

Velocity estimation via registration-guided least-squares inversion

Hyoungsu Baek*, Henri Calandra†, Laurent Demanet*

* *Department of Mathematics, MIT, Cambridge, MA 02139, USA. Email:*

hbaek@math.mit.edu, laurent@math.mit.edu

† *TOTAL Exploration & Production, Avenue Larribau, 64018 Pau, France. Email:*

henri.calandra@total.com

(November 27, 2013)

Running head: **Registration guided least-squares inversion**

ABSTRACT

Least-squares acoustic waveform inversion often suffers from a very narrow basin of attraction near the global minimum. In order to mitigate this problem, this paper introduces an iterative inversion scheme where the notion of proximity of two traces is not the usual least-squares distance, but instead involves registration as in image processing. Observed data are matched to predicted waveforms via piecewise-polynomial warpings, obtained by solving a nonconvex optimization problem in a multiscale fashion from low to high frequencies. This multiscale process requires defining low-frequency augmented signals in order to seed the frequency sweep at zero frequency. Custom adjoint sources are then defined from the warped waveforms. The new method, referred to as Registration-guided least-squares, is successfully applied to a few scenarios of model velocity estimation in the transmission setting. We show that the new method can converge to the correct model in situations where conventional least-squares inversion suffers from cycle-skipping and converges to a

spurious model.

INTRODUCTION

Waveform inversion via non-linear least-squares (LS) minimization (Tarantola and Valette, 1982) is effective when the starting model is accurate (Virieux and Operto, 2009), but otherwise suffers from stalled convergence to spurious local minima due to the oscillatory nature of the data and nonlinearity. The presence of local minima in seismic inversion was clearly demonstrated with the so-called Camembert example (Gauthier et al., 1986). In order to prevent convergence to a local minimum, frequency sweeps in full waveform inversion (FWI) were proposed by many authors including Bunks et al. (1995) and Pratt (1999), and consist in fitting data from low to high frequencies. However, the lack of low-frequency data, or their corruption by noise, often hinders this frequency sweep approach. Accurate initial models are typically found using traveltimes tomography (Bregman et al., 1989; Pratt and Gouly, 1991; Prieux et al., 2013), which are then improved upon by waveform inversion. Instead of taking two separate steps for inversion, there have also been efforts to combine traveltimes tomography and waveform inversion to exploit the advantages of both methods: convexity of traveltimes tomography and high resolution of waveform inversion (Luo and Schuster, 1991). Gee and Jordan also took advantage of robust traveltimes information rather than sensitive amplitude in the seismogram (Gee and Jordan, 1992). Fichtner et al. (2008) proposed an objective functional which minimizes envelope and phase misfits using the time-frequency representation of traces with the flexibility of emphasizing phase first and then envelope misfits later. Bozdağ et al. (2011) showed that FWI using misfits defined with instantaneous phase and envelope reduces the non-linearity of waveform modeling. In the same line of research, an objective functional defined by the energy in the crosscorrelation of observed and predicted data was proposed and studied by Van Leeuwen and Mulder (2010), though an analysis of their method was provided in Baek and Demanet

(2013).

In this paper, we propose a method which implicitly extracts phase information by solving auxiliary seismogram registration sub-problems. The resulting method recovers the velocity model in some transmission scenarios, without traveltimes picking, and even when the data only contain high frequencies. We formulate the registration problem with piecewise polynomials that can be found from the comparison of non-linearly transformed waveforms such as envelopes.

A way to incorporate this new kinematic information into waveform inversion is to replace the adjoint source that appears in the model update, normally the difference $u - d$ between the predicted data u and observed data d , by a geometrically meaningful quantity that does not suffer from cycle-skipping. The motivation for correcting the adjoint source is that the phases of d are in general off by more than one wave period in comparison to those of u . In contrast, we define *fractionally warped data* \tilde{d} so that their phases match those of the prediction u to within a small fraction of a period. This concept will of course be given a precise definition in the sequel. We demonstrate through numerical inversion examples that least-squares misfit optimization with the warped data can have a much enlarged basin of attraction. We refer to our method as *registration-guided least-squares* (RGLS) inversion.

The necessity of extracting shifts or warping in many applications has given rise to many schemes under different names. For example, ideas related to registration include traveltimes delay based on crosscorrelations (Luo and Schuster, 1991), image registration using optimal transport (Haker and Tannenbaum, 2001), curve registration (Ramsay and Li, 1998), registration using local similarities (Fomel and Jin, 2009; Fomel and van der Baan, 2010), and dynamic time warping for speech pattern matching (Sakoe and Chiba, 1978). Finding

traveltime discrepancies between two traces is not a trivial problem, especially for multiple waves with different traveltime discrepancies. Maggi et al. (2009), for example, developed an automated algorithm to select time windows to extract time shifts from isolated waves with iterative tomographic inversion in mind. Liner and Clapp (2004) pointed that trace alignment is not a mere translation but a time warping, and used a global optimization algorithm used in amino acid alignment for seismic trace processing and interpretation. Hale (2013) further improved a different dynamic programming method developed for speech recognition with proper constraints, recovering time shifts which are a few times larger than a period/wavelength in a stable and robust manner. Kennett and Fichtner (2012) defined a generalized mapping between traces, introducing transfer operators which map seismograms in a similar way to our piecewise polynomial mapping.

In order to find the best warping between an observed trace d and the corresponding predicted trace u , we formulate and solve an optimization problem which, not unlike FWI, is itself nonconvex. The highly oscillatory nature of the traces is also what makes seismogram registration nontrivial. We show that the nonconvexity of the matching problem is tractable and can be handled by a continuation strategy, where the match is realized scale-by-scale in a careful, iterative fashion. The traces d and u usually do not contain useful low frequencies in exploration seismology, so the seeding problem of this multiscale iteration is as much an issue here as in classical FWI. We propose to solve this problem by introducing nonlinearly transformed signals which, by construction, contain low-frequency components. We refer to these convenient, nonphysical nonlinearly-transformed signals as *low-frequency augmented* (LFA) signals. The LFA transformation can be thought of as an ad-hoc pre-processing of the traces so as to create low frequencies, yet maintain much of the information at high frequencies. Subsequently, seismogram registration is realized through the match of the

LFA of d to the LFA of u by a warping function of limited complexity, such as a piecewise polynomial.

This paper is part of the community’s broad effort to enlarge the basin of attraction of FWI by replacing least-squares by other objective functions, or by directly modifying the adjoint source, as in Luo and Schuster (1991); Gee and Jordan (1992); Fichtner et al. (2008); Van Leeuwen and Mulder (2010); Shah et al. (2010). To the best of our knowledge, however, no studies have yet proposed to modify the adjoint source *by replacing observed data with time-warped predicted data* for this purpose. Moreover, we illustrate the benefits of considering piecewise polynomials (as an alternative to dynamic warping, for instance) to define mappings between different images or traces. Finally, the idea of transforming traces to generate low frequencies (including 0 - 5 Hz) seems to have been mostly overlooked by the community. The work of Shin and Ha (2008) and Shin and Cha (2009) is an important exception, where the LFA is realized by a decaying exponential, but we are unaware that the type of nonlinearity that we consider in this paper had been previously used for the purpose of frequency augmentation.

The paper is organized as follows. We start by explaining the motivation behind modifying the adjoint source to the adjoint state equation. We then detail trace a registration method. Seismogram registration at the trace level is demonstrated with synthetic noisy and noiseless data. The RGLS inversion method is then tested in several transmission cases, with Gaussian high/low velocity models and a smoothed Marmousi model. We show that the LS and RGLS methods behave significantly differently. We finish by discussing the limits of the proposed method as well as comparisons with other similar methods for inversion and registration. In a nutshell, seismogram registration requires comparable traces, which explains why we consider transmission rather than reflection examples in this paper.

