



Research article

A Multi-Objective Optimization Framework for Joint Inversion

Lennox Thompson^{1*} Aaron A. Velasco¹ and Vladik Kreinovich²

¹ Department of Geological Sciences, University of Texas at El Paso (UTEP), 500 W. University Avenue, El Paso, TX 79968, USA

² Department of Computer Science, University of Texas at El Paso (UTEP), 500 W. University Avenue, El Paso, TX 79968, USA

* **Correspondence:** lethompson@miners.utep.edu;

Abstract: Different geophysical data sets such as receiver functions, surface wave dispersion measurements, and first arrival travel times, provide complementary information about the Earth structure. To utilize all this information, it is desirable to perform a *joint inversion*, i.e., to use all these datasets when determining the Earth structure. In the ideal case, when we know the variance of each measurement, we can use the usual Least Squares approach to solve the joint inversion problem. In practice, we only have an approximate knowledge of these variances. As a result, if a geophysical feature appears in a solution corresponding to these approximate values of variances, there is no guarantee that this feature will still be visible if we use the actual (somewhat different) variances.

To make the joint inversion process more robust, it is therefore desirable to repeatedly solve the joint inversion problem with different possible combinations of variances. From the mathematical viewpoint, such solutions form a Pareto front of the corresponding multi-objective optimization problem.

Keywords: Teleseismic, Receiver Functions, Primal-Dual Interior Point Method

1. Introduction

In this paper, we describe how to combine multiple geophysical datasets for the purpose of assisting in better determining physical properties of the Earth structure. The need for combining different datasets comes from the fact that different datasets provide complementary information about the Earth structure. By jointly inverting multiple geophysical datasets, we combine these complementary pieces of information and thus, we get a more accurate description of the Earth structure; see, e.g., (Vozoff and Jupp 1975, Julia et al. 2000, Shen et al. 2003, Colombo and De Stefano 2007, Maceira and Ammon 2009, Shearer 2009, Stein and Wysession 2003).

Specifically, we combine receiver functions, surface wave dispersion measurements, and *P*-wave

travel times. The need to use different datasets comes from the fact that there are two types of seismic waves that travel through the Earth: the body waves and the surface waves. Both types of waves provide different sensitivities and information about the Earth structure, since they are sampling, correspondingly, the interior and surface of the Earth. The information collected from the body waves travels deeper into the Earth and translates into teleseismic P -wave receiver functions. In order to obtain information about the Earth surface, surface waves are analyzed, in our case, by means of surface waves dispersion. Receiver functions resolve discontinuities (impedance contrasts) in seismic velocities, and provide good measurement of crustal thickness, without providing a good average of shear wave velocity. On the other hand, we have surface (Love and Rayleigh) waves whose energy is concentrated near the Earth's surface, and provide good average of absolute shear wave velocity, without a good shear-wave velocity contrasts in layered structures (Julia et al. 2000, Maceira and Ammon 2009, Shearer 2009, Stein and Wysession 2003, Cho et al. 2007, Obrebski et al. 2010). Therefore these two data sets provide complementary information about the Earth structure. Seismic first-arrival travel times are complementary to the other datasets because the travel time enable us to recover the causative slowness of the Earth structure (Lees and Vandecar 1991).

For each dataset, we usually know the relative variance (uncertainty of data) of different measurement results from this dataset, and thus, we can use the Least Squares method to find the corresponding Earth model. For multiple datasets, we can sometimes still use the Least Squares Method to process all these datasets – provided that we know the variances of different measurements from different datasets. In practice, however, we usually only have an approximate knowledge of these variances. So, instead of producing a single model, several models corresponding to different possible variances are generated. If all these models – corresponding to different possible values of variances – contain a certain geophysical feature, then we can be certain that this feature is also present in the actual Earth model (which corresponds to the actual (unknown) values of the variances).

From the mathematical viewpoint, the task of computing all these models is equivalent to computing the Pareto front of the corresponding Multi-Objective Optimization problem (Kozlovskaya 2000), where different objective functions correspond to different datasets.

In addition to producing all these models, it is also desirable to produce a "typical" model, so that we only look for features which are present on this typical model. In this paper, we use methods for selecting such a typical model as described in (Sambridge 1999a, Sambridge 1999b, Kozlovskaya 2000).

This paper has the following structure. In Section 2, we describe, in detail, the need for multi-objective optimization. In Section 3, we show how to solve the corresponding optimization problems. In Section 4, we briefly describe the corresponding geophysical datasets. The results of applying multi-objective optimization technique to these datasets are shown and discussed in Section 5. Finally, Section 6 contains conclusions.

2. Need for Multi-Objective Optimization

2.1. Inverse Problems: Brief Reminder

In many real-life situations, we are interested in the values of the quantities x_1, x_2, \dots , which are difficult (or even impossible) to measure directly. For example, in geophysics, one of our main objectives is to find the shear velocities x_i at different 3-D points i (i.e., at different depths and at

different geographic locations). To find these values, we measure easier-to-measure quantities y_1, y_2 , etc., which are related to $x = (x_1, \dots, x_n)$ by a known relation $y = F(x)$, and then use the measured quantities $y = (y_1, \dots, y_m)$ to find the desired values x .

In particular, in geophysics, to find shear velocities x_i , we can calculate the teleseismic receiver functions y^{RF} , surface wave dispersion velocities y^{SW} , travel times y^{TT} , etc. For each of these types of data, if we know the velocity model x , then we can predict the Earth's response by using the corresponding (known) operator F : $y^{RF} = F^{RF}(x)$, $y^{SW} = F^{SW}(x)$, $y^{TT} = F^{TT}(x)$, etc.

Measurements are never absolutely accurate, there is always some measurement inaccuracy, there is always some level of noise preventing us from measuring the corresponding quantities exactly. A usual way to estimate parameters in the presence of noise is to use the Least Squares method, i.e., to find the values x that minimize the expression

$$\sum_{i=1}^m \frac{(F_i(x) - y_i)^2}{\sigma_i^2}, \quad (1)$$

where σ_i is the standard deviation of the noise (measurement inaccuracy) of the i -th measurement.

In some cases, all the available data points come from measurements of the same type, obtained by using the same methodology and similar instrumentation. For example, we may only have travel times, or only surface wave dispersion velocities. In such cases, it is reasonable to assume that all these measured values have the same accuracy, i.e., that $\sigma_i = \text{const}$. Under this assumption, minimizing the expression (1) is equivalent to minimizing the sum of the squares

$$\|F(x) - y\|^2 \stackrel{\text{def}}{=} \sum_{i=1}^m (F_i(x) - y_i)^2. \quad (2)$$

2.2. Need to Take Constraints into Account

Sometimes, the models x obtained by an appropriate minimization are not physically meaningful. For example, in geophysics, some models x predict higher velocities in the crust and lower velocities in the mantle, contrary to known geophysical models. In other cases, when we expect a smooth dependence on a signal on time, the reconstructed signal x can exhibit abrupt non-physical changes.

To avoid such non-physical solutions, it is desirable to explicitly take into account the corresponding constraints. For example, in most practical problems, there are known physical bounds on the values of the quantities x_i . In particular, in geophysics, for each depth, we can approximate the lower bound and the upper bound on possible values of shear velocity x_i at this depth.

In precise terms, for every i , we know the bounds a_i and b_i such that $a_i \leq x_i \leq b_i$.

Restrictions on smoothness can be described as known bounds Δ_{ij} on the difference between the values x_i and x_j for nearby points of time or at nearby spatial locations: $-\Delta_{ij} \leq x_i - x_j \leq \Delta_{ij}$.

Under the corresponding constraints, the optimization problem (2) takes the form

$$\begin{aligned} \min_x \|F(x) - y\|^2 \\ \text{s.t. } g(x) \geq 0. \end{aligned} \quad (3)$$

where $g(x)$ is a vector consisting of the corresponding constraints. For example, to describe bounds a_i and b_i on the values x_i , we use constraints $x_i - a_i \geq 0$ and $b_i - x_i \geq 0$.

Constraints corresponding to smoothness can also be expressed in the form $g(x) \geq 0$, with the corresponding components of the vector $g(x)$ having the form $\Delta_{ij} - (x_i - x_j)$ and $(x_i - x_j) - (-\Delta_{ij})$.

Comment. Traditionally, researchers avoid non-physical non-smooth velocity models by adding a regularization term $\lambda \|Lx\|^2$ to the minimized function; see, e.g., (Tikhonov and Arsenin 1977). The problem with this term is that it is not clear how to select λ , and different values of λ lead to different solutions; see, e.g., (Hansen 2010, Vogel 2002). Because of this problem, in this paper, instead of using regularization, we explicitly formulate constraints that need to be satisfied. For example, the desired smoothness is described as a bound on the differences $x_i - x_j$.

2.3. Joint Inversion: Idealized Case

As noted earlier, measurements of different type usually provide complementary information and it is, therefore, beneficial to use measurement results of all the types.

When we use measurements of different types t, t' , etc., then while it is reasonable to assume that all the measurements i of the same type t have the same standard deviation $\sigma_i = \sigma^t$, standard deviations of measurements of different types are, in general, different: $\sigma^t \neq \sigma^{t'}$. Let us first consider the idealized case, when we assume that we know the accuracy σ^t of measurements of type t .

