Geophysical Journal International

Geophys. J. Int. (2013) **195,** 1745–1762
Advance Access publication 2013 September 17

doi: 10.1093/gji/ggt326

Constrained optimization framework for joint inversion of geophysical data sets

A. Sosa,[1,2] A. A. Velasco,[3] L. Velazquez,[1,4] M. Argaez[1,4] and R. Romero[5]

[1] *Program in Computational Science, The University of Texas at El Paso, El Paso, TX* 79968-0514, *USA. E-mail: uasosa@icesi.edu.co*
[2] *Universidad ICESI, Cali, Valle del Cauca, Colombia*
[3] *Department of Geological Sciences, The University of Texas at El Paso, El Paso, TX* 79968-0514, *USA*
[4] *Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX* 79968-0514, *USA*
[5] *Department of Computer Science, The University of Texas at El Paso, El Paso, TX* 79968-0514, *USA*

Accepted 2013 August 13. Received 2013 August 12; in original form 2013 March 6

SUMMARY

Many experimental techniques in geophysics advance the understanding of Earth processes by estimating and interpreting Earth structure (e.g. velocity and/or density structure). Different types of geophysical data can be collected and analysed separately, sometimes resulting in inconsistent models of the Earth depending on the data used. We present a constrained optimization approach for a joint inversion least-squares (LSQ) algorithm to characterize 1-D Earth's structure. We use two geophysical data sets sensitive to shear velocities: receiver function and surface wave dispersion velocity observations. We study the use of bound constraints on the regularized inverse problem, which are more physical than the regularization parameters required by conventional unconstrained formulations. Specifically, we develop a constrained optimization formulation that is solved with a primal-dual interior-point (PDIP) method, and validate our results with a traditional, unconstrained formulation that is solved with a truncated singular value decomposition (TSVD) for a set of numerical experiments with synthetic crustal velocity models. We conclude that the PDIP results are as accurate as those from the regularized TSVD approach, are less affected by noise, and honour the geophysical constraints.

Key words: Numerical solutions; Inverse theory; Computational seismology.

1 INTRODUCTION

The search of meaningful Earth (velocity and density) models requires efficient and robust computational techniques, especially in the context of increasing data collection of multiple types of data. Motivated by the current and novel computational optimization techniques for solving inverse problems (El-Bakry *et al.* 1996; Forsgren *et al.* 2002; Nocedal & Wright 2006) and the advantages of integrating multiple data sets (Haber & Oldenburg 1997; Julia *et al.* 2000, 2005; Gallardo-Delgado *et al.* 2003; Gallardo & Meju 2004; Maceira & Ammon 2009; Moorkamp *et al.* 2010, 2011; Bodin *et al.* 2012), joint inversion of multiple types of data sources can better determine physical properties of the Earth.

Joint inversion in geophysics designates that the simultaneous optimization of several objective functions, such as ℓ_2- misfit functions, allows for estimating a model that explains all data sets at once. Since the objective function is expected to be less subject to local minima, this approach reduces intrinsic non-uniqueness of the problem (Colombo & De Stefano 2007). The simultaneous inversion of multiple data sets is not a new concept, and different approaches have been developed as joint inversion of independent data sets. For example, cooperative inversion (Lines *et al.* 1988) manually adjusts values to improve the fit to the data, yet the approach becomes usually biased to the input model. Also, a weighted scheme involving all the data sets simultaneously can be used for seismic travel times and gravity data (Lees & Vandecar 1991), for DC resistivity and seismic data (Gallardo & Meju 2004), receiver functions and surface wave dispersion (Julia *et al.* 2000), and receiver functions, surface wave dispersion, and magnetotelluric data (Moorkamp *et al.* 2010, 2011). For both approaches (cooperative and weighted), the main assumption in most of the cases is that the data sets comprised in the inversion sample similar geological boundaries. Recently, Bodin *et al.* (2012) demonstrated that Transdimensional Bayesian Inversion is a new approach well suited for combining data sets of different quality, since they allow the flexible parametrizations that are often crucial for combining data that have inherently different sensitivities to structure.

Joint inversion can be considered successful if the following conditions hold (Julia *et al.* 2000): each data set samples the same propagating medium (consistency), and the combination of the data sets increases the resolution of the inverted model (complementarity). The success of this type of inversions, assuming that the above conditions hold, relies on the complementarity between the data sets imposing better physical constraints, while increasing the resolution of the final model. However, difficulties arise for highly non-linear misfit functions and large-dimensional model spaces, which include: (1) the identification of the appropriate weights and the level of influence of each data set over the final inverted model, (2) the lack of complementarity among the data sets, (3) the presence of (spurious) solutions not geophysically consistent and (4) the inherent ill-conditioning of the inverse problem, which necessitates an appropriate choice of regularization and smoothing constraints. In this work we focus on addressing the last two difficulties.

Joint inversion optimization allows for the integration of multiple data sets that have different sensitivities and resolutions. Complementary information that constrains the physics of the problem can be incorporated to the inverse process, and help us to resolve ambiguities. In this paper, we present an alternative scheme, based on a constrained optimization framework, for a joint inversion algorithm (Julia *et al.* 2000). We apply the algorithm to two geophysical data sets, teleseismic *P*-wave receiver functions and surface wave dispersion velocities, for finding a mutually consistent estimate of 1-D Earth structure (e.g. Julia *et al.* 2000, 2005). We characterize the Earth as a layered 1-D structure using seismic shear velocities as a model parameter. Our optimization approach uses bound constraints, which are more physical than the smoothing constraints required in most of the conventional formulations of the inverse problem, and that may require a considerable amount of computations to properly select regularization parameters. We perform synthetic tests to show that our constrained optimization approach, not only can be as accurate as the traditional unconstrained formulation of the inverse problem, but also that our results honour the geophysical constraints while being less affected by noise. We conclude with a discussion on the relevance of the parameters required by the methods with respect to the inversion performance, possible disadvantages of both approaches and the potential applications of our constrained optimization approach.

2 OPTIMIZATION APPROACHES FOR GEOPHYSICAL INVERSION

Many global optimization methodologies have been developed and implemented for inverse problems in geophysics, including genetic algorithms (Shibutani *et al.* 1996; Chang *et al.* 2004; Moorkamp *et al.* 2010), niching genetic algorithm (Lawrence & Wiens 2004), simulated annealing (Vinnik *et al.* 2004, 2006), very fast simulated annealing (Zhao *et al.* 1996), Monte Carlo methods (Sambridge & Mosegaard 2002), and the neighbourhood algorithm (Sambridge 1999a; Bannister *et al.* 2003). These techniques efficiently search a large multidimensional model space and provide complex earth models, yet some are not accessible to linearized schemes (Kozlovskaya *et al.* 2007). Furthermore, the models obtained by global optimization approaches minimize a misfit measure without need of linearization or computation of derivatives, avoiding entrapment of the solution in local minima of the objective function. The increased information implies (1) high computational cost and (2) similar quality among data sets (Moorkamp *et al.* 2010). The first point has become less a concern due to the computational resources currently available, but similar data quality can be difficult using multiple data sets, as it is the choice of regularization parameters for smoothing constraints.

Constrained and unconstrained optimization algorithms have been previously used for solving joint inversion problems (Lines *et al.* 1988). For these strategies, the conventional formulation of the inverse problem is usually presented as an unconstrained optimization formulation. Thus no equality or inequality constraints over the variables are considered, and/or these constraints are replaced by terms added to the objective function to penalize constraint violations. In general, these formulations provide excellent fit to the data but may result in unrealistic models, requiring the use of regularization terms added to the objective function. Using a constrained approach has the advantage of giving control over the model parameters, thus restricting the model parameters to a narrower range of values (Lines *et al.* 1988). By using *a priori* knowledge, model ambiguity can be reduced (Musil *et al.* 2003). A general constrained optimization formulation will consist of both, inequality and equality constraints over the model parameters. Special care must be taken over the constraints, for example, these should not be excessively rigorous to avoid exclusion of potential solutions.

Implementation of equality and inequality constraints is difficult (Backus & Gilbert 1967). The standard implementation uses a Lagrangian method (e.g. Gallardo & Meju 2004). Other work uses a quadratic programming approach based on stable simplex-type solvers, but this is usually computationally expensive (Gallardo *et al.* 2004). Musil *et al.* (2003) use the simplex method to solve a mixed-integer linear programming approach for joint inversion of unconnected physical properties. Li & Oldenburg (2003) use interior-point methods for single inversions in an unconstrained formulation, while Burstedde & Ghattas (2009) use primal-dual active sets for full-waveform inversion. A powerful constrained optimization method known as primal-dual interior point (PDIP) provides for an alternative methodology to solve linear programming problems that became widely implemented after Karmarkar (1984), but not for joint inversion of geophysical data sets in a constrained optimization framework.

Optimization algorithms provide a single solution to a non-unique inverse problem similar to those encountered in geophysical applications. Although any approach may fit the data within errors, the single solution may not represent physically reasonable models. For example the non-uniqueness of receiver functions inversion for 1-D *S*-wave velocity structure (Ammon *et al.* 1990) can been addressed with a Bayesian framework where the objective is to create an ensemble of models that represents the posterior probability distribution, that is, the probability of model parameters, given the observed data (Sambridge 1999b; Lucente *et al.* 2005; Piana Agostinetti & Malinverno 2010; Bodin *et al.* 2012).

3 JOINT INVERSION FORMULATION

In order to characterize the Earth's structure, we estimate the distribution of a physical property that affects seismic wave propagation, such as shear wave (*S*-wave) velocities, which we use for our forward computation. Our approach is a particular case of constrained optimization. In our case, we introduce inequality constraints only over the model parameter to control the model space, therefore introducing intrinsic regularization to the inverse problem. We can compute other properties such as density, compressional (*P*-wave) velocity, and/or layer thickness by using empirical or analytical relationships. For the joint inversion optimization approach, we use the following forward and inverse formulations.

From a given experiment that provides a layered shear velocity distribution $x \in R^n$, we evaluate a forward linear or non-linear operator $F \in R^m$, at the given velocity x, to predict the Earth's response. Here F relates the data and the model space as follows:

$$F(x) = [F_1(x), \ldots, F_m(x)], x = (x_1, \ldots, x_n), (m \gg n), \tag{1}$$

where m is the number of measurements or observations, and n is the number of fixed thickness plane layers. Thus, if we know the velocity model in advance, we can predict the Earth's response corresponding to that model, by using the forward operator F. Given an observed data vector, $y \in R^m$, the inverse problem consists of finding the unknown velocity model, x such that $F(x)$ approximates y as much as possible, that is,

$$\min_x \frac{1}{2} \|F(x) - y\|^2 = \min_x \frac{1}{2} \sum_{i=1}^{m} [F_i(x) - y_i]^2. \tag{2}$$

The inverse problem is generally posed as an unconstrained weighted non-linear least squares (LSQ) problem (Julia *et al.* 2000; Musil *et al.* 2003; Wilson & Aster 2005; Maceira & Ammon 2009; Moorkamp *et al.* 2011; Bailey *et al.* 2012). In our case, we expect to match simultaneously data from different geophysical domains: receiver functions (RF) and surface wave dispersion velocities (SW). Therefore we reformulate (2) as,

$$\min_x \frac{1}{2} \left\|F^{\mathrm{SW}}(x) - y^{\mathrm{SW}}\right\|^2 + \frac{1}{2} \left\|F^{\mathrm{RF}}(x) - y^{\mathrm{RF}}\right\|^2 + \frac{\lambda}{2} \|Lx\|^2, \tag{3}$$

where the first two terms are non-linear ℓ_2- misfit functions between the forward operators (F^{RF}, F^{SW}) and the observations (y^{RF}, y^{SW}), corresponding to RF and SW, respectively. The last term represents regularization with an *a priori* smoothness parameter, λ, and a discrete derivative operator matrix, L, included to avoid sharp velocity changes in adjacent layers, or to smooth velocity variations. For receiver functions, F^{RF} represents the numerical computation of synthetic waveforms (Ammon *et al.* 1990), and for surface waves, F^{SW} represents the numerical evaluation of dispersion velocities (Maceira & Ammon 2009).