GUIDED LEAST-SQUARES WITH A MODIFIED ADJOINT SOURCE

Full waveform inversion (FWI), in its standard form, tries to minimize the least-squares misfit

$$J[m] = \frac{1}{2} \sum_{s,r} \int |S_{s,r}u_s(x, t) - d_s(x_r, t)|^2 dt, \quad (1)$$

where $m(x)$ denotes the squared slowness and $S_{s,r}$ is a sampling operator; predicted and observed data at a shot s and at a receiver x_r are denoted by $u_s = \mathcal{F}_s[m]$, and d_s , respectively. For notational simplicity, the subscripts s and r are omitted whenever it does not cause confusion. Moreover, the sampling operator $S_{s,r}$ is omitted when the predicted data $S_{s,r}u_s$ is compared with the corresponding observed data d_s ; the residual $S_{s,r}u_s(x, t) - d_s(x_r, t)$ may be written as $u - d$. In this paper the forward operator $\mathcal{F}_s[m]$ maps a squared slowness m to data $u_s(x_r, t)$ through the acoustic wave equation,

$$m \frac{\partial^2 u_s}{\partial t^2} = \Delta u_s + f_s(x, t), \quad (2)$$

where $f_s(x, t)$ is a source term. The adjoint-state method generates the gradient of $J[m]$

$$\frac{\delta J}{\delta m}[m] = - \sum_s \int q_s(x, t) \frac{\partial^2 u_s}{\partial t^2}(x, t) dt, \quad (3)$$

where the adjoint field q_s is propagated backward in time from the receiver positions in the medium m using the data residual as right-hand side (Plessix, 2006).

It is well known that the nonconvexity of $J[m]$ is particularly pronounced when the data are oscillatory. More specifically, when the time difference between corresponding arrivals in $S_{s,r}u_s$ and d_s is larger than a half period, the steepest descent direction of the data misfit may result in increasing those time differences, consequently increasing the model error but still decreasing the misfit error. In order to guide the iterations in a better direction, we propose to change the residual $S_{s,r}u_s - d_s$ in the adjoint wave equation by replacing d_s

by a version of $S_{s,r}u_s$ transported “part of the way” toward d_s . We denote these virtual, transported data by \tilde{d}_s and refer to them as *fractionally warped data*. The rationale behind \tilde{d}_s is that its arrivals can now be less than a half period apart from those in $S_{s,r}u_s$.

In order to generate a good candidate of fractionally warped data, we propose to find piecewise cubic polynomials $p(t)$ and $A(t)$ so as to have a good match $d_s(t) \approx A(t)S_{s,r}u_s(p(t))$, then define fractionally warped data as

$$\tilde{d}_s(t) = [A(t)]^\alpha S_{s,r}u_s((1 - \alpha)t + \alpha p(t)) \quad (4)$$

with some very small $0 < \alpha \ll 1$. When the parameter α is close to 1, $\tilde{d}_s(t)$ are located near the observed data $d_s(t)$; such \tilde{d}_s , which are far away from the predicted data $S_{s,r}u_s(t)$, result in cycle-skipping. Hence, a small α close to 0 is used to define $\tilde{d}(t)$. A generic way to find a proper α seems to be $\alpha T_d < \frac{1}{2}T_p$, i.e., $\alpha < \frac{T_p}{2T_d}$, where T_d and T_p are the largest traveltime discrepancy in a shot and the wave period, respectively.

Let us remark that it is the prediction which is transported toward the observed data, and not the other way around. A definition of \tilde{d}_s in place of $S_{s,r}u_s$ and a large $\alpha \simeq 1$ is also possible, but we believe that this choice would be inferior. The substitution of d_s by \tilde{d}_s is illustrated in Figure 1. The underlying assumption of our proposed method is that the observed data d_s are a warped version of the corresponding predicted data $S_{s,r}u_s$. However, we are aware that there are cases in which such an assumption does not hold; different background velocity models result in not only warping waves in time but also the appearance or the disappearance of waves. The limitations are commented on in the discussion section.

We can now write the registration-guided least-squares method (RGLS) as the following 2-level local optimization problem: assume m_{k-1} is known from the previous iteration, then

obtain m_k from one gradient step for the least-squares misfit

$$J_k[m] = \frac{1}{2} \sum_{s,r} \int |S_{s,r}u_s(x,t;m) - \tilde{d}_s(x_r,t;m_{k-1})|^2 dt, \quad (5)$$

where

$$\tilde{d}_s(x_r,t;m_{k-1}) := A^\alpha(t)S_{s,r}u_s(x,(1-\alpha)t + \alpha p(t);m_{k-1}), \quad (6)$$

and the warping parameters are optimal in the sense that

$$(p(t), A(t)) = \operatorname{argmin} W_{LFA}[p, A], \quad (7)$$

where the objective function W_{LFA} is defined in the next section. The expression of W_{LFA} only involves m_{k-1} , not m , hence $J_k[m]$ only depends on m via $u_s(x,t;m)$. The gradient of $J_k[m]$ is obtained in a standard fashion via the adjoint-state method, as mentioned earlier, as the migration operator applied to the adjoint source $S_{s,r}u_s - \tilde{d}_s$.

Notice that our update consisting of the migrated image of $S_{s,r}u_s - \tilde{d}_s$ can be interpreted either as a modified gradient for the objective functional 1, or as the gradient for the modified objective functional 5. In either case, it is clear that these updates are not expected to be gradients of any single objective function – hence the term local optimization.

The introduction of fractionally warped data and a modified adjoint source changes a conventional LS inversion algorithm slightly by adding a registration step before the backward modeling step. The registration step is highly parallelizable and can be sped up exploiting data redundancy; a detailed cost analysis is given in the discussion section. The overall computational cost for the RGLS method is slightly higher than that of the LS method, by a few percentage in our parallelized implementation. An RGLS inversion algorithm is presented in the next section after we state the sub-problem of RGLS inversion, i.e., the seismogram registration problem.

SEISMOGRAM REGISTRATION

The warping $p(t)$, the amplitude $A(t)$, and the value of α are chosen so that fractionally warped data $\tilde{d}(t)$ have a similar shape to that of the prediction $u(t)$ but phase discrepancies smaller than a half of a wave period. In order to find $p(t)$ and $A(t)$, we propose a non-convex optimization scheme similar to image registration (Glasbey and Mardia, 1988; Zitová and Flusser, 2003). The proposed FWI method therefore transfers a part of non-convexity, which results from the large traveltime discrepancies, to the registration problem at the trace level.

Statement of the optimization problem

In order to find $A(t)$ and $p(t)$ we propose to solve the following least-squares minimization problem for each trace: find $p(t)$ and $A(t)$ piecewise cubics that minimize

$$W[p, A] = \frac{1}{2} \int |d(t) - A(t)u(p(t))|^2 dt + \frac{\lambda}{2} \int |p(t) - t|^2 dt, \quad (8)$$

where λ is a weighting parameter for a regularization term which enforces $p(t)$ to stay as an one-to-one mapping. The letter W stands for “warping”.

This registration problem is non-convex and suffers from the same cycle-skipping phenomenon as conventional least-squares FWI does, due to the oscillatory nature of the predicted and observed data. Simulated annealing (Kirkpatrick, 1984) or other Monte Carlo methods (Wenzel and Hamacher, 1999) for global optimization could be performed but they are not tried in this paper. Instead, the minimization is carried out in a multiscale fashion by restricting the data $d(t)$ and its prediction $A(t)u(p(t))$ to a slowly growing subset of frequencies, from the zero frequency to successively higher frequencies, as in frequency domain FWI (Plessix, 2009). However, the observed data d usually have small energy in the low fre-

quency band and may be corrupted by noise. In order to start the sweep at zero frequency, we use modified traces $D(t)$ and $U(t)$, manufactured to contain low frequencies, instead of $d(t)$ and $u(t)$. We call $D(t)$ and $U(t)$ *low-frequency augmented* (LFA) signals. Hence, the optimization problem 8 becomes: find $p(t)$ and $A(t)$ piecewise cubics that minimize

$$W_{\text{LFA}}[p, A] = \frac{1}{2} \int |D(t) - A(t)U(p(t))|^2 dt + \frac{\lambda}{2} \int |p(t) - t|^2 dt. \quad (9)$$

We point out that registration with the observed and predicted data $u(t), d(t)$ instead of their LFA versions $U(t), D(t)$ fails to recover the correct time shifts. Frequency sweeping from low frequencies around 0 Hz using appropriate LFA transformations seems to be crucial for successful registration and inversion. Here are four reasonable possibilities for defining an LFA signal, $U(t)$, from a band-limited signal $u(t)$:

$$U_h = u(t) + |u(t) + i(\mathcal{H}u)(t)|, \quad (10a)$$

$$U_s = u^2(t), \quad (10b)$$

$$U_a = |u(t)|, \quad (10c)$$

where i is $\sqrt{-1}$ and \mathcal{H} is the Hilbert transform, defined in the frequency domain as $\widehat{\mathcal{H}u}(\omega) = -i \operatorname{sgn}(\omega)\widehat{u}(\omega)$ (Benedetto, 1997), where $\operatorname{sgn}(\cdot)$ is the sign function.