In this case, we can still use the Least Squares expression (1) to find the desired model x . By grouping together measurements of different type, we can rewrite the expression (1) in the following form:

$$\sum_{i=1}^m \frac{(F_i(x) - y_i)^2}{\sigma_i^2} = \sum_{t=1}^T \sum_{i \in t} \frac{(F_i(x) - y_i)^2}{(\sigma^t)^2} = \sum_{t=1}^T \frac{1}{(\sigma^t)^2} \cdot \left(\sum_{i \in t} (F_i(x) - y_i)^2 \right) \quad (4)$$

where T is the total number of different types of measurements, and the notation $i \in t$ indicates that the i -th measurement is of type t .

We can rewrite this expression as

$$\sum_{t=1}^T c_t^2 \cdot \|F^t(x) - y^t\|^2, \quad (5)$$

where $c_t \stackrel{\text{def}}{=} \frac{1}{\sigma^t}$, y^t is the list (tuple) consisting of all measurements of type t , and $F^t(x)$ is the list consisting of all the corresponding values $F_i(x)$.

To find the desired solution x , we must minimize this expression under the constraint $g(x) \geq 0$.

2.4. Reformulation of the Problem

The more measurements of a given type t we have, the larger the contribution of these measurements to the solution. In general, all the terms $\frac{(F_i(x) - y_i)^2}{\sigma_i^2}$ in the sum (4) are approximately of the same type, so we can conclude that the joint contribution of all the measurements of type t is proportional to the number m_t of all measurement results of this type.

To compare the importance of measurements of different types, it is useful to define *relative importance* as a ratio of the importance m_t of this type to the overall value $\sum_{t'} m_{t'}$, i.e., as $\eta_t \stackrel{\text{def}}{=} \frac{m_t}{\sum_{t'} m_{t'}}$. These *influence parameters* η_t are non-negative numbers that add up to 1.

To better understand the minimized expression (5), it makes sense to show the explicit dependence on the influence parameters. This can be done, e.g., by making sure that each term in the new formula is proportional to η_t ; this way we will see that the larger the influence parameter, the larger the influence of this term. To do that, we replace each term c_t^2 with $\eta_t \cdot k_t^2$. From the condition that $c_t^2 = \eta_t \cdot k_t^2$, we conclude that $k_t^2 = \frac{c_t^2}{\eta_t}$. Since $c_t = \frac{1}{\sigma^t}$ and $\eta_t = \frac{m_t}{\sum_{t'} m_{t'}}$, we have

$$k_t = \sqrt{\frac{c_t^2}{\eta_t}} = \sqrt{\frac{1}{(\sigma^t)^2 \cdot m_t}} \cdot \sqrt{\sum_{t'} m_{t'}}. \quad (6)$$

Thus, each term $c_t^2 = \eta_t \cdot k_t^2$ has the form $c_t^2 = w_t^2 \cdot C$, where $w_t \stackrel{\text{def}}{=} \sqrt{\frac{\eta_t}{(\sigma^t)^2 \cdot m_t}}$ and $C \stackrel{\text{def}}{=} \sum_{t'} m_{t'}$. So, the minimized expression (5) takes the form

$$C \cdot \sum_{t=1}^T w_t^2 \cdot \|F^t(x) - y^t\|^2. \quad (7)$$

The location of the minimum does not change if we divide all the values of a function by the same positive coefficient C . Therefore, minimizing the expression (7) is equivalent to minimizing a simpler expression

$$\sum_{t=1}^T w_t^2 \cdot \|F^t(x) - y^t\|^2. \quad (8)$$

2.5. General Case: A Description

In the previous sections, we considered the ideal case, when we know the exact variance σ_i^2 of each measurement i . In this case, we can use the usual Least Squares approach to solve the joint inversion problem.

In practice, we only have an *approximate* knowledge of these variances. For example, for each measurement type t , we only know an approximate value $\tilde{\sigma}^t$ of the corresponding standard deviation σ^t .

A traditional approach to such situations is to use these approximate values $\tilde{\sigma}^t$ and solve the corresponding optimization problem. The problem with this approach is that, if a geophysical features appears in the solution corresponding to these approximate values of variances, there is no guarantee that this feature will still be visible if we use the actual (somewhat different) variances.

It is desirable to separate artifacts that are due to the specific choice of variances from the phenomena that occur no matter what variances we use. For this purpose, we wish to repeatedly solve the joint inversion problem with different possible combinations of variances. If a certain geological feature is visible in *all* these solutions, then we can be confident that this feature is also present in the actual solution corresponding to the actual (unknown) values of the variances – i.e., that it is the feature of the actual Earth structure.

2.6. General Case: Analysis of the Problem

In the previous sections, we described the need to minimize the expression

$$\sum_{t=1}^T \frac{1}{(\sigma^t)^2} \cdot \|F^t(x) - y^t\|^2 \quad (9)$$

under the condition $g(x) \geq 0$, where σ^t are the known standard deviations. We showed that this equivalent to minimizing the expression (8), where $w_t = \sqrt{\frac{\eta_t}{(\sigma^t)^2 \cdot m_t}}$ and $\eta_t = \frac{m_t}{\sum_{t'} m_{t'}}$.

In situations when we only know an approximate value $\tilde{\sigma}^t$, the traditional approach would be to use this approximate value, i.e., to minimize the expression

$$\sum_{t=1}^T \frac{1}{(\tilde{\sigma}^t)^2} \cdot \|F^t(x) - y^t\|^2, \quad (10)$$

or, equivalently, to minimize the expression (8), in which

$$w_t = \sqrt{\frac{\eta_t}{(\tilde{\sigma}^t)^2 \cdot m_t}}, \quad (11)$$

with the same values $\eta_t = \frac{m_t}{\sum_{t'} m_{t'}}$ of the influence parameters.

As we have mentioned, a more appropriate approach is to minimize expressions (9) corresponding to *all possible* combinations of standard deviations σ^t . Let us show that:

- each such constraint minimization problem can be equivalently reformulated into the form (8) with the weights (11) if we select *different* values of the influence parameters $\eta_t > 0$, and that
- for each combination of influence parameters $\eta_t > 0$ with $\sum_t \eta_t = 1$, there exist values $\sigma^t > 0$ for which the corresponding optimization problem (8) is equivalent to the original optimization problem (9).

Indeed, for each combination of values σ^t , let us take

$$\eta_t \stackrel{\text{def}}{=} \frac{1}{c} \cdot \frac{m_t \cdot (\tilde{\sigma}^t)^2}{(\sigma^t)^2},$$

where the normalization coefficient c is chosen as

$$c \stackrel{\text{def}}{=} \sum_t \frac{m_t \cdot (\tilde{\sigma}^t)^2}{(\sigma^t)^2}.$$

In this case, $\eta_t > 0$ and $\sum_t \eta_t = 1$. For the corresponding weights (11), we get

$$w_t^2 = \frac{\eta_t}{(\tilde{\sigma}^t)^2 \cdot m_t} = \frac{c}{(\sigma^t)^2}. \quad (12)$$

Thus, minimizing the expression (8) is equivalent to minimizing the expression

$$c \cdot \sum_{t=1}^T \frac{1}{(\sigma^t)^2} \cdot \|F^t(x) - y^t\|^2,$$

which, in its turn, is equivalent to minimizing the expression (9).

Vice versa, for each combination of weights $\eta_t > 0$ for which $\sum_t \eta_t = 1$, we can take

$$\sigma^t = \sqrt{\frac{m_t}{\eta_t \cdot \sum_{t'} m_{t'}}} \cdot \tilde{\sigma}^t. \quad (13)$$

For these standard deviations σ^t , the expression (9) takes the form

$$\sum_{t=1}^T (w'_t)^2 \cdot \|F^t(x) - y^t\|^2, \quad (14)$$

where we denoted

$$(w'_t)^2 \stackrel{\text{def}}{=} \sum_{t'} m_{t'} \cdot \frac{\eta_t}{(\tilde{\sigma}^t)^2 \cdot m_t}. \quad (15)$$

From the formula (11), we conclude that $w_t^2 = \frac{\eta_t}{(\tilde{\sigma}^t)^2 \cdot m_t}$ and thus, that the formula (15) takes the form

$(w'_t)^2 = C \cdot w_t^2$, where $C \stackrel{\text{def}}{=} \sum_{t'} m_{t'}$. Thus, the expression (14) takes an equivalent form

$$C \cdot \sum_{t=1}^T w_t^2 \cdot \|F^t(x) - y^t\|^2, \quad (16)$$

and the minimization of this expression is indeed equivalent to minimizing the expression (8).

The equivalence is proven.

Comment. The above analysis holds when we know the approximate values $\tilde{\sigma}^t$ of the corresponding accuracies σ^t , but we know no guaranteed bounds on these accuracies. In some practical situations, in addition to the approximate values $\tilde{\sigma}^t$, we also know the bounds $\underline{\sigma}^t$ and $\bar{\sigma}^t$, for which $\underline{\sigma}^t \leq \sigma^t \leq \bar{\sigma}^t$. In this case, instead of all possible values η_t of the corresponding influence parameter, we only need to consider possible values – and we can use the above formulas relating η_t with σ^t to transform bounds on σ^t into the corresponding bounds on η_t .

