-1}$, the slower the convergence of the algorithm. In order to circumvent this difficulty, one can resort to a preconditioned version of the algorithm (Repetti et al., 2012). However, this was not found to be useful in our experiments.

Algorithm 4 Primal-dual algo. M+LFBF to solve (4.5).

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

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

for $i = 0, 1, \dots$ **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 $l = 1 : \mathcal{L}$ **do**

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

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

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 $l = 1 : \mathcal{L}$ **do**

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

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

end for

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

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

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

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

4.5 Results

4.5.1 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. Realistic synthetic data are obtained from a modeled seismic trace with primaries y. 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, $\forall p \in \{p', \ldots, p' + P_j - 1\}$, $\overline{h}_j^{(n_t)}(p) = \eta_j^{(n_t)}/P_j$ (cf. Eq. (4.2)). The time-varying filters are thus unambiguously defined, at a given time n_t, by the constants $\eta_j^{(n_t)}$. 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.

More specifically, we used a standard practice in geophysics. The simulated data is based on the simulated data used in (Ventosa et al., 2012). A model of earth layer is built, a series of spikes that mimic a reflectivity sequence (Ricker, 1943), and wave propagation is emulated by propagation of so-called seismic wavelets, here a Ricker wavelet (a shifted second derivative of a Gaussian, (Ryan, 1994)). Such codes are effective to simulate the seismic sensor response, by allowing only primary reflections, or both primary and multiple reflections. By varying the wave celerity in medium, this allows us to emulate different types of multiples. From this modeling, we extract two slightly different and clean multiple signals. The medium is chosen so that the multiples partly overlap the primaries. We now forget about the model, and consider the two modeled multiples as templates.

To test the proposed methodology, we now build two families of time-varying filters, as explained in this chapter. At each time index, the filters have uniform coefficients. These constants follow laws, varying along the time index, given in Figure 4.2 - p. 65. The time-varying filters associated with the first templates have higher amplitude at the beginning, while the second have higher amplitude at the end. In the middle, both filters possess the same amplitude.

The simulated multiple is composed of the sum of the two templates, each weighted by one of the two time-varying filters. Thus we obtain a synthetic trace with isolated primaries, and primaries mixed with multiple reflexions. Such simple models are not fully representative of seismic reality, as for phantoms in image processing. They nevertheless permit to test, calibrate and evaluate algorithms, as the ground truth is known. So it is important to note that the templates are generated by seismic wave propagation, and not chosen in an ad-hoc manner.

4.5.2 Qualitative results on synthetic data

We choose here two filter families with lengths $P_0 = 10$ and $P_1 = 14$. The two filters evolve complementally in time, emulating a bi-modal multiple mixture at two different depths. They are combined with the two templates in a multiple signal depicted at the fifth row from the top of Figure 4.3 - p. 73. Data and template were designed in order to mimic the time and frequency contents of seismic signals. This figure also displays the other signals of interest, known and unknown, when the noise standard deviation is $\sigma = 0.08$. We aim at recovering weak primary signals, potentially hidden under both

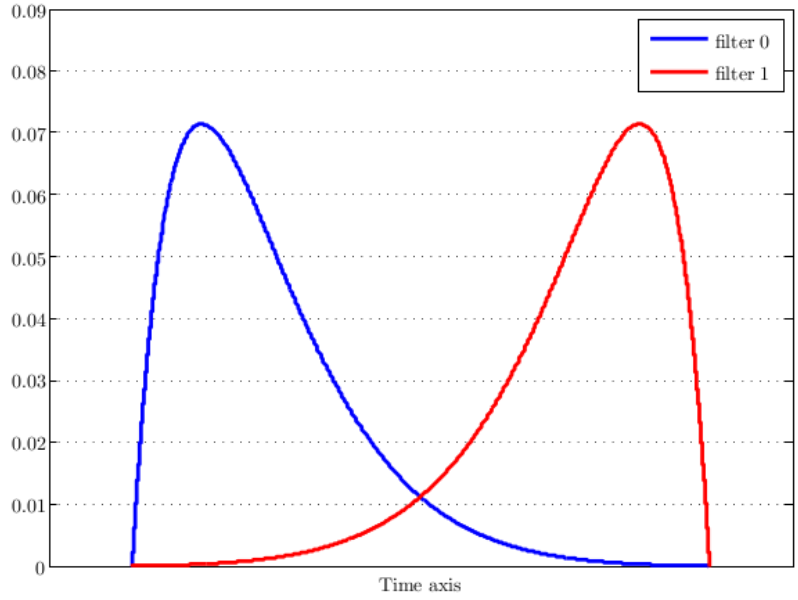


Figure 4.2: *Synthetic filter profiles.*

multiple and random perturbations, from the observed signal z 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 Figure 4.3 - p. 73 (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 Figure 4.5 - p. 75. 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.

For the proposed method, we choose the following initial settings. An undecimated wavelet frame transform with 8-length Symlet filters is performed on 4 resolution levels ($\mathcal{L} = 5$) . The loss functions ψ and $(\widetilde{\varphi}_l)_{1\leq l\leq \mathcal{L}}$ in (4.5) are chosen as $\psi = \|\,.\,\|^2$ and $\widetilde{\varphi}_l = |\,.\,|$. 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 (Combettes and Pesquet, 2007b, p. 1356). The best matching power for the chosen wavelet tight frame yields the taxicab metric or ℓ_1-norm, as illustrated in Figure 4.4 - p. 74.

The constraints C_1 and C_2 are chosen according to (4.6) and (4.7), 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 propose to infer those constants from other methods. In practice, alternative cruder filtering or restoration algorithms indeed exist, with the same purpose. They are often 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 (Ventosa et al., 2012). Finally, $\widetilde{\rho}$ is chosen as the $\ell_{1,2}$-norm.

Figure 4.5 - p. 75 provides close-ups of delimited areas with weak primaries. With a low noise level (Figure 4.5-top - p. 75), 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 Figure 4.5-bottom - p. 75. 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 above, the ability to restore — albeit imperfectly — spurs of strongly hidden primaries (cf. Figure 4.5 - p. 75, between indices 540 and 600) is of paramount interest for seismic exploration in greater depths.

4.5.3 Quantitative results on synthetic 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 Symlet 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 Table 4.1 - p. 68 and Table 4.2 - p. 69 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 4.2 - p. 69 as well) indicate the best result for a given decomposition level. Numbers in italics denote the best SNRs obtained, irrespective of the number of wavelet levels. The standard deviation tables are combined with Table 4.1 - p. 68 and Table 4.2 - p. 69 in Table 4.5 - p. 72 as a "significance index" of the outcome.

We first exemplify results in the leftmost part of Table 4.1 - p. 68 (Haar wavelet), for the first two rows. For the ℓ_1-norm, we observe a primary restoration improvement of 2.2 dB ($21.3 - 19.1$ dB, 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 additional 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 4.2 - p. 69, 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).

Average differences allow the observation of global trends. In practice, consistent results, taking into account SNR dispersion, are more important. Assuming that the denoised realizations follow a Gaussian distribution, we now study the difference Gaussian pdf between frame and basis results. 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 4.5 - p. 72 reports the normalized difference significance index $t_{f/b} = \mu_{f/b}/\sigma_{f/b}$, reminiscent of the Student's test. It is associated with the probability $\pi_{f/b}$ that, in an outcome of the realizations, the basis SNR is superior to the frame SNR. Table 4.3 - p. 70 and Table 4.4 - p. 71 show the standard deviations for SNRs from estimations of y in Table 4.1 - p. 68 and of s in Table 4.2 - p. 69, respectively. The Haar wavelet, ℓ_1-norm, primary restoration improvement of $\mu_{f/b} = 2.2$ dB yields $\sigma_{f/b} = 0.30$. Hence, $t_{f/b} = 7.2$. The interpretation of this significance is illustrated with the abacus in Figure 4.6 - p. 76, with $\pi_{f/b}$ associated to the shaded area. 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 four 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 4.5 - p. 72-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 4.5 - p. 72 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 Symlet 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.

			Noise levels (σ)							
			0.01		0.02		0.04		0.08	
wavelet	$\widetilde{\rho}$	$F \setminus \mathcal{L}$	4	5	4	5	4	5	4	5
Haar	ℓ_1	b	19.1	*20.6*	19.2	*19.8*	16.5	*16.7*	12.3	*12.4*
		f	21.3	*21.8*	20.3	*20.8*	18.2	*18.5*	14.7	*15.0*
	ℓ_2	b	20.2	*21.6*	20.1	*20.5*	16.9	*16.9*	12.5	*15.0*
		f	**21.9**	***22.4***	**21.1**	***21.3***	**18.5**	***18.8***	**14.9**	***15.2***
	$\ell_{1,2}$	b	18.9	*20.3*	18.6	*19.5*	16.2	*16.5*	12.2	*12.3*
		f	21.3	*21.8*	20.1	*20.6*	18.0	*18.3*	14.6	*14.9*
Symlet	ℓ_1	b	*20.9*	20.8	*20.6*	20.1	*18.0*	18.0	*14.3*	14.2
		f	21.8	*22.0*	21.3	*21.6*	20.1	*20.2*	17.4	*17.5*
	ℓ_2	b	21.1	*22.1*	*21.2*	21.0	18.3	*18.3*	*14.4*	14.4
		f	**22.3**	***22.7***	**22.1**	***22.2***	**20.6**	***20.7***	**17.7**	***17.7***
	$\ell_{1,2}$	b	20.1	*21.1*	*20.4*	20.0	*17.9*	17.7	*14.2*	14.1
		f	21.8	*22.1*	21.2	*21.5*	19.9	*20.0*	17.3	*17.3*
Daubechies	ℓ_1	b	19.9	*20.3*	*20.1*	20.0	*17.9*	17.9	*14.1*	13.9
		f	21.8	*22.0*	21.5	*21.6*	20.2	*20.3*	*17.6*	17.5
	ℓ_2	b	21.4	*22.3*	21.4	*21.5*	*18.4*	18.21	*14.3*	14.0
		f	**22.5**	***22.8***	**22.3**	***22.4***	***20.8***	**20.7**	***17.9***	**17.7**
	$\ell_{1,2}$	b	19.2	*20.4*	*19.5*	19.2	*17.6*	17.5	*14.0*	13.2
		f	21.9	*22.1*	21.3	*21.6*	20.0	*20.1*	*17.5*	17.3

Table 4.1: *SNR, averaged over* 100 *noise realizations for the estimations of* y.

4.5.4 Comparative evaluation: synthetic data

In addition to the above objective and subjective results, we perform a comparative evaluation with the empirical algorithm proposed in (Ventosa et al., 2012). 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 in overlapping windows taken in the complex scalogram, i.e. the complex-valued, discretized, continuous wavelet transform. This algorithm was successfully tested against industry standards. It is quantitatively faster, but it does not permit the introduction of prior knowledge (Salaun

			Noise levels (σ)							
			0.01		0.02		0.04		0.08	
wavelet	$\widetilde{\rho}$	$F \setminus \mathcal{L}$	4	5	4	5	4	5	4	5
Haar	ℓ_1	b	26.3	*27.6*	25.3	*25.8*	22.1	*22.3*	18.4	*18.4*
		f	27.0	*28.3*	25.6	*25.7*	22.1	*22.2*	18.4	*18.4*
	ℓ_2	b	27.4	*28.4*	**25.7**	***26.0***	21.8	*21.9*	17.7	*17.8*
		f	**28.4**	***28.7***	**25.7**	*25.8*	21.8	*21.8*	*17.8*	*17.8*
	$\ell_{1,2}$	b	26.2	*27.4*	25.0	*25.7*	22.2	***22.4***	18.6	***18.6***
		f	28.0	*28.4*	25.5	*25.8*	**22.3**	*22.3*	***18.6***	*18.6*
Symlet	ℓ_1	b	*27.6*	27.6	***25.8***	25.5	*22.3*	22.2	*18.5*	18.5
		f	28.0	*28.1*	25.5	*25.6*	22.1	*22.1*	18.4	*18.4*
	ℓ_2	b	28.2	***28.6***	*25.8*	25.7	20.1	*21.8*	*17.8*	*17.8*
		f	**28.3**	*28.5*	25.6	***25.6***	*21.8*	*21.8*	*17.8*	17.7
	$\ell_{1,2}$	b	26.9	*27.7*	***25.8***	25.5	***22.5***	**22.4**	***18.7***	**18.7**
		f	28.0	*28.1*	25.5	***25.6***	*22.3*	22.2	*18.6*	18.6
Daubechies	ℓ_1	b	26.7	*27.1*	*25.3*	25.3	*22.1*	21.2	18.4	*18.4*
		f	28.0	*28.1*	25.5	*25.6*	22.1	*22.1*	18.4	*18.4*
	ℓ_2	b	27.9	*28.5*	**25.8**	***25.9***	21.8	*21.8*	*17.8*	*17.8*
		f	**28.4**	***28.6***	25.6	*25.6*	21.8	*21.8*	*17.8*	17.8
	$\ell_{1,2}$	b	26.5	*27.1*	25.1	*25.3*	*22.2*	*22.2*	***18.6***	**18.6**
		f	28.0	*28.2*	25.5	*25.6*	***22.3***	**22.3**	***18.6***	18.6

Table 4.2: *SNR, averaged over* 100 *noise realizations for the estimations of* s.

et al., 2015) on signal sparsity or filter regularity. Figure 4.7 - p. 77 presents synthetic 2D seismic data, constructed similarly to the previous 1D traces in Section VI-B, with a high noise level ($\sigma = 0.08$). Vertical traces are stacked laterally to form a 2D image. 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, 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.

4.5.5 Comparative evaluation: real data

The previous simulated example is a little bit simplistic. We finally compare our algorithm with (Ventosa et al., 2012) on a portion of a real seismic data set. Recorded and multiple template data belong to the same marine seismic survey processed in (Ventosa et al., 2012). The recorded seismic data is displayed in Figure 4.8a - p. 78. The main objective is to uncover a potential primary, masked by strong multiple events that mostly contribute to the observed total seismic signal energy. The primary appears partially

			Noise levels (σ)							
			0.01		0.02		0.04		0.08	
wavelet	$\widetilde{\rho}$	$F \setminus \mathcal{L}$	4	5	4	5	4	5	4	5
Haar	ℓ_1	b	0.26	*0.22*	***0.22***	0.26	***0.33***	0.35	***0.43***	**0.46**
		f	***0.16***	0.18	*0.25*	0.27	*0.37*	*0.37*	0.53	*0.52*
	ℓ_2	b	0.26	*0.24*	*0.26*	*0.26*	***0.33***	0.35	*0.44*	**0.46**
		f	0.29	***0.16***	0.24	***0.23***	*0.34*	***0.34***	*0.51*	*0.51*
	$\ell_{1,2}$	b	0.26	*0.24*	0.36	*0.30*	0.37	*0.36*	*0.44*	0.47
		f	*0.17*	0.19	*0.25*	0.31	*0.39*	0.40	*0.52*	0.53
Symlet	ℓ_1	b	0.27	***0.19***	*0.31*	0.32	*0.41*	0.43	***0.53***	**0.59**
		f	***0.19***	0.22	*0.34*	0.39	*0.54*	0.58	*0.72*	0.78
	ℓ_2	b	0.26	***0.19***	***0.25***	**0.26**	**0.39**	***0.38***	***0.53***	**0.59**
		f	0.34	*0.21*	*0.34*	0.35	*0.46*	0.49	*0.65*	0.72
	$\ell_{1,2}$	b	0.28	*0.21*	*0.34*	0.36	*0.44*	0.47	*0.55*	**0.59**
		f	*0.21*	0.22	*0.37*	0.41	*0.58*	0.63	*0.75*	0.82
Daubechies	ℓ_1	b	0.29	*0.23*	*0.35*	0.38	*0.44*	0.47	***0.53***	0.55
		f	***0.19***	**0.22**	*0.34*	0.38	*0.56*	0.59	*0.74*	0.78
	ℓ_2	b	0.27	*0.26*	***0.29***	**0.36**	***0.42***	**0.43**	***0.53***	**0.54**
		f	0.32	*0.26*	*0.34*	**0.36**	*0.48*	0.53	*0.67*	0.72
	$\ell_{1,2}$	b	0.31	*0.26*	0.47	*0.39*	*0.47*	0.48	***0.53***	0.55
		f	*0.21*	**0.22**	*0.37*	0.40	*0.59*	0.62	*0.77*	0.82

Table 4.3: *Standard deviations for SNRs from estimations of y in Table 4.1 - p. 68.*

as a wiggling, horizontal stripes in the bottom part of the figure, on the right side. Geologically speaking, it should be re-linked with the left side of the picture, to one of the dimmed sloping stripes. By looking at differences between the recorded data and the multiple template in Figure 4.8b - p. 78, the trace of the flat primary may appear more obvious. Templates are obtained by different involved seismic modeling techniques, whose details (Pica et al., 2005b) are beyond the scope of the work. The core of adaptive multiple removal techniques in seismic data boils down to locally adapting the patterns in Figure 4.8b - p. 78 in location and amplitude to the data in Figure 4.8a - p. 78. Once adapted, the approximate patterns may be subtracted from the observed signal, with the hope of unveiling previously hidden signals.

The efficiency of a seismic data processing algorithm is difficult to assess, due to the absence of ground truth. One of the challenges of present seismic data processing resides in the ability to identify deeper target. To this aim, either noisier data sets or broadband seismic acquisitions are being address by geophysical signal processing. Figure 4.8-middle - p. 78 thus compares the results obtained with (Ventosa et al., 2012) (c) and the proposed algorithm (d). Although the random noise is apparently highly heteroskedastic, both methods are able to successfully retrieve the weak primary below the multiple level, especially of the left side of the figure. The method in (Ventosa et al.,

			Noise levels (σ)							
			0.01		0.02		0.04		0.08	
wavelet	$\widetilde{\rho}$	$F \setminus \mathcal{L}$	3	4	3	4	3	4	3	4
Haar	ℓ_1	b	0.19	*0.17*	0.23	*0.21*	*0.26*	0.27	*0.30*	0.31
		f	*0.17*	0.18	*0.23*	*0.23*	*0.26*	*0.26*	*0.30*	*0.30*
	ℓ_2	b	*0.20*	*0.20*	*0.25*	*0.25*	*0.26*	*0.26*	0.30	*0.29*
		f	0.26	*0.20*	*0.25*	*0.25*	*0.26*	*0.26*	*0.29*	0.30
	$\ell_{1,2}$	b	0.18	*0.17*	0.24	*0.23*	*0.26*	*0.26*	*0.30*	0.31
		f	*0.16*	0.17	*0.22*	*0.22*	*0.25*	0.26	*0.30*	*0.30*
Symlet	ℓ_1	b	0.25	*0.14*	0.23	*0.22*	0.27	*0.26*	*0.31*	*0.31*
		f	*0.17*	0.18	*0.22*	0.24	*0.25*	0.26	*0.30*	*0.30*
	ℓ_2	b	0.24	*0.19*	*0.24*	*0.24*	*0.26*	*0.26*	*0.29*	0.30
		f	0.28	*0.20*	*0.24*	*0.24*	*0.25*	*0.25*	*0.29*	*0.29*
	$\ell_{1,2}$	b	0.27	*0.18*	0.25	*0.23*	*0.27*	*0.27*	*0.30*	*0.30*
		f	*0.16*	0.17	*0.21*	0.23	*0.25*	0.26	*0.30*	*0.30*
Daubechies	ℓ_1	b	0.18	*0.15*	*0.21*	0.23	*0.25*	0.26	*0.30*	0.31
		f	*0.16*	0.17	*0.22*	0.23	*0.25*	0.26	*0.29*	0.30
	ℓ_2	b	*0.19*	0.20	*0.25*	*0.25*	*0.26*	*0.26*	*0.30*	*0.30*
		f	0.25	*0.22*	0.24	*0.23*	*0.25*	*0.25*	*0.29*	*0.29*
	$\ell_{1,2}$	b	0.19	*0.17*	0.24	*0.23*	*0.26*	0.27	*0.30*	*0.30*
		f	*0.16*	*0.16*	*0.21*	0.22	*0.25*	0.26	*0.30*	*0.30*

Table 4.4: *Standard deviations for SNRs from estimations of s in Table 4.2 - p. 69.*

2012) may suffer from a little more pre-echo above the primary in the top-left corner of Figure 4.8(c) - p. 78, while a remnant $-45°$ shadow affects the proximal multiple removal in its central part of Figure 4.8d - p. 78.

The increased robustness to noisier seismic data is estimated with a wide-band Gaussian noise, added to the seismic field data. The outcome is illustrated in Figure 4.8-bottom - p. 78. While the primary can still be tracked with (Ventosa et al., 2012) in Figure 4.8e - p. 78, it dims inside the ambient noise on the left-most side. The proposed template-based multiple filtering is more robust to noise, reflecting its practical potential. Naturally, the anisotropic, oriented nature of seismic data, and the directional diversity of primaries and multiples, suggests an extension to oriented frames in two dimensions as presented in the newt chapter.

4.6 Conclusion

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

			Noise levels (σ)							
			0.01		0.02		0.04		0.08	
wavelet	$\widetilde{\rho}$	$F \setminus \mathcal{L}$	4	5	4	5	4	5	4	5
Haar	ℓ_1	b	*7.3*	*4.2*	*3.0*	**2.6**	*3.4*	*3.5*	**3.5**	**3.8**
		f	*4.3*	*2.8*	*2.7*	*2.3*	**3.5**	**3.7**	**3.5**	**3.8**
	ℓ_2	b	**7.7**	**5.1**	**3.4**	*2.5*	*3.2*	*3.3*	**3.5**	*3.7*
		f	*2.6*	**4.2**	*1.5*	**3.0**	*3.0*	*3.2*	*3.5*	*3.4*
	$\ell_{1,2}$	b	*3.0*	*1.9*	**2.1**	*2.9*	**3.9**	**3.8**	**3.9**	**3.6**
		f	**4.9**	*3.3*	*1.6*	*2.7*	*2.7*	*2.8*	*3.3*	*3.2*
Symlet	ℓ_1	b	*5.6*	**5.3**	*2.8*	*3.0*	*3.2*	*3.3*	*3.8*	*3.8*
		f	*2.6*	*1.3*	*2.1*	*1.7*	**3.8**	**3.7**	**4.2**	*4.1*
	ℓ_2	b	**7.2**	*5.1*	**3.1**	**4.3**	*3.2*	*3.3*	*3.7*	**4.2**
		f	*2.7*	*2.9*	0.7	-0.1	0.0	-0.3	0.0	0.0
	$\ell_{1,2}$	b	*3.2*	*1.1*	0.0	-0.6	0.0	-0.1	0.0	0.0
		f	**7.5**	**4.2**	*1.6*	0.1	0.1	-0.2	0.1	0.0
Daubechies	ℓ_1	b	*1.3*	**2.5**	*-1.1*	0.3	-0.6	-0.3	-0.2	-0.1
		f	0.3	-0.2	-0.8	-0.3	**4.7**	-0.1	0.0	0.0
	ℓ_2	b	**3.5**	*1.6*	-0.9	0.2	-0.6	-0.3	-0.1	-0.2
		f	*5.3*	**4.4**	0.5	0.8	-0.2	**2.6**	-0.1	-0.1
	$\ell_{1,2}$	b	*1.5*	0.1	-0.5	-0.7	-0.1	-0.1	0.0	0.0
		f	**6.0**	**4.4**	**1.5**	**1.2**	0.1	0.1	0.0	0.0

Table 4.5: *Normalized difference significance index.*

adaptive filter design under appropriate regularity and amplitude coefficient 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 adaptive filter design. Its performance is interesting for instance in recovering weak signals buried under both strong random and structured noise. Provided appropriate templates are obtained, this structured-pattern filtering algorithm could be useful in other application areas, e.g. acoustic echo-cancellation in sound and speech, non-destructive testing where transmitted waves may rebound at material interfaces (e.g. ultrasounds), or pattern matching in images. In the next chapter, two-dimensional directional multiscale approaches (Jacques et al., 2011) are considered and may provide sparser representations for seismic data.

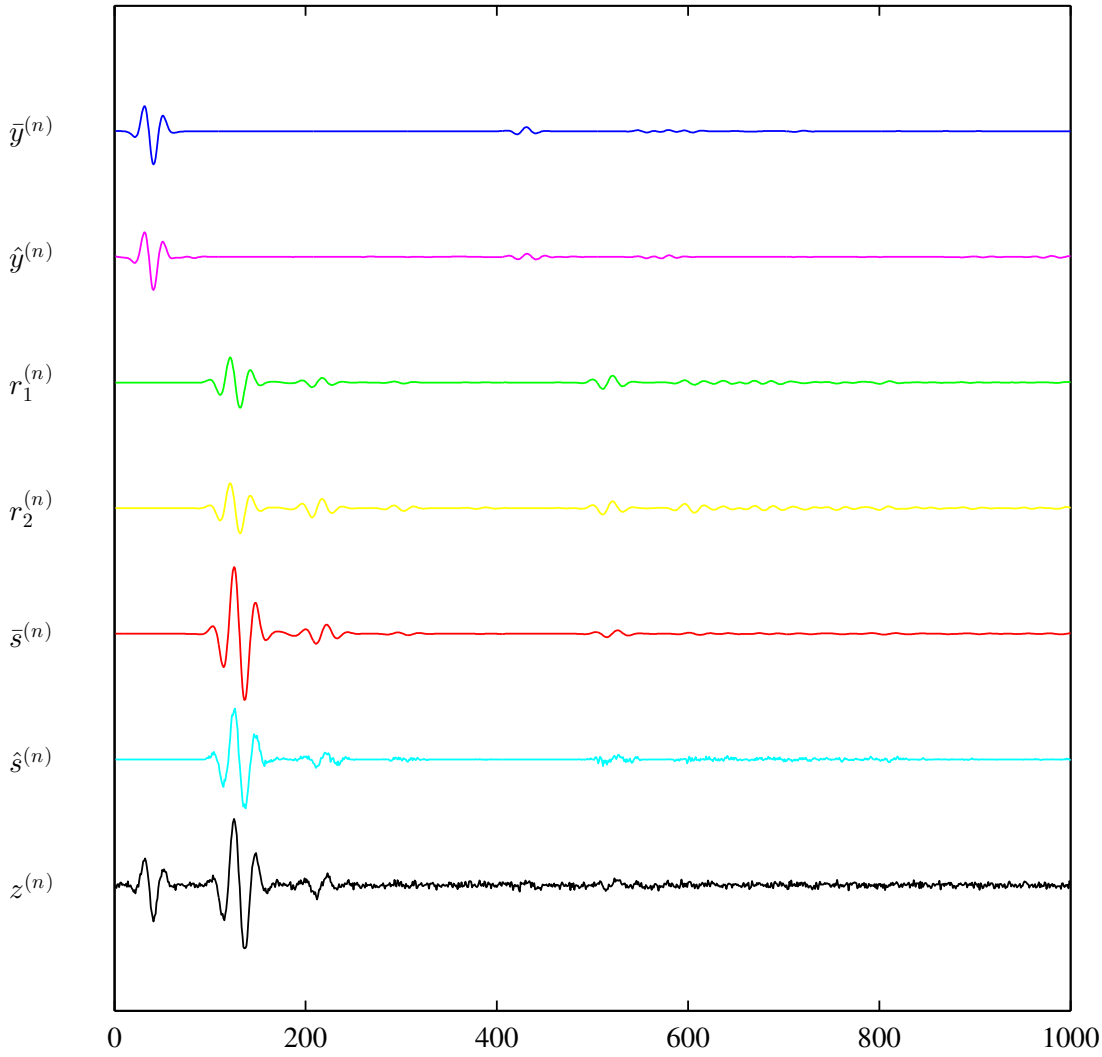
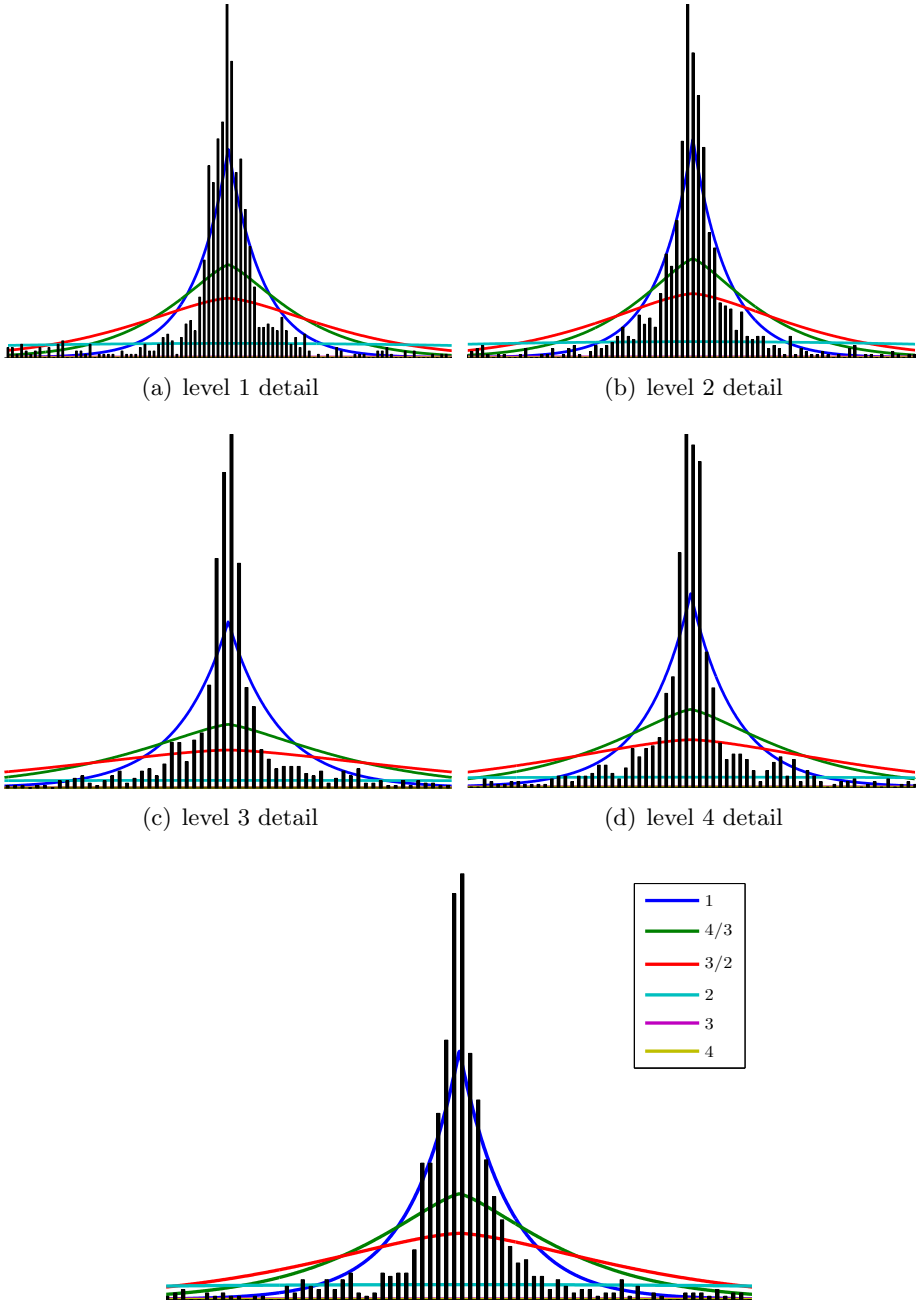


Figure 4.3: *Considered simulated seismic signals with noise level* $\sigma = 0.08$. *From top to bottom: primary (unknown)* $\bar{y}$, *estimated* $\hat{y}$, *first template* r_0, *second template* r_1, *multiple (unknown)* $\bar{s}$, *estimated* $\hat{s}$, *and observed signal* z.