The Hilbert transform completes any real signal with an imaginary part, so that $u + i\mathcal{H}u$ is an “analytic” signal in the sense of having no negative frequency component. The amplitude $\sqrt{u^2(t) + (\mathcal{H}u(t))^2}$ has the interpretation of an envelope for the signal $u + i\mathcal{H}u$. The Hilbert transform is a classical tool in signal processing; it is typically used for demodulation in seismic inversion (Bozdağ et al., 2011).

All three LFA transformations generate strong low-frequency signals as shown in Figure 2, enabling the frequency sweep from zero frequency. Our comparison of three LFA trans-

formations shows that U_h is a particularly good LFA transformation in terms of frequency content, convergence of frequency sweeping, and registration errors.

The piecewise cubic polynomials $p(t)$ and $A(t)$ can be written as

$$p(t) = \sum_{k=1}^n \rho_k \phi_k(t) = [\phi_1(t) \ \phi_2(t) \dots \phi_n(t)] [\rho_1 \ \rho_2 \ \dots \ \rho_n]^T \quad (11)$$

and

$$A(t) = \sum_{k=1}^n \theta_k \phi_k(t) = [\phi_1(t) \ \phi_2(t) \dots \phi_n(t)] [\theta_1 \ \theta_2 \ \dots \ \theta_n]^T, \quad (12)$$

respectively. The superscript T denotes the transposed. Global basis functions with compact support in multiple subintervals are denoted by $\phi_k(t)$, $k = 1, 2, 3, \dots, n$, while both $p(t)$ and $A(t)$ are represented in each subinterval as a cubic. The errors in the approximation of smooth functions with the piecewise cubic polynomials are proportional to $O(h^4)$, while those of linear interpolations are $O(h^2)$, where h is the length of a subinterval. The number of global basis functions is proportional to the recording length of traces and the complexity of mapping functions. This paper does not address the problem where the spline nodes could also be determined by optimization. The column vectors $[\rho_1, \rho_2, \dots, \rho_n]^T$ and $[\theta_1, \theta_2, \dots, \theta_n]^T$ are denoted by $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$, respectively. Their components are the $2n$ parameters to be determined per trace.

The gradient and the Hessian matrix of W_{LFA} in problem 9 with respect to $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ can be found analytically, e.g.,

$$\frac{\partial W_{\text{LFA}}}{\partial \rho_i} = \int [D - AU(p)] [-AU'(p)\phi_i] + \lambda(p-t)\phi_i dt, \quad (13)$$

$$(\mathbf{H}_{\boldsymbol{\rho}})_{ij} = \frac{\partial^2 W_{\text{LFA}}}{\partial \rho_i \partial \rho_j} = \int \left[[AU'(p)]^2 - [D - AU(p)][AU''(p)] + \lambda \right] \phi_i \phi_j dt, \quad (14)$$

where $U'()$ and $U''()$ are the first and second order derivatives of U . Similar expressions can be derived for the gradient and Hessian with respect to the coefficients of $A(t)$. The integrals are computed approximately using the trapezoidal rule as a quadrature.

Optimization strategy and algorithm for seismogram registration

We detail a way to resolve the non-convexity issue of the optimization problem in expression 8 or 9 for seismogram registration in **Algorithm 1**.

The collection of discrete frequencies from 0 to ω_i is denoted by $\Omega_i = [0, \omega_i]$. We create M such sets, $\Omega_1, \Omega_2, \dots, \Omega_M$, where $\Omega_1 \subset \Omega_2 \subset \dots \subset \Omega_M$ and $\omega_1 < \omega_2 < \dots < \omega_M$. Below, $LPF_k(\cdot)$ denotes the application of a low-pass filter with passband $\Omega_k = [0, \omega_k]$. At the k^{th} outer iteration step, both LFA traces D and U are low-pass filtered to the frequency range $\omega \in \Omega_k$, resulting in the LF signals D_k and U_k . We let $W_{LFA,k}$ for the expression of W_{LFA} with D_k and U_k in place of D and U . The maximum frequency ω_M in the outer loop is set to a frequency below the central frequency of the source signature used to generate data.

Algorithm 1 Seismogram registration.

Input: traces $u(t)$ and $d(t)$

Initialize: $p(t) = t$, $A(t) = 1$

LFA: $D(t) \leftarrow LFA(d(t))$, $U(t) \leftarrow LFA(u(t))$

for $k = 1, 2, \dots, M$ **do**

Filter: $D_k(t) \leftarrow LPF_k(D(t))$, $U_k(t) \leftarrow LPF_k(U(t))$

while not converged **do**

Compute: $\frac{\partial W_{LFA,k}}{\partial \rho}$, $\frac{\partial W_{LFA,k}}{\partial \theta}$, and the Hessians \mathbf{H}_ρ , \mathbf{H}_θ of the functional $W_{LFA,k}$.

Newton step: $\rho \leftarrow \rho - \mathbf{H}_\rho^{-1} \frac{\partial W_{LFA,k}}{\partial \rho}$, $\theta \leftarrow \theta - \mathbf{H}_\theta^{-1} \frac{\partial W_{LFA,k}}{\partial \theta}$

end while

end for

Output: $p(t)$, $A(t)$

Computational aspect of registration and modification to LS inversion

Registering every trace can be time-consuming because the computational cost is proportional to the number of shots, of receivers, and of samples per trace. A more detailed cost analysis of seismogram registration is found below in the discussion section. The problem being highly parallelizable, traces in shots can be registered independently; our implementation of the registration step is parallelized over shots. As the grid size of a computational domain gets larger, the extra computational cost for the registration step gets smaller compared to that of forward/backward modeling steps. Moreover, the registration step can be sped up by tuning the parameters in the registration algorithm, exploiting data redundancy. For example, skipping registrations in some traces and interpolating piecewise polynomials can be done and the frequency sweeping range can be customized. Hence, the seismogram registration step in our implementation takes less time than a forward/backward modeling step.

The introduction of fractionally warped data for a modified adjoint source requires to solve a nonconvex optimization problem every iteration. It may give an impression that RGLS inversion including seismogram registration is unwieldy. However, Algorithm 2 is a minor modification of the conventional LS inversion algorithm from the point of view of complexity. The new steps for the RGLS method are marked with underlines; the steps are highly parallelizable and unexpensive compared to the forward/backward modeling steps. As a result, our implementation of the RGLS method takes slightly longer than the conventional LS method.

Algorithm 2 RGLS inversion algorithm.

Input: initial model m_0 and observed data $d(x_r, t)$

for $i = 0, 1, 2, \dots, N - 1$ **do**

Forward modeling: obtain $u(x, t; m_i)$ for the velocity model m_i

Seismogram registration: find $A(t), p(t)$ from $u(x_r, t; m_i)$ and $d(x_r, t)$
by minimizing the objective functional 9.

Fractionally warped data: $\tilde{d} \leftarrow [A(t)]^\alpha u((1 - \alpha)t + \alpha p(t))$

Backward modeling: obtain $q(x, t)$ by backpropagating $u(x_r, t; m_i) - \tilde{d}$

Imaging condition: $\delta m = - \int q(x, t) \frac{\partial^2 u}{\partial t^2}(x, t; m_i) dt$

Model updating: $m_{i+1} = m_i - \beta \delta m$

end for

Output: m_N

Examples of registration of synthetic traces

Here, we demonstrate the registration capability of non-convex optimization using the non-linear formulation 10a.

The underlying assumption of the registration idea is that two traces can be mapped via piecewise cubic polynomials. The mapping function $p(t)$ is also enforced to remain one-to-one thanks to the penalty term $\|p(t) - t\|^2$ in objective functional 8. If a true mapping does not satisfy these conditions, the solution is not guaranteed. However, the following numerical examples show that our registration method is robust and works well, even when some of the conditions are not met. Our first example demonstrates the capability of registration when the mapping is smooth and can be well-approximated by piecewise cubic polynomials. The second example shows that the frequency sweeping scheme makes

registration insensitive to random noise in the data. The third example is more challenging in that the waveforms are quite different and one of them is not a transported version of one trace contrary to our assumption.