2.7. General Case: Resulting Optimization Problem

So, we arrive at the following equivalent reformulation of our problem: for all possible combination of influence factors $\eta_t > 0$ for which $\sum_t \eta_t = 1$, we compute the weights

$$w_t = \sqrt{\frac{\eta_t}{(\tilde{\sigma}^t)^2 \cdot m_t}},$$

and then minimize the expression

$$\sum_{t=1}^T w_t^2 \cdot \|F^t(x) - y^t\|^2$$

under the corresponding constraints $g(x) \geq 0$.

This minimization problem can be reformulated as

$$\min_x \|F(x) - y\|^2, \quad (17)$$

where

$$F(x) = W \begin{pmatrix} F^1(x) \\ F^2(x) \\ \cdots \\ F^t(x) \\ \cdots \\ F^T(x) \end{pmatrix} \in \mathbb{R}^m, \quad y = W \begin{pmatrix} y^1 \\ y^2 \\ \cdots \\ y^t \\ \cdots \\ y^T \end{pmatrix} \in \mathbb{R}^n$$

and $W = \text{diag}(w_t)$ is a diagonal matrix whose elements will be called *weights*.

In particular, when we reconstruct the values of shear velocities from Receiver Functions (RF), Surface Waves (SW), and Travel Times (TT), the corresponding minimize functional $J(x)$ takes the form

$$J = w_{RF}^2 \|F^{RF}(x) - y^{RF}\|^2 + w_{SW}^2 \|F^{SW}(x) - y^{SW}\|^2 + w_{TT}^2 \|F^{TT}(x) - y^{TT}\|^2,$$

i.e., equivalently, form (17), with

$$F(x) = W \begin{pmatrix} F^{SW}(x) \\ F^{RF}(x) \\ F^{TT}(x) \end{pmatrix} \in \mathbb{R}^m, \quad y = W \begin{pmatrix} y^{SW} \\ y^{RF} \\ y^{TT} \end{pmatrix} \in \mathbb{R}^n$$

and

$$W = \text{diag}(w_i), w_i = \sqrt{\frac{\eta_1}{(\tilde{\sigma}_i)^2 p}}, i = 1, \dots, p, w_i = \sqrt{\frac{\eta_2}{(\tilde{\sigma}_i)^2 q}}, i = p + 1, \dots, p + q, \\ w_i = \sqrt{\frac{1 - (\eta_1 + \eta_2)}{(\tilde{\sigma}_i)^2 r}}, i = p + q + 1, \dots, p + q + r, \quad (18)$$

$\tilde{\sigma}_i$ is the approximate standard deviation of each point, and p , q , and r are the number of RF, SW, and TT observations (Sosa et al. 2013).

2.8. Relation to Multi-Objective Optimization

When we have observations of only one type t , then, to find the corresponding model, we minimize the function $\|F^t(x) - y^t\|^2$. Minimizing this function is equivalent to minimizing the expression

$$f_t(x) \stackrel{\text{def}}{=} \frac{1}{(\tilde{\sigma}^t)^2 \cdot m_t} \cdot \|F^t(x) - y^t\|^2.$$

In situations when we have observations of different type, and when we only know the approximate values $\bar{\sigma}^t$ of the corresponding accuracies, we need to minimize expressions (8), i.e., equivalently, expressions of the type

$$\sum_t \eta_t \cdot f_i(x). \quad (19)$$

corresponding to all possible combinations $\eta_t > 0$ for which $\sum_t \eta_t = 1$.

It is known that, under reasonable conditions, the resulting set of solutions can be described in terms of the corresponding *multi-objective optimization problem* (MOP), namely, the problem of optimizing

$$f(x) \stackrel{\text{def}}{=} (f_1(x), f_2(x), \dots, f_i(x), \dots, f_T(x)).$$

For solving multi-objective problems, a natural idea is to generate the *Pareto optimal set* (also known as *Pareto front*), i.e., the set of all the values x for which it is not possible to improve all the criteria $f_i(x)$. In precise terms, the Pareto optimal set $P(x)$ is defined as

$$P(x) = \{x \in \Omega : \nexists x' \in \Omega (f(x') < f(x)), \quad (20)$$

where Ω is the set of all possible solutions that satisfy the corresponding constraints, and $f(x') < f(x)$ means that

$$\forall t (f_t(x') \leq f_t(x)) \& \exists t (f_t(x') < f_t(x)). \quad (21)$$

Under reasonable conditions, elements of the Pareto set can be obtained by finding the minima of all the functions (19) corresponding to all possible weights η_t adding to 1, and, vice versa, each such minimum is an element of the Pareto set $P(x)$.

In these terms, we can say that what we want in the general case, when we only know the approximate values of the corresponding accuracies, is to find the Pareto set of the multi-objective problem in which we minimize the criteria

$$f_i(x) = \text{const} \cdot \|F^t(x) - y_i\|^2$$

corresponding to measurements of different types t .

In particular, in our geophysical example, we want to minimize the three criteria $f_{FR}(x) = \text{const} \cdot \|F^{RF}(x) - y^{RF}\|^2$, $f_{SW}(x) = \text{const} \cdot \|F^{SW}(x) - y^{SW}\|^2$, and $f_{TT}(x) = \text{const} \cdot \|F^{TT}(x) - y^{TT}\|^2$.

2.9. How to Generate a “Typical” Solution

When we have data of several types, and we only know approximate values of the corresponding accuracies, our recommendation is to generate solutions corresponding to all possible combinations of actual accuracies. The number of such solutions is huge, and meaningfully analyzing all these solutions is difficult. It is therefore desirable to select, among these solutions, a solution which is, in some sense, typical.

The expectation is that, in general, this “typical” solution will only have features that all other solutions have. Thus, when we look for features common for all possible solutions, a good idea is to first analyze this typical solution, and then to check whether the features that we found on this solution are indeed present in all other solutions as well.

In multi-objective optimization, there are several possible ways of generating such a “typical” solution; see, e.g., (Sambridge 1999a, Sambridge 1999b, Kozlovskaya 2000). For example, once

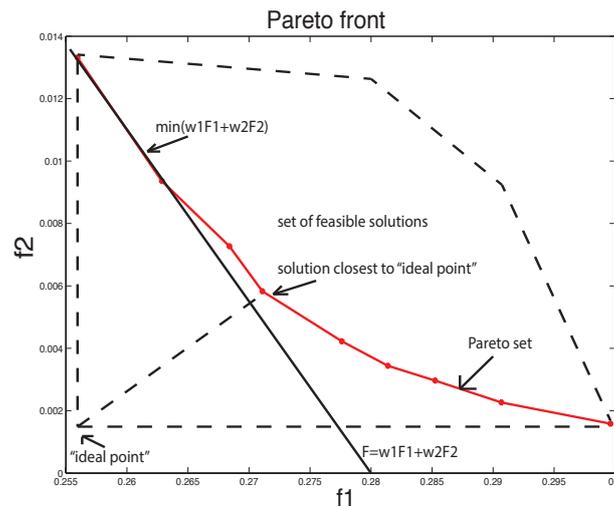


Figure 1. Illustration of the solution set or Pareto front, which is, defined as the weights times the perspective objective functions.

we find all solutions x corresponding to different combinations, then, for each criterion f_t , we can find the smallest value f_t^{\min} and the largest value f_t^{\max} . The smallest values form an *ideal point* $f^{\min} = (f_1^{\min}, \dots, f_t^{\min}, \dots, f_T^{\min})$. We then select a solution x which is the closest to this ideal point. Specifically, we normalize each differences $f_t(x) - f_t^{\min}(x)$ to the interval $(0,1)$ by dividing it by $f_t^{\max}(x) - f_t^{\min}(x)$, and then we minimize the corresponding normalized distance. In other terms, we select a solution x for which the distance

$$d^2(f^{\min}, f(x)) = \sum_{t=1}^T \left(\frac{f_t(x) - f_t^{\min}(x)}{f_t^{\max}(x) - f_t^{\min}(x)} \right)^2 \quad (22)$$

is the smallest possible.

3. Solving the Corresponding Constraint Optimization Problems

In the proposed approach, we need to solve several minimization problems, corresponding to different combinations of the influence parameters η_t . For each such combination, we need to find the value $\|F(x) - y\|^2$ under the constraint $g(x) \geq 0$. Let us show how to solve the corresponding constraint optimization problems.

3.1. Linearization

In most practical situations, we know the approximate values \bar{x}_1 of the corresponding quantities x , i.e., values for which $x \approx \bar{x}_0$. Since these values are close to each other, the difference $x - \bar{x}_0$ is small and thus, we can expand the functions $F(x)$ and $g(x)$ into Taylor series in this difference and safely ignore terms which are quadratic (or higher order), and only keep the first order Taylor approximation. Once we solve this first-order approximation problem, we get a better approximation \bar{x}_1 . We can use this approximation as the basis of a new linearization and get an even more accurate approximation, etc.