For simplicity, we introduce:

$$F(x) = W \begin{bmatrix} F^{\mathrm{SW}}(x) \\ F^{\mathrm{RF}}(x) \end{bmatrix} \in R^m, y = W \begin{bmatrix} y^{\mathrm{SW}} \\ y^{\mathrm{RF}} \end{bmatrix} \in R^m, m = p + q, \tag{4}$$

where we have a weighted diagonal matrix W,

$$W = \mathrm{diag}(w_i), w_i = \sqrt{\frac{\eta}{\sigma_i^2 p}}, \qquad i = 1, \ldots, p, w_i = \sqrt{\frac{1-\eta}{\sigma_i^2 q}}, \qquad i = p+1, \ldots, m,$$

$\mathrm{diag}(w_i)$ denotes a diagonal matrix used to equalize the contribution of each data set with respect to physical units and number of data points, $\eta \in [0, 1]$ is a data set influence parameter that measures the reliability of each data set, σ_i^2 is the approximate standard deviation of each point, and p and q are the number of RF and SW observations respectively. In this work, we assume a typical uncertainty value σ_i^2 of 0.05 (km s^{-1}) for SW and 0.01 (s^{-1}) for RF observations (Julia *et al.* 2000). Hence, we use (4) to rewrite (3) as follows,

$$\min_x \frac{1}{2} \|F(x) - y\|^2 + \frac{\lambda}{2} \|Lx\|^2, \tag{5}$$

The addition of *a priori* information into the regularization term is a well-known technique to improve the condition of the inverse problem (e.g. Tikhonov & Arsenin 1977), sometimes mentioned as a LSQ method with damping (Marquardt 1963). However, how to compute the best value for the parameter λ remains as an open question. Hansen (1987) and Vogel (2002) refer to different techniques that can be used to properly select this parameter. In some cases we require more than the single use of the parameter λ to get reliable solutions; therefore the need of a stronger regularization term may arise. The discrete derivative operator matrices L_i as the ones below, allow us to introduce such type of regularization (Hansen 2010),

$$L_1 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}_{(n-1)\times n}, L_2 = \begin{bmatrix} -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 2 & -1 \end{bmatrix}_{(n-2)\times n}.$$

These matrices further constrain velocities in adjacent layers to avoid rapid velocity variations that may not be physically possible. Therefore *a priori* information about model structure is introduced into the problem by minimizing its roughness or as a trade-off between data misfit and roughness (Constable *et al.* 1987). Julia *et al.* (2000) and Maceira & Ammon (2009) use these types of matrices to help keeping the inversion stable, by compensating the antismoothing effect of the receiver function operator when poor quality data for surface waves is present. The influence parameter, η, can be used to test the compatibility of the data sets comprised in the inversion (Julia *et al.* 2000). For instance, we show in Section 5 how different choices of this parameter affect the final inversion results. In this work, we use η to perform single inversions; in other words, we run the algorithm in absence of one data set. For example, if $\eta = 0$ we use SW dispersion velocities only. In the next subsections, we present two different methodologies to solve the inverse problem by using seismic shear velocities as the model parameter x.

4 METHODOLOGY

The inverse problem is usually posed in an unconstrained form as presented in (5). However, due to the usual highly non-linear behaviour of the operator F, the numerical complexity associated to the computation of higher order derivatives represents a major difficulty. Hence, we solve (5) iteratively as a linearized LSQ (Julia *et al.* 2000; Maceira & Ammon 2009; Bailey *et al.* 2012). To this end, we use a first order Taylor approximation of the operator F around some suitable model, x_k, this is $F(x) \approx F(x_k) + F'(x_k)(x - x_k)$ where $F'(x_k)$ is the matrix with the partial derivatives of F. Thus, we rewrite problem (5) as,

$$\min_x \frac{1}{2} \left\| F'(x_k)\, x + b \right\|^2 + \frac{\lambda}{2} \|Lx\|^2, \tag{6}$$

where $b = F(x_k) - y - F'(x_k)\, x_k$. Even if $F'(x_k)$ is not full column rank, for a given regularization term, there is a unique least squares solution that solves the symmetric positive definite linear system:

$$\underbrace{\left[F'(x_k)^T F'(x_k) + \lambda L^T L\right]}_{H} x = F'(x_k)^T b, \tag{7}$$

which are called the 'normal equations'. From the well-known algorithms to solve this equation, we use the 'truncated singular value decomposition' (TSVD; Wiggins 1972) of the system matrix H, which is also a method of regularization (Hansen 1987). We have adopted this approach for solving (5) to validate our PDIP approach, since TSVD is still a conventional method for solving these problems (Julia *et al.* 2000; Maceira & Ammon 2009; Bailey *et al.* 2012).

In addition to this Gauss–Newton approach, a Levenberg–Marquardt stabilized approach may be considered. The linear least squares results in the normal equations that are better conditioned and then solved much more efficiently by Cholesky factorization. Also, instead of solving the normal equations, the linear least-squares problems may be solved by: (1) using an iterative method, such as LSQR (Musil *et al.* 2003). Conjugate gradient (CG) and non-linear CG (NLCG), directly to (5) (Wilson & Aster 2005; Moorkamp *et al.* 2011; Zhdanov *et al.* 2012), or a limited memory BFGS method (Moorkamp *et al.* 2011). However, for all of these possible choices, regularization parameters must be computed to stabilize the inversion. Such computations are not required in general for our approach outlined below, while adjustment of the physical bound constraints may be needed.

4.1 Truncated Singular Value Decomposition

The expression obtained using a TSVD to compute x can be written as

$$x = \sum_{i=1}^{\tau} \left(\frac{u_i b}{\sigma_i} \right) v_i, \tag{8}$$

where $U \Sigma V = U \mathrm{diag}(\sigma_i)\, V$ is the SVD factorization of the system matrix H, U and V are orthogonal matrices of sizes $m \times m$ and $n \times n$, respectively, and $\Sigma = \mathrm{diag}(\sigma_i)$ is a $m \times n$ diagonal matrix with all singular values σ_i of the matrix H. Here, τ denotes the numerical rank of the matrix H that leads us to a truncated regularized solution whenever a good choice of τ improves the condition of the inverse problem, in other words when the ratio σ_1/σ_τ becomes moderate. In Section 6.3 we explore the threshold that defines τ, such as we can expect a stable inversion. This approach results computationally intense for large-scale systems, and lacks a specific algorithm to compute both the regularization parameter, λ, and the truncation factor, τ. The parameter λ usually represents the trade-off between resolution and stability, by selecting the value that shows an optimal balance (Hansen 2010). We explore this trade-off in Section 6.3.

4.2 PDIP method

We propose a constrained optimization framework to solve (6), the linearized version of the inverse problem (5). The addition of bound constraints over the model parameter reduces model ambiguity (Lines *et al.* 1988) due to the use of *a priori* knowledge to restrict model

parameters to a smaller range of values (Musil *et al.* 2003). Moreover, with this strategy, we remove the regularization term in (6), that is, we set $\lambda = 0$, and instead we add appropriate physical bound constraints over the model parameter x in the inverse problem:

$$\begin{aligned} &\min_x \frac{1}{2} \left\| F'(x_k)x + b \right\|^2 \\ &\text{s.t.} \quad g(x) \geq 0, \end{aligned} \tag{9}$$

where $g(x) = [\, x - v_{\min} \;\; v_{\max} - x \,]^T$ allows us to add physical bounds corresponding to a certain minimum and maximum velocities, that is, $v_{\min} \leq x \leq v_{\max}$. We rewrite problem (9) in a standard non-linear programming form as follows:

$$\begin{aligned} &\min_x \frac{1}{2} \left\| F'(x_k)x + b \right\|^2 \\ &\text{s.t.} \quad g(x) - s = 0 \\ &\qquad s \geq 0, \end{aligned} \tag{10}$$

where $s \in R^{2n}$ is a slack variable. The methodology used to solve (10) is a well-known continuation or homotopy strategy (Karmarkar 1984). In the early 1980s, interior point methods became popular in the optimization community after the success of non-linear programming algorithms to solve large linear programs (minimization of a linear function subject to linear constraints; Forsgren *et al.* 2002; Nocedal & Wright 2006). These linear programs were, and still are usually solved by means of the simplex method (Dantzig 1963). Interior point methods transformed dramatically the theory and practice in the optimization field (Forsgren *et al.* 2002; Nocedal & Wright 2006). The idea of interior resides in the fact that at all times the model parameters satisfy the inequality constraints. By the early 1990s, one of the subclasses of interior point methods known as primal dual appears as the most theoretically elegant and robust approach for this methodology (El-Bakry *et al.* 1996; Forsgren *et al.* 2002).

In this work, we use PDIP methods to solve (10). To this end we need to define the Lagrangian function associated to (10):

$$L(x, z, s) = \frac{1}{2} \left\| F'(x_k)x + b \right\|^2 - [g(x) - s]^T z, \quad (s, z) > 0, \tag{11}$$

where $z \in R^{2n}$ is the Lagrange multiplier associated to the inequality constraints. In these methods, we use a *perturbation* parameter $\mu > 0$, to define the perturbed Karush–Kuhn–Tucker (KKT) or necessary conditions as follows,

$$\begin{bmatrix} \nabla_x L(x, z, s) \\ \nabla_z L(x, z, s) \\ SZe - \mu e \end{bmatrix} = \begin{bmatrix} F'(x_k)^T (F'(x_k)x + b) - \nabla g(x)^T z \\ g(x) - s \\ SZe - \mu e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad (s, z) > 0, \tag{12}$$

where $S = \operatorname{diag}(s_1, \ldots, s_{2n})$, $Z = \operatorname{diag}(z_1, \ldots, z_{2n})$, $e = (1, \ldots, 1) \in R^{2n}$, and ∇ denotes the gradient operator. Next, we define the following Newton system associated to (12),

$$\begin{bmatrix} F'(x_k)^T F'(x_k) & \nabla g(x)^T & 0 \\ \nabla g(x) & 0 & -I \\ 0 & S & Z \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \\ \Delta s \end{bmatrix} = - \begin{bmatrix} F'(x_k)^T (F'(x_k)x + b) - \nabla g(x)^T z \\ g(x) - s \\ SZe - \mu e \end{bmatrix}, \qquad (g(x), s, z) > 0. \tag{13}$$

Here, the system's matrix is called the Jacobian, and the last block of equations in this linear system is known as the complementarity conditions, a fundamental ingredient of PDIP methods (El-Bakry *et al.* 1996). We use a methodology known as 'path-following' strategy to solve (13), as defined in Argaez & Tapia (2002). In this methodology, there are two main components: (1) forcing positivity of the iterates, x_k, to keep them in the interior, and (2) a globalization strategy to guarantee descent directions and progress towards the constraints. For a perturbation parameter $\mu > 0$, and working from the interior, that is, $[g(x), z, s] > 0$, we apply a linesearch Newton's method (Nocedal & Wright 2006) to the perturbed KKT conditions (12), which leads us to the reduced linear system

$$\begin{aligned} &\begin{bmatrix} -F'(x_k)^T F'(x_k) & \nabla g(x)^T \\ \nabla g(x) & Z^{-1}S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = \begin{bmatrix} F'(x_k)^T (F'(x_k)x + b) - \nabla g(x)^T z \\ \mu Z^{-1} e - g(x) \end{bmatrix}, \qquad (g(x), z, s) > 0. \\ &\Delta s = -s + \mu Z^{-1} e - Z^{-1} S \Delta z, \end{aligned} \tag{14}$$

We solve for $\Delta v = (\Delta x, \Delta z)$ and then repeat the process until an optimal solution x, is obtained as the perturbation parameter μ goes to zero. Note that unlike the non-symmetric and usually highly indefinite system (13), the reduced system (14) is symmetric with its size reduced considerably, and more likely to be better conditioned. For large-scale problems, the use of direct methods like TSVD to solve (14), or (7), can be computationally intense. Other approaches may be used, such as LSQR, or a Krylov space iterative method [e.g. implementing a variant of the CG algorithm (Saad 2000)]. However the performance of these methods is beyond the scope of this paper.