Our first example shows the registration of two noiseless synthetic traces containing many reflected waves. One of the two traces is obtained from a numerical experiment with the Marmousi velocity model. The other trace is the result of applying a warping map $p : t \mapsto t + 0.15 \exp\left(-8(t/T_c - 1)^2\right)$, where T_c is half the recording time. Unless otherwise stated, registration is performed from zero frequency with the LFA signal U_h in equation 10a. Figure 3(b) shows a good registration match.

For the second example, we test seismogram registration with noisy synthetic data. A synthetic trace is obtained from a numerical experiment with the Marmousi model and is used as reference data. The trace is then transported using the same function as used in example 1 in order to generate predicted data. Two independent realizations of Gaussian white noise with mean 0. and standard deviation 0.05 are respectively added to the two traces. This noise level is about 35 % of the original traces in the root-mean-square sense (rms). Figure 4 shows excellent registration results in the presence of strong noise.

The third example uses two traces obtained from numerical experiments with two distinct velocity models. An observed trace is obtained from the Marmousi velocity model $V_{Marmousi}(x, z)$ while we use a different velocity model $V_{pred}(x, z) = V_{Marmousi}(x, z) - 0.15z$ for the predicted trace. Due to the reduction in velocity, the predicted data lag behind the observed data up to a few wave periods in the coda. A good, though not perfect agreement is observed between the observed trace and the transported trace as shown in Figure 5.

For the three numerical examples above, sweeping up to half of the central (dominant)

frequency suffices for convergence. Most kinematic discrepancies between the observed data and the predicted data are fixed after sweeping up to 5 Hz, where central frequencies of the traces are above 15 Hz. The signals in the above examples have a recording time of about 4 s with the sampling period about 1 ms, resulting in about 4000 samples per trace. For signals of such length, 4 subintervals are enough for the piecewise polynomials $p(t)$ and $A(t)$ in equations 11-12.

Although above three examples demonstrate robust registration with reasonable accuracy in the presence of strong random noise and in the case of different waveforms, there are cases where registration gives wrong results, i.e., shifting waveforms in the wrong direction. This scenario often happens when the number of waves differs in the two traces. Another case would be strong noise with energy in the sweeping frequency band.

NUMERICAL EXAMPLES OF FULL WAVEFORM INVERSION

In this section, we demonstrate the potential of registration-guided least-squares (RGLS) optimization for waveform inversion in transmission settings with synthetic velocity models. Specifically, we compare convergence of the RGLS method quantitatively with that of the LS method through examples 1 to 3 in Table 1. We point out that both RGLS and LS inversion are done in the time domain without a frequency sweeping. The frequency sweepings from zero frequency are performed in seismogram registration as explained in the previous section. The first two examples involve models with a Gaussian lens, while the third example involves a smoothed Marmousi model. We plot 1) true vs. converged velocity models, 2) data misfit vs. iteration count for both LS and RGLS, and 3) rms values of $V_T - V_k$ for some examples, where V_T is the true velocity model and V_k is the k^{th} step velocity model. By data misfit, we simply mean the least-squares misfit in expression 1.

The acoustic wave equation is discretized with a 4th order accuracy finite difference scheme in space. For the time discretization, the explicit 2nd order leap-frog scheme is used. Perfectly matched layers (PML) surround the computational domain (Berenger, 1994).

Inversion example 1: Gaussian lens

Example 1 compares the performance of RGLS and LS optimization using the velocity model V_H and V_L plotted in Figure 6(a,d). The initial models are homogeneous at $V_{init}(x, z) = 5100, 6000$ m/s, i.e., without any a priori knowledge about the true models. These true and initial velocity models are chosen to result in traveltimes discrepancies which are as large as 3.4 (4.5) wave periods in rays starting from the center of the left boundary to the opposite side, passing through the center of model $V_H(V_L)$, respectively.

The computational domain is 2500 m x 2500 m; the grid size is 501 by 501 with a distance of 5 m between grid points along both directions. The total number of shots used for computing the update is 196 with the distance between sources 50 m. For any source on one of four sides, we consider a fine sampling of 750 receivers with spacing 10 m on the other three sides as marked with magenta triangles in Figure 6(a). A Ricker wavelet with center frequency 50 Hz is used as an acoustic source.

High-velocity Gaussian lens: The RGLS method correctly updates the model velocity by increasing it near the center. The updates of the RGLS method (100 iterations) are free of artifacts and the velocity models look very close to the true models. See Figure 6(b). However, the LS method converges to a wrong model as shown in Figure 6(c). In particular, the velocity at the center of the converged model is around 3500 m/s, which is much lower than the initial velocity 5100 m/s and the true velocity 6200 m/s. However,

notice that the four corners seem to be updated properly. There is no cycle-skipping there: predicted data with short travel times are within a half of a period from corresponding observed data around the corners.

The changes in data misfit and model rms error over iteration are compared in Figures 7(a-b). The RGLS method temporarily increases the data misfit, but decreases the model rms error, as the predicted data correctly move towards the observed data during the first 20 to 30 iterations. For the LS method, however, the data misfit decreases but the model velocity error increases gradually, as shown in Figure 7(b).

Low-velocity Gaussian lens: As in the high-velocity Gaussian lens case, the RGLS recovers the true model much better than the LS method as shown in Figures 6(e-f). The LS method updates parts of the domain near the boundary correctly as well as the corners, as shown in Figure 6(f). Interestingly, a hundred more iterations result in a much better velocity model with large errors only at the center. The area of the error zone at the center gets smaller as the iterations proceed. Quantitatively, both data and model errors are reduced by both the RGLS and LS methods as shown Figures 7(c-d). However, RGLS is much faster; LS updates the model in a slow, piecemeal way from the boundary inwards.

Inversion example 2: Noisy Gaussian lens

We test the RGLS method with a more complicated true velocity model shown in Figure 8(a). The true velocity model in this example contains noise generated by convolution of a Gaussian kernel with an array of normally distributed random numbers as well as the high-velocity Gaussian lens. Other configurations are the same as in the high-velocity Gaussian lens of the previous example: the initial model, the source and receiver locations,

the acoustic source, and the grid size.

The medium-scale details of the model are successfully recovered by RGLS optimization as shown in Figure 8(b). The data misfit and model rms error are shown in Figure Figure 9. RGLS optimization stalls the data misfit after 63 iterations: the inversion then switches from RGLS to LS. Note that the LS method is close to the special case $\alpha = 1$ in the construction of fractionally warped data, hence the late-game switch to LS is more of a parameter adjustment than an ad-hoc fix. Switching to LS is safe because observed data are now within a fraction of a wavelength of predicted data. Using a velocity model with stronger randomness would make the RGLS method fail, because the observed data would contain many refracted waves that the predicted data do not contain.

Inversion example 3: Smoothed Marmousi

A more realistic velocity model is used to demonstrate the RGLS method: a smoothed Marmousi model. The physical dimension of the velocity model is 9096 m x 2976 m which is discretized into 380 x 125 grid points with spacing 24 m in both directions. The simulation consists of 75 shots with 120 m spacing between the sources; the positions of some sources are marked with white asterisks in Figure 10(b). The number of receivers is 616 receivers in total per shot with 24 m spacing and some of them are marked with magenta triangles in Figure 10(b); 376 receivers at $z = 2952$ m and 120 receivers at $x = 48$ and 9072 m each. The data are sampled at the receivers for 6 s with a time step size 1 ms. The Ricker wavelet with peak frequency 10 Hz is used as a source. The initial model has linearly increasing velocity from 1500 m/s near the surface to 3000 m/s at the bottom. A smoothed Marmousi model shown in Figure 10(a) is used to generate observed data. Initial traveltimes discrepancies at

the bottom receivers are around 2.7-5.3 wave periods.

A converged velocity model using the RGLS method is plotted in Figure 10(c), showing recovered features of the true model. As inversion switches from RGLS to LS in the previous example 2, LS inversion followed 96 RGLS iterations. For the RGLS method, both the model rms error and data misfit decrease by an order and two orders of magnitude, respectively. A hump is also observed in the data misfit plot of the RGLS method as in the previous examples. The LS method, however, could not recover the background velocity model, not to mention most reflectors of the true model as shown in Figure 10(d).