On each of these iterations, we start with an approximate model \bar{x}_k , and then we use the first order Taylor approximation of the operators F and g around \bar{x}_k :

$$F(x) \approx F(\bar{x}_k) + F'(\bar{x}_k)\Delta x = F(\bar{x}_k) + F'(\bar{x}_k)(x - \bar{x}_k), \quad (23)$$

$$g(x) \approx g(\bar{x}_k) + \nabla g^T(\bar{x}_k)(x - \bar{x}_k), \quad (24)$$

where $F'(\bar{x}_k)$ is the matrix formed by the partial derivatives of F and ∇g is the matrix formed by the partial derivatives of different components of $g(x)$.

Substituting the expressions (23) and (24) in the corresponding constraint optimization problem, we get the following linearized constraint optimization problem:

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F'(\bar{x}_k)x + r(\bar{x}_k)\|^2 \\ \text{s.t.} \quad & g(\bar{x}_k) + \nabla g^T(\bar{x}_k)(x - \bar{x}_k) \geq 0, \end{aligned} \quad (25)$$

where $r(\bar{x}_k) \stackrel{\text{def}}{=} F(\bar{x}_k) - y - F'(\bar{x}_k)\bar{x}_k$.

3.2. Primal Dual Interior-Point Method

To solve the linearized problem (25), we will use the Primal Dual Interior-Point method; see, e.g., (Sosa et al. 2013, Nocedal and Wright 2006). To use this method, we first reformulate our problem in a standard form as follows:

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F'(\bar{x}_k)x + r(\bar{x}_k)\|^2 \\ \text{s.t.} \quad & g(\bar{x}_k) + \nabla g^T(\bar{x}_k)(x - \bar{x}_k) - s = 0 \\ & s \geq 0 \end{aligned} \quad (26)$$

where components of $s \in \mathbb{R}^{2n}$ are called *slack variables*.

Then we define the Lagrange function associated to problem (26) as:

$$l(\bar{x}_k, z, s, w) = \frac{1}{2} \|F'(\bar{x}_k)x + r(\bar{x}_k)\|^2 - (g(\bar{x}_k) + \nabla g^T(\bar{x}_k)(x - \bar{x}_k) - s)^T z - s^T w \quad (27)$$

with the Lagrangian multipliers $z, w \in \mathbb{R}^{2n}$, $(z, w) \geq 0$. For a given perturbation parameter $\mu > 0$, the perturbed Karush-Kuhn-Tucker (KKT) or necessary conditions are given by:

$$\hat{F}(\bar{x}_k, z, s, w) = \begin{pmatrix} F'(\bar{x}_k)^T (F'(\bar{x}_k)x + r(\bar{x}_k)) - \nabla g^T(\bar{x}_k)z \\ g(\bar{x}_k) + \nabla g^T(\bar{x}_k)(x - \bar{x}_k) - s \\ z - w \\ S W e - \mu e \end{pmatrix} = 0, \quad (28)$$

where

$$\hat{F} : \mathbb{R}^{n+2n+2n} \longrightarrow \mathbb{R}^{n+2n+2n} \quad S = \text{diag}(s_1, \dots, s_{2n}), \quad W = \text{diag}(w_1, \dots, w_{2n}) \quad (29)$$

and $e = (1, \dots, 1) \in \mathbb{R}^{2n}$.

The formula (28) implies, in particular, that $z - w = 0$, and thus $z = w$. Hence the perturbed KKT system (28) is rewritten as

$$\hat{F}(x, z, s, w) = \begin{pmatrix} F'(\bar{x}_k)^T (F'(\bar{x}_k)x + r(\bar{x}_k)) - \nabla g^T(\bar{x}_k)z \\ g(\bar{x}_k) - s \\ SZe - \mu e \end{pmatrix} = 0; \quad (30)$$

thus the Jacobian associated to (30) is then computed as

$$F' \begin{pmatrix} x \\ z \\ s \end{pmatrix} = \begin{bmatrix} F'(\bar{x}_k)^T F'(\bar{x}_k) & -\nabla g^T(\bar{x}_k) & 0_{n \times n} \\ \nabla g(\bar{x}_k) & 0_{n \times m} & -I_{m \times m} \\ 0_{m \times n} & S & Z \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \\ \Delta s \end{bmatrix} = - \begin{bmatrix} \nabla_x l(x, z, s) \\ g(\bar{x}_k) - s \\ SZe - \mu e \end{bmatrix}$$

The system (31) can be simplified further by eliminating the third block of equations as follows. From the last block of equation in (31), we have

$$S \Delta z + Z \Delta s = -SZe + \mu e,$$

therefore

$$Z \Delta s = -SZe + \mu e - S \Delta z$$

$$\Delta s = -s + \mu Z^{-1} e - Z^{-1} S \Delta z,$$

and then

$$\nabla g^T(\bar{x}_k) \Delta x - \Delta s = \nabla g^T(\bar{x}_k) \Delta x + s - \mu Z^{-1} e + Z^{-1} e + Z^{-1} S \Delta z = -\nabla g^T(\bar{x}_k) x + s$$

$$\nabla g^T(\bar{x}_k) \Delta x + Z^{-1} S \Delta z = \mu Z^{-1} e - g(\bar{x}_k)$$

which allow us to write the reduced linear system

$$\begin{bmatrix} -F'(\bar{x}_k)^T F'(\bar{x}_k) & \nabla g^T(\bar{x}_k) \\ \nabla g(\bar{x}_k) & Z^{-1} S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = \begin{bmatrix} \nabla_x l(x, z, s) \\ Z^{-1} \mu e - g(\bar{x}_k) \end{bmatrix}$$

4. Geophysical Datasets

4.1. Receiver Functions

A receiver function is simply a time series representation of the Earth's response relative to an incoming P -wave propagating near a recording station. A representation of a receiver function is indicated in figure 2, with different incoming P -to- S converted waves and a time series representation of the Earth response underneath a seismic station. Positive or negative spike amplitudes represent positive or negative seismic velocity contrast. A receiver function technique can model the structure of the earth by using seismograms from three component (vertical, north, and east) seismic stations from teleseismic earthquakes. The receiver function technique takes advantage of the fact that part of the energy of seismic P waves is converted into S waves at discontinuities along the ray path (Bashir et al. 2011, Dzierma et al. 2011), and has been utilized in many studies; see, e.g., (Wilson et al. 2005, Wilson and Aster 2005, Bailey et al. 2012, Hansen et al. 2013). For data collection and processing,

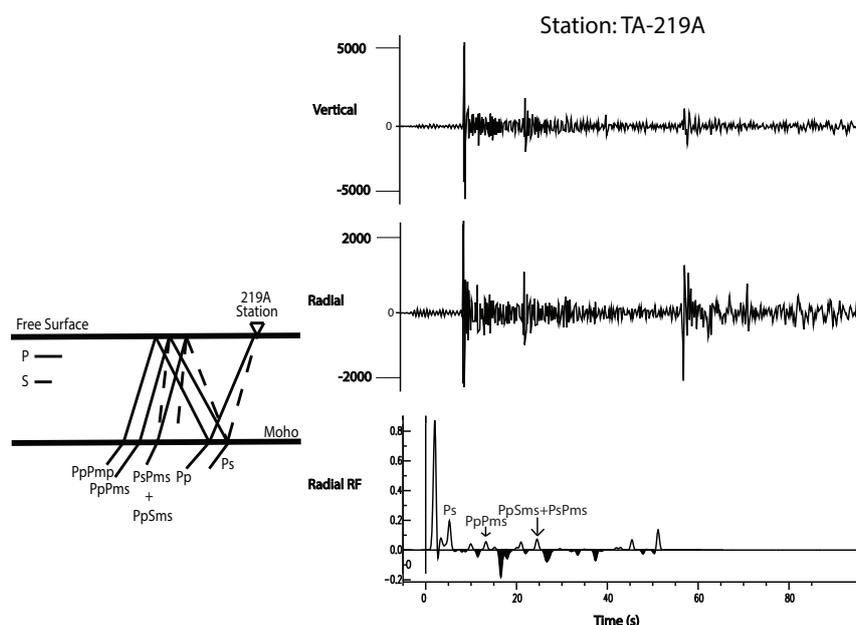


Figure 2. (Left) Illustration of a simplified ray diagram, which identifies the P_s , converted phases, which comprise the receiver function for a single layer. (Right) Vertical and radial seismograms and the corresponding receiver function resulting from the deconvolution of the vertical component from the radial component.

we use the Standing Order for Data (SOD) (Owens et al. 2004, Bailey et al. 2012) to request three component seismograms for P -wave arrivals and for events with a minimum magnitude 5.5, depth in the range of 1–600 km, and an epicentral distance ranging from 30° to 95° ; see, e.g., (Bailey et al. 2012).

Receiver functions were first applied in the late 1970s at solitary stations to obtain local one-dimensional structural estimates (Langston 1981). Since then, there was an increase in the number of stations deployed seismic experiments. It is now possible to generate detailed two or three-dimensional images of structures, such as the Moho and upper mantle transition zone discontinuities near 410 km and 670 km depth; see, e.g., (Wilson 2003).