(a) level 1 detail (b) level 2 detail

(c) level 3 detail (d) level 4 detail

(e) approximation

Figure 4.4: *Generalized Gaussian modeling of seismic data wavelet frame decomposition with different power laws.*

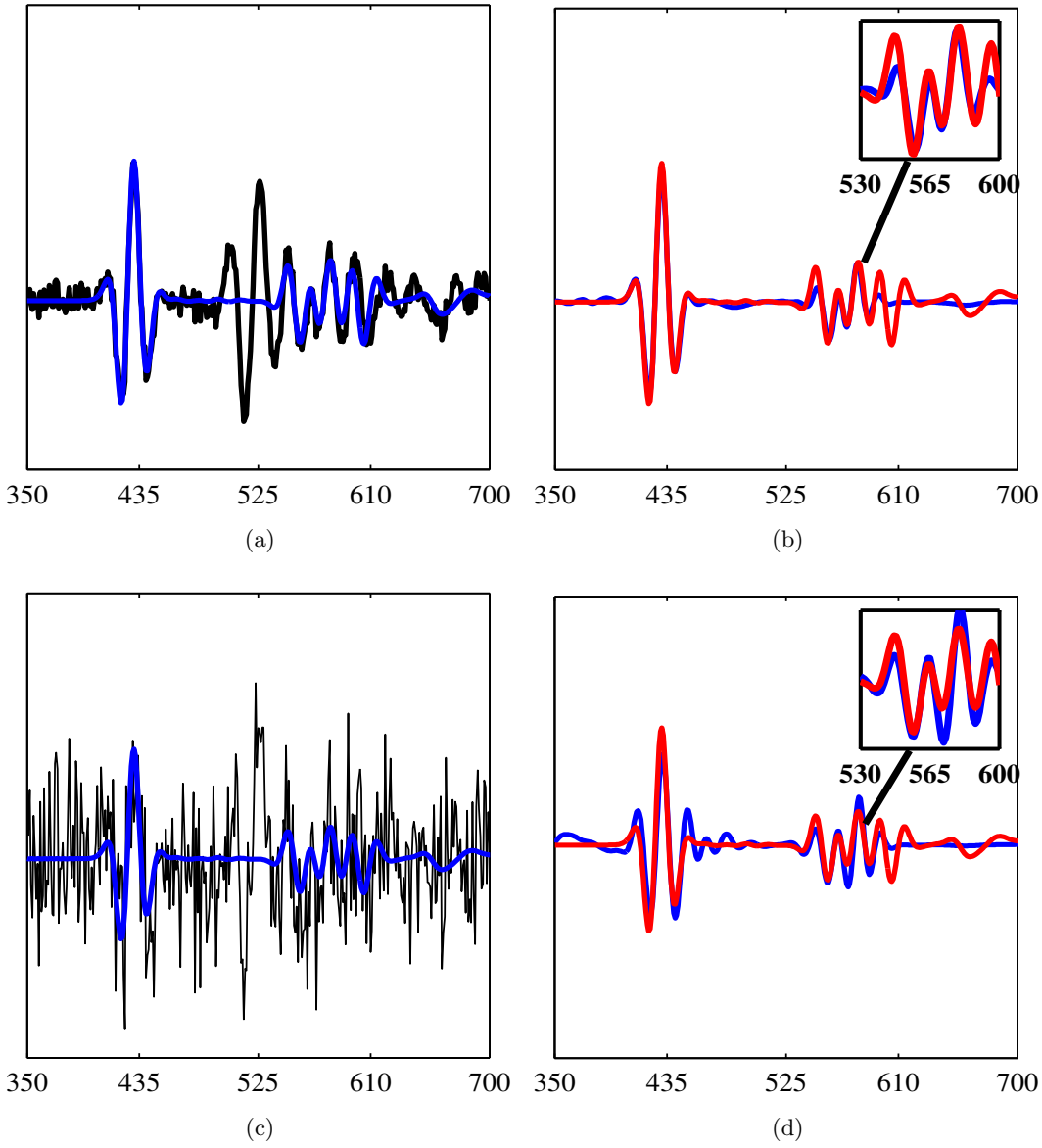


Figure 4.5: *Close-up when $\widetilde{\rho}$ is the $\ell_{1,2}$-norm; (a) input data z (black line $\sigma = 0.01$), primary $\bar{y}$ (blue line), (b) output separated primary $\hat{y}$ (red line) and primary $\bar{y}$ (blue line); (c) input data z (black line $\sigma = 0.08$), primary $\bar{y}$ (blue line), (d) output separated primary $\hat{y}$ (red line) and primary $\bar{y}$ (blue line).*

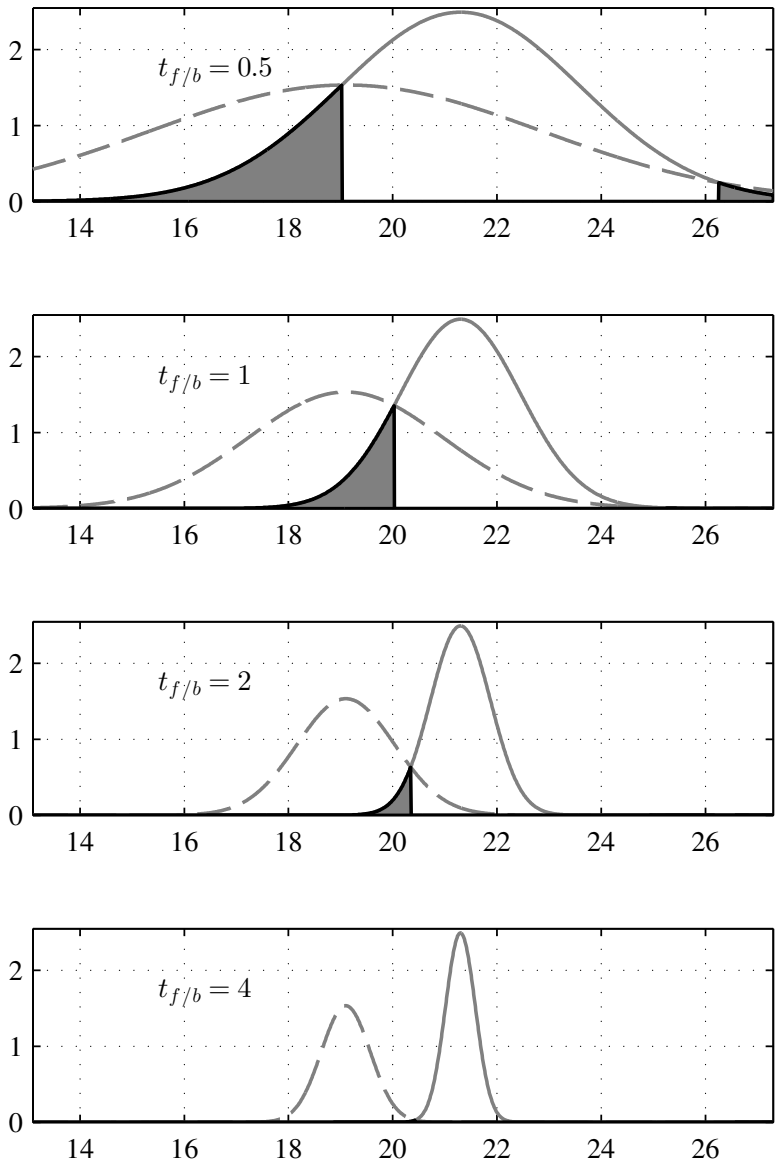
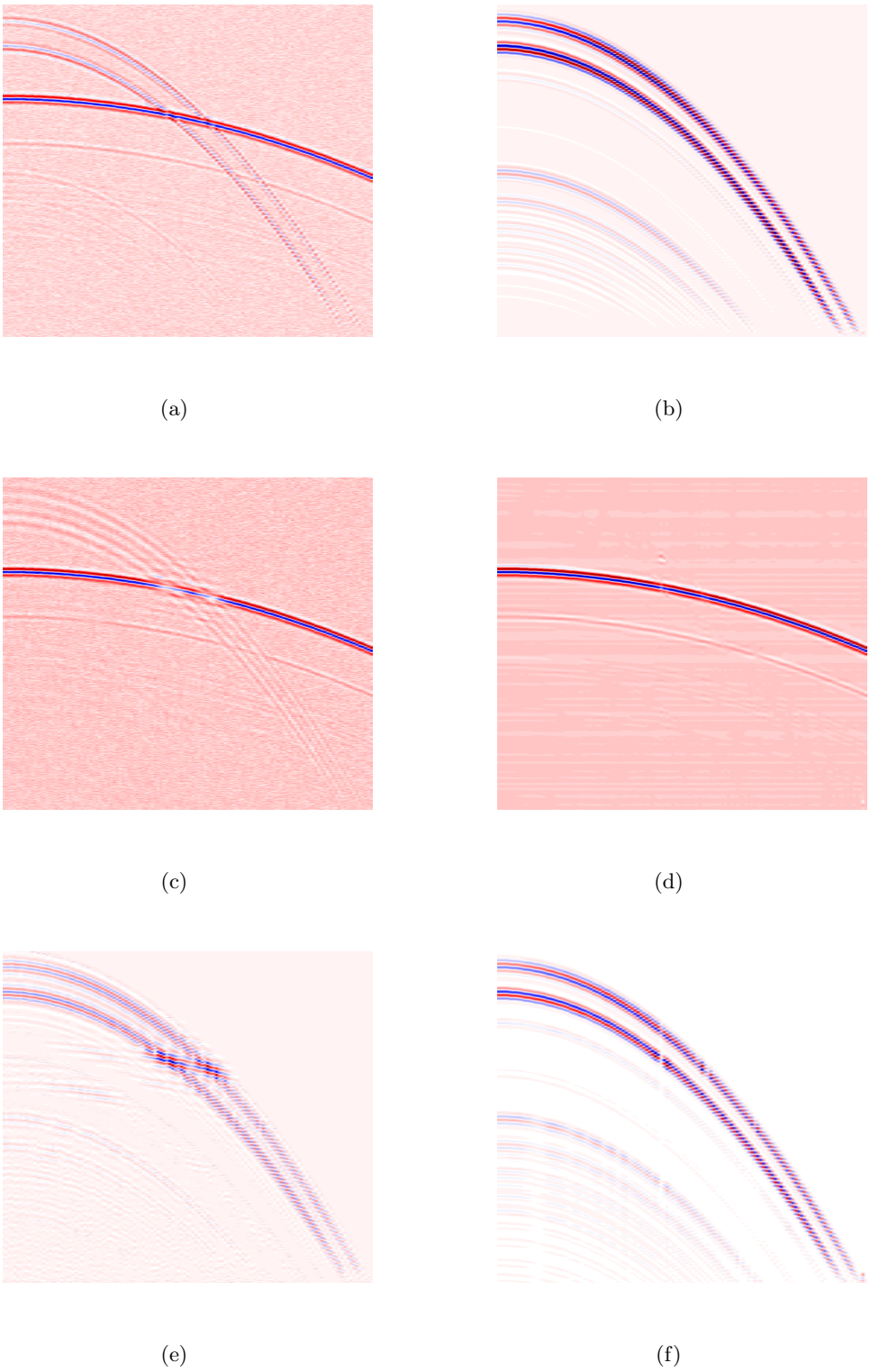


Figure 4.6: *Significance index abacus with different "significance levels" (and probability $\pi_{f/b}$, shaded): 0.5 (0.17), 1 (0.14), 2 (0.05) and 4 (0.0013).*



(a) (b)

(c) (d)

(e) (f)

Figure 4.7: *Data composed by three events ($\sigma = 0.08$), one primary and two multiples, SNR = 1.71dB (a); multiples composed by two estimated events (b); output separated primaries with (Ventosa et al., 2012), SNR = 3.11 dB (c) and our method, SNR = 16.77 dB (d); output adapted multiples with (Ventosa et al., 2012), SNR = 3.1 dB (e) and our method, SNR = 15.44 dB (f).*

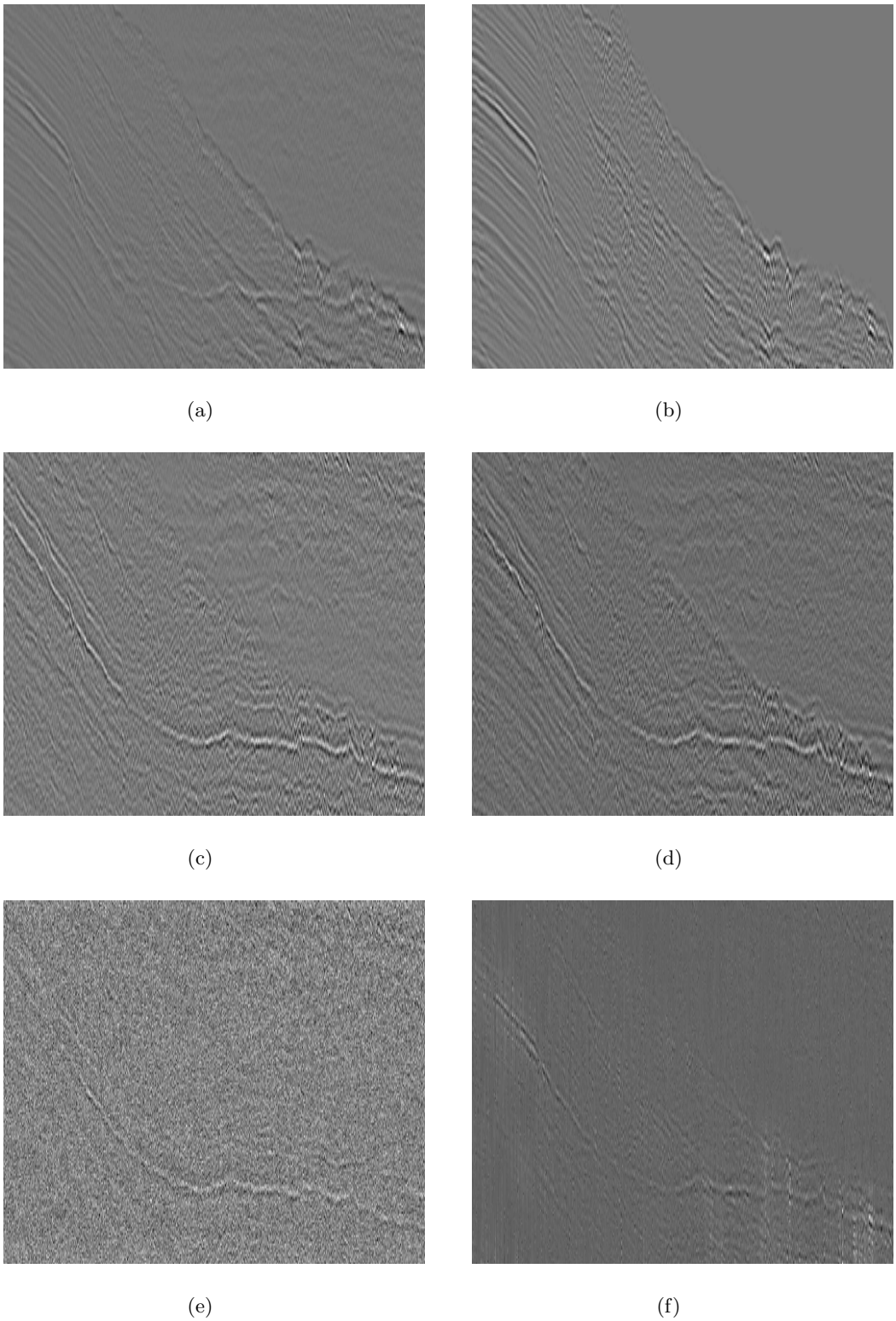


Figure 4.8: *Portion of a common receiver gather: (a) recorded seismic data with a partially appearing primary, (b) multiple wavefield template; subtraction results, low field-noise case: primaries (separated from multiples) with (c) (Ventosa et al., 2012), (d) the proposed method; subtraction results, high field-noise case: primaries (separated from multiples) with (e) (Ventosa et al., 2012), (f) the proposed method.*

- Chapter 5 -

Multiple removal in 2D

As mentioned earlier, seismic data consist of a collection of seismic traces. These collections may appear in different dimensions. 2D seismic images would illustrate a single slice of the subsurface (see Figure 5.1 - p. 80 - image from (Hill, 2014)), 3D seismic a volume of the earth and 4D seismic depicts a 3D volume at different times in the life of an oil and/or gas field. In previous chapters, we explored the problem for 1D seismic data with wavelet-frame-based adaptive filtering. Our motivation was a thorough assessment of transforms and optimization tools. However, it is often considered that addressing such data in an higher dimension is often beneficial. The main drive relates to the adjunction of additional dimensions, for instance spatial. It arises from the nature of the acquisition, where the first dimension reflects time (wave propagation), while other dimensions are governed by space (the sensors's location in an array of sensors) or time. Generally, geological structures present a lateral continuity that reflects in wave propagation. It turns into an apparent anisotropy in seismic data. The aim of this chapter is to address an extension of our previous approach in the second spatial dimension.

5.1 Introduction

Most of the seismic methods are implemented on 2D, 3D or 4D data. However several seismic data methods are designed in 1D (in time) (Ventosa et al., 2012; Pham et al., 2014b) or space, sometimes in a separable fashion. In the latter case, we only considered the depth without taking into account additional structures. Actually, a seismic sensor yields an acoustic signal similar to sounds or ultrasounds. In contrast, seismic images do not share all the characteristics of standard or natural images. More precisely, all pixels of standard images are taken at a same time, thus the image points are related to an homogeneous space location. A common practice in image processing is to use separable transforms, nearly separable but directional ones (Chaux et al., 2006), or more anisotropic representations (Jacques et al., 2011) such as curvelets, shearlets or with finer frequency decomposition as in (Aujol et al., 2003). However, the distinct nature of rows

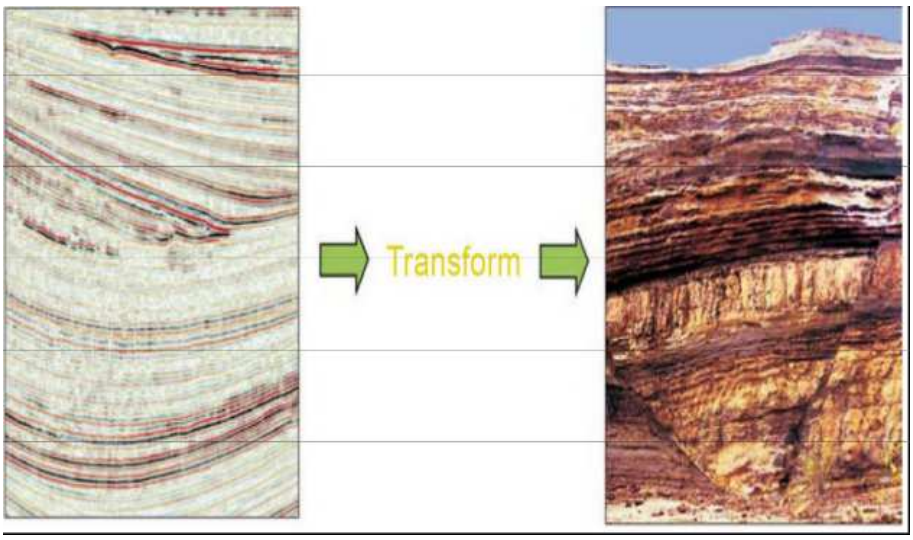


Figure 5.1: *Seismic data gives the explorationist a picture of the geology of the subsurface (image from (Hill, 2014)).*

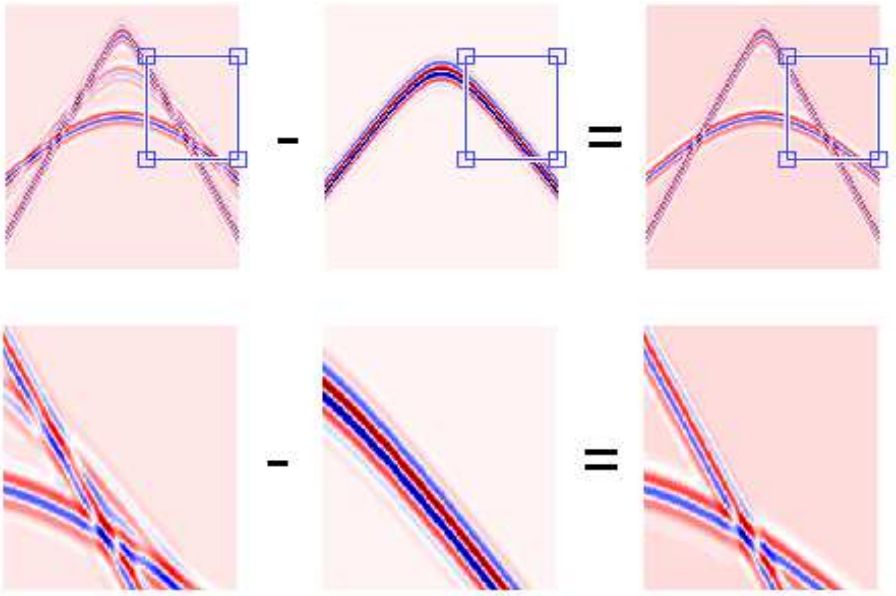


Figure 5.2: *From left to right: observed image z, multiple $\overline{s}$, and primary $\overline{y}$.*

and columns in seismic data suggests the use of different transforms along time and space. This chapter focuses on the use of the hybrid dual-tree M-band wavelet. We first review some works in multiple removal for 2D seismic data.

In (Barnes, 1996), the author observed that, the processing in two-dimensional frequency and bandwidth is appropriate for the migrated data (the process by which seismic events are geometrically re-located in either space or time to the subsurface location the event occurred in rather than the location where it was recorded at the surface, thereby creating a more accurate image of the subsurface (Sheriff, 1989, Chapter 21)) by using the 2D Hilbert transform and 2D complex trace attributes. For unmigrated data (opposite of migrated data), one-dimensional frequency and bandwidth are appropriate. Multiples can be attenuated by filtering in the $f - k$ (Wu and Wang, 2011) or $\tau - p$ domain (Yilmaz, 2001). The 2D Hilbert transform has also been proposed in (Karsli et al., 2006). Separating waves in general or multiple removal in particular is related to the wave propagation, (Andersson and Duchkov, 2013) extended a structure tensor using a one-way (in time) wave equation with a constant velocity as a tool for estimating a two-dimensional field. However it can be applied to generic images not related to wave propagation. Indeed, seismic images possess geometric regularity that advises a 2D approach for improved performance. As a result, multiscale geometric transforms such as curvelet frames (Herrmann and Verschuur, 2004; Neelamani et al., 2010; Ma and Plonka, 2010; Donno et al., 2010) have been employed in geophysical processing.

As mentioned above, adaptive multiple removal is often performed in two or three dimensional seismic data, via standard ℓ_2 or more robust ℓ_1-norms, with local multidimensional matching FIR filters. The multidimensionality of the filters is thought to ensure lateral continuity in seismic events. A quite opposite direction was taken in (Ventosa et al., 2012), with an emphasis on a frequency- and shift-insensitive complex wavelet transform frame, associated with simple unary (one-tap) complex filters, in 1D only. Counter-intuitively, it was able to perform similarly to more classical 2D matching techniques. In other words, a careful partnership between sparse representations and adaptive filtering was deemed beneficial in 1D, with respect to traditional 2D methods. An interpretation of this observation is that a more accurate spectral processing in 1D may yield perceptual continuity often credited to 2D methods. Indeed, 2D matching filters resemble directional filters (Donno, 2011). To account for additional properties, including statistical distributions for primaries (Costagliola et al., 2012) and slow filter variations (Le Touzé et al., 2009), (Pham et al., 2014b,a) pursued seismic data adaptive filtering with 1D wavelet frames.

The remainder of the chapter is organized as follows: Section 5.2 presents the considered data modelization. Section 5.3 proposes the approach to solve the problem. Then, the implemented algorithm is detailed in Section 5.4. Simulation results are drawn in Section 5.5 for both simulated and real field seismic data. We then demonstrate the effectiveness of our proposed 2D method through the comparisons with the previous 1D method (Pham et al., 2014b) and the industrially-validated method (Ventosa et al., 2012). This method demonstrated good performance on field data with (Salaun et al., 2015). Finally, Section 5.6 concludes the chapter.

5.2 Model description

This section aims at describing the model accounting for multiple reflections in 2D seismic data. As mentioned above, a line of seismic sensors delivers a two-dimensional image. Each column is formed by a 1D temporal signal acquired by one sensor:

$$z^{(\mathbf{n})} = \overline{s}^{(\mathbf{n})} + \overline{y}^{(\mathbf{n})} + b^{(\mathbf{n})} \tag{5.1}$$

with $\mathbf{n} = (n_t, n_o)$, where $n_t \in \mathbb{N}_t \triangleq \{0, \ldots, N_t - 1\}$ is the time index, $n_o \in \mathbb{N}_o \triangleq \{0, \ldots, N_o - 1\}$ is the sensor index, and $\mathbf{n} \in \boldsymbol{\mathcal{N}} \triangleq \{(n_t, n_o) | \, n_t \in \mathbb{N}_t, \, n_o \in \mathbb{N}_o\}$. Observed data $z = (z^{(\mathbf{n})})_{\mathbf{n}\in\boldsymbol{\mathcal{N}}}$ (Figure 5.2-left - p. 80) is composed of the primary (Figure 5.2-right - p. 80) $\overline{y} = (\overline{y}^{(\mathbf{n})})_{\mathbf{n}\in\boldsymbol{\mathcal{N}}}$ (2D data of interest, unknown), multiples (Figure 5.2-middle - p. 80) $(\overline{s}^{(\mathbf{n})})_{\mathbf{n}\in\boldsymbol{\mathcal{N}}}$ (sum of undesired reflected data) and additive noise $(b^{(\mathbf{n})})_{\mathbf{n}\in\boldsymbol{\mathcal{N}}}$. One assumes that genuine multiples $(\overline{s}^{(\mathbf{n})})_{\mathbf{n}\in\boldsymbol{\mathcal{N}}}$ can be estimated as a local, weighted sum of template candidates:

$$\overline{s}^{\mathbf{n}} = \sum_{j=0}^{J-1} \sum_{p=p'}^{p'+P_j-1} \overline{h}_j^{(\mathbf{n})}(p) r_j^{(n_t-p,n_o)}. \tag{5.2}$$

Here $(r_j^{(\mathbf{n})})_{\mathbf{n}\in\boldsymbol{\mathcal{N}},0\leq j<J}$ denotes the available templates and $\overline{h}_j^{(\mathbf{n})}$ is an unknown impulse response (with P_j tap coefficients) corresponding to template r_j, at time n_t and sensor n_o and where $p' \in \{-\min_{0\leq j<J} P_j + 1, \ldots, 0\}$. It must be emphasized that the dependence w.r.t. the time index n_t of the impulse response implies that the filtering process is not time and space invariant, although it can be assumed slowly varying in practice. Eq. (5.2) can be expressed more concisely as

$$\overline{s} = \sum_{j=0}^{J-1} R_j \overline{h}_j \tag{5.3}$$

by appropriately defining vectors $\overline{s}$, $\overline{h}_j$ and matrix R_j. More precisely,

$$\overline{s} = \left[\overline{s}^{(0,0)}, \, \ldots \, , \overline{s}^{(N_t-1,0)}, \, \ldots \, , \overline{s}^{(0,N_o-1)}, \, \ldots \, , \overline{s}^{(N_t-1,N_o-1)}\right]^\top, \tag{5.4}$$

$$\overline{h}_j = \left[\overline{h}_j^{(0)\top}, \, \ldots \, , \overline{h}_j^{(N_o-1)\top}\right]^\top \tag{5.5}$$

where

$$\begin{aligned} \overline{h}_j^{(n_o)} = \Big[&\overline{h}_j^{(0,n_o)}(p'), \, \ldots \, , \overline{h}_j^{(0,n_o)}(p' + P_j - 1), \, \ldots \, , \\ &\overline{h}_j^{(N_t-1,n_o)}(p'), \, \ldots \, , \overline{h}_j^{(N_t-1,n_o)}(p' + P_j - 1)\Big]^\top. \end{aligned} \tag{5.6}$$

The matrix R_j is defined as

$$R_j = \begin{bmatrix} R_j^{(0)} & 0 & \dots & 0 \\ 0 & R_j^{(1)} & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & R_j^{(N_o-1)} \end{bmatrix} \tag{5.7}$$

where

$$R_j^{(n_o)} = \begin{bmatrix} R_j^{(0,n_o)} & 0 & \dots & 0 \\ 0 & R_j^{(1,n_o)} & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & R_j^{(N_t-1,n_o)} \end{bmatrix} \tag{5.8}$$

and $(R_j^{(n_t,n_o)})_{n_t\in\mathbb{N}_t}$ are given by

$$\begin{bmatrix} R_j^{(0,n_o)} \\ \vdots \\ R_j^{(N_t-1,n_o)} \end{bmatrix} = \begin{bmatrix} r_j^{(-p',n_o)} & \dots & r_j^{(0,n_o)} & 0 & \dots & 0 \\ r_j^{(-p'+1,n_o)} & \dots & & r_j^{(0,n_o)} & 0 \ \dots & 0 \\ \vdots & & & & & \\ r_j^{(N_t-1,n_o)} & r_j^{(N_t-2,n_o)} & & \dots & & r_j^{(N_t-P_j,n_o)} \\ 0 & r_j^{(N_t-1,n_o)} & & \dots & & r_j^{(N_t-P_j+1,n_o)} \\ \vdots & & & & & \\ 0 & \dots & 0 & r_j^{(N_t-1,n_o)} & \dots & r_j^{(N_t-P_j-p',n_o)} \end{bmatrix}. \tag{5.9}$$

One can note that each matrix R_j is a block diagonal matrix of size $Q \times QP_j$ where $Q = N_tN_o$ and that, for every $n_o \in \{0,\dots,N_o-1\}$, $R_j^{(n_o)}$ is a block diagonal matrix for which the concatenation of its block diagonal elements is a Tœplitz matrix of size $N_t \times P_j$. Model (5.1) can be written more concisely as

$$z = \mathbf{R}\overline{\mathbf{h}} + \overline{y} + b \tag{5.10}$$

where $\mathbf{R} = [R_0 \dots R_{J-1}] \in \mathbb{R}^{Q\times QP}$, $P = \sum_{j=0}^{J-1} P_j$ and $\overline{\mathbf{h}} = [\overline{h}_0^\top \dots \overline{h}_{J-1}^\top]^\top \in \mathbb{R}^{QP}$.

5.3 Proposed approach

We propose a variational framework where we aim at simultaneously estimating the filter tap coefficients of the nonstationary process $\mathbf{h}$ and the primary y. The multiple removal problem is thus formulated as the following constrained convex minimization problem

$$\underset{y\in\mathbb{R}^Q,\, \mathbf{h}\in C_1\cap C_2}{\text{minimize}} \ \|z - y - \mathbf{R}\mathbf{h}\|^2 + \iota_D(Fy) + \iota_{C_3}(L\mathbf{h})\,, \tag{5.11}$$

where $L \in \mathbb{R}^{\widetilde{K}\times QP}$ is a linear operator and $F \in \mathbb{R}^{KN_o\times Q}$ models a frame operator (e.g. directional wavelets (Jacques et al., 2011)). Generally, seismic data exhibits directional features corresponding to subsurface interfaces. Linear, hyperbolic and parabolic structures, at least piecewise, are observed. Due to the anisotropic nature of seismic data, we choose F as a hybrid dual-tree M-band wavelet (with different wavelets along space and time (Chaux et al., 2006, 2007b)). Their decorrelation properties (Chaux et al., 2007c) have proven efficient in textured image denoising (Chaux et al., 2008; Pustelnik et al., 2012), for instance, compared to curvelets whose redundancy is more important. We now turn our attention to the choice of the convex sets $C = C_1 \cap C_2 \cap C_3$ and D.

5.3.1 A priori information on primary signal

Frame coefficients, denoted by x, are related to the primary signal y through the relation $x = Fy$. Frame coefficients exhibit specific subband structures that are exploited here. Therefore, the constraint can be split by defining a partition of $\{1,\dots,KN_o\}$ denoted by $\{\mathbb{K}_l \mid l \in \{1,\dots,\mathcal{L}\}\}$ where $\mathcal{L}$ and $\mathbb{K}_l$ are defined dependently on the transform. More precisely, we investigate two kinds of $\mathcal{L}$ and $\mathbb{K}_l$

1. If F is an orthogonal basis or a shift-invariant (SI) wavelet frame: $\mathcal{L}$ corresponds to the number of subbands and $\mathbb{K}_l$ is the l-th subband with $D_l = \{(x_k)_{k\in\mathbb{K}_l} \mid \sum_{k\in\mathbb{K}_l} \varphi_\ell(x_k) \leq \beta_l\}$

2. If F is a hybrid dual-tree wavelet: $\mathcal{L}/2$ corresponds to the number of subbands and for every $l \in \{1,2,\dots,\mathcal{L}/2\}$, $\mathbb{K}_l$ and $\mathbb{K}_{l+\mathcal{L}/2}$ are the l-th subband corresponding to the primal and dual coefficients, respectively. Then, one can choose, for every $l \in \{1,\dots,\mathcal{L}/2\}$

$$D_l = \left\{ (x_k)_{k\in\mathbb{K}_l\cup\mathbb{K}_{l+\mathcal{L}/2}} \mid \sum_{k\in\mathbb{K}_l} \varphi_\ell(x_k) \leq \beta_l \text{ and } \sum_{k\in\mathbb{K}_{l+\mathcal{L}/2}} \varphi_\ell(x_k) \leq \beta_l \right\}$$

where, $\beta_l \in]0,+\infty[$, and $\varphi_l : \mathbb{R} \to [0,+\infty[$ is a lower-semicontinuous convex function. For example, a natural choice for φ_l is the ℓ_1-norm (the coefficients follow a Gauss-Laplace distribution as demonstrated in Figure 4.4 - p. 74). Figure 5.3 - p. 85 shows the primal and the dual trees of the dual-tree wavelet transform of the primary wave only, with a Symlet of length 8 along time and Daubechies of length 4 along space over 4 decomposition levels.

5.3.2 A priori information on adaptive filter

The convex set C introduces a priori knowledge on filter tap coefficients. As mentioned earlier, the filters here are assumed to vary not only along the time index n_t but also along the sensor index n_o. However, the filter variations are usually slow. This can be

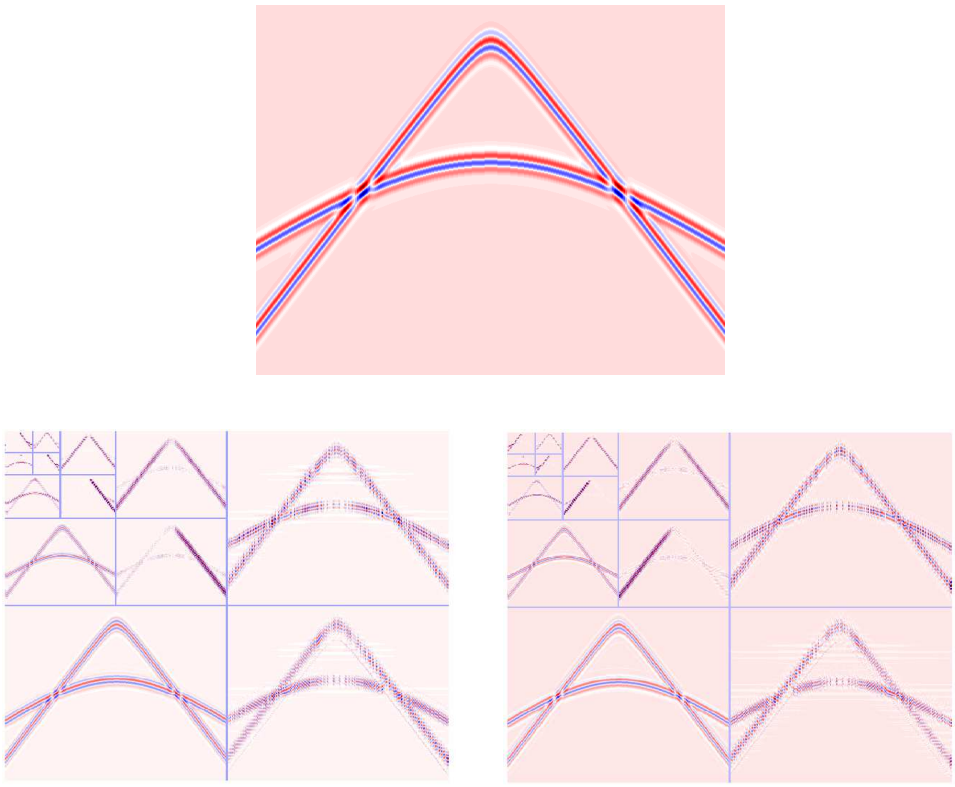


Figure 5.3: *Primary data (top); dual-tree M-band wavelet: primal coefficients (bottom-left), and dual coefficients (bottom-right).*

modeled by the following two constraints:

$$\big(\forall(n_t, n_o, j, p)\big) \qquad |h_j^{(n_t+1,n_o)}(p) - h_j^{(n_t,n_o)}(p)| \leq \varepsilon_{j,p}^{(n_o)} \tag{5.12}$$

$$|h_j^{(n_t,n_o+1)}(p) - h_j^{(n_t+1,n_o)}(p)| \leq \varepsilon_{j,p}^{(n_t)}. \tag{5.13}$$

These inequalities link:

1. corresponding coefficients of the impulse response which are estimated between two consecutive time samples (in other words, we impose consecutive filters in time to take close values). The bound $\varepsilon_{j,p}^{(n_o)} \in]0, +\infty[$ may depend on the shape of the expected filter. For example, its dependence on the coefficient indexes p and n_o may enable a larger (resp. smaller) difference for filter coefficients taking larger (resp. smaller) values;

2. corresponding coefficients of the impulse response which are estimated between two neighboring sensor samples (in other words, we impose close filters in space to take close values). The bound $\varepsilon_{j,p}^{(n_t)} \in]0, +\infty[$ may also depend on the shape of the expected filter.

