Time Domain Least Squares Migration and Dimensionality Reduction

par

Valentin Tschannen

Rapport de stage en vue de l'obtention du Diplôme de Master STE à **L'Ecole et Observatoire des Sciences de la Terre** de l'Université de Strasbourg, parcours excellence

Stage effectué au *Seismic Laboratory for Imaging and Modeling* de l'Université de Colombie Britannique

Maîtres de stage: *Zhilong Fang* et *Felix J. Herrmann*

Vancouver BC, Canada, Juin 2014

Abstract

Least-squares seismic migration (LSM) is a wave equation based linearized inversion problem relying on the minimization of a least-squares misfit function, with respect to the medium perturbation, between recorded and modeled wavefields. Today's challenges in Hydrocarbon exploration are to build high resolution images of more and more complicated geological reservoirs, which requires to handle very large systems of equations. The extreme size of the problem combined with the fact that it is ill-conditioned make LSM not yet feasible for industrial purposes. To overcome this "curse of dimensionality", dimension reduction and divide-and-conquer techniques that aim to decrease the computation time and the required memory, while conserving the image quality, have recently been developed. By borrowing ideas from stochastic optimization and compressive sensing, the imaging problem is reformulated as an L_1-regularized, sparsity promoted LSM. The idea is to take advantage of the compressibility of the model perturbation in the curvelet domain and to work on series of smaller subproblems each involving a small randomized subset of data. We try two different subset sampling strategies, artificial randomized simultaneous sources experiments ("supershots") and drawing sequential shots firing at random source locations. These subsets are changed after each subproblem is solved. In both cases we obtain good migration results at significantly reduced computational cost. Application of these methods to a complicated synthetic model yields to encouraging results, underlining the usefulness of sparsity promotion and randomization in time stepping formulation.

Keywords : Wave equation migration, sparsity promotion, compressive sensing, stochastic optimization.

Résumé

La migration par les moindres carrés de données de sismique réflexion au moyen de l'équation d'onde est une procédure d'inversion linéarisée qui cherche à minimiser, au sens de la norme L_2, l'écart entre les champs d'ondes enregistrés et modélisés. Le défi aujourd'hui pour l'exploration des hydrocarbures est d'être capable d'obtenir des images de hautes résolutions de réservoirs situés dans des zones géologiques de plus en plus complexes. Cependant, la taille extrême des données, couplée au fait que le problème est mal conditionné, rendent cette approche impraticable dans l'industrie pour des modèles 3D encore aujourd'hui. Pour palier à cette limitation, plusieurs méthodes ont récemment été développées afin de diminuer les temps de calculs ainsi que la mémoire nécessaire, sans pour autant sacrifier la qualité des résultats. En empruntant des idées venant de l'optimisation stochastique et de l'acquisition comprimée, le problème d'imagerie peut être reformulé pour y inclure une régularisation au moyen de la norme L_1 et pour prendre en compte le fait que la perturbation du modèle peut être compressée en passant dans le domaine des curvelettes. En travaillant avec des sous problèmes construits à partir d'un sous ensemble des données, nous présentons ici deux méthodes d'échantillonnages du champ d'onde. La première consiste à simuler artificiellement des "super-sources" construites comme étant la somme de sources séquentielles munies d'un poids aléatoire dans la procédure d'inversion. L'autre méthode consiste à utiliser des sources séquentielles déclenchées à des positions aléatoires le long du profil. Après que chaque sous problème ait été résolu, nous renouvelons l'échantillonnage. Dans les deux cas nous obtenons de bons résultats pour des temps de calculs significativement inférieurs. Ces méthodes sont appliquées à un modèle synthétique compliqué et mènent à des résultats encourageants, ce qui montre l'utilité d'utiliser l'échantillonnage aléatoire ainsi que de prendre en compte la compressibilité de la perturbation du modèle dans le domaine des curvelettes.

Mots-clefs : Migration par l'équation d'onde, optimisation stochastique, acquisition comprimée.

Contents

List of Figures

Acknowledgment

This report is the result of the four month I spent at the Seismic Laboratory for Imaging and Modeling of the University of British Columbia. I first would like to thank its director, Pr. Felix Herrmann, for accepting me in the group and funding my stay. I also thank Miranda Joyce and Jane Peng for their welcome and their support when I arrived in Vancouver. I am particularly grateful toward my supervisor Zhilong Fang, who has been very available and always took time to answer my questions and help me with my project. Finally, I thank the members of the SLIM group, I have learned a lot during my visit and discussing about their different research areas has sharpen my interest for the fields of seismic imaging and modeling.

The bulk of my work is based on the following software: SPOT (`http://www.cs.ubc.ca/labs/scl/spot/`), $SPGl_1$ (`http://www.cs.ubc.ca/labs/scl/spgl1/`), IWAVE (`http://www.trip.caam.rice.edu/software/iwave/`), Madagascar (`http://www.ahay.org/wiki/Main_Page`) and CurveLab (curvelet.org).

Current industrial sponsors of the Seismic Laboratory for Imaging and Modeling are BG Group, BGP, CGG, Chevron, ConocoPhillips, Ion, Petrobras, PGS, Statoil, Total SA., WesternGeco, Woodside.

1 Introduction

Modern seismic exploration techniques rely on the collection of massive data volumes in order to provide high resolution images of increasingly complicated regions of the Earth. Seismic inversion extracts information about the subsurface from surface or borehole data, by adjusting model parameters to predict this data via numerical simulation. The least-squares approach to inversion, known as Full Waveform Inversion (FWI), is a non-linear data-fitting procedure that seeks to minimize the root mean square difference between predicted and observed waveforms [10]. The simplest mathematical statement of this least-squares approach (ignoring regularization, data and model representation details,...) is:

$$\underset{\mathbf{m}}{\text{minimize}}\ \Phi(\mathbf{m}) := \left\{ \frac{1}{2K} \sum_{i=1}^{K} \|\mathbf{d}_i - \mathcal{F}_i[\mathbf{m}, \mathbf{q}_i]\|_2^2 = \frac{1}{2} \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m}, \mathbf{Q}]\|_F^2 \right\} \tag{1}$$

with $\mathbf{d}_i$ representing the vectorized shot records, at the receivers positions, of the Earth response to sources $\mathbf{q}_i$. The vectors $\mathbf{q}_i$ encode the location and the signature of the i^{th} shot experiment. $\mathcal{F}_i[\mathbf{m}, \mathbf{q}_i]$, $i = 1...K$, are the nonlinear forward operators of the wave equation composed with a restriction of the full solution operators to the receivers positions, and $K = N_t.N_s$, with N_t, N_s the number of time samples and sources, respectively. The modeling operator $\boldsymbol{\mathcal{F}} : \boldsymbol{\mathcal{M}} \rightarrow \boldsymbol{\mathcal{D}}$ maps the model space $\boldsymbol{\mathcal{M}}$ (a set of possible models of the subsurface) of dimension n to the data space $\boldsymbol{\mathcal{D}}$, a Hilbert space of dimension $N_s.N_r.N_t$ (with N_r the number of receivers) of possible data sets with a Frobenius norm $\|.\|_F$ that serves as an error measure (define as $\|A\|_F = \sqrt{trace(A^\dagger A)}$, with † the complex conjugate transpose). In the acoustic constant-density case, this operator is parametrized by the unknown velocity model $\mathbf{m}$ and involves the inversion of a large system of equations. Starting from an initial guess of the subsurface parameters, the fitting scheme is repeated iteratively by updating the model and starting again the minimization procedure. It can be written as

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \delta m_k \tag{2}$$

where the model residual δm_k is called the searched direction in the model space and α_k is the step-length. In this report we are interested in the computation of the model residual δm. This particular approach which geometrically re-locate seismic events in space to the location the event occurred in the subsurface is called least-squares migration (LSM), it belongs to the family of pre-stack depth wave equation migration. Within the framework of Born approximation we write the unknown earth model $\mathbf{m}$ as the sum of an approximate model $\mathbf{m}_0$ known a-priori and a small model perturbation δm. LSM then seeks to find a model perturbation that satisfies the following linearized equation:

$$\frac{\partial \mathcal{F}[\mathbf{m}_0, \mathbf{Q}]}{\partial m} \delta m \approx \delta d. \tag{3}$$

Here, we define the misfit vector $\delta d \in \mathbb{R}^{N_s N_r N_t}$ as $\delta d = \mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m}_0, \mathbf{Q}]$. This vector represents the difference at the receiver positions between the recorded pressure field $\mathbf{D}$ and the modeled pressure field $\boldsymbol{\mathcal{F}}[\mathbf{m}_0, \mathbf{Q}]$ for each source-receiver pair of the survey. The partial derivative of the forward operator with respect to the model parameters are called the Fréchet derivatives or sensitivity kernels.

