Geophysical Journal (1989) **96**, 151–162

Generalized non-linear elastic inversion with constraints in model and data spaces

Philip M. Carrion
School of Geophysical Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

Accepted 1988 July 7. Received 1988 July 6; in original form 1988 February 18

SUMMARY

Seismic data are non-linearly related to model parameters such as seismic velocities. However, seismic inversion is usually considered in a linear approximation. Such techniques as the Born inversion were recently applied to seismic data.

Non-linear inversion is more complicated and involves extensive calculations. Non-linear inversion was developed in the frame work of an unconstrained optimization procedure. It uses as *a priori* information an initial model and probability distribution functions in the data and model spaces (This *a priori* information is called 'soft' bounds).

In this paper, we propose a new technique for solving a constrained non-linear inversion. This technique will allow us to use *a priori* information not only in terms of 'soft' bounds, but 'hard' bounds as well (usually giving more stable and accurate solutions).

Non-linear inversion is considered as an iterative procedure which involves a dual transform at each iteration. A dual transform allows for considering the problem in terms of the Lagrangian multipliers. The number of Lagrangian multipliers is equal to the number of available data and thus, significantly reduces the dimension of the problem (this is true for underdetermined problems only). However, the most important property of the dual transform is that it allows us to consider a constrained problem as an unconstrained problem.

Another important property is that proper constraints incorporate small-wave numbers in the generalized inversion. It is shown that conventional (unconstrained non-linear inversion) is a special case of the constrained non-linear inversion developed in this paper if the truncation operator is represented by the identity matrix.

1 INTRODUCTION

One of the most important problems in geophysics is to map geological structures from the recorded data. Data are usually measured at the surface of the Earth or in boreholes (land survey). In marine seismology, receivers are placed in towing arrays or on the seafloor.

The Earth can be interpreted as a bandpass filter: this precludes measuring low-frequency and high-frequency responses. In exploration geophysics, for example, we usually deal with the data in the range of 5–70 Hz which means that our data are bandlimited.

The bandlimitation of the data is the major cause of high instabilities when 1-D exact inversion is applied to the observed data. Recall that exact inversion requires that the source be the Dirac δ-function and that there be no attenuation in the medium. If the Dirac-type source is used in the loss-free 1-D medium, then it is possible to uniquely recover the acoustic impedance using, for example, a layer-stripping approach (Robinson 1982; Santosa & Schwetlick 1982; Bube & Burridge 1983; Yagle & Levy 1984 among others). It was also established that missing low frequencies are responsible for high instabilities, whereas missing high frequencies affect the resolution and do not cause unstable results (Carrion & Patton 1983; Santosa, Symes & Raggio 1985; Gray & Symes 1985).

Seismic data are usually corrupted with noise which leads to unstable results especially if the seismic wavelet is not minimum-phase (Carrion 1987).

If we consider a layered medium, then instead of one experiment with a point source, we can consider a number of plane-wave experiments with different angles of incidence. This can be done by plane-wave decomposition (PWD) or slant stacking (Stoffa *et al.* 1981; Treitel, Gutowski & Wagner 1982; Brysk & McCowan 1986a). PWD allows for effective separation of post-critical, critical and pre-critical arrivals.

It was shown that larger offset arrivals fill in missing low-frequencies and critical arrivals fill in missing zero-frequency (Carrion 1987). Moreover, larger offset data fill in missing low-frequencies, so that it becomes possible to separately recover density and velocity profiles from a pair of experiments with different angles of propagation (Carrion 1985; Yagle 1985; Santosa & Symes 1985; Brysk & McCowan, 1986b). It was shown that the reconstruction of densities is poorer than velocities. (Both velocity and density profiles have spikes even for perfect data which on density profiles are larger than on the velocity profiles.)

These spikes can be considered as noise. In order to suppress these spurious spikes, a number of pairs of plane-wave experiments are used and the results are averaged.

When the medium cannot be presented as a stack of homogeneous layers, then inversion becomes much more complicated.

In order to simplify the issue, the problem can be linearized and presented as an inversion of small perturbations of the known background velocity. If the background velocity is chosen to be constant, then linearized inversion and migration become equivalent. In particular, a linearized 2.5 inversion with a constant background becomes identical to the Stolt migration (Stolt 1978). Linearized inversion was developed by Cohen & Bleistein (1977) and Bleistein, Cohen & Haigin (1985). Linearized multiparameter inversion was recently developed by Weglein, Violette & Keho (1986).

For variable background velocity, Clayton & Stolt (1981) introduced a two-step procedure which is based on downward extrapolation of the recorded data using a downward extrapolation operator based on the WKBJ presentation of the wave equation. This operator removes the effect of a variable background and pushes reflectors located above the new datum level into negative times (an important property of downward extrapolation operators). However, these operators can handle only pre-critical events and are not defined near the turning points of propagating plane waves.

Clayton & Stolt (1981) showed that downward extrapolation removes the effects of a variable background and inversion for velocity and density can then be performed as for a constant background.

Bleistein & Gray (1985) extended the earlier results of Cohen & Bleistein to the case of variable background as a function of depth.

The mathematical relation between migration and inversion was given by Beylkin (1985) who showed that linearized inversion in a high-frequency approximation is directly related to the inversion of causal Radon transforms (CRT). This means that inversion is also related to tomography when we try to reconstruct parameters from their projections. Beylkin (1985) showed that migration can be presented as the first term of the expansion of Fourier integral operators. He also showed that the migration operator is able to recover not only the location of discontinuities but the magnitudes of these discontinuities. This is true under two assumptions: (1) the relation between the data and the model can be approximated by the Radon transform and (2) complete data (when the object is surrounded by receivers with infinitesimal distance between geophones).

Bleistein (1987) generalized Beylkin's results and showed that approximate reflection coefficients can be recovered from the Kirchhoff data. His results as well as Beylkin derivations are true for high-frequency approximations. Parsons (1986) showed that WKBJ reflection coefficients typically aproximate true amplitudes very well.

Carrion (1987) showed that for limited apertures, WKBJ reflection coefficient can be improved by using a generalized inversion (least-squares data fitting) if these WKBJ reflection coefficient are taken as the initial model.

Chung, Carrion & Beylkin (1987) showed that operators derived by Beylkin & Bleistein have an important property which is that they do not create images in the wrong locations. In other words, wavefront sets of the image and the true object partially coincide. This is true for any Kirchhoff-type operator.

Inversion becomes very complicated if it is considered in a non-linear framework (see an excellent review paper by Stolt & Weglein 1985).

One of the approaches to solving non-linear inverse problems is to invert a Born series for a variable background velocity (Weglein 1982).

Yagle (1986) showed an inversion technique for the multidimensional Schrödinger potential. Earlier, Coen, Cheney & Weglein (1984) demonstrated an inversion technique for a variable density medium based on the inversion of the 2-D Marchenko equation developed by Cheney (1984).

Recent work on inversion includes Carrion & VerWest (1987) who considered inversion in lossy medium. They showed that the quality parameter Q can be uniquely recovered from normal incidence seismograms if this parameter is frequency independent.

Meadows & Coen (1986) developed a technique for inversion in anisotropic medium. This technique is based on the state-space approach developed by Shiva & Mendel (1983) and Aminzàdeh (1984).

2 NON-LINEAR GENERALIZED INVERSION

2.1 Definitions

Recently, substantial interest in the geophysical community was devoted to non-linear generalized inversion. This is also explained by the close relation of generalized inversion to seismic tomography and migration.

In particular, Tarantola (1984a) and Lailly (1984) showed that linearized least-squares inversion is related to the Kirchhoff migration. Carrion (1987) showed that the linearized least-squares can be reduced to the Backus–Gilbert (B–G) inversion.

Linearized inversion was very well described by Lines & Treitel (1984). Treitel & Lines (1982) established a direct relation between linear least-squares and deconvolution.

Let us consider the elastic wave equation which governs the propagation of stress in elastic media:

$$\rho(\mathbf{x})U_i(\mathbf{x}) = B_i(x) + \tau_{ij,j}(\mathbf{x}), \tag{1}$$

where $U_i(\mathbf{x})$ is the ith component of the 3-D displacement vector, $\mathbf{x} \in R^3$ is a coordinate vector in a 3-D space, $\mathbf{B}$ is the vector of body forces, τ_{ij} is the stress tensor and ρ is the density of the medium.