The associated closed convex sets are defined as

$$\begin{aligned} C_1 =& \left\{ \mathbf{h} \in \mathbb{R}^{QP} \mid \forall (j, n_t, n_o, p),\ |h_j^{(n_t+1,n_o)}(p) - h_j^{(n_t,n_o)}(p)| \leq \varepsilon_{j,p}^{(n_o)} \right\} \\ C_2 =& \left\{ \mathbf{h} \in \mathbb{R}^{QP} \mid \forall (j, n_t, n_o, p),\ |h_j^{(n_t,n_o+1)}(p) - h_j^{(n_t,n_o)}(p)| \leq \varepsilon_{j,p}^{(n_t)} \right\} \end{aligned} \tag{5.14}$$

where $(\varepsilon_{j,p}^{(n_o)}, \varepsilon_{j,p}^{(n_t)}) \in [0, +\infty[^2$.

For computational issues, the convex set C_1 (resp. C_2) can be expressed as the intersection of two convex subsets $C_1 = C_{1,1} \cap C_{1,2}$ and $C_2 = C_{2,1} \cap C_{2,2}$:

$$\begin{aligned} C_{1,1} &= \left\{ \mathbf{h} \mid \forall p, \forall n_o, \forall n_t \in \left\{ 0, \ldots, \left\lfloor \frac{N_t}{2} \right\rfloor - 1 \right\} : \quad \left| h^{(2n_t+1,n_o)}(p) - h^{(2n_t,n_o)}(p) \right| \leq \varepsilon_{j,p}^{(n_o)} \right\} \\ C_{1,2} &= \left\{ \mathbf{h} \mid \forall p, \forall n_o, \forall n_t \in \left\{ 1, \ldots, \left\lfloor \frac{N_t - 1}{2} \right\rfloor \right\} : \quad \left| h^{(2n_t,n_o)}(p) - h^{(2n_t-1,n_o)}(p) \right| \leq \varepsilon_{j,p}^{(n_o)} \right\} \end{aligned} \tag{5.15}$$

$$\begin{aligned} C_{2,1} &= \left\{ \mathbf{h} \mid \forall p, \forall n_t, \forall n_o \in \left\{ 0, \ldots, \left\lfloor \frac{N_o}{2} \right\rfloor - 1 \right\} : \quad \left| h^{(n_t,2n_o+1)}(p) - h^{(n_t,2n_o)}(p) \right| \leq \varepsilon_{j,p}^{(n_t)} \right\} \\ C_{2,2} &= \left\{ \mathbf{h} \mid \forall p, \forall n_t, \forall n_o \in \left\{ 1, \ldots, \left\lfloor \frac{N_o - 1}{2} \right\rfloor \right\} : \quad \left| h^{(n_t,2n_o)}(p) - h^{(n_t,2n_o-1)}(p) \right| \leq \varepsilon_{j,p}^{(n_t)} \right\}. \end{aligned} \tag{5.16}$$

Moreover, some additional a priori information can be added directly on the vector of filter coefficients $\mathbf{h}$ which we will discuss in detail in the next section. This amounts to defining a new convex set C_3 as a lower level set of some proper lower-semicontinuous convex function ρ, by setting:

$$C_3 = \left\{ d \in \mathbb{R}^{\widetilde{K}} \mid \rho(d) \leq \lambda \right\}$$

where $\lambda \in]0, +\infty[$, $d = L\mathbf{h}$ and $\rho \colon \mathbb{R}^{\widetilde{K}} \to [0, +\infty[$ is a lower-semicontinuous convex function. Now let us focus on the choice of L:

1. The simplest choice is $L = \mathbf{R}$, it relates to the sparsity of multiple reflections.
2. A more complex choice is $L = \widetilde{F}\mathbf{R}$ where $\widetilde{F} : \mathbb{R}^Q \to \mathbb{R}^{\widetilde{K}}$ is an analysis frame operator.

5.4 Proximal algorithm

We choose to employ a primal-dual algorithm (Komodakis and Pesquet, 2014) an example of which is the M+LFBF algorithm 5 (Combettes and Pesquet, 2012).
For the convergence of M+LFBF algorithm, at each iteration i the stepsize $\gamma^{[i]}$ must be

Algorithm 5 M+LFBF

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

Set $y^{[0]} \in \mathbb{R}^{Q}$, $\mathbf{h}^{[0]} \in \mathbb{R}^{QP}$, $v^{[0]} \in \mathbb{R}^{KN_o}$, $u_1^{[0]} \in \mathbb{R}^{QP}$, $u_2^{[0]} \in \mathbb{R}^{QP}$, $u_3^{[0]} \in \mathbb{R}^{\widetilde{K}}$

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

Gradient computation

$s_1^{[i]} = y^{[i]} - \gamma^{[i]} \left(\mathbf{R}\mathbf{h}^{[i]} + y^{[i]} - z + F^* v^{[i]}\right)$

$t_1^{[i]} = \mathbf{h}^{[i]} - \gamma^{[i]} \left(\mathbf{R}^\top (\mathbf{R}\mathbf{h}^{[i]} + y^{[i]} - z) + L^* u_3^{[i]} + u_1^{[i]} + u_2^{[i]}\right)$

Projection computation

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

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

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

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

end for

$t_{2,1}^{[i]} = u_1^{[i]} + \gamma^{[i]} \mathbf{h}^{[i]}$

$t_{2,2}^{[i]} = u_2^{[i]} + \gamma^{[i]} \mathbf{h}^{[i]}$

$t_{2,3}^{[i]} = u_3^{[i]} + \gamma^{[i]} L \mathbf{h}^{[i]}$

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

$w_{2,m}^{[i]} = t_{2,m}^{[i]} - \gamma^{[i]} \Pi_{C_m}(\gamma^{[i]})^{-1} t_{2,m}^{[i]})$

end for

Averaging

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

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

$\left(q_{1,1}^{[i]}(k)\right)_{k\in\mathbb{K}_l} = \left(w_1^{[i]}(k) + \gamma^{[i]} x_2^{[i]}(k)\right)_{k\in\mathbb{K}_l}$

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

end for

$q_{2,1}^{[i]} = w_{2,1}^{[i]} + \gamma^{[i]} t_1^{[i]}$

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

$q_{2,3}^{[i]} = w_{2,3}^{[i]} + \gamma^{[i]} L t_1^{[i]}$

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

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

end for

Update

$y^{[i+1]} = y^{[i]} - \gamma^{[i]} \left(\mathbf{R} t_1^{[i]} + s_1^{[i]} - z + F^* w_1^{[i]}\right)$

$\mathbf{h}^{[i+1]} = \mathbf{h}^{[i]} - \gamma^{[i]} \left(\mathbf{R}^\top (\mathbf{R} t_1^{[i]} + s_1^{[i]} - z) + L^* w_{2,3}^{[i]} + w_{2,1}^{[i]} + w_{2,2}^{[i]}\right)$

end for

chosen so as to satisfy the following rule: Let $\beta = 2(1+|||\mathbf{R}|||^2)+\sqrt{4+\max(|||F|||^2, |||L|||^2)}$ and $\epsilon \in]0, \frac{1}{1+\beta}[$, then $\gamma^{[i]} \in [\epsilon, \frac{1-\epsilon}{\beta}]$. In this case, we have

$$|||\mathbf{R}||| = \left(\sum_{j=0}^{J-1} |||R_j|||^2 \right)^{1/2}$$

where $|||R_j||| = \max_{\mathbf{n}\in\boldsymbol{\mathcal{N}}} \left|\left| R_j^{(\mathbf{n})} \right|\right|$ for every $j \in \{0, \ldots, J-1\}$.

5.5 Results

The object of interest in this work is the primary waveforms. We thus privileged the display of results on primary reflections. However, numerical evaluations are performed on both primary and multiples.

5.5.1 Comparative evaluation: synthetic data

Our method is evaluated on the 2D synthetic seismic dataset presented in Figure 5.4 - p. 89. It is constructed as in Chapter 4 with $N_o = 512$ seismic traces of length $N_t = 512$ with filters of lengths $P_0 = 4$ and $P_1 = 5$. The constraint sets $C_{1,1}$, $C_{1,2}$, $C_{2,1}$ and $C_{2,2}$ are defined by (5.15) and (5.16) where, for every (j, p, n_t, n_o), $\varepsilon_{j,p}^{(n_o)} = 0.05$ and $\varepsilon_{j,p}^{(n_t)} = 0.0001$. The observed image z (with $\sigma = 0.04$) is presented at the top of Figure 5.4 - p. 89, in which top and bottom hyperbolas correspond to primaries, while the medium one corresponds to a multiple event. Regarding convex sets D and C_3, we choose $\varphi_l = |\cdot|$, $L = \mathbf{R}$ and we investigate three possible choices for ρ:

1. ℓ_1-norm

$$\left(\forall d \in \mathbb{R}^{N_t \times N_o}\right) \quad \rho(d) = \ell_1(d) = \sum_{n_o=0}^{N_o-1} \sum_{n_t=0}^{N_t-1} |d^{(n_t,n_o)}|,$$

2. ℓ_2-norm

$$\left(\forall d \in \mathbb{R}^{N_t \times N_o}\right) \quad \rho(d) = \ell_2(d) = \left(\sum_{n_o=0}^{N_o-1} \sum_{n_t=0}^{N_t-1} (d^{(n_t,n_o)})^2 \right)^{1/2},$$

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

$$\left(\forall d \in \mathbb{R}^{N_t \times N_o}\right) \quad \rho(d) = \ell_{1,2}(d) = \sum_{n_o=0}^{N_o-1} \left(\sum_{n_t=0}^{N_t-1} (d^{(n_t,n_o)})^2 \right)^{1/2}.$$

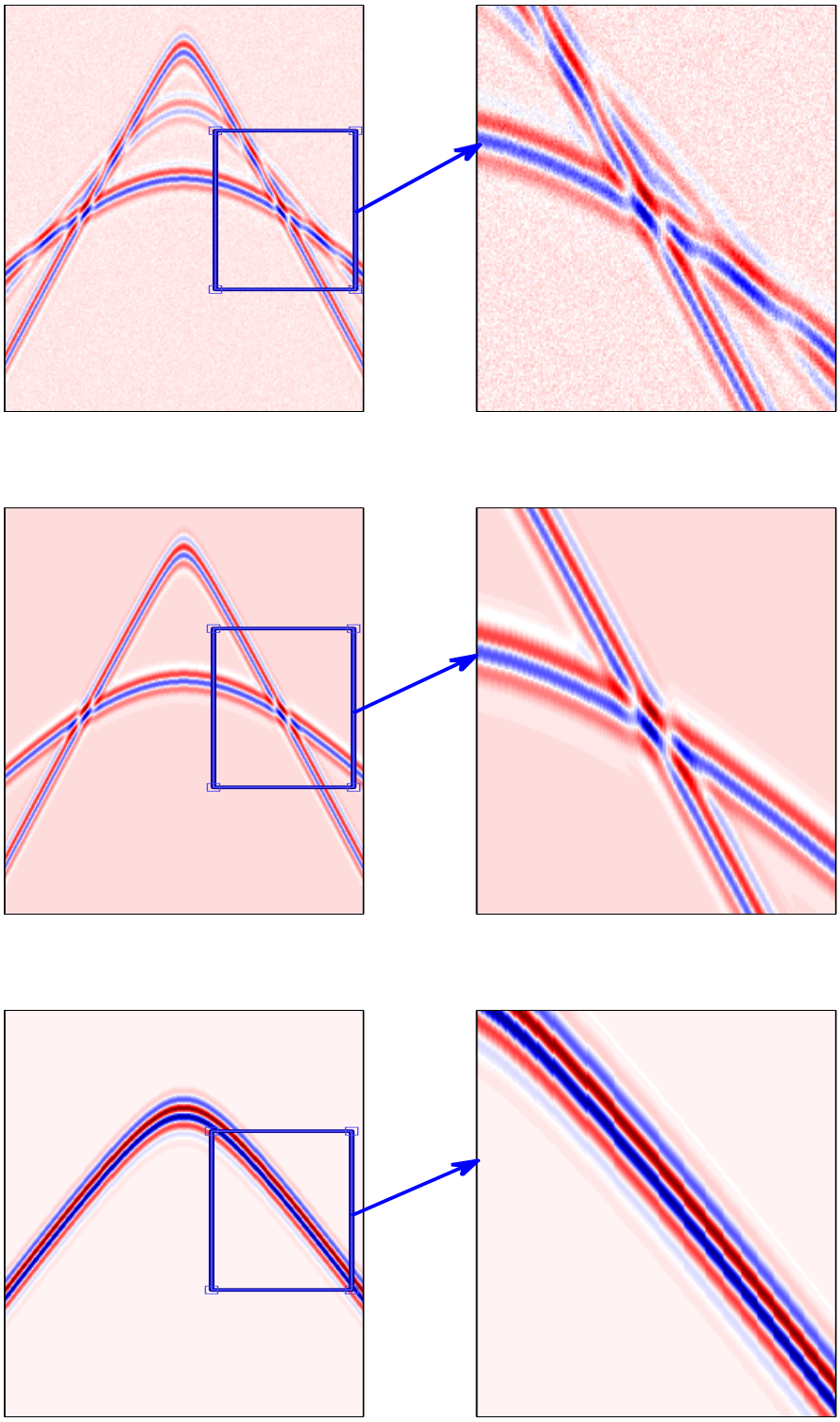


Figure 5.4: *From top to bottom: observed image z with $\sigma = 0.04$, primary $\overline{y}$ and multiple $\overline{s}$.*

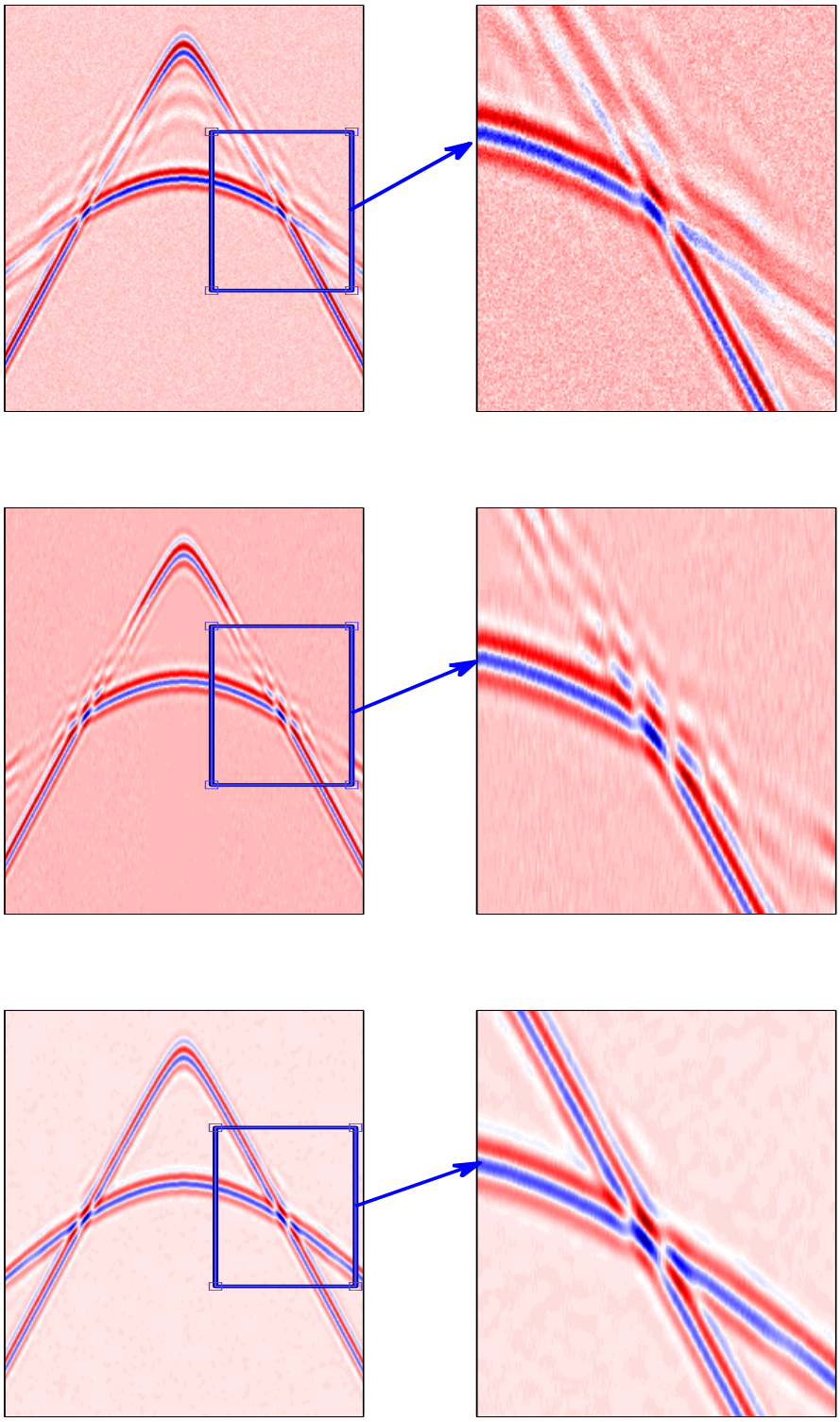


Figure 5.5: *From top to bottom: estimated primary y with (Ventosa et al., 2012), with 1D version in Chapter 4 and with the proposed 2D version.*

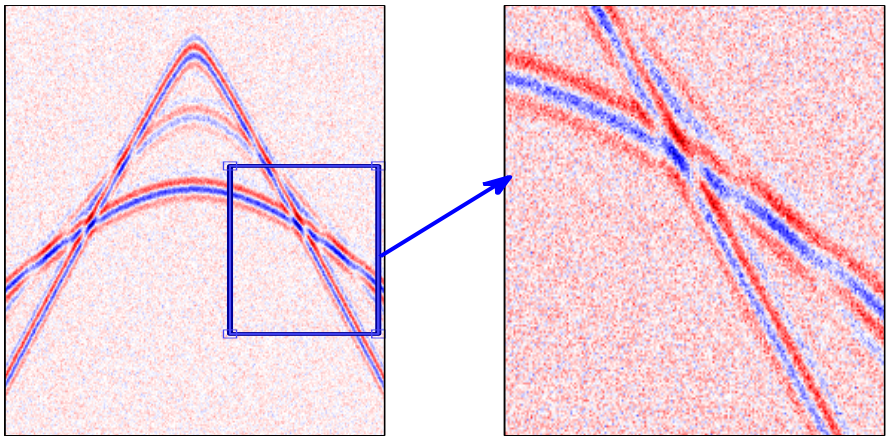


Figure 5.6: *Observed image z with $\sigma = 0.16$.*

We compute projections onto the convex sets C and D as in Chapter 4. The transforms are chosen as follows. In the time (vertical or sensor) and in the space directions, we choose a Symlet of length 8 and a Daubechies wavelet of length 4, respectively. The overall transform F may either be: 1D shift-invariant (SI, length-8 Symlet along time only), 2D critical and orthogonal, 2D shift-invariant (fully redundant) or 2D dual-tree (twice redundant). In the latter case, the given wavelets yield the primal tree, the dual being obtained by the Hilbert transform of the aforementioned wavelets. In Figure 5.5 - p. 90, the top row shows estimated primaries obtained by (Ventosa et al., 2012).
The SI frame (Figure 5.5 - p. 90, second row), although more redundant, exhibits a higher level of remnant noise and a blur crossing, as opposed to the bottom row, where primaries are clearly recovered. More interesting is the stronger noise condition in Figure 5.6 - p. 91 ($\sigma = 0.16$). The obtained results from this high-noise observed data that correspond to the estimated primaries, are shown in Figure 5.7 - p. 92. One can observe that the estimated primary by the algorithm in Chapter 4 may be not better than the one by (Ventosa et al., 2012) (less noise, but lost information on the primary). While the obtained primary by the 2D proposed method is also good, which demonstrates the robustness to noise of this method.
The comparison is pursued in a more complete and objective manner with ℓ_1 and ℓ_2 residual errors for primaries $\overline{y} - \hat{y}$ in Table 5.1 - p. 93 and for multiples $\overline{s} - \hat{s}$ in Table 5.2 - p. 94, the smaller being the better. The blue (resp. red) indicates the minimum value of ℓ_1 (resp. ℓ_2) for each ρ. We test the proposed method against different levels of noise: $\sigma \in \{0.04, 0.08, 0.16\}$. We see that 2D versions generally further reduce modeling errors, as expected. Moreover, a loose ranking is has follows: (Ventosa et al., 2012) < SI-1D < 2D orthogonal basis < SI-2D < dual-tree. We cannot clearly conclude about a positive specific effect of ℓ_1 or ℓ_2-norms, however the smallest

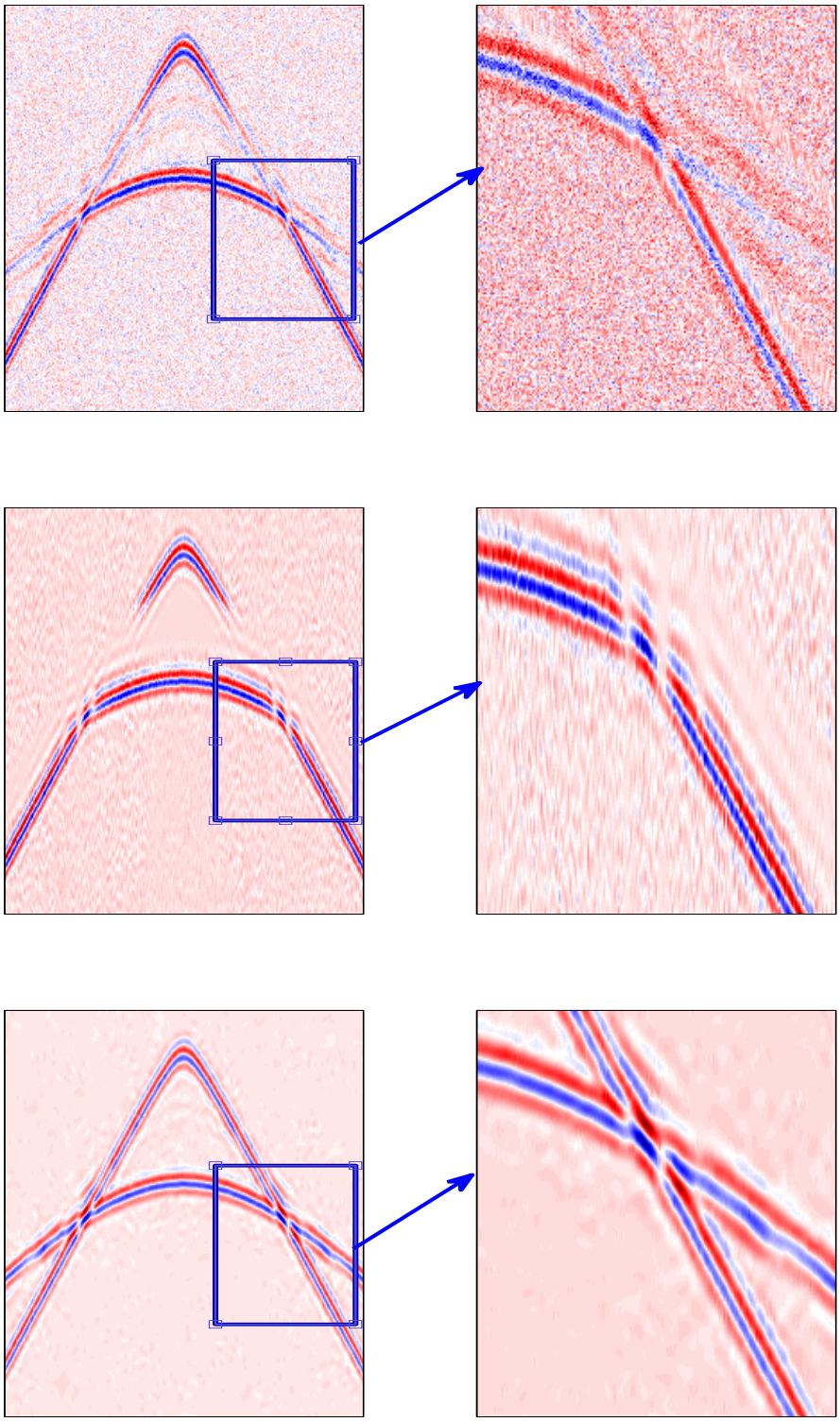


Figure 5.7: *From top to bottom: estimated primary y with (Ventosa et al., 2012), with 1D version in Chapter 4 and with the proposed 2D version.*

		σ	0.04	0.08	0.16
$\overline{y} - z$		$\ell_1(\times 10^2)$	3.88	6.89	13.09
		$\ell_2(\times 10^2)$	5.42	8.80	16.43
(Ventosa et al., 2012)		$\ell_1(\times 10^2)$	4.32	7.15	13.01
		$\ell_2(\times 10^2)$	6.70	9.50	16.50
1D version in Chapter 4		$\ell_1(\times 10^2)$	2.06	2.87	3.75
		$\ell_2(\times 10^2)$	5.36	6.28	7.14
$\rho = \ell_2$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	1.30	1.94	2.73
		$\ell_2(\times 10^2)$	2.67	3.73	4.91
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	1.06	1.38	1.91
		$\ell_2(\times 10^2)$	3.01	3.25	3.89
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.77	1.29	1.65
		$\ell_2(\times 10^2)$	1.83	3.31	3.66
$\rho = \ell_1$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	1.31	1.92	2.77
		$\ell_2(\times 10^2)$	2.57	3.59	5.05
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.88	1.26	1.86
		$\ell_2(\times 10^2)$	2.16	2.69	3.63
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.79	1.09	1.58
		$\ell_2(\times 10^2)$	1.86	2.34	3.33
$\rho = \ell_{1,2}$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	1.19	1.77	2.63
		$\ell_2(\times 10^2)$	2.22	3.11	4.62
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.96	1.35	1.99
		$\ell_2(\times 10^2)$	2.68	3.22	4.33
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.71	1.02	1.50
		$\ell_2(\times 10^2)$	1.44	1.95	2.86

Table 5.1: *Comparison of the estimated primaries with the 2D proposed version$^{(*)}$ in using three different 2D wavelet transforms, over three noise levels, and three a priori functions $\rho \in \{\ell_2, \ell_1, \ell_{1,2}\}$, with (Ventosa et al., 2012) and with the 1D version in Chapter 4. (These results correspond to the data in Figure 5.4 - p. 89).*

absolute or quadratic residuals are always obtained with the $\ell_{1,2}$ norm. This confirms that mild redundancy is beneficial when coupled with appropriate constraints.
To better emulate more complex geophysical configurations, we consider another synthetic data (see Figure 5.8 - p. 96). This data resembles a lot the data described above, however, there are other remnants where the structure is more complex with several overlapping hyperbolas. The results are shown in Figure 5.9 - p. 97 (estimated primaries). Figure 5.10 - p. 99 and Figure 5.11 - p. 100 illustrate for a stronger noise case ($\sigma = 0.16$). Once again, despite the more complicated data-structure with several overlapping waves, our proposed 2D method gives good results and remains also the top performer compared to two other methods. The comparison of the experimental results are presented in Table 5.3 - p. 98 (for estimated primaries) and Table 5.4 - p. 101

		σ	0.04	0.08	0.16
$\overline{s}-z$		$\ell_1(\times 10^2)$	6.18	8.96	14.73
		$\ell_2(\times 10^2)$	11.28	13.24	19.18
(Ventosa et al., 2012)		$\ell_1(\times 10^2)$	1.85	1.97	2.28
		$\ell_2(\times 10^2)$	5.51	5.68	6.31
1D version in Chapter 4		$\ell_1(\times 10^2)$	1.31	1.71	2.25
		$\ell_2(\times 10^2)$	5.38	6.42	7.88
$\rho=\ell_2$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	0.52	0.67	0.97
		$\ell_2(\times 10^2)$	2.14	2.79	3.80
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.41	0.49	0.61
		$\ell_2(\times 10^2)$	1.96	2.24	2.63
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.39	0.49	0.74
		$\ell_2(\times 10^2)$	1.72	2.09	3.07
$\rho=\ell_1$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	0.50	0.75	0.92
		$\ell_2(\times 10^2)$	2.28	3.08	3.57
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.72	0.78	0.84
		$\ell_2(\times 10^2)$	3.02	3.11	3.28
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.37	0.83	0.90
		$\ell_2(\times 10^2)$	1.69	3.33	3.56
$\rho=\ell_{1,2}$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	0.35	0.42	0.56
		$\ell_2(\times 10^2)$	1.63	1.90	2.42
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.48	0.58	0.81
		$\ell_2(\times 10^2)$	2.53	2.88	3.73
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.27	0.35	0.55
		$\ell_2(\times 10^2)$	1.20	1.55	2.32

Table 5.2: *Comparison of the estimated multiples with the 2D proposed version$^{(*)}$ in using three different 2D wavelet transforms, over three noise levels, and three a priori functions $\rho \in \{\ell_2, \ell_1, \ell_{1,2}\}$, with (Ventosa et al., 2012) and with the 1D version in Chapter 4. (These results correspond to the data in Figure 5.4 - p. 89).*

(for estimated multiples). The best results were obtained by using 2D proposed method, however, due to the nonstationarity of the data, and the presence of structured noise, the statistical analysis is not straightforward in comparing the performance of transforms. The previous experiments were performed with synthetic data. Its construction is not related to actual subsurface modelling or wave propagation. However, it exhibits event shapes that help in understanding the proposed method and the algorithm. Additionally, it permits a statistical analysis of the results that cannot be done with real data in the absence of ground truth. In the next subsection, we show the performance of our proposed method on real data.

5.5.2 Comparative evaluation: real data

We finally test our 2D proposed method for the real data provided by Jean Charléty (IFP Energies nouvelles). Figure 5.12-top - p. 102 is a recorded seismic data with a partially appearing primary (arrow) and at bottom a cropped square zone and Figure 5.13 - p. 103 displays two multiple wavefield templates. For this data, we choose $L = \widetilde{F}\mathbf{R}$, where $\widetilde{F}$ is a 2D orthogonal basis with a Symlet of length 8 over 3 decomposition levels, ρ is a mixed $\ell_{1,2}$-norm. F is a 2D dual-tree with a Symlet of length 8 along time and Daubechies of length 4 along space over 3 decomposition levels, and $\varphi = |\cdot|$. Figure 5.14 - p. 104 and Figure 5.15 - p. 105 show the estimated primaries obtained by (Ventosa et al., 2012), by 1D version in Chapter 4 and by 2D proposed method. Three methods are able to successfully retrieve a weak primary below the multiple level. On Figure 5.14-top - p. 104, arrows indicate remnants of multiples, parallel to the primary, that have been incorrectly subtracted. Figure 5.14-bottom - p. 104 contains more noise and more artifacts. The primary is better delineated with the directional dual-tree wavelet frame, with a reduced level of remaining noise and multiple interferences.