DISCUSSION

Our registration method is similar to Hale’s dynamic warping with strain constraints (Hale, 2013; Ma and Hale, 2013) in that both solve the least-squares misfit nonconvex optimization problem for the warpings. Both methods assume that one of two corresponding traces can be mapped with a smooth warping function. One advantage of our method over Hale’s is that it handles amplitude changes as well as time shifts. A second advantage is that thanks to the frequency sweeping, smoothing traces or shifts is not necessary in the presence of random noise. Our method, however, seems to be more expensive due to the frequency sweeping and the Hilbert transform. The computational complexity of the seismogram registration per trace is $O(n_k(N_t \log N_t + n_i n^2 N_t))$ for the number of samples N_t , where n_k , n , and n_i are the number of frequency sweeping steps, global basis functions and Newton iterations, respectively. The number of frequency sweeping steps is $n_k = \frac{f_p T_f}{m}$, where f_p , T_f , m are the peak frequency, final recording time, and increment in frequency index. Since most cases have a larger $n_i n^2$ than $\log N_t$, the complexity per shot can be simplified further to $O(n_k n_i n^2 n_r N_t)$, where n_r is the number of receivers. For a fixed number

of receivers, n_r , the computational cost for forward/backward modeling with an explicit finite difference method, $O(N_x N_z N_t)$, gets more expensive than that of the registration, $O(n_k n_i n^2 n_r N_t)$, as the grid size, $N_x N_z$, gets larger. Although our warping algorithm has not yet been extended to 2D or 3D sections of the dataset, it is a reasonable direction for future research.

We compare our RGLS method with three related methods: correlation-based traveltime tomography (TT) (Luo and Schuster, 1991; Tromp et al., 2005), waveform inversion (WI) (Tarantola and Valette, 1982; Bunks et al., 1995), phase/envelope (PE) misfit inversion (Fichtner et al., 2008; Bozdağ et al., 2011). For a more complete treatment of each method, we refer the reader to these references. Comprehensive comparisons among three methods, i.e. TT, WI, and PE, are well documented in Bozdağ et al. (2011). The RGLS method does not need phase isolation as in the TT method since the whole waveform is used to implicitly determine traveltime discrepancies or phase difference. Thanks to the frequency sweeping in seismogram registration, traveltime differences larger than a half period can be recovered and cycle-skipping can be overcome as in TT. WI and PE do not need phase isolation either, but suffer from cycle-skipping problems which can be avoided by inversion of long-period waves first, followed by short-period waves. In our study, successful registration of data with Gaussian random noise is demonstrated and we expect that the updates are only moderately affected by noise in the data because the adjoint sources, $u - \tilde{d}$, are made of only synthetic noise-free data, u and \tilde{d} .

A drawback of our RGLS method is that observed and synthetic data must be comparable, i.e., can be paired; this assumption can be broken even in both the transmission and the reflection settings. This assumption is shared by both TT and PE, but not by WI. We found that adapting registration ideas to waveform inversion in the reflection setting is

particularly challenging. Modeled data are closer to matching observed data kinematically in the reflection case than in the transmission case, because the LS gradient updates produce ad-hoc reflectors in the wrong locations to balance the wrong medium. As a result, traveltimes discrepancies no longer seem to be the dominant effect to correct. We attempted to use the alternating update methods proposed by Clément et al. (2001) and Xu et al. (2012) to deal with reflection data, but without much success.

It seems natural to extend the idea of LFA to waveform inversion:

$$J_{LFA}[m] = \frac{1}{2} \sum_{s,r} \int |S_{s,r}LFA\{u_s\}(x, t) - LFA\{d_s\}(x_r, t)|^2 dt. \quad (15)$$

Since $LFA\{u_s\}(x, t)$ has low frequency components near zero, the frequency sweeping can be done with the following equivalent form in the frequency domain:

$$J_{LFA}[m] = \frac{1}{2} \sum_{s,r} \int_0^{\omega'} |S_{s,r}LFA\{u_s\}(x, \omega) - LFA\{d_s\}(x_r, \omega)|^2 d\omega, \quad (16)$$

where ω' can be much smaller than frequencies available at the data d_s . A study of this new objective function is under way.

Lastly, we point out that the “fractional” character of the warping used to generate the adjoint source in this paper is in the same spirit as a solution proposed in Sava and Biondi (2004), where image perturbations in the image domain are replaced by their linearized version to mitigate lack-of-convexity issues beyond the Born approximation. Similar suggestions of updates are also proposed in recent work by Fei and Williamson (2010) and Albertin (2011). Those updates are generated from residuals, which are obtained by taking differences between an image and its infinitesimally modified version for the same reason we take the modified adjoint source instead of the conventional one.

CONCLUSIONS

We present registration guided least-squares (RGLS) as a way to mitigate the cycle-skipping problem in FWI, thereby extending the basin of attraction to the global minimizer. The successful application of the RGLS method to seismic inversion problems is demonstrated in the transmission setting, where the conventional LS method often converges to a wrong model. The proposed method substitutes a transported version of the prediction, referred to as fractionally warped data, for the observed data in the conventional least-squares misfit residual. In order to generate transported data, mappings in the form of piecewise polynomials are found through a non-convex optimization formulation. The non-convex optimization problem is tackled in a multiscale manner similar to frequency sweep/continuation in frequency domain FWI. In order to create the low frequencies which may be absent in data, low-frequency augmented (LFA) signals are proposed and demonstrated to provide a satisfying alternative to the raw seismograms for the registration step. A method using the envelope property of the Hilbert transform is proposed for this LFA transformation. Three inversion examples using seismogram registration and the RGLS method show that the proposed method decreases model errors monotonically while it allows the data misfit to increase temporarily prior to eventual convergence.

ACKNOWLEDGMENTS

The authors are grateful to the authors of Madagascar; and to TOTAL S.A. for support.

REFERENCES

- Albertin, U., 2011, An improved gradient computation for adjoint wave-equation reflection tomography: 81st Annual International Meeting, SEG, Expanded Abstracts, 3969–3973.
- Baek, H., and L. Demanet, 2013, The failure mode of correlation focusing for model velocity estimation: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4704–4708.
- Benedetto, J. J., 1997, Harmonic analysis and applications: CRC-Press.
- Berenger, J.-P., 1994, A perfectly matched layer for the absorption of electromagnetic waves: *Journal of Computational Physics*, **114**, 185 – 200.
- Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: *Geophysical Journal International*, **185**, 845–870.
- Bregman, N. D., R. C. Bailey, and C. H. Chapman, 1989, Crosshole seismic tomography: *GEOPHYSICS*, **54**, 200–215.
- Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *GEOPHYSICS*, **60**, 1457–1473.
- Clément, F., G. Chavent, and S. Gómez, 2001, Migration-based traveltime waveform inversion of 2-d simple structures: A synthetic example: *GEOPHYSICS*, **66**, 845–860.
- Fei, W., and P. Williamson, 2010, On the gradient artifacts in migration velocity analysis based on differential semblance optimization: 80th Annual International Meeting, SEG, Expanded Abstracts, 4071–4076.
- Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge, 2008, Theoretical background for continental- and global-scale full-waveform inversion in the timefrequency domain: *Geophysical Journal International*, **175**, 665–685.
- Fomel, S., and L. Jin, 2009, Time-lapse image registration using the local similarity at-

- tribute: *GEOPHYSICS*, **74**, no. 2, A7–A11.
- Fomel, S., and M. van der Baan, 2010, Local similarity with the envelope as a seismic phase detector: 80th Annual International Meeting, SEG, Expanded Abstracts, 1555–1559.
- Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: *GEOPHYSICS*, **51**, 1387–1403.
- Gee, L. S., and T. H. Jordan, 1992, Generalized seismological data functionals: *Geophysical Journal International*, **111**, 363–390.
- Glasbey, C. A., and K. V. Mardia, 1988, A review of image-warping methods and improved gradient computation for adjoint wave equation reflection tomography: *Journal of Applied Statistics*, **25**, 155–171.
- Haker, S., and A. Tannenbaum, 2001, Optimal mass transport and image registration: *Proceedings IEEE Workshop on Variational and Level Set Methods in Computer Vision*, 29–36.
- Hale, D., 2013, Dynamic warping of seismic images: *GEOPHYSICS*, **78**, no. 2, S105–S115.
- Kennett, B. L. N., and A. Fichtner, 2012, A unified concept for comparison of seismograms using transfer functions: *Geophysical Journal International*, **191**, 1403–1416.
- Kirkpatrick, S., 1984, Optimization by simulated annealing: *Quantitative studies: Journal of Statistical Physics*, **34**, 975–986.
- Liner, C. L., and R. G. Clapp, 2004, Nonlinear pairwise alignment of seismic traces: *The Leading Edge*, **23**, 1146–1150.
- Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: *GEOPHYSICS*, **56**, 645–653.
- Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: *GEOPHYSICS*, **78**, no. 6, R223–R233.

