Duality Methods for Waveform Inversion

Guy Chavent

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Duality Methods for Waveform Inversion

Guy Chavent

N ° 2975

September 11, 1996

THEME 4

ISSN 0249-6399

INRIA
ROCQUENCOURT

Duality Methods for Waveform Inversion *

Guy Chavent *

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet ESTIME

Rapport de recherche n°2975 — September 11, 1996 — 28 pages

Abstract: We show in this paper how the application of convex duality leads to various reformulations of the classical non optimizable least squares approach to waveform prestack inversion. These reformulations search all for i) a background velocity model and ii) a reflectivity model defined in the time domain, and linked to the usual depth reflectivity model by prestack depth migration. We show that the data themselve are a good approximation of this optimal time reflectivity unknown when a quantitative migration is used.

Hence duality provides a synthetic view of various approaches to prestack inversion, building up a bridge between concepts as different a priori as waveform inversion by minimization of data misfit and migration velocity analysis.

From a pratical point of view, it provides objective functions for the determination of the velocity background (data misfit for the Migration Based Travel Time reformulation, norm of the stack of the prestack migrated sections for the Multiple Migration reformulation), whose evaluation and optimization by local gradient technique is now feasible, thus eliminating the need for travel time picking or analysis of coherency panels.

Key-words: duality, seismic inversion, migration, inverse problem

(Résumé : tsvp)

Manuscript prepared for the Svenfest held in Golden, Colorado, March 6-8 1994.

* CEREMADE, Université de Paris-Dauphine, pl. du Maréchal De Lattre De Tassigny, F-75775 Paris Cédex 16, and INRIA-Rocquencourt, Domaine de Voluceau, B.P. 105, F-78153 Le Chesnay Cédex.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Méthodes de Dualité pour l'inversion sismique

Résumé : Nous montrons dans ce papier comment la dualité convexe permet de reformuler le problème classique de l'inversion sismique par moindres carrés, qui est connu pour être non optimisable. Ces reformulations cherchent toutes à déterminer i) un modèle de réflectivité défini dans le domaine temps, et lié à la réflectivité habituelle en profondeur par un opérateur de migration. La dualité permet ainsi d'avoir une vue synthétique des diverses approches de l'inversion avant sommation, montrant par exemple le lien entre des concepts à priori aussi différents que l'approche moindres carrés et l'analyse des vitesses de migration. D'un point de vue pratique, elle fournit des fonctions objectif pour la détermination des vitesses de référence (erreur sur les données pour la reformulation en Temps de Parcours Basé sur la Migration, norme de la somme des sections migrées avant sommation pour la reformulation par Migration Multiples), dont le calcul numérique et l'optimisation par des méthodes de gradient est aujourd'hui possible, ce qui élimine le recours au pointé d'horizons sismiques et à l'analyse des panneaux de cohérence.

Mots-clé : dualité, inversion sismique, migration, problème inverse

1 Introduction

The determination of the earth structure from the recording of reflected or transmitted seismic data is one of the important goals of the geophysic. Classically, this determination is performed by applying, usually iteratively, imaging tools and background determination tools.

The objective of this paper is to formulate methods for full waveform inversion, which solve simultemeously the imaging and background determination problems, and are both feasible and cost effective, as opposed to least squares inversion à la Tarantola [22] which suffers from non optimizability, and to the Differential Semblance Optimization (DSO) method of Symes [19] which suffers from being computationaly expensive.

Before we develop this further and present our approach, let us fix some terminology, which we will use throughout the paper :

- we call reflectivity (and denote by r) the vector of parameters describing the part of the earth model which, in the chosen forward modelisation, is responsable mainly for the reflection and diffraction of energy. For example, if an acoustic model with fixed density is used, the parameter r describes the high frequencies component of the slowness (called also "rough slowness").
- we call propagator (and denote by π) the vector of parameters describing mainly the kinematic of energy propagation. For example, in an acoustic model, the parameter π describes the low frequencies component of the slowness (called also the "smooth slowness"). But it could also describe the smooth component of the velocity ! This propagator is often referred to as the "background" in the geophysical literature, but we shall prefer the term "propagator" in the sequel, as it is much more descriptive of the physical effects this parameter is associated with.

The reflectivity and the propagator are usually determined separately :

- for a given propagator, depth-migration operators produce a good image of the reflectivity r (in the sense that the reflectors in the image are located at the same place as in r, but with completely wrong amplitudes if no special care is taken) provided the correct propagator is used.
- the classical tools for the determination of the propagator π fall roughly into three classes :
  - stacking velocity analysis (NMO)
  - (reflection) tomography, wich tries to determine the propagator by adjusting the kinematic of picked (reflection) events.
  - migration velocity analysis, where the propagator is adjusted, usually manually, by requiring that images for the different shots or offset are as close as possible (coherency panels).

In many cases, a clever combination of these two techniques allows a good determination of the earth structure. However, when this latter becomes more complex (salt dome, fractures...), travel time picking for example becomes difficult, and the classical techniques may fail.

So in the 80' s, when the development of the power of computers made it clear that solving numerically the acoustic equation would become more and more a routine step, great expectations were placed in the least squares waveform inversion approach, which conceptually solved at the same time the imaging and propagator determination problems, and could in principle work as well for complex models :

$$Min_\nu J(\nu) = \frac{1}{2} \parallel d - c_\nu \parallel^2 \qquad (LS)$$

where d is the collection of seismic data (common shot gather, or common offset gather...) to be inverted, and where c_ν is made of the corresponding synthetics computed from ν using the chosen forward model. This approach has been proposed as early as 1979 (cf [1]) for 1-D problems, and 1982 (cf [22]) for multidimensional problems.

It has however turned out since, that despite the fact that the formulation (LS) is perfectly satisfying from a conceptual point of view, it is not usable from an operational point of view, as the least-squares objective function $J(\nu)$ possesses many parasitic local minima, and a very narrow domain of attraction for the global minimum, and which becomes smaller when the frequency of the data increases.

So in attempts to solve numerically (LS) by mean of a local (gradient) optimization algorithm, the error in the current slowness leads the algorithm to stop prematurely in a local minimum. Convergence to the global minimum is ensured only for initial guesses of slowness which are close enough to the true slowness so that they generate in the synthetics phase shifts smaller than a quarter of a wavelength. Such initial guesses are of course practically out of reach. Recently an attempt has been made (cf [3]) to "squeeze" the largest amount of information out of (LS) by generalizing to 2-D medium the frequency continuation technique introduced in [13] for tabular 1-D medium. However this rehabilitation of the least-square formulation (LS) as an operational method for seismic data inversion has limitations due to the lack of very low frequency components in the data. Simulated annealing and genetic algorithms have also been tested (cf. [16], [17]) in order to overcome the problem of local minima, but their use is limited by the large number of function evaluation (and hence of forward simulation) which they require.

There are two explanations for the difficulties encountered in the minimization of the least-squares objective function $J(\nu)$:

the phase-shift problem : an error in the low frequency component of ν causes phase-shift in the synthetic time sections. As soon as this phase-shift becomes larger than a half-cycle, the least squares algorithm will try to match together parts of the synthetics and data which differ by one or more cycle skip, leading thus the algorithm to stop in one meaningless local minimum.

the destructive interference problem : the gradient of J with respect to ν is known to be the stack of all migrated shotgathers with the current estimate of the slowness ν (cf (5) below and paragraph 4). When the reflectivity of the medium under investigation is rich enough, the migrated shot gather will stack incoherently and produce a "white noise" type section, unless the smooth part of the current slowness (propagator) is very close to the true one. Hence the gradient section carries little useful information on how to update the slowness.