5.6 Conclusion

This chapter proposed an adaptive filtering method allowing the suppression of multiple in 2D seismic data. The 2D anisotropic structure is taken into account and tackled using geometric multiscale representations. The adaptive filter is performed under appropriate regularity and amplitude concentration constraints. We tested three different measures for ρ, such as ℓ_1, ℓ_2 and $\ell_{1,2}$ norms. The $\ell_{1,2}$-norm seems better suited to design adaptive filters. Simulations are carried out on both synthetic and real seismic data. In all cases, the 2D version with $\rho = \ell_{1,2}$ with 2D dual-tree transform demonstrated very good performance. The comparison of three methods shows that the proposed 2D method improves the robustness to noise and better preserve structural information. Furthermore, instead of using continuous or shift-invariant redundant wavelet transforms that yield huge redundancy, and thus a very large number of coefficients, we employ a 2D dual-tree transform which produces an only twice oversampled transform. The mild redundancy could be beneficial regarding the huge volumes of seismic data, and the results are actually better. From the obtained results and with the same argument, we hope that the

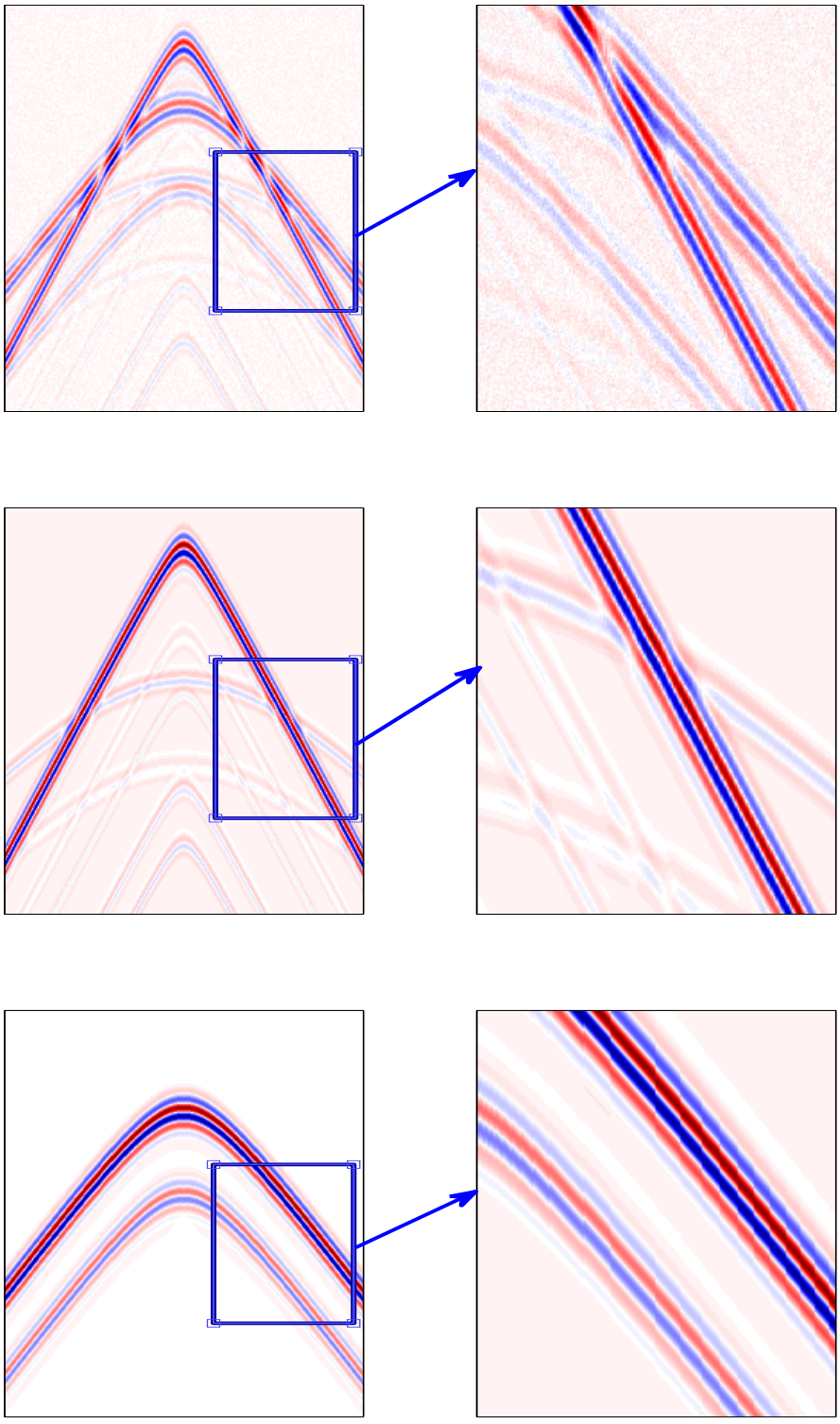


Figure 5.8: *From top to bottom: observed image z with $\sigma = 0.04$, primary $\overline{y}$ and multiple $\overline{s}$.*

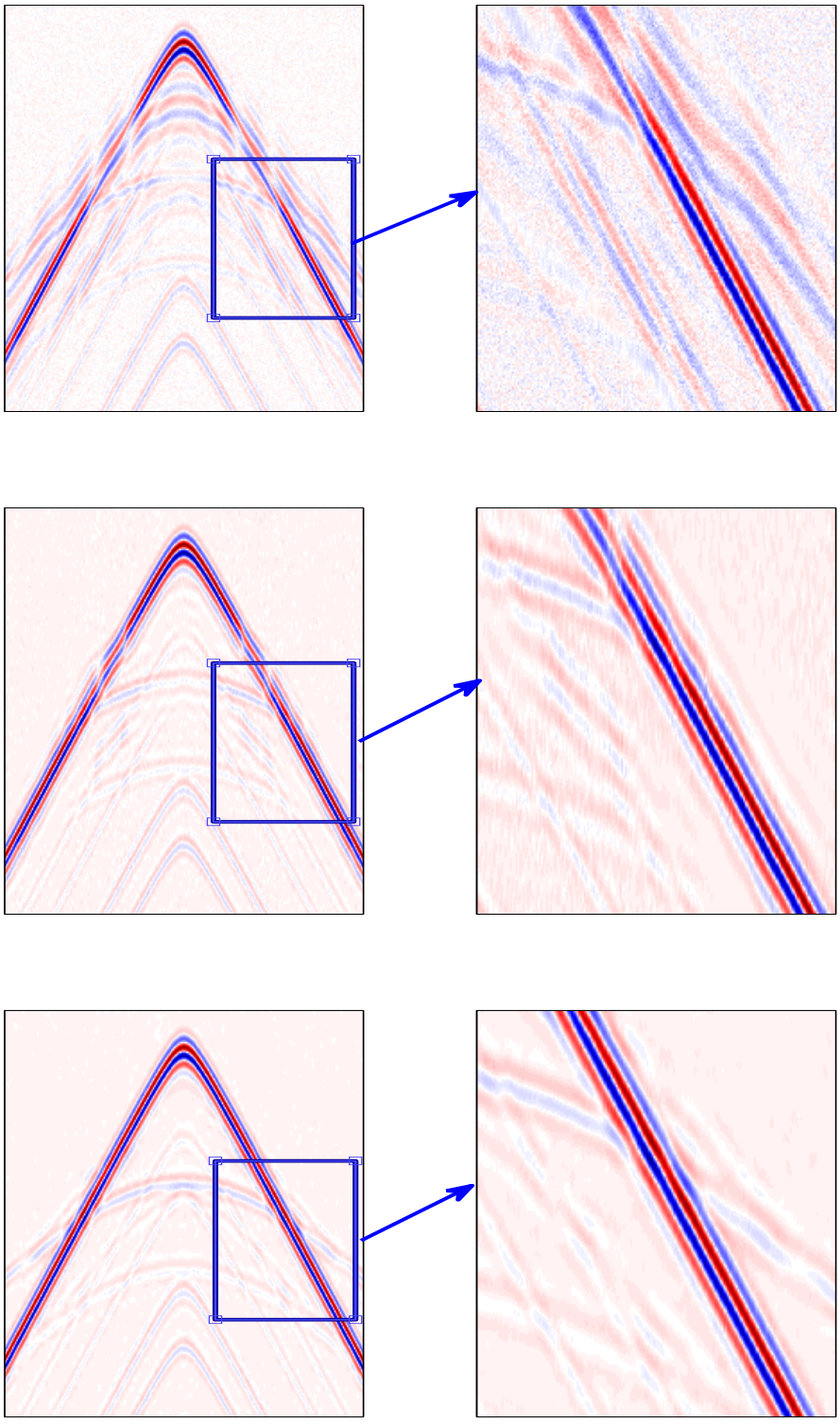


Figure 5.9: *From top to bottom: estimated primary y with (Ventosa et al., 2012), with 1D version in Chapter 4 and with the proposed 2D version.*

		σ	0.04	0.08	0.16
$\overline{y} - z$		$\ell_1(\times 10^2)$	3.88	6.89	13.09
		$\ell_2(\times 10^2)$	5.42	8.80	16.43
(Ventosa et al., 2012)		$\ell_1(\times 10^2)$	5.38	7.87	13.36
		$\ell_2(\times 10^2)$	8.97	11.14	17.30
1D version in Chapter 4		$\ell_1(\times 10^2)$	2.05	2.91	4.07
		$\ell_2(\times 10^2)$	4.75	6.26	8.06
$\rho = \ell_2$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	1.53	2.27	3.34
		$\ell_2(\times 10^2)$	2.56	3.59	5.33
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	1.19	1.69	2.42
		$\ell_2(\times 10^2)$	2.34	2.93	4.01
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	1.07	1.41	1.96
		$\ell_2(\times 10^2)$	2.20	2.50	3.20
$\rho = \ell_1$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	1.66	2.33	3.37
		$\ell_2(\times 10^2)$	2.96	3.78	5.40
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	1.23	1.70	2.39
		$\ell_2(\times 10^2)$	2.51	2.93	3.86
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	1.14	1.47	2.00
		$\ell_2(\times 10^2)$	2.42	2.66	3.30
$\rho = \ell_{1,2}$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	1.51	2.25	3.32
		$\ell_2(\times 10^2)$	2.48	3.54	5.27
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	1.10	1.58	2.32
		$\ell_2(\times 10^2)$	1.76	2.36	3.49
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.95	1.31	1.87
		$\ell_2(\times 10^2)$	1.60	2.07	2.83

Table 5.3: *Comparison of the estimated primaries with the 2D proposed version$^{(*)}$ in using three different 2D wavelet transforms, over three noise levels, and three a priori functions $\rho \in \{\ell_2, \ell_1, \ell_{1,2}\}$, with (Ventosa et al., 2012) and with the 1D version in Chapter 4. (These results correspond to the data in Figure 5.8 - p. 96).*

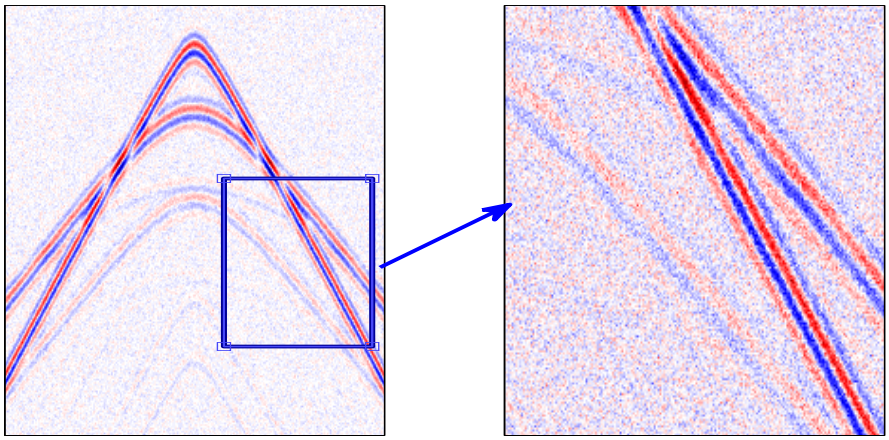


Figure 5.10: *Observed image z with $\sigma = 0.16$.*

proposed method could be extended to 3D representations and their associated filters (van Borselen and Schonewille, 2005; Baumstein and Hadidi, 2006; Beasley et al., 2012; Brittan et al., 2011; McHugo et al., 2008; Pica et al., 2005a; Zhang and Wang, 2011) and provide better results.

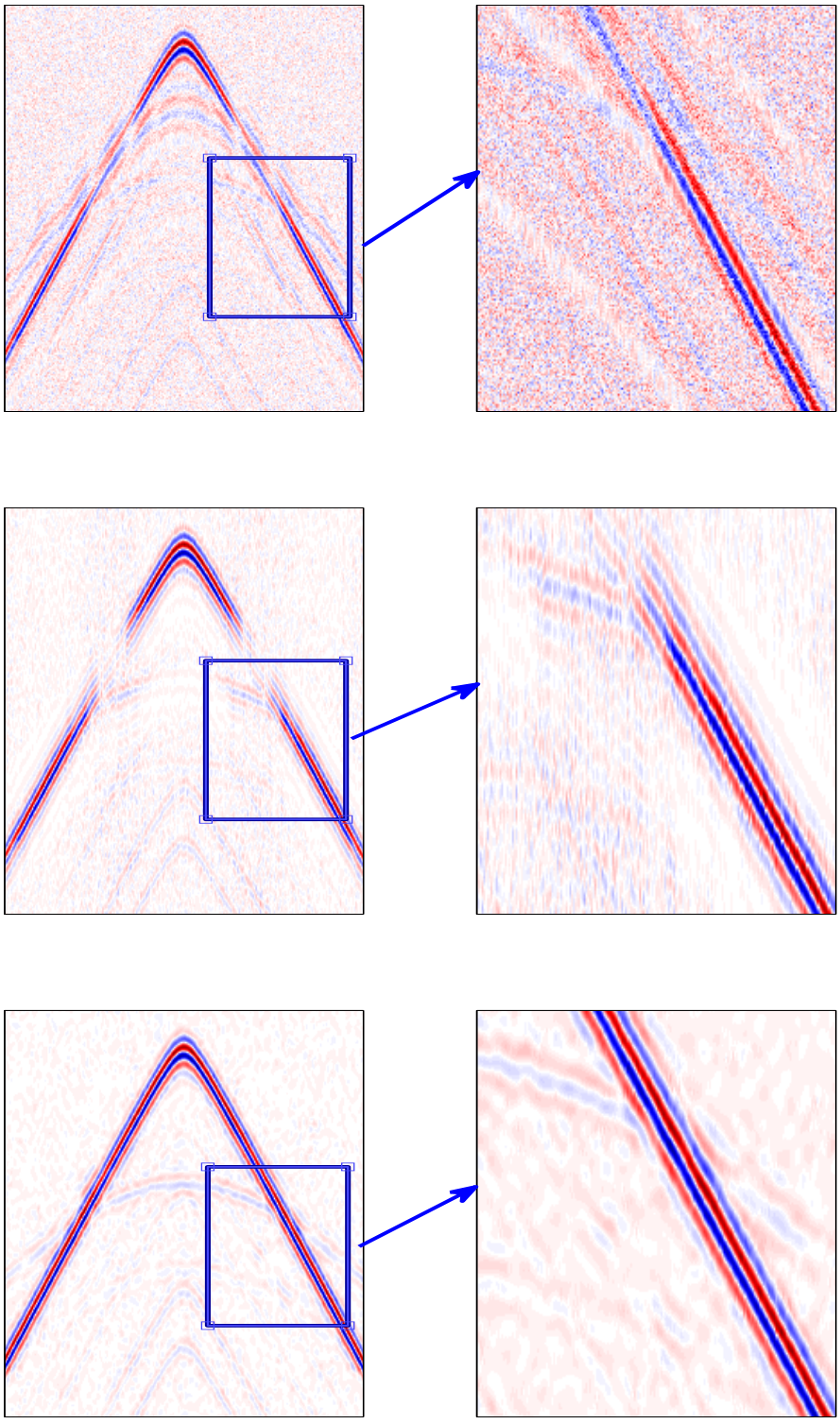


Figure 5.11: *From top to bottom: estimated primary y with (Ventosa et al., 2012), with 1D version in Chapter 4 and with the proposed 2D version.*

		σ	0.04	0.08	0.16
$\overline{s} - z$		$\ell_1(\times 10^2)$	7.10	9.81	15.59
		$\ell_2(\times 10^2)$	15.18	16.68	21.68
(Ventosa et al., 2012)		$\ell_1(\times 10^2)$	3.46	3.63	4.10
		$\ell_2(\times 10^2)$	8.17	8.34	9.02
1D version in Chapter 4		$\ell_1(\times 10^2)$	1.55	2.24	3.34
		$\ell_2(\times 10^2)$	4.92	6.77	9.35
$\rho = \ell_2$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	0.54	0.64	0.86
		$\ell_2(\times 10^2)$	1.88	2.20	2.89
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.58	0.66	0.87
		$\ell_2(\times 10^2)$	2.10	2.33	2.94
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.58	0.62	0.77
		$\ell_2(\times 10^2)$	2.04	2.13	2.54
$\rho = \ell_1$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	0.72	0.75	0.93
		$\ell_2(\times 10^2)$	2.46	2.58	3.08
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.64	0.68	0.82
		$\ell_2(\times 10^2)$	2.30	2.34	2.68
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.68	0.71	0.84
		$\ell_2(\times 10^2)$	2.30	2.35	2.70
$\rho = \ell_{1,2}$	orthogonal basis$^{(*)}$	$\ell_1(\times 10^2)$	0.52	0.60	0.78
		$\ell_2(\times 10^2)$	1.77	2.05	2.57
	SI frame$^{(*)}$	$\ell_1(\times 10^2)$	0.45	0.49	0.65
		$\ell_2(\times 10^2)$	1.36	1.43	1.89
	M-band dual-tree$^{(*)}$	$\ell_1(\times 10^2)$	0.44	0.50	0.64
		$\ell_2(\times 10^2)$	1.34	1.54	1.95

Table 5.4: *Comparison of the estimated multiples with the 2D proposed version$^{(*)}$ in using three different 2D wavelet transforms, over three noise levels, and three a priori functions $\rho \in \{\ell_2, \ell_1, \ell_{1,2}\}$, with (Ventosa et al., 2012) and with the 1D version in Chapter 4. (These results correspond to the data in Figure 5.8 - p. 96).*

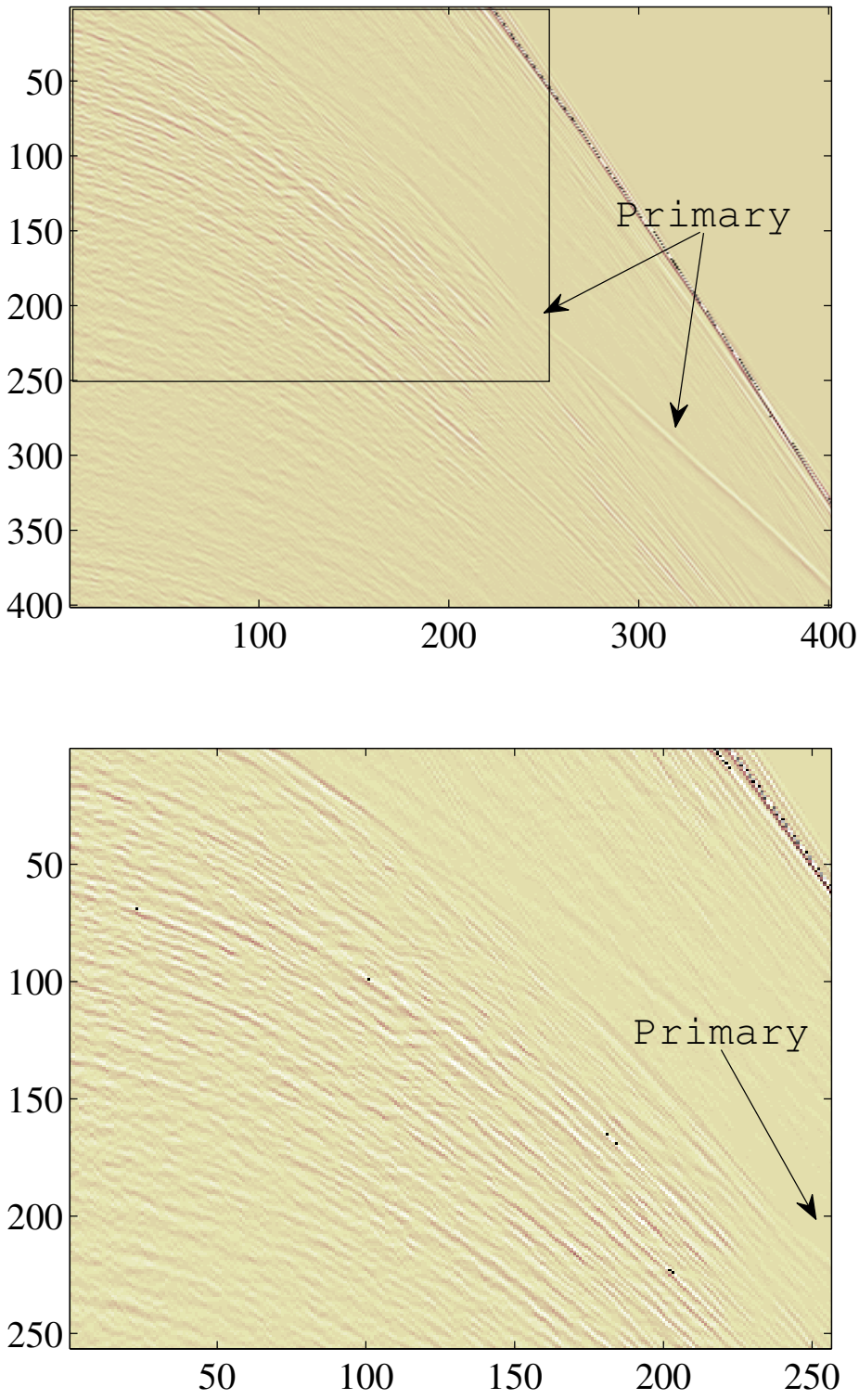


Figure 5.12: *From top to bottom: recorded seismic data with a partially appearing primary and a cropped of recorded seismic data.*

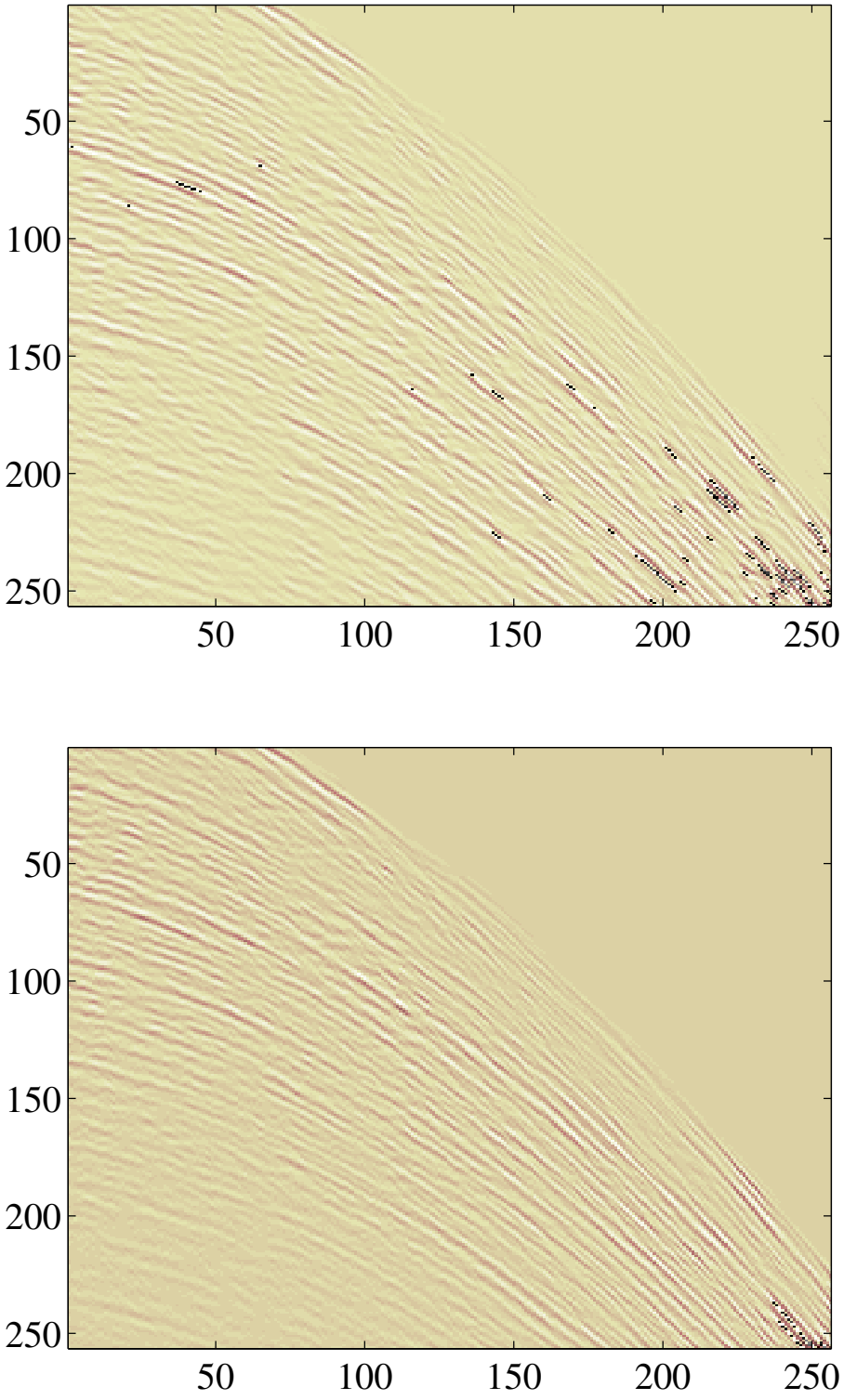


Figure 5.13: *Tow multiple wavefield templates.*

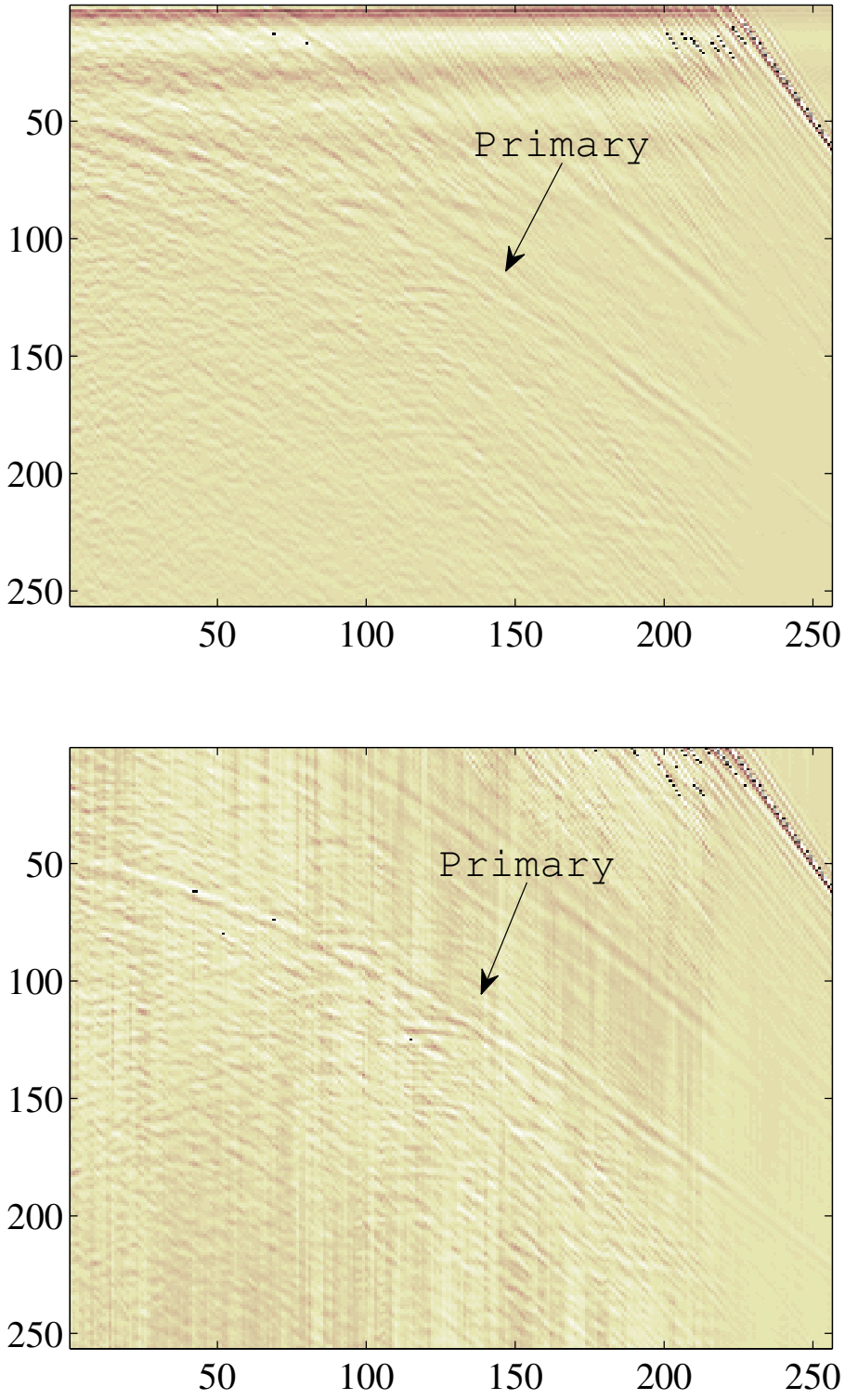


Figure 5.14: *From top to bottom: estimated primary obtained with (Ventosa et al., 2012) and with 1D version in Chapter 4.*

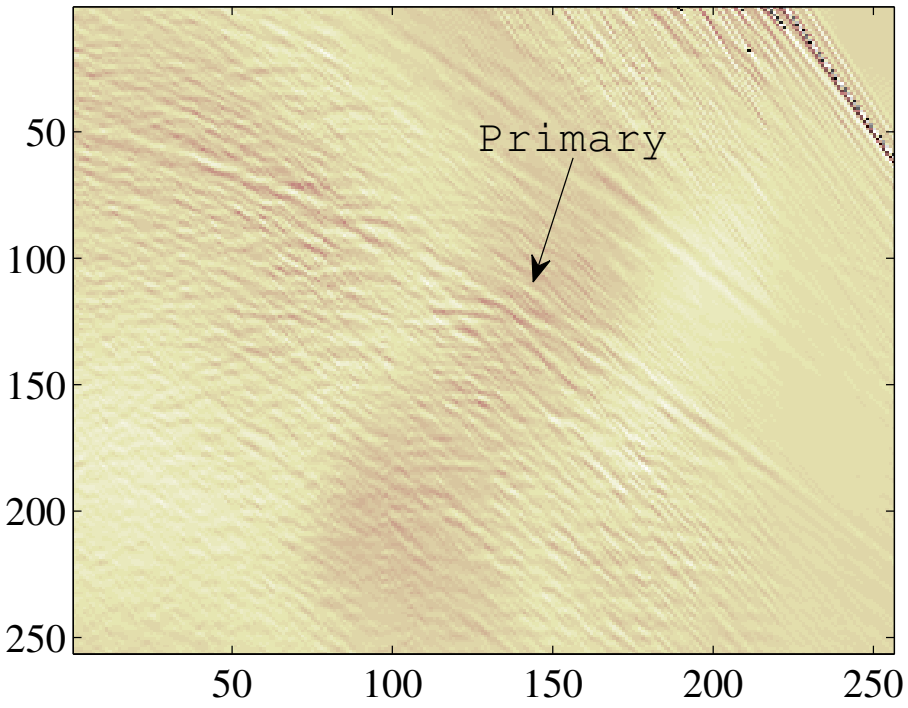


Figure 5.15: *Estimated primary obtained with the proposed 2D version.*

- Chapter 6 -

Sparse Blind Deconvolution with Smoothed ℓ_1/ℓ_2 Regularization

Blind deconvolution plays a central role in the field of signal and image processing in general and of geophysical processing in particular. Indeed several geophysical issues may be formulated as blind deconvolution problems. For instance, (Li and Lu, 2013; Liu and Dragoset, 2013; Kaplan and Innanen, 2008) use this formalism to address multiple removal. As in the previous chapters we have separated multiples from primaries, a next goal would be to deconvolve them from the seismic source. As the source (the seismic wavelet) is rarely known precisely, blind seismic deconvolution is considered here. It aims at improving the temporal resolution of seismic data in the shape of a sparser sequence. Seismic deconvolution has generated a lot of seminal works on sparsity-promoting measures, including ℓ_1 and even ℓ_1/ℓ_2 penalties (Claerbout, 1977; Gray, 1979). This chapter is restricted to the use of an ℓ_1/ℓ_2 penalty. The ℓ_1/ℓ_2 ratio regularization function has shown good performance for retrieving sparse signals in a number of recent works, in the context of blind deconvolution. Indeed, it benefits from a scale invariance property much desirable in the blind context. However, the ℓ_1/ℓ_2 function raises some difficulties when solving the nonconvex and nonsmooth minimization problems resulting from the use of such a penalty term in current restoration formulations. In this work, we propose a new penalty based on a smooth approximation to the ℓ_1/ℓ_2 function. In addition, we develop a proximal-based algorithm to solve variational problems involving this function and we derive theoretical convergence results. We demonstrate the effectiveness of our method through a comparison with a recent alternating optimization strategy dealing with the exact ℓ_1/ℓ_2 term, on both applications to seismic data (1D) and image blind deconvolution (2D).

6.1 Introduction

As alluded to in Section **??**, many experimental settings are modeled as inverse problems. They resort to estimating an unknown signal $\overline{x} \in \mathbb{R}^N$ from observations $y \in \mathbb{R}^N$, through

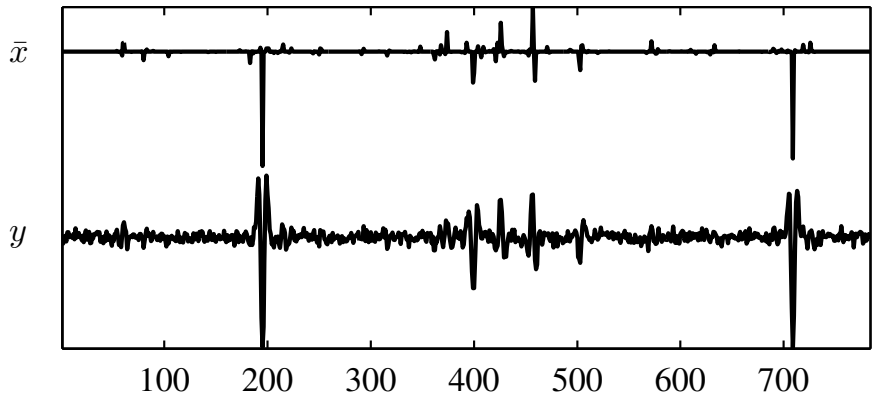


Figure 6.1: *Unknown seismic signal $\bar{x}$ (top), blurred/noisy observation y (bottom).*

the measurement process:

$$y = H(\overline{h}, \overline{x}) + b\,, \tag{6.1}$$

an illustration of which is provided in Figure **??**. Here, $\overline{h} \in \mathbb{R}^S$ represents an impulse response (e.g. a linear sensor response or a "blur" convolutive point spread function), $H : \mathbb{R}^S \times \mathbb{R}^N \to \mathbb{R}^N : (h, x) \mapsto h * x$, where $*$ denotes a discrete-time convolution operator (with appropriate boundary processing), and $b \in \mathbb{R}^N$ is a realization of a random variable modeling an additive noise. Standard approaches, such as Wiener filtering and its statistical SURE-based extensions (Pesquet et al., 2009), aim at minimizing criteria based on the squared Euclidean norm (ℓ_2^2). However, the use of the sole least squares data fidelity term is prone to noise sensitivity and the addition of an ℓ_2^2 regularization often leads to over-smoothed estimates. The deconvolution problem becomes blind, even more ill-posed, when the blur kernel $\overline{h}$ is unknown, and needs to be estimated as well as the target signal. Applications include communications (equalization or channel estimation) (Haykin, 1994), nondestructive testing (Nandi et al., 1997), geophysics (Kaaresen and Taxt, 1998; Takahata et al., 2012; Pham et al., 2014b), image processing (Kundur and Hatzinakos, 1996a,b; Kato et al., 1999; Ahmed et al., 2014), medical imaging and remote sensing (Campisi and Egiazarian, 2007). Blind deconvolution, being an underdetermined problem, often requires additional hypotheses. A usual approach seeks estimates $(\widehat{x}, \widehat{h}) \in \mathbb{R}^N \times \mathbb{R}^S$ of $(\overline{x}, \overline{h})$ as minimizers of the sum of a data fidelity term and additional regularization terms on the signal and on the blur kernel. Such regularization functions account for a priori assumptions one imposes on original sought objects, like sparsity, and ensure the stability of the solution. Blind deconvolution is subject to scaling ambiguity, and suggests scale-invariant contrast functions (Comon, 1996; Moreau and Pesquet, 1997).