Receiver functions are derived using deconvolution (Liggoria and Ammon 1991), a mathematical method used to filter a signal and isolate the superimposed harmonic waves. Specially, receiver functions are calculated by deconvolving the vertical component of a seismogram from the radial component, resulting in the identification of converted phases where there is an impedance contrast (crustal-mantle boundary) (Shearer, 2009).

4.2. Surface Wave Dispersion

Surface waves in general differ from body waves in many respects: they travel slower, lower frequencies, largest amplitudes, and their velocities are in fact dependent on frequency (Shearer 2009). The surface wave velocities vary with respect to depth being sampled by each period of the surface wave. The sampling by each period of the surface wave is known as dispersion (Shearer 2009). Valuable information can be inferred by measuring surface wave dispersion because it will allow you to

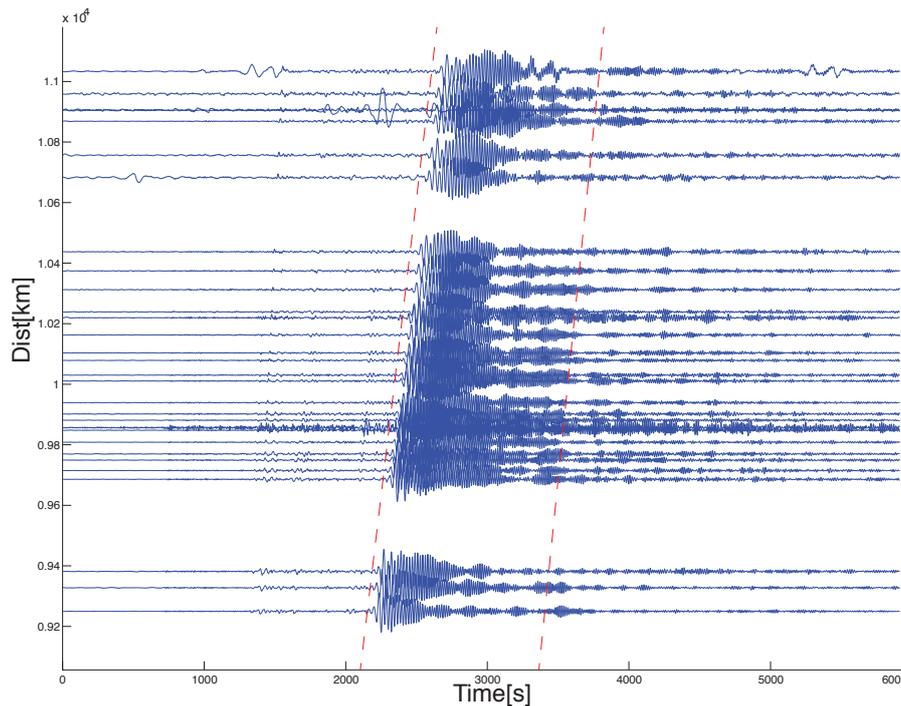


Figure 3. Example of surface wave Rayleigh waveforms for all stations used for this research. Dashed red line illustrates the window of where the phase curve passes through the stations.

be able to better understand the Earth's crustal and mantle velocity structure (Laske Masters and Reif 2000, Obrebski et al. 2010, Sosa et al. 2013). In particular, Love and Rayleigh wave group dispersion observations generally account for average velocity structure as a function of depth (Julia et al. 2000, Maceira and Ammon 2009). The dispersion curves for surface waves are extracted from station records of three component seismograms for different frequencies and distances, by using reduction algorithms that rely on spectral analysis techniques. The important fact here is that, 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. Figure 4 provides an example of teleseismic Rayleigh waveforms for the 12/01/2010 event. The Rayleigh's principle states that the phase velocity perturbation, denoted by $\frac{\delta c}{c}$, can be viewed as a function of $(K_\alpha, K_\beta, K_\rho)$, the sensitivity coefficients for P wave velocity, S wave velocity and density, respectively, i.e.,

$$\frac{\delta c(T)}{c(T)} = \int \left(K_\alpha \frac{\delta \alpha(z)}{\alpha(z)} + K_\beta \frac{\delta \beta(z)}{\beta(z)} + K_\rho \frac{\delta \rho(z)}{\rho(z)} \right) \quad (31)$$

where T is the period and z is the depth. By investigating sensitivity function variation in depth, the relative contribution of each property to dispersion can be shown. This subject is beyond the scope of our work, thus we just mention here that such analysis allows geophysicists to show that the relative contribution of P wave velocity, and density to dispersion is smaller than the one for S wave velocity (Julia et al. 2000). That is, surface wave dispersion is much more sensitive with respect to S wave velocity, and therefore we have established the dependence of this data set on shear wave velocity.

For this research, a Matlab-based software package – that automatically downloads, analyzes, and

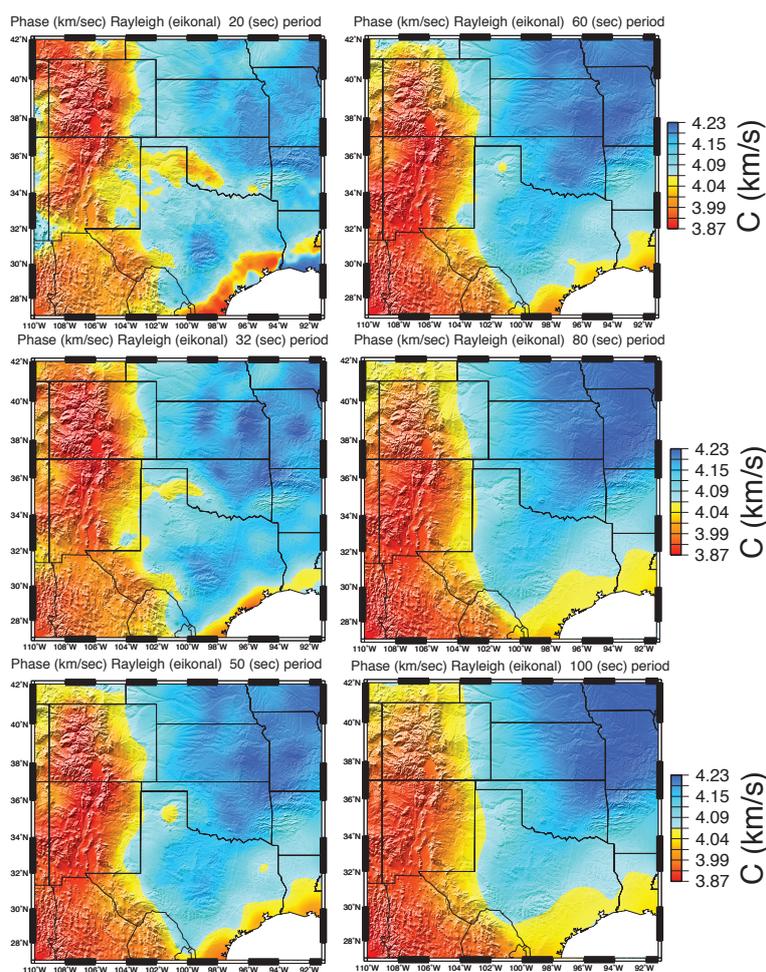


Figure 4. Surface wave phase map (Rayleigh) from USArray data used for this research. Red colors indicate low phase velocities and blue colors as high phase velocities for the Texas region.

measures phase as well as amplitude of surface waves to generate surface-wave tomography maps – was used to construct figure 5 describing tomographic images of the Texas region. The Automated Surface-Wave Phase-Velocity Measuring System (ASWMS) was the matlab package developed by (Jin and Gaherty, 2014) that developed this automated cross-correlation based method to generate surface-wave tomography of the entire U.S. The ASWMS tool was used to see what geological signatures that we can resolve using surface wave phase data.