The PDIP algorithm is initialized with an (interior) velocity model x_0, that is, $[g(x_0), z_0, s_0] > 0$. We begin with an iterative process until the ℓ_2- norm of the right-hand side in (14) becomes smaller than certain tolerance value $\varepsilon > 0$; this is $[\, \nabla_x L(x_k, z_k, s_k) \;\; \mu Z^{-1} e - g(x_k) \,]^T \leq \varepsilon$. At each iteration, we compute the perturbation parameter $\mu = \frac{g(x)^T z}{2n} > 0$, and then solve for $\Delta v_k = (\Delta x_k, \Delta z_k)$ from the system (14) to set

Δs_k To enforce positivity, once we compute the Newton step from (14), we correct it with a factor $\hat{\alpha} \in (0, 1]$ to guarantee that the inequalities are strictly positive; we accomplished this task by using the following formula (El-Bakry *et al.* 1996; Nocedal & Wright 2006).

$$\widehat{\alpha} = 0.9995 \min \left\{ \frac{-1}{\min\left\{\frac{\Delta x}{g(x)}\right\}}, \frac{-1}{\min\left\{\frac{\Delta z}{z}\right\}}, \frac{-1}{\min\left\{\frac{\Delta s}{s}\right\}} \right\}.$$

Once we have computed the enforce positivity factor, $\hat{\alpha}$, in order to monitor progress to an optimal solution, we use a linesearch globalization strategy that requires a merit function M. This function has as a role to determine a step length $\alpha \in (0, \hat{\alpha}]$ that provides sufficient decrease of the objective function, while satisfying the inequality constraints. This is possible whenever the Newton direction results to be a descent direction for the merit function M at the current interior point $v_k = (x_k, z_k)$. We address this issue by implementing the Armijo condition as shown in Nocedal & Wright (2006), where the step length, α, can be chosen such that:

$$M(v_k + \alpha \Delta v_k) \leq M(v_k) + 10^{-4} \alpha \nabla M(v_k)^T \Delta v_k. \tag{15}$$

In the simplest case, we can select the objective function as the merit function; however we choose the modified augmented Lagrangian function introduced by Argaez & Tapia (2002), as the merit function due to its robustness:

$$M(x, z; \theta) = L(x, z, s) + \theta \varphi(x, z), \tag{16}$$

where θ is a non-negative penalty parameter, and the second term is the penalty function to the complementarity condition,

$$\varphi(x, z) = g(x)^T z - \mu \sum_{i=1}^{2n} \log [G(x) Ze]_i. \quad \{G(x) = \text{diag}[g(x)]\}. \tag{17}$$

It can be shown that the Newton direction is a descent direction for the penalty function; therefore, we can find θ such that the Newton direction is a descent direction for the merit function, this is

$$\nabla M(v_k)^T \Delta v_k < 0.$$

Finally, we compute our update $v_{k+1} = v_k + \alpha \Delta v_k$ to start again the process if we do not reach convergence or if no significant difference between iterates of the model parameter exists, in other words if $\| \frac{x_{k+1} - x_k}{x_{k+1}} \| < 10^{-5}$.

5 NUMERICAL EXPERIMENTATION

We briefly describe the two synthetic data sets used in the inversion: receiver functions and surface waves. In general, receiver functions can be used to resolve discontinuities (impedance contrasts) in seismic velocities, and provide good measurement of crustal thickness, without providing a good average of shear wave velocity. Surface (Love and Rayleigh) waves, provide a good average of absolute shear wave velocity, without good shear wave velocity contrasts in layered structures (e.g. Julia *et al.* 2000). Therefore, these two data sets can be considered as complementary and consistent, as long as we sample the same medium while assuming similar geological boundaries. Since both data sets are mainly sensitive to shear wave velocity structure (Julia *et al.* 2000), we can assume that the forward operator F depends non-linearly in our model parameter: the shear velocities $x_i, i = 1, \ldots, n$ that represent the different velocities of a half-space with n horizontal layers.

5.1 Receiver functions

Receiver functions give a good measurement of depth velocity discontinuities and crustal thickness. Different techniques can be used to compute receiver functions, for instance the spectral water level deconvolution technique (Langston 1979) and the time domain iterative deconvolution technique (Ligorria & Ammon 1999), which we implemented in this work. The resulting receiver function is a time series that can be viewed as a linear combination of delta functions, in which major negative (or positive) spike amplitudes correspond to a decrease (or increase) of seismic velocity, respectively. The time separation between *Ps* and *P* phases can be used to estimate crustal thickness, *h,* given the average crustal velocities (Zhu & Kanamori 2000),

$$h = \frac{t_{P_S}}{\sqrt{\frac{1}{v_s^2} - p^2} - \sqrt{\frac{1}{v_p^2} - p^2}},$$

where p represents the ray parameter of the incident wave, and (v_p, v_s) denotes the compressional and shear velocities, respectively.

This estimation has a trade-off between thickness and crustal velocities. However, since t_{Ps} represents the differential traveltime of S waves with respect to P waves in the crust, the dependence of h on v_p is not as strong as on our model parameter $x = v_s$ (or more precisely on the ratio v_p/v_s). Therefore, we consider receiver functions to be more sensitive to shear wave velocity contrasts. On the other hand, crustal thickness estimated only from the delay time of the Moho-Ps converted phase trade-off strongly with the v_p/v_s ratio. Since the average velocity cannot be resolved clearly, information from the surface waves can be used to better constrain the average shear velocities.

5.2 Surface wave dispersion

In general, surface waves dominate seismograms as the largest amplitude waves from an earthquake, and are observed at lower frequencies than the body waves. Furthermore, surface wave velocities vary depending on the depth range sampled by each period, making the surface wave dispersion valuable for studying Earth's structure. In general, dispersion curves are extracted from three component seismograms at a station for different frequencies and distances using reduction algorithms that rely on spectral analysis techniques (e.g. Maceira & Ammon 2009). Based on Rayleigh's principle, surface wave velocities are more sensitive to *S*-wave velocity, although they are also theoretically sensitive to *P*-wave velocity and density (Julia *et al.* 2000). This principle can be used to show that the phase velocity perturbation $\delta c/c$ can be viewed as a function of the sensitivity coefficients for *P*-wave velocity, *S*-wave velocity, and density. By investigating sensitivity function variation with respect to depth, the relative contribution of *P*-wave velocity, and density to dispersion can be shown to be smaller than the one for *S*-wave velocity (Julia *et al.* 2000). Therefore, surface wave dispersion is more also sensitive with respect to *S*-wave velocity.

5.3 Synthetic models

We implemented the joint inversion algorithm for four different Earth's crustal synthetic models. We divided each model in layers of 1 km thickness. For each synthetic velocity model, we compute two different dispersion curves, Rayleigh and Love, for either phase or group velocities. For each curve we sample 35 points or periods from 10 to 70 s. Simultaneously, we compute receiver functions for three different ray parameters *p*, with 2048 data points each. We create both data sets synthetically and noise free, by using standard algorithms (Herrmann 2002), based on the crustal models obtained from CRUST 2.0 (Laske *et al.* 2000). Fig. 1 shows a synthetic crustal (Rift) model used to create three receiver functions corresponding to three different ray parameters [$p = 0.03, 0.05,$ and $0.07\ (\mathrm{s\,km^{-1}})$], and the surface wave dispersion curves corresponding to group and phase dispersion curves, for both Rayleigh and Love waves.

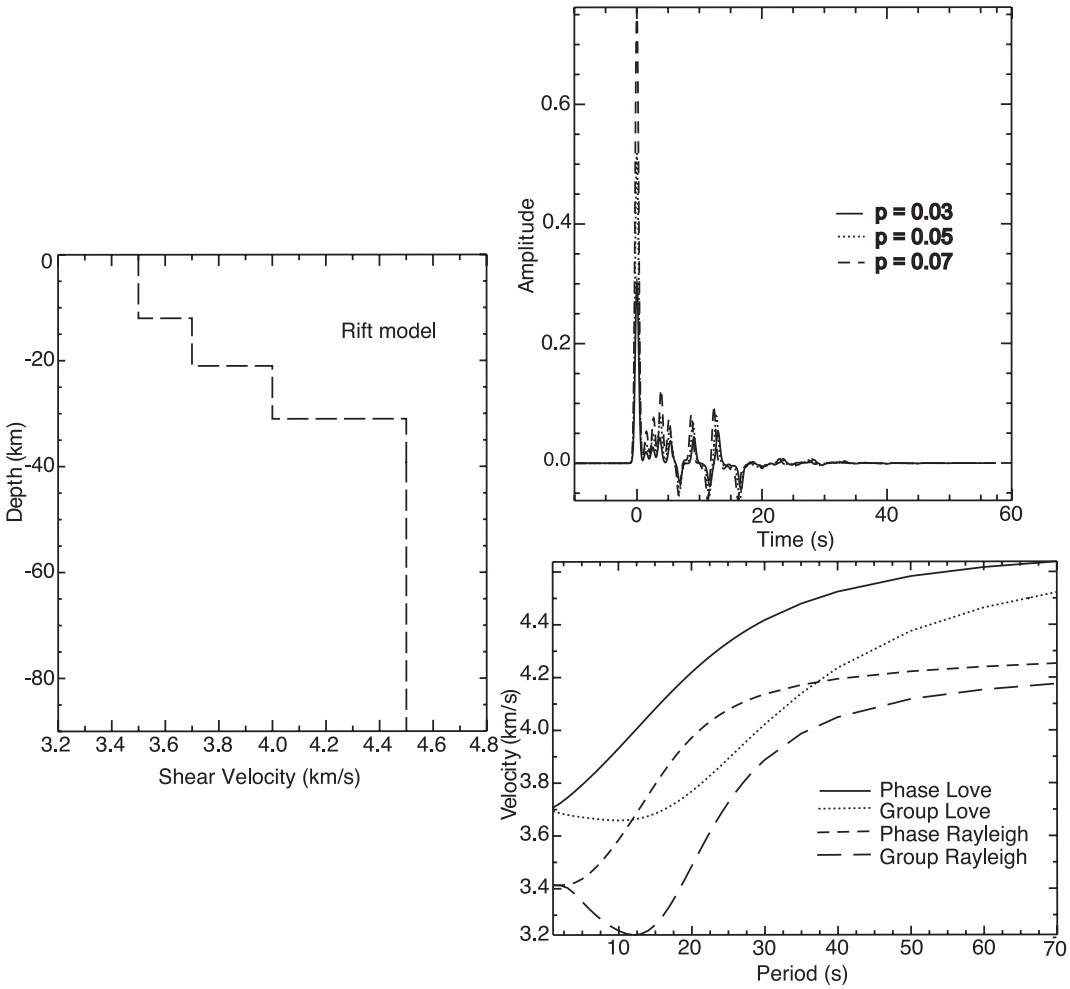

Figure 1. Left-hand panel: *S*-wave synthetic velocity model (Rift) (Laske *et al.* 2000). Right-hand top panel: Receiver functions for three different ray parameters *p*, right-hand bottom panel: surface wave dispersion curves (Love and Rayleigh) computed for the velocity model on the left-hand side. Both synthetic data sets were created noise free.

6 TEST RESULTS

We solved the inversion problems (6) and (9) by using the TSVD and PDIP method respectively, when inverting for the shear wave velocity $x = v_s$ (km s^{-1}) as the model parameter. The purpose of these tests is to validate the results obtained by the PDIP methods, when compared to those obtained by a conventional strategy used to solve the linearized version of the inverse problem (5). For both data sets, *P*-wave velocity can be inferred by assuming a constant value of the ratio v_p/v_s, as $v_p = (v_p/v_s)v_s = (v_p/v_s)x$. We estimate the density ρ(g /cm^3) from the resulting v_p velocity through some empirical relation, for example $\rho = 0.32v_p + 0.77$ introduced by Birch (1961). In this work, we do not attempt to quantify uncertainty. Thus, we assume some typical uncertainty values σ_i^2 for both data sets: 0.05 (km s^{-1}) for SW and 0.01 (s^{-1}) for RF observations.