It should be mentioned that equation (1) describes the propagation of compressional and shear-waves in elastic media.

For small values of strain, the dependence between stress and strain can be presented via Hooke's law:

$$\tau_{ij} = c_{ijkl}e_{kl}, \tag{2}$$

where τ_{ij} is the stress tensor, e_{kl} is the strain tensor and c_{ijkl} is the elastic fourth-order tensor which can be expressed (this is true only for isotropic media) through two Lamé constants λ and μ:

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \tag{3}$$

Actually, equation (1) is a non-linear equation with respect to the unknown phase velocity of compressional $V_p(\mathbf{x})$ and shear $V_s(\mathbf{x})$ waves which are determined by Lamé constants:

$$V_p = \sqrt{\frac{\lambda + 2\mu}{\rho}} \tag{4}$$

and

$$V_s = \sqrt{\frac{\mu}{\rho}}. \tag{5}$$

Equation (1) can be presented in a non-linear functional form

$$U_i = U_i(V_p, V_s, \rho) \tag{6}$$

which simply means that the displacement vector in elastic media is uniquely defined by the distribution of compressional velocity, shear velocity and density in the medium (for fixed sources).

If this equation is considered at some surface S where the receivers are located, then our problem will be to find V_p, V_s and ρ from the measured response at the surface. Thus, the generalized non-linear elastic inversion is to find all elastic parameters from the seismic response measured at a surface $S(\xi_r)$ and non-linearly related to these parameters via equation (6). Recall that for high-frequency approximations and small perturbations of parameters, non-linear equation (6) reduces to the Radon transform.

2.2 Least-squares and linear models

Equation (6) defines the non-linear relation of the displacement vector and the medium parameters V_p, V_s and ρ.

Suppose that we are looking for a solution $\mathbf{m}$

$$\mathbf{m} = \begin{pmatrix} V_p \\ V_s \\ \rho \end{pmatrix} \tag{7}$$

which is close enough to the reference model $\mathbf{m}_0$. Equation (6) can then be linearized using the Taylor expansion:

$$U_i(m) \approx U_i(\mathbf{m}_0) + \mathbf{F}_{ij}(m_j - m_{j,0}), \tag{8}$$

where $\mathbf{F}_{ij}$ is the matrix of Frechét derivatives (derivatives of U with respect to m):

$$\mathbf{F}_{ij} = \frac{\partial U_i}{\partial m_j}. \tag{9}$$

Equation (8) can be rewritten as:

$$\mathbf{U}_s(\mathbf{m}) = \mathbf{F}(\mathbf{m} - \mathbf{m}_0), \tag{10}$$

where U_s is the backscattered response associated with the perturbed model:

$$\mathbf{U}_s = \mathbf{U}(\mathbf{m}) - \mathbf{U}(\mathbf{m}_0). \tag{11}$$

The non-linear problem (6) is now linearized and a solution $\mathbf{m}$ can be found from (10). However, in order to solve equation (10), the inverse of the Frechét matrix should be found. In general, the inverse of $\mathbf{F}$ does not exist and thus, to find $\mathbf{m}$ from (10) is not possible. Instead, we are trying to find a generalized solution to (10) using least-squares or other norm (for example, L_1-norm). In least-squares, the mathematics is straightforward and moreover, we deal with convex functionals which will greatly facilitate the derivation of the main results of this paper. Conventional least-squares can be written as:

$$E(\mathbf{m}) = \|\mathbf{U}_s - \mathbf{F}(\mathbf{m} - \mathbf{m}_0)\|^2_{L_2}. \tag{12}$$

In order to solve the least-squares problem, the objective function $E(\bullet)$ should be minimized. This means that we should apply ∇ to equation (12) and set it to zero. This is the necessary condition for making the objective function minimum. This yields:

$$\mathbf{F}^T\mathbf{U}_s = \mathbf{F}^T\mathbf{F}(\mathbf{m} - \mathbf{m}_0). \tag{13}$$

Equation (13) is called the normal equation from which $\mathbf{m}$ can be estimated (provided that the inverse of the square matrix $\mathbf{F}^T\mathbf{F}$ exists).

2.3 Null vectors

Minimization of equation (12) will not give us a unique solution because of null vectors of the Frechét matrix $\mathbf{F}$ which satisfy the following equation:

$$\mathbf{F}(\delta\mathbf{m}) = 0, \tag{14}$$

where

$$\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0. \tag{15}$$

A generalized solution to (12) can be sought in the following form:

$$\delta\mathbf{m} = \delta\mathbf{m}_g + c\delta\mathbf{m}^0, \tag{16}$$

where $\mathbf{m}_g$ is a solution which does not correspond to zero eigenvalue, $\mathbf{m}^0$ corresponds to zero eigenvalues and c is a constant. It is possible to prove that null vectors are perpendicular to $\mathbf{m}_g$. For this reason, let us consider the following chain of expressions:

$$(\mathbf{m}^0, \lambda\mathbf{m}_g) = (\mathbf{m}^0, \mathbf{F}\mathbf{m}_g) = (\mathbf{F}^T\mathbf{m}^0, \mathbf{m}_g) = 0, \tag{17}$$

where λ is a non-zero eigenvalue of the Frechét matrix. This means that all solutions to the conventional least-squares are located at the hyperplane parallel to the null vector with distance from the null vectors being $|\delta\mathbf{m}_g|$. It is also important to recall that the objective function in (12) is a convex function but not in a strict sense. This means that it might have a number of minima which are due to the existence of null vectors.

2.4 The choise of reference velocity

In order to solve a minimization problem, the reference model $\mathbf{m}_0$ should be chosen.

If we are solving non-linear problems, then the reference model is updated in the course of iterations.

Let us consider now a kth iteration

$$\mathbf{U}_{s,k} = \mathbf{F}_k\delta\mathbf{m}_k + \boldsymbol{\epsilon}_k, \tag{18}$$

where ϵ_k represents the noise vector at kth iteration. In this equation, $\mathbf{U}_{s,k}$ represents the calculated response at the kth iteration:

$$\mathbf{U}_{s,k} = \begin{pmatrix} U_{s,k}(\xi_{r,1}) \\ U_{s,k}(\xi_{r,2}) \\ \vdots \\ U_{s,k}(\xi_{r,N}) \end{pmatrix}, \tag{19}$$

where $U_{s,k}(x_{r,i})$ denotes the data measured at the location of the ith receiver and associated with the model at the kth iteration. The Jacobian matrix $\mathbf{F}$ of the Frechét derivatives is a rectangular matrix of which the column vectors span the data space and row vectors span the model space. In equation (18), $\delta\mathbf{m}_k$ is a discrete set of values of perturbations in the model space:

$$\delta\mathbf{m}_k = \begin{pmatrix} \delta m_{k,1} \\ \vdots \\ \delta m_{k,M} \end{pmatrix}. \tag{20}$$

Equation (18) is a linear equation and the matrix $\mathbf{F}$ can be factored using the singular value decomposition (SVD):

$$\mathbf{F} = \mathbf{D}\mathbf{\Lambda}\mathbf{M}^T. \tag{21}$$

In this equation, $\mathbf{D}$ is an $N \times N$ matrix, $\mathbf{M}$ is an $M \times M$ matrix and $\mathbf{\Lambda}$ is a p-diagonal matrix of singular values, where $p = \min(M, N)$. For underdetermined systems, M is larger than N and thus, $p = N$.

Symes & Santosa (1987) and Carrion (1987) found that if the background reference model has discontinuities, these discontinuities add large singular values to the matrix $\mathbf{F}$. If the background model produces reflections or multiples, then this leads to a significant increase in large singular values of the matrix $\mathbf{F}$. Each reflection or multiple adds large singular values to the perturbed singular value spectrum. When large singular values are added to $\mathbf{F}$ by the background velocity, results become very inaccurate and necessitates a significant number of iterations (Gauss–Newton iterations, for example) before any results are obtained. Sometimes iterations do not converge at all (Carrion 1987). This means that in order to obtain accurate results using linearized inversion, the background velocity should be chosen in such a way that it does not generate reflections and multiples which add large singular values to the perturbed model. This will also assure the convergence to a generalized solution.