In this paper, we focus on the solution of the phase shift problem. So we can expect that the proposed methodology will be effective on problems for which destructive interference is not too severe, for example when the reflectivity pattern is simple in term of number of reflectors. This was confirmed numerically for the algorithms derived in paragraphs 4 and 5 ([8], [10], [6], [7], [15]). In the last years, a few approaches have been proposed to overcome part or all of the above difficulties ([8], [10], [6], [19], [18], [7]). All the above mentioned approaches take into account the fact that, because of the lack of low frequencies of the source, the information on the long wavelength of the slowness ν is available in the surface data d only through the kinematic of the reflected energy, and they all give a different mathematical treatment to the propagation and reflection effects. This requires to replace the search for the slowness ν by the search for the "propagator unknown" π (which describes the kinematic of energy propagation) and a "reflectivity unknown" r (which is responsible for the reflection of energy).
So the first step of our construction consists in rewriting the least-squares problem (LS) in term of the (propagator, reflectivity) unknows :

$$Min_{\pi \in C, r} J(\pi, r) = \frac{1}{2}\|d - c_{\pi,r}\|^2 \qquad (P)$$

where we have made explicit the constraint that π should belong to a set C of admissible propagators, and where $c_{\pi,r}$ is the synthetic computed using the chosen forward model with propagator π and reflectivity r. By construction of π and r, an earth with $r = 0$ should produce no sizeable reflections, hence we shall suppose throughout the paper that :

$$c_{\pi,r} = 0 \quad \text{at } r = 0. \qquad (1)$$

Another advantage of the (propagator, reflectivity) choice of unknows is that, for a fixed propagator π, the synthetic $c_{\pi,r}$ depends almost linearly from r , the sole non-linearity being caused by the multiple reflections. Hence the linearized model around a given propagator π and a zero reflectivity :

$$c_{\pi,r} = B(\pi)r \qquad (2)$$

with

$$B(\pi) = \frac{\partial c_{\pi,r}}{\partial r} \quad \text{at } (\pi, r = 0) \qquad (3)$$

is in many circumstances a useful approximation to the full nonlinear model, and the objective function $J(\pi, r)$ is almost quadratical with respect to the reflectivity r.
Because of the lack of illumination of certain parts of the earth, the linearized forward model $B(\pi)$ is usually not injective, or at least extremely ill-conditioned. So in order to stabilize the "easy part" of problem (P), i.e. the minimization with respect to r for a fixed π, one is naturally led to replace (P) by its regularized version

$$Min_{\pi \in C, r} J_\varepsilon(\pi, r) = \frac{1}{2}\|d - c_{\pi,r}\|^2 + \frac{\varepsilon^2}{2}\|r\|^2 \qquad (P_\varepsilon)$$

where $\varepsilon > 0$ is the regularization parameter (this is a way to compute an approximation of the minimum-norm reflectivity solution of (P)).

Of course, problems (P) and (P_ε) are still as poorly behaved with respect to π as (LS) was with respect to ν : a small error in the propagator π produces, for a fixed depth reflectivity r, large phase shifts in the synthetics $c_{\pi,r}$, which prevent the actual solution of (P) and (P_ε) by local optimization methods.

A first solution to the phase-shift problem would be to replace the minimization of $J_\varepsilon(\pi, r)$ by that of the redudred objective function

$$J_\varepsilon^\sharp(\pi) = J_\varepsilon(\pi, \hat{r}_\varepsilon(\pi)) \qquad (4)$$

where $\hat{r}_\varepsilon(\pi)$ is the minimizer of $J_\varepsilon(\pi, r)$ with respect to r for the given propagator π. It has been shown (cf [18]) that this new function tends to be, at least for noiseless data, unimodal, so that its minimization by local gradient techniques is feasible. However, because of the inner optimization loop involved in (4), the evaluation of $J_\varepsilon^\sharp$, and a fortiori its minimization, are computationally expensive. This solution to the phase shift problem, together with a special choice of the reflectivity space and regularization term aimed at overcoming the destructive interference problem, has been successfully used in the Differential Semblance Optimization method (cf [19] [18]).

One other way to overcome the phase-shift problem would be to have the reflectivity unknown not defined, as r, in the depth domain (where the "right" location of a reflector depends on the availability of the "right" propagator), but rather in the time domain (where the "right" location of a reflection is expected to be intrinsically related to the data themselves, and independent of the value of the propagator). The advantages of looking at reflectors in the time domain was recognized very early in the preferred use of "time migration", and more recently in tomography.

A natural way to associate to a time reflectivity section s (which lives in the same space as data the d, and hence contains as many subsections $s_1 ... s_N$ as shots or offsets...) a depth reflectivity section r is to use the transposed $B(\pi)^*$ of the linearized forward model, as it was recognized by Lailly (cf [14]) and Tarantola (cf [21]) that :

$$r = B(\pi)^* s \qquad (5)$$

is the stack of the migrated sections of each shot or offset obtained using the propagator π. Replacing r by s is the equivalent, for a multidimensional problem, of the travel-time change of variable which, in the one-dimensional case, is known to make the solution of (P) numerically feasible (cf [1]). This Migration-Based Travel Time change of unknown (5) was the intuitive basis of the MBTT reformulation of the least-square problem (P) (cf [10]), which has proved to be both legitimate and numerically efficient as it allowed the determination of the propagator π by local gradient techniques (cf [6] [7] [15]).

We show in paragraph 2 of this paper how the theory of convex duality (cf [11]) gives a systematic way of introducing such a time reflectivity unknown s. It leads to the definition of a dual objective function

$$DSI_\varepsilon(\pi, s) = \frac{1}{2}\|B(\pi)^* s\|^2 - (s, d) + \frac{\varepsilon^2}{2}\|s\|^2, \tag{6}$$

which we call the Dual Similarity Index, as we see from (5) that its first term is the squared norm of the stack of the individually migrated subsections of the reflectivity unknown s, and hence is a measure of the similarity or coherency of these migrated sections.

For a given π, the minima of J with respect to r and DSI with respect to s are related by

$$Min_r J(\pi, r) + \varepsilon^2 Min_s DSI(\pi, s) = 0 \tag{7}$$

with the minimizers $\hat{r}_\varepsilon$ and $\hat{s}_\varepsilon$ satisfying (5) i.e. :

$$\hat{r}_\varepsilon = B(\pi)^* \hat{s}_\varepsilon. \tag{8}$$

Hence we are led naturally to define a new dual problem

$$Max_{\pi \in C} Min_s DSI(\pi, s) = \frac{1}{2}\|B(\pi)^* s\|^2 - (s, d) + \frac{\varepsilon^2}{2}\|s\|^2 \qquad (DP_\varepsilon)$$

which, in sight of (7)(8), is equivalent to (P_ε) in the sense that

$$(P_\varepsilon) \Longleftrightarrow \{(DP_\varepsilon) \quad \text{followed by} \quad \hat{r}_\varepsilon = B(\hat{\pi}_\varepsilon)^* \hat{s}_\varepsilon\} \tag{9}$$

This is our first main result, namely that the numerically intractable least-squares problem (P_ε) is equivalent to the new dual problem (DP_ε) stated in term of propagator and time reflectivity unknowns, and hence more amenable to numerical solution by local optimization techniques.

We shall then see, in paragraph 3, how to define, via a proper scaling, the parameters r which describe the high frequency part of the slowness in such a way that $B(\pi)^*$ is a quantitative migration operator (cf [4] , [9]), which restores at best the absolute amplitudes of migrated events, and allows to re-simulate at best the part of the data which is in the

range of the forward model. A quantitative migration satisfies hence

$$\left\{ \begin{array}{l} B(\pi)\ B(\pi)^* \simeq P_\pi \\ \text{where } P_\pi = \text{ projection on } Im(B(\pi)) \text{ in the data space.} \end{array} \right. \tag{10}$$

In that case, we will show that the data d themselves are a good approximation of the optimal dual unknown $\hat{s}_\varepsilon$:

$$\hat{s}_\varepsilon \simeq d. \tag{11}$$

This is our second main result : the simultaneous use of a dual time reflectivity unknown and a quantitative migration allows to replaces the search for $(\hat{\pi}_\varepsilon, \hat{r}_\varepsilon)$, both of them being unknown, by the search for $(\hat{\pi}_\varepsilon, \hat{s}_\varepsilon)$ where $\hat{s}_\varepsilon$ is almost already known !

Then we will see in paragraph 4 that the dual problem (DP_ε) is very closely related to a migration velocity analysis approach based on a global coherency measure defined as the norm of the stack of the migrated sections (as in the Multiple Migration Fitting method of [8] and the nonlinear inversion for velocity of [12]). This shows that two apparently unrelated approaches for the determination of the propagator or background π, namely least squares waveform inversion (problem (P) or (P_ε)) and migration velocity analysis are in fact mathematically dual methods. It provides a mathematical basis for migration velocity analysis, and suggests one specific coherency criterion (the norm of the stack of migrated section for the parametrization of the reflectivity which provides a quantitative migration) for which the equivalence with waveform inversion will hold.

In paragraph 5, we investigate the relations with the Migration-Based Travel Time approach of [10] and [7], which consists in plugging the new time reflectivity unknown s in the old formulation (P).

Finally, we shall see in paragraph 6 how the duality results produce a new formula for the evaluation of the objective function of the Differential Semblance Optimization of [19].