To illustrate how receiver functions and surface wave dispersion velocities complement each other, we present the inversion results for the data sets created from the rift velocity model (Fig. 2). In each case, the two figures on the left represent a single inversion, this is, an inversion involving only one data set ($\eta = 0$ or $\eta = 1$). We performed these inversions to show the resolution that each data set can provide for a particular target velocity model, and to expose non-uniqueness of the inverted model. The joint inversion of both data sets (right) provides a better approximation to the target model as expected.

For each 1-D velocity model inverted, Figs 3–6 show the relative error (rms) with respect to the target model $\|\frac{x_k - x^*}{x^*}\|$, the residual error between the predictions and the observations $\|\frac{F(x_k)-y}{y}\|$, and the average shear velocity (the grey line represents average velocity of the model). We use the average shear velocity as a measure of how close the estimated models are to the true average velocity. This test can be used to check for spurious solutions. On the right of each figure, we show the target, initial, and estimated models given by each method: TSVD estimated models in dashed lines, and PDIP estimated models in solid lines. We show the best velocity model approximation for both methods with all models. When using the PDIP method, the upper and lower bounds over all the models x are defined as $v_{\min} = 3.2$ (km s^{-1}) and $v_{\max} = 4.8$ (km s^{-1}). Since the objective of the numerical experimentation is to have the smaller relative error (rms), our approximations should be closer to the target model. We set the maximum number of iterations to six for all the inversions and for both methods. As we show below, the number of iterations needed for the both method to converge sometimes results in less than six; however, we continue the inversion until the sixth iteration in the aim of comparison between the methods.

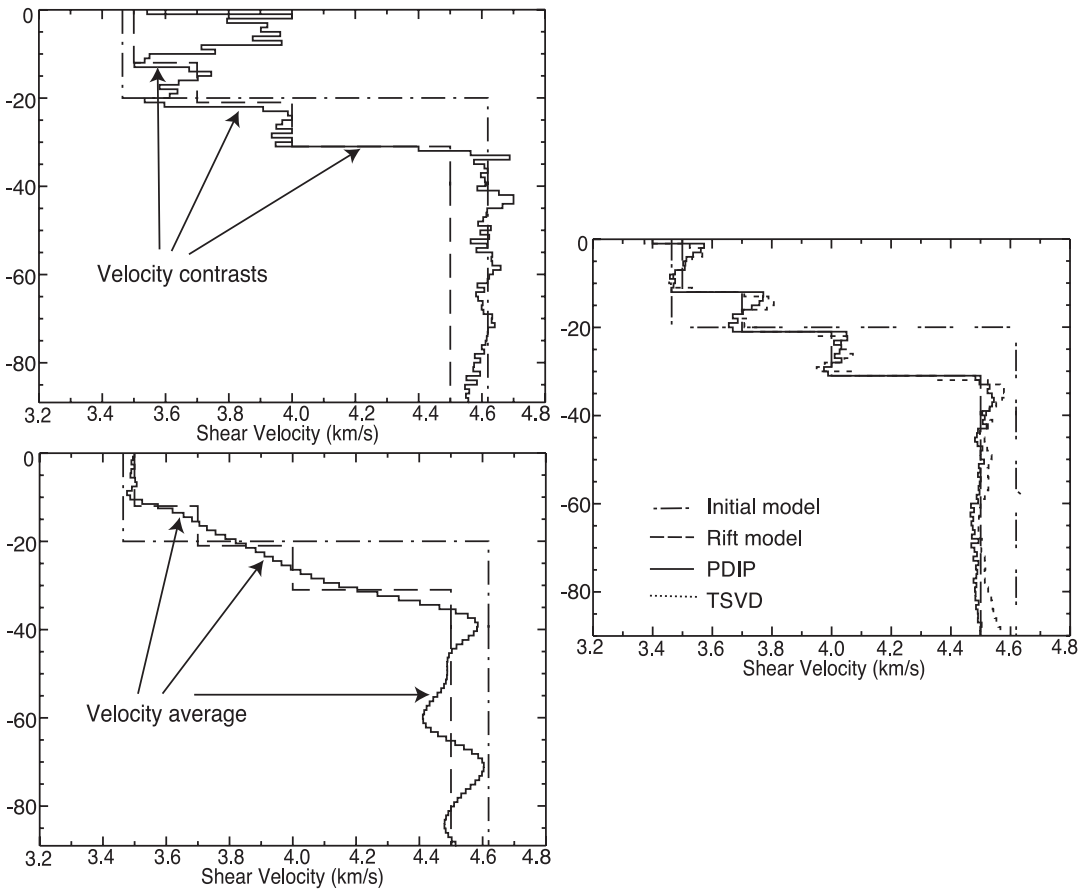

Figure 2. The single inversion of receiver functions (left-hand top panel) identifies velocity contrasts, while single inversion of surface waves (left-hand bottom panel) gives information on the average velocities at different depths. The joint inversion (right-hand panel) of these two data sets ($\eta = 0.5$) combines all this information and provides a substantial improvement in the final estimated model.

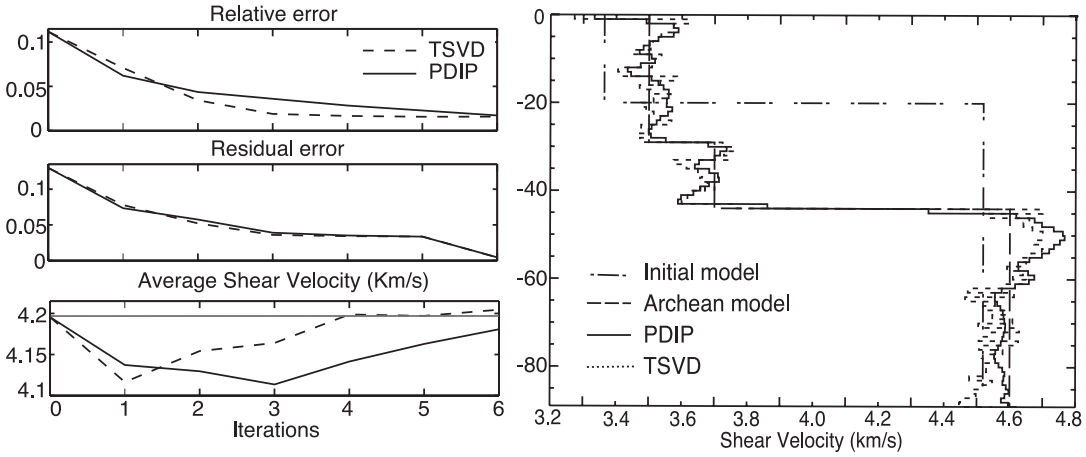

Figure 3. Archean model. Notice that the residual errors are about the same at each iteration, with faster convergence for TSVD method. This means that the rms minimal error obtained from the TSVD method is reached first than for the PDIP approximation.

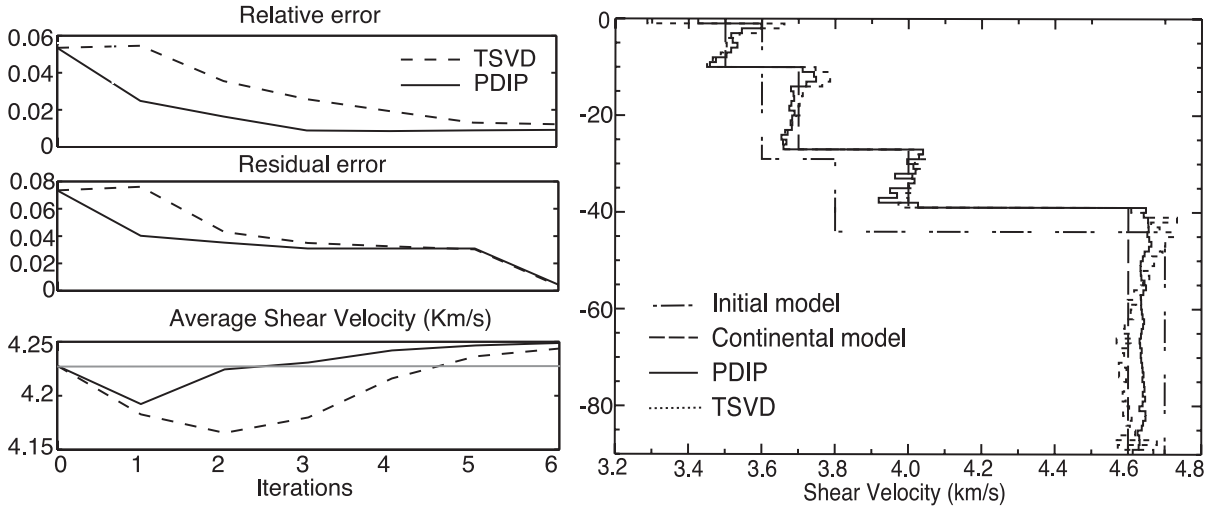

Figure 4. Continental model. The relative rms and residual errors are close at the last two iterations for both methods, but PDIP minimize better the rms and converges faster than TSVD. Both average velocities come closer to the true average velocity at the end of the inversion.

6.1 Synthetic velocity model tests

We begin with an initial guess of the velocity model, x_0, that usually is based on *a priori* geophysical knowledge. Since we know the synthetic target models, x^*, we have the freedom to select an appropriate initial model. We use the first synthetic model (Archean) as the initial guess for the next two models, Continental and Orogen, and a two half-space model for the Archean and the Rift models. Once we compute the observation vectors y^{SW} and y^{RF}, we iterate by evaluating the forward problem for the different approximations x_k, and solve the inverse problem either with the TSVD method for the unconstrained formulation (6) or the PDIP method for the constrained formulation (9). Once the selected method returns the updated model x_{k+1}, we check for convergence by requiring the residual error to be less than certain tolerance $\varepsilon > 0$, this is $\| \frac{F(x_k)-y}{y} \| \leq \varepsilon$. Finally we restart the iteration until it converges or reaches a maximum number of iterations. Six iterations are enough to provide an acceptable reduction in the rms value (as usual with most inversion methods that use an optimization framework).

We plot results for the different models in Figs 3–6. We observe that for all the models but the Archean (Fig. 3), the relative error and the number of joint inversion iterations required for convergence results smaller for PDIP than for TSVD. However, the accuracy obtained by the approximated Archean model with PDIP methods is still comparable with the one provided by TSVD at the final iteration. Therefore, our approach for solving the inverse problem with the PDIP method appears to be as good as the traditional unconstrained regularized inversion that uses TSVD. Our results for TSVD are obtained with a single selection of the smoothing parameter λ, therefore better results may be obtained when a search for the best parameters is made with a suite of inversions. To this end, we carry out further tests in Section 6.3 for a single value of the influence parameter η, to show the best choice of λ which is crucial when noise is present in the data.

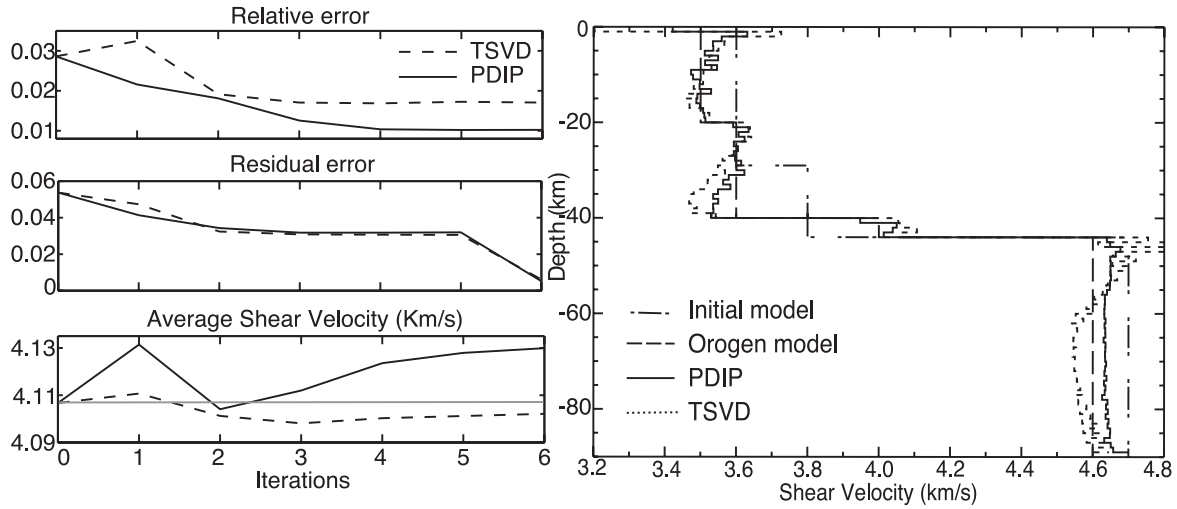

Figure 5. Orogen model. The relative error is smaller for PDIP at all stages of the algorithm, while the residual error is similar for both methods. The TSVD estimated velocity is closer to the true average velocity.