4.3. Delay Travel Times

For this research, we used TauP toolkit ($\tau(p)$) and Antelope (BRTTO) database program to acquire the delayed travel times from the Array Network Facility (ANF) seismic catalog. Figure 6, shows an example of some of the delayed travel time data used for this study that were acquired from the ANF seismic catalog. The TauP toolkit program that we used to acquire the delayed travel times, uses the spherical Earth geometry into the computation. The TauP toolkit uses the relation of Snell's law,

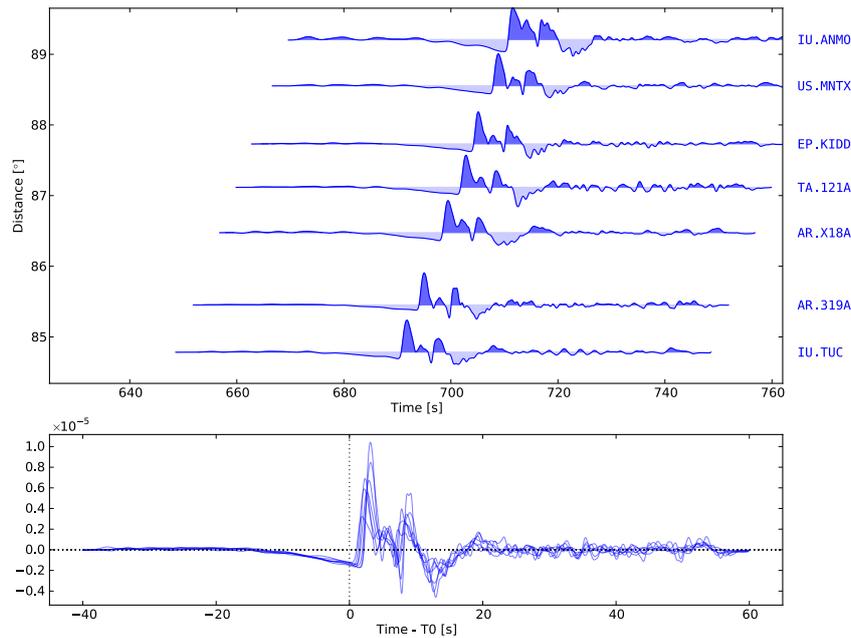


Figure 5. Example plot shows P -delay times for different USArray Stations that were used for this research from an earthquake during 09/15/2011. (Second) plot shows the difference between the predicted and measured travel times and illustrates the P -wave delay.

according to which, for each ray k , the ratio $\frac{\sin i_{kj}}{x_{kj}} = \frac{\Delta T_k}{l_{kj}}$ remains constant along the k -th ray, i.e., does not depend on j . This ratio is known as the constant ray parameter, and is usually denoted by p_k . We use this law to compute delayed travel times ΔT_k (i.e., components of the vector $F^{TT}(x)$):

$$\Delta T_k = \sum_{j=0}^n \frac{h_j}{x_{kj} \cos(i_{kj})}, \quad (32)$$

where h_j is the thickness of the j -th layer.

By using the Snell's law as mentioned earlier, the incidence angle i_{kj} can be rewritten as:

$$i_{kj} = \sin^{-1}(p_k x_j) \quad (33)$$

Equation (32) can be rewritten as

$$\Delta T_k = \sum_{j=0}^n \frac{h_j}{x_{kj} \cos(i_{kj})} = \sum_{j=0}^n h_j (x_{kj} \cos(\sin^{-1}(p_k x_{kj})))^{-1} \quad (34)$$

Using Snell's law and rewriting equation (32), we obtained the partial derivatives which are needed to use the primal dual interior point method mentioned earlier:

$$\frac{\partial \Delta T_k}{\partial x_{kj}} = \frac{-h_j}{x_{kj}^2 \cos(\sin^{-1}(p_k x_{kj}))^2} \left(\cos(\sin^{-1}(p_k x_{kj})) - \frac{(p_k x_{kj})^2}{\sqrt{1 - (p_k x_{kj})^2}} \right) \quad (35)$$

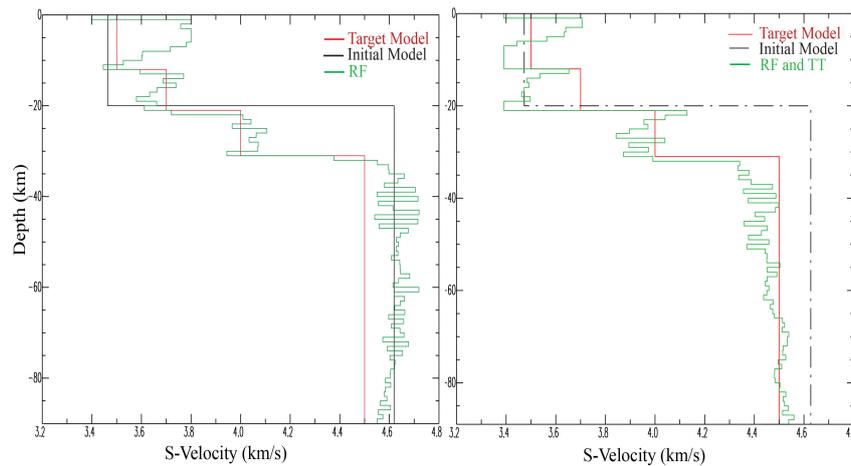


Figure 6. (Left) Single data inversion of receiver functions and synthetic rift model. (Right) joint inversion using receiver functions and travel times to obtain a better estimate of the rift model (target model).

Table 1. Relative RMS and residuals errors comparison for MOP joint inversion of surface waves and receiver function when travel times are added.

MOP Joint Inversion	RMS	Residual Error
RF & SW	0.00878717	0.00284017
RF & SW & TT	0.00739198	0.00030573

When ray paths between the source and the receiver are short enough, Earth's curvature is known to be negligible, which provides us with the importance of utilizing equation (35) for our computation of partial derivatives of T .

5. Results and Discussion

Based on the joint inversion results using multiple geophysical datasets, the compatibility of the datasets provides better estimates of the target model based on numerical experiments with the datasets and the synthetic rift model. In order to illustrate how receiver functions and surface wave dispersion velocities complement each other, we present in figures 6-8 the inversion results for the data sets created from the rift velocity model.

In figure 6, the joint inversion of both RFs and TT data sets gives a better approximation to the target model as expected based on the compatibility between the two datasets. The single inversion of receiver functions (left top) identifies the velocity interfaces (fig 6), while single inversion of surface waves (fig 7) gives information on the average velocities at different depths. For figure 7, the joint inversion of SWs and TT also provides a better approximation of the target model indicated in red.

The joint inversion of RFs, SWs, and TT provides more accuracy and compatibility when performing joint inversion of the three geophysical datasets as shown in figure 8.

A synthetic rift velocity model that we developed was used as our initial test for the MOP joint inversion scheme. For the MOP joint inversion scheme, a comparison was done with the rift velocity

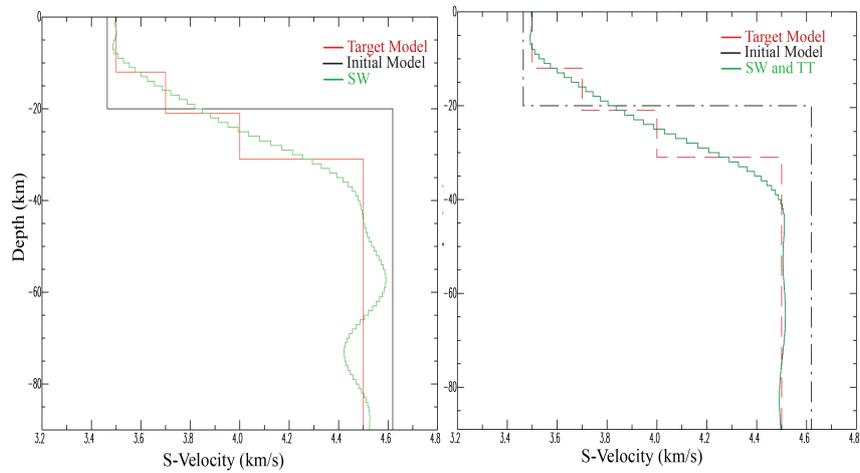


Figure 7. (Left) Single data inversion of surface wave dispersion measurements and synthetic rift model. (Right) joint inversion using surface wave dispersion measurements and travel times to obtain a better estimate of the rift model (target model).

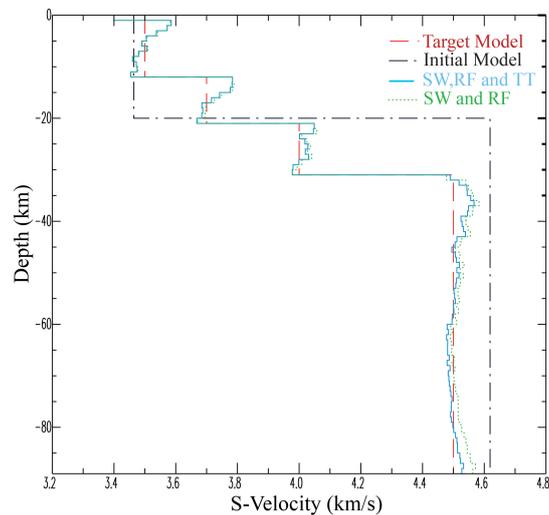


Figure 8. Joint inversion using receiver functions, surface wave dispersion measurements, and travel times to obtain a better estimate of the rift model (target model).