- Maggi, A., C. Tape, M. Chen, D. Chao, and J. Tromp, 2009, An automated time-window selection algorithm for seismic tomography: *Geophysical Journal International*, **178**, 257–281.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503.
- , 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: *GEOPHYSICS*, **74**, no. 6, WCC149–WCC157.
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: *GEOPHYSICS*, **64**, 888–901.
- Pratt, R. G., and N. R. Goult, 1991, Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data: *GEOPHYSICS*, **56**, 208–224.
- Prieux, V., G. Lambaré, S. Operto, and J. Virieux, 2013, Building starting models for full waveform inversion from wide-aperture data by stereotomography: *Geophysical Prospecting*, 109–137.
- Ramsay, J. O., and X. Li, 1998, Curve registration: *Journal of the Royal Statistical Society Series B*, **60**, 351–363.
- Sakoe, H., and S. Chiba, 1978, Dynamic programming algorithm optimization for spoken word recognition: *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 43–49.
- Sava, P., and B. Biondi, 2004, Wave-equation migration velocity analysis. I. theory: *Geophysical Prospecting*, **52**, 593–606.
- Shah, N., M. Warner, L. Guasch, I. Stekl, and A. Umpleby, 2010, Waveform inversion of surface seismic data without the need for low frequencies: 80th Annual International

- Meeting, SEG, Expanded Abstracts, 2865–2869.
- Shin, C., and Y. H. Cha, 2009, Waveform inversion in the laplacefourier domain: *Geophysical Journal International*, **177**, 1067–1079.
- Shin, C., and W. Ha, 2008, A comparison between the behavior of objective functions for waveform inversion in the frequency and laplace domains: *GEOPHYSICS*, **73**, no. 5, VE119–VE133.
- Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: *Reviews of Geophysics*, **20**, 219–232.
- Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: *Geophysical Journal International*, **160**, 195–216.
- Van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation travelttime tomography: *Geophysical Journal International*, **182**, 1383–1394.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *GEOPHYSICS*, **74**, no. 6, WCC1–WCC26.
- Wenzel, W., and K. Hamacher, 1999, Stochastic tunneling approach for global minimization of complex potential energy landscapes: *Physical Review Letters*, **82**, no. 15, 3003–3007.
- Xu, S., D. Wang, F. Chen, G. Lambaré, and Y. Zhang, 2012, Inversion on reflected seismic wave: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–7.
- Zitová, B., and J. Flusser, 2003, Image registration methods: a survey: *Image and Vision Computing*, **21**, 977–1000.

LIST OF TABLES

1 Velocity models and data types for inversion examples. Reference velocity models V_H , V_L and V_R are $V_H(x, z) = 5200 + 900 \exp(-|(x, z) - (1250, 1250)|^2/10^6)$ and $V_L(x, z) = 5500 - 900 \exp(-|(x, z) - (1250, 1250)|^2/10^6)$, and $V_R(x, z) = 5000 + 900 \exp(-|(x, z) - (1250, 1250)|^2/10^6)$, respectively.

LIST OF FIGURES

1 Replacement of observed data with fractionally warped data. (a) Fractionally warped trace which is a mapped (transported slightly) version of the given predicted data towards the observed data. We call the new traces *fractionally warped data*. (b) Comparison of observed and predicted traces. (c) Comparison of fractionally warped and predicted traces.

2 Spectra of signals transformed via LFA transformations 10(a-c). (a) $u(t)$ and its envelope $U_e(t) = |u(t) + i\mathcal{H}(u)(t)|$, (b) Comparison of a wavelet $u(t)$ and its LFA transformation $U_h(t)$ obtained using the Hilbert transform in equation 10a, (c) Comparison of spectra of the wavelet $u(t)$ and of the envelope signal $U_e(t)$, (d) Comparison of spectra of the wavelet $u(t)$ and of the LFA signal $U_h(t)$, (e) Comparison of spectra of the wavelet $u(t)$ and of the LFA signal $U_s(t) = u^2(t)$, (f) Comparison of spectra of the wavelet $u(t)$ and of the LFA signal $U_a(t) = |u(t)|$. The blue solid (red dashed) lines in (c)-(f) respectively correspond to spectra of the wavelet $u(t)$ (of the LFA signals).

3 Registration example 1: (a) Observed (blue) and predicted (red) traces before registration. A synthetic trace is generated using the Marmousi velocity model and the other trace is created by warping with a known mapping. The black arrows indicate corresponding waves. (b) Two traces after registration. The predicted trace (red) is mapped to the observed trace (blue).

4 Registration example 2: matching noisy synthetic traces. (a) Observed (blue) and predicted (red) traces before mapping. A synthetic trace is generated using the Marmousi velocity model and the other trace is created by applying a known mapping. The black arrows in the top figure connect corresponding peaks. (b) Two traces after the predicted trace (red) is mapped to the observed trace (blue).

5 Registration example 3: matching traces with different amplitudes and phases from the Marmousi model and a modified Marmousi model. (a) Observed (blue) and predicted (red) trace. The black arrows indicate corresponding waves. (b) Two traces after registration. The predicted (red) trace is transported towards the observed (blue) trace and its amplitude is decreased.

6 Inversion example 1: plots of velocity models: (a) true model with sources and receivers marked in white and maganta, respectively, (b) converged model of RGLS optimization, (c) converged model of LS optimization. The white asterisks and magenta triangles in (a) mark the location of a few sources and receivers, respectively. Inversion example 2: plots of velocity models: (d) true model, (e) converged model of RGLS optimization, (f) converged model of LS optimization.

7 Inversion example 1: convergence of model (velocity) rms error $V_k - V_{true}$ (a) and data misfit J (b). Inversion example 2: convergence of model (velocity) rms error $V_k - V_{true}$ (c) and data misfit J (d).

8 Inversion example 2: Plots of velocity models of (a) true model, (b) of converged model of RGLS optimization, and (c) of converged model of LS optimization.

9 Inversion example 2: convergence of model (velocity) rms error $V_k - V_{true}$ (a) and data misfit J (b).

10 Inversion example 3: Plots of velocity models of (a) true model,(b) of initial model with sources and receivers marked in white and magenta, (c) of converged model of RGLS optimization, and (d) of converged model of LS optimization. The white asterisks and magenta inverted triangles in (b) indicate sources on the top and receivers at three sides, respectively.

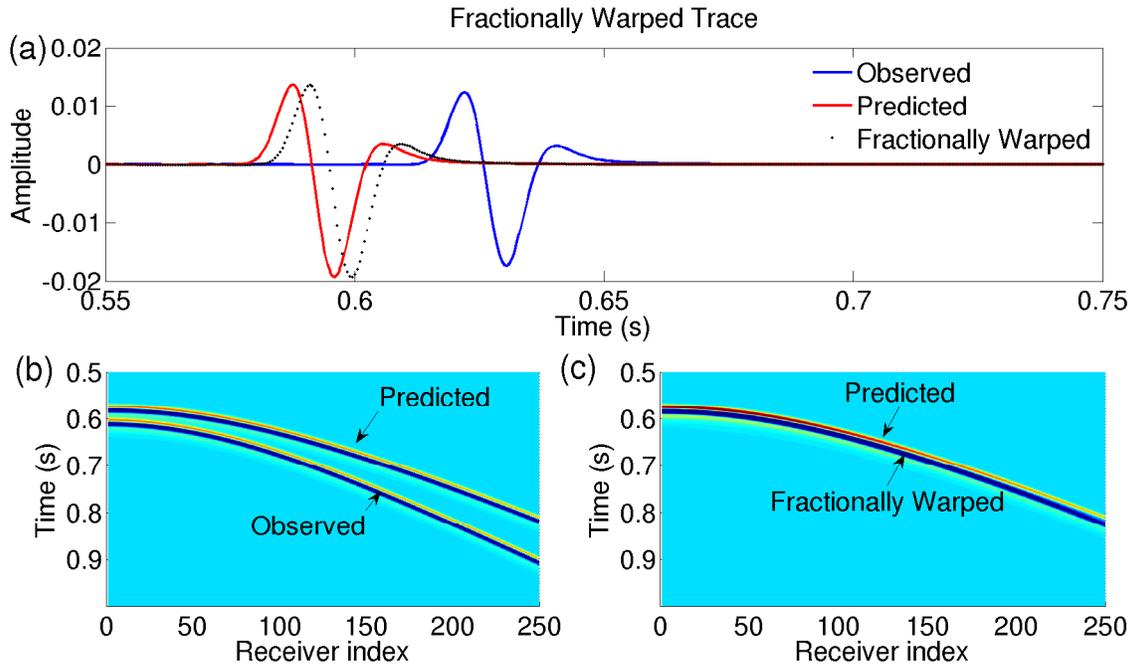


Figure 1: Replacement of observed data with fractionally warped data. (a) Fractionally warped trace which is a mapped (transported slightly) version of the given predicted data towards the observed data. We call the new traces *fractionally warped data*. (b) Comparison of observed and predicted traces. (c) Comparison of fractionally warped and predicted traces.