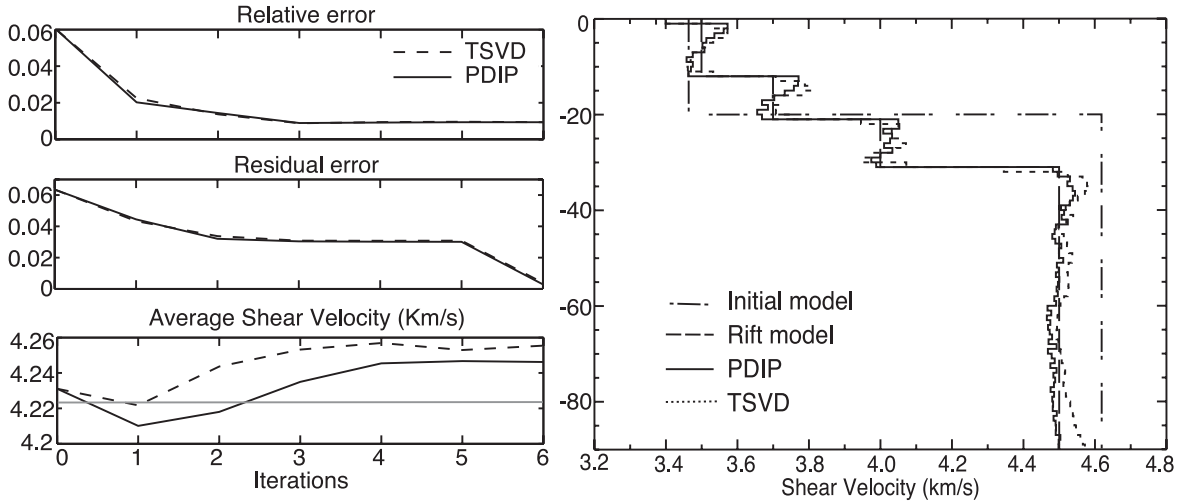

Figure 6. Rift model. The relative error and the residual error are within the same order of accuracy for both methods. The average is better resolved for PDIP until the third iteration, where the smaller relative rms error is reached.

6.2 Parameter selection tests

We performed a variety of experiments to test the best possible selection of parameters for both types of inversion methods. We use the Rift model previously presented as the test model for all of the experiments in this subsection. Different choices of the influence parameter η can help us to improve resolution in the final model. However, this process usually becomes expensive since it requires several suites of inversions, and an automatic procedure to obtain such parameter does not exist currently. Thus, we restrict the discussion to a combination of the two data sets when $\eta = 0.5$; in other words, we assume that both data sets are equally reliable for the inversion. In Table 1, we show how the relative and residual error changes as we select different values for η.

Table 1. Relative rms and residuals errors associated to different choices of the influence parameter η for the joint inversion. Our inverted Rift model was obtained by using PDIP as the inversion method. Note that the best choice for the influence parameter is $\eta = 0.75$.

η	rms	Residual error
0	3.2670E–02	2.3046E–01
0.25	7.9577E–03	1.1330E–02
0.5	7.4820E–03	3.0111E–03
0.75	6.6809E–03	1.6802E–03
1	1.6308E–02	7.6348E–04

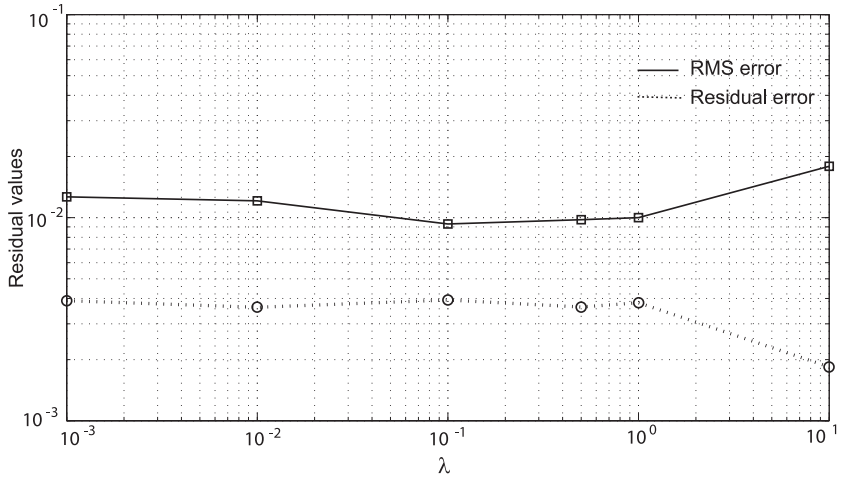

Figure 7. Logarithmic scale plot of the relative rms and residuals errors associated to different choices of the smoothing parameter λ. We applied the joint inversion algorithm over the Rift model when solving the inverse problem by using the TSVD method.

Table 2. Relative rms and residuals errors associated to different choices of the truncation parameter τ for the joint inversion over the Rift model when solving the inverse problem by using the TSVD method.

τ	rms	Residual error
10^{-12}	9.3036E–03	3.9254E–03
$10^{-5} \leq \tau \leq 10^{-2}$	≈9.3036E–03	≈3.9254E–03
5×10^{-2}	2.0526E–01	3.5542E–01
10^{-1}	–	–

6.2.1 TSVD parameters

We test different values of the regularization parameter λ. As λ moves away from zero, the regularization is too strong, thus introducing excessive smoothing to the final approximated model. The results appear to be similar to inverting the surface wave data alone (Fig. 2). The excessive smoothing may blur the ability of the receiver function to identify large velocity contrasts. When λ approaches zero, however, the algorithm becomes unstable; hence, we may not achieve convergence to a meaningful 1-D velocity model. In all the tests, we set the truncation parameter to be $\tau = 10^{-12}$, since for smaller values the inversion becomes unstable due to the high frequency components associated to the smaller eigenvalues (Hansen 1987). As we show in the next section this choice appears to be the best for our experiments. The best choice obtained from the suite of inversions for $\lambda \in \{0.001,0.01,0.1,0.5,1,10\}$ consisted on setting $\lambda = 0.1$, since for this value we obtain the smaller rms error, which is the inversion's goal, while having a good fit to the data. We present our results for relative and residual error in Fig. 7.

The value of τ is generally selected as a threshold to exclude those singular values close to machine precision. To test the effect of the truncation parameter over the inversion results, we use a similar strategy by defining $\lambda = 0.1$, our best value for λ, to be used for all the experiments and choosing τ from the set $\{10^{-12}, 10^{-5}, 10^{-3}, 10^{-2}, 10^{-1}\}$. The choice of $\tau = 10^{-12}$ works best and similar results were obtained up to $\tau = 10^{-2}$. However if $\tau > 10^{-2}$ we do not achieve convergence, which is probably a consequence of removing too many information provided by the eigenvectors corresponding to the greater singular values of the system matrix H in eq. (7). We summarize our numerical findings in Table 2.

If we implement smoothing with no truncation we cannot attain convergence. The high frequency components of the singular values introduce strong variations into the calculated receiver function waveform after the third iteration. As a result, the receiver function is distorted and the surface waves deviate considerably from the first approximations, no longer matching the true synthetics. With truncation applied but no smoothing, we still can achieve convergence, but with less accuracy to the true earth model (Fig. 8). Hence, we realize the relevance of the truncation parameter, τ, to achieve convergence compared to the smoothing parameter, λ, mainly intended to smooth out the final approximations. When neither truncation nor smoothing is applied, we do not have stability or convergence of the algorithm.

6.2.2 PDIP parameters

In Section 4.2, we showed how to compute the perturbation parameter μ, and the Lagrangian multipliers (z, s) associated to the PDIP method. However, we wish to test the impact that different choices of PDIP bounds $\nu_{\min}$ and $\nu_{\max}$ may have over the performance of the method. We use the following pair choices $(\nu_{\min},\nu_{\max}) \in \{(0,10), (3,5), (3.4,4.8)\}$. We realized that as the lower and upper bound approach the true maximum and minimum values of our test rift earth model, the performance of the method improves considerably, as presented in Table 3. The improvement is achieved by reducing the number of iterations to converge, and by diminishing the relative rms and residuals error.

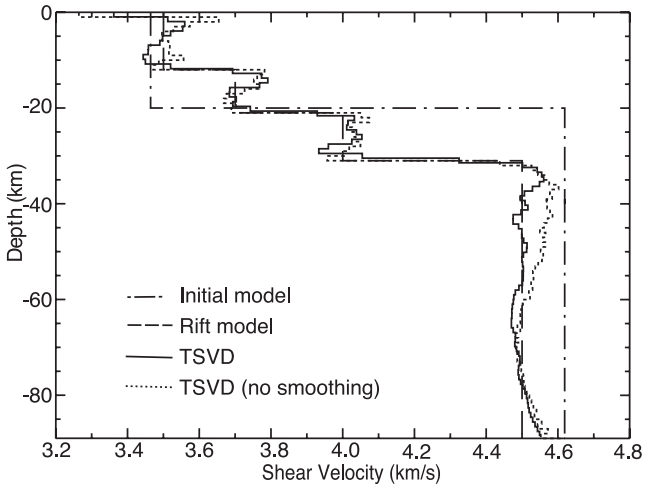

Figure 8. Numerical results obtained for the TSVD method when we use truncation without smoothing.

Table 3. Relative rms and residuals errors associated to different choices of the lower an upper velocity bounds over our model x. The joint inversion for the Rift model was solved by using the PDIP method. The last row shows the results when adding smoothing with $\lambda = 0.1$.

ν_{min}	ν_{max}	rms	Residual error
0.0	100.0	1.9327E–00	9.6538E–02
3.0	5.0	1.2789E–02	3.9357E–03
3.4	4.8	9.3143E–03	2.8394E–03
3.4	4.8	9.5301E–03	2.8411E–03

We tested also the effect of the regularization term as presented for the unconstrained case function. This implies an additional matrix added to the first block of the system matrix in eq. (14). The impact of this regularization over the inversion results does not represent an improvement with respect to the original results obtained for PDIP without that term (see last row in Table 3).

6.3 Noise synthetic tests

To add realistic noise to the data and check robustness of the methods, we follow the approach of Julia *et al.* (2000). The most common sources of contamination for surface wave dispersion are identified by Julia *et al.* (2000) as multipathing and modal superposition effects. Therefore, to contaminate the dispersion observations, the lower periods in the dispersion curves are shifted to higher velocities by adding 0.01 and 0.05 km s^{-1}, while higher periods are reduced by the same velocity amounts (Fig. 9, top panel). For receiver functions, Julia *et al.* (2000) note that the noise is usually introduced by scattering; thus, to contaminate our synthetic receiver functions, we add the transverse component of the receiver function corresponding to field data from USArray station 226A, showed in Fig. 13, to the radial component of our synthetic receiver function in Fig. 1. We present one of the resulting contaminated receiver functions together with the one that is noise free in Fig. 9(bottom panel).

Since an adjustment of the smoothing parameter λ is necessary to improve the numerical results, we run a suite of inversions to update the value of λ in the inversion according to the level of noise. We set the influence parameter to $\eta = 0.5$, since similar curves were obtained for other values of η, and the smoothing parameter λ is selected by comparing the rms values and the residual errors obtained for each value of λ. We keep the truncation threshold as $\tau = 10^{-12}$ in all cases. From this suite of inversions, we found $\lambda = 1$ to be the best choice for both cases where the dispersion curves are shifted by 0.01 km s^{-1} and then by 0.05 km s^{-1}. We present the trade-off in Fig. 10. For PDIP we do not use smoothing nor truncation; however, we changed the physical bounds to be $\nu_{min} = 3.4$ and $\nu_{max} = 4.8$, as a measure to counteract the noise effect over the data. We display the numerical results in Figs 11–12.