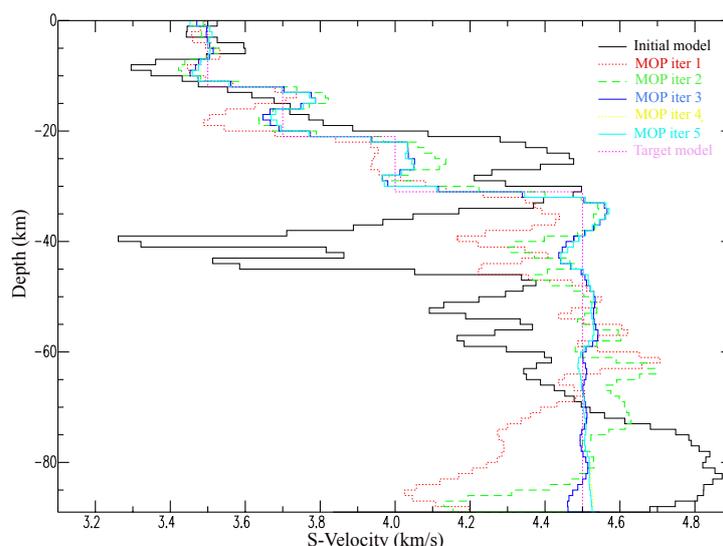


Figure 9. 1-D joint inversion results for a synthetic 1-D orogen velocity model. The target velocity model in purple is the rift velocity model. The black color represents the initial orogen model. The other colors represent the improvement of the 1-D shear wave velocity models matching with the target model (purple).

model and the initial velocity test model. The synthetic rift model was the best velocity model used for this joint inversion scheme. Different synthetic Earth velocity models were used but overall, the rift model was the best. Numerous tests were performed to test the compatibility and complementary nature of the multiple geophysical datasets based from the results in figures 9-11. The algorithm used to perform the joint inversion of the multiple geophysical datasets using the multi-objective joint inversion scheme was written in FORTRAN 77 and coupled with a C code that performs the Multi-Objective Optimization method, based on the work of [25].

6. Conclusion

We apply an inversion scheme that expands a joint-inversion constraint least-squares (LSQ) algorithm used to characterize a one-dimensional Earth's structure. We utilize the Multi-Objective Optimization technique to perform joint inversion of multiple data sets (receiver functions, surface wave dispersion, and travel times) to develop 3-D shear wave models like in figure 12 (e.g., Thompson et al., submitted for publication). By jointly inverting these three geophysical data sets, we improve the model's accuracy. In the ideal situation when we know the relative accuracy of different datasets, we can formulate the joint inversion problem as a (single) least-square optimization problem. In practice, however, we only know approximate values of these accuracies; so, for inversion, we use the Multi-Objective Optimization Problem (MOP) approach. Different combinations of weights were incorporated in the MOP inversion scheme in order to map the Pareto Set (Solution Space) corresponding to receiver functions, surface wave dispersion measurements, and travel times. From the Pareto Set, the MOP technique performs a direct search method that selects the most feasible solution from a set of alternative solutions from the model space (Sambridge 1999a, Sambridge 1999b, Kozlovskaya 2000).

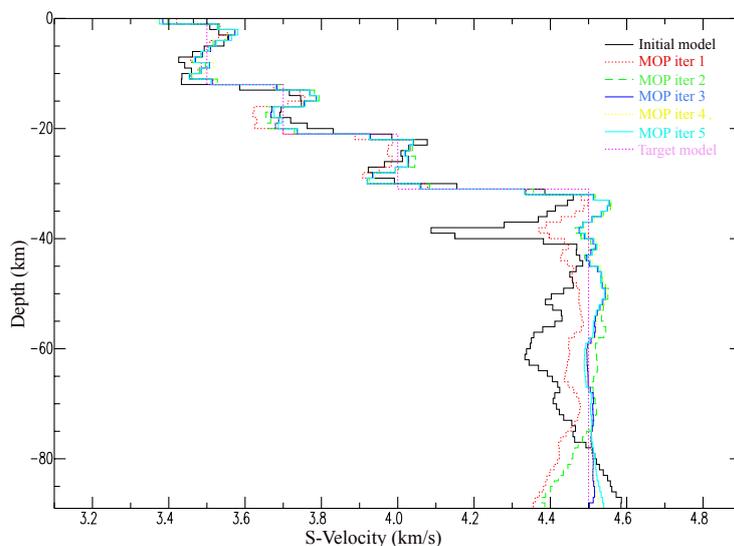


Figure 10. 1-D joint inversion results for a synthetic 1-D continental velocity model. The target velocity model in purple is the rift velocity model. The black color represents the initial continental model. The other colors represent the improvement of the 1-D shear wave velocity models matching with the target model (purple).

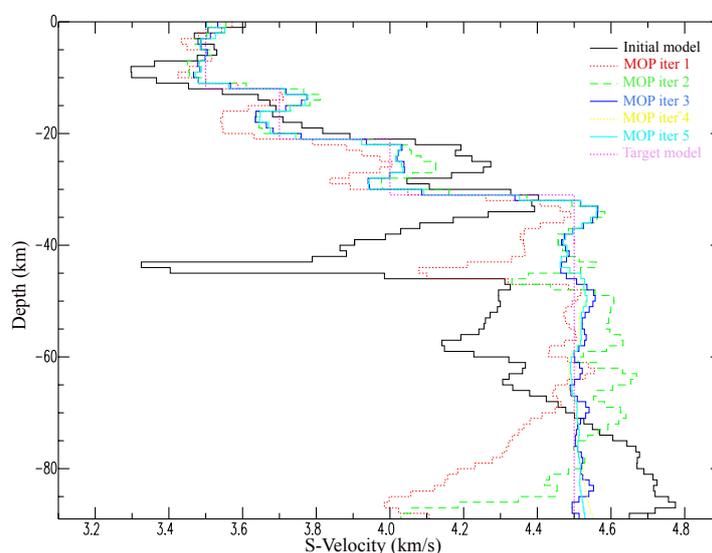


Figure 11. 1-D joint inversion results for a synthetic 1-D archean velocity model. The target velocity model in purple is the rift velocity model. The black color represents the initial archean model. The other colors represent the improvement of the 1-D shear wave velocity models matching with the target model (purple).

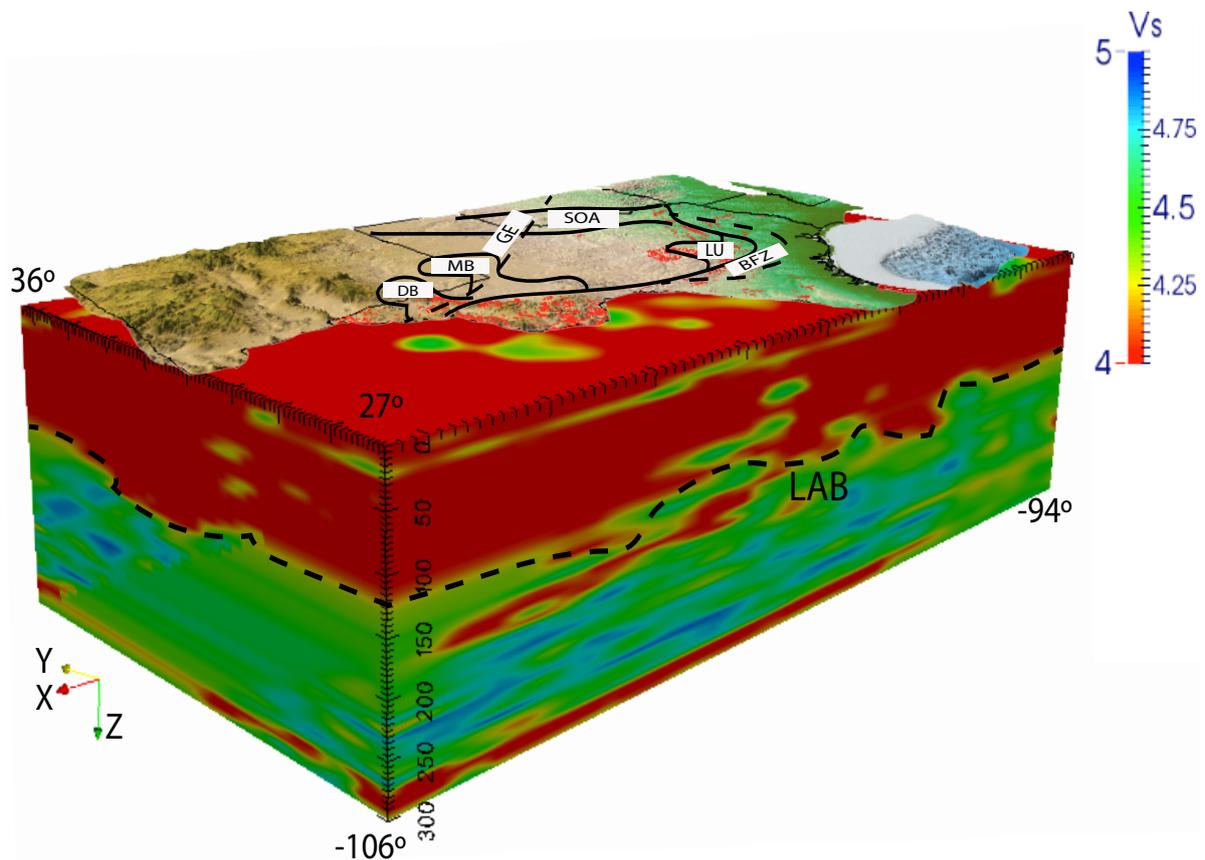


Figure 12. 3-D shear wave model of the Texas region from the surface to 300km depth. The 3-D shear wave model is the final product obtain by the MOP joint inversion scheme using the three geophysical datasets: receiver functions, surface wave dispersion measurements, & travel times. Red colors represent low shear wave velocities (4 km/s) and blue colors represent high shear wave velocities (5 km/s). DB: Delaware Basin, MB: Midland Basin, GE: Grensville Orogeny, SOA: South Oklahoma Aulcagen, LU: Llano Uplift, BFZ: Balcones Fault Zone.