2 The Imaging Problem

2.1 Forward Modeling

The forward problem is dedicated to the estimation of seismic wavefields for a given subsurface model. In our work we use the software framework IWAVE++ [9] to perform the modeling. In the most general case we would have to deal with the anisotropic, visco-elastic wave equation that links the model parameters corresponding to the 21 elastic moduli, the density and some memory variables that characterizes the anelastic behavior of the subsurface. Because the forward problem requires the solution of a very large system of PDEs we need to restrict ourselves to the case of the acoustic approximation of the wave equation. Assuming the density variations are negligible, the equation in the time domain writes as:

$$\nabla^2 \mathbf{p} - \frac{1}{c^2}\frac{\partial^2 \mathbf{p}}{\partial t^2} = \mathbf{S} \tag{4}$$

where $\mathbf{p}(\mathbf{x}, t)$ is the pressure field, $\mathbf{S}(\mathbf{x}, t)$ is a time dependent source term known spatially as punctual sources. In practice we used a Ricker wavelet to simulate the excitation of the medium. For the medium density $\rho(\mathbf{x})$ and the bulk modulus $\kappa(\mathbf{x})$ we define the compressional wave propagation velocity

$$c(\mathbf{x}) = \sqrt{\frac{\kappa}{\rho}} \tag{5}$$

$\mathbf{x}$ and t denote the spatial coordinates and the time.

To numerically solve these PDEs, IWAVE uses a finite difference method where partial derivatives are transform into finite difference expressions (we used a 2^{nd} order in time and 4^{th} order in space) after having discretized the medium into a regular mesh. The computation is done on parallel by distributing sources over processors.

Because seismic imaging in time domain requires the cross-correlation, for each source, of the incident wavefield from source to receivers and the back propagated residues from the receivers to the source [10] at each node of the mesh; IWAVE uses checkpoints that require to save the incident wavefield only at some given nodes of the mesh and then evaluate it at the other nodes via an interpolation method.

Free surfaces boundary conditions are implemented on the top side of the model to mimic the interface solid-air or water-air and absorbing boundary conditions are implemented on the other sides to simulate an infinite medium by avoiding numerical reflexion artifacts.

IWAVE requires the model parameters to be stored in the RSF format of Madagascar and the acquisition geometry parameters in an SU file. Several functions were written to do the interface between the Matlab Scripts and the IWAVE code written in C/C++.

2.2 Classical Least-Squares Local Optimization Approach

In this section we present the simplest approach toward least-squares migration, ignoring the probabilistic maximum likelihood or generalized inverse formulations [13]. In our case, the model $\mathbf{m} \in \mathcal{M}$ is real valued and represents the bulk modulus and the buoyancy defined at each node of the numerical mesh.

We start from the least-squares norm $\phi(\mathbf{m})$ of the misfit vector defined in equation 1 and write it as:

$$\phi(\mathbf{m}) = \frac{1}{2}\delta d^\dagger \delta d. \tag{6}$$

This is referred as the misfit function or the objective function. Here the symbol † denotes the adjoint operator or transpose complex conjugate operator (although the misfit vector is real valued in the time domain we will continue with this notation).

As this approach is local, we will seek a minimum of the objective function in the vicinity of a starting model or a-priory model $\mathbf{m}_0$. In order to linearize the objective function we will restrain our calculations to the Born approximation where a linear relationship between the model and wavefield perturbations is assumed. Thus, we can write the model $\mathbf{m}$ as the sum of the starting model $\mathbf{m}_0$ plus a perturbation model δm : $\mathbf{m} = \mathbf{m}_0 + \delta m$. We also assume to be in the constant density case therefore the model parameter is only the bulk modulus.

We can operate a second order Taylor-Lagrange development of the objective function in the vicinity of $\mathbf{m}_0$ (for $\mathbf{m}$ is real valued)

$$\phi(\mathbf{m}) = \phi(\mathbf{m}_0) + \frac{\partial\phi(\mathbf{m}_0)}{\partial m}\delta m + \frac{1}{2}\frac{\partial^2\phi(\mathbf{m}_0)}{\partial m^2}\delta m^2 + \mathcal{O}(\mathbf{m}^3). \tag{7}$$

In order to find the minimum of this objective function upon all possible perturbation models δm, we want to find the one that annuls its gradient. That is to say, we want to find δm such that

$$\nabla\phi(\mathbf{m}) \approx \frac{\partial\phi(\mathbf{m}_0)}{\partial m} + \frac{\partial^2\phi(\mathbf{m}_0)}{\partial m^2}\delta m = 0. \tag{8}$$

This gives us the perturbation model vector

$$\delta m = -\left[\frac{\partial^2\phi(\mathbf{m}_0)}{\partial m^2}\right]^{-1}\frac{\partial\phi(\mathbf{m}_0)}{\partial m}. \tag{9}$$

The derivative of $-\phi(\mathbf{m}_0)$ with respect to the model parameters $\mathbf{m}$ leads us to the search of the model perturbation in a descent direction of the misfit function at point $\mathbf{m}_0$. If we refer to the definition of the misfit function (Eq.6) we obtain:

$$\frac{\partial\phi(\mathbf{m}_0)}{\partial m} = -\mathcal{R}_e\left[\mathbf{J}_0^\dagger \delta d\right], \tag{10}$$

where $\mathbf{J}_0 = \frac{\partial\boldsymbol{\mathcal{F}}[\mathbf{m}_0,\mathbf{Q}]}{\partial\mathbf{m}}$ is the sensitivity or Fréchet derivative matrix and $\mathcal{R}_e$ denotes the real part.

The second derivative of the misfit function is the Hessian, it defines the curvature of the misfit function at $\mathbf{m}_0$. From (6) we obtain the following expression of the Hessian:

$$\frac{\partial^2\phi(\mathbf{m}_0)}{\partial m^2} = \mathcal{R}_e\left[\mathbf{J}_0^\dagger\mathbf{J}_0\right] + \mathcal{R}_e\left[\frac{\partial\mathbf{J}_0^t}{\partial m^t}(\delta d^*...\delta d^*)\right], \tag{11}$$

where t and * denotes the transpose and the complex conjugate. Of the two terms of the Hessian matrix, the first is straightforward to compute, whereas the second term is difficult to compute. Moreover, if the residuals are small or if the forward equation is linearized this last term is small [8]. Therefore, we introduce the approximate Hessian $\mathbf{H}_0 = \mathcal{R}_e\left[\mathbf{J}_0^\dagger\mathbf{J}_0\right]$. The term $\mathbf{J}_0^\dagger$ can be seen as the reverse time migration operator and the term $\mathbf{J}_0$ as the linearized Born scattering operator [10]. This approximate Hessian is an n × n real matrix (with n the number

of parameters), symmetric (i.e $\mathbf{H}_0^t = \mathbf{H}_0$) and positive definite (i.e $x^\dagger \mathbf{H}_0 x > 0 \; \forall x \in \mathcal{M} \setminus \{\emptyset\}$) [8]. The inversion of the Hessian corresponds to the physical idea of correcting for wrong amplitudes in the subsurface image because of the uneven illumination by the incident waves.

From (10) and (11) we obtain the following expression of the perturbation model:

$$\delta m = -\mathcal{R}_e \left\{ \left[\mathbf{J}_0^\dagger \mathbf{J}_0 \right]^{-1} \right\} \mathcal{R}_e \left[\mathbf{J}_0^\dagger \delta d \right]. \tag{12}$$