2 Duality and first main result

We postpone to the next paragraph the precise definition of the reflectivity and propagator unknows r and s, and concentrate in this paragraph on the mathematical construction of the dual problem. We shall call :

$$\left\{ \begin{array}{lcll} R & = & \text{space of reflectivities} & r \\ \Pi & = & \text{space of propagators} & \pi \\ C \subset \Pi & = & \text{set of physically admissible propagators,} & \\ D & = & \text{space of data} & d \end{array} \right. \tag{12}$$

and suppose throughout this paper that

$$\begin{cases} R, D, \Pi & \text{are finite dimensional Hilbert spaces} \\ C & \text{is a closed bounded convex subset of } \Pi \end{cases} \qquad (H1)$$

where the finite dimensionality hypothesis is made in order to avoid technical difficulties (it will be at least satisfied by the discretized problem !), and where C denotes a subset of physically admissible propagators π, which accounts for lower and upper bounds on the slowness for example.
In order to be able to compute explicitly the dual formulation of (P_ε) , we shall use the linearized approximation (2), (3) of the forward model, and suppose that :

$$B \in \mathcal{C}^1(\mathbf{C}; \mathcal{L}(R, D)) \qquad (H2)$$

where $\mathcal{C}^1$ demotes the space of continuously derivable functions, and $\mathcal{L}(R, D)$ the space of (continuous) linear applications from R to D. Hypothesis $(H2)$ will be satisfied by discretized problems. We suppose also that we are given a regularization parameter

$$\varepsilon > 0 \qquad (H3)$$

for the approximate selection of the minimum norm solution r of (P)
Then we have the :

Proposition 1. *Under hypothesis (H1), (H2), (H3) the linearized regularized problem* (P_ε) *admits (possibly many) solutions* $(\hat{\pi}_\varepsilon, \hat{r}_\varepsilon)$. *If there exist two distinct solutions, they differ necessarily by their propagator* π.

Proof. for any given $\pi \in C$, the minimization of (P_ε) with respect to r is a quadratical problem, which has a unique solution $\hat{r} = r(\pi)$ depending continuously on $B(\pi)$, and hence on π by hypothesis $(H2)$. Solving (P_ε) amounts then to minimize the continuous function $J_\varepsilon(\pi, \hat{r}(\pi))$ on the compact set C, which is known to have at least solution. □

Application to (P_ε) of the Fenschell Rockfellar theorem on convex duality (see Appendix) leads to :

Proposition 2. *(First main result : Dual formulation)*
Let hypothesis $(H1), (H2), (H3)$ *hold. Then the original problem*

$$Min_{\pi \in C} Min_r J_\varepsilon(\pi, r) = \frac{1}{2}\|d - B(\pi)r\|_D^2 + \frac{\varepsilon^2}{2}\|r\|_R^2 \qquad (P_\varepsilon)$$

and the dual problem

$$Max_{\pi \in C} Min_s DSI_\varepsilon(\pi, s) = \frac{1}{2}\|B(\pi)^* s\|^2 - < s, d >_D + \frac{\varepsilon^2}{2}\|s\|_D^2 \qquad (DP_\varepsilon)$$

admit both solutions, and are equivalent in the sense that, for any $\pi \in C$:

$$Min_r J_\varepsilon(\pi, r) + \varepsilon^2 Min_s DSI_\varepsilon(\pi, s) = 0 \qquad (13)$$

which implies that, if $(\hat{\pi}_\varepsilon, \hat{r}_\varepsilon)$ *is a solution of* (P_ε)*, then there exists a* $\hat{s_\varepsilon}$ *such that* $(\hat{\pi_\varepsilon}, \hat{s_\varepsilon})$ *is solution of* (DP_ε)*, and conversely. For a given triple* $(\hat{\pi_\varepsilon}, \hat{r_\varepsilon}, \hat{s_\varepsilon})$ *of solutions of* (P_ε) *and* (DP_ε) *the duality relations read :*

$$\left\{ \begin{array}{l} \hat{r_\varepsilon} = B(\hat{\pi_\varepsilon})^* \hat{s_\varepsilon} \\ \varepsilon^2 \hat{s_\varepsilon} = d - B(\hat{\pi_\varepsilon}) \hat{r}_\varepsilon \end{array} \right. \tag{14}$$

We first take a look at the objective function $DSI_\varepsilon(\pi, s)$ of the new dual problem (DP_ε). Its first term is the norm of the stack of individually migrated shot or offset gathers (cf (5)), which is the similarity index

$$SI(\pi) = \frac{1}{2} \|B(\pi)^* s\|^2 \tag{15}$$

used in the Multiple Migration Fitting (MMF) approach to migration velocity analysis (cf. [8]). The other terms depend only on s, not on π. This is why we have called "Dual Similarity Index"the objective function of (DP_ε).

The question arises also of the limit, when $\varepsilon \to 0$, of a solution $\hat{\pi}_\varepsilon, \hat{r}_\varepsilon, \hat{s}_\varepsilon$ of (P_ε) and (DP_ε). A natural candidate for the limit is $(\hat{\pi}, \hat{r}, \hat{s})$ where $(\hat{\pi}, \hat{r})$ is a solution of (P) such that $\hat{r}$ is of minimum norm. Using the hypothesis that the data d is attainable with the linearized model for some propagator, and that $\hat{r}$ is of minimum norm $\hat{\pi}, \hat{r}$ and $\hat{s}$ satisfy :

$$\left\{ \begin{array}{l} B(\hat{\pi})\hat{r} = d, \\ \hat{r} = B(\hat{\pi})^* \hat{s}, \\ \hat{s} \in ImB(\hat{\pi}). \end{array} \right. \tag{16}$$

The convergence of $(\hat{\pi}_\varepsilon, \hat{r}_\varepsilon, \hat{s}_\varepsilon)$ to $(\hat{\pi}, \hat{r}, \hat{s})$ will be studied in the next paragraph.

We have shown in the lower part of figure 1 (with the switches set to $L =$ "linear") the block-diagram for the solution of the dual problem (DP_ε). The large boxes on the left and right sides are the local optimization subroutines for the maximization with respect to the propagator π (left) and the minimization with respect to the time reflectivity s (right). The convention in this diagram is that letters in small (duplicated) squares correspond to time sections $(d, c, s,$ etc...) ; letters in small circles correspond to depth sections $(\pi, m,$ etc...) and letters in small diamonds correspond to real numbers $(J, SI,$ etc...). The current propagator π proposed by the left optimizer is used as slowness ν in the Direct Migrations box, which is in charge of producing a quantitative migrated depth section $m = \nabla_r J = B(\pi)^* s$ of the current time reflectivity section s proposed by the right optimizer.

We describe shortly here the operations which are performed in the Direct Migrations box. For each shot or offset, the source is forward propagated with the (smooth) slowness $\nu = \pi$ (which produces a $c \simeq 0$ synthetic as the slowness background is smooth), the time reflectivity section s is backward propagated with the same background, and the two fields are correllated. These correllations are stacked for all shots or offset, in order to produce $\nabla_\nu J$ at $\nu = \pi$. This depth section (in the circle above the Direct Migration box) is not direcly used in the dual formulations, but it will allow for the minimization of the "old"

least-squares problem (LS) if desired at the final stage of the data inversion. For the dual formulations, the $\nabla_\nu J$ section is first projected onto V_r, then scaled by application of W according to (21) in order to produce $m = \nabla_r J = B(\pi)^* s$ (see paragraph 3).

The norm of m is the similarity index SI (cf (15)) which is the first term of the dual similarity index DSI. The left loop of the figure performs a migration velocity analysis by maximization of SI with respect to the propagator π for a fixed time reflectivity s. Where s is chosen to be the data d themselves, this reduces to the Multiple Migration Fitting algorithm of section 4, where we will provide a more detailed analysis of this left loop.

The right loop of the figure performs the minimization of the dual objective function DSI with respect to the time reflectivity s for a given current propagator π. The DSI objective function is obtained by adding to SI the duality terms (compare (DP_ε) and (15)). The gradient of the duality terms is obvious, and the gradient of SI with respect to s is easily seen to be $B(\pi)m$, which corresponds to the "Forward Linearized Wave Equation" box at the bottom right of the center box.

In the solution of the dual problem (DP_ε) the left loop is the outer one : given a current value of π, the right loop has to be ran to compute $\hat{s}_\varepsilon(\pi)$. By comparing (13) and (4) we obtain :

$$J_\varepsilon^\sharp(\pi) = -\varepsilon^2 DSI_\varepsilon(\pi, \hat{s}_\varepsilon(\pi)), \tag{17}$$

and we see that running the right loop provides a new algorithm for the computation of the reduced objective function $J_\varepsilon^\sharp(\pi)$. Then π has to be updated in order to minimize $J_\varepsilon^\sharp(\pi)$ or equivalently to maximize $DSI_\varepsilon(\pi, \hat{s}_\varepsilon(\pi))$. As $\hat{s}_\varepsilon(\pi)$ is going to be much less depending on π as $\hat{r}_\varepsilon(\pi)$ was, one can expect that this updating can be done simply by using the gradient of DSI_ε with respect to π, for $s = \hat{s}_\varepsilon(\pi)$ fixed, which is the left loop of the figure. On the contrary, it was found numerically [20] that the updating of π by the gradient of $J^\sharp(\pi)$ computed by differentiation of $J_\varepsilon(\pi, \hat{r}_\varepsilon(\pi))$ could not be performed by neglecting the term $\partial J_\varepsilon/\partial r.\ \partial \hat{r}_\varepsilon/\partial \pi$: of course $\partial J_\varepsilon/\partial r$ is small as $\hat{r}_\varepsilon$ is an (approximate) minimizer of J_ε, but $\partial \hat{r}_\varepsilon/\partial \pi$ is very large as perturbating π moves the singularities (reflectors) in the inverted sections $\hat{r}_\varepsilon(\pi)$.