2.5 'Damped' least-squares

In order to avoid the problem with null vectors and non-uniqueness of a generalized solution, we can modify the objective function E:

$$E(\mathbf{m}) = \|\mathbf{U}_s - \mathbf{F}\delta\mathbf{m}\|_{L_2}^2 + \lambda_L \|\delta\mathbf{m}\|_{L_2}^2. \tag{22}$$

In this equation, λ_L is a positive number and is the so-called 'damping' parameter (see e.g. Lines & Treitel 1984).

The most important characteristic of the modified objective function is that it becomes a convex function in a strict sense and thus, has only one minimum (provided that λ_L is not zero). This also guarantees the uniqueness of a generalized solution since the minimization of (22) eliminates the influence of null vectors.

'Damped' least-squares were used in geophysics to obtain generalized solutions. Keys & Weglein (1983) applied the generalized inversion to a 1-D model. They also showed that the linearized least-squares are directly related to the Born inversion.

Keys (1986) showed that results of the generalized inversion depend on the value of the 'damping' parameter λ_L. Similar results were reported by Foster & Carrion (1986) who applied generalized least-squares inversion to slant stacks.

'Damped' least-squares are widely used in tomographic reconstruction of velocity anomalies (see e.g. Stork & Clayton 1985; Fawcett & Clayton 1984; Bishop *et al.* 1985).

Suppose now that the reference model generates multiples and reflections which we will denote $U_{s,B}$. Equation (22) will then be modified:

$$E(\mathbf{m}) = \|\mathbf{U}_s + \mathbf{U}_{s,B} - \mathbf{F}\delta\mathbf{m}\|_{L_2}^2 + \lambda_L \|\delta\mathbf{m} + \delta\mathbf{m}_B\|_{L_2}^2. \tag{23}$$

The minimization of E in equation (23) leads to the following equation

$$\mathbf{L}^T\mathbf{U}_{s,B} = (\mathbf{F}^T\mathbf{F} + \lambda_L\mathbf{I})\delta\mathbf{m}_B, \tag{24}$$

where $\delta\mathbf{m}_B$ is 'noise' which appears due to reflections and multiples associated with the reference model. We can see from this equation that the magnitude of 'noise' will depend on the operator itself and on the value of the data associated with the background model.

Therefore, for 'damped' least-squares, it is also necessary to chose the reference background in such a way that it does not generate multiples and reflections associated with large singular values.

Another problem with 'damped' least-squares is that the 'damping' is undetermined *a priori* and usually found using trial and error techniques. However, a solution strongly depends on the choice of this parameter. In geophysics, however, we usually chose this parameter as large as possible to be sure that a chosen solution is feasible.

2.6 *A priori* information

In this section, we will consider how *a priori* information can be incorporated into the generalized inversion. The idea of inversion is to recover the subsurface parameters from the recorded data. This is the so-called 'pure' inversion. However, in a typical geophysical experiment, some additional information is always available. For example, in a typical marine experiment, we collect the data in terms of pressure using towing arrays of hydrophones. The data are always corrupted with noise and thus, it is useful to have some *a priori* information in the data space. Along with the data space, it is always possible to introduce some *a priori* information in the model space. For example, in a marine experiment, the water layer does not support shear-wave, all model parameters (velocities and densities) are positive defiant and shear-waves velocities at any fixed point of space are less than compressional velocities. Besides that, we always know high and low bounds of compressional and shear velocities. For layered media, these bounds can be estimated from conventional velocity analysis using hyperbolic or elliptic moveouts. For more complicated media, low and high bounds can be found from migration velocities.

The data can be treated as random variable. Random variable can be defined by the probability distribution function P. If the probability distribution function is known, then the variance can be determined from:

$$\sigma^2 = \int d\mathbf{U}_s\{U_s - \mathbf{E}(U_s)\}^2 P(\mathbf{U}_s), \tag{25}$$

where $\mathbf{E}(\bullet)$ is the expected (mean) value of a random variable. Covariance matrix $\mathbf{C}^d$ in the data space is defined as:

$$\mathbf{C}^d_{ij} = \int d\mathbf{U}_s[U_{s,i} - \mathbf{E}(U_{s,i})][U_{s,j} - \mathbf{E}(U_{s,j})]P_d(\mathbf{U}_s). \tag{26}$$

In the model space, the covariance matrix can be written as:

$$\mathbf{C}^m_{ij} = \int d\mathbf{m}[m_i - \mathbf{E}(m_i)][m_j - \mathbf{E}m_j]P_m(\mathbf{m}). \tag{27}$$

When we define the probability function P, then variance and covariance can be defined. It is conventional to use the term 'soft' bounds for determining the probability function in the data space or in the model space. When the probability function is determined in the model space, this will mean that the probability distribution of a solution $\mathbf{m}$ in the model space is known. This is also an example of a 'soft' bound.

Recall that there are an infinite number of probability functions which satisfy the same covariance matrix. We will use this when we discuss the possibility of 'softening' 'hard' bounds.

Tarantola (1984b) recently presented a technique for non-linear inversion with 'soft' bounds. Usually, the probability function is chosen to be Gaussian, since the probability distribution of a sum of random variables with different probability distributions tends to Gaussian if the number of variables is large enough.

Besides 'soft' bounds, we also can incorporate *a priori* information in terms of 'hard' bounds. 'Hard' bounds can be imposed on a solution in the model space. For example:

$$A_i + M_{i,0} \leq m_i \leq B_i + m_{i,0} \tag{28}$$

or in the data space

$$\boldsymbol{\epsilon}^T \mathbf{C}^d \boldsymbol{\epsilon} \leq \gamma, \tag{29}$$

where γ can be chosen with a 95 confidence region according to a chi-squared distribution (Carrion, Auyeung & Mersereau 1988).

2.7 Softening 'hard' bounds

In this section, we will consider how 'hard' bounds can be softened. 'Hard' bounds restrict the length of a solution vector in the model space. 'Hard' bounds can be smoothed if instead of limiting the length of a solution, we determine the probability distribution of the solution in the model space. Suppose this distribution is Gaussian

$$P_m(\mathbf{m}) = \frac{1}{\sqrt{2\pi\delta_m^2}} \exp\left\{-\frac{(\mathbf{m} - \mathbf{E}(\mathbf{m}))^2}{2\sigma_m^2}\right\} \tag{30}$$

which defines the probability in the model space with variance σ_m. The probability that a solution lies between $\mathbf{m}_0$

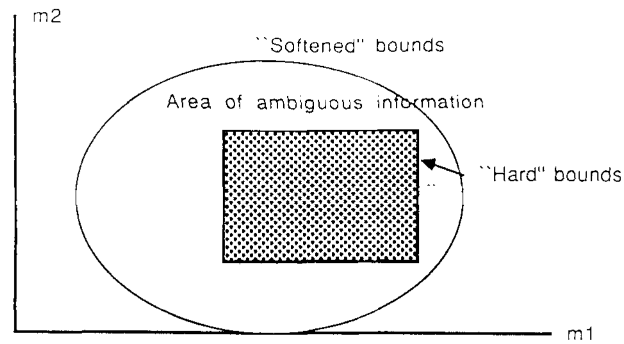

Figure 1. Softening 'hard' bounds leads to adding some ambiguous information which might contradict the correct information about the possible length of a solution.

and $\mathbf{m}_0 + \delta\mathbf{m}$ is

$$P_m(\delta\mathbf{m}) = P_m(\mathbf{m})\delta\mathbf{m}. \tag{31}$$

'Hard' bounds specify the largest length of the model vector

$$|\delta\mathbf{m}| \leq A. \tag{32}$$

In order to find the variance of a solution, equation (31) should be integrated from $-\infty$ to ∞. The square-root of this integral will give the confidence interval, which can be much larger than A which specifies 'hard' bounds in equation (32). This means that 'soft' bounds may introduce incorrect information about a solution. (See also Backus 1987.)