Because the size of the problem is very big (10 of millions of unknowns for 2D data) we must resort to an iterative algorithm to find an approximate numerical solution of equation (12). We use a conjugate gradient method to solve the system of linear equations $\mathbf{H}_0 \delta m = \mathcal{R}_e \left\{ \mathbf{J}_0^\dagger \delta d \right\}$ with the unknown δm. By definition, two non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be conjugate with respect to the Hessian matrix if $\langle u, v \rangle_{\mathbf{H}_0} = u^t \mathbf{H}_0 v = 0$. By symmetry of the inner product, if $\mathbf{u}$ is conjugate to $\mathbf{v}$ then $\mathbf{v}$ is conjugate to $\mathbf{u}$. We denote by δm_0 the initial guess of the solution and δm_* the exact solution of the system. In practice we take $\delta m_0 = 0$. If we consider that $\mathrm{P} = \left\{ p_k : \forall i \neq k, i, k \in [1, n] \, , \langle p_i, p_k \rangle_{\mathbf{H}_0} = 0 \right\}$ is a set of n mutually conjugate directions then $\mathbf{P}$ is a basis of $\mathbb{R}^n$. This means that we can express the exact solution as:

$$\delta m_* = \sum_{i=1}^{n} \alpha_i p_i \tag{13}$$

with α_i a family of real coefficients. Here we want to choose wisely the conjugate vectors p_i to obtain a good approximate of δm_* for $i = 1..n'$ with $n' << n$. Starting with δm_0 we search for the solution and in each iterations we need a metric to tell us whether we are closer to the unknown solution δm_*. This metric comes from the fact that δm_* is also the unique minimizer of the following quadratic function $\mathfrak{f}(\delta m)$:

$$\mathfrak{f}(\delta m) = \frac{1}{2} \delta m^t \mathbf{H}_0 \delta m - \delta m^t \mathcal{R}_e \left\{ \mathbf{J}_0^\dagger \delta d \right\}, \tag{14}$$

and therefore if $\mathfrak{f}(\delta m)$ becomes smaller after an iteration it means that we are closer to δm_*. As this quadratic function is convex we choose p_0 to be the negative of the gradient of $\mathfrak{f}$ at point δm_0. Its gradient is $\mathbf{H}_0 \delta m - \mathcal{R}_e \left\{ \mathbf{J}_0^\dagger \delta d \right\}$ so we obtain $p_0 = \mathcal{R}_e \left\{ \mathbf{J}_0^\dagger \delta d \right\} - \mathbf{H}_0 \delta m_0$.

Let $\mathbf{r}_k$ be the residual of the gradient of $\mathfrak{f}$ at the k^{th} step:

$$\mathbf{r}_k = \mathcal{R}_e \left\{ \mathbf{J}_0^\dagger \delta d \right\} - \mathbf{H}_0 \delta m_k \tag{15}$$

we want to build the directions p_k to be conjugate to each other. This gives the following expression:

$$p_k = \mathbf{r}_k - \sum_{i<k} \frac{p_i^t \mathbf{H}_0 \mathbf{r}_k}{p_i^t \mathbf{H}_0 p_i} p_i \tag{16}$$

and the next optimal location is given by

$$\delta m_{k+1} = \delta m_k + \alpha_k p_k \tag{17}$$

with

$$\alpha_k = \frac{p_k^t \mathbf{r}_{k-1}}{p_k^t \mathbf{H}_0 p_k} \tag{18}$$

the iterations are repeated until a satisfactory small enough residual is obtained.

The results obtained on a simple horizontally layered model (1D case) are shown below. The dimensions are 1000 m $\times$ 1000 m. The buoyancy is set to be constant and the bulk modulus is parametrized on a 200 $\times$ 200 grid with a grid size of 5 m, it varies from 25GPa to 33GPa. The synthetic data are computed with IWAVE from 20 source positions sampled at an interval of 50 m and 100 receivers sampled with an interval of 10 m. The time of the simulation is 0.805 s with a time step of 0.00115 s, and the shot impulsion is modeled by a 30 Hz Ricker wavelet. The data residuals are given by the difference between simulations with the true and smoothed bulk models. The migration took 6h10min to compute on the cluster, using 2 nodes and 4 processors per node (ppn).

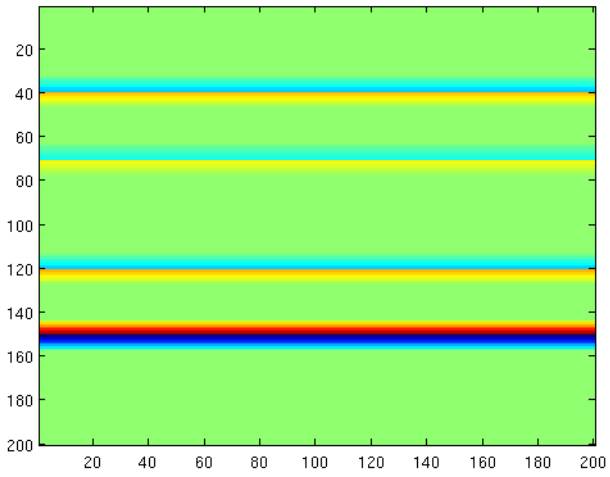


Figure 1: True model perturbation

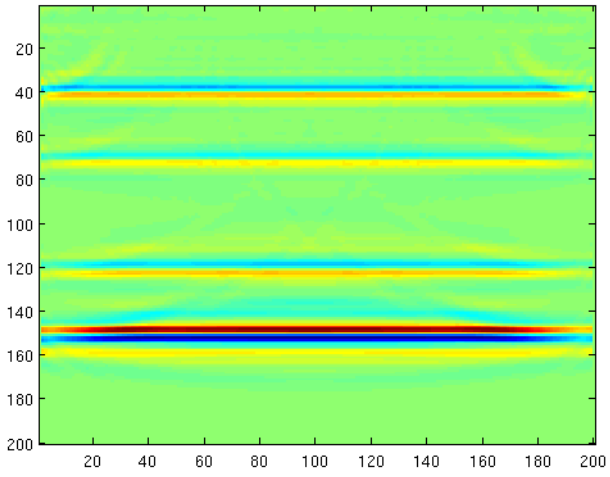


Figure 2: Model perturbation reconstructed by a conjugate gradient method

In figure 2 we can see that the perturbation is well recovered and that the resolution at the boundaries of the model is good. However, we notice some pretty strong artifacts in the shallow and deep interfaces.

Because of the size of the problem this approach to LSM isn't suitable for 3D or even large 2D models. In the following we present an other formulation that aims to make LSM computationally more affordable.

2.3 Dimensionality Reduction

2.3.1 Stochastic Optimization

The purpose of reformulating the LSM as a stochastic optimization problem is to reduce the cost involved in the evaluation of the gradient of the misfit by randomly drawing a batch of realizations of the objective function. When replacing the simple shot data by random simultaneous shots (see section 2.4.1) we can dramatically reduce the number of PDE solves (see e.g [12]). Two different approaches can be combined to formulate this problem, stochastic approximation techniques and sampled average approximation.

The stochastic formulation is based on the following identity [12] :

$$\|A\|_F^2 = E_w\left(w^T A^T A w\right) = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} w_k^T A^T A w_k, \tag{19}$$

where E_w denotes the expectation over the random vectors w. This vectors need to be chosen such that $E_w(ww^T) = I$, the identity matrix. The objective function as defined in equation (1) is now formulate as

$$\Phi(\mathbf{m}, \mathbf{Q}, \mathbf{D}) = E_w \Phi(\mathbf{m}, \mathbf{Q}w, \mathbf{D}w). \tag{20}$$

The Sampled Average Approximation replaces the expectation over w by an ensemble average

$$\Phi_K(\mathbf{m}) = \frac{1}{K} \sum_{k=1}^{K} \Phi(\mathbf{m}, \mathbf{Q}w_k, \mathbf{D}w_k). \tag{21}$$

The batch size K will determine the computation cost and the amount of artifacts introduced during the inversion. Stochastic Approximation algorithms are centered on the idea of picking a new random realization in each iteration of the algorithm. The random vector w is changed over iterations. While decreasing the size of the problem, these techniques might introduce source cross-talk in the image and the noise in the data may be amplified by randomly combining shots. An other problem is that the well known results of these approaches are derived for the case of a convex misfit function which isn't the case for seismic imaging problems.

2.3.2 Sparsity and Curvelet Transform

We call sparse matrix a matrix populated primarily with zeros coefficients. A signal is said to be compressible when it has only a few large coefficients after being transform in a carefully chosen domain basis. Finding a good sparsifying domain can allow the use of sophisticated sparsity-promoting solvers that reduce the required computational resources for imaging purposes [5].

The wave equation reduced Hessian $\mathbf{H}_0$ (i.e the demigration/migration operator) has a null space because of the constrains in the acquisition geometry, the limited band-width of the recorded signal and the complexity of the subsurface and therefore requires regularization to stabilize its inversion (at least for 3D data). The transform domain sparsity leads to a concentration of the update's energy into few large transform domain coefficients. It serves as a prior

that fills the null-space. To guarantee a fast decay for the magnitude sorted coefficients on the updates, we require the transform to detect wavefronts, possibly with conflicting dips, and to be nearly invariant under wave propagation. Rigorously, we would need to measure the sparsity of a matrix by counting each one of it's non-zero coefficients. Such a problem belongs to the class of NP-hard (Non-deterministic Polynomial-time hard) problems. However in practice we prefer to use the l_1 norm which is more suitable for algorithmic use. Using the l_1 norm turns the sparsity promotion into a convex problem wich is easier to solve than a NP-hard and that can lead to the same solution under certain conditions (see e.g [4]).