When a detailed enough propagator π has been found, one can replace the linearized model by the full acoustic wave equation by changing the two switches from linear to non-linear, and run the upper loop for the solution of the original problem (LS), as the gradient of J_ε needed for its minimization is obtained as intermediate result in the "Direct Migrations" box as we have seen above.

The least-squares algorithm (LS) has been shown to be able to produce very sharp images under the stringent condition that it is started with slowness profile which provides a very good description of the propagation of energy (cf [3]). Hence performing a few "non linear iterations" with the upper loop may allow to provide the "final touch" to the interpretation of data, by taking into account all multiples for example.

3 Quantitative Migration and second main result

We recall first how the propagator/reflectivity unknowns (π, r) can be defined from the material unknowns which describe the earth. We shall consider the case where an acoustic model is used, and suppose for simplicity that the density is known, so that the earth is described by its slowness distribution ν. The propagator/reflector splitting of ν into (π, r) is usually based on the decomposition of the space V of slowness into

$$V = V_s + V_r \tag{18}$$

where V_s a space of smooth functions and V_r a space of "rough" or rapidly oscillating functions. We refer to [5] for more details on possible choices for V, V_s and V_r, and on the case where the density is not known.
The propagator space Π is then chosen to contain only the low frequency components of the slowness :

$$\Pi = V_s, \quad \pi = \nu_s \tag{19}$$

The space R for the reflectivity vectors r which parametrize the high-frequency components ν_r of the slowness can be chosen as

$$R = V \tag{20}$$

with the parametrization :

$$\nu_r = P_r r_0, \quad r_0 = Wr \tag{21}$$

where :

$$\left\{ \begin{array}{ll} P_r & \text{is the projection in } V \text{ on } V_r \text{ parallely to } V_s \\ r_0 & \text{is the parameter of "natural" or unscaled reflectivity} \\ W & \text{is a (self adjoint) scaling operator acting on functions of V} \\ r & \text{is the sought vector of scaled reflectivities} \end{array} \right. \tag{22}$$

Of course, the choice (20) for the reflectivity space leads to an overparametrization of V_r. The motivation for this choice is that a scaling operator maps a function of V into a function of V, (which is not true for V_r), making thus easy the introduction of the scaling operator W. Other choices are possible, as for example the more natural choice $R = V_r$, but the definition of the scaling operator W is slighly more complicated as it requires the choice of a basis in V_r.
The scaling operator W can be chosen freely, and we take now advantage of this to enhance the performance of the migration operator $B(\pi)^*$.

Definition 1. *The operator $B(\pi)^* \in \mathcal{L}(D; R)$ is a quantitative migration operator if there exists $0 \leq \eta(\pi) < 1$ such that*

$$\| B(\pi)^* B(\pi) - P_{\pi,*} \| \leq \eta(\pi) \tag{23}$$

where :

$$\left\{ \begin{array}{ll} P_{\pi,*} = & \textit{Orthogonal Projection on } ImB(\pi)^* \textit{ in } R \\ \| \ \| = & \textit{Matrix norm associated to the Euclidian norm in } R \\ & \textit{(spectral radius for symmetric matrices)} \end{array} \right. \tag{24}$$

The motivation for this definition is clear : if $s \in ImB(\pi)$ is a time section generated by the reflectivity r, i.e. $s = B(\pi)r$, one wants that the migrated section $B(\pi)^* s$ gives a good estimate of the part of that reflectivity which is in $ImB(\pi)^*$, i.e. of the minimum norm reflectivity which has generated d. The smaller η, the better the quantitative migration ! The interest of these migrations for the waveform inversion problem is their ability to resimulate the original time section when they are used as input in the forward model :

Proposition 3. *Let $B(\pi)^* \in \mathcal{L}(D;R)$ be a quantitative migration operator. Then*

$$\|B(\pi)B(\pi)^* - P_\pi\| = \|B(\pi)^* B(\pi) - P_{\pi,*}\| \leq \eta(\pi) \tag{25}$$

where

$$P_\pi = \textit{ Orthogonal Projection on } ImB(\pi) \textit{ in } D. \tag{26}$$

Proof. Equality (25) follows immediately from (23) and the singular value decomposition of the matrix $B(\pi)$.. □

So, if $B(\pi)$ is a quantitative migration operator and s an element of the data space D, the migration of s with $B(\pi)^*$ produces a reflectivity $r = B(\pi)^* s$ which resimulates, when used as input in the linearized forward model $B(\pi)$, a synthetic section $c = B(\pi)r$ which is close to the projection of s on the range of $B(\pi)$.

We discuss now the construction of quantitative migrations. Let us denote by $B_0(\pi)$ the forward linearized map when the natural or unscaled reflectivity parameter r_0 is used :

$$B_0(\pi) = \frac{\partial c_{\pi,r_0}}{\partial r_0}(\pi, r_0 = 0). \tag{27}$$

It was recognized very early in [14] (in the case of a linearized acoustic model) and in [21] (in the case of a the high frequency approximation) that the operator $B_0(\pi)^*$ was a migration operator ; however, because of the attenuation occuring during the propagation of the forward and backward propagated fields involved in the calculation of $B_0^*(\pi)$, the amplitudes of deep events are strongly attenuated ; hence $B_0^*(\pi)$ alone is a very poor migration operator.

But we can play with the scaling W in order to compensate at best for this attenuation. From (21) we see that the forward linearized map $B(\pi)$ with respect to the scaled reflectivity parameters r is related to $B_0(\pi)$ by

$$B(\pi) = B_0(\pi)W, \tag{28}$$

and the corresponding migration operators are related by :

$$B^*(\pi) = WB_0(\pi)^*. \tag{29}$$

So, according to definition 1 ; $B^*(\pi)$ will be a quantitative migration if one is able to find a scaling W such that :

$$\|WB_0(\pi)^*B_0(\pi)W - P_{\pi,*}\| \leq \eta(\pi) < 1 \tag{30}$$

By chance, $B_0(\pi)^*B_0(\pi)$ is close to a diagonal matrix - this is why migration works ! - so that (30) can be satisfied reasonably well.

When a full acoustic forward model is used, the matrix $B_0(\pi)^*B_0(\pi)$ is numerically out of reach, but simple considerations show that using a scaling W proportional to square root of depth leads, when the norm in the data space compensates for the geometrical spreading, to a quantitative migration. This was confirmed numerically in ([5]), where it was also shown that the subspace V_r had to be chosen orthogonal in V to the space V_s of propagators $\pi = \nu_s$. This orthogonality allows to minimize the travel-time perturbations induced by the addition of ν_r to ν_s which is required to reconstruct the full slowness ν, prior one performs a forward modeling step with the full acoustic model.

When a $Born + rays$ or $WKBJ$ approximation is used as forward model, the matrix associated to $B_0(\pi)^*B_0(\pi)$ can be computed explicitely, for a source of finite bandwith and a given geometry of sources and receivers with finite aperture. Then one can chose for W the inverse of the mass-lumped matrix $B_0(\pi)^*B_0(\pi)$, which leads to a quantitative Kirchhoff migration formula (cf [9]).

Notice that a quantitative migration operator $B(\pi)^*$ produces, whenever applied to a time section d, a quantitative estimation of the scaled reflectivity r which has produced the component of d in the image of $B(\pi)$ or $B_0(\pi)$ (they are the same). If one wants an estimation of the *unscaled reflectivity* r_0, one has to multiply by W the estimation of the scaled reflectivity r ; hence the operator :

$$WB(\pi)^* = W^2B_0(\pi)^* \tag{31}$$

is a *quantitative migration operator for the estimation of the unscaled reflectivity* r_0. Formula (31) expresses the equivalence between scaling the reflectivity unknown (as we did here) and post multiplying the usual gradient-migration operator $B_0(\pi)^*$ by an amplitude correction factor W^2 aimed at restoring the true amplitudes of the reflectors (this was the presentation used in [9] [4]).