Suppose now that we choose the probability distribution function in the model space as follows:

$$P_m(\mathbf{m}) = \begin{cases} \text{Gaussian} & \text{for } |\mathbf{m}| \leq A \\ 0 & \text{elsewhere} \end{cases}. \tag{33}$$

This probability function satisfies 'hard' bounds (32). The covariance matrix, however, with this probability function will be the same if other probability functions are considered. However, other probability functions will contradict 'hard' bounds (32) (Fig. 1).

Therefore, in later sections, we will consider non-linear inversion of seismic data with 'hard' bounds.

2.8 Hardening 'soft' bounds

In the previous section, we saw that softening 'hard' bounds can lead to the introduction of some ambiguous information which will contradict 'hard' bounds constraints. This will lead to inaccurate and sometimes erroneous solutions.

In this section, we will discuss what might happen if 'soft' bounds are hardened. Let us introduce 'soft' bounds constraints in terms of the covariance operator $\mathbf{C}^\epsilon$ which describes white noise in the data space. This operator will then be a diagonal identity matrix. The expression

$$\boldsymbol{\epsilon}^T \cdot \boldsymbol{\epsilon} = R_\epsilon^2 \tag{34}$$

will describe a multidimensional sphere with radius $|R_\epsilon|$. If the noise is not white, then equation (34) with a covariance matrix $\mathbf{C}^\epsilon$ will describe an ellipsoid. 'Soft' bounds in terms of white noise can be hardened if the sphere (34) is bounded by a sphere of radius R^* which is larger than R_ϵ ($R^* > R_\epsilon$).

These hardened constraints can then be written as

$$\boldsymbol{\epsilon}^T \cdot \boldsymbol{\epsilon} \leq (R^*)^2. \tag{35}$$

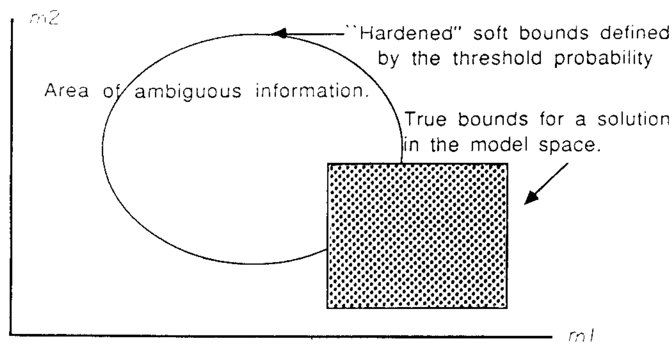

Figure 2. Hardening 'soft' bounds depends on the choice of the threshold of probability distribution function and thus, may introduce erroneous information about solution.

We see that hardening 'soft' bounds in the data space alone will not give us any explicit restrictions on the length of a solution. Let us discuss what happens if we harden 'soft' bounds in the model space. For this reason, let us consider a probability function $P_m(\mathbf{m})$. Then the probability that a solution les between $\mathbf{m}_0$ and $m + \mathbf{m}_0$ is $P(\mathbf{m})\delta m$. In order to harden 'soft' bounds (the probability function P), we should define a threshold probability P^*. The threshold probability function means that if the probability in the model space is less than P^*, the probability should be set to zero. This we will call 'hardening' soft bounds in the model space.

$$P(\mathbf{m}) = \begin{cases} P(\mathbf{m}) & \text{if } P \geq P^* \\ 0 & \text{elsewhere} \end{cases}. \qquad (36)$$

The threshold probability will define the confidence interval and thus, will put some constraints on the length of a solution vector. However, this maximum length of a solution vector will depend on our choice of the threshold and thus, is subjectively biased. Fig. 2 describes this statement. We see that softening 'hard' bounds or hardening 'soft' bounds can introduce ambiguous information about a solution. Instead, 'hard' bounds should be introduced in the inversion independently.

3 NON-LINEAR GENERALIZED INVERSION WITHOUT CONSTRAINTS

3.1 Basic principles of unconstrained non-linear inversion

In this section, we will consider an unconstrained non-linear generalized inversion. Recently, this type of inversion was developed by Tarantola (1984b), Gauthier, Virieux & Tarantola (1986), McAulay (1986), Pan, Phinney & Odom (1986) and Mora (1987), among others.

The idea of the unconstrained generalized inversion is the following. We will try to fit the observed data $\mathbf{D}$ by the computed response $U(\mathbf{m})$. We will consider the objective function $E(\bullet)$ which can be written as:

$$E(\mathbf{m}) = \|\mathbf{D} - U(\mathbf{m})\|_{L_2}^2 + \lambda_L \|\delta\mathbf{m}\|_{L_2}^2 \qquad (37)$$

which can be also written as:

$$E(\mathbf{m}) = (\mathbf{D} - U(\mathbf{m}))^T(\mathbf{C}^d)^{-1}(\mathbf{D} - U(\mathbf{m})) + \lambda_L \delta\mathbf{m}^T(\mathbf{C}^m)^{-1}\delta\mathbf{m} \qquad (38)$$

(assuming that functions are real. If functions are not real, then the transpose operation should be changed on adjoint).

Let us apply ∇ to equation (38). This will yield:

$$\mathbf{m} - \mathbf{m}_0 = \lambda_L^{-1}\mathbf{C}^m\mathbf{F}^T(\mathbf{C}^d)^{-1}(\mathbf{D} - \mathbf{U}(\mathbf{m})) \qquad (39)$$

which is similar to Tarantola (1984b, equation 14). The next step is to regularize equation (39). For this reason, we will write equation (39) in the 'Tichonov' form:

$$\mathbf{m} = \mathbf{m}_0 + [\lambda_L\mathbf{I} + \mathbf{C}^m\mathbf{F}^T(\mathbf{C}^d)^{-1}\mathbf{F}]^{-1} \times (\mathbf{C}^m)\mathbf{F}^T(\mathbf{C}^d)^{-1}[\mathbf{D} - U(\mathbf{m}) + \mathbf{F}(\mathbf{m} - \mathbf{m}_0)]. \qquad (40)$$

This equation can be solved using an iteration procedure

$$\mathbf{m}_{k+1} = \mathbf{m}_0 + [\lambda_L\mathbf{I} + (\mathbf{C}^m)\mathbf{F}_k^T(\mathbf{C}^d)^{-1}\mathbf{F}_k)^{-1}]\mathbf{C}^m) \times \mathbf{F}_k^T(\mathbf{C}^d)^{-1}[\mathbf{D} - U(\mathbf{m}) + \mathbf{F}_k(\mathbf{m}_k - \mathbf{m}_0)]. \qquad (41)$$

This equation is also similar to one suggested by Tarantola (1986). In equation (41), the 'damping' parameter λ_L stabilizes the inversion since it removes small eigenvalues from zero. We should also mention that since covariance is a symmetric positive definite matrix, then the operator $\lambda_L\mathbf{I} + \mathbf{C}_m\mathbf{F}^T\mathbf{C}^{-1}\mathbf{F}$ has always inverse (provided that the damping parameter is a positive number).

3.2 The small wave number problem in unconstrained inversion

It was shown that non-linear unconstrained inversion does not efficiently recover small wave numbers (see Gauthier *et al.* 1986: Mora 1987).

There are two basic reasons why non-linear generalized unconstrained inversion is not able to reconstruct small wave numbers. The first reason is that the seismic data are band-limited and thus, information about low-frequencies is missing. Low-frequencies and thus, small wave numbers can be picked up from larger offset events.

Secondly, travel-time curves for limited aperture CDP data have much less information about small vertical wave numbers than transmitted data.