A sparsity operator that is suitable for seismic problems [4] is the mirror extended discrete curvelet transform [1]. Curvelets decompose wavefields as a superposition of highly anisotropic localized and multiscaled waveforms. In the Figures 3a and 3b are ploted the sorted absolute value coefficients of the model residual δm and of the transformation of δm in the curvelet domain. We can see that the proportion of significant coefficients is much less for the curvelet-transformed model residual.

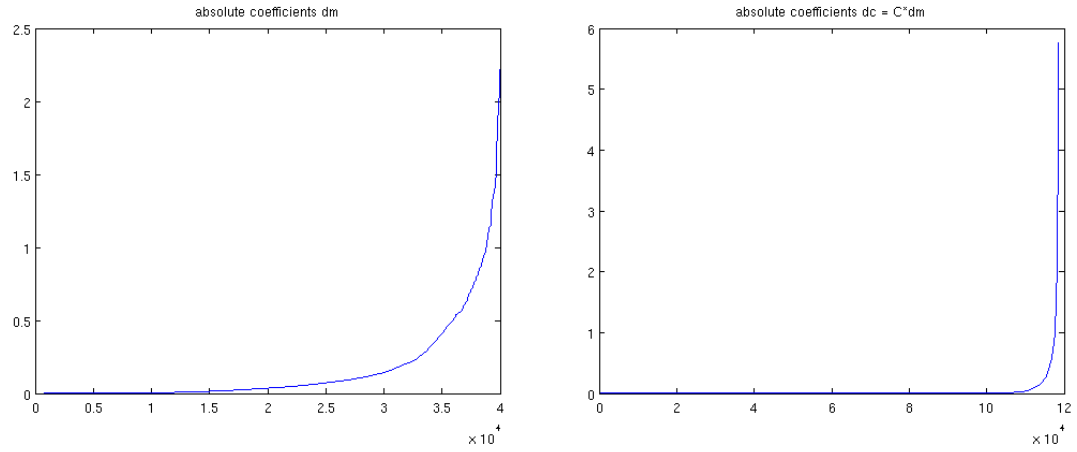


Figure 3: Absolute value of the coefficients of (a) δm. (b) the transformed coefficients of δm in the curvelet domain.

2.3.3 Compressive Sensing

Compressive sensing (CS) is a signal processing technique for efficiently recovering a signal by finding a solution to an under-determined linear system. This takes advantage of the signal's sparseness or compressibility in some domain, allowing the entire signal to be reconstructed with fewer samples than the Shannon-Nyquist theorem requires [2],[3]. The number of spatial samples required by CS to achieve a certain accuracy scales logarithmically with the sampling grid size. Thus, in seismic exploration where the number of measurements that need to be obtained (according to Shannon-Nyquist sampling theory) is prohibitively large, the theory of CS is especially invaluable. The success of the dimensionality reduction according to CS hinges on an incoherent sampling strategy where coherent aliases are turned into relatively harmless white Gaussian noise.

2.4 L_1 Regularized, Sparsity-Promoting Least-Squares Migration

In this section we seek an approximate solution for the l_1-regularized least-squares problem by promoting sparsity in a transform domain. We take advantage of the sparsity of the model perturbation in the curvelet domain and define $\delta c = \mathbf{C}\delta m$ the transform of the model residual, where $\mathbf{C}$ is the curvelet transform operator. We now reformulate equation 3 as the following constrained convex optimization problem [7][6]:

$$\underset{\delta c}{\text{minimize}} \ \|\delta c\|_1 \ \ \textit{subject to} \ \ \left\|\delta d - \nabla \boldsymbol{\mathcal{F}}[\mathbf{m}_0, \mathbf{Q}]\mathbf{C}^H \delta c\right\|_F \leq \sigma, \tag{22}$$

with $\mathbf{C}^H$ the inverse of the curvelet transform, σ is an estimation of the noise level and δd as defined in section 2.2. $\nabla \boldsymbol{\mathcal{F}}[\mathbf{m}_0, \mathbf{Q}]$ is the linear Fréchet derivative of the forward operator (or first order Born scattering operator).This operator is generally overdetermined and has a null space caused by complex overburdens. This problem is known as the Basis Pursuit Denoise (BPDN).

Because the objective function formulated from equation 22 is non-differentiable, we do not solve it directly. Instead, this problem is solved with $SPGl_1$ [11]. $SPGl_1$ is a high performance large scale solver for sparsity promoting problems. It alternatively solves a series of non-linear sub-problems known as Least Absolute Shrinkage and Selection Operator (LASSO):

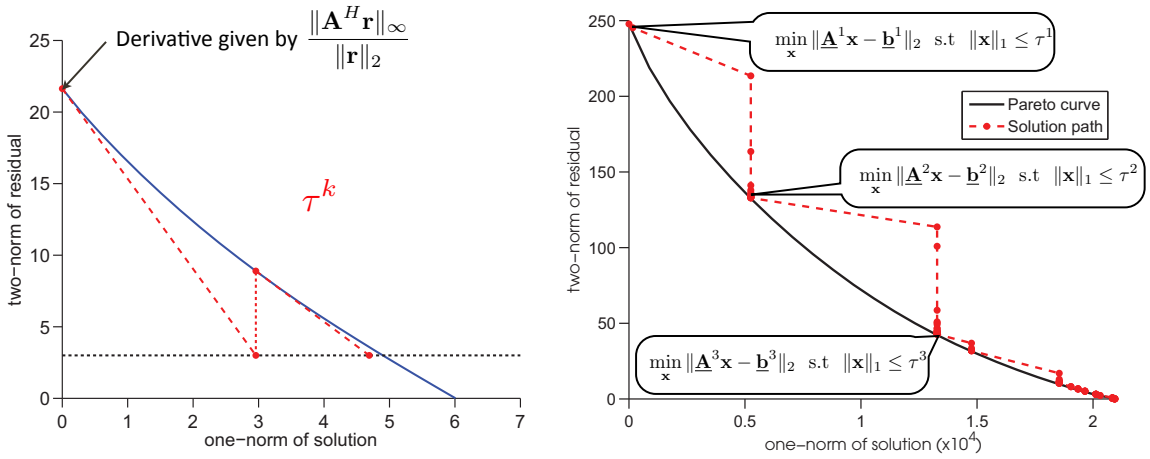
$$\underset{\delta c}{\text{minimize}} \ \frac{1}{2}\left\|\delta d - \nabla \boldsymbol{\mathcal{F}}[\mathbf{m}_0, \mathbf{Q}]\mathbf{C}^H \delta c\right\|_F^2 \ \ \textit{subject to} \ \ \|\delta c\|_1 \leq \tau, \tag{23}$$

where the parameter τ is increased intelligently at each new LASSO solving. In this method, the Pareto boundary (i.e the trade-off curve delineating feasible and infeasible solutions as a function of the l_2 norm of the data misfit and the model's l_1 norm) is exploited to compute the relaxations coefficient τ^k (that causes the LASSO and BPDN to share the same solution) by an inexact Newton root finding method that uses the convexity and smoothness (i.e continuous differentiability) of the Pareto curve [11]. The figures 4a and 4b illustrates this principle and the corresponding solution path(with the notations of [5]). The red dotted line in figure 4b shows a realization of one solve of a BPDN problem, starting with an initial guess of zero. At this stage the reader should consider that the linear systems of figure 4b are the same for every steps and that only the τ's are updated.

Each iteration of the Newton root-finding method requires the approximate evaluation of the Pareto curve at point τ^k. This evaluation involves the minimization of the LASSO problem for τ^k. For solving these individual LASSO problems, $SPGl_1$ uses a Spectral Projected Gradient method (SPG). SPG includes projecting iterates onto the feasible set $\left\{\delta c \mid \|\delta c\|_1 \leq \tau^k\right\}$ via the projection operator:

$$\boldsymbol{\mathcal{P}}_{\tau^k}(v) = \arg\min_{\delta c} \|v - \delta c\|_2 \ \ \textit{subject to} \ \ \|\delta c\|_1 \leq \tau^k \tag{24}$$

which gives the projection of a n-vector v onto the one-norm ball with radius τ^k. Each iteration of the algorithm searches the projected gradient path $\boldsymbol{\mathcal{P}}_{\tau^k}(\delta c_k - \alpha g_k)$, where g_k is the current gradient of the function $\left\|\delta d - \nabla \boldsymbol{\mathcal{F}}[\mathbf{m}_0, \mathbf{Q}]\mathbf{C}^H \delta c_k\right\|_F^2$. The step length α is selected by a non monotone line search algorithm.