For future work, we plan to incorporate gravity into our inversion scheme to obtain a more constrained Earth structure and to better determine physical properties of the Earth structure.

Acknowledgments

We would like to take the time to thank the computational science, mathematical science, and computer science departments from University of Texas at El Paso (UTEP). We would also like to thank Ezer Patlan, and Dr. Anibal Sosa for all of their contributions to this work. This work was sponsored by the NSF CREST under Grant Cybershare HRD-0734825.

References

1. Bashir, L., S.S. Gao, K.H. Liu, and K. Mickus (2011). Crustal structure and evolution beneath the Colorado Plateau and the southern Basin and Range Province: Results from receiver function and gravity studies. *Geochem. Geophys. Geosyst.*, 12, Q06008, doi:10.1029/2011GC003563.
2. Bailey, I.W., M.S. Miller, K. Liu, and A. Levander (2012). Vs 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:10.1029/2011JB0085.
3. Cho, K. H., R. B. Herrmann, C. J. Ammon, and K. Lee (2007). Imaging the upper crust of the Korean Peninsula by surface-wave tomography. *Bulletin of the Seismological Society of America*, 97, pp 198–207.
4. Colombo, D., and M. De Stefano (2007). Geophysical modeling via simultaneous joint inversion of seismic, gravity, and electromagnetic data: Application to prestack depth imaging. *The Leading Edge*, 26, pp 326–331.
5. Dzierma, Y., W. Rabbel, M.M. Thorwart, E.R. Flueh, M.M. Mora, and G.E. Alvarado (2011). The steeply subducting edge of the Cocos Ridge: evidence from receiver functions beneath the northern Talamanca Range, south-central Costa Rica. *Geochem. Geophys. Geosyst.* 12: doi:10.1029/2010GC003477.
6. Gurrola, H., E. G. Baker, and B.J. Minster (1995). Simultaneous time-domain deconvolution with application to the computation of receiver functions. *Geophys. J. Int.*, 120, pp. 537–543.
7. Jin, G., and J. B. Gaherty (2014), Surface Wave Measurement Based on Cross-correlation, *Geophys. J. Int.*, submitted.
8. Haber, E., and D. Oldenburg (1997). Joint inversion: a structural approach. *Inverse Problems*, 13, pp. 63–77.
9. Hansen, P.C. (2010). *Discrete Inverse Problems: Insight and Algorithms*, 225pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.
10. Hansen, S.M., K.G. Dueker, J.C. Stachnik, R.C. Aster, and K.E. Karlstrom (2013). A rootless rockies - Support and lithospheric structure of the Colorado Rocky Mountains inferred from CREST and TA seismic data. *Geochem. Geophys. Geosyst.*, 14, 2670–2695, doi:10.1002/ggge.20143.

11. Julia, J., C. J. Ammon, R. Hermann, and M. Correig (2000). Joint inversion of receiver function and surface wave dispersion observations. *Geophys. J. Int.*, 142, pp. 99–112.
12. Kozlovskaya, E. (2000). An algorithm of geophysical data inversion based on non-probabilistic presentation of a-prior information and definition of pareto-optimality. *Inverse Problems*, 16, pp. 839–861.
13. Langston, C. A. (1981). Evidence for the subducting lithosphere under southern Vancouver Island and western Oregon from teleseismic P wave conversions. *J. Geophys. Res.*, 86, pp. 3857–3866.
14. Laske, G., G. Masters and C. Reif (2000). Crust 2.0. The Current Limits of Resolution for Surface Wave Tomography in North America. *EOS Trans AGU*, 81, F897, <http://igpppublic.ucsd.edu/gabi/ftp/crust2/>
15. Lees, J.M. and J. C. Vandecar (1991). Seismic tomography constrained by bouguer gravity anomalies: Applications in western Washington. *PAGEOPH*, 135, pp 31–52.
16. Ligorria, J. P., and C. J. Ammon (1999), Iterative deconvolution and receiver-function estimation, *Bull. Seismol. Soc. Am.*, 89(5), pp 1395–1400.
17. Maceira, M., and C.J. Ammon (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.
18. Nocedal, J. and S.J. Wright (2006). *Numerical Optimization*. 2nd edn. Springer, New York, NY.
19. Obrebski, M., S. Kiselev, L. Vinnik, and J. P. Montagner (2010). Anisotropic stratification beneath Africa from joint inversion of SKS and P receiver functions. *J. Geophys. Res.*, 115, B09313, doi:10.1029/2009JB006923.
20. Owens, T. J., H. P. Crotwell, C. Groves, and P. Oliver-Paul (2004). SOD: Standing Order for Data. *Seismol. Res. Lett.*, 75, 515–520.
21. Sambridge, M. (1999a). Geophysical inversion with a neighborhood algorithm I: searching a parameter space. *Geophys. J. Int.*, 138, pp. 479–494.
22. Sambridge, M. (1999b). Geophysical inversion with a neighborhood algorithm II: appraising the ensemble. *Geophys. J. Int.*, 138, pp. 727–746.
23. Shearer, P.M. (2009). *Introduction to Seismology*, Second Edition, Cambridge University Press, Cambridge.
24. Shen W., M. H. Ritzwoller, and V. Schulte-Pelkum (2013). A 3-D model of the crust and uppermost mantle beneath the Central and Western US by joint inversion of receiver functions and surface wave dispersion. *J. Geophys. Res. Solid Earth*, 118, doi:10.1029/2012JB009602.
25. Sosa, A., A.A. Velasco, L. Velasquez, M. Arguez, and R. Romero. (2013). Constrained Optimization framework for joint inversion of geophysical data sets. *Geophys. J. Int.*, 195, pp. 197–211.
26. Stein, S., and M. Wysession (2003). *An Introduction to Seismology Earthquakes and Earth Structure*, Blackwell, Maiden, Mass.
27. Thompson, L., A. A. Velasco, V. Kreinovich, R. Romero, and A. Sosa, 2016, 3-D Shear Wave Based Models of the Texas Region Using 1-D Constrained Multi-Objective Optimization, *Journal of Geophysical Research*, (submitted for publication).

28. Tikhonov, A. N., and V.Y. Arsenin (1977). *Solution of Ill-Posed Problems*. VH Winston & Sons, Washington, D.C.
29. Vogel, C. R. (2002). *Computational Methods for Inverse Problems*. SIAM FR23, Philadelphia.
30. Vozoff, K. and D. L. B. Jupp (1975). Joint inversion of geophysical data. *Geophys. J. Roy Astr. Soc.*, 42, pp. 977–991.
31. Wilson, D. (2003). Imagining crust and upper mantle seismic structure in the southwestern United States using teleseismic receiver functions. *Leading Edge* 22, pp. 232–237.
32. Wilson, D., and R. Aster (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.*, B05305, doi:10.1029/2004JB003430.
33. Wilson, D., R. Aster, J. Ni, S. Grand, M. West, W. Gao, W.S. Baldrige, and S. Semken (2005). Imaging the structure of the crust and upper mantle beneath the Great Plains, Rio Grande Rift, and Colorado Plateau using receiver functions. *J. Geophys. Res.*, 110, B05306, doi:10.1029/2004JB003492.

For more questions regarding reference style, please refer to the Citing Medicine.

Appendix

Joint Inversion Algorithm (Two Datasets)

1. Given an initial velocity model $(V_s)_0 = x_0$ *RF* observations y^{RF} , *SW* dispersion observations y^{SW} , and a max. number of iterations l .
2. for $\eta = 0, 0.1, 0.2, \dots$, until 1.0 do
3. Evaluate y^{RF}, y^{SW}
4. for $k = 0, 1, 2, \dots$, until l do
5. Evaluate $F(x_k), F'(x_k)$ and compute $b = F(x_k) - y - F'(x_k)x_k$
6. Solve Equation (15) by using PDIP method.
7. if $\|F(x_k) - y\| \leq \epsilon$ then
8. break
9. end if
10. Go to Step 4
11. end for
12. end for

Joint Inversion Algorithm (Three Datasets)

1. Given an initial velocity model $(V_s)_0 = x_0$ *RF* observations y^{RF} , *SW* dispersion observations y^{SW} , y^{TT} and a max. number of iterations l .
2. for $\eta = 0, 0.1, 0.2, \dots$, until 1.0 do
3. Evaluate y^{RF}, y^{SW}
4. for $k = 0, 1, 2, \dots$, until l do
5. Evaluate $F(x_k), F'(x_k)$ and compute $b = F(x_k) - y - F'(x_k)x_k$

-
6. Solve Equation (15) by using PDIP method.
 7. if $\|F(x_k) - y\| \leq \epsilon$ then
 8. break
 9. end if
 10. Go to Step 4
 11. end for
 12. end for



©2016, Lennox Thompson et al., licensee AIMS Press.
This is an open access article distributed under the
terms of the Creative Commons Attribution License
(<http://creativecommons.org/licenses/by/4.0>)