However, the most important problem is that small wave number information which is contained in CDP data is not enough to accommodate non-linear generalized unconstrained inversion. Thus, the most crucial problem in non-linear unconstrained generalized inversion is where to get more information about small wave numbers and how to incorporate this information into the inversion procedure. Fig. 3 illustrates this statement. Mora (1987) suggested adding small wave numbers by including the VSP data which

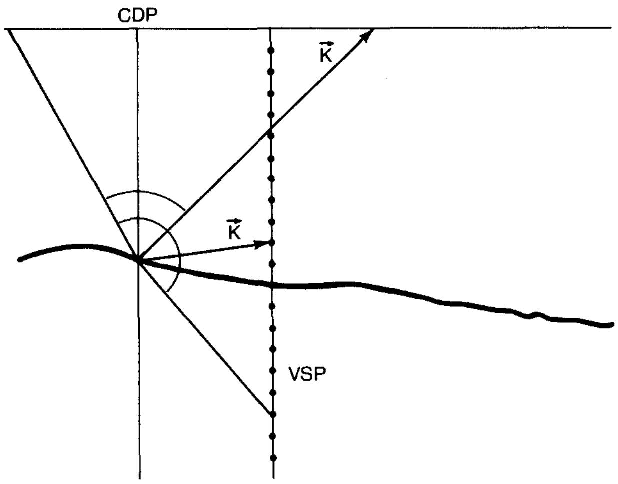

Figure 3. CDP data which have hyperbolic trajectories do not contain small vetical wave numbers, whereas borehole data contain smaller vertical wave numbers.

have a substantial contribution from direct transmitted waves. Thus, the problem is not only to include large apertures but to also include transmitted arrivals from borehold data.

In the following sections, we will see how small wave numbers can be obtained from 'hard' bounds which means that only CDP data can accommodate constrained non-linear generalized inversions.

4 CALCULUS OF CONVEX FUNCTIONALS

4.1 Convex sets and functions

In this section, we will recall the major properties of the calculus of convex functionals. A subset C is said to be convex if $ax+(1-a)y$ belongs to this subset whenever x and y are in the subset and $0 \le a \le 1$. This simply means that for every pair of points in C the straight line segment linking these points lies wholly within C (Fig. 4).

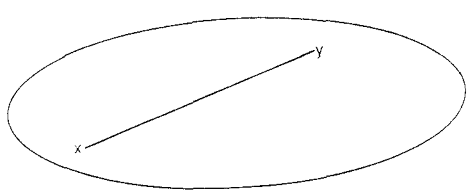

Figure 4. An example of a convex sheet.

When we try to solve constrained optimization problems, it is important to introduce convex functions defined on convex subsets.

One of the important characteristics of convex functions is the epigraph of functions. (Fig. 5) depicts a convex and a non-convex function.

The function f is called a convex function if the epigraph of this function, epi (f), is a convex subset and a concave function if f is convex.

If f is a proper convex function, then:

$$f(ax+(1-a)y) \le af(x)+(1-a)f(y) \qquad (42)$$

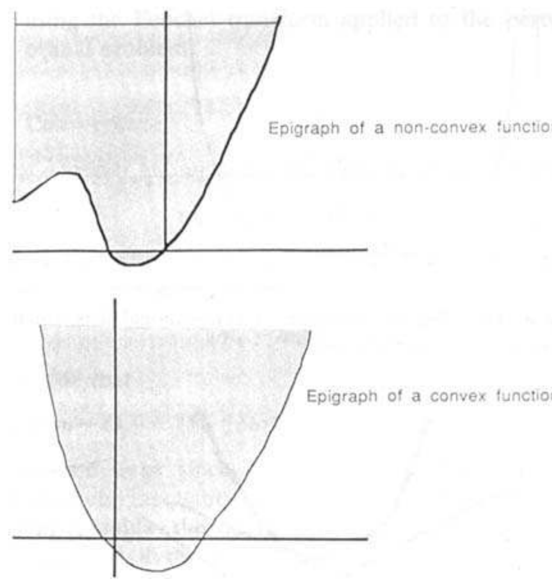

Figure 5. Epigraph of a non-convex function (top) and a convex function (bottom).

whenever x and y belong to the effective domain of this function and $0 \le a \le 1$.

Remark. The effective domain dom $f(x)$ is a set bounded by this function itself where f is not infinity. ($|f(x)| < \infty$). The function is convex when its effective domain dom (f) is convex.

4.2 Example 1

Let us prove that the objective function E in equation (12) is convex. Since in this equation, $\mathbf{U}_s$ does not depend on the model m, it is sufficient to prove that $\|\mathbf{F}\delta\mathbf{m}\|$ is convex. We will consider function

$$f=\mathbf{F}\{ax+(1-a)y\}. \qquad (43)$$

We will rewrite this equation in the L_2 norm and apply the triangular inequality. This yields:

$$\|\mathbf{F}\{ax+(1-a)y\}\| \le |a|\ \|\mathbf{F}x\| + |1-a|\ \|\mathbf{F}y\| \qquad (44)$$

which proves that the function is convex. However, in order to analyse this function in terms of strict convexity, we should show that this function has only one minimum. We see that it is not the case. As discussed before, it is possible to see that any null vector can be added or subtracted from a solution and the whole objective function E does not change. This means that this function is convex but not in a strict sense. Fig. 6 illustrates this statement.

4.3 Affine minorants and Fenchel transform

Hyperplanes (lines in the 2-D plane) which intersect the effective domain of the convex function at most at the boundary points are called affine minorants. Fig. 7 shows affine minorants of a convex function.

Let us consider hyperplanes (lines in the 2-D plane)

$$Z(x)=\mathbf{c}^T\mathbf{x}-b, \qquad (45)$$

where $\mathbf{c}$ is a constant vector which defines the angle of

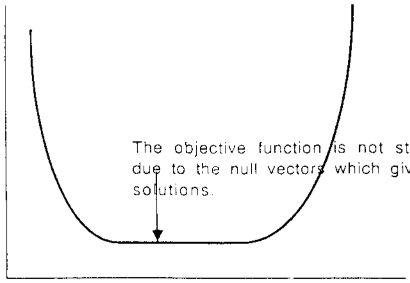

Figure 6. The objective function of conventional least-squares is convex but not in a strict sense.

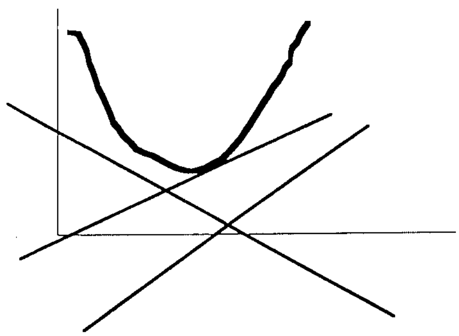

Figure 7. Affine minorants of a convex function.

inclination of the line and b is the intercept. It is clear that in order that $Z(x)$ be a minorant, the following inequality should hold:

$$\mathbf{c}^T\mathbf{x} - b - f(x) \le 0. \quad (46)$$

Let us determine now the least value b can take to remain a minorant (for fixed $\mathbf{c}$). It is possible to see that the least value for b will be defined by a minorant which intersects with the function itself and thus, satisfies the following equation:

$$f^*(\mathbf{c}) = \sup_x (\mathbf{c}^T\mathbf{x} - f(\mathbf{x})). \quad (47)$$

(See e.g. Ekeland & Turnbull 1983). This least value of b is called the Fenchel transform of the convex function $f(x)$ (see Fig. 8).

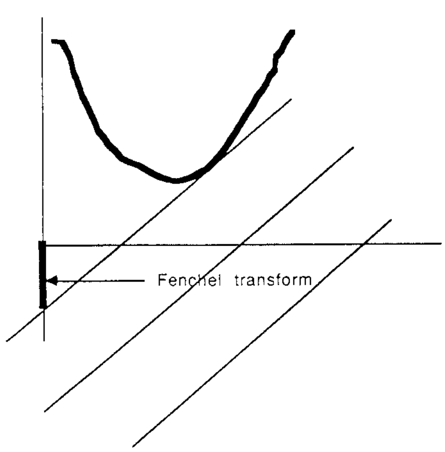

Figure 8. Minorants of the convex function and its Fenchel transform which finds the least intercept of all minorants with the same angle of inclination.