We can now study the convergence, when $\varepsilon \to 0$, of the regularized solution $\hat{\pi}_\varepsilon, \hat{r}_\varepsilon, \hat{s}_\varepsilon$. We consider fot this a sequence $\varepsilon_n \to 0$ when $n \to \infty$, and denote by $\pi_n = \hat{\pi}_{\varepsilon n}, r_n = \hat{r}_{\varepsilon n}, s_n = \hat{s}_{\varepsilon n}$ corresponding solutions of $(P_{\varepsilon n})$ and $(DP_{\varepsilon n})$, and by B_n the matrix $B(\pi_{\varepsilon n})$:

Proposition 4. *Let hypothesis* $(H1)(H2)(H3)$ *hold, and suppose that* $B(\pi)^*$ *is a quantitative migration operator such that*

$$\textit{for any } \pi \in C, \quad \eta(\pi) \leq \eta < 1. \tag{32}$$

Suppose that d is noiseless, in the sense that :

$$\min_{\pi \in C, r} \frac{1}{2}\|B(\pi)r - d\|^2 = 0 \tag{33}$$

Then, when $n \to \infty$, there exists a subsequence of π_n, r_n and s_n such that

$$\pi_n \to \hat{\pi}, \quad r_n \to \hat{r}, \tag{34}$$

$$P_n s_n \to \hat{s}, \quad \varepsilon_n^2 (I - P_n s_n) = (I - P_n)d \tag{35}$$

where $\hat{\pi}, \hat{r}, \hat{s}$ are optimal propagator, depth reflectivity and time reflectivity solution of the inverse problem associated to d, which satisfy equations (16), and where P_n is the orthogonal projection on the range of $B_n = B(\hat{\pi}_n)$.

(See appendix B for the proof).

The use of a quantitative migration operator in the dual formulation brings an enormous simplification to the solution of the inverse problem :

Proposition 5. *(second main result) Let hypothesis and notations of proposition (4) hold. Then any optimal time reflectivity $\hat{s}$ (associated to the original problem (P) by (16)), and $\hat{s}_{\varepsilon_n} = s_n$ (associated to the regularized problem (DP_{ε_n})) are related to the seismic data d by :*

$$\|d - \hat{s}\| \leq \frac{\eta}{1 - \eta}\|d\|, \tag{36}$$

$$P_n s_n \to \hat{s} \text{ when } n \to \infty \tag{37}$$

$$\|P_n(d - (1 + \varepsilon_n^2)s_n\| \leq \frac{\eta}{1 + \varepsilon_n^2 - \eta}\|d\| \tag{38}$$

If, moreover, when n goes to infinity :

$$dist\,(d, ImB_n) = \|(I - P_n)d\| = o(\varepsilon_n^2) \tag{39}$$

Then one has

$$s_n \to \hat{s} \text{ when } n \to \infty \tag{40}$$

(See appendix C for the proof).

Proposition 5 shows the importance of choosing an as-good-as possible quantitative migration operator (i.e. an as small as possible η) for the solution of the waveform inversion problem by duality : this makes it possible to replace the search for $(\hat{\pi}_\varepsilon, \hat{r}_\varepsilon)$, both of them

being unknown, by the search for $(\hat{\pi}_\varepsilon, \hat{s}_\varepsilon)$, where $\hat{s}_\varepsilon$ is almost known.

However, in its form (DP_ε), the dual problem does not allow to take directly advantage of this, as it requires, for any current guess π of the propagator, the calculation of $\hat{s}_\varepsilon(\pi)$ via an inner loop on s. The only gain sofar in replacing (P_ε) by (DP_ε) is in the expected simplification in the calculation of the gradient of the reduced objective function $J^\sharp(\pi)$ with respect to the propagator π, as we have seen in paragraph 2 ; but (DP_ε) still possesses an inner optimization loop which makes it expensive to solve.

So we devote the two next paragraphs to explore reformulations of (DP_ε) which provide, by "freezing" the dual variable s at its near-optimal value d, objective functions for the determination of the propagator π which do not involve inner iterations.

4 Duality and Migration Velocity Analysis

In the dual formulation (DP_ε), the outer optimization loop is on π, and the inner one on s. One may be willing to take advantage of the fact that, by proposition (5), one knows that $\hat{P}_\varepsilon \hat{s}_\varepsilon \simeq \hat{P}_\varepsilon d$, where $\hat{P}_\varepsilon$ is the orthogonal projection on $ImB(\hat{\pi}_\varepsilon)$, so that the migrated sections corresponding to $\hat{s}_\varepsilon$ and d satisfy :

$$B(\hat{\pi}_\varepsilon)^* \hat{s}_\varepsilon \simeq B(\hat{\pi}_\varepsilon)^* d. \tag{41}$$

One way of doing this is to exchange the order of the loops, and to replace the resolution of (DP_ε) by that of

$$Min_s Max_{\pi \in C} DSI_\varepsilon(\pi, s) \tag{DMF_ε}$$

where the acronym DMF stands for Dual Migration Fitting. As s is now in the outer loop, we can start the solution of (DMF_ε) by fixing s to the value d which, as we have seen in (41), provides a good approximation of the first term of DSI_ε, and solve for the propagator π by running the inner loop, which reads now :

$$Max_{\pi \in C} SI(\pi) = \frac{1}{2} \|B(\pi)^* d\|^2 \tag{MMF}$$

and coincides with the Multiple Migrations Fitting approach of [8] and the two-step Monte Carlo approach of [12] to migration velocity analysis, where the propagator π is defined by maximization of a similarity index defined as the norm of the stacked migrated section. It is interesting to note that this similarity index, which was originally introduced somewhat empirically in order to quantify the geophysical intuition that "good velocity model produce coherent migrations", is receiving now a mathematical legitimacy, as its maximization coincides with the inner loop of the (DMF_ε) formulation for $s = d$. In particular, the duality approach indicates that the migration operator to be used in the (MMF) objective function for its minimizer to be an approximation of the optimal propagator is $B^*(\pi) = WB_0^*(\pi)$ where W has been chosen as in paragraph 3 in such a way that $W^2 B_0^*(\pi)$ is a quantitative migration for the estimation of the unscaled reflectivity r_0 (cf (31)).

The good new about replacing (DP_ε) by (DMF_ε) is the

Proposition 6. *let* $(H1),(H2),(H3)$ *hold. Then the Dual Migration Fitting problem* (DMF_ε) *admits a solution.*

We skip the proof of this proposition, which results simply from the fact thet C is compact and B continuous.
But the bad new is that, in general, (DMF_ε) is not necessarily equivalent to (DP_ε) as one only has

$$Min_s Max_{\pi\in C} DSI_\varepsilon(\pi,s) \geq Max_{\pi\in C} Min_s DSI_\varepsilon(\pi,s). \tag{42}$$

A well-known condition for the equivalence is the

Proposition 7. *let* $(H1),(H2),(H3)$ *hold. If the dual similarity index has a saddle point , i.e. if*

$$\begin{cases} \exists \hat{\pi}_\varepsilon \in C, \hat{s}_\varepsilon \in D \text{ such that }, \forall \pi \in C, \forall s \in D \\ DSI_\varepsilon(\pi,\hat{s}_\varepsilon) \leq DSI_\varepsilon(\hat{\pi}_\varepsilon,\hat{s}_\varepsilon) \leq DSI(\hat{\pi}_\varepsilon,s) \end{cases} \tag{43}$$

then the problems (DP_ε) *and* (DMF_ε) *are equivalent.*

We don't know if the existence of such a saddle point can be proven when $B(\pi)$ is the forward linearized wave equation. But there are some reasons for which this may in some case be true. First, the right inequality, in (43) is automatically satisfied if one chooses for $(\hat{\pi}_\varepsilon,\hat{s}_\varepsilon)$ the solution of (DP_ε), which is equivalent to (IP_ε), so that the geophysical interpretation of this $\hat{\pi}_\varepsilon$ is the propagator which gives the best fit to the data. Second, the left inequality in (43) will be satisfied for the above choice of $\hat{\pi}_\varepsilon$ for all propagators π which give a migrated section of $\hat{s}_\varepsilon$ with a smaller norm than the one obtained when migrating with the best propagator $\hat{\pi}_\varepsilon$. One can then hope that, if the reflectivity of the medium is rich enough, and the shots or illuminations numerous enough, "any" perturbation π of $\hat{\pi}_\varepsilon$ will to destroy the coherency of the migrated sections at some places and hence lead to a decrease of the norm of the stacked migrated section, in which case the left inequality of (43) would be satisfied also.