A decade ago, a Taxicab-Euclidean norm ratio (ℓ_1/ℓ_2) arose as a sparseness measure (Zibulevsky and Pearlmutter, 2001; Hoyer, 2004; Hurley and Rickard, 2009; Barak et al., 2014), used in NMF (Nonlinear Matrix Factorization) (Mørup et al., 2008). Earlier mentions of a one-norm/two-norm ratio deconvolution appeared in geophysics (Gray, 1978). It has since been used to constrain sharp images through wavelet frame coefficients (Ji et al., 2012), or for sparse recovery (Demanet and Hand, 2015). Such a regularization

term is moreover suggested in (Benichoux et al., 2013) to avoid common pitfalls in blind sparse deconvolution.

Recently, (Krishnan et al., 2011) proposed an alternating minimization algorithm to deal with the ℓ_1/ℓ_2 regularization function. Its originality consists of transforming the ℓ_1/ℓ_2 nonconvex regularization term into a convex ℓ_1 regularization function. This is done in a reweighted fashion, by fixing the denominator ℓ_2 from the previous iterate. An iterative shrinkage-thresholding algorithm finally solves the remaining ℓ_1 regularized problem. Although the convergence of this approach has not been deeply investigated, it appears to work in practice. More recently, (Esser et al., 2013) proposed a scaled projected gradient algorithm for minimizing a smooth approximation to the ℓ_1/ℓ_2 function, however limited to the case when the sparse signal to retrieve takes nonnegative values. We generalize this idea to a parametrized Smoothed One-Over-Two (SOOT) penalty for signed, real data. We present a novel efficient method based on recent results in nonconvex optimization combining an alternating minimization strategy with a forward-backward iteration (Bolte et al., 2014; Chouzenoux et al., 2013). Moreover, we accelerate the convergence of our algorithm by using a Majorize-Minimize (MM) approach (Sotthivirat and Fessler, 2002; Chouzenoux et al., 2013, 2014). Section **??** introduces the minimization problem. Section **??** describes the proposed method and provides convergence results. The algorithm performance, compared with (Krishnan et al., 2011), is discussed in Section **??** for seismic data blind deconvolution (1D) and in Section **??** for blind image deconvolution (2D). Some conclusions are drawn in Section **??**.

6.2 Sparsity measure ℓ_1/ℓ_2 and its surrogates

6.2.1 Motivation on the sparsity measure ℓ_1/ℓ_2

In signal representation, there are many ways to *practically* define sparsity. For instance, *a signal is sparse if the number of the non-zero coefficients in its representation is small compared to its dimension* (Zonoobi et al., 2011), that is related to the ℓ_0-"quasi-norm". However, the optimization problem with the ℓ_0-penalty is nonconvex and generally impossible to solve as its solution usually leads to an intractable combinatorial search. A relaxation for the ℓ_0-"quasi-norm" is the ℓ_p-norm/quasi-norm ($0 < p \leq 1$). We instead consider the ratio of the ℓ_1-norm to the ℓ_2-norm that we denote $\varphi_0 = \ell_1/\ell_2$, and we then compare with different measures, via some simple examples.

Example 6.2.1 Assume that we have two signals x and y of length $2N$ (Figure **??** - p. **??**), where

$$x_n = \frac{1}{2N} \text{ for every } n \in \{-N+1, \cdots, 0, 1, \ldots, N\} \text{ and}$$

$$y_n = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \in \{-N+1, \cdots, -1, 1, \cdots, N\}. \end{cases}$$

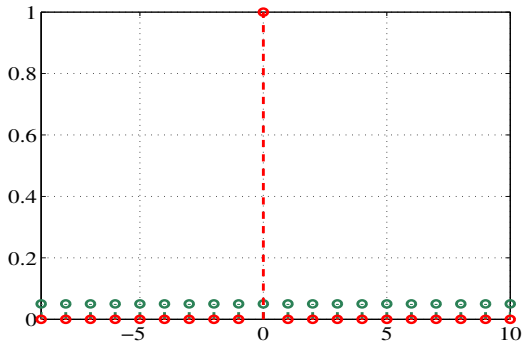


Figure 6.2: *Plots of signals x (green) and y (red), with $N = 10$.*

We have:

- $\ell_1(x) = \ell_1(y) = 1$
- $\ell_0(x) = 2N > \ell_0(y) = 1$
- $\varphi_0(x) = \sqrt{2N} > \varphi_0(y) = 1$.

It is clear that y is sparse, but x is not. From the values obtained by different measures (ℓ_1, ℓ_0, φ_0) for x and y, we observe that, nothing conclusive can be said about the sparsity of x and y with ℓ_1-norm. However, φ_0 may be a better sparsity measure. A parametric family of sparse to non sparse signals is given next.

Example 6.2.2 Let x be a signal of length $N = 1000$ with $x_n \in [0, 1]$ for every $n \in \{0, 1, \ldots, N-1\}$. We now consider the evaluation of φ_0 for power laws $x \to x^p$, $(p > 0)$. Figure **??** - p. **??** illustrates the x^p signal where on the top of figure: $p = 100$ (left), x^p is sparse with $\varphi_0(x^p) = 4.44$ and $p = 0.01$ (right), x^p is not sparse with $\varphi_0(x^p) = 31.61$. In the bottom of figure: the representation of different signals when p varies (left) and the evaluation of φ_0 when p varies (right). We see that, the bigger p, the smaller $\varphi_0(x^p)$.

Figure **??** - p. **??** is a visualization of the ℓ_0-"quasi-norm" (top-left), ℓ_1-norm (top-right), $\ell_{\frac{1}{2}}$-quasi-norm (bottom-left) and φ_0 for a two-dimensional vector. The morphology of φ_0 is the most similar to that of ℓ_0. However, the ratio φ_0 is difficult to use, we thus propose to replace the nonsmooth function φ_0 by a more maniable smooth approximation. In this work, we investigate two kinds of smooth approximation (surrogate functions):

1. The first one is a natural way to replace ℓ_1 and ℓ_2 by their smooth approximations $\ell_{1,\alpha}$ (sometimes called hybrid ℓ_1-ℓ_2 or hyperbolic penalty) and $\ell_{2,\eta}$, where

$$\ell_{1,\alpha}(x) = \sum_{n=1}^{N} \left(\sqrt{x_n^2 + \alpha^2} - \alpha \right), \; \ell_{2,\eta}(x) = \sqrt{\sum_{n=1}^{N} x_n^2 + \eta^2}$$

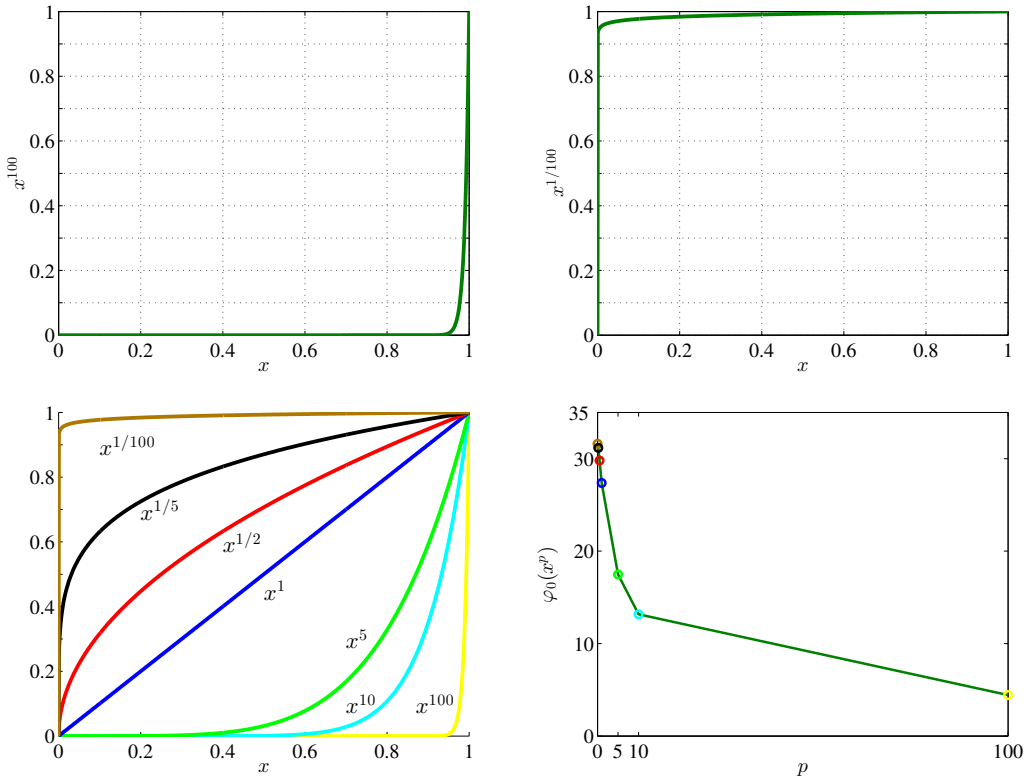


Figure 6.3: *Evaluation of φ_0 for power laws.*

with $(\alpha, \eta) \in]0, +\infty[^2$. Note that ℓ_1 and ℓ_2 are recovered for $\alpha = \eta = 0$. Thus, the considered function is

$$\widetilde{\varphi}(x) = \lambda \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)} \tag{6.2}$$

where, $\lambda \in]0, +\infty[$.

2. The second one is a little bit more sophisticated

$$\varphi(x) = \lambda \log\left(\frac{\ell_{1,\alpha}(x) + \beta}{\ell_{2,\eta}(x)}\right), \tag{6.3}$$

with $(\lambda, \beta, \alpha, \eta) \in]0, +\infty[^4$.

The log function both makes the penalty easier to handle and, through its concavity, strengthens the flattening of the ℓ_1/ℓ_2 function toward sparsity. We now turn our attention to the properties of these functions.

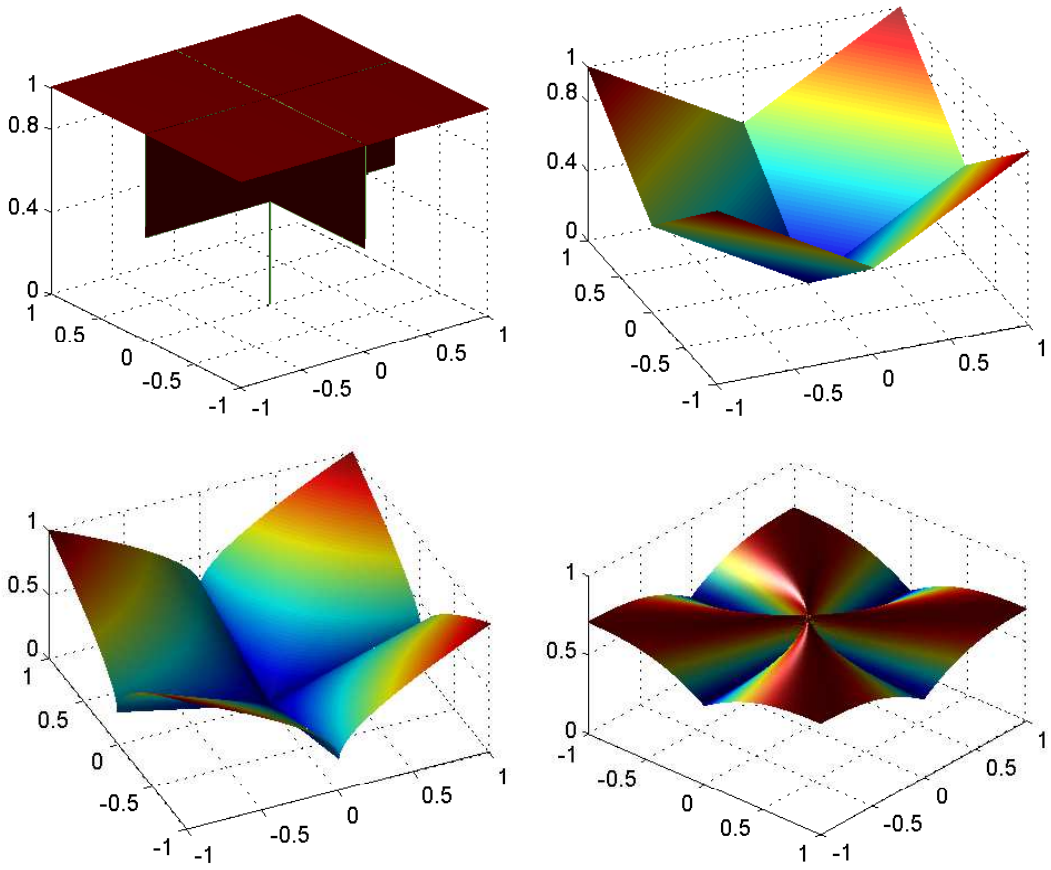


Figure 6.4: *Measures for the case of $N = 2$: ℓ_0-"quasi-norm" at the top left, ℓ_1-norm at the top right, $\ell_{1/2}$-quasi-norm at the bottom left and ℓ_1/ℓ_2-function at the bottom right.*

6.2.2 Properties of functions ℓ_1/ℓ_2 and its surrogates

We concisely state three results that will be necessary thereafter. The proofs are given in the appendix.

Proposition 6.2.3 *For all $\alpha > 0$, $\eta > 0$ and $x \in \mathbb{R}^N$,*

$$\frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}^2(x)} \leq \frac{1}{2(\alpha + \frac{\eta}{\sqrt{N}})}. \tag{6.4}$$

The proof is given in Appendix A.1.

Proposition 6.2.4 *$\widetilde{\varphi}$ has a $\widetilde{\mu}$-Lipschitzian gradient on $\mathbb{R}^N$ (i.e., for every $(z, z') \in (\mathbb{R}^N)^2$, $\|\nabla\widetilde{\varphi}(z) - \nabla\widetilde{\varphi}(z')\| \leq \widetilde{\mu}\|z - z'\|$), with*

$$\widetilde{\mu} = \lambda \left(1 + \frac{4}{3\sqrt{3}}\right) \frac{1}{\alpha\eta}. \tag{6.5}$$

The proof is given in Appendix A.2.

Proposition 6.2.5 *φ has a μ-Lipschitzian gradient on $\mathbb{R}^N$, with*

$$\mu{=}\lambda\left(\max\left\{\frac{1}{\beta\alpha},\frac{1}{\eta^2}\right\}+\frac{1}{2\alpha\beta}+\frac{1}{2\eta^2}\right). \tag{6.6}$$

The proof is given in Appendix A.3.

6.3 Optimization model

6.3.1 Optimization tools

Our minimization strategy relies on two optimization principles. Let $U \in \mathbb{R}^{M\times M}$ be a symmetric positive definite (SPD) matrix. Firstly, we define the U-weighted proximity operator (Hiriart-Urruty and Lemaréchal, 1993, Sec. XV.4), (Combettes and Vũ, 2013) of a proper, lower semicontinuous, convex function $\psi\colon \mathbb{R}^M \to]-\infty,+\infty]$ at $z \in \mathbb{R}^M$, relative to the metric induced by U, and denoted by $\mathrm{prox}_{U,\psi}(z)$, as the unique minimizer of $\psi + \frac{1}{2}\|\cdot - z\|_U^2$, where $\|.\|_U$ denotes the weighted Euclidean norm, i.e., $(\forall z \in \mathbb{R}^M)$ $\|z\|_U = \left(z^\top U z\right)^{1/2}$. When U is equal to I_M, the identity matrix of $\mathbb{R}^{M\times M}$, then $\mathrm{prox}_{\mathrm{I}_M,\psi}$ reduces to the original definition of the proximity operator in (Moreau, 1965). Secondly, we introduce the Majoration-Minimization (MM) principle:

Definition 6.3.1 *Let $\zeta\colon \mathbb{R}^M \to \mathbb{R}$ be a differentiable function. Let $z \in \mathbb{R}^M$. Let us define, for every $z' \in \mathbb{R}^M$,*

$$q(z',z) = \zeta(z) + (z'-z)^\top \nabla\zeta(z) + \frac{1}{2}\|z'-z\|^2_{U(z)},$$

where $U(z) \in \mathbb{R}^{M\times M}$ is a SPD matrix. Then, $U(z)$ satisfies the majoration condition for ζ at z if $q(\cdot,z)$ is a majorant of the function ζ at z, i.e., for every $z' \in \mathbb{R}^M$, $\zeta(z') \leq q(z',z)$.

If function ζ has an L-Lipschitzian gradient on a convex subset $C \subset \mathbb{R}^M$, then, for every $z \in C$, a quadratic majorant of ζ at z is trivially obtained by taking $U(z) = L\,\mathrm{I}_M$.

We now shall give some facts concerning "semi-algebraic functions" and subsets which we need to use in this chapter.

Definition 6.3.2 *A semi-algebraic subset of $\mathbb{R}^N$ (Bierstone and Milman, 1988) is a finite union of sets of the form*

$$\{x \in \mathbb{R}^N \;:\; f_i(x) = 0,\, g_j(x) < 0, i \in \mathrm{I}, j \in \mathrm{J}\}$$

where I, J *are finite and for every $i \in \mathrm{I}$ and $j \in \mathrm{J}$, f_i, $g_j : \mathbb{R}^N \to \mathbb{R}$ are real polynomial functions. A mapping is semi-algebraic if its graph is a semi-algebraic set.*

Remark 6.3.1 *Semi-algebraicity is a property satisfied by a wide class of functions. In particular, it is satisfied for function φ defined in* (**??**)*, for standard numerical implementations of the* log *function.*

6.3.2 Proposed variational approach

From now on, definitions and properties apply for every $x = (x_n)_{1\leq n\leq N} \in \mathbb{R}^N$ and $h \in \mathbb{R}^S$, unless otherwise stated. We propose to define an estimate $(\widehat{x}, \widehat{h})$ of $(\overline{x}, \overline{h})$ as a minimizer of the following penalized criterion:

$$G(x,h) = \rho(x,h) + g(x,h) + \varphi(x), \tag{6.7}$$

where $\rho(x,h) = \frac{1}{2}\|h * x - y\|^2$ is the least-squares objective function, g introduces additional a priori information on the sought objects. Finally, as provided in Proposition **??**, φ has a Lipschitzian gradient on $\mathbb{R}^N$, which is a desirable property for deriving an efficient algorithm to minimize (**??**). In the following, we assume that g can be split as

$$g(x,h) = g_1(x) + g_2(h), \tag{6.8}$$

where g_1 and g_2 are proper, lower semicontinuous, convex functions, continuous on their domain. Thanks to these properties, the problem can be addressed with the Block Coordinate Variable Metric Forward-Backward algorithm (Chouzenoux et al., 2013). Moreover, in practice, h and x have different properties and this choice permits to take into account independently the a priori infomation on the sought objects. Moreover, we denote by

$$f(x,h) = \rho(x,h) + \varphi(x), \tag{6.9}$$

the smooth part of the criterion, and $\nabla_1 f(x,h) \in \mathbb{R}^N$ (resp. $\nabla_2 f(x,h) \in \mathbb{R}^S$) the partial gradient of f with respect to the variable x (resp. h) computed at (x,h).

6.4 Proposed alternating optimization method

6.4.1 Smoothed ℓ_1/ℓ_2 (SOOT) algorithm

To minimize (**??**), one can exploit the block-variable structure of G by using an alternating forward-backward algorithm (Luo and Tseng, 1992; Bolte et al., 2010; Xu and Yin, 2013; Bolte et al., 2014; Chouzenoux et al., 2013). At each iteration $k \in \mathbb{N}$, this algorithm updates x^k (resp. h^k) with a gradient step on $f(\cdot, h^k)$ (resp. $f(x^{k+1}, \cdot)$) followed by a proximity step on g_1 (resp. g_2).

We use this alternating minimization method combined with an MM strategy, as described in (Chouzenoux et al., 2013). For every $(x,h) \in \mathbb{R}^N \times \mathbb{R}^S$, let us assume the existence of SPD matrices $A_1(x,h) \in \mathbb{R}^{N\times N}$ and $A_2(x,h) \in \mathbb{R}^{S\times S}$ such that $A_1(x,h)$

Algorithm 6 SOOT algorithm.

For every $k \in \mathbb{N}$, let $J_k \in \mathbb{N}^*$, $I_k \in \mathbb{N}^*$ and let $(\gamma_1^{k,j})_{0\leq j\leq J_k-1}$ and $(\gamma_2^{k,i})_{0\leq i\leq I_k-1}$ be positive sequences. Initialize with $x^0 \in \operatorname{dom} g_1$ and $h^0 \in \operatorname{dom} g_2$.

Iterations:

For $k = 0, 1, \ldots$

$x^{k,0} = x^k,\ h^{k,0} = h^k,$

For $j = 0, \ldots, J_k - 1$

$\widetilde{x}^{k,j} = x^{k,j} - \gamma_1^{k,j} A_1(x^{k,j}, h^k)^{-1} \nabla_1 f(x^{k,j}, h^k),$

$x^{k,j+1} = \operatorname{prox}_{(\gamma_1^{k,j})^{-1} A_1(x^{k,j},h^k),g_1}\left(\widetilde{x}^{k,j}\right),$

$x^{k+1} = x^{k,J_k}.$

For $i = 0, \ldots, I_k - 1$

$\widetilde{h}^{k,i} = h^{k,i} - \gamma_2^{k,i} A_2(x^{k+1}, h^{k,i})^{-1} \nabla_2 f(x^{k+1}, h^{k,i}),$

$h^{k,i+1} = \operatorname{prox}_{(\gamma_2^{k,i})^{-1} A_2(x^{k+1},h^{k,i}),g_2}\left(\widetilde{h}^{k,i}\right),$

$h^{k+1} = h^{k,I_k}.$

(resp. $A_2(x,h)$) satisfies the majoration condition for $f(\cdot, h)$ at x (resp. $f(x,\cdot)$ at h). Then, the SOOT algorithm for the minimization of (**??**) is described in Algorithm **??**.

Note that PALM algorithm (Bolte et al., 2014) is recovered as a special case if $J_k \equiv I_k \equiv 1$ and, at each iteration, the Lipschitz constant of $\nabla_1 f(\cdot, h^k)$ (resp. $\nabla_2 f(x^{k+1}, \cdot)$) is substituted for $A_1(x^{k,0}, h^k)$ (resp. $A_2(x^{k+1}, h^{k,0})$). However, recent works on variable metric strategies (Chouzenoux et al., 2013, 2014) show that the use of more judicious preconditioning matrices can significantly accelerate the convergence of the algorithm. An example of such matrices is proposed in Section **??**. Moreover, we show in our experimental part the practical interest in terms of convergence speed of taking the number of inner loops $(I_k)_{k\in\mathbb{N}}$ or $(J_k)_{k\in\mathbb{N}}$ greater than one.

The convergence of Algorithm **??** can be derived from the general results established in (Chouzenoux et al., 2013):

Proposition 6.4.1 *Let $(x^k)_{k\in\mathbb{N}}$ and $(h^k)_{k\in\mathbb{N}}$ be sequences generated by Algorithm **??**. Assume that:*

1. *There exists $(\underline{\nu}, \overline{\nu}) \in]0, +\infty[^2$ such that, for all $k \in \mathbb{N}$,*

$$(\forall j \in \{0, \ldots, J_k - 1\}) \quad \underline{\nu}\, \mathrm{I}_N \preceq A_1(x^{k,j}, h^k) \preceq \overline{\nu}\, \mathrm{I}_N,$$
$$(\forall i \in \{0, \ldots, I_k - 1\}) \quad \underline{\nu}\, \mathrm{I}_S \preceq A_2(x^{k+1}, h^{k,i}) \preceq \overline{\nu}\, \mathrm{I}_S\,.$$

2. *Step-sizes $(\gamma_x^{k,j})_{k\in\mathbb{N},0\leq j\leq J_k-1}$ and $(\gamma_h^{k,i})_{k\in\mathbb{N},0\leq i\leq I_k-1}$ are chosen in the interval $[\underline{\gamma}, 2-\overline{\gamma}]$ where $\underline{\gamma}$ and $\overline{\gamma}$ are some given positive real constants.*

3. *g is a semi-algebraic function.*

*Then, the sequence $(x^k, h^k)_{k\in\mathbb{N}}$ converges to a critical point $(\widehat{x}, \widehat{h})$ of (**??**). Moreover, $\left(G(x^k, h^k)\right)_{k\in\mathbb{N}}$ is a nonincreasing sequence converging to $G(\widehat{x}, \widehat{h})$.*

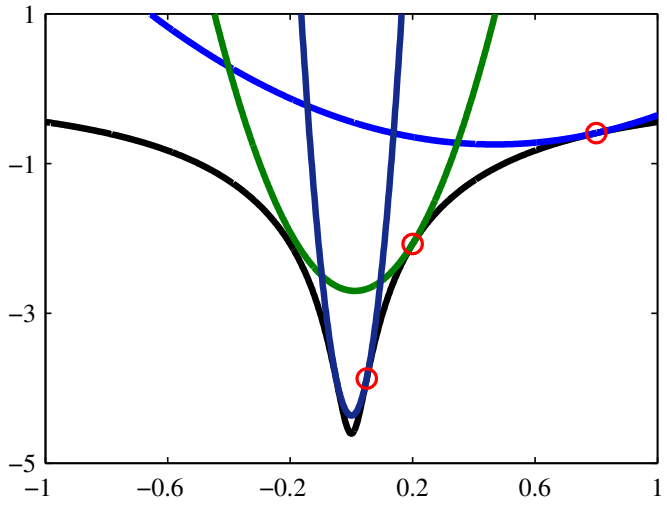


Figure 6.5: *Function φ (in black) and its majorization functions at $x = 0.8$ (in blue), $x = 0.2$ (in purple) and $x = 0.05$ (in green).*

6.4.2 Construction of quadratic majorants

The numerical efficiency of the SOOT algorithm **??** relies on the use of quadratic majorants providing tight approximations to the criterion and whose curvature matrices are simple to compute. The following proposition allows us to propose SPD matrices A_1 and A_2 for building majorizing approximations of f with respect to x and h.

Proposition 6.4.2 *For every $(x, h) \in \mathbb{R}^N \times \mathbb{R}^S$, let*

$$A_1(x,h) = \left(L_1(h) + \frac{9\lambda}{8\eta^2}\right) \mathrm{I}_N + \frac{\lambda}{\ell_{1,\alpha}(x) + \beta} A_{\ell_{1,\alpha}}(x),$$
$$A_2(x,h) = L_2(x)\, \mathrm{I}_S,$$

where

$$A_{\ell_{1,\alpha}}(x) = \mathrm{Diag}\left(\left((x_n^2 + \alpha^2)^{-1/2}\right)_{1 \le n \le N}\right), \tag{6.10}$$

and $L_1(h)$ (resp. $L_2(x)$) is a Lipschitz constant for $\nabla_1 \rho(\cdot, h)$ (resp. $\nabla_2 \rho(x, \cdot)$).[1] *Then, $A_1(x,h)$ (resp. $A_2(x,h)$) satisfies the majoration condition for $f(\cdot, h)$ at x (resp. $f(x, \cdot)$ at h).*

Proof. Let us decompose $\varphi = \varphi_1 + \varphi_2$ $(\forall\, x \in \mathbb{R}^N)$ with, $\varphi_1(x) = \lambda \log\left(\ell_{1,\alpha}(x) + \beta\right)$ and $\varphi_2(x) = -\lambda \log\left(\ell_{2,\eta}(x)\right)$. It then suffices to prove that, for every $x \in \mathbb{R}^N$,

(i) $A_{\varphi_1}(x) = \frac{\lambda}{\ell_{1,\alpha}(x)+\beta} A_{\ell_{1,\alpha}}(x)$ satisfies the majoration condition for φ_1 at x,

[1] Such Lipschitz constants are straightforward to derive since ρ is a quadratic cost.

(ii) φ_2 has a μ-Lipschitzian gradient, with $\mu = \frac{9\lambda}{8\eta^2}$.

On the one hand, setting $\tau(x) = \ell_{1,\alpha}(x) + \beta$, we have (Allain et al., 2006)

$$\tau(x') \leq \tau(x) + (x' - x)^\top \nabla\tau(x) + \frac{1}{2}\|x' - x\|^2_{A_{\ell_{1,\alpha}}(x)}, \tag{6.11}$$

for every $x' \in \mathbb{R}^N$, where $A_{\ell_{1,\alpha}}(x)$ is given by (**??**).

On the other hand, for every $(u, v) \in]0, +\infty[^2$,

$$\log v \leq \log u + \frac{v}{u} - 1 = \log u + \frac{v - u}{u}. \tag{6.12}$$

By taking $v = \tau(x') > 0$ and $u = \tau(x) > 0$, and by combining (**??**) and (**??**), we obtain

$$\varphi_1(x') \leq \varphi_1(x) + \frac{\lambda}{\tau(x)}(x' - x)^\top \nabla\tau(x) + \frac{1}{2}(x' - x)^\top \frac{\lambda}{\tau(x)} A_{\ell_{1,\alpha}}(x)(x' - x).$$

Thus, Statement (i) is proved by remarking that $\nabla\varphi_1(x) = \frac{\lambda}{\tau(x)}\nabla\tau(x)$ and $A_{\varphi_1}(x) = \frac{\lambda}{\tau(x)} A_{\ell_{1,\alpha}}(x)$. On the other hand, the Hessian of φ_2 is given by

$$\nabla^2\varphi_2(x) = \frac{2\lambda}{\ell^4_{2,\eta}(x)} xx^\top - \frac{\lambda}{\ell^2_{2,\eta}(x)}\, \mathrm{I}_N\,.$$

Noting that $\ell^2_{2,\eta}(x) = \|x\|^2 + \eta^2$, and applying the triangular inequality yields

$$\left|\left|\left|\nabla^2\varphi_2(x)\right|\right|\right| \leq \frac{2\lambda\|x\|^2}{(\|x\|^2 + \eta^2)^2} + \frac{\lambda}{\|x\|^2 + \eta^2} = \chi(\|x\|),$$

where $\chi\colon u \in [0, +\infty[\mapsto \lambda\frac{3u^2+\eta^2}{(u^2+\eta^2)^2}$. The derivative of χ is given, for every $u \in [0, +\infty[$, by

$$\dot\chi(u) = \lambda\frac{2u}{(u^2 + \eta^2)^3}(\eta^2 - 3u^2),$$

thus χ is an increasing function on $[0, \eta/\sqrt{3}]$ and a decreasing function on $]\eta/\sqrt{3}, +\infty[$, and $\max_{u\in[0,+\infty[}\chi(u) = \chi\left(\eta/\sqrt{3}\right) = \frac{9\lambda}{8\eta^2}$. Hence, the proof of Statement (ii). ■

Figure **??** - p. **??** gives an illustration of this majorant for $N = 1$.

Remark 6.4.1 *Note that, the Lipschitz constant for $\nabla_2\rho(x, \cdot)$ is zero when x is a vector of zeros. Thus, to ensure that condition 1 of Proposition* **??** *holds, we add a very small positive value $\widetilde{\beta}$ to the Lipschitz constant i.e. $A_2(x, h) = (L_2(x) + \widetilde{\beta})\,\mathrm{I}_S$.*

Algorithm 7 Overall[1] algorithm (Krishnan et al., 2011) (1D Overall adapted algorithm).

For every $k \in \mathbb{N}$, let $J \in \mathbb{N}^*$, $I_k \in \mathbb{N}^*$, let $\kappa \in]0, +\infty[$ and $\delta = 10^{-3}$. Initialize with $x^0 \in \operatorname{dom} \varphi_0$ and $h^0 \in \operatorname{dom} g_2$.

Iterations:

For $k = 0, 1, \ldots$

- $x^{k,0} = x^k$, $h^{k,0} = h^k$,
- $d = \frac{\kappa}{2}\|H(h^k, x^{k,0}) - y\|_2^2 + \varphi_0(x^{k,0})$,
- While $(\delta > 10^{-4})$
  - $c = d$,
  - For $j = 0, \ldots, J-1$
    - $\beta = \kappa\|x^{k,j,0}\|_2$
    - $x^{k,j,0} = x^{k,j}$
    - For $\ell = 0, \ldots, L-1$
      - $\widetilde{x}^{k,j,\ell} = x^{k,j,\ell} - \delta H^*(h^k, x^{k,j,\ell})\left(H(h^k, x^{k,j,\ell}) - y\right)$
      - $x^{k,j+1,\ell} = \operatorname{prox}_{\delta\beta\|\cdot\|_1}(\widetilde{x}^{k,j,\ell})$
    - $x^{k,j+1} = x^{k,j,L}$
  - $d = \frac{\kappa}{2}\|H^*(h^k, x^{k,j+1}) - y\|_2^2 + \varphi_0(x^{k,j+1})$
  - If $(d > 3c)$
    - $x^{k,0} = x^k$
    - $\delta = \frac{\delta}{2}$
  - Else
    - exit loop
- $x^{k+1} = x^{k,J}$.
- For $i = 0, \ldots, I_k - 1$
  - $\widetilde{h}^{k,i} = h^{k,i} - \gamma_h^{k,i} A_2(x^{k+1}, h^{k,i})^{-1} \nabla_2 f(x^{k+1}, h^{k,i})$,
  - $h^{k,i+1} = \operatorname{prox}_{(\gamma_h^{k,i})^{-1} A_2(x^{k+1}, h^{k,i}), g_2}\left(\widetilde{h}^{k,i}\right)$,
- $h^{k+1} = h^{k,I_k}$.