4.4 Example 2

Let us calculate the Fenchel transform of the function $f(x)$ given as:

$$f(x) = \begin{cases} x^2 + 1 & \text{if } x \le 0 \\ x^3 + 1, & \text{if } x \ge 0 \end{cases}. \quad (48)$$

Let us apply the Fenchel transform to this function:

$$f^*(c) = \sup_x \{cx - f(x)\}. \quad (49)$$

The point where the function $cx - f(x)$ is maximum for $x < 0$ is $x = c/2$. For positive x, the point when the function $cx - f(x)$ is maximum is $x = \sqrt{c/3}$. Then the Fenchel transform of $f(x)$ is

$$f^*(c) = \begin{cases} c^2/4 - 1 & \text{if } c < 0 \\ \dfrac{2(c)^{3/2}}{\sqrt{3} - 1} & \text{for } c \ge 0 \end{cases}. \quad (50)$$

4.5 Primal and dual problems

Let us formulate the primal problem. We will try to minimize the model vector subject to some constraints (using 'hard' bounds constraints).

The primal problem can be formulated as follows:

$$PRIM(\delta\mathbf{m}) = \begin{cases} \frac{1}{2}\,\|\delta\mathbf{m}\|^2_{L_2}, & \text{if } A_i \le \delta m_i \le B_i,\ \boldsymbol{\epsilon}^T(\mathbf{C}^d) - 1\boldsymbol{\epsilon} \le \gamma \\ \infty & \text{otherwise} \end{cases} \quad (51)$$

(see e.g. Carrion *et al.* 1988). The primal problem (51) can be presented as a sum of two primal problems

$$PRIM(\delta\mathbf{m}) = PRIM_1(\delta\mathbf{m}) + PRIM_2(\boldsymbol{\epsilon}), \quad (52)$$

where

$$PRIM_1(\delta\mathbf{m}) = \begin{cases} \frac{1}{2}\,\|\delta\mathbf{m}\|^2_{L_2}, & \text{if } A_i \le \delta m_i \le B_i \\ \infty & \text{otherwise} \end{cases} \quad (53)$$

and

$$PRIM_2 = \begin{cases} 0, & \text{if } \boldsymbol{\epsilon}^T(\mathbf{C}^d)^{-1}\boldsymbol{\epsilon} \le \gamma \\ \infty & \text{otherwise} \end{cases}. \quad (54)$$

Along with the primal problem, we will consider the perturbation problem which can be written as:

$$PERT(\delta\mathbf{m}, \sigma) = PRIM_1(\delta\mathbf{m}) + PRIM_2(\boldsymbol{\epsilon} + \boldsymbol{\sigma}). \quad (55)$$

It is possible to see that (see e.g. Ekeland & Turnbull 1983) the primal perturbed problem can be transformed to the dual problem $PERT^*$ with respect to unknown parameters which are called Lagrangian multipliers.

$$PERT^*(\mathbf{F}^T\lambda) = -PRIM_1^*(\mathbf{F}^T\lambda) - PRIM_2^*(-\lambda), \quad (56)$$

where $PRIM_1^*$ and $PRIM_2^*$ are Fenchel transforms of two primal problems in the model and data space ($PRIM_1$ and $PRIM_2$, respectively). Fenchel transforms of these functions were presented by Carrion *et al.* (1987). Here we will write only the final result:

$$PRIM_1^* = -\tfrac{1}{2}\,\|\mathbf{T}\mathbf{F}^T\boldsymbol{\lambda}\|^2_{L_2} + \boldsymbol{\lambda}^T\mathbf{F}\mathbf{T}\mathbf{F}^T\boldsymbol{\lambda} \quad (57)$$

and

$$PRIM_2^*(-\boldsymbol{\lambda}) = \sqrt{\gamma\boldsymbol{\lambda}^T\mathbf{C}^d\boldsymbol{\lambda}} - \mathbf{U}_s^T\boldsymbol{\lambda}. \quad (58)$$

Then, the dual problem can be presented as:

$$PERT^*(\lambda) = \tfrac{1}{2}\,\|\mathbf{TF}^T(\lambda)\|_{L_2}^2 - \boldsymbol{\lambda}^T\mathbf{FTF}^T\boldsymbol{\lambda} + \mathbf{U}_s^T\boldsymbol{\lambda} - \sqrt{\lambda\boldsymbol{\lambda}^T\mathbf{C}^d\boldsymbol{\lambda}}. \tag{59}$$

The primal problem (51) with constraints in the model and data space can be expressed in terms of the dual problem with respect to unknown Lagrangian multipliers λ.

We solve the maximization problem (59) and find optimal Lagrangian multipliers. When optimal Lagrangian multipliers are found, the model vector is computed from:

$$\mathbf{m} = \mathbf{m}_0 + \mathbf{TF}^T\boldsymbol{\lambda}, \tag{60}$$

where $\mathbf{T}$ is the so-called truncation operator which is defined as:

$$\mathbf{Tc} = \begin{cases} c_i, & \text{if } A_i < \delta m_i < B_i \\ A_i & \text{if } \delta m_i < A_i \\ B_i & \text{if } B_i < \delta m_i \end{cases} \tag{61}$$

In the next section, we will discuss how the constrained elastic non-linear problem can be solved.

5 SOLUTION TO A CONSTRAINED ELASTIC PROBLEM

5.1 Iteration procedure for solving the non-linear problem

In the previous sections, we considered how a constrained generalized inverse problem can be solved via the dual transform. Let us now develop an algorithm for solving non-linear generalized constrained inversion.

We will define the generalized non-linear inversion as: to find the model vector $\mathbf{m}$ which satisfies the non-linear wave equation (1) which can be written in the functional form (6). We deal with the measured data D and will minimize the objective function

$$E(m) = \|\mathbf{m} - \mathbf{m}_0\|_{L_2}^2 \tag{62}$$

subject to

$$A_i \geq \{\mathbf{m} - \mathbf{m}_0\}_i \leq B_i \tag{63}$$

and

$$(D - U(m))^T\mathbf{C}^{-1}(D - U(m)) \leq \gamma. \tag{64}$$

Equation (64) can be written as:

$$[D_s - \mathbf{F}_k(\mathbf{m}_k - m_0]^T\mathbf{C}^{-1}[D_s - \mathbf{F}_k(\mathbf{m}_k - m_0] \leq \gamma. \tag{65}$$

This equation shows that if we know $\mathbf{F}_k$ at each iteration, we can solve a non-linear problem iteratively. It is interesting to note that (65) guarantees that a solution will converge.

However, the problem (62)–(63) and (65) has a solution expressed in the dual space as:

$$\mathbf{m}_k = \mathbf{m}_0 + \mathbf{TF}_k^T\lambda_k. \tag{66}$$

This suggests the following steps in solving the non-linear constrained problem:

1. The generalized inverse problem presented in terms of the primal problem (62)–(64)
2. The primal non-linear problem is linearized using the computed Jacobian matrix–Frechét derivatives.
3. For each linear iteration, the dual problem is solved using the Fenchel transform applied to the perturbed primal problem.

5.2 Convergence

Let us consider an iteration procedure in the dual space:

$$\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{TF}_k^T\lambda_k. \tag{67}$$

Suppose that with growing k, iterations diverge. This means that we can choose any number L that

$$\|m_k - m_0\| = L. \tag{68}$$

If it is true that

$$\|\mathbf{F}\delta m - \mathbf{D}_s\| \geq \|\mathbf{F}\|\,\|\delta m\| - \|\mathbf{D}_s\| = L\,\|\mathbf{F}\| - \|\mathbf{D}_s\| \tag{69}$$

and can be large since L can be arbitrarily large. This contradicts the inequality (65). This means that if the primal problem is stable, the dual problem is stable as well. It is also clear that if the primal problem converges, the dual problem converges as well. One of the problems which remains open is the convergence of the primal non-linear problem. Convergence will certainly depend upon the non-linear functional $U(m)$.