(a) **(b)**

Figure 4: $SPGl_1$. **(a) Newton root finding using convexity and smoothness of the Pareto curve [11]. (b) Path followed by solving series of LASSO, the region above the Pareto curve is the feasibility region for the BPDN problem [11].**

Before exploiting some interesting concepts of randomized dimensionality-reduction and divide-and-conquer techniques, we show the results obtained using the same configuration as in section 2.2.

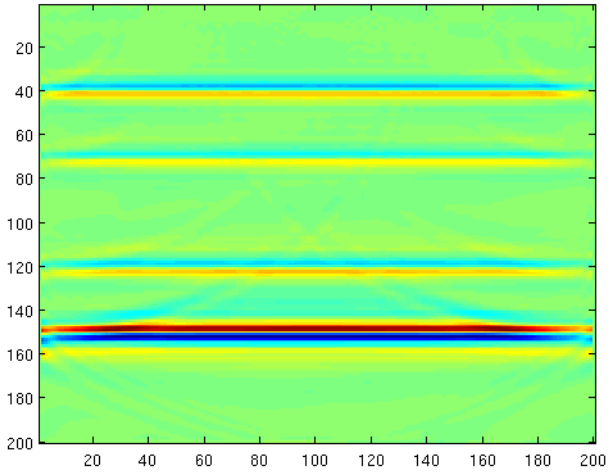


Figure 5: Model perturbation reconstructed with $SPGl_1$ for the initial model of Fig.1

We can see on figure 5 that the perturbation is well recovered even on the edge of the model. We notice that there are less artifacts than in figure 2. The computation time was 5h55min on the cluster, using 2 nodes and 8 ppn.

In the next subsection we present dimension reduction techniques to decrease the problem size and reduce the computational complexity of the least-squares migration. By combining ideas from stochastic optimization and compressed sensing, we take advantage of the sparsity-promoting solver $SPGl_1$ to work on a series of smaller subproblems each involving a randomized subset of data [5][7]. The idea is to work with compressive seismic experiments that consist of collections of small number of simultaneous shots, called supershots (see Fig.6a). These

supershots are made of randomized superpositions of sequential sources.

2.4.1 Randomly weighted simultaneous sources

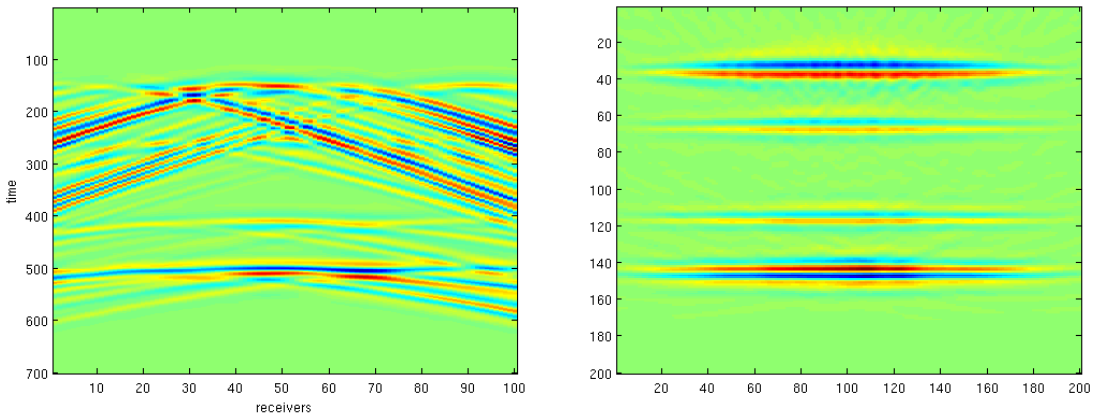
Equation 23 is dimensionally reduced by the subsampling matrices $\mathbf{W}$. We choose $\mathbf{W} \in \mathbb{R}^{(N_s')\times(N_s)}$ to be a Gaussian matrix with $\mathbf{W}\mathbf{W}^H$ near unitary with incoherent off-diagonals. We now replace $\delta d \in \mathbb{R}^{(N_s,N_rN_t)}$ and $\mathbf{Q} \in \mathbb{R}^{(N_s,N_t)}$ by $\mathbf{W}\delta d \in \mathbb{R}^{(N_s'N_rN_t)}$ and $\mathbf{WQ} \in \mathbb{R}^{(N_s'Nt)}$.

Equation 23 may know be reformulated as:

$$\underset{\delta c}{\text{minimize}}\ \frac{1}{2}\left\|\mathbf{W}\delta d - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}_0,\mathbf{WQ}]\mathbf{C}^H\delta c\right\|_F^2 \quad subject\ to \quad \|\delta c\|_1 \leq \tau. \tag{25}$$

By constructing the supershots using $N_s' \ll N_s$ we reduce significantly the problem size, in fact the number of PDE solves for each iteration of the solution of equation 25 is slimmed down by a factor of K'/K [5]. In theory the reduced Born-scattering operator should now be undetermined, however for time stepping formulation the very large size of the time series may cause this operator to still be overdetermined and therefore we may not take full advantage of sparse recovery. This approach takes advantage of the linearity of equation 25 with respect to the source experiments. When multiplying the wavefield residual and the source terms by a Gaussian matrix $\mathbf{W}$, we shape the subsampling artifacts (e.g aliasing and sources cross-talk) into white noise that we can separate from the signal using sparsity-promoting recovery (this idea comes from CS, see e.g [12]). Physically, it's like if we gave a random weight to the amplitude of each source wavelet. The N_s' supershots are then consider as the simultaneous firing of N_s sequential sources (see Fig.6a).

In addition, we take advantage of the way $SPGl_1$ works (see Fig.4b) to redefine the matrix $\mathbf{W}$ each time the Pareto curve is reached and solve a new LASSO problems using the current (curvelet-domain transformed) model residual (hence the upper indices for $\mathbf{A}$ and $\mathbf{b}$ in Fig.4b). This idea, coming from stochastic-average approximations, leads to better results because these renewals remove possible correlations between the sampling matrix and the current estimate of the curvelet-domain bulk perturbation and introduce more information in the migration [5].



(a) Randomly weighted pressure field residual. **(b) Migration with randomly weighted simultaneous shots.**

Figure 6

For the migration we used 5 supershots and run the job on the cluster using 1 node and 6 processors. The computing time was 2h17min. The quality of the results of Fig.6b compare to Fig.5 is good. The computation time is much less than for the classic approach and the amount of cross-talk intoduced during the inversion is reasonably low. We can also see on Fig.7 the improvement brought when doing redraws.

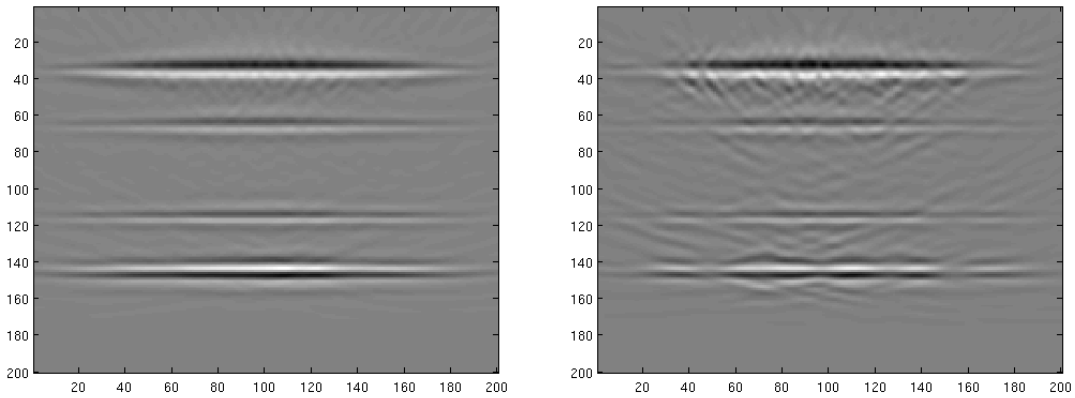


Figure 7: Migration results with weighted supershots with (left) and without (right) redraws.

2.4.2 Randomly phase encoded simultaneous sources

An other possible way of doing compressive experiments is to work with supershots composed of sequential sources with random phase encoding. For $\theta \in [0, 2\pi]^{N_t, Ns, N_s'}$ we define our encoding operator by $\mathbf{P} = \mathcal{F}(\mathcal{R}_e(\mathcal{F}^H(exp(i * \theta))))$ where $\mathcal{F}$ denotes the Fourier transform along time. We keep only the real part of the time transform of the complex exponential to not introduce terms that are incoherent with the physics of the problem. After applying this operator to the source terms and the wavefiel residuals in the Fourier domain and stacking along the sources, we come back in the time domain and take the real part of our signals. We also apply a random sign to each time series measured for a given source/receiver couple. Figures 8 and 9 show exemples of such encoding.