The interfaces are well characterized by both methods; PDIP has stronger reverberations likely due to the exaggerated peaks in the contaminated RF. However the fit to the dispersion curves in both methods is good enough to identify the correct depths of the velocity contrasts. The effect of both noise tests shows up considerably in the deeper layers. Both methods fail to identify properly the lower layers, especially when SW are shifted by 0.05 km s^{-1} (Fig. 12, right-hand panel). For TSVD the extra smoothing constraint has a similar effect over the added noise like the physical bounds on PDIP.

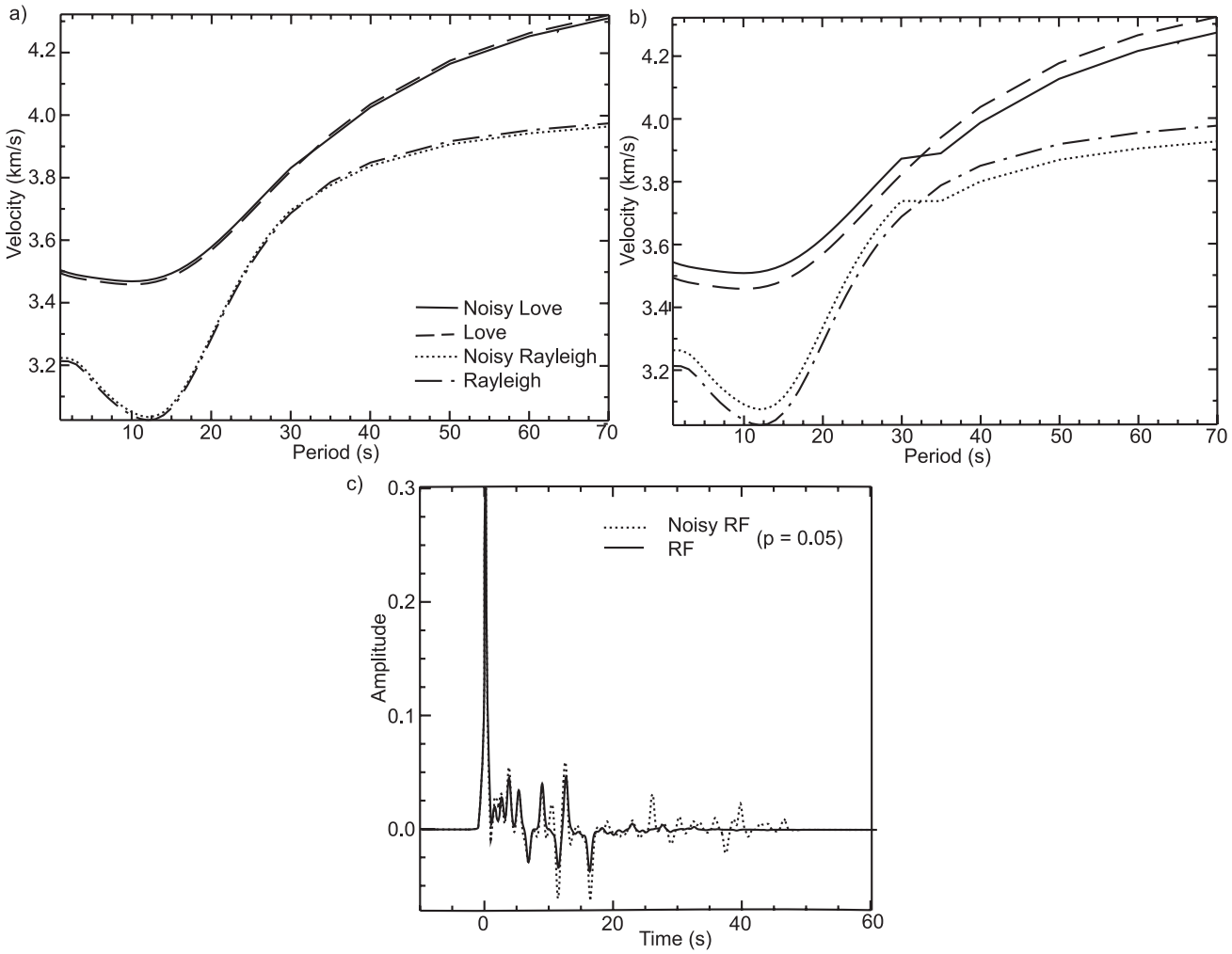

Figure 9. Synthetic data associated to the Rift model with realistic noise added. Dispersion curves with and without correlated noise obtained (a) by adding 0.01 km s^{-1} to the lower periods, and by reducing the higher periods in 0.01 km s^{-1}, (b) by using 0.05 km s^{-1} instead of 0.01 km s^{-1}. (c) Receiver function and contaminated receiver function corresponding to the ray parameter $p = 0.05$.

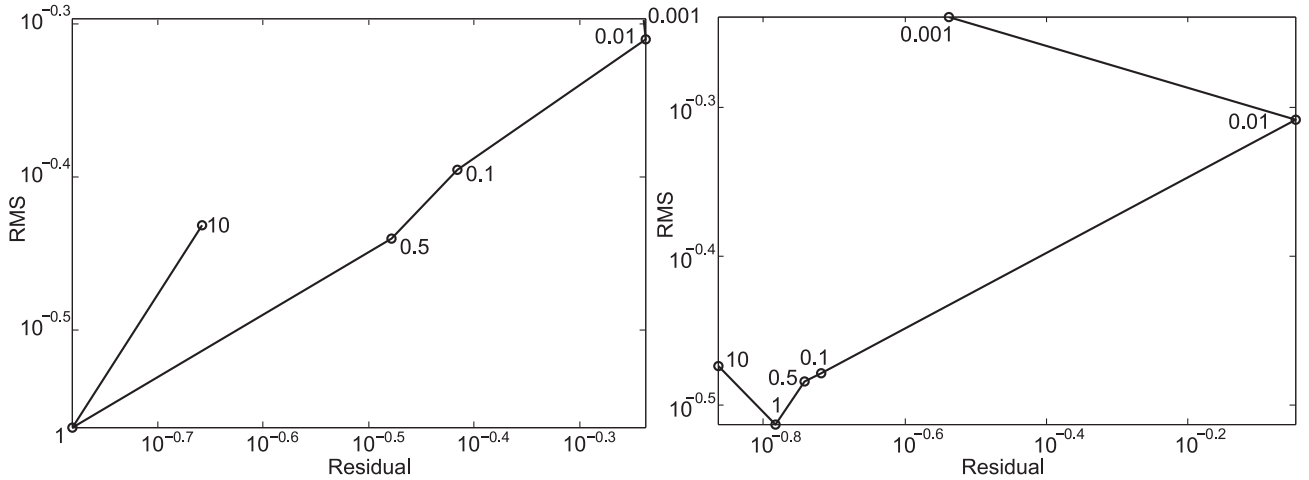

Figure 10. Logarithmic scale plot of the relative rms and residuals errors associated to different choices of the smoothing parameter λ, when dispersion curves are shifted by (a) 0.01 km s^{-1} and (b) by 0.05 km s^{-1}.

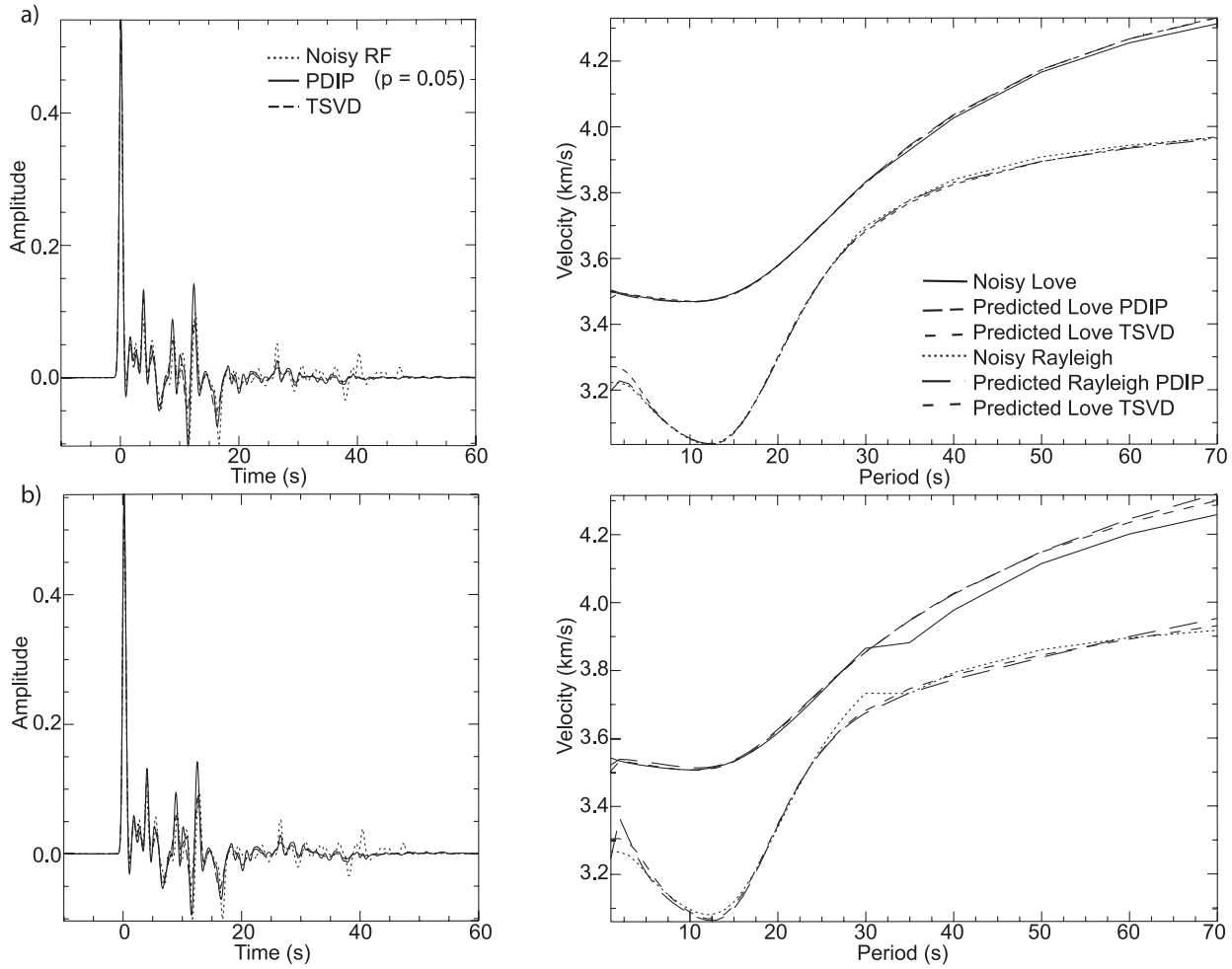

Figure 11. Inverted receiver functions and dispersion curves resulting from: (a) shift of surface waves by 0.01 km s^{-1}, and (b) by 0.05 km s^{-1}. The fit of both methods is similar for the RF synthetics and the dispersion curves. We note that the PDIP receiver function approximation fit better the exaggerated RF peaks and troughs between 10 and 20 s. The TSVD estimation of the dispersion curves is slightly better than for the PDIP ones, probably as a consequence of the extra smoothing constraint.

6.4 Real data application

To illustrate the capabilities of our 1-D joint inversion approach by using the PDIP method, we have used real data from one station of the EarthScope Transportable Array (USArray) and another one from the Colorado Plateau-Rio Grande Rift-Great Plains Seismic Transect (LA RISTRA) passive experiment, which are located in the North-American Rio Grande Rift. We use high quality group (Rayleigh and Love) surface wave dispersion observations [http://www.eas.slu.edu/eqc/eqc_research/NATOMO] and teleseismic *P*-wave receiver functions (Thompson *et al.* 2013). We show the results for only these two stations, since our objective is to present the quality of our results rather than to perform a complete velocity structure study for the region where the stations are located.