6.5 Evaluation on seismic signal deconvolution

6.5.1 Problem statement

As some of the earliest mentions of ℓ_1/ℓ_2 deconvolution appeared in geophysics (Gray, 1978), blind seismic deconvolution (or inversion (Osman and Robinson, 1996; Ulrych and Sacchi, 2006)) is a natural application. The sparse seismic signal $\overline{x}$, of length $N = 784$, on the top of Figure **??** - p. **??** is composed of a sequence of spikes termed primary reflection coefficients (Walden and Hosken, 1986). This reflectivity series indicates, in reflection seismology at normal incidence, the travel time of seismic waves between two seismic reflectors, and the amplitude of the seismic events reflected back to the sensor. The observed seismic trace y displayed in Figure **??**-bottom - p. **??** follows Model (**??**).

In this context, the blur $\overline{h}$ is related to the generated seismic source. We use here a band-pass "Ricker" seismic wavelet (or Mexican hat (Ricker, 1940), Figure 2.5-left - p. 19) of size $S = 41$ (Figure **??** - p. **??**) with a frequency spectrum concentrated between 10 and 40 Hz. The additive noise w is a realization of a zero-mean white Gaussian noise with variance σ^2. Since the reflectivity series is sparse, but limited in amplitude, we choose g_1 as the indicator function of the convex hypercube $[x_{\min}, x_{\max}]^N$. Similarly, as the seismic wavelet possesses finite energy, g_2 is equal to the indicator function of the set $\mathcal{C} = \{h \in [h_{\min}, h_{\max}]^S \mid \|h\| \leq \delta\}$, where $\delta > 0$, and $h_{\min}$ (resp. $h_{\max}$) is the minimum (resp. maximum) value of $\overline{h}$.

6.5.2 Numerical results

Figure **??** - p. **??** presents the variations of the reconstruction time, in seconds, with respect to the number of inner-loops $J_k \equiv J$, with $I_k \equiv 1$ and noise level $\sigma = 0.03$. The reconstruction time corresponds to the stopping criterion $\|x^k - x^{k-1}\| \leq \sqrt{N} \times 10^{-6}$. One can observe that the best compromise in terms of convergence speed is obtained for an intermediate number of inner-loops, namely $J = 71$. Note that the quality of the reconstruction is stable for each choice of J.

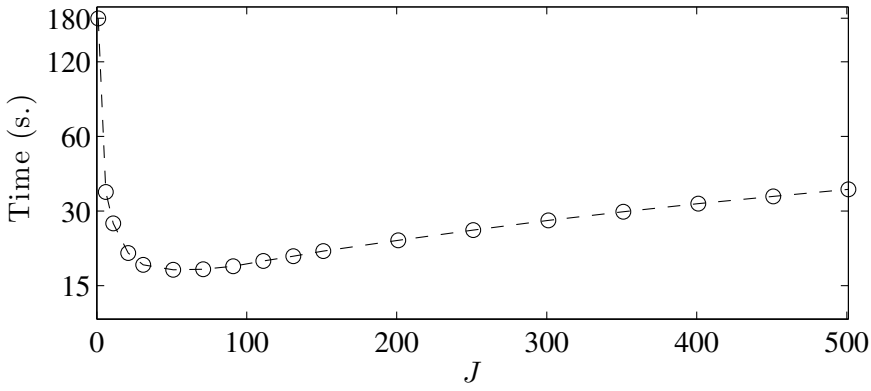


Figure 6.6: *Reconstruction time for different numbers of inner-loops $J_k \equiv J$ (average over thirty noise realizations).*

We gather comparisons of the SOOT algorithm **??** with Overall[1] algorithm **??** in Table **??**, where the same initialization strategy has been used for both algorithms: x^0 is a constant-valued signal such that $x^0 \in [x_{\min}, x_{\max}]^N$, and h^0 is a centered Gaussian filter, such that $h^0 \in \mathcal{C}$. Results presented in this table, for each noise level σ, are averaged over two hundred noise realizations. The regularization parameters of Overall[1] algorithm **??** and $(\lambda, \alpha, \beta, \eta) \in]0, +\infty[^4$ of (**??**) are adjusted so as to minimize the ℓ_1 norm between the original and the reconstructed signals. We also set, for every $k \in \mathbb{N}$, $J_k = 71$ and $I_k = 1$. If both methods yield tremendous improvements in ℓ_2 and ℓ_1 norms, the SOOT algorithm **??** exhibits better results, for all noise levels, for both $\overline{x}$

and $\overline{h}$ estimates, especially in terms of ℓ_1 norm. Interestingly, the SOOT algorithm **??** is also significantly faster in this application.

Noise level (σ)			0.01	0.02	0.03
Observation error		ℓ_2 ($\times 10^{-2}$)	7.14	7.35	7.68
		ℓ_1 ($\times 10^{-2}$)	2.85	3.44	4.09
Signal error	Overall[1] algorithm **??**	ℓ_2 ($\times 10^{-2}$)	1.23	1.66	1.84
		ℓ_1 ($\times 10^{-3}$)	3.79	4.69	5.30
	SOOT algorithm **??**	ℓ_2 ($\times 10^{-2}$)	1.09	1.63	1.83
		ℓ_1 ($\times 10^{-3}$)	3.42	4.30	4.85
Kernel error	Overall[1] algorithm **??**	ℓ_2 ($\times 10^{-2}$)	1.88	2.51	3.21
		ℓ_1 ($\times 10^{-2}$)	1.44	1.96	2.53
	SOOT algorithm **??**	ℓ_2 ($\times 10^{-2}$)	1.62	2.26	2.93
		ℓ_1 ($\times 10^{-2}$)	1.22	1.77	2.31
Time (s.)	Overall[1] algorithm **??**		106	61	56
	SOOT algorithm **??**		56	22	18

Table 6.1: *Comparison between Overall[1] algorithm* **??** *and SOOT algorithm* **??** *for $\overline{x}$ and $\overline{h}$ estimates (Intel(R) Xeon(R) CPU E5-2609 v2@2.5GHz using Matlab 8).*

Table **??** presents the variances associated with the averaged results of Table **??**. It can be noticed that the variance on the error values are of similar order of magnitude for both algorithms (typically around 10% of the error averages), which confirms the statistical consistency of the presented results. While the variance of the signal errors may be a little bit larger for the SOOT algorithm than for Overall[1] algorithm **??**, the opposite behavior being observed for the kernel errors. SOOT algorithm appears to be more stable than Overall[1] algorithm **??**, in terms of computational time.

The performance is further assessed by subjective results for $\sigma = 0.03$. The plots of $\overline{x}$ and $\widehat{x}$ can be found below (Figure **??** - p. **??**). One can observe that the reconstructed signals for both algorithms are visually similar, with the exception of a few under-estimated peaks for Overall[1] algorithm **??**. We found however that it is difficult to see the differences between SOOT and Overall[1] algorithm **??** results on this figure. For this reason, Figure **??** - p. **??** shows the residual error of the sparse signal estimation $\overline{x} - \widehat{x}$, for a given noise realization, where $\widehat{x}$ is estimated with Overall[1] algorithm **??** in (a), and with SOOT algorithm **??** in (b). It appears, in this example, that the error is smaller using SOOT algorithm **??**. The estimated blur kernels look similar for both methods, as displayed in Figure **??** - p. **??**.

6.6 Evaluation in blind image deconvolution

Similar ideas can be reused in the context of blind image deconvolution. The main difference resides in that natural images are not considered directly sparse. The ℓ_1/ℓ_2

Noise level (σ)			0.01	0.02	0.03
Observation error		ℓ_2 ($\times 10^{-4}$)	3.52	7.02	10.42
		ℓ_1 ($\times 10^{-4}$)	2.98	5.55	8.06
Signal error	Overall[1] algorithm ??	ℓ_2 ($\times 10^{-3}$)	1.35	2.35	2.09
		ℓ_1 ($\times 10^{-4}$)	2.17	3.36	3.49
	SOOT algorithm ??	ℓ_2 ($\times 10^{-3}$)	2.92	3.44	2.74
		ℓ_1 ($\times 10^{-4}$)	3.88	4.81	4.07
Kernel error	Overall[1] algorithm ??	ℓ_2 ($\times 10^{-3}$)	2.65	6.24	6.36
		ℓ_1 ($\times 10^{-3}$)	1.99	3.96	4.35
	SOOT algorithm ??	ℓ_2 ($\times 10^{-3}$)	2.94	2.75	3.43
		ℓ_1 ($\times 10^{-3}$)	2.04	2.25	2.86
Time (s.)	Overall[1] algorithm ??		54	45	46
	SOOT algorithm ??		11	4	6

Table 6.2: *Variances on error values and computation times, for algorithms Overall[1] algorithm ?? and SOOT algorithm ?? (Intel(R) Xeon(R) CPU E5-2609 v2@2.5GHz using Matlab 8).*

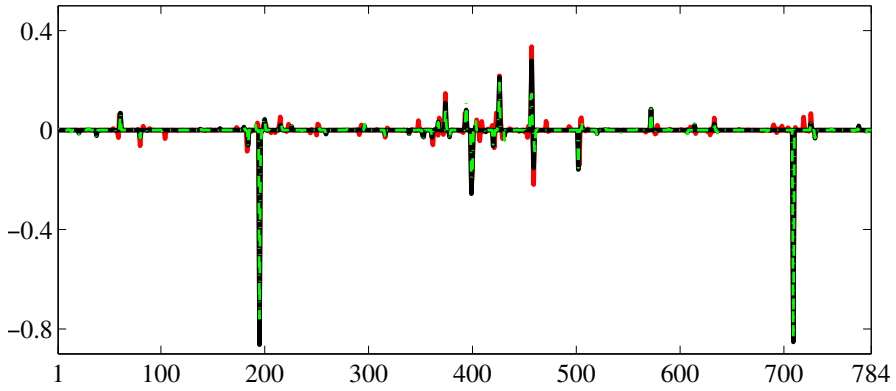


Figure 6.7: *Original signal $\overline{x}$ (continuous red), estimated $\widehat{x}$ with SOOT algorithm ??(continuous black) and Overall[1] algorithm ?? (dashed green).*

function is thus not efficient for measuring sparsity in all their frequency bands. Working on higher frequencies related to contours is more appropriate. At a given resolution, the ℓ_1/ℓ_2 function or its surrogate is applied to the horizontal and vertical gradients. We recast the approach in (Krishnan et al., 2011) in the following two main steps. The first step solves the following optimization problem:

$$\underset{u\in\mathbb{R}^{2N^2},h\in\mathbb{R}^S}{\text{minimize}} \ \frac{\kappa}{2}\|\widetilde{H}(h,u)-v\|^2+\Psi(u)+g(u,h)\,. \tag{6.13}$$

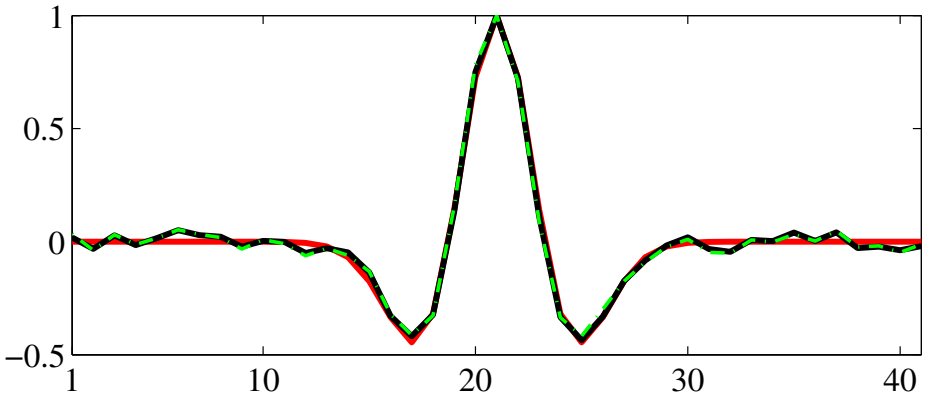


Figure 6.8: *Original blur $\overline{h}$ (continuous red), estimated $\widehat{h}$ with SOOT algorithm* **??** *(continuous black) and Overall*[1] *algorithm* **??** *(dashed green).*

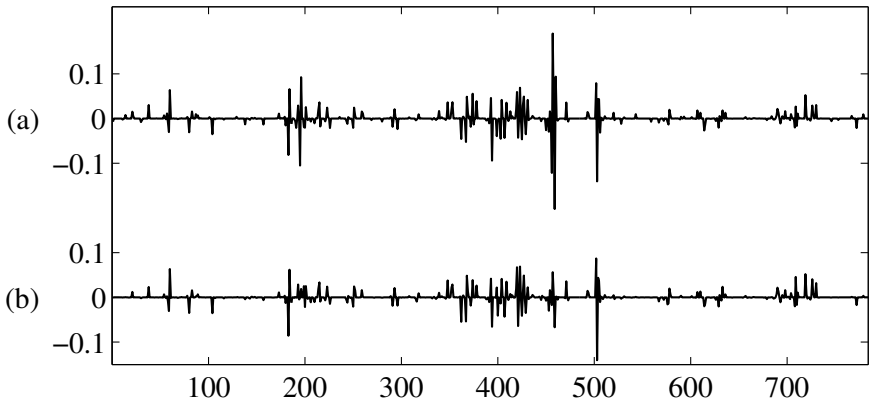


Figure 6.9: *Signal estimation error $\overline{x}-\widehat{x}$ with estimates $\widehat{x}$ given by Overall*[1] *algorithm* **??** *(a) and SOOT algorithm* **??** *(b).*

Here, ∇_x and ∇_y denote the gradient operators in the horizontal and vertical directions, respectively, $v = \begin{bmatrix}\nabla_x \\ \nabla_y\end{bmatrix} y$. With $u = \begin{bmatrix}u_x \\ u_y\end{bmatrix} \in \mathbb{R}^{2N^2}$, the operator $\widetilde{H} : \mathbb{R}^{2N^2} \to \mathbb{R}^{2N^2} : u \mapsto \begin{bmatrix}H(h,u_x) \\ H(h,u_y)\end{bmatrix}$ combines blur and gradient. The penalties are $\Psi(u) = \varphi(u_x) + \varphi(u_y)$ and $g(u,h) = \iota_{[u_{\min},u_{\max}]^{2N^2}}(u) + \iota_D(h)$ with $D = \left\{h \in [0,+\infty[^S : \sum_{s=0}^{S-1} h_s = 1\right\}$ and we introduce $f(u,h) = \frac{\kappa}{2}\|\widetilde{H}(h,u) - v\|^2 + \Psi(u)$. Once the blur has been estimated, the second step solves a non-blind image deconvolution problem. More precisely for these two steps:

❶: As in (Fergus et al., 2006), a coarse-to-fine loop over a pyramid of image resolutions achieves a multiscale estimation of the blur. We refer to (Krishnan et al., 2011, Section 3.1.3) for additional details.

❷: The non-blind deconvolution involves a criterion composed of a sum of a quadratic function and a convex function composed with linear operators. It can be minimized properly with PPXA+ (Pesquet and Pustelnik, 2012) or primal-dual approaches such as the M+LFBF algorithm (Combettes and Pesquet, 2012).

Our 2D adaptation (hereafter denoted by SOOT^2) plugs the SOOT algorithm in the first step of the Overall algorithm (Algorithm **??**). The latter is rewritten in a more compact form, as detailed in Algorithm **??**.

Remark 6.6.1 *For every* $(u,h) \in \mathbb{R}^{2N^2} \times \mathbb{R}^S$ *and* $\gamma \in]0,+\infty[$ *, let* $A_1(u,h)$ *be the majorant matrix of* $f(\cdot,h)$ *at* u*. Then,*

$$\mathrm{prox}_{(\gamma)^{-1}A_1(u,h),\iota_{[u_{\min},u_{\max}]^{2N^2}}} = \Pi_{[u_{\min},u_{\max}]^{2N^2}}\,.$$

Similarly, the majorant matrix of $f(u,\cdot)$ *at* h*,* $A_2(u,h)$ *is a diagonal matrix whose diagonal elements all equal the same scalar value (the Lipschitz constant for* $\nabla_2 f(u,\cdot)$*). Then,*

$$\mathrm{prox}_{(\gamma)^{-1}A_2(u,h),\iota_D} = \Pi_D,$$

which can be easily computed.

We compare our results to (Krishnan et al., 2011) in Figure **??** - p. **??** and Figure **??** - p. **??**. We observe that we better estimate the blur than the Overall algorithm. Moreover, the Overall algorithm took over 7 minutes for the example shown in Figure **??** - p. **??** and over 2 minutes for the example shown in Figure **??** - p. **??**. SOOT^2 only takes 45 seconds and 7 seconds, respectively, on the same CPU. The deblurred images are visually similar for both algorithms. We gain 1.83 dB in terms of SNR for the first test (Figure **??** - p. **??**) and 0.86 dB for the second test (Figure **??** - p. **??**). A thorough comparative assessment of SOOT^2 and the algorithm from (Krishnan et al., 2011) for blind image deconvolution however deserves additional evaluations on speed, objective and subjective quality results.

6.7 Conclusion

The proposed SOOT algorithm **??** and its 2D adaptation SOOT^2 in algorithm **??** for minimizing a modified ℓ_1/ℓ_2 penalized criterion have been demonstrated to be quite effective in blind deconvolution applications on seismic reflectivity data and on images, respectively. In addition, SOOT offers theoretically guaranteed convergence. The performance of the proposed method is demonstrated through a comparison with a recent alternating optimization strategy dealing with the exact ℓ_1/ℓ_2 term, on both blind sparse signal and image deconvolution. The application of the method using a non-quadratic data fidelity term, in association with more sophisticated preconditioning matrices, is finally of main interest.

Algorithm 8 Overall algorithm (Krishnan et al., 2011).

For every $k \in \mathbb{N}$, let $J \in \mathbb{N}^*$, let $(\kappa, \tau) \in]0, +\infty[^2$, and $\delta = 10^{-3}$. Initialize with $u_x^0 \in \operatorname{dom} \varphi_0 \circ \nabla_x$, $u_y^0 \in \operatorname{dom} \varphi_0 \circ \nabla_y$ and $h^0 \in \operatorname{dom} g_2$.
Explanations for the steps ❶ and ❷ can be found in the text.

1. Blind estimation of blur

❶ **Loop over coarse-to-fine levels**

Alternate:

For $k = 0, 1, \ldots, K$

- $u_x^{k,0} = u_x^k$, $u_y^{k,0} = u_y^k$, $h^{k,0} = h^k$,
- $d = \frac{\kappa}{2}\left(\|H(h^k, u_x^{k,0}) - v_x\|_2^2 + \|H(h^k, u_y^{k,0}) - v_y\|_2^2\right) + \varphi_0(u_x^{k,0}) + \varphi_0(u_y^{k,0})$,
- Update sharp high-frequency image
- While $(\delta > 10^{-4})$
  - $c = d$,
  - For $j = 0, \ldots, J-1$
    - $\beta_x = \kappa \|u_x^{k,j,0}\|_2$ and $\beta_y = \kappa \|u_y^{k,j,0}\|_2$
    - $u_x^{k,j,0} = u_x^{k,j}$ and $u_y^{k,j,0} = u_y^{k,j}$
    - For $\ell = 0, \ldots, L-1$
      - $\widetilde{u}_x^{k,j,\ell} = u_x^{k,j,\ell} - \delta H^*(h^k, u_x^{k,j,\ell})\left(H(h^k, u_x^{k,j,\ell}) - v_x\right)$
      - $u_x^{k,j+1,\ell} = \operatorname{prox}_{\delta\beta_x\|\cdot\|_1}\left(\widetilde{u}_x^{k,j,\ell}\right)$
      - $\widetilde{u}_y^{k,j,\ell} = u_y^{k,j,\ell} - \delta H^*(h^k, u_y^{k,j,\ell})\left(H(h^k, u_y^{k,j,\ell}) - v_y\right)$
      - $u_y^{k,j+1,\ell} = \operatorname{prox}_{\delta\beta_y\|\cdot\|_1}\left(\widetilde{u}_y^{k,j,\ell}\right)$
    - $u_x^{k,j+1} = u_x^{k,j,L}$ and $u_y^{k,j+1} = u_y^{k,j,L}$
  - $d = \frac{\kappa}{2}\left(\|H(h^k, u_x^{k,j+1}) - v_x\|_2^2 + \|H(h^k, u_y^{k,j+1}) - v_y\|_2^2\right) + \varphi_0(u_x^{k,j+1}) + \varphi_0(u_y^{k,j+1})$
  - If $(d > 3c)$
    - $u_x^{k,0} = u_x^k$ and $u_y^{k,0} = u_y^k$
    - $\delta = \frac{\delta}{2}$
  - Else
    - exit loop
- $u_x^{k+1} = u_x^{k,J}$ and $u_y^{k+1} = u_y^{k,J}$
- Update blurring matrix (Krishnan et al., 2011)
- $h^{k+1} = \underset{h \in D}{\text{minimize}}\ \frac{\kappa}{2}\left(\|H(h, u_x^{k+1}) - v_x\|_2^2 + \|H(h, u_y^{k+1}) - v_y\|_2^2\right) + \|h\|_1$

$\widehat{h} = h^K$

2. Non-blind image deconvolution

❷ $\widehat{x} = \underset{x \in \mathbb{R}^N}{\text{minimize}}\ \frac{\kappa}{2}\|H(\widehat{h}, x) - y\|_2^2 + \tau \left\| \begin{bmatrix} \nabla_x \\ \nabla_y \end{bmatrix} x \right\|_1$

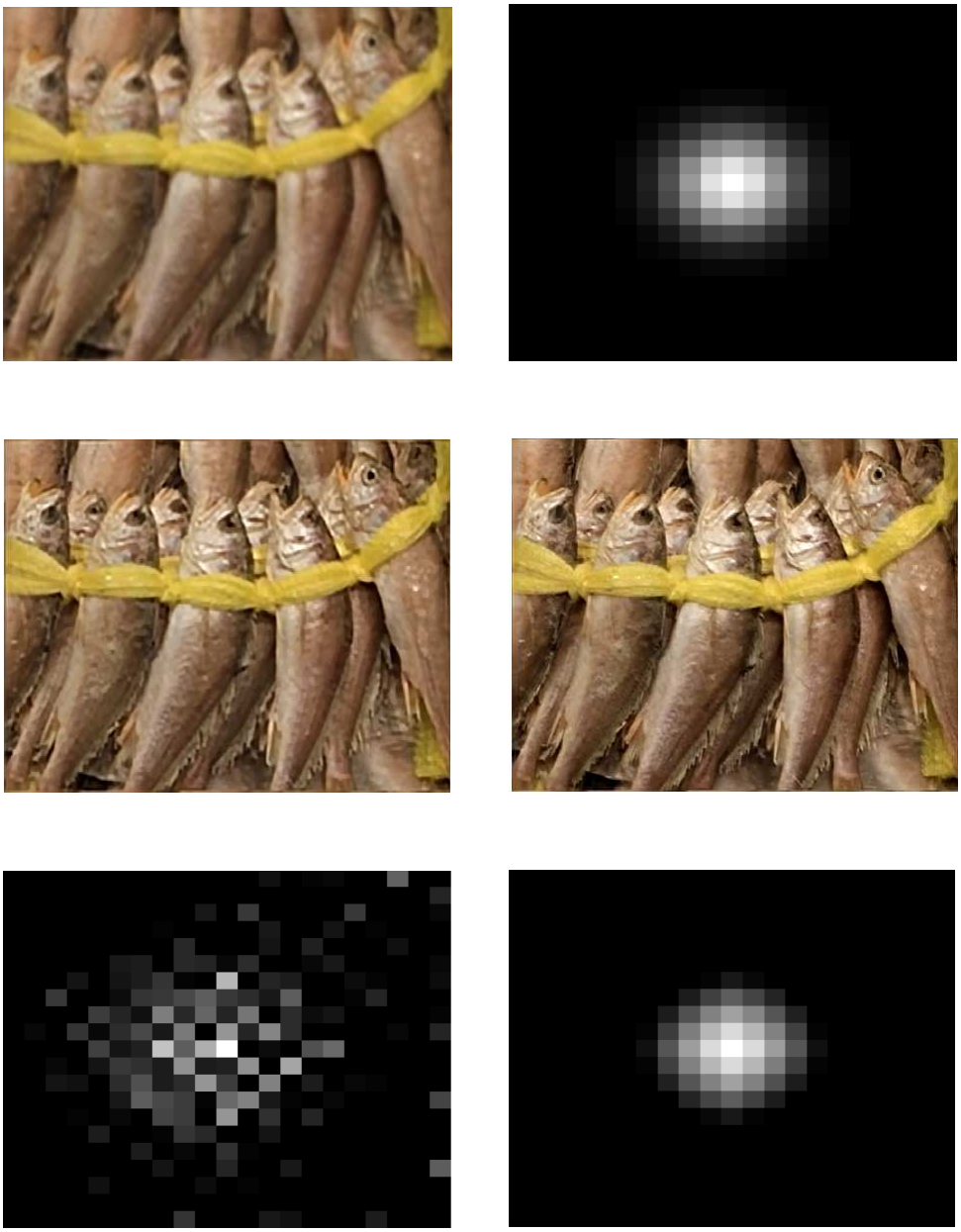


Figure 6.10: *Top-left: Input blurry image; top-right: original blur; middle-left: deblurred with Overall algorithm* **??** *(SNR* $= 20.97\,\mathrm{dB}$*) middle-right: deblurred with SOOT*^{2} *algorithm* **??** *(SNR* $= 22.80\,\mathrm{dB}$*); bottom-left: estimated blur with Overall algorithm* **??** *and bottom-right: estimated blur with SOOT*^{2} *algorithm* **??**.

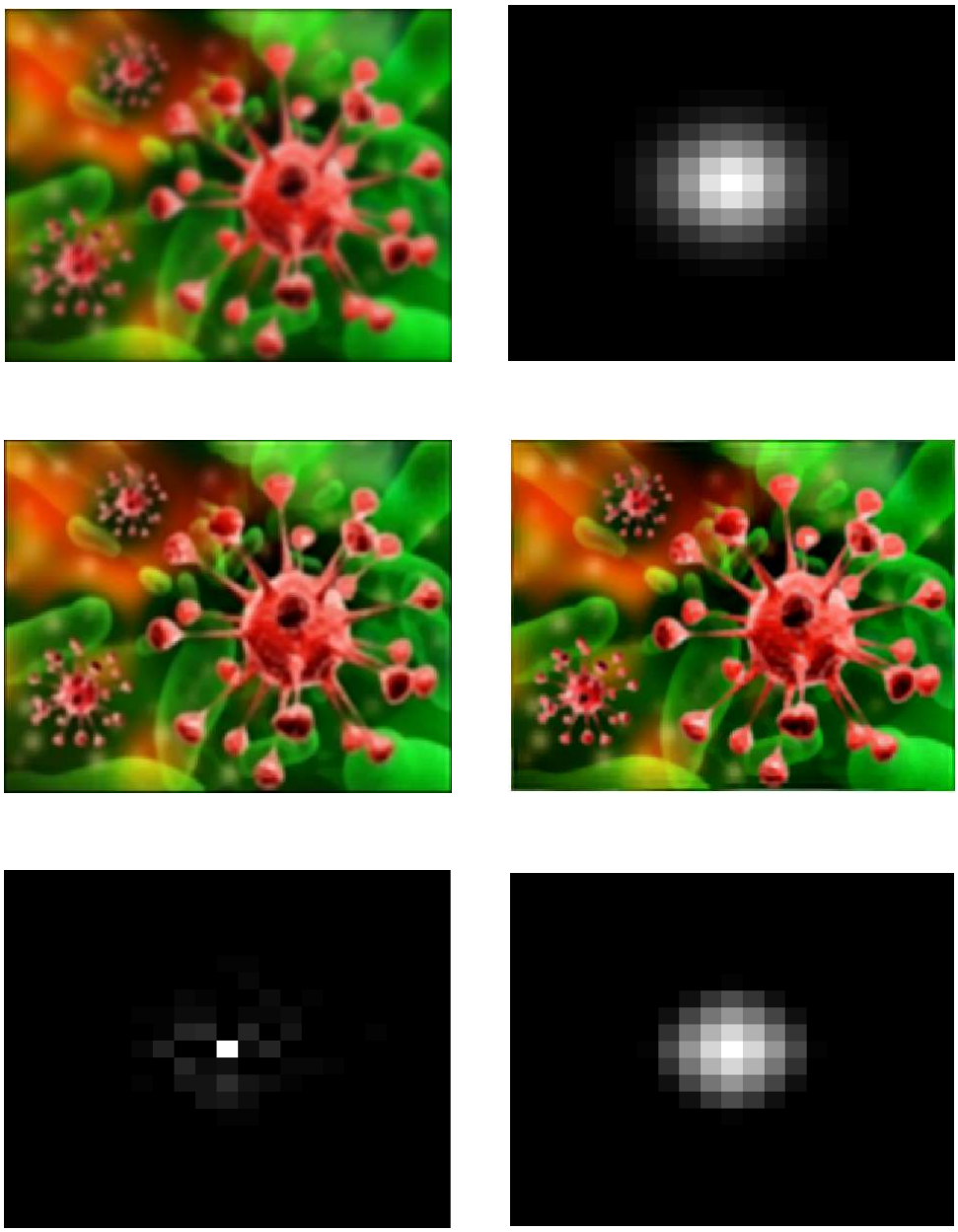


Figure 6.11: *Top-left: Input blurry image; top-right: original blur; middle-left: deblurred with Overall algorithm* **??** *(SNR* $= 15.20\,\mathrm{dB}$*); middle-right: deblurred with SOOT*^{2} *algorithm* **??** *(SNR* $= 16.06\,\mathrm{dB}$*); bottom-left: estimated blur with Overall algorithm* **??** *and bottom-right: estimated blur with SOOT*^{2} *algorithm* **??**.

Algorithm 9 SOOT^2 (2D SOOT adapted algorithm).

For every $k \in \mathbb{N}$, let $J_k \in \mathbb{N}^*$, $I_k \in \mathbb{N}^*$, let $(\kappa, \tau) \in]0, +\infty[^2$ and let $(\gamma_1^{k,j})_{0 \leq j \leq J_k - 1}$ and $(\gamma_2^{k,i})_{0 \leq i \leq I_k - 1}$ be positive sequences. Initialize with $u^0 \in [u_{\min}, u_{\max}]^{2N^2}$ and $h^0 \in D$. Explanations for the steps ❶ and ❷ can be found in the text.

1. Blind estimation of blur

❶ **Loop over coarse-to-fine levels**

Alternate:

$\delta^0 = 1,\ k = 0,$

While $\left(\delta^k > \sqrt{S} \times 10^{-6}\right)$

$u^{k,0} = u^k,\ h^{k,0} = h^k,$

Update sharp high-frequency image

For $j = 0, \ldots, J_k - 1$

$\widetilde{u}^{k,j} = u^{k,j} - \gamma_1^{k,j} A_1(u^{k,j}, h^k)^{-1} \nabla_1 f(u^{k,j}, h^k),$

$u^{k,j+1} = \Pi_{[u_{\min}, u_{\max}]^{2N^2}} \left(\widetilde{u}^{k,j}\right),$

$u^{k+1} = u^{k,J_k}$

Update blurring matrix

For $i = 0, \ldots, I_k - 1$

$\widetilde{h}^{k,i} = h^{k,i} - \gamma_2^{k,i} A_2(u^{k+1}, h^{k,i})^{-1} \nabla_2 f(u^{k+1}, h^{k,i})$

$h^{k,i+1} = \Pi_D \left(\widetilde{h}^{k,i}\right),$

$h^{k+1} = h^{k,I_k},$

$\delta^{k+1} = \|h^{k+1} - h^k\|,$

$k = k + 1.$

$\widehat{h} = h^k$

2. Non-blind image deconvolution

❷ $\widehat{x} = \underset{x \in \mathbb{R}^N}{\text{minimize}}\ \frac{\kappa}{2} \|H(\widehat{h}, x) - y\|_2^2 + \tau \left\| \begin{bmatrix} \nabla_x \\ \nabla_y \end{bmatrix} x \right\|_1$

- Chapter 7 -

Contributions and perspectives

This thesis focused mainly on two problems. In the first part, we investigated the problem of multiple removal in seismic data within the framework of convex optimization. In the second part, we provided a novel penalty based on the ℓ_1/ℓ_2 function leading to a nonconvex optimization problem and applied it to blind deconvolution of seismic data and images. Both can be seen as facets of source separation, more or less informed.

7.1 Contributions

Multiple removal: A geophysical 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. These templates are partly inaccurate — in delay, amplitude and frequency — and should be adapted in a time-varying fashion before being subtracted from the recorded data. We proposed a methodology for primary/multiple adaptive separation based on these approximate templates. This framework presented for the first time a relatively generic framework for multiple reflection filtering with:

A noise prior: By taking into account additional random noise into the generic model, we argue that our proposed method also effectively gets rid of more incoherent noise.

Sparsity constraints on signal frame coefficients: The proposed approach evaluated on seismic data using different orthogonal wavelet bases and tight frames, and various sparsity for wavelet coefficients. Especially, the priors anisotropic structure of 2D seismic data was taken into account by using the hybrid dual-tree M-band wavelets with different wavelets along space and time. The experiments on 2D seismic data showed that the choice of the transformation plays an important role.

Slow variation modeling of the adaptive filters: In the literature, no work in multiple removal has endeavored a more systematic study of the smooth variation of the adaptive filters. In this thesis, we proposed a formulation allowing a family of penalties to be applied to the adaptive FIR filters. More precisely, we expected the adapted filters to exhibit bounded variations from one time index to the next one.