The lower part of figure 1, which we used to illustrate the algorithm for the solution of (DP_ε) in paragraph 3, can also describe the solution of (DMF_ε) by exchanging the order in which the left and right loops are performed : the initial guess s^{init} to the time reflectivity is set equal to the data d, and the left loop on the propagator is run to obtain a first estimate of the background (which shows that performing a migration velocity analysis via MMF is the first step of Dual Migration Fitting). It is hoped, as d is a good approximation to $\hat{s}_\varepsilon$, that this first optimization run will produce a good propagator. The numerical results obtained in [8] with this first MMF step concern an example with a very simple reflectivity pattern

and very few illumination, which does not allow to conclude, but the results optained in [12] on real data with a more rich reflectivity and a larger number of shots indicates that the MMF step alone can provide a propagator of satisfying quality. If this quality is not good enough, one can enhance the time reflectivity s by running a few iterations of the right loop, and so on.

Notice that the left optimizer is itself surrounded by a small loop, which corresponds to a hierarchical or multiscale optimization, which is very natural in this context as it does not make sense to try to fine-tune the local details of the propagator (background) as long as the general trend has not been properly set. So the propagator space V_s, which is a subspace of the space V of slowness as we have seen in section 3, is decomposed on a range of scales going from 0 (coarser) to SCAVS (finer), and the optimization is performed from the coarser to the finer scale, the propagator obtained at a given scale being used as initial guess for the optimization at the next scale.

5 Migration Based Travel Time Formulation

The MBTT formulation (cf [10] [4] [7][15]) can also be seen as a way of taking advantage of the formula (14) of proposition 3, which shows that parameterizing the depth reflectivity r by the time reflectivity s using the migration operator $B(\pi)^*$

$$r = B(\pi)^* s \tag{44}$$

does not lead to underparametrization when a linearized forward model is used. But instead of using the Max/Min formulation (DP_ε) of the dual problem, this change of variable is plugged into the forward model in (P_ε) in order to keep a Min/Min problem. The regularization term $\varepsilon^2/2\|r\|_R^2$ in (P_ε) was introduced in order that the solution r_ε of (P_ε) is an approximation to the minimum norm solution $\hat{r}$ of (P). But (44) gives directly a parametrization of this minimum norm solution, as $ImB(\pi)^*$ is the orthogonal subspace to $\ker B(\pi)$. Hence the regularization term can be dropped when the parametrization (44) is used, leading thus to the formulation :

$$Min_{\pi\in C} Min_s J(\pi, s) = \frac{1}{2}\|d - B(\pi)B(\pi)^* s\|^2 \qquad (MBTT)$$

Proposition 8. *let $(H1)$ hold. Then $(MBTT)$ is equivalent to the search of a solution of (P) with minimum norm reflectivity.*

One advantage of $(MBTT)$ over (DP_ε) is that one can use the full non-linear acoustic forward model, by replacing in $(MBTT)$ the objective function $J(\pi, s)$ by

$$J(\pi, s) = \frac{1}{2}\|d - c_{\pi, B^*(\pi)s}\|^2 \tag{45}$$

This is the way the $(MBTT)$ approach hes been implemented in [10] , [4] [7]. In this case, the fact that the $r \rightarrow s$ change of unknown is not underparametrizating has to be checked numerically.

One other advantage of $(MBTT)$ over (DP_ε) is that it is a minimization problem instead of a maxi/minimization one. This allows to perform the loops on π and s in any convenient order, and to take advantage of the fact that the data d themselves are a good approximation to $\hat{s}$ as we has seen in proposition (5) : one will first minimize $J(\pi, s)$ with respect to π, for $s = d$ fixed. In that case, because of the proposition (3) on quantitative migration operators, we see that $B(\pi)B^*(\pi)d$ is close to the projection of the data d onto the range of the forward model $B(\pi)$, so that

$$J(\pi, d) \simeq \text{ dist } (d, ImB(\pi)) = J^\sharp(\pi). \tag{46}$$

Hence the first step of $(MBTT)$ is close to the minimization of the reduced objective function $J^\sharp(\pi)$; we see here the power of the duality approach : it allows to replace the evaluation of $J^\sharp(\pi)$, which is very expensive as it requires an inner optimization loop, by that of $J(\pi, d)$, which requires only a fixed and finite number of operations !

The first numerical results with a full acoustic forward modelling (cf [7]) indicate that the minimization of $J(\pi, s)$ with respect to π, for $s = d$ fixed, results in a satisfactory propagator π. An additional round of optimization with respect to s, for the previously obtained π, produces a better restitution of the amplitudes in the resimulated data. However, a second round of optimization with respect to π, using the above enhanced value of s, did not lead to a better definition of π.

The same approach used with a simplified Born + rays forward modelling (cf [15]) has been sucessfully applied to both synthetic and real data made of a few thousand of traces.

6 A Dual Formulation of Differential Semblance Optimization

The differential Semblance Optimization (DSO) was introduced (cf [19]) in order to overcome the problem of local minima encountered when one tries to determine the propagator π by minimization of the data misfit function as in $(LS), (P)$ or (P_ε). We recall first this formulation, which was developed for the linearized wave equation with known density. The propagator π is the smooth slowness at which the linearization occurs, and the reflectivity is the slowness perturbation in the right hand side of the linearized equation.

The first feature of DSO is that it relaxes the usual formulation (P_ε) by replacing the search for one reflectivity unknown r by the search for as many unknowns as shots. Hence the reflectivity space for DSO is

$$R^{DSO} = R^{NSHOT} \tag{47}$$

and the fact that "the earth is unique" is accounted for by penalization in the objective function :

$$J_\varepsilon^\propto(\pi,r)=\frac{1}{2}\|d-B(\pi)r\|^2+\frac{\varepsilon^2}{2}\{\|r\|^2+\frac{1}{\propto^2}\|Dr\|^2\} \tag{48}$$

where

$$Dr=(r_2-r_1,...r_{NSHOT}-r_{NSHOT-1})\in R^{NSHOT-1} \tag{49}$$

is the semblance operator. This is justified by the

Proposition 9. *Let $\pi,\varepsilon>0$ be given, and let $\hat{r}_\varepsilon\in R$ and $\hat{r}_\varepsilon^\propto\in R^{DSO}$ be the minimizers of J_ε and $J_\varepsilon^\propto$. Then one has*

$$\hat{r}_\varepsilon^\propto\rightarrow(\hat{r}_\varepsilon,\hat{r}_\varepsilon,...) \text{ when } \quad \propto\rightarrow 0 \tag{50}$$

Proof. This results directly from the classical regularization theory. □

The second feature of DSO is that it performs a nested optimization of $J_\varepsilon^\propto$, with the inner loop on r and the outer loop on π. The DSO formulation writes :

$$Min_{\pi\in C}\{J^{DSO}(\pi)=Min_r J_\varepsilon^\propto(\pi,r)\} \tag{DSO}$$

The function J^{DSO} has been shown to behave in a nice unimodal manner with respect to π , which makes it a useful tool for the determination of the propagator π by local optimization technique. However, its definition as a Min makes it expensive to evaluate, inasmuch the whole minimization whith respect to r has to be performed from scratch again each time the value of π is changed, even slightly (because of the phase shifts in depth, using $\hat{r}(\pi)$ as starting value for the evaluation of $J^{DSO}(\pi+\delta\pi)$ does not do any good, the starting value $\hat{r}=0$ works usually even better !).

The duality technique provides an alternative expression of $J^{DSO}(\pi)$:

Proposition 10. *Let hypothesis $(H1)(H2)(H3)$ hold. Then $J^{DSO}(\pi)$ can be calculated by*

$$J^{DSO}(\pi)=\varepsilon^2 Max_{s\in D}\left\{ <s,d>-\frac{\varepsilon^2}{2}\|s\|^2-\frac{1}{2}\|B(\pi)^*s\|^2_{-\propto} \right\} \tag{51}$$

where

$$\|r\|^2_\propto=\|r\|^2_{R^{DSO}}+\frac{1}{\propto^2}\|Dr\|^2=<M_\propto r,r>_{R^{DSO}} \tag{52}$$

$$\|r\|^2_{-\propto}=<M_\propto^{-1}r,r>_{R^{DSO}} \tag{53}$$

Proof. formula (51) results directely from proposition 3 for the choice

$$\begin{aligned} F(c) &= \frac{1}{2}\|d-c\|^2 \\ G(r) &= \frac{\varepsilon^2}{2}\{\|r\|^2_{R^{DSO}} + \frac{1}{\propto^2}\|Dr\|^2\} \\ &= \frac{\varepsilon^2}{2}\|r\|^2_{\propto} = \frac{\varepsilon^2}{2} < M_{\propto} r, r >_{R^{DSO}} \end{aligned}$$

and with the dual variable

$$s = b/\varepsilon^2.$$

□

The advantage of formulation (51) is that $J^{DSO}(\pi)$ is obtained by a maximization with respect to a time section s ; it can then be expected that $\hat{s}(\pi)$ will be a good starting value for the maximization for the evaluation of $J^{DSO}(\pi + \delta\pi)$, so that only a few iterations in s will be necessary. The disadvanteges of formulation (51) is that the optimization variable s for the inner loop has the size of a collection of time sections, and hence is very large ! Moreover, the evaluation of the norm $\| \quad \|_{-\propto}$ of the collection of migrated sections requires the solution of a linear system, which may limit the efficiency of the dual approach.