For both of the stations used as examples, the initial velocity model, x_0, corresponds to the AK-135 model of Kennett *et al.* (1995), starting at 10 km depth and distributed at a 2 km interval up to 70 km depth, then at a 5 km interval up to 250 km and finally at 10 km until 420 km. Each joint inversion includes at least three receiver function bins created according to an average ray parameter with a width of approximately 0.01 s km^{-1} between 0.04 and 0.07 s km^{-1}. The number of receiver functions employed to create these stacks is not less than 25 per ray parameter. Each receiver function consists on 820 data points for a time range from −5 to 80 s. Love and Rayleigh group velocities have from 50 to 75 dispersion measurements, with periods between 5 and 140 s. The upper and lower bounds over the velocity models were set to be $\nu_{\min} = 3.1$ (km s^{-1}) and $\nu_{\max} = 4.9$ (km s^{-1}). More details about the processing of the data sets can be found in Thompson *et al.* (2013).

We selected stations 226A and NM08 from the USArray and LA RISTRA experiment, respectively as a form of validation of our results, since its proximity (less than 25 km) will allow us to compare coherence among our predicted values. Figs 13–14 displays crustal and upper mantle 1-D velocity structure computed by using the constrained joint inversion algorithm for stations 226A and NM08. Also, we show the acceptable fit to receiver function observations, and to Love and Rayleigh wave group dispersion curves up to 80 (s). We note the similarities between the dispersion curves in both stations. This is a consequence of the short distance between the stations. However, the joint inversion shows the differences in the velocity models due to the receiver function data contribution. In both cases, the three major interfaces above

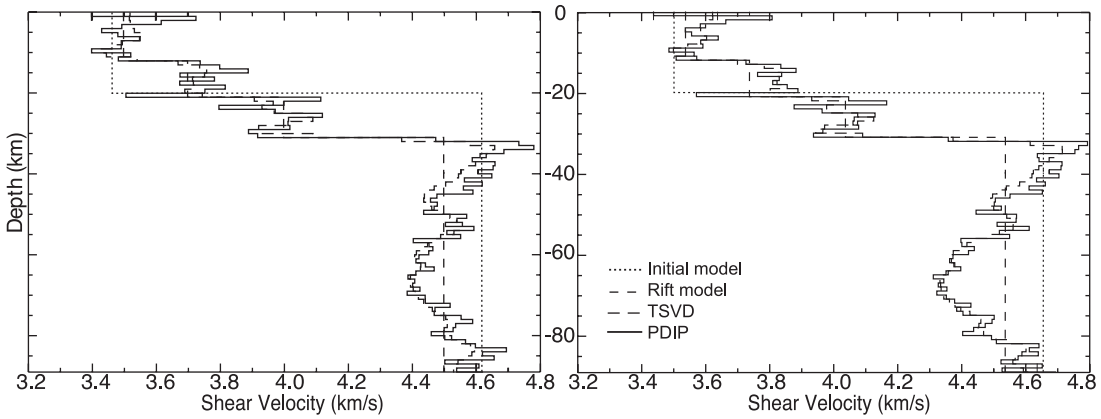

Figure 12. Rift velocity model approximation obtained from contaminated RF and SW dispersion curves. Both methods reach about the same level of convergence.

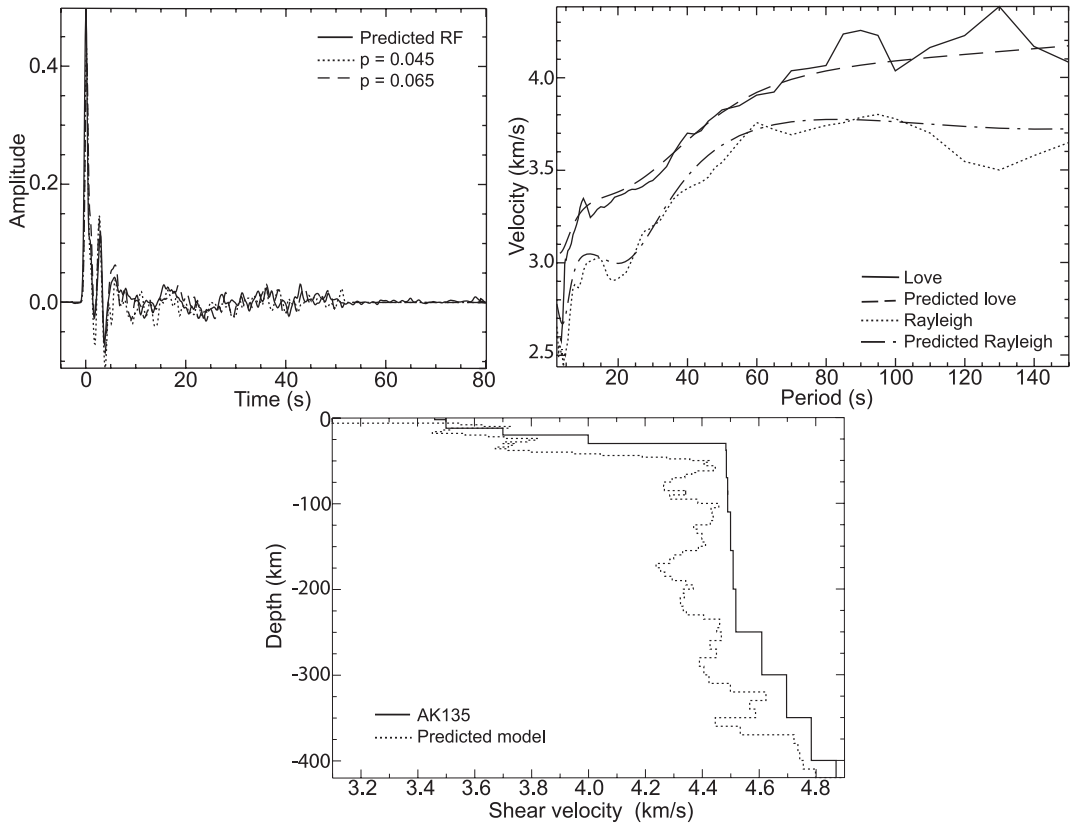

Figure 13. 1-D joint inversion results for station 226A. Bottom panel: the final *S*-wave velocity profile (dashed line) provided by our joint inversion approach, using ak135 (Kennett *et al.* 1995) (solid line) as the initial model. Top panels: fit to the RF observations (left) from our synthetic approximation (continuous line) for a ray parameter $p = 0.0446$ (small dashes) and $p = 0.0654$ (dashes), and to Love and Rayleigh group dispersion observations (right).

100 km are well determined, which show coherence in the numerical results due to the proximity between stations. The anomalously low velocity shown in Fig. 14 for station NM08 may be a consequence of the deeper peaks about 20 s in its corresponding RF. The velocities in station 226A are slightly slower than the standard model AK135 as expected near the end of the Rift, and the gradient is quite similar to the one of the standard model.

7 DISCUSSION

We expanded a joint inversion algorithm (Julia *et al.* 2000) to include the PDIP method for solving a constrained formulation of the inverse problem (5). We validate our constrained optimization approach with a conventional TSVD methodology (Julia *et al.* 2000; Maceira & Ammon 2009; Bailey *et al.* 2012). As expected, the geophysical information provided by joint inversion of receiver function and surface

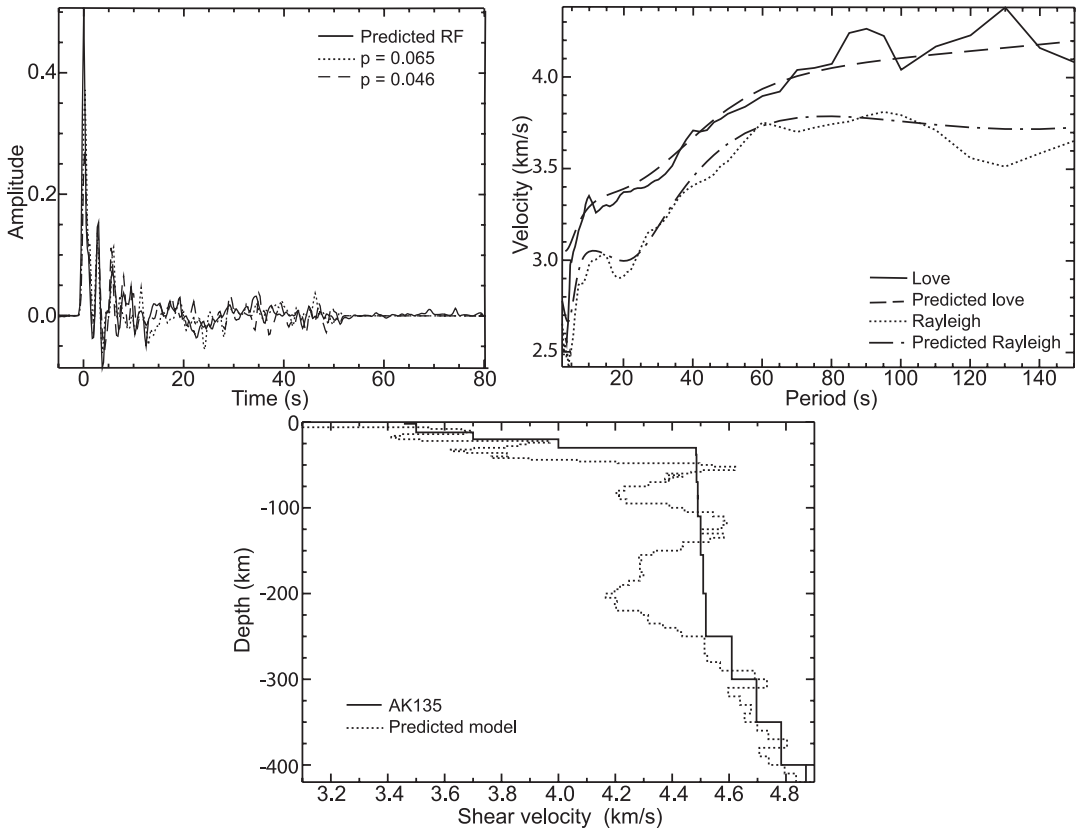

Figure 14. 1-D joint inversion results for station NM08. Bottom panel: the final *S*-wave velocity profile (dashed line) provided by our joint inversion approach, using ak135 (Kennett *et al.* 1995) (solid line) as the initial model. Top panels: fit to the RF observations (left) from our synthetic approximation (continuous line) for a ray parameter $p = 0.0446$ (dashes) and $p = 0.0654$ (small dashes), and to Love and Rayleigh group dispersion observations (right).

waves dispersion data complement each other. Therefore, we obtained an improvement in the final approximation to the target model, with respect to single inversions for both methodologies.

We showed that PDIP methods do not require the use of either a regularization parameter, λ, or a truncation parameter, τ, to obtain competitive results in terms of accuracy and robustness, that is, stability of the results in the presence of realistic noise. Also we were able to use real data from two nearby seismic stations located at the Rio Grande Rift (NA), and obtain numerical results that were in good agreement with the data and between the stations themselves. Instead of using regularization parameters for smoothing constraints, we carry the regularization by including *a priori* information in the form of physical bounds. The addition of these bounds over the model parameter, x, helps us to reduce the model space thus avoiding spurious solutions. We believe that this is an advantage, and a major difference with respect to conventional unconstrained formulations (Haber & Oldenburg 1997; Julia *et al.* 2000, 2005; Colombo & De Stefano 2007; Maceira & Ammon 2009; Moorkamp *et al.* 2011) in which the constraints are all included into the objective function, and also from the constrained formulations of Gallardo & Meju (2004), Backus & Gilbert (1967), and Lines *et al.* (1988), which rely on a quadratic programming approach with expensive simplex-type schemes, or Lagrange multipliers methods without any globalization strategy (Menke 1984). Our rationale proceeds from a possible general formulation in the unconstrained case,

$$\min_x \left[f(x) + \sum_i \lambda_i R_i(x) \right]$$

in which f is usually a misfit function, for example, $f(x) = \frac{1}{2} \|F'(x_k)x + b\|^2$, and relies in the capability of the terms $\lambda_i R_i$ to introduce *a priori* information to regularize the objective function. In our approach, we attempt to use an explicit formulation of the constraints as in (9), where we use the more convenient physical bound constraints instead of the traditional smoothing constraints that require the computation of the regularization parameters λ_i.