Concentration metrics on the filters: The a priori information on the filter coefficients h can also be exploited. Since the energy of these coefficients is limited, the simple norms such as ℓ_1 or ℓ_2-norms or more sophisticated one such as a mixed $\ell_{1,2}$ or nuclear norms are used.

We 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 coefficient concentration constraints. 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 adaptive filter design. The **Bayesian framework** provided in this work could also serve to develop other statistical approaches for multiple removal, e.g. by using Markov Chain Monte-Carlo methods. The designed **primal-dual algorithm (M+LFBF)** solved a constrained minimization problem that alleviated standard regularization issues in finding hyperparameters. The approach demonstrated significantly good performance in low signal-to-noise ratio conditions, both for simulated and real field seismic data.

Blind deconvolution: In the context of blind deconvolution, a relatively common assumption is that the signal of interest has a sparse representation. According to (Donoho et al., 1995), an ideal measure of sparsity is the ℓ_0 measure. However, due to the nonconvexity of the ℓ_0 pseudo-norm, different alternative measures of sparsity have been proposed. Among them the ℓ_1 measure is the most popular, because it is convex. In this thesis we proposed a new penalty based on a smooth approximation to the ℓ_1/ℓ_2 function. This work was the result of a collaboration with Audrey Repetti and Emilie Chouzenoux. We developed a proximal-based algorithm to solve variational problems involving this function named **SOOT algorithm**. This algorithm is based on a new BC-VMFB algorithm for minimizing the sum of:

- a **nonconvex smooth** function,
- a **nonconvex non necessarily smooth** function.

We derived theoretical **convergence results** both on iterates and function values. The blocks are updated according to a flexible **quasi-cyclic** rule. The acceleration of the convergence is obtained thanks to the choice of matrices $(A_{j,\ell}(x_\ell))_{\ell\in\mathbb{N}}$ based

on the **MM principle**. The performance of the proposed method is demonstrated through a comparison with a recent alternating optimization strategy dealing with the exact ℓ_1/ℓ_2 term, on the applications, both on seismic data and image blind deconvolution.

7.2 Perspectives

Hyperplane constraints: Concerning the multiple removal problem, 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 (Ventosa et al., 2012; Song et al., 2013) 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 (Walden and Hosken, 1986) of the subsurface, sparsity promotion suggests the use of sparsifying transforms (e.g. wavelet/curvelet frames (Herrmann and Verschuur, 2004; Neelamani et al., 2010)), potentially combined with robust norms (approximate ℓ_1 in (Guitton and Verschuur, 2004)), quasi-norms or source separation methods (Donno, 2011; Duarte et al., 2012). 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 the context of this thesis, we incorporated 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 reformulated the previous approach as a constrained minimization problem. This allows us to more easily determine data-based parameters. The various constraint efficiency in wavelet frame subbands for the primary signal may be exploited. Convex sets defined as appropriate hyperplanes can outperform standard ℓ_1-ball constraints. The minimization problem (4.4) can be reexpressed as:

$$\underset{y\in\mathbb{R}^{N_t},\,\mathbf{h}\in\mathbb{R}^{N_tP}}{\text{minimize}}\; f\left(y,\mathbf{h}\right)+\iota_S\left(y+\mathbf{R}\mathbf{h}-z\right)+\iota_{C_1\cap C_2}(\mathbf{h})+\iota_D(Fy) \tag{7.1}$$

where $f\colon \mathbb{R}^{N_t}\times\mathbb{R}^{N_tP}\to\,]-\infty,+\infty] : (u,v)\mapsto\alpha\rho(v)+(1-\alpha)\phi(u)$, with $\alpha\in[0,1]$, $\phi:\mathbb{R}^{N_t}\to\,]-\infty,+\infty]$ and $\rho\colon\mathbb{R}^{N_tP}\to\,]-\infty,+\infty]$ are two convex functions.
$S=\left\{w\in\mathbb{R}^{N_t}\mid\|w\|^2\leq N_t\sigma^2\right\}$ and σ^2 is the variance of Gaussian noise b. C_1 and C_2 are defined in (4.6) and (4.7). As defined in 4.3.3.3 the convex set D can be splitted as $D_1\times\cdots\times D_{\mathcal{L}}$. However, another kind of convex sets $(D_l)_{l\in\{1,\ldots,\mathcal{L}\}}$ possibly more original consists of defining hyperplanes: for every $l\in\{1,\ldots,\mathcal{L}\}$, $D_l=\{(x_k)_{k\in\mathbb{K}_l}\mid\sum_{k\in\mathbb{K}_l}\varphi_l((FLz)_k)x_k=\beta_l\}$, where $L\in\mathbb{R}^{N_t\times N_t}$ is an appropriate

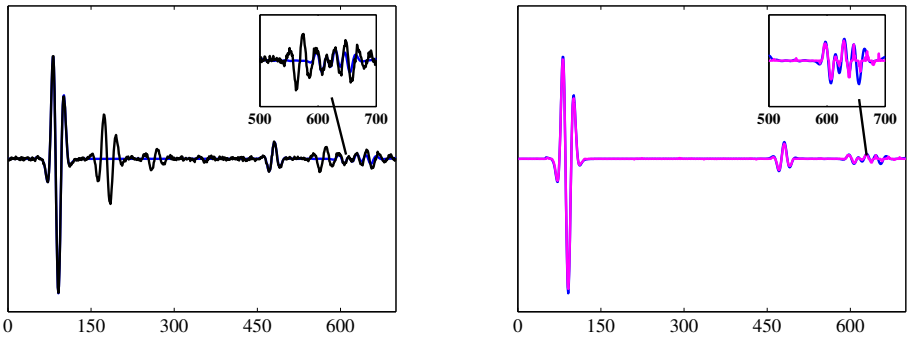


Figure 7.1: *Left: Observed signal z (black; $\sigma = 0.01$), original $\bar{y}$ (blue); Right: Estimated signal $\hat{y}$ (magenta), original signal $\bar{y}$ (blue). D is the intersection of two hyperplanes defined from the identity and the sign functions.*

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

$$L = \lambda_1 \operatorname{Diag}\left((1 + \lambda_1 + \lambda_2 \|\mathbf{R}^{(0)}\|^2)^{-1}, \ldots, (1 + \lambda_1 + \lambda_2 \|\mathbf{R}^{(N_t-1)}\|^2)^{-1}\right) \quad (7.2)$$

where $(\lambda_1, \lambda_2) \in]0, +\infty[^2$ and for every $n_t \in \{0, \ldots, N_t - 1\}$, $\mathbf{R}^{(n_t)}$ denotes the n_t-th row of matrix $\mathbf{R}$.
Now let us show some preliminary simulation tests on synthetic seismic data. Note that we use the same synthetic data as the one in Chapter 4. φ is chosen to be the ℓ_1-norm, and ρ is the squared ℓ_2-norm. 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 for D where, in the first case (inequality constraint), $\varphi_l \equiv \varphi$ where φ is either the ℓ_1, ℓ_2 or the ℓ_∞-norm. In the second case, $\varphi_l \equiv \varphi$ where φ is either the identity or the sign function; furthermore, L is chosen according to (7.2) 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+LFBF algorithm for the primary signal in the case when $\sigma = 0.01$, are displayed in Figure 7.1-right - p. 132. These results were published in (Pham et al., 2014a). From these the figures, one can note that the stochastic part is accurately removed, even if some residual noise remain for instance when the signal is of small amplitude. Table 7.1 - p. 133 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 $\varphi = 0$ 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.

These results allow us to integrate a large panel of hard constraints corresponding

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 7.1: *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" defined as in 4.3.3.3 and lower table part: hyperplane constraints.*

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. We now study the class of hyperplane constraints of interest as well as their inner parameters. In this context, the variance of Gaussian noise b is known. Further study would start from the experimental setting presented here. The first goal would be to **estimate the variance of Gaussian noise from the observed signal**. A possibility could be to use the method proposed in (Roberts et al., 1998). The second goal would be to **construct the wavelet constraints using a mix of statistical and convex-analytical tools**. This can be further improved by applying interpolation methods (Combettes and Pesquet, 2004; Chaux et al., 2008; Pesquet et al., 2009). The third goal could be to **extend the proposed method to two dimensional data** and then **compare with other algorithms** as (Chambolle and Pock, 2011; Vũ, 2013; Condat, 2013; Combettes and Pesquet, 2011).

Echo cancellation An echo is the repetition of a waveform due to reflection from points where the characteristics of the medium, through which the wave propagates, change. Echo cancellation is an important feature of communication systems. Figure 7.2 - p. 134 shows a concept of acoustic echo, while Figure 7.3 - p. 134 is a simple example of signal (sound) with/without echo. Assuming that the signal on the line from the speaker B to the speaker A at time $n_t \in]0, +\infty[$, denoted by $y_B(n_t)$ is composed of the speech of speaker B, $x_B(n_t)$, plus the echo of speaker A, $x_A^{\mathrm{echo}}(n_t)$, we have

$$\left(\forall n_t \in \mathbb{R}_+^*\right) \quad y_B(n_t) = x_B(n_t) + x_A^{\mathrm{echo}}(n_t).$$

When speaker A is talking, and speaker B is listening and silent, and only echo is present from line B to A, we have

$$y_B(n_t) = x_A^{\mathrm{echo}}(n_t).$$

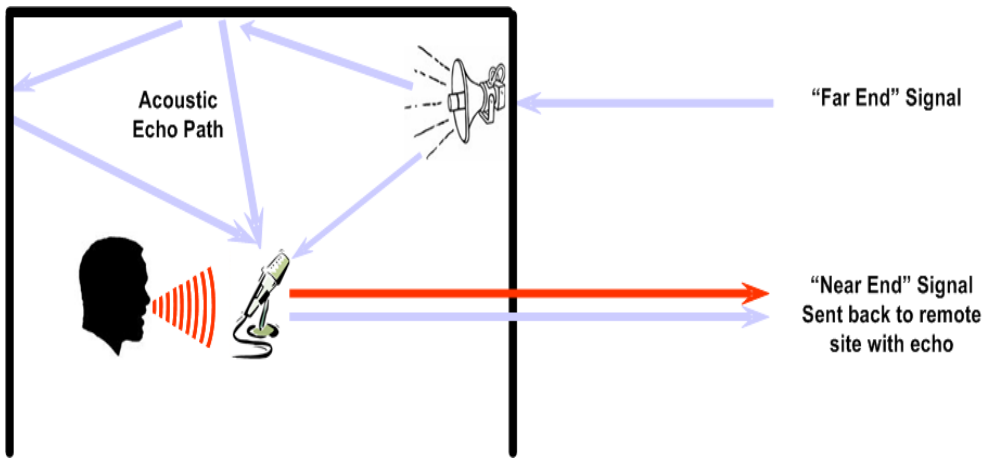


Figure 7.2: *Concept of acoustic echo.*

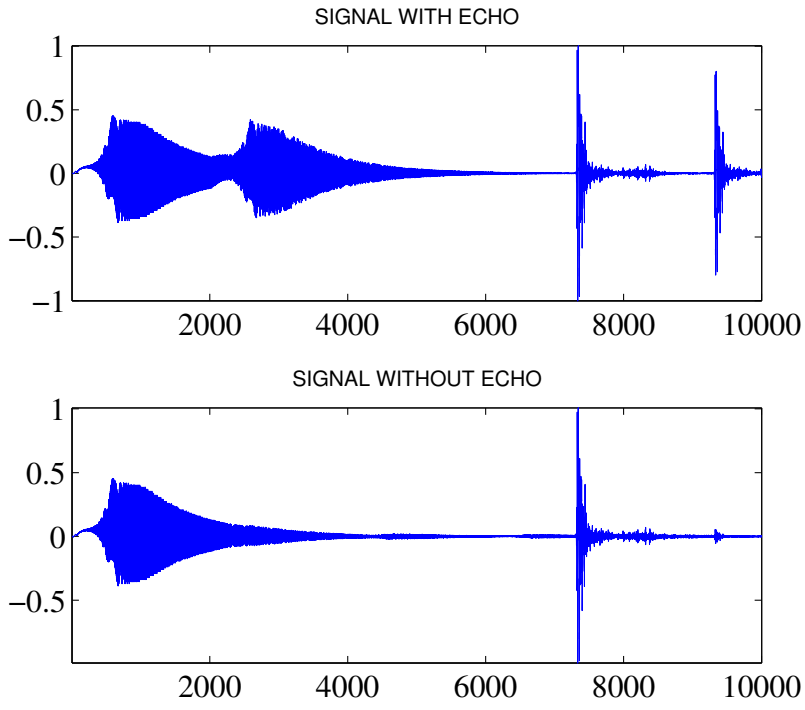


Figure 7.3: *From top to bottom: signal with echo and signal without echo.*

In (Vaseghi, 2006, Chapter 15), the authors supposed that the echo synthesizer is a FIR filter. The filter output estimate of the echo signal can be expressed as

$$\widehat{x}_A^{\text{echo}}(n_t) = \sum_{p=0}^{P-1} h_{n_t}(p) x_A(n_t - p)$$

where, $h_{n_t}(p)$ are the time-varying coefficients of an adaptive FIR filter, $\widehat{x}_A^{\text{echo}}(n_t)$ is an estimate of the echo of speaker A on the line from speaker B to speaker A, and $x_A(n_t)$ is the speech of speaker A. Then, they find the filter coefficients which are adapted to minimize the energy of an error signal defined as

$$\underset{h_{n_t}}{\text{minimize}} \, \|x_B(n_t) + x_A^{\text{echo}}(n_t) - \sum_{p=0}^{P-1} h_{n_t}(p) x_A(n_t - p)\|^2.$$

According to these assumptions, we consider a problem

$$y_B(n_t) = x_B(n_t) + x_A^{\text{echo}}(n_t) + b(n_t)$$

where $b(n_t)$ is an additional noise. Assuming that the echo of speaker A $x_A^{\text{echo}}(n_t)$ is related to his speech $x_A(n_t)$ through an FIR filter

$$x_A^{\text{echo}}(n_t) = \sum_{p=0}^{P-1} \overline{h}_{n_t}(p) x_A(n_t - p)$$

where $\overline{h}_{n_t}(p)$ is an unknown impulse response. A main difference with our approach is however that echo cancellation in acoustic applications must be performed in real-time based on adaptive filtering techniques. Then this problem appears to be related to multiple removal with a strong emphasis on online estimation (Mossi et al., 2011; Joo, 2014; Moller and Heusdens, 2013; Lin and Lee, 2005).

Image blind deconvolution It is well known that, in seismic blind deconvolution, the reflectivities themselves are considered very sparse. Thus, the smooth approximation ℓ_1/ℓ_2 can be directly used on the signal as a sparseness measure. However, in the context of image restoration, the image often has a sparse representation with respect to preassigned operator. For instance, our proposed method for image blind deconvolution assumes that the image admits a sparse gradient prior and then, the blind kernel estimation step is performed on the high frequencies of the image. In future work we will extend the proposed method to a more sophisticated linear operator (e.g. a basis or frame...) for which, the sparsity is not directly linked to the image (pixels) but to a representation of it (coefficients). The question raised is, how to find the quadratic majorant of $\varphi(Lx)$, where $\varphi(\cdot) = \lambda \log\left(\frac{\ell_{1,\alpha}(\cdot)+\beta}{\ell_{2,\eta}(\cdot)}\right)$ and L is a linear operator from $\mathbb{R}^N$ to $\mathbb{R}^M$? The following simple proposition shows the relationship between the quadratic majorants of φ and $\varphi \circ L$.

Proposition 7.2.1 *Let $\zeta : \mathbb{R}^M \to \mathbb{R}$ be a differentiable function. Let $z \in \mathbb{R}^M$. For every $z' \in \mathbb{R}^M$ we have a SPD matrix $U(z) \in \mathbb{R}^{M\times M}$ such that*

$$q(z', z) = \zeta(z) + (z' - z)^\top \nabla\zeta(z) + \frac{1}{2}\|z' - z\|^2_{U(z)},$$

$q(z', z)$ is a quadratic majorant of the function ζ. Then,

$$\widetilde{q}(z', z) = \widetilde{\zeta}(z) + (z' - z)^\top \nabla\widetilde{\zeta}(z) + \frac{1}{2}\|z' - z\|^2_{\widetilde{U}(z)},$$

where $\widetilde{U}(z) = L^\top U(z)L$ is a quadratic majorant of the function $\widetilde{\zeta} = \zeta \circ L$.

Sparseness measure ℓ_p/ℓ_q $(p \leq 1 < q)$ One can find in the literature various sparsity measures. In (Hurley and Rickard, 2009) several commonly-used sparsity measures are compared based on six intuitive attributes (Robin Hood, Scaling, Rising Tide, Cloning, Bill Gates, and Babies). This paper shows that the $pq-$mean measure (or ℓ_p/ℓ_q, with $p \leq 1 < q$) satisfies all six criteria. The ℓ_1/ℓ_2 function is one of $pq-$mean cases. In future work we want to extend our SOOT algorithm to a general $pq-$mean measure.

$$\varphi(x) = \lambda \log\left(\frac{\ell_{p,\alpha}(x)}{\ell_{q,\eta}(x)}\right) \tag{7.3}$$

where, $p < q$, $(\lambda, \alpha, \eta) \in]0, +\infty[^3$, $\alpha \leq \eta$ and $\ell_{p,\alpha}(x)$, $\ell_{q,\eta}(x)$ are defined as follows:

$$\ell_{p,\alpha}(x) = \left(\sum_{n=1}^{N}(x_n^2 + \alpha^2)^{p/2}\right)^{1/p} \tag{7.4}$$

and

$$\ell_{q,\eta}(x) = \left(\sum_{n=1}^{N}(x_n^2 + \eta^2)^{q/2}\right)^{1/q} \tag{7.5}$$

Let $(p, q) \in]0, +\infty[^2$ such that, $1 \leq p < q$, one can prove that, for every $0 < \alpha \leq \eta$

$$\ell_{p,\alpha}(x) \leq \vartheta\, \ell_{q,\eta}(x) \tag{7.6}$$

where $\vartheta = N^{1/p-1/q}$.

Image shadow removal In image processing, image shadow removal is an important topic. Shadows are generated by a local and relative absence of light and they are caused by a local decrease in the amount of light that reaches a surface or by a local change in the amount of light rejected by a surface toward the observer. Insights into this problem may apply to other vision problems as well (such as segmentation, enhancement,...) (Finlayson et al., 2006). According to (Tappen et al., 2005) an observed image $y \in \mathbb{R}^{N\times M}$ is a product of two images: one illumination image

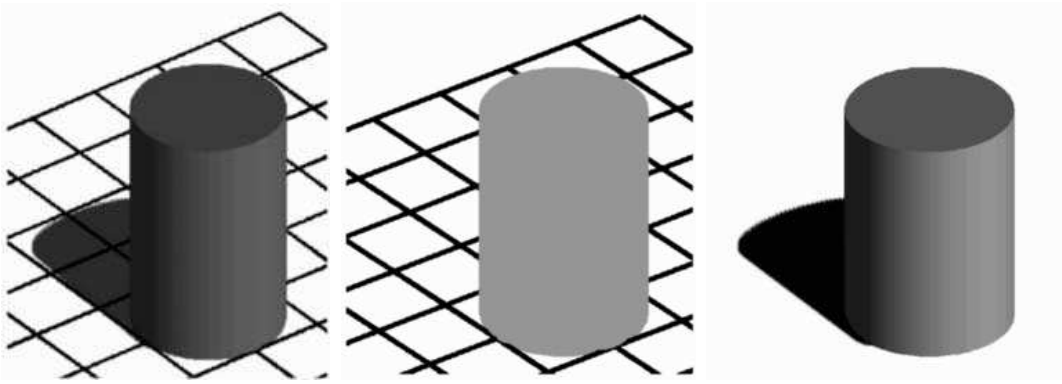


Figure 7.4: *From left to right: observed image y, reflectance image r and illumination image x.*

and one reflectance image. Thus, for every $(i,j) \in \{1,\cdots,N\} \times \{1,\cdots,M\}$, let $y(i,j)$ be the observed image, $r(i,j)$ be the reflectance image and $x(i,j)$ be the illumination image, these three images are related by the following equation

$$y(i,j) = r(i,j)x(i,j). \tag{7.7}$$

Note that, when $x(i,j) = 1$ then $y(i,j) = r(i,j)$. Figure 7.4 - p. 137 (Weiss, 2001) illustrates the decomposition of an observed image into a product of reflectance image and illumination image. One can suppose that the values of these images are positive, thus this equation can be rewritten in the log domain as follows

$$Y = R + X \tag{7.8}$$

where, $Y(i,j) = \log(y(i,i))$, $X(i,j) = \log(x(i,i))$ and $R(i,j) = \log(r(i,i))$. Normally, the shadows appear in the illumination image, assuming that $X = \overline{X} + S$ where S is the shadow of $\overline{X}$ and related to $\overline{X}$ by a FIR filter h i.e. $S = h * \overline{X}$. We now have the problem

$$Y = R + H\overline{X} \tag{7.9}$$

where, $H : \mathbb{R}^N \to \mathbb{R}^N : \overline{X} \mapsto \overline{X} + h * \overline{X}$. The FIR filter h can be related to the direction of light-source, position of camera, etc. If h is known, then we have a problem related to multiple removal, otherwise, the problem can be seen as a mixture of multiple removal and "blind" deconvolution. Figure 7.5 - p. 138 shows a simple example on the "multiple" removal image, where Barbara image is degraded by the head of Lena. We adapt our code for multiple removal problem in seismic data to recover Barbara image. Note that, the model of Lena's head is supposed to be known.

Figure 7.5: *From top to bottom: Observed image z (with Lena's head inclusion), original image y and estimated image $\hat{y}$ with the M+LFBF algorithm 5 (on the left are the full versions and on the right are their zoomed versions).*

- Appendix A -

Auxiliary proofs

A.1 Proof of Proposition ??

The proof of this result is equivalent to prove that the following expression is positive:

$$
\begin{aligned}
&\ell_{2,\eta}^2(x) - 2\left(\alpha + \frac{\eta}{\sqrt{N}}\right)\ell_{1,\alpha}(x) \\
=&\sum_{n=1}^{N} x_n^2 + \eta^2 - 2\left(\alpha + \frac{\eta}{\sqrt{N}}\right)\sum_{n=1}^{N}\left(\sqrt{x_n^2+\alpha^2} - \alpha\right) \\
=&\sum_{n=1}^{N}\left(x_n^2 + \frac{\eta^2}{N} - 2\left(\alpha + \frac{\eta}{\sqrt{N}}\right)\left(\sqrt{x_n^2+\alpha^2} - \alpha\right)\right) \\
=&\sum_{n=1}^{N}\left((x_n^2+\alpha^2) - 2\left(\alpha + \frac{\eta}{\sqrt{N}}\right)\left(\sqrt{x_n^2+\alpha^2}\right) + 2\alpha\left(\alpha + \frac{\eta}{\sqrt{N}}\right) - \alpha^2 + \frac{\eta^2}{N}\right) \\
=&\sum_{n=1}^{N}\left((x_n^2+\alpha^2) - 2\left(\alpha + \frac{\eta}{\sqrt{N}}\right)\left(\sqrt{x_n^2+\alpha^2}\right) + \left(\alpha + \frac{\eta}{\sqrt{N}}\right)^2\right) \\
=&\sum_{n=1}^{N}\left(\sqrt{x_n^2+\alpha^2} - \alpha - \frac{\eta}{\sqrt{N}}\right)^2 \\
\geq&0.
\end{aligned}
$$

The minimum is reached when $x_n = \sqrt{\frac{\eta^2}{N} + 2\frac{\alpha\eta}{\sqrt{N}}}$ or $x_n = -\sqrt{\frac{\eta^2}{N} + 2\frac{\alpha\eta}{\sqrt{N}}}$ for every $n \in \{1, \ldots, N\}$. ■

A.2 Proof of Proposition ??

To simplify some expressions, let us denote

$$u(x) = \Big(\frac{x_1}{(x_1^2+\alpha^2)^{1/2}}, \frac{x_2}{(x_2^2+\alpha^2)^{1/2}}, \cdots, \frac{x_N}{(x_N^2+\alpha^2)^{1/2}}\Big)^\top.$$

Now, we calculate the gradient and the Hessian of $\ell_{1,\alpha}(x)$ and $\ell_{2,\eta}(x)$,

$$\nabla \ell_{1,\alpha}(x) = \left(\frac{x_1}{(x_1^2+\alpha^2)^{1/2}}, \cdots, \frac{x_N}{(x_N^2+\alpha^2)^{1/2}}\right)^\top = u(x) \tag{A.1}$$

$$\nabla^2 \ell_{1,\alpha}(x) = \operatorname{Diag}\left(\frac{\alpha^2}{(x_1^2+\alpha^2)^{3/2}}, \cdots, \frac{\alpha^2}{(x_N^2+\alpha^2)^{3/2}}\right) \tag{A.2}$$

$$\nabla \ell_{2,\eta}(x) = \left(\frac{x_1}{\left(\sum_{n=1}^N x_n^2+\eta^2\right)^{1/2}}, \cdots, \frac{x_N}{\left(\sum_{n=1}^N x_n^2+\eta^2\right)^{1/2}}\right) = \frac{1}{\ell_{2,\eta}(x)} x \tag{A.3}$$

$$\begin{aligned}\nabla^2 \ell_{2,\eta}(x) &= \frac{1}{\ell_{2,\eta}(x)}\,\mathrm{I}_N - \frac{1}{\ell_{2,\eta}(x)^2} x \nabla l_{2,\eta}(x)^\top \\ &= \frac{1}{\ell_{2,\eta}(x)}\,\mathrm{I}_N - \frac{1}{\ell_{2,\eta}(x)^3} x x^\top.\end{aligned} \tag{A.4}$$

From (A.1), (A.2), (A.3), and (A.4), we obtain the gradient of $\widetilde{\varphi}(x)$,

$$\begin{aligned}\nabla \widetilde{\varphi}(x) &= \frac{1}{\ell_{2,\eta}(x)} \nabla \ell_{1,\alpha}(x) - \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^2} \nabla \ell_{2,\eta}(x) \\ &= \frac{1}{\ell_{2,\eta}(x)} u(x) - \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^3} x\end{aligned} \tag{A.5}$$

and its Hessian is

$$\begin{aligned}\nabla^2 \widetilde{\varphi}(x) =& \frac{1}{\ell_{2,\eta}(x)} \nabla^2 \ell_{1,\alpha}(x) - \frac{1}{\ell_{2,\eta}(x)^2} u(x) \nabla \ell_{2,\eta}(x)^\top \\ & - \frac{1}{\ell_{2,\eta}(x)^3} \Big(x \nabla \ell_{1,\alpha}(x)^\top + \ell_{1,\alpha}(x)\,\mathrm{I}_N\Big) + 3 \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^4} x \nabla \ell_{2,\eta}(x)^\top \\ =& \frac{1}{\ell_{2,\eta}(x)} \nabla^2 \ell_{1,\alpha}(x) - \frac{1}{\ell_{2,\eta}(x)^3} u(x) x^\top \\ & - \frac{1}{\ell_{2,\eta}(x)^3} x u(x)^\top - \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^3}\,\mathrm{I}_N + 3 \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^5} x x^\top \\ =& \frac{1}{\ell_{2,\eta}(x)} \nabla^2 \ell_{1,\alpha}(x) - \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^3}\,\mathrm{I}_N \\ & - \frac{1}{\ell_{2,\eta}(x)^3} \Big(u(x) x^\top + x u(x)^\top\Big) + 3 \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^5} x x^\top.\end{aligned}$$

Then, the Hessian of $\widetilde{\varphi}(x)$ can be compactly written as follows

$$\nabla^2\widetilde{\varphi}(x) = D(x) + z(x)x^\top + xz(x)^\top \tag{A.6}$$

where

$$D(x) = \frac{1}{\ell_{2,\eta}(x)} \operatorname{Diag}\left(\left(\frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}} - \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^2}\right)_{1\le n\le N}\right)$$

and

$$z(x) = \frac{3}{2}\frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^5}x - \frac{1}{\ell_{2,\eta}(x)^3}u(x).$$

By the triangle inequality we can deduce that

$$\begin{aligned}\left|\!\left|\!\left|\nabla^2\widetilde{\varphi}(x)\right|\!\right|\!\right| &= \left|\!\left|\!\left|D(x) + z(x)x^\top + xz(x)^\top\right|\!\right|\!\right| \\ &\le |\!|\!|D(x)|\!|\!| + \left|\!\left|\!\left|z(x)x^\top + xz(x)^\top\right|\!\right|\!\right|. \end{aligned}\tag{A.7}$$

We now turn our attention to the first term of (A.7), as $D(x)$ is a diagonal matrix, it follows that

$$|\!|\!|D(x)|\!|\!| = \frac{1}{\ell_{2,\eta}(x)}\max_{1\le n\le N}\left|\frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}} - \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^2}\right|.$$

Using Proposition **??**. we have

$$0 < \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^2} \le \frac{1}{2(\alpha+\frac{\eta}{\sqrt{N}})}.$$

Moreover

$$0 < \frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}} \le \frac{1}{\alpha}.$$

On the one hand,

$$|\!|\!|D(x)|\!|\!| \le \frac{1}{\eta}\max\left\{\frac{1}{\alpha}, \frac{1}{2(\alpha+\frac{\eta}{\sqrt{N}})}\right\} = \frac{1}{\eta\alpha}. \tag{A.8}$$

On the other hand, the last term of (A.7) is

$$\begin{aligned}\left|\!\left|\!\left|z(x)x^\top + xz(x)^\top\right|\!\right|\!\right|^2 &\le \left(\left|\!\left|\!\left|z(x)x^\top\right|\!\right|\!\right| + \left|\!\left|\!\left|xz(x)^\top\right|\!\right|\!\right|\right)^2 \\ &\le \left(\|z(x)\|\|x\| + \|x\|\|z(x)\|\right)^2 \\ &= 4\|x\|^2\|z(x)\|^2 \\ &= 4\|x\|^2\sum_{n=1}^N\left(\frac{3}{2}\frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^5}x_n - \frac{1}{\ell_{2,\eta}(x)^3}\frac{x_n}{(x_n^2+\alpha^2)^{1/2}}\right)^2 \\ &= 4\frac{\|x\|^2}{\ell_{2,\eta}(x)^6}\sum_{n=1}^N x_n^2\left(\frac{3}{2}\frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^2} - \frac{1}{(x_n^2+\alpha^2)^{1/2}}\right)^2.\end{aligned}$$

For every $x \in \mathbb{R}^N$ we have, $0 < \frac{1}{\sqrt{x_n^2+\alpha^2}} \leq \frac{1}{\alpha}$ and $0 < \frac{\ell_{1,\alpha}(x)}{\ell_{2,\eta}(x)^2} \leq \frac{1}{2(\alpha+\frac{\eta}{\sqrt{N}})}$ (using Proposition **??**). Then

$$\begin{aligned}\left|\!\left|\!\left| z(x)x^\top + xz(x)^\top \right|\!\right|\!\right|^2 &\leq 4\frac{\|x\|^4}{\ell_{2,\eta}(x)^6}\max\left\{\frac{1}{\alpha^2}, \frac{9}{4}\left(\frac{1}{2(\alpha+\frac{\eta}{\sqrt{N}})}\right)^2\right\} \\ &= 4\frac{\|x\|^4}{\left(\|x\|^2+\eta^2\right)^3}\frac{1}{\alpha^2} = \theta(\|x\|^2),\end{aligned}$$

where $\theta : u \in [0, +\infty[\mapsto \frac{4}{\alpha^2}\frac{u^2}{(u+\eta^2)^3}$. The derivative of θ is given, for every $u \in [0, +\infty[$, by

$$\dot{\theta}(u) = \frac{4}{\alpha^2}\frac{u(2\eta^2-u)}{(u+\eta^2)^4},$$

thus θ is an increasing function on $[0,\, 2\eta^2]$ and a decreasing function on $[2\eta^2,\, +\infty[$, and

$$\max_{u\in[0,+\infty[} \theta(u) = \theta(2\eta^2) = \frac{16}{27\alpha^2\eta^2}. \tag{A.9}$$

Finally, from (A.9) and (A.8) we then deduce that, for every $x \in \mathbb{R}^N$

$$\begin{aligned}\left|\!\left|\!\left| \nabla^2\widetilde{\varphi}(x) \right|\!\right|\!\right| &\leq \left|\!\left|\!\left| D(x) \right|\!\right|\!\right| + \left|\!\left|\!\left| z(x)x^\top + xz(x)^\top \right|\!\right|\!\right| \\ &\leq \frac{1}{\alpha\eta} + \frac{4}{\sqrt{27}\alpha\eta} \\ &= \left(1+\frac{4}{3\sqrt{3}}\right)\frac{1}{\alpha\eta}\end{aligned} \tag{A.10}$$

which leads to the conclusion. ■

A.3 Proof of Proposition **??**

Firstly let us calculate the gradient and Hessian of $\ell_{2,\eta}(x)^2$,

$$\nabla\ell_{2,\eta}(x)^2 = 2x \tag{A.11}$$

$$\nabla^2\ell_{2,\eta}(x)^2 = 2\,\mathrm{I}_N\,. \tag{A.12}$$

By definition of function φ, we have

$$\varphi(x) = \lambda\log\left(\frac{\ell_{1,\alpha}(x)+\beta}{\ell_{2,\eta}(x)}\right) = \lambda\left(\log(\ell_{1,\alpha}(x)+\beta) - \frac{1}{2}\log\ell_{2,\eta}(x)^2\right).$$

We now compute the gradient and Hessian of $\varphi(x)$

$$\nabla\varphi(x) = \lambda\frac{\nabla\ell_{1,\alpha}(x)}{\ell_{1,\alpha}(x)+\beta} - \frac{\lambda}{2}\frac{\nabla\ell_{2,\eta}(x)^2}{\ell_{2,\eta}(x)^2} \tag{A.13}$$

$$\begin{aligned}\nabla^2\varphi(x) =& \frac{\lambda}{(\ell_{1,\alpha}(x)+\beta)^2}\left(\nabla^2\ell_{1,\alpha}(x)\,(\ell_{1,\alpha}(x)+\beta) - \nabla\ell_{1,\alpha}(x)\nabla\ell_{1,\alpha}(x)^\top\right)\\ &- \frac{\lambda}{2\ell_{2,\eta}(x)^4}\left(\nabla^2\ell_{2,\eta}(x)^2\ell_{2,\eta}(x)^2 - \nabla\ell_{2,\eta}(x)^2\nabla(\ell_{2,\eta}(x)^2)^\top\right).\end{aligned} \tag{A.14}$$