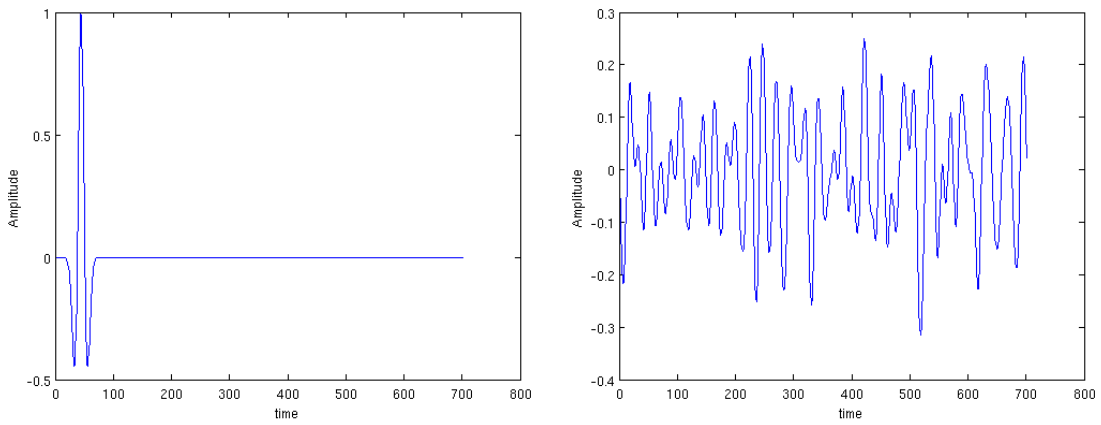


Figure 8: Original and phase encoded Ricker wavelet.

Fig.10a shoes that this way of doing migration can be efficient, very few artifacts are introduced during the inversion and results have reasonable quality when compare with figures 5 ans 2.

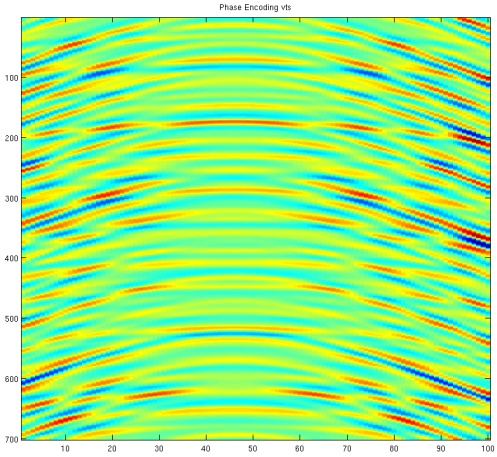
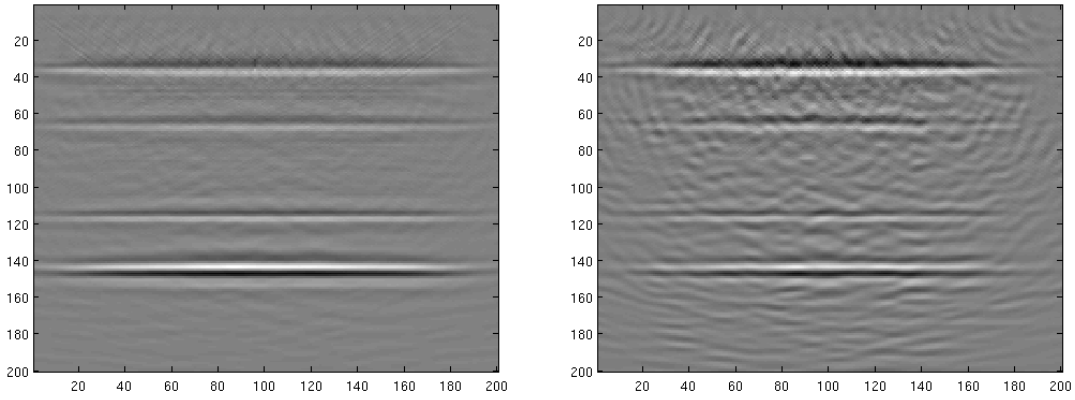


Figure 9: Phase encoded common shot gather.



(a) **(b)**

Figure 10: Migration results by phase encoding with (left) and without (right) redraws.

Unfortunately, these methods relie on fixed-spread acquisition where each source sees the same receivers, limiting this approach to land and ocean bottom acquisitions. In the next section we present a technique which allows dimensionality reduction for marine data.

2.4.3 Random Subsets of Sequential Shots

Instead of working with simultaneous shots, we can work with random subsets of sequential sources [7]. We now use single sources firing at random positions along the profile. We define the restriction matrix $\mathbf{R}$ by the Kronecker product: $\mathbf{R} = \mathbf{R}^{\Sigma} \otimes \mathbf{I}_{N_r} \otimes \mathbf{I}_{N_t} \in \mathbb{R}^{(N'_s,Ns)\times(N_r,N_r)\times(N_t,N_t)}$. $\mathbf{R}^{\Sigma}$ selects $N'_s \ll N_s$ rows uniform randomly amongst $[1...N_s]$. The same way we redraw the subsampling matrix for the construction of the supershots, we choose a new subset of sequential shots each time the Pareto curve is reached and a new BP problem is being solved (see Fig.4b).

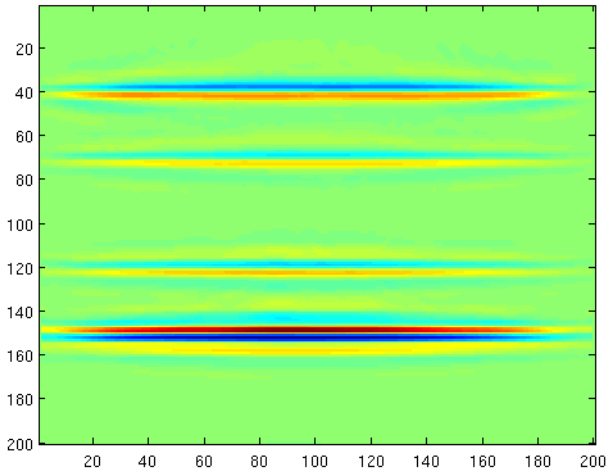
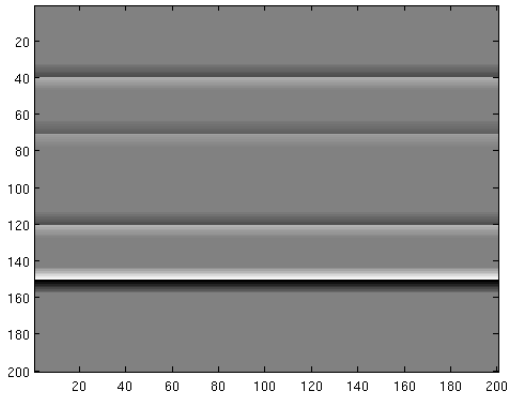


Figure 11: Migration with random subsampling.

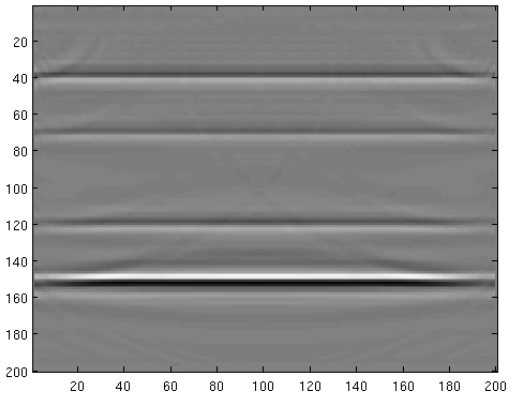
Using the model described for the figure 1 we run the algorithm using 5 shots (that is 25% of the total number) fired at random locations on the profile. The computing time with 1 nodes and 6 ppn was 2h29min. Compare to the migrated image obtained using all the shots (Fig.5 and 2), the recovering of the model perturbation in this case can be considered as good, with a computation much faster.