We acknowledge that for TSVD, the computational complexity required to solve the inverse problem is larger than the one required by other methods previously mentioned. In some cases, these methods may be more efficient than PDIP for the size of the problems studied in this work. However, for large-scale problems, arising when we increase the size of the data sets or the number of models, grid search type algorithms become computationally intensive (Simmons *et al.* 2011). PDIP methods, on the other hand, are known for their capabilities to solve large-scale problems (Forsgren *et al.* 2002; Nocedal & Wright 2006; Burstedde & Ghattas 2009). The additional work on PDIP comes in the exploration for appropriate velocity bounds, which might need to be changed with respect to depth, and a new set of initial values for the Lagrangian multipliers that may need to be computed.

8 CONCLUSIONS

We present a constrained optimization framework for joint inversion of two synthetic geophysical data sets solved by using PDIP methods. We validate the numerical results obtained with this approach against those obtained with a conventional TSVD inversion method. We show that the addition of explicit inequality constraints, which are more physical than the typical smoothing constraint terms in the objective function, results effective to handle the introduction of *a priori* information, and to regularize the inversion. This feature helps us to reduce ambiguities and the influence of noise over the inversion results, without computing regularization parameters for smoothing constraints. Our work explores an approach well known by the optimization community but not generally implemented by the geophysicists to solve this type of inverse problems. We believe that our approach stands as a good alternative when looking for a robust inversion method, being less affected by noise in the data, while it also represents savings in computational costs, since there is no need for searching the best regularization parameters, as required for conventional optimization approaches.

ACKNOWLEDGEMENTS

This research project is based on work supported by the National Science Foundation under Grants No. HRD 0734825 and CNS 0923442, and the Program in Computational Science. We thank the Department of Mathematical Sciences, and the first author thanks the support of the Universidad ICESI (Cali, Colombia).

REFERENCES

Ammon, C.J., Randall, G.E. & Zandt, G., 1990. On the nonuniqueness of receiver function inversion, *J. geophys. Res.,* **95,** 15 303–15 318.

Argaez, M. & Tapia, R., 2002. On the global convergence of a modified augmented Lagrangian linesearch interior-point Newton method for nonlinear programming, *J. Optim. Theory,* **114,** 255–272.

Backus, G.E. & Gilbert, J.F., 1967. Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. astr. Soc.,* **13,** 247–276.

Bailey, I.W., Miller, M.S., Liu, K.J. & Levander, A., 2012. V-S and density structure beneath the Colorado Plateau constrained by gravity anomalies and joint inversions of receiver function and phase velocity data, *J. geophys. Res.,* **117,** B02313, doi:10.1029/2011JB008522.

Birch, F., 1961. Composition of the Earth's mantle, *Geophys. J. R. astron. Soc.,* **4,** 295–311.

Bannister, S., Yu, J., Leitner, B. & Kennett, B.L.N., 2003. Variations in crustal structure across the transition from West to East Antarctica, Southern Victoria Land, *Geophys. J. Int.,* **155**(3), 870–880.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117**(B2), doi:10.1029/2011JB008560.

Burstedde, C. & Ghattas, O., 2009. Algorithmic strategies for full waveform inversion: 1D experiments, *Geophysics,* **74**(6), WCC37–WCC46.

Chang, S., Baag, C. & Langston, C., 2004. Joint analysis of teleseismic receiver functions and surface wave dispersion using the genetic algorithm, *Bull. seism. Soc. Am.,* **94**(2), 691–704.

Colombo, D. & De Stefano, M., 2007. Geophysical modeling via simultaneous joint inversion of seismic, gravity, and electromagnetic data: application to prestack depth imaging, *Leading Edge,* **26,** 326–331.

Constable, S.C., Parker, R.L. & Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics,* **52,** 289–300.

Dantzig, G.B., 1963. *Linear Programming and Extensions,* Princeton University Press.

El-Bakry, A.S., Tapia, R., Tsuchiya, A. & Zhang, Y., 1996. On the formulation and theory of the primal-dual newton interior-point method for nonlinear programming, *J. Optim. Theory,* **89,** 507–541.

Forsgren, A., Gill, P.E. & Wright, M.H., 2002. Interior point methods for nonlinear optimization, *SIAM Rev.,* **44**(4), 525–597.

Gallardo, L.A. & Meju, M.A., 2004. Joint two-dimensional DC resistivity and seismic traveltime inversion with cross-gradients constraints, *J. geophys. Res.,* **109,** B03311, doi:10.1029/2003JB002716.

Gallardo, L.A., Meju, M.A. & Pérez-Flores, M.A., 2004. A quadratic programming approach for joint image reconstruction: mathematical and geophysical examples, *Inverse Problems,* **21,** 435–452.

Gallardo-Delgado, L.A., Pérez-Flores, M.A. & Gómez-Treviño, E., 2003. A versatile algorithm for joint 3-D inversion of gravity and magnetic data, *Geophysics,* **68,** 949–959.

Haber, O. & Oldenburg, D., 1997. Joint inversion: a structural approach, *Inverse Problems,* **13,** 63–77.

Hansen, P.C., 1987. The truncated SVD as a method for regularization, *BIT,* **27,** 534–553.

Hansen, P.C., 2010. *Discrete Inverse Problems,* SIAM.

Herrmann, R.B., 2002. Computer programs in seismology, Available at: http://www.eas.slu.edu/eqc/eqccps.html (last accessed 6 May 2012).

Julia, J., Ammon, C.J., Herrmann, R. & Correig, M., 2000. Joint inversion of receiver function and surface wave dispersion observations, *Geophys. J. Int.,* **142,** 99–112.

Julia, J., Ammon, C.J. & Nyblade, A.A., 2005. Evidence for mafic lower crust in Tanzania, East Africa, from joint inversion of receiver function and Rayleigh wave dispersion velocities, *Geophys. J. Int.,* **162,** 555–569.

Karmarkar, N., 1984. A new polynomial algorithm for linear programming, *Combinatorics,* **4,** 373–395.

Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.,* **122**(1), 108–124.

Kozlovskaya, E., Vecsey, L., Plomerova, J. & Raita, T., 2007. Joint inversion of multiple data types with the use of multiobjective optimization problem formulation and application to the seismic anisotropy investigation, *Geophys. J. Int.,* **171,** 761–779.

Langston, C.A., 1979. Structure under Mount Rainier, Washington, inferred from teleseismic body waves, *J. geophys. Res.,* **84,** 4749–4762.

Laske, G., Masters, G. & Reif, C., 2000. Crust 2.0: the current limits of resolution for surface wave tomography in North America, *EOS, Trans Am. geophys. Un.,* **81,** F897, Available at: http://igpppublic.ucsd.edu/∼gabi/crust2.html (last accessed 12 September 2013).

Lawrence, J. & Wiens, D., 2004. Combined receiver-function and surface wave phase-velocity inversion using a niching genetic algorithm: application to Patagonia, *Bull. seism. Soc. Am.,* **94**(3), 977–987.

Lees, J.M. & Vandecar, J.C., 1991. Seismic tomography constrained by bouguer gravity anomalies: applications in western Washington, *PAGEOPH,* **135,** 31–52.

Li, Y. & Oldenburg, D.W., 2003. Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method, *Geophys. J. Int.,* **152,** 251–265.

Ligorria, J.P. & Ammon, C.J., 1999. Iterative deconvolution and receiver-function estimation, *Bull. seism. Soc. Am.,* **89,** 1395–1400.

Lines, L.R., Schultz, A.K. & Treitel, S., 1988. Cooperative inversion of geophysical data, *Geophysics,* **53,** 8–20.

Lucente, F., Piana Agostinetti, N., Moro, M., Selvaggi, G. & Di Bona, M., 2005. Possible fault plane in a seismic gap area of the southern Apennines (Italy) revealed by receiver function analysis, *J. geophys. Res.,* **110,** B04307, doi:10.1029/2004JB003187.

Maceira, M. & Ammon, C.J., 2009. Joint inversion of surface wave velocity and gravity observations and its application to central Asian basins s-velocity structure, *J. geophys. Res.,* **114,** B02314, doi:10.1029/2007JB0005157.

Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters, *SIAM J. Appl. Math.,* **11,** 431–441.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory*, (textbook), Academic Press.

Moorkamp, M., Jones, A.G. & Fishwick, S., 2010. Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data, *J. geophys. Res.,* **115,** B04318, doi:10.1029/2009JB0006369.

Moorkamp, M., Heincke, B., Jegen, M., Roberts, A.W. & Hobbs, R.W., 2011. A framework for 3-D joint inversion of MT, gravity and seismic refraction data, *Geophys. J. Int.,* **184,** 477–493.

Musil, M., Maurer, H.R. & Green, A.G., 2003. Discrete tomography and joint inversion for loosely connected or unconnected physical properties: application to crosshole seismic and georadar data sets, *Geophys. J. Int.,* **153,** 389–402.

Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization,* 2nd edn, Springer Verlag.

Piana Agostinetti, N. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181**(2), 858–872.

Saad, Y., 2000. *Iterative Methods for Sparse Linear Systems*, SIAM.

Sambridge, M., 1999a. Geophysical inversion with a neighbourhood algorithm I. Searching a parameter space, *Geophys. J. Int.,* **138**(2), 479–494.

Sambridge, M., 1999b. Geophysical inversion with a neighbourhood algorithm II. Appraising the ensemble, *Geophys. J. Int.,* **138**(3), 727–746.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 1009, doi:10.1029/2000RG000089.

Shibutani, T., Sambridge, M. & Kennett, B., 1996. Genetic algorithm inversion for receiver functions with application to the crust and uppermost mantle structure beneath eastern Australia, *Geophys. Res. Lett.,* **23,** 1829–1832.

Simmons, N., Myers, S. & Johannesson, G., 2011. Global-scale P wave tomography optimized for prediction of teleseismic and regional travel times for Middle East events: 2. Tomographic inversion, *J. geophys. Res.,* **116,** B04305, doi:10.1029/2010JB007969.

Thompson, L., Velasco, A.A. & Hussein, M., 2013. Geophysical constraints on the crustal structure of the Southern Rio Grande Rift, *Geophys. Res. Lett.,* in press.

Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions of Ill-Posed Problems,* W.H. Whinston & Sons.

Vinnik, L., Reigber, C., Aleshin, I., Kosarev, G., Kaban, M., Oreshin, S. & Roecker, S., 2004. Receiver function tomography of the central Tien Shan, *Earth planet. Sci. Lett.,* **225**(1–2), 131–146.

Vinnik, L., Aleshin, I., Kaban, M., Kiselev, S., Kosarev, G., Oreshin, S. & Reigber, C., 2006. Crust and mantle of the Tien Shan from data of the receiver function tomography, *Izv. Phys. Solid Earth,* **42**(8), 639–651.

Vogel, C.R., 2002. *Computational Methods for Inverse Problems*, SIAM.

Wiggins, R.A., 1972. The general linear inverse problem: implication of surface waves and free oscillations for Earth structure, *Rev. Geophys.,* **10,** 251–285.

Wilson, D. & Aster, R., 2005. Seismic imaging of the crust and upper mantle using regularized joint receiver functions, frequency–wave number filtering, and multimode Kirchhoff migration, *J. geophys. Res.,* **110,** B05305, doi:10.1029/2004JB00343.

Zhao, L., Sen, M., Stoffa, P. & Frohlich, C., 1996. Application of very fast simulated annealing to the determination of the crustal structure beneath Tibet, *Geophys. J. Int.,* **125**(2), 355–370.

Zhdanov, M.S., Gribenko, A. & Wilson, G., 2012. Generalized joint inversion of multimodal geophysical data using Gramian constraints, *Geophys. Res. Lett.,* **39,** L09301, doi:10.1029/2012GL051233.

Zhu, L. & Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions, *J. geophys. Res.,* **105,** 2969–2980.