Baek, Calandra, & Demanet –

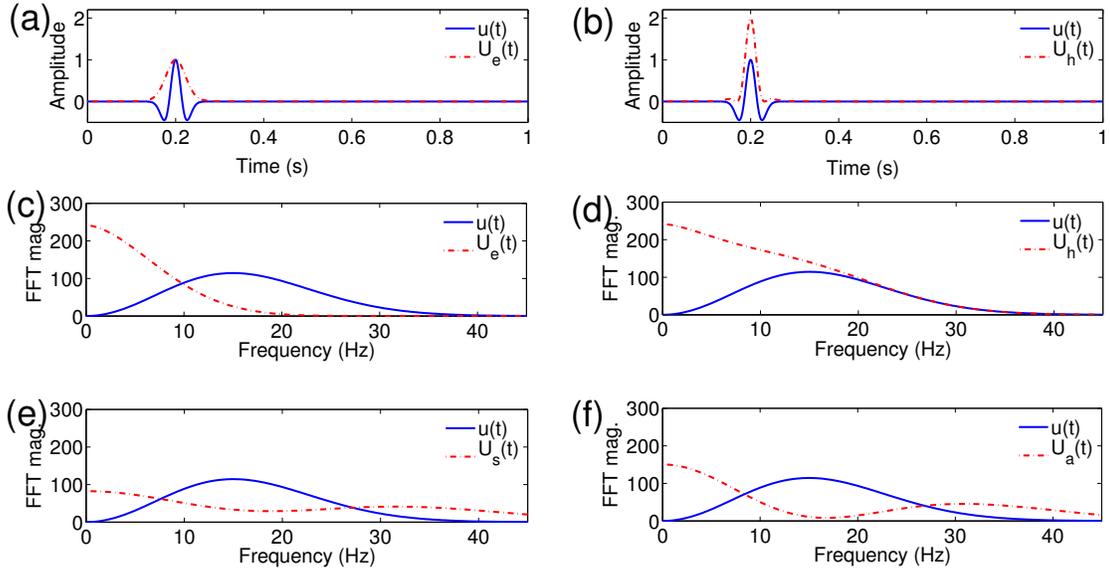


Figure 2: Spectra of signals transformed via LFA transformations 10(a-c). (a) $u(t)$ and its envelope $U_e(t) = |u(t) + i\mathcal{H}(u)(t)|$, (b) Comparison of a wavelet $u(t)$ and its LFA transformation $U_h(t)$ obtained using the Hilbert transform in equation 10a, (c) Comparison of spectra of the wavelet $u(t)$ and of the envelope signal $U_e(t)$, (d) Comparison of spectra of the wavelet $u(t)$ and of the LFA signal $U_h(t)$, (e) Comparison of spectra of the wavelet $u(t)$ and of the LFA signal $U_s(t) = u^2(t)$, (f) Comparison of spectra of the wavelet $u(t)$ and of the LFA signal $U_a(t) = |u(t)|$. The blue solid (red dashed) lines in (c)-(f) respectively correspond to spectra of the wavelet $u(t)$ (of the LFA signals).

Baek, Calandra, & Demanet –

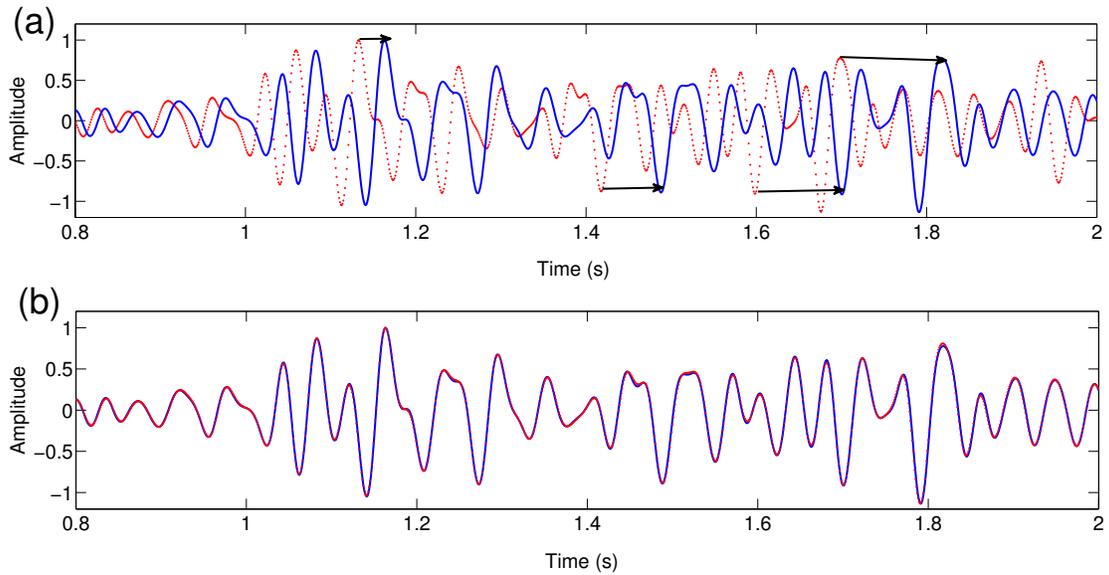


Figure 3: Registration example 1: (a) Observed (blue) and predicted (red) traces before registration. A synthetic trace is generated using the Marmousi velocity model and the other trace is created by warping with a known mapping. The black arrows indicate corresponding waves. (b) Two traces after registration. The predicted trace (red) is mapped to the observed trace (blue).

Baek, Calandra, & Demanet –

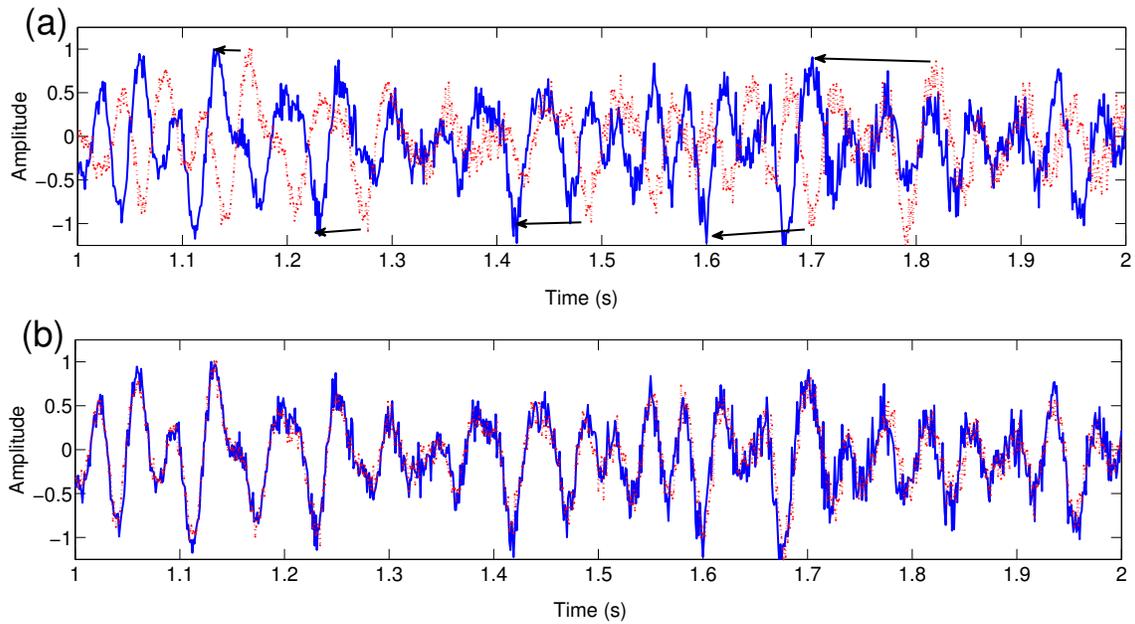


Figure 4: Registration example 2: matching noisy synthetic traces. (a) Observed (blue) and predicted (red) traces before mapping. A synthetic trace is generated using the Marmousi velocity model and the other trace is created by applying a known mapping. The black arrows in the top figure connect corresponding peaks. (b) Two traces after the predicted trace (red) is mapped to the observed trace (blue).