2.5 Comparison of the results

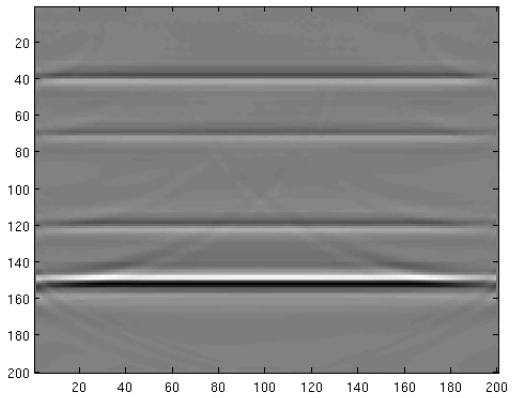
All the images of figure 12 are plotted with the same color scale. We first see that the results are similar for the classic migration solved with a conjugate gradient method (Fig.12b) and the reformulated migration with sparsity constrain using all the shots (Fig.12c). The migrated images are close to the true perturbation but we note that there are some strong artifacts that could be problematic if performing a migration on a complicated model. In figure 12d is plotted the migration done with random shot locations. The results are good even if the amplitude recovery is a beat low. We can even see an improvement in term of quality compare to the migration with all the shots since most of the strong artifacts have disappeared. The migration with simultaneous random shots in figures 12e and 12f also give good results even if the top part of these images are a bit noisy.



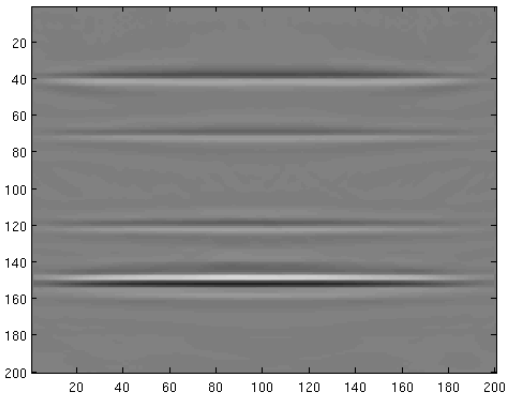
(a) True perturbation



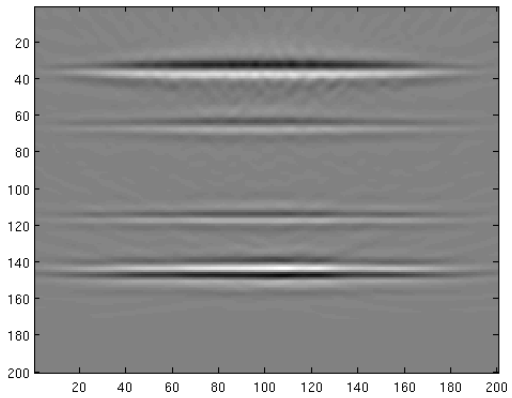
(b) Conjugate gradient method



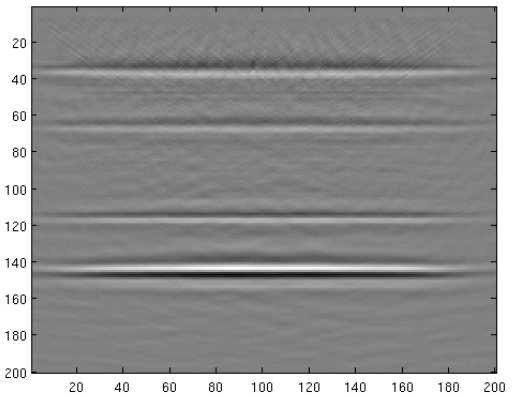
(c) SPG method with all the shots



(d) SPG method with random shots



(e) SPG method with weighted supershots

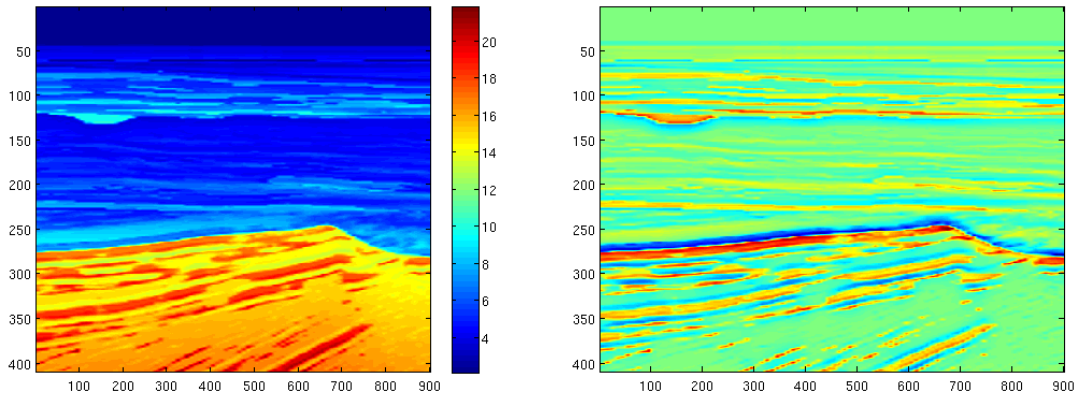


(f) SPG method with phase encoded supershots

Figure 12: Migration results for a simple 1D model.

3 Results with the BG Compass model

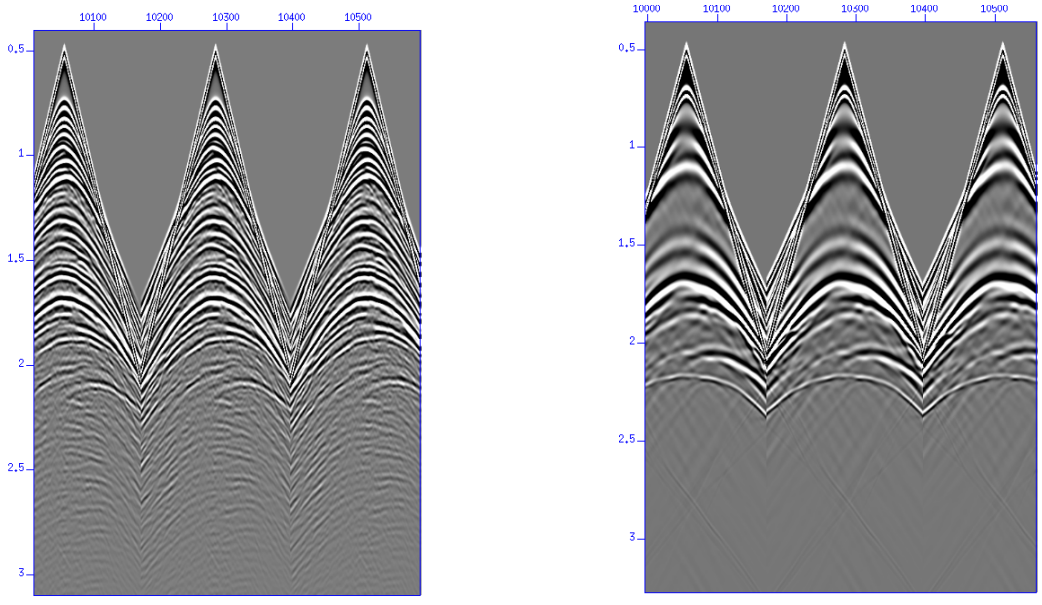
The velocity model Compass was provided to SLIM by the British Gas.



(a) **(b)**

Figure 13: (a) Compass acoustic model. (b) True perturbation.

The simulations are done with 91 shots and 226 receiver positions sampled at 50m and 20m interval, yielding a maximum offset of 4550m. The source pulse are modeled using a Ricker wavelet with a central frequency at 20Hz, and the simulations are carried out over 4s with a time sample of 0.004s. The grid is 901×409 with a grid size of 5m.

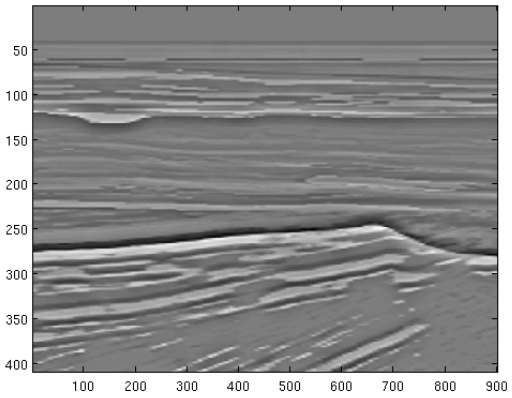


(a) **(b)**

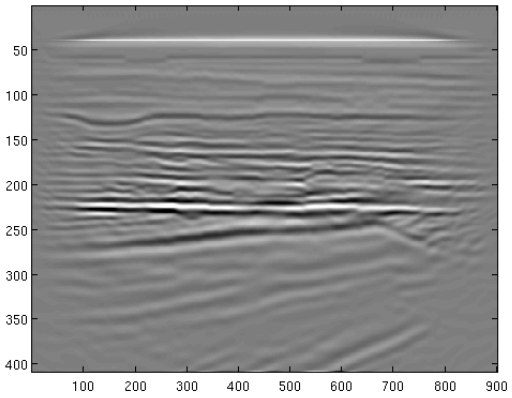
Figure 14: Synthetic shot gathers obtain with the (a) true velocity model. (b) Smooth model.