5.3 Computation of Frechét derivatives

Computation of the Frechét derivatives has been discussed in several papers for acoustic or elastic set-up (see Tarantola 1984b; Gauthier *et al.* 1986; Pan, Phinney & Odom 1988; Mora, 1987). Let us recall basic principles of calculations of the Frechét derivatives. Besides the Frechét matrix, we should calculate the transpose to the Frechét matrix. This can be done using the equality

$$(\mathbf{F}\delta\mathbf{m}, \mathbf{U}) = (\delta\mathbf{m}, \mathbf{F}^T\mathbf{U}). \tag{70}$$

Frechét derivatives should be calculated at each iteration. When the Frechét derivatives are calculated and the transposed Frechét matrix is found, a solution is sought from equation (66).

5.4 Small wave numbers from constraints

In this section, we will consider how small wave numbers can be obtained from constraints. Suppose that we solve unconstrained generalized problem. Then we should set $\mathbf{T} = \mathbf{I}$. Then for the unconstrained problem

$$\mathbf{m}_k = \mathbf{m}_0 + \mathbf{F}_k^T\boldsymbol{\lambda}_k. \tag{71}$$

Suppose that the initial model is taken to be homogeneous. Then if for CDP data we do not have small wave numbers, the result will not have small wave numbers either. In particular, if in the course of calculations, $\mathbf{F}_k^F\boldsymbol{\lambda}_k$ becomes small, the final result will be almost homogeneous (flat) corresponding to the lack of small wave numbers.

Suppose now that we have constraints in the form of 'hard' bounds. For small values of $\mathbf{F}_k\boldsymbol{\lambda}_k$ we will then have

$$[\mathbf{TF}_k^T\lambda_k] = A_i \tag{72}$$

and thus, a result will not be flat (since information about small wave numbers is incorporated in the constraints). Fig. 9 illustrates this concept.

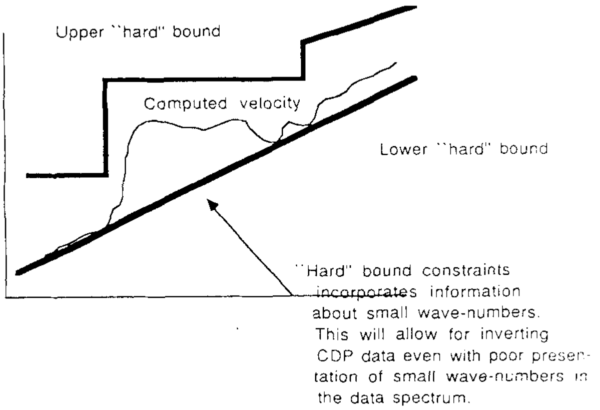

Figure 9. Small wave numbers in constrained generalized inversion can be incorporated in terms of proper constraints.

6 EXAMPLE: 1-D INVERSION

Let us illustrate the proposed technique on a simple example of 1-D data. More complicated numerical examples related to inversion of 2-D and 3-D data will be published elsewhere. However, the convergence can be easier seen on a simple example of 1-D inversion. Fig. 10 shows the acoustic impedance as function of detph. Fig. 11 shows constraints imposed on a solution. The constraints were chosen to satisfy positiveness of the acoustic impedance function (lower constaint). Higher constraint was chosen in the form of the ramp function which prevent the acoustic impedance to rapidly grow in the course of calculations. Fig. 12 represents the initial guess which does not include low velocity zones.

Figure 13 is a result after seven iterations. Finally Fig. 14 depicts the acoustic impedance after 16 iterations. It

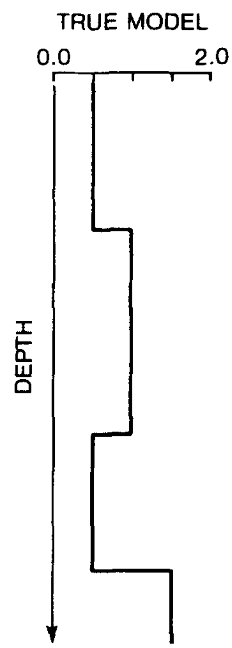

Figure 10. Acoustic impedance. Model.

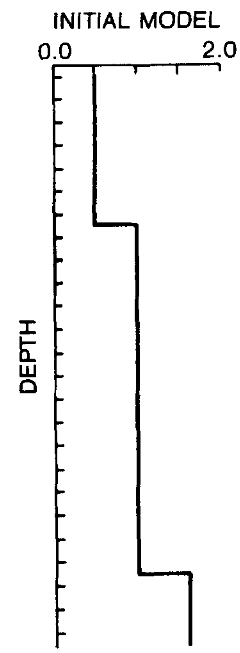

Figure 12. Initial guess which does not contain low velocity zones.

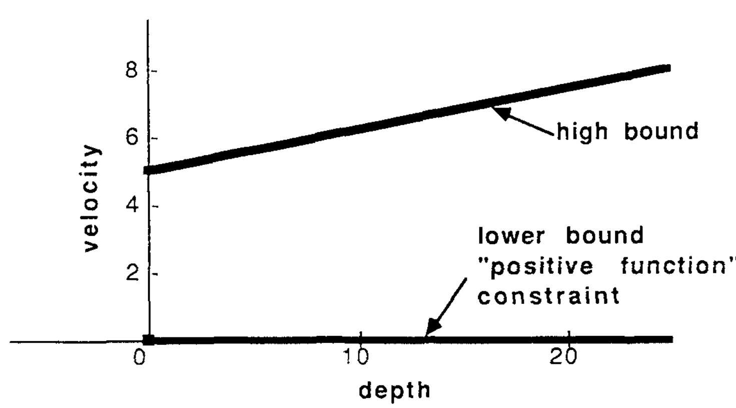

Figure 11. Constraints in the model space. The lower constraint makes velocity to remain positive in the course of iterations. The higher constraint in the form of 'ramp' function precludes the velocity rapidly grow in positive direction.

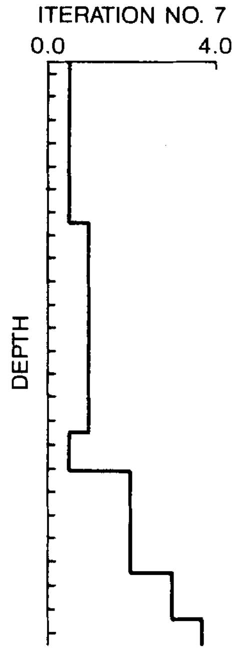

Figure 13. Acoustic impedance after seven iterations.

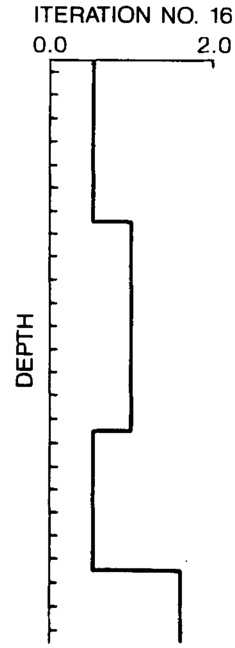

Figure 14. After 16 iterations acoustic impedance converges to the true model.

matches the model very well. Although a non-impulsive source was used for inversion, results do not show instabilities related to the band-limited nature of the source.

7 CONCLUSION

In this paper, I considered a constrained generalized non-linear inversion. It was shown that the generalized non-linear problem with constraints can be reduced to an unconstrained generalized problem in the dual space. In the dual space, constraints are introduced in terms of the truncation operator which becomes the identity matrix for unconstrained problems.

The problem in the dual space is considered in terms of the optimal Lagrangian multipliers which are calculated directly using the Fenchel transform applied to the primal problem.

The role of constraints in exploration geophysics is to incorporate maximum *a priori* information into the inversion procedure. It is well-known that, in general, CDP data are poorly represented by small vertical wave numbers. Thus, unconstrained inversion of only CDP data usually gives inaccurate results. Therefore, it is always useful to incorporate VSP data in the generalized unconstrained inversion. The generalized unconstrained inversion becomes a mixture of tomography and non-linear least-squares filtering of the CDP data.

Properly chosen constraints incorporate small wave numbers into generalized inversion and thus, is possible to invert CDP data without using tomographic (transmission) arrivals.

Another advantage of the method is that it reduces the dimension of the problem (as discussed by Carrion *et al.* 1987). This property is especially useful for underdetermined systems where the number of unknowns is much more than the number of available data.