7 Conclusion

Convex duality allows to replace the untractable classical least-square formulation (LS) of waveform inversion by various reformulations : the dual problem (DP), the dual migration fitting (DMF), the Migration based travel time $(MBTT)$, which have all in common that they search for a propagator or background unknown π and a time reflectivity unknown s similar in nature to the data to be inverted. The use of quantitative migrations in the dual formulation ensures moreover that the data d themselves are a good approximation of the optimal time reflectivity $\hat{s}$. The (DMF) and $(MBTT)$ formulations take full advantage of this : by freezing in a first step s to its nearly optimal value d, they produce an objective function which depends only on the propagator or background π :

- the first step of (DMF) is multiple migration fitting (MMF), which amounts to perform a migration velocity analysis by maximization of a properly defined coherency criteria. The equivalence, under certain hypothesis, of (DMF) and the initial least-square problem (LS) shows that waveform inversion and migration velocity analysis (MVA) are in fact mathematically dual problems, and the equivalence is constructive in the sense that duality indicates how to define the MVA coherency criterion for this equivalence to hold.

- The first step of $(MBTT)$ consists in minimizing the misfit between the data and their resimulation, obtained by migration followed by modelling. When a linearized forward model and a quantitative migration are used, this is very close to searching for the propagator π which minimizes the reduced objective function $J^\sharp(\pi)$, which measures the distance of the data d to the range of the forward modelling operator $B(\pi)$.

So duality provides a unifying framework for waveform inversion and migration velocity analysis. Its practical importance comes from the fact that it suggests various objective functions for the determination of propagator or backgrounds, which are all optimizable by local gradient technique.

Acknowledgments

I would like to thank Sven Treitel, of Amoco Production Research, who gave me the opportunity of visiting his lab since many years. What I learned there on the real side of migration and inversion had for sure a strong influence on the (slow !) maturation process which led to this paper.
This paper was rewritten while I was visiting the Institute for Mathematics and its Applications (IMA) in Minneapolis, which I thank for its hospitality and great environment.
I thank Muriel De Bianchi (INRIA, France) for the patient typing of the numerous succesive versions of this paper, and Sherry Vanscoyoc (Amoco's Technology Center, Tulsa) for her skilled drawing of the block diagram which illustrate this paper.

References

[1] A. Bamberger, G. Chavent, and P. Lailly. About the stability of the inverse problem in a 1-D wave equation - application to the interpretation of seismic profiles. *Journal of Applied Mathematics and Optimization*, 5:1–47, 1979.

[2] C. Bunks. Paraxial wave equation inversion of seismic data by migration and remodeling. 1993. submitted to Geophysical Prospecting.

[3] C. Bunks, F. Mohamed-Saleck, S. Zaleski, and G. Chavent. Multiscale seismic waveform inversion. 1993. submitted to Geophysics.

[4] G. Chavent and F. Clément. Quantitative migrations and resimulation techniques for waveform inversion via a Migration-Based Travel Time formulation. *Submitted to Geophysics*, 1993. (preliminary version : INRIA Report n°1839, Jan. 93).

[5] G. Chavent and F. Clément. Separating propagation and reflection parameters in the acoustic wave equation. *Submitted to Geophysics*, 1993. (preliminary version : INRIA Report n°1839, Jan. 93).

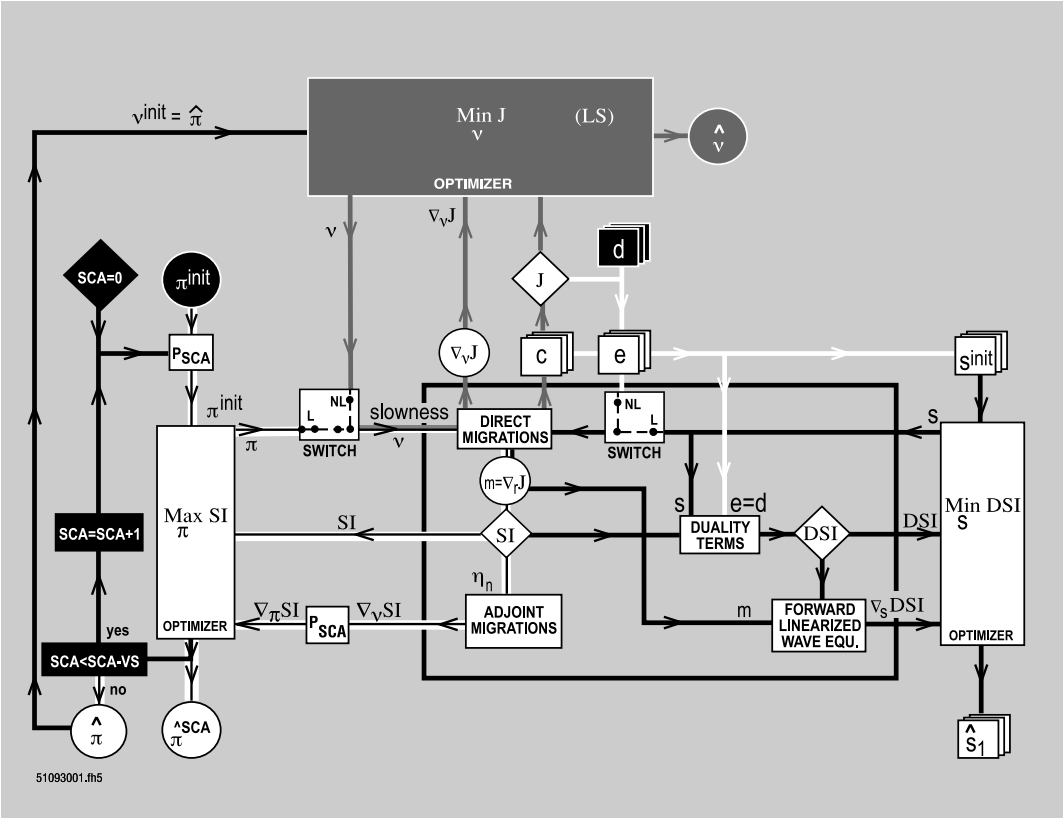

Figure 1: Block-diagram for the dual problem (DP_ε) and the dual migration fitting (DMF_ε) (bottom part of the figure) and for the classical least squares waveform inversion (LS) (upper part of the figure) in the case of a known density (the ε indices are omitted).

[6] G. Chavent, F. Clément, and S. Gomez. En route to waveform inversion by MBTT formulation : optimization for one shot and gradient for multiple shots. 1994. Preprint in Annual Report PSI Consortium 1993.

[7] G. Chavent, F. Clément, and S. Gomez. Acoustic waveform inversion by MBTT formulation. 1995. Submitted to Geophysics.

[8] G. Chavent and C. A. Jacewitz. Determination of background velocities by multiple migration fitting. *Geophysics*, 60-2:476–490, 1995.

[9] G. Chavent and R. E. Plessix. A time domain derivation of optimal and suboptimal Kirchoff quantitative migration via the least-squares approach. 1995. Submitted to Geophysics (preliminary version : technical report 1928 INRIA 1993.

[10] F. Clément. A Migration-Based Travel Time formulation for the inversion of 2-D seismic reflection data. In *1st Int. Conf. on Mathematical and Numerical Aspects of Wave Propagation Phenomena*, pages 455–461. INRIA and SIAM, 1991.

[11] I. Ekeland and R. Témam. Convex analysis and variational problems. *Dunod*, 1974. Paris.

[12] S. Jin and R. Madariaga. Nonlinear velocity inversion by a two-step Monte Carlo method. *Geophysics*, 59-4:577–590, 1994.

[13] P. Kolb, F. Collino, and P. Lailly. Pre-stack inversion of a 1-D medium. 74(3):498–508, 1986. Proceedings of the IEEE.

[14] P. Lailly. The seismic inverse problem as a sequence of before stack migrations. In Robinson E. Bednar J., Rodner R. and Weglein A., editors, *Conf. on Inverse Scattering : Theory and Application*. SIAM, 1983.

[15] R. E. Plessix, Y. H. De Roeck, and G. Chavent. Automatic and simultaneous migration velocity analysis and waveform inversion of real data using an MBTT/UKBJ formulation. Oct 1995. Extended Abstract SEG Fall Meeting, Houston.