The model we use here is a 2D slice taken from the 3D synthetic one. It contains details in all dimensions at many spatial frequencies from seismic down to log scale. A variety of dipping and faulted beds where included as well as a gas cloud. As we are still working to improve the workflow, we present here temporary results. For quality improvement, we do some post-

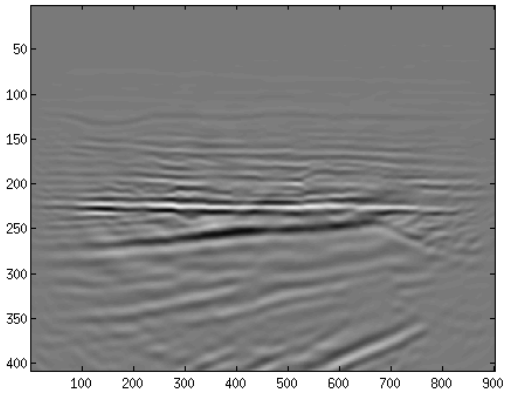
processing such as taking the vertical derivative, doing depth weighting and mute coarse scale curvelet coefficients (for attenuation of low frequency artifacts introduced by curvelets).



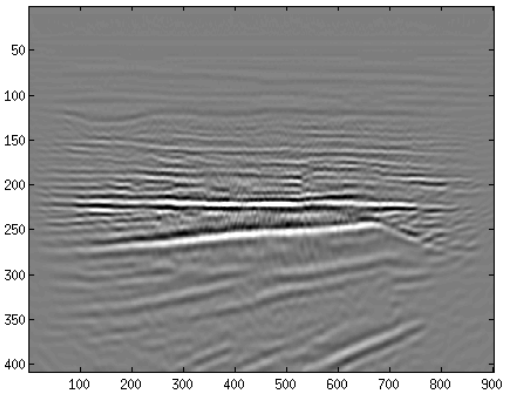
(a) True perturbation



(b) Reverse Time Migration



(c) Migration with 15 random shots with redraws



(d) 15 weighted supershots with redraws

Figure 15: Migration results for the BG model.

4 Conclusion

During this internship a framework to perform time stepping least-squares migration with IWAVE has been built. By following previous work done in the frequency domain, we have tried different dimension reduction methods to decrease the computational burden of wave equation based linearized imaging while aiming for high resolution. By combining random sub/compressive sampling with curvelets properties, we reduce the size of the problem and aim to concentrate information about seismic events in few large number of coefficients while distributing coherent noise uniformly over numerous small coefficients. Then, by using a sparsity promoting solver we restore the amplitudes of geological reflectors while removing crosstalk and aliasing artifacts. Additionally, we highlight the improvement brought when renewing the random sampling after each sub-problem is solved in order to remove possible correlations between the subsampling and the model perturbation update. Both randomized subsets and compressive simultaneous experiments led to encouraging results while significantly decreasing the computation time and the memory requirement. Results obtained on the BG model are temporary and we will continue working on this project. Other methods such as continuous random-noise sources experiments are to be investigated. It will also be interesting to do a careful comparison between time-harmonic and time LSM dimension reduction techniques.

References

[1] Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006.

[2] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[3] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[4] Felix J Herrmann, Michael P Friedlander, and Ozgur Yilmaz. Fighting the curse of dimensionality: compressive sensing in exploration seismology. *Signal Processing Magazine, IEEE*, 29(3):88–100, 2012.

[5] Felix J. Herrmann and Xiang Li. Efficient least-squares imaging with sparsity promotion and compressive sensing. *Geophysical Prospecting*, 60(4), 2012.

[6] Felix J. Herrmann, Xiang Li, Aleksandr Y. Aravkin, and Tristan van Leeuwen. A modified, sparsity promoting, Gauss-Newton algorithm for seismic waveform inversion. In *Proc. SPIE*, volume 2011, 08 2011.

[7] X. Li, A. Aravkin, T. van Leeuwen, and F. Herrmann. Fast randomized full-waveform inversion with compressive sensing. *GEOPHYSICS*, 77(3):A13–A17, 2012.

[8] Gerhard Pratt, Changsoo Shin, et al. Gauss–newton and full newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362, 1998.

[9] William W. Symes, Dong Sun, and Marco Enriquez. From modelling to inversion: designing a well-adapted simulator. *Geophysical Prospecting*, 59(5):814–833, 2011.

[10] Albert Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984.

[11] E. van den Berg and M. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2009.

[12] Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann. Seismic waveform inversion by stochastic optimization. *International Journal of Geophysics*, 2011, 12 2011.

[13] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *GEOPHYSICS*, 74(6):WCC1–WCC26, 2009.

Appendices

A - Spectral projected gradient for LS_τ [11]

Algorithm 1 $SPGl_1$, for solving $\min_x \|Ax-b\|_2$ s.t $\|x\|_1 \leq \tau$

Require: x, τ, δ

Set minimum and maximum step lengths $0 < \alpha_{min} < \alpha max$

Set initial step length $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$ and sufficient descent parameter $\gamma \in (0,1)$

Set an integer line search history length $M \geq 1$

Set initial iterates: $x_0 \leftarrow \mathbf{P}_\tau[x], r_0 \leftarrow b - Ax_0, g_0 \leftarrow -A^T r_0$

$l \leftarrow 0$

while not converged **do**

$\delta_l \leftarrow \|r_l\|_2 - (b^T r_l - \tau \|g_l\|_\infty)/\|r_l\|_2$ // *Compute duality gap*

if $\delta_l < \delta$ **then**

break // *exit if converged*

end if

$\alpha \leftarrow \alpha_l$ // *initial step length*

while not converged **do**

$\overline{x} \leftarrow \mathbf{P}_\tau[x_l - \alpha g_l]$ // *candidate line search iterate*

$\overline{r} \leftarrow b - A\overline{x}$ // *update the corresponding residual*

if $\|\overline{r}\|_2^2 \leq \max_{j \in [0, min\{k, M-1\}]} \|r_{l-j}\|_2^2 + \gamma(\overline{x} - x_l)^T g_l$ **then**

break // *exit line search*

else

$\alpha \leftarrow \alpha/2$ // *decrease step length*

end if

end while

$x_{l+1} \leftarrow \overline{x}, r_{l+1} \leftarrow \overline{r}, g_{l+1} \leftarrow -A^T r_{l+1}$ // *update iterates*

$\Delta x \leftarrow x_{l+1} - xl, \Delta g \leftarrow g_{l+1} - gl$

if $\Delta x^T \Delta g \leq 0$ **then** // *update the Barzilai-Borwein step length*

$\alpha_{l+1} \leftarrow \alpha_{max}$

else

$\alpha_{l+1} \leftarrow min\left\{\alpha_{max}, max[\alpha_{min}, (\Delta x^T \Delta x)/(\Delta x^T \Delta g)]\right\}$

end if

$l \leftarrow l + 1$

end while

return $x_\tau \leftarrow x_l, r_\tau \leftarrow r_l$

B - Dimensionality-reduced LSM with sparsity promotion (adapted from [7])

Algorithm 2 Dimensionality-reduced LSM with sparsity promotion

$k \leftarrow 0$; $\delta c^k \leftarrow \delta c_0$ // *initial perturbation (often 0)*

while not converge **do**

$\left\{\mathbf{D}^k, \mathbf{Q}^k\right\} \leftarrow \{\mathcal{S}\mathbf{D}, \mathcal{S}\mathbf{Q}\}$ with $\mathcal{S}$ random sampler // *indep. draw.*

$\delta d^k \leftarrow \mathbf{D}^k - \nabla \boldsymbol{\mathcal{F}}[\mathbf{m}, \mathbf{Q}^k]$ // *residual*

$\tau^k \leftarrow \left\|\delta d^k\right\|_F / \left\|\mathbf{C}\nabla \boldsymbol{\mathcal{F}}^*[\mathbf{m}, \mathbf{Q}^k]\delta d^k\right\|_\infty$ // *update* τ

$\delta c^k \leftarrow \arg\min_{\delta c^k} \frac{1}{2}\left\|\delta d^k - \nabla \boldsymbol{\mathcal{F}}[\mathbf{m}, \mathbf{Q}^k]\mathbf{C}^H \delta c^k\right\|_F^2$ subject to $\left\|\delta c^k\right\|_1 \leq \tau^k$

$\delta c^{k+1} \leftarrow \delta c^k$

$k \leftarrow k + 1$

end

$\delta m = \mathbf{C}^H \delta c$ // *from curvelet domain to model space*