Using Equations (A.1), (A.2), (A.11), and (A.12), we obtain

$$\begin{aligned}\nabla^2\varphi(x) =& \frac{\lambda}{\ell_{1,\alpha}(x)+\beta}\operatorname{Diag}\left(\left(\frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}}\right)_{1\le n\le N}\right) - \frac{\lambda}{\ell_{2,\eta}(x)^2}\,\mathrm{I}_N\\ &- \frac{\lambda}{(\ell_{1,\alpha}(x)+\beta)^2}u(x)u(x)^\top + \frac{2\lambda}{\ell_{2,\eta}(x)^4}xx^\top.\end{aligned} \tag{A.15}$$

Thus, using the triangle inequality, we can deduce that

$$\begin{aligned}|||\nabla^2\varphi(x)||| \le& \left|\left|\left|\frac{\lambda}{\ell_{1,\alpha}(x)+\beta}\operatorname{Diag}\left(\left(\frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}}\right)_{1\le n\le N}\right) - \frac{\lambda}{\ell_{2,\eta}(x)^2}\,\mathrm{I}_N\right|\right|\right|\\ &+ \frac{\lambda}{(\ell_{1,\alpha}(x)+\beta)^2}\|u(x)\|^2 + \frac{2\lambda}{\ell_{2,\eta}(x)^4}\|x\|^2\\ =&\lambda\max_{1\le n\le N}\left|\frac{1}{\sum_{n=1}^N\left(\sqrt{x_n^2+\alpha^2}-\alpha\right)+\beta}\frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}} - \frac{1}{\sum_{n=1}^N x_n^2+\eta^2}\right|\\ &+ \frac{\lambda}{\left(\sum_{n=1}^N(\sqrt{x_n^2+\alpha^2}-\alpha)+\beta\right)^2}\sum_{n=1}^N\frac{x_n^2}{x_n^2+\alpha^2} + 2\lambda\frac{\sum_{n=1}^N x_n^2}{\left(\sum_{n=1}^N x_n^2+\eta\right)^2}.\end{aligned} \tag{A.16}$$

We need to find an upper bound for (A.16). First observe that, for every $1 \le n \le N$,

$$\begin{aligned}\frac{x_n^2}{x_n^2+\alpha^2} &= \frac{\left(\sqrt{x_n^2+\alpha^2}-\alpha\right)\left(\sqrt{x_n^2+\alpha^2}+\alpha\right)}{x_n^2+\alpha^2}\\ &= \left(\sqrt{x_n^2+\alpha^2}-\alpha\right)\frac{\left(\sqrt{x_n^2+\alpha^2}+\alpha\right)}{x_n^2+\alpha^2}.\end{aligned}$$

Since for every $n \in \{1,\dots,N\}$, $\left(\sqrt{x_n^2+\alpha^2}-\alpha\right) \geq 0$, $0 < \frac{\left(\sqrt{x_n^2+\alpha^2}+\alpha\right)}{x_n^2+\alpha^2} \leq \frac{2}{\alpha}$ and $0 < \frac{1}{\left(\sum_{n=1}^N\left(\sqrt{x_n^2+\alpha^2}-\alpha\right)+\beta\right)^2} \leq \frac{1}{\beta^2}$, then

$$\frac{1}{\left(\sum_{n=1}^N(\sqrt{x_n^2+\alpha^2}-\alpha)+\beta\right)^2}\sum_{n=1}^N\frac{x_n^2}{x_n^2+\alpha^2} \leq \frac{2}{\alpha}\frac{\sum_{n=1}^N\left(\sqrt{x_n^2+\alpha^2}-\alpha\right)}{\left(\sum_{n=1}^N(\sqrt{x_n^2+\alpha^2}-\alpha)+\beta\right)^2} \leq \frac{2}{\alpha}\frac{1}{4\beta} = \frac{1}{2\alpha\beta}. \tag{A.17}$$

Moreover, $0 < \frac{1}{\sum_{n=1}^N x_n^2+\eta^2} \leq \frac{1}{\eta^2}$ and $0 < \frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}} \leq \frac{1}{\alpha}$ for every $n \in \{1,\dots,N\}$, thus

$$\max_{1\leq n\leq N}\left|\frac{1}{\sum_{n=1}^N\left(\sqrt{x_n^2+\alpha^2}-\alpha\right)+\beta}\frac{\alpha^2}{(x_n^2+\alpha^2)^{3/2}} - \frac{1}{\sum_{n=1}^N x_n^2+\eta^2}\right| \leq \max\left\{\frac{1}{\beta\alpha}, \frac{1}{\eta^2}\right\} \tag{A.18}$$

and

$$0 \leq \frac{2\sum_{n=1}^N x_n^2}{\left(\sum_{n=1}^N x_n^2+\eta^2\right)^2} \leq \frac{1}{2\eta^2}. \tag{A.19}$$

Finally, from (A.17), (A.18), and (A.19), we also have

$$|||\nabla^2\varphi(x)||| \leq \lambda\left(\max\left\{\frac{1}{\beta\alpha}, \frac{1}{\eta^2}\right\} + \frac{1}{2\alpha\beta} + \frac{1}{2\eta^2}\right) \tag{A.20}$$

which yields the conclusion. ■

List of Figures

List of Tables

Bibliography

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

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

Ahmed, A., Recht, B., and Romberg, J. (2014). Blind deconvolution using convex programming. *IEEE Trans. Inform. Theory*, 60(3):1711–1732.

Ahmed, I. (2007). 2D wavelet transform-domain adaptive subtraction for enhancing 3D SRME. In *Annual International Meeting*, volume 26, pages 2490–2494. Soc. Expl. Geophysicists.

Alghamdi, M. A., Alotaibi, A., Combettes, P. L., and Shahzad, N. (2014). A primal-dual method of partial inverses for composite inclusions. *Optim. Lett.*, 8:2271–2284.

Allain, M., Idier, J., and Goussard, Y. (2006). On global and local convergence of half-quadratic algorithms. *IEEE Trans. Image Process.*, 15(5):1130–1142.

Andersson, F. and Duchkov, A. A. (2013). Extended structure tensors for multiple directionality estimation. *Geophys. Prospect.*, 61(6):1135–1149.

Attouch, H. and Bolte, J. (2010). Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457.

Attouch, H., Bolte, J., and Svaiter, B. F. (2013). Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Progr. (Ser. A)*, 137(1-2):91–129.

Aujol, J.-F., Aubert, G., and Blanc-Féraud, L. (2003). Wavelet-based level set evolution for classification of textured images. *IEEE Trans. Image Process.*, 12(12):1634–1641.

Barak, B., Kelner, J., and Steurer, D. (2014). Rounding sum-of-squares relaxations. In *Proc. ACM Symp. Theo. Comput. (STOC)*, New York, NY, USA.

Barnes, A. E. (1996). Theory of 2-D complex seismic trace analysis. *Geophysics*, 61(1):264–272.

Baumstein, A. and Hadidi, M. T. (2006). 3D surface-related multiple elimination: data reconstruction and application to field data. *Geophysics*, 71(3):E25–E33.

Bauschke, H. H. and Combettes, P. L. (2011). *Convex analysis and monotone operator theory in Hilbert spaces*. CMS books in mathematics. Springer.

Beasley, C. J., Dragoset, B., and Salama, A. (2012). A 3D simultaneous source field test processed using alternating projections: a new active separation method. *Geophys. Prospect.*, 60(4):591–601.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202.

Benichoux, A., Vincent, E., and Gribonval, R. (2013). A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, pages 6108–6112, Vancouver, BC, Canada.

Berkhout, A. J. and Verschuur, D. J. (2006). Focal transformation, an imaging concept for signal restoration and noise removal. *Geophysics*, 71(6):A55–A59.

Bierstone, E. and Milman, P. D. (1988). Semianalytic and subanalytic sets. *Inst. Hautes Études Sci. Publ. Math.*, 67(1):5–42.

Bioucas-Dias, J. and Figueiredo, M. (2007). A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.*, 16(12):2992–3004.

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

Bolte, J., Combettes, P. L., and Pesquet, J.-C. (2010). Alternating proximal algorithm for blind image recovery. In *Proc. Int. Conf. Image Process.*, pages 1673–1676, Hong-Kong, China.

Bolte, J., Daniilidis, A., and Lewis, A. (2007). The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.*, 17(4):1205–1223.

Bolte, J., Sabach, S., and Teboulle, M. (2014). Proximal alternating linearized minimization fon nonconvex and nonsmooth problems. *Math. Programm.*, 146(1-2):459–494.

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

Brittan, J., Martin, T., Bekara, M., and Koch, K. (2011). 3D shallow water demultiple — extending the concept. *First Break*, 29(9):97–101.

Buttkus, B. (1975). Homomorphic filtering — theory and practice. *Geophys. Prospect.*, 23(4):712–748.

Cai, J.-F., Candès, E. J., and Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20(4):1956–1982.

Campisi, P. and Egiazarian, K., editors (2007). *Blind Image Deconvolution: Theory and Applications*. CRC Press.

Candès, E. J., Li, X., Ma, Y., and Wright, J. (2011). Robust principal component analysis? 58(3):1–37.

Chaâri, L., Pustelnik, N., Chaux, C., and Pesquet, J.-C. (2009). Solving inverse problems with overcomplete transforms and convex optimization techniques. In Goyal, V. K., Papadakis, M., and Van De Ville, D., editors, *Proc. SPIE, Wavelets*, volume 7446, San Diego, CA, USA.

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

Chaux, C. and Blanc-Féraud, L. (2012). Wavelet-based hyperparameter estimation for solving inverse problems. Technical report.

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

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

Chaux, C., Duval, L., and Pesquet, J.-C. (2006). Image analysis using a dual-tree M-band wavelet transform. *IEEE Trans. Image Process.*, 15(8):2397–2412.

Chaux, C., Pesquet, J.-C., and Duval, L. (2007b). 2D dual-tree complex biorthogonal M-band wavelet transform. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, volume 3, pages 845–848, Honolulu, HI, USA.

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

Chen, G. H.-G. and Rockafellar, R. T. (1997). Convergence rates in forward-backward splitting. *SIAM J. Optim.*, 7(2):421–444.

Chen, K. and Sacchi, M. D. (2014). Robust reduced-rank filtering for erratic seismic noise attenuation. *Geophysics*, 80(1):V1–V11.

Chouzenoux, E., Idier, J., and Moussaoui, S. (2011). A majorize-minimize strategy for subspace optimization applied to image restoration. *IEEE Trans. Image Process.*, 20(6):1517–1528.

Chouzenoux, E., Pesquet, J.-C., and Repetti, A. (2013). A block coordinate variable metric forward-backward algorithm. Technical report.

Chouzenoux, E., Pesquet, J.-C., and Repetti, A. (2014). Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. *J. Optim. Theory Appl.*, 162(1):107–132.

Claerbout, J. F. (1977). Parsimonious deconvolution. Technical report, SEP-13.

Claerbout, J. F. and Muir, F. (1973). Robust modeling with erratic data. *Geophysics*, 38(5):826–844.

Combettes, P. L., Condat, L., Pesquet, J.-C., and Vu, B. C. (2014). A forward-backward view of some primal-dual optimization methods in image recovery. In *Proceedings of the IEEE International Conference on Image Processing*, Paris, France.

Combettes, P. L. and Pesquet, J.-C. (2004). Wavelet-constrained image restoration. *Int. J. Wavelets Multiresolut. Inf. Proc.*, 2(4):371–389.

Combettes, P. L. and Pesquet, J.-C. (2007a). A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J. Sel. Topics Signal Process.*, 1(4):564–574.

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

Combettes, P. L. and Pesquet, J.-C. (2008). A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6):27.

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

Combettes, P. L. and Pesquet, J.-C. (2012). 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.

Combettes, P. L. and Vũ, B. C. (2013). Variable metric quasi-Fejér monotonicity. *Nonlinear Anal.*, 78:17–31.

Combettes, P. L. and Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200.

Comon, P. (1996). Contrasts for multichannel blind deconvolution. *IEEE Signal Process. Lett.*, 3(7):209–211.

Condat, L. (2013). A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158(2):460–479.

Costagliola, S., Mazzucchelli, P., and Bienati, N. (2011). Hybrid norm adaptive subtraction for multiple removal. In *Proc. EAGE Conf. Tech. Exhib.*, Vienna, Austria. European Assoc. Geoscientists Eng.

Costagliola, S., Mazzucchelli, P., and Bienati, N. (2012). Matched filtering with orthogonal constraints. In *Proc. EAGE Conf. Tech. Exhib.*, Copenhagen, Denmark. European Assoc. Geoscientists Eng.

Cucker, F. and Smale, S. (2002). Best choices for regularization parameters in learning theory: on the bias-variance problem. *Found. Comput. Math.*, 2(4):413–428.

da Costa Filho, C. A. (2013). Applications of independent component analysis to the attenuation of multiple reflections in seismic data. Master's thesis, Universidade Estadual de Campinas.

Dai, Y. H. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim*, 10:177–182.

de Hoop, M. V., Smith, H., Uhlmann, G., and van der Hilst, R. D. (2009). Seismic imaging with the generalized Radon transform: a curvelet transform perspective. *Inverse Problems*, 25(2):025005.

Deift, P. and Zhou, X. (1993). A steepest descent method for oscillatory Riemann–Hilbert problems: asymptotics for the MKdV equation. *Ann. of Math*, pages 295–368.

Demanet, L. and Hand, P. (2015). Scaling law for recovering the sparsest element in a subspace. *Inf. Inference*, 3(4):295–309.

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

Donno, D., Chauris, H., and Noble, M. (2010). Curvelet-based multiple prediction. *Geophysics*, 75(6):WB255–WB263.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: Asymptopia? *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 57(2):301–369.

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

Eckstein, J. and Bertsekas, D. P. (1992). On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programm.*, 55(1–3):293–318.

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

Esser, E., Lou, Y., and Xin, J. (2013). A method for finding structured sparse solutions to non-negative least squares problems with applications. *SIAM J. Imaging Sci.*, 6(4):2010–2046.

Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., and Freeman, W. (2006). Removing camera shake from a single photograph. In *Proc. ACM SIGGRAPH Comput. Graph.*, volume 25, pages 787–794, Boston, Massachusetts USA.

Finlayson, G. D., Hordley, S. D., Lu, C., and Drew, M. S. (2006). On the removal of shadows from images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1):59–68.

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

Geman, D. and Reynolds, G. (1992). Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):367–383.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, Baltimore, MD, USA.

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

Grave, E., Obozinski, G., and Bach, F. (2011). Trace Lasso: a trace norm regularization for correlated designs. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, pages 2187–2195, Granada, Spain.

Gray, W. C. (1978). Variable norm deconvolution. Technical Report SEP-14, Stanford Exploration Project.

Gray, W. C. (1979). A theory for variable norm deconvolution. Technical report, Stanford university.

Guitton, A. and Symes, W. (2003). Robust inversion of seismic data using the Huber norm. *Geophysics*, 68(4):1310–1319.

Guitton, A. and Verschuur, D. J. (2004). Adaptive subtraction of multiples using the L_1-norm. *Geophys. Prospect.*, 52:27–38.

Hadamard, J. (1923). *Lectures on Cauchy's problem in linear partial differential equations*. New Haven Yale University Press.

Hampson, D. (1986). Inverse velocity stacking for multiple elimination. *Can. J. Explor. Geophys.*, 22(1):44–55.

Haykin, S., editor (1994). *Blind Deconvolution*. Prentice Hall.

Herrmann, F. J. and Verschuur, D. J. (2004). Curvelet imaging and processing: adaptive multiple elimination. In *Proc. CSEG Nat. Conv.*, Calgary, Canada. Canadian Soc. Expl. Geophysicists.

Herrmann, F. J. and Verschuur, D. J. (2005). Robust curvelet-domain primary-multiple separation with sparseness constraints. In *Proc. EAGE Conf. Tech. Exhib.*, Madrid, Spain. European Assoc. Geoscientists Eng.

Hill, K. B. (2014). A seismic oil and gas primer. Hill Geophysical Consulting.

Hiriart-Urruty, J.-B. and Lemaréchal, C. (1993). *Convex Analysis and Minimization Algorithms*. Springer-Verlag.

Hoyer, P. (2004). Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.*, 5:1457–1469.

Huo, S. and Wang, Y. (2009). Improving adaptive subtraction in seismic multiple attenuation. *Geophysics*, 74(4):59–67.

Hurley, N. and Rickard, S. (2009). Comparing measures of sparsity. *IEEE Trans. Inform. Theory*, 55(10):4723–4741.

Jacobson, M. W. and Fessler, J. A. (2007). An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Trans. Image Process.*, 16(10):2411–2422.

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

Ji, H., Li, J., Shen, Z., and Wang, K. (2012). Image deconvolution using a characterization of sharp images in wavelet domain. *Appl. Comput. Harmon. Analysis*, 32(2):295–304.

Joo, J. H. (2014). Acoustic echo cancellation techniques for far-end telephony speech recognition in barge-in situation. *Contemp. Eng. Sci.*, 7(22):1113–1120.

Kaaresen, K. F. and Taxt, T. (1998). Multichannel blind deconvolution of seismic signals. *Geophysics*, 63(6):2093–2107.

Kabanikhin, S. I. (2008). Definitions and examples of inverse and ill-posed problems. *J. Inv. Ill-Posed Problems*, 16(4):317–357.

Kaplan, A. and Tichatschke, R. (1998). Proximal point methods and nonconvex optimization. *J. Global Optim.*, 13(4):389–406.

Kaplan, S. T. and Innanen, K. A. (2008). Adaptive separation of free-surface multiples through independent component analysis. *Geophysics*, 73(3):V29–V36.

Karsli, H., Dondurur, D., and Çifçi, G. (2006). Application of complex-trace analysis to seismic data for random-noise suppression and temporal resolution improvement. *Geophysics*, 71(3):V79–V86.

Kato, M., Yamada, I., and Sakaniwa, K. (1999). A set-theoretic blind image deconvolution based on hybrid steepest descent method. *IEICE Trans. Fund. Electron. Comm. Comput. Sci.*, E82-A(8):1443–1449.

Komodakis, N. and Pesquet, J.-C. (2014). Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *PREPRINT*.

Kowalski, M. (2009). Sparse regression using mixed norms. *Appl. Comput. Harmon. Analysis*, 27(3):303–324.

Kreimer, N., Stanton, A., and Sacchi, M. D. (2013). Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction. *Geophysics*, 78(6):V273–V284.

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

Krishnan, D., Tay, T., and Fergus, R. (2011). Blind deconvolution using a normalized sparsity measure. In *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pages 233–240, Colorado Springs, CO, USA.

Kundur, D. and Hatzinakos, D. (1996a). Blind image deconvolution. *IEEE Signal Process. Mag.*, 13(3):43–64.

Kundur, D. and Hatzinakos, D. (1996b). Blind image deconvolution revisited. *IEEE Signal Process. Mag.*, 13(6):61–63.

Lange, K. (2010). *Numerical Analysis for Statisticians*. Springer, New York.

Lavergne, M. (1986). *Méthodes sismiques*. Publications de l'Institut français du pétrole. Cours de l'École nationale supérieure du pétrole et des moteurs. Éd. Technip, Paris.

Le Touzé, G., Nicolas, B., Mars, J. I., and Lacoume, J.-L. (2009). Matched representations and filters for guided waves. *IEEE Trans. Signal Process.*, 57(5):1783–1795.

Li, Z.-X. and Lu, W.-K. (2013). Adaptive multiple subtraction based on 3D blind separation of convolved mixtures. *Geophysics*, 78(6):V251–V266.

Lin, D., Young, J., Huang, Y., and Hartmann, M. (2004). 3D SRME application in the Gulf of Mexico. In *Annual International Meeting*, volume 23, pages 1257–1260, Denver, CO, USA. Soc. Expl. Geophysicists.

Lin, Y. and Lee, D. D. (2005). Bayesian regularization and nonnegative deconvolution (brand) for acoustic echo cancellation. In *Proc. IEEE Workshop Appl. Signal Process. Audio Acoustics*, Mohonk Mountain House New Paltz, New York.

Lions, P. L. and Mercier, B. (1979). Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979.

Liu, K.-H. and Dragoset, W. H. (2012). Adaptive multiple separation based on information maximization. In *Annual International Meeting*, pages 1–6, Las Vegas, NV, USA. Soc. Expl. Geophysicists.

Liu, K.-H. and Dragoset, W. H. (2013). Blind-source separation of seismic signals based on information maximization. *Geophysics*, 78(4):V119–V130.

Lu, W.-K. and Liu, L. (2009). Adaptive multiple subtraction based on constrained independent component analysis. *Geophysics*, 74(1):V1–V7.

Luo, Z. Q. and Tseng, P. (1992). On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optim. Theory Appl.*, 72(1):7–35.

Ma, J. and Plonka, G. (2010). The curvelet transform — a review of recent applications. *IEEE Signal Process. Mag.*, 27(2):118–133.

Mallat, S. (2009). *A wavelet tour of signal processing: the sparse way*. Academic Press, San Diego, CA, USA, 3rd edition.

Mars, J., Mari, J.-L., Lacoume, J.-L., and Glangeaud, F. (2004). *Traitement du signal pour géologues et géophysiciens : Techniques avancées 3*. Publications de l'Institut Français du Pétrole. Éd. Technip.

McHugo, S., Webb, B., Grechishnikova, T., and Whitebread, R. (2008). 3D multiple attenuation compensates for irregular Black Sea bed. *Offshore*.

Mei, Y. and Zou, Z. (2010). A weighted adaptive subtraction for two or more multiple models. In *Annual International Meeting*, volume 29, pages 3488–3492, Denver, CO, USA. Soc. Expl. Geophysicists.

Moller, S. and Heusdens, R. (2013). Objective estimation of speech quality for communication systems. *Proc. IEEE*, 101(9):1955–1967.

Moni, A., Bean, C. J., Lokmer, I., and Rickard, S. (2012). Source separation on seismic data: Application in a geophysical setting. *IEEE Signal Process. Mag.*, 29(3):16–28.

Moreau, E. and Pesquet, J.-C. (1997). Generalized contrasts for multichannel blind deconvolution of linear systems. *IEEE Signal Process. Lett.*, 4(6):182–183.

Moreau, J. J. (1965). Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299.

Morlet, J. (1975). Seismic tomorrow, interferometry and quantum mechanics. In *Annual International Meeting*, Denver, CO, USA. Soc. Expl. Geophysicists.

Morlet, J., Arens, G., Fourgeau, E., and Giard, D. (1982). Wave propagation and sampling theory—part I: Complex signal and scattering in multilayered media. *Geophysics*, 47(2):203–221.

Mørup, M., Madsen, K. H., and Hansen, L. K. (2008). Approximate L_0 constrained non-negative matrix and tensor factorization. In *Proc. Int. Symp. Circuits Syst.*, pages 1328–1331, Seattle, WA , USA.

Mossi, M. I., Yemdji, C., and Evans, N. (2011). Non-linear acoustic echo cancellation using online loudspeaker linearization. In *Proc. IEEE Workshop Appl. Signal Process. Audio Acoustics*, pages 97–100, New Paltz, New York.

Nandi, A. K., Mampel, D., and Roscher, B. (1997). Blind deconvolution of ultrasonic signals in nondestructive testing applications. *IEEE Trans. Signal Process.*, 45(5):1382–1390.

Neelamani, R., Baumstein, A., and Ross, W. (2008). Complex curvelet-based adaptive subtraction of several noise templates. In *Annual International Meeting*, pages 3650–3655, Las Vegas, NV, USA. Soc. Expl. Geophysicists.

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

Nesterov, Y. (2013). Gradient methods for minimising composite functions. 140(1):125–161.

Nocedal, J. and Wright, S. J. (2006). *Numerical optimization*. Springer Verlag, New York, 2nd edition.

Nowak, E. J. and Imhof, M. G. (2006). Amplitude preservation of Radon-based multiple-removal filters. *Geophysics*, 71(5):V123–V126.

Nuzzo, L. and Quarta, T. (2004). Improvement in GPR coherent noise attenuation using $\tau - p$ and wavelet transforms. *Geophysics*, 69(3):789–802.

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

Osman, O. S. and Robinson, E. A., editors (1996). *Seismic Source Signature Estimation and Measurement*, volume 18 of *Geophysics Reprint Series*. Soc. Expl. Geophysicists, Tulsa, OK, USA.

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

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

Pesquet, J.-C. and Pustelnik, N. (2012). A parallel inertial proximal optimization method. *Pac. J. Optim.*, 8(2):273–305.

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

Pham, M. Q., Chaux, C., Duval, L., and Pesquet, J.-C. (2014a). A constrained-based optimization approach for seismic data recovery problems. In *Proc. Int. Conf. Acoust. Speech Signal Process.*, pages 2377–2381, Florence, Italy.

Pham, M. Q., Duval, L., Chaux, C., and Pesquet, J.-C. (2014b). A primal-dual proximal algorithm for sparse template-based adaptive filtering: Application to seismic multiple removal. *IEEE Trans. Signal Process.*, 62(16):4256–4269.

Pica, A. (1992). Behaviour of prestack Backus filters for multiple suppression. In *Proc. EAEG Meeting*, Paris, France. European Assoc. Geoscientists Eng.

Pica, A., Poulain, G., David, B., Magesan, M., Baldock, S., Weisser, T., Hugonnet, P., and Herrmann, P. (2005a). 3D surface-related multiple modeling. *The Leading Edge*, 24:292–296. Special section: Multiple attenuation.

Pica, A., Poulain, G., David, B., Magesan, M., Baldock, S., Weisser, T., Hugonnet, P., and Herrmann, P. (2005b). 3D surface-related multiple modeling, principles and results. In *Annual International Meeting*, volume 24, pages 2080–2083, Houston, TX, USA. Soc. Expl. Geophysicists.

Pokrovskaia, T. and Wombell, R. (2004). Attenuation of residual multiples and coherent noise in the wavelet transform domain. In *Proc. EAGE Conf. Tech. Exhib.*, Paris, France. European Assoc. Geoscientists Eng.

Pustelnik, N., Pesquet, J.-C., and Chaux, C. (2012). Relaxing tight frame condition in parallel proximal methods for signal restoration. *IEEE Trans. Signal Process.*, 60(2):968–973.

Raguet, H., Fadili, J., and Peyré, G. (2013). A generalized forward-backward splitting. *SIAM J. Imaging Sci.*, 6(3):1199–1226.

Renaut, R. A., Hntynková, I., and Mead, J. (2010). Regularization parameter estimation for large-scale Tikhonov regularization using a priori information. *Comput. Stat. Data Anal.*, 54(12):3430–3445.

Repetti, A., Chouzenoux, E., and Pesquet, J.-C. (2012). 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.

Repetti, A., Pham, M. Q., Duval, L., Chouzenoux, E., and Pesquet, J.-C. (2015). Euclid in a taxicab: Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization. *IEEE Signal Process. Lett.*, 22(5):539–543.

Ricker, N. (1940). The form and nature of seismic waves and the structure of seismograms. *Geophysics*, 5(4):348–366.

Ricker, N. (1943). Further developments in the wavelet theory of seismogram structure. *Bull. Seismol. Soc. Amer.*, 33(3):197–228.

Ristow, D. and Kosbahn, B. (1979). Time-varying prediction filtering by means of updating. *Geophys. Prospect.*, 27(1):40–61.

Roberts, S. J., Husmeier, D., Rezek, I., and Penny, W. D. (1998). Bayesian approaches to Gaussian mixture modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1133–1142.

Robinson, E. A. and Treitel, S. (1967). Principles of digital Wiener filtering. *Geophys. Prospect.*, 15(3):311–332.

Rockafellar, R. T. (1976). Monotone operator and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898.

Rockafellar, R. T. and Wets, R. J.-B. (2005). *Variational Analysis*. Fundamental Principles of Mathematical Sciences. Berlin: Springer-Verlag, third corrected printing edition.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268.

Ryan, H. (1994). Ricker, Ormsby, Klauder, Butterwork — a choice of wavelets. *CSEG Recorder*, pages 8–9.

Sacchi, M. D. (1997). Reweighting strategies in seismic deconvolution. *Geophys. J. Int.*, 129(3):651–656.

Sacchi, M. D., Velis, D. R., and Cominguez, A. H. (1994). Minimum entropy deconvolution with frequency-domain constraints. *Geophysics*, 59(6):938–945.

Salaun, N., Hardouin, D., Huard, I., and Balanca, J. (2015). Adaptive subtractive methods for low frequency multiples on broadband data: A case study offshore West Africa (check subtractive). In *Proc. EAGE Conf. Tech. Exhib.* European Assoc. Geoscientists Eng. Submitted.

Sanchis, C. and Hanssen, A. (2011). Multiple-input adaptive seismic noise canceller for the attenuation of nonstationary coherent noise. *Geophysics*, 76(6):V115–V126.

Sheriff, R. (1973). *Encyclopedic Dictionary of Exploration Geophysics*. Society of Exploration Geophysicists.

Sheriff, R. E. (1989). *Geophysical Methods*. Prentice Hall.

Song, J., Verschuur, E., and Chen, X. (2013). Comparing three feedback internal multiple elimination methods. *J. Appl. Geophys.*, 95(0):66–76.

Sotthivirat, S. and Fessler, J. A. (2002). Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms. *IEEE Trans. Image Process.*, 11(3):306–317.

Spitz, S., Hampson, G., and Pica, A. (2009). Simultaneous source separation using wave field modeling and PEF adaptive subtraction. In *Proc. EAGE Marine Seismic Workshop*, Limassol, Cyprus. European Assoc. Geoscientists Eng.

Sra, S. (2011). Nonconvex proximal splitting: batch and incremental algorithms. Technical Report 2, Max Planck Institute for Intelligent Systems, Tübingen, Germany.

Takahata, A. K., Nadalin, E. Z., Ferrari, R., Duarte, L. T., Suyama, R., Lopes, R. R., Romano, J. M. T., and Tygel, M. (2012). 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.

Taner, M. T. (1980). Long period sea-floor multiples and their suppression. *Geophys. Prospect.*, 28(1):30–48.

Taner, M. T., O'Doherty, R. F., and Koehler, F. (1995). Long period multiple suppression by predictive deconvolution in the $x-t$ domain. *Geophys. Prospect.*, 43(4):433–468.

Tappen, M. F., Freeman, W. T., and Adelson, E. H. (2005). Recovering intrinsic images from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(9):1459–1472.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of Ill-posed problems*. W.H. Winston.

Trad, D., Ulrych, T., and Sacchi, M. (2003). Latest views of the sparse Radon transform. *Geophysics*, 68(1):386–399.

Tseng, P. (2000). A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446.

Ulrych, T. J. and Sacchi, M. D. (2006). *Information-Based Inversion and Processing with Applications*, volume 36 of *Handbook of Geophysical Exploration: Seismic Exploration*.

Vũ, B. C. (2013). A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681.

van Borselen, R., Baardman, R., Martin, T., Goswami, B., and Fromyr, E. (2012). An inversion approach to separating sources in marine simultaneous shooting acquisition — application to a Gulf of Mexico data set. *Geophys. Prospect.*, 60(4):640–647.

van Borselen, R. and Schonewille, M. (2005). 3D surface-related multiple elimination: acquisition and processing solutions. *The Leading Edge*, 24:260–268. special section: Multiple attenuation.

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

Vaseghi, S. V. (2006). *Advanced digital signal processing and noise reduction*. John Wiley & Sons.

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

Verschuur, D. and Berkhout, A. (1992). Adaptive surface related multiple elimination. *Geophysics*, 57(9):1166–1177.

Verschuur, D. J. (2006). *Seismic multiple removal techniques: past, present and future*. EAGE Publications.

Walden, A. T. (1985). Non-Gaussian reflectivity, entropy, and deconvolution. *Geophysics*, 50(12):2862–2888.

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

Wang, Y. (2003). Multiple subtraction using an expanded multichannel matching filter. *Geophysics*, 68:346–354.

Weglein, A. B., Gasparotto, F. A., Carvalho, P. M., and Stolt, R. H. (1997). An inverse-scattering series method for attenuating multiples in seismic reflection data. *Geophysics*, 62(6):1975–1989.

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

Weiss, Y. (2001). Deriving intrinsic images from image sequences. In *Proc. IEEE Int. Conf. Comput. Vis.*, volume 2, pages 68–75, Vancouver, British Columbia, Canada.

Weisser, T., Pica, A., Herrmann, P., and Taylor, R. (2006). Wave equation multiple modelling: acquisition independent 3D SRME. *First Break*, 24:75–79.

Wu, M. and Wang, S. (2011). A case study of $f - k$ demultiple on 2D offshore seismic data. *The Leading Edge*, 30(4):446–450.

Xu, Y. and Yin, W. (2013). A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.*, 6(3):1758–1789.

Yilmaz, Ö. (2001). *Seismic data analysis: processing, inversion, and interpretation of seismic data*. Soc. Expl. Geophysicists.

Yunpeng Li, D. P. H. (2008). Sparse long-range random field and its application to image denoising. *Proc. Eur. Conf. Comput. Vis.*, 5304:344–357.

Zeng, Z., Chan, T.-H., Jia, K., and Xu, D. (2012). Finding correspondence from multiple images via sparse and low-rank decomposition. In *Proc. Eur. Conf. Comput. Vis.*, volume 7576, pages 325–339.

Zhang, N. and Wang, Y. (2011). 2D/3D seismic multiple attenuation in the deep water of South China Sea. In *Proc. EAGE Conf. Tech. Exhib.*, Vienna, Austria. European Assoc. Geoscientists Eng.

Zibulevsky, M. and Pearlmutter, B. A. (2001). Blind source separation by sparse decomposition in a signal dictionary. *Neural Comput.*, 13(4):863–882.

Zonoobi, D., Kassim, A. A., and Venkatesh, Y. V. (2011). Gini index as sparsity measure for signal reconstruction from compressive samples. *IEEE J. Sel. Topics Signal Process.*, 5(5):927–932.