[16] J. A. Scales, M. L. Smith, and T. L. Fischer. Global optimization methods for multimodal inverse problems. *Journal of Computational Physics*, 103:258–268, 1992.

[17] P. L. Stoffa and M. K. Sen M. Nonlinear multiparameter optimization using genetic algorithms : inversion of plane-wave seismograms. *Geophysics*, 56(11):1794–1810, 1991.

[18] W. W. Symes. The reflection inverse problem for acoustic waves. In L. Halpern G. Cohen and P. Joly, editors, *Mathematical and numerical aspects of wave propagation phenomena*, pages 423–433. SIAM, 1991.

[19] W. W. Symes and J.J. Carrazone. Velocity inversion by Differential Semblance Optimization. *Geophysics*, 56:654–663, 1991.

[20] W. W. Symes and Michel Kern. Inversion of reflection seismograms by differential semblance analysis : algorithm structure and synthetic example. *Geophysical Prespecting*, 42:565–614, 1994.

[21] A. Tarantola. Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015, 1984.

[22] A. Tarantola and B. Valette. Generalized non-linear inverse problems solved using the least-squares criterion. *Review of Geophysics and Space Physics*, 20:219–232, 1982.

A Convex Duality

In order to transform (P_ε) by duality, we recall the following duality result (see for example [11] for a detailed exposition) :

Proposition 11. *(Fenschel and Rockafellar) : Let* $F : D \to \mathbb{R}, G : R \to \mathbb{R} \cup +\infty$ *and* $A \in \mathcal{L}(R; D)$ *satisfy :*

$$F \text{ is convex continuous, } \quad lim_{\|b\|_D \to \infty} \quad F(b) = +\infty \tag{D1}$$

$$G \text{ is positive, convex, lower semi continuous and not identical to } \quad +\infty \tag{D2}$$

$$lim_{\|x\| \to \infty} \quad F(Ar) + G(r) = +\infty. \tag{D3}$$

Then

$$Inf_{r \in R} \{F(Ar) + G(r)\} + Inf_{b \in D} \{G^*(A^*b) + F^*(-b)\} = 0 \tag{54}$$

where $F^* : D \to \mathbb{R}$ *and* $G^* : R \to \mathbb{R} \cup +\infty$ *are the convex conjugate functions of* F *and* G*, defined by :*

$$F^*(b) = Sup_{c \in D} \{< b, c >_D - F(c)\} \tag{55}$$

$$G^*(q) = Sup_{r \in R} \{< q, r >_R - G(r)\} \tag{56}$$

Moreover, the two optimization problems in (54) have solutions $\hat{r}, \hat{b}$ *which satisfy the duality relations :*

$$\begin{cases} A^*\hat{b} \in \partial G(\hat{r}) \\ -\hat{b} \in \partial F(A\hat{r}) \end{cases} \tag{57}$$

where $\partial f(x)$ denotes the subdifferential of a convex function f. When f is differentiable, ∂f contains a single vector, namely the gradient of f at point x. It is then straightforward to specialize this to the linearized problem (P_ε) by choosing, for a given π :

$$\left\{ \begin{array}{l} F(c) = \dfrac{1}{2}\|d-c\|^2, \\ G(r) = \dfrac{\varepsilon^2}{2}\|r\|^2, A = B(\pi). \end{array} \right. \tag{58}$$

A simple calculation gives then

$$\left\{ \begin{array}{l} F^*(b) =< b, d > + \dfrac{1}{2}\|b\|_D^2 \\ G^*(q) = \dfrac{1}{2\varepsilon^2}\|q\|_R^2 \end{array} \right. \tag{59}$$

and hypothesis $(H1), (H2), (H3)$ ensure that, for any $\pi \in C$, the duality hypothesis $(D1), (D2), (D3)$ of proposition (11) are satisfied. So if we replace the abstract dual variable b by the more convenient one

$$s = b/\varepsilon^2 \tag{60}$$

(which will be the time reflectivity unknown we were looking for in the introduction), the proposition (11) rewrites as proposition (2).

B Proof of proposition 4

By definition of π_n, R_n one has :

$$\left\{ \frac{1}{2}\|d - B_n r_n\|^2 + \frac{\varepsilon_n^2}{2}\|r_n\|^2 \leq \frac{1}{2}\|d - B(\pi)r\|^2 + \frac{\varepsilon_n^2}{2}\|r\|^2 \text{for any } \pi \in C \text{ and } r \in R. \right. \tag{61}$$

By chosing $\pi = \hat{\pi}$ and $r = \hat{r}$ one finds that :

$$\|d - B_n r_n\| \leq \varepsilon_n \|\hat{r}\|, \tag{62}$$

$$\|r_n\| \leq \|\hat{r}\|. \tag{63}$$

Then, by definition of s_n one has :

$$(B_n B_n^* + \varepsilon_n^2 I)s_n = d$$

which rewrites, by adding and substracting the projection $P_n s_n$:

$$(P_n + \varepsilon_n^2 I)s_n + (B_n B_n^* - P_n)s_n = d,$$

and splits, by projection on the range of B_n and on its orthogonal subspace, into :

$$[(1+\varepsilon_n^2)I + (B_n B_n^* - P_n)]P_n s_n = P_n d \tag{64}$$

and :

$$\varepsilon_n^2 (I - P_n) s_n = (I - P_n) d. \tag{65}$$

Equation (65) proves the right part of (35). Then, as $\|B_n B_n^* - P_n\| \leq \eta < 1$, (64) implies that

$$\frac{1}{1+\varepsilon_n^2+\eta}\|P_n d\| \leq \|P_n s_n\| \leq \frac{1}{1+\varepsilon_n^2-\eta}\|P_n d\| \tag{66}$$

$$\|P_n d - (1+\varepsilon_n^2)P_n s_n\| \leq \eta \|P_n s_n\|. \tag{67}$$

Using (63) (66) and the fact that C is compact, one can extract from π_n, r_n and s_n subsequences, still denoted by π_n, r_n and s_n, such that

$$\begin{cases} \pi_n & \to \tilde{\pi} \in C, \\ r_n & \to \tilde{r} \in R, \\ P_n s_n & \to \tilde{s} \in D. \end{cases} \tag{68}$$

As $\pi \to B(\pi)$ continuous, this implies that

$$B_n \to \tilde{B} = B(\tilde{\pi}) \text{ in } \mathcal{L}(R; D),$$

and we can pass to the limit in (61) :

$$\begin{cases} \frac{1}{2}\|d - \tilde{B}\tilde{r}\|^2 & \leq \frac{1}{2}\|d - B(\pi)r\|^2 \\ \text{for any } \pi \in C & \text{and } r \in R \end{cases}$$

which proves that

$$d = \tilde{B}\tilde{r}. \tag{69}$$

But, for each n one has

$$r_n = B_n^* s_n = B_n^* P_n s_n$$

which, by passing to the limit, shows that

$$\tilde{r} = \tilde{B}^* \tilde{s}. \tag{70}$$

From (69)(70) we see that $\tilde{\pi}, \tilde{r}, \tilde{s}$ satisfy (16), which ends the proof of the proposition.

C Proof of proposition 5

From (16) we see that

$$B(\hat{\pi})B(\hat{\pi})^* \hat{s}, = d \quad \text{with } \hat{s} \in ImB(\hat{\pi}).$$

Hence, if $\hat{P}$ denotes the orthogonal projection on $ImB(\hat{\pi})$:

$$[\hat{P} + (B(\hat{\pi})B(\hat{\pi})^* - \hat{P})]\hat{s} = d$$

which rewrites, as $\hat{s}$ and d are in $ImB(\hat{\pi})$:

$$[I + (B(\hat{\pi})B(\hat{\pi})^* - \hat{P})]\hat{s} = d. \tag{71}$$

But $\|B(\hat{\pi})B(\hat{\pi})^* - \hat{P}\| \leq \eta < 1$, so that $\|(B(\hat{\pi})B(\hat{\pi})^* - \hat{P})^{-1}\| \leq 1/(1-\eta)$, and :

$$\|\hat{s}\| \leq \frac{1}{1-\eta}\|d\| \tag{72}$$

But from (71) we see also that :

$$\|d - \hat{s}\| \leq \eta\|\hat{s}\|$$

which, combined with (72) gives (13).

The convergence (37) of $P_n s_n$ towards $\hat{s}$ has already been proven in the proposition (4), and the formula (38) results directely from formula (66) (67) of the proof of proposition (4).

Finally, under hypothesis (39), we see from the right part of (35) in proposition (4) that

$$\|(I - P_n)s_n\| \to 0 \quad \text{when } n \to \infty, \tag{73}$$

which together with (37) implies (40). This ends the proof of proposition (5).

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LES NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399