Baek, Calandra, & Demanet –

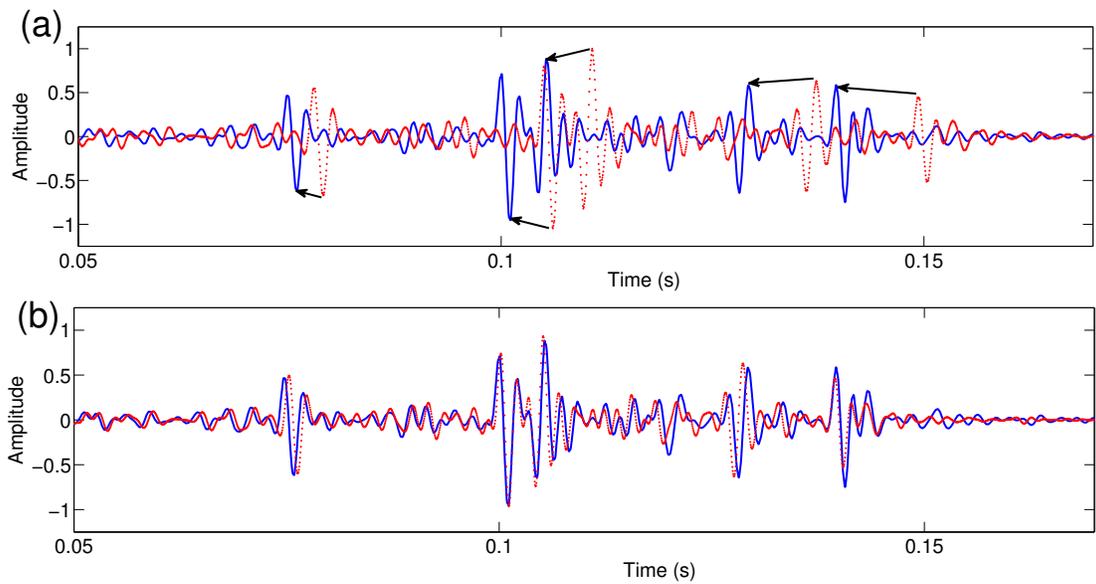


Figure 5: Registration example 3: matching traces with different amplitudes and phases from the Marmousi model and a modified Marmousi model. (a) Observed (blue) and predicted (red) trace. The black arrows indicate corresponding waves. (b) Two traces after registration. The predicted (red) trace is transported towards the observed (blue) trace and its amplitude is decreased.

Baek, Calandra, & Demanet –

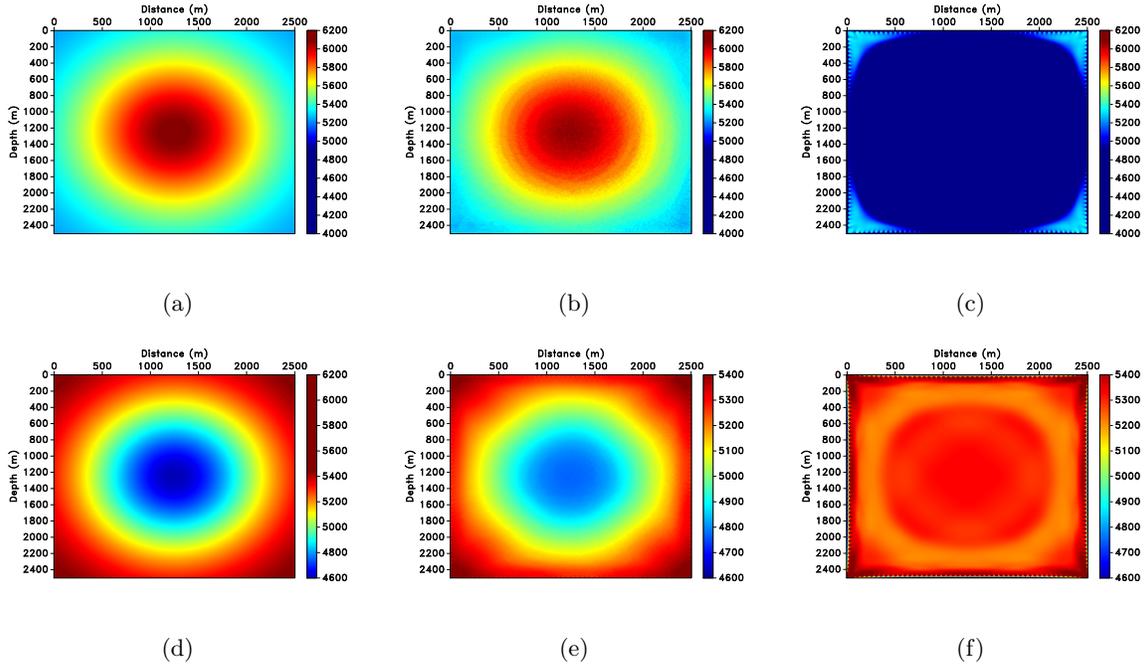
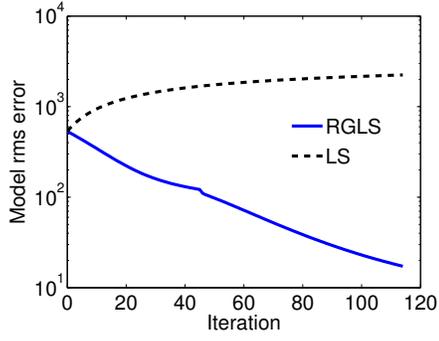
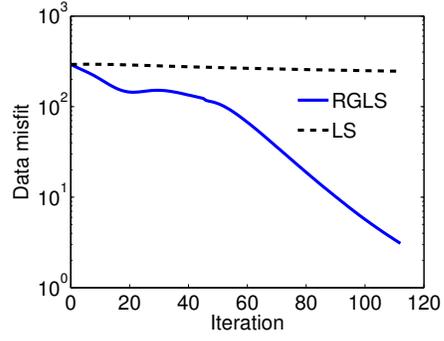


Figure 6: Inversion example 1: plots of velocity models: (a) true model with sources and receivers marked in white and magenta, respectively, (b) converged model of RGLS optimization, (c) converged model of LS optimization. The white asterisks and magenta triangles in (a) mark the location of a few sources and receivers, respectively. Inversion example 2: plots of velocity models: (d) true model, (e) converged model of RGLS optimization, (f) converged model of LS optimization.

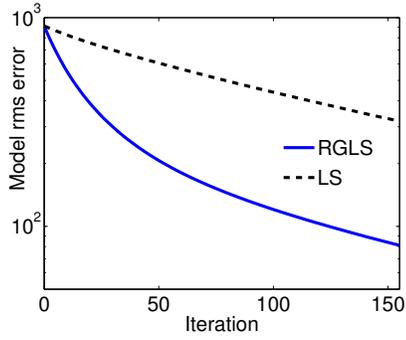
Baek, Calandra, & Demanet –



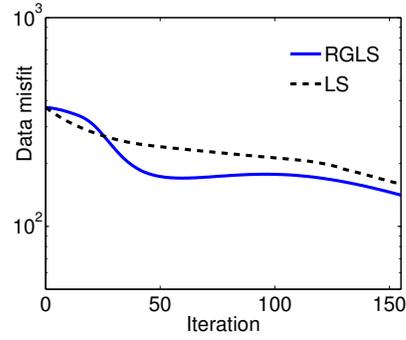
(a)



(b)



(c)



(d)

Figure 7: Inversion example 1: convergence of model (velocity) rms error $V_k - V_{true}$ (a) and data misfit J (b). Inversion example 2: convergence of model (velocity) rms error $V_k - V_{true}$ (c) and data misfit J (d).