REFERENCES

Aminzadeh, F., 1984. Derivation of layer parameters of an elastic medium from reflection coefficient matrices, *Geophys. Prosp.*, **32,** 819–827.

Backus, G., 1988. Comparing hard and soft prior bounds in geophysical problems, Technical Report, Scripps Oceanographic Institute.

Beylkin, G., 1985. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform, *J. Math. Phys.*, **26,** 99–108.

Bishop, T., Bube, K., Cutler, R., Langan, R., Love, P., Resnick, J., Shuey, R., Spindler, D. & Wyld, H., 1985. Tomographic determination of velocity and depth in laterally varying media, *Geophysics* **50,** 903–923.

Bleistein, N., Cohen, J. & Haigin, F., 1985. Computational and asymptotic aspects of velocity inversion, *Geophysics,* **50,** 1253–1565.

Bleistein, N. & Gray, S., 1985. An extension of the Born inversion technique to a depth dependent reference profile, *Geophys. Prospect.*, **33,** 99–1022.

Bleistein, N., 1987. On the imaging of reflectors in the Earth, *Geophysics,* **52,** 931–942.

Brysk, H. & McCowan, D., 1986a. A slant stack procedure for point source data, *Geophysics,* **51,** 1370–1386.

Brysk, H. & McCowan, D. 1986b. Direct inversion of slant-stacked seismic data. Part I. Synthetic seismogram results, *Bull. seism. Soc. Am.*, **76,** 815–835.

Bube, K. & Burridge, R., 1983. The one-dimensional inverse problem of reflection seismology, *SIAM Rev.*, **4,** 497–559.

Carrion, P. & Patton, W., 1983. Criteria for resolution and reconstruction of the acoustic impedance, *J. geophys. Res.*, **88,** 10349–10358.

Carrion, P., 1985. Computation of velocity and density profiles of acoustic media with vertical inhomogeneities using the method of characteristics applied to the slant stacked data, *J. acoust. Soc. Am.*, **77,** 1370–1376.

Carrion, P., Auyeung, C. & Mersereau, R., 1988. Constrained travel-time tomography, Presented at the 58th SEG meeting, Anaheim, CA.

Carrion, P. & VerWest, B., 1987. A procedure for inverting seismic data to obtain *Q*-profiles, *Inverse probl.*, **3,** 65–71.

Carrion, P., 1987. *Inverse Problems and Tomography in Acoustics and Seismology,* Penn. Publ. Co, Atlanta, GA.

Chung, W., Carrion, P. & Beylkin, G., 1987. Wavefront sets of solutions to linearized inverse scattering problems, *Inverse Prob.*, **3**, 683–690.

Clayton, R. & Stolt, R., 1981. A Born WKBJ inversion method for acoustic reflection data, *Geophysics*, **46**, 1559–1567.

Coen, S., Cheney, M., & Weglein, A., 1984. Velocity and density of a two-dimensional acoustic medium on point source surface data, *J. Math. Phys.*, **25**, 1857–1861.

Cohen, J. & Bleistein, N., 1977. An inverse method for determining small variations in propagation speed, *SIAM J. appl. Math.* **32**, 784–799.

Ekeland, I. & Turnbull, T., 1983. *Infinite-dimensional Optimization and Convexity*, University of Chicago Press.

Fawcett, J. & Clayton, R., 1984. Tomographic reconstruction of velocity anomalies, *Bull. seism. Soc. Am.*, **74**, 2201–2219.

Foster, D. & Carrion, P., 1986. Generalized Born inversion of seismic reflection data, *Geophys. J. R. astr. Soc.*, **85**, 329–347.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Non-linear inversion of multi-offset seismic reflection data, *Geophysics*, **51**, 1387–1403.

Gray, S. & Symes, W., 1985, Stability considerations for one-dimensional inverse problems, *Geophys. J. R. astr. Soc.*, **80**, 149–163.

Keys, R., 1986. An application of Marquardt's procedure to the seismic inverse problem, *Proc. IEEE*, **74**, 477–486.

Keys, R. & Weglein, A., 1983. Generalized linear inversion and the first Born theory for acoustic media, *J. Math. Phys.*, **24**, 1444–1449.

Lines, L. & Treitel, S., 1984. A review of least-squares inversion and its application to geophysical problems, *Geophys. Prospect.*, **32**, 159–186.

Lailly, P., 1984. Migration methods: partial but efficient solutions to the seismic inverse problems, in *Inverse Problems of Acoustic and Elastic Waves*, ed. Santosa, F., *et al.* SIAM.

McAulay, J. 1986. Plane-layer prestack inversion in the presence of surface reverberations, *Geophysics*, **51**, 1789–1800.

Meadows, M. & Coen, S., 1986. Exact inversion of plane-layered isotropic and anisotropic elastic media by the state-space approach, *Geophysics*, **51**, 2031–2050.

Mora, P., 1987. Non-linear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, **52**, 1211–1228.

Pan, G., Phinney, R. & Odom, R., 1988. Full-waveform inversion of plane wave seismograms in stratified acoustic media: theory and feasibility, *Geophysics*, **53**, 21–31.

Parsons, R., 1986. Estimating reservoir mechanical properties using constant offset images of reflection coefficients and incident angles, *Presented at the 56th Annual SEG Meeting, Houston, TX.*

Robinson, E., 1982. Spectral approach to geophysical inversion by Lorentz, Fourier, and Radon transform, *Proc. IEEE*, **70**, 1039–1054.

Santosa, F. & Schwetlick, M., 1982. The inversion of acoustical impedance profiles by methods fo characteristics. *Wave Motion*, **4**, 99–110.

Santosa, F. & Symes, W., 1985. The determination of a layered acoustic medium via multiple impedance profile inversions from plane wave responses, *Geophys. J. R. astr. Soc.*, **81**, 175–195.

Santosa, F., Symes, W. & Raggio, G., 1985. Inversion of bandlimited reflection seismograms using stacking velocities as constraints, *Tech. Rep. No.* 29, Cornell University.

Shiva, M. & Mendel, J., 1983. Non-normal incidence inversion: existence of solution, *Geophys. Prosp.*, **31**, 888–914.

Stoffa, P., Buhl, P., Diebold, J. & Wenzel, F., 1981. Direct mapping of seismic data to the domain of intercept time and ray parameter: a plane wave decomposition, *Geophysics*, **46**, 255–268.

Stolt, R., 1978. Migration by Fourier transform, *Geophysics*, **43**, 23–48.

Stolt, R. & Weglein, A., 1985. Migration and inversion of seismic data, *Geophysics*, **30**, 2358–2472.

Stork, C. & Clayton, R., 1985. The iterative tomographic and migration reconstruction of seismic images, *Presented at the 55th Annual SEG Meeting, Washington, DC.*

Tarantola, A., 1986. *Inverse Problem Theory. Methods for Data Fitting and Model parameter Estimation*, Elsevier, New York.

Tarantola, A., 1984a. Linearized inversion of seismic data, *Geophys. Prosp.*, **32**, 998–1015.

Tarantola, A., 1984b, Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.

Treitel, S., Gutowsky, P. & Wagner, D., 1982. Plane wave decomposition of seismograms, *Geophysics*, **47**, 1375–1401.

Treitel, S. & Lines, L., 1982. Linear inverse theory and deconvolution, *Geophysics*, **47**, 1153–1159.

Weglein, A., Violette, P. & Keho, T., 1986. Using a constant background multi-parameter Born theory to obtain exact inversion goals, *Geophysics*, **51**, 1069–1074.

Weglein, A., 1982. Near field inverse scattering formalism for three dimensional wave equation: the inclusion of *a priori* velocity information, *J. acoust. Soc. Am.*, **71**, 1179–1182.

Yagle, A. & Levy, B., 1984. Application of the Shur algorithm to the inverse problems for a layered acoustic medium, *J. accoust. Soc. Am.*, **76**, 301–308.

Yagle, A., 1985. A layer-stripping solution of inverse seismic problems, *PhD thesis*, Department of EE, MIT.

Yagle, A., 1986. Differential and integral methods for multidimensional inverse scattering problems, *J. Math. Phys.*, **27**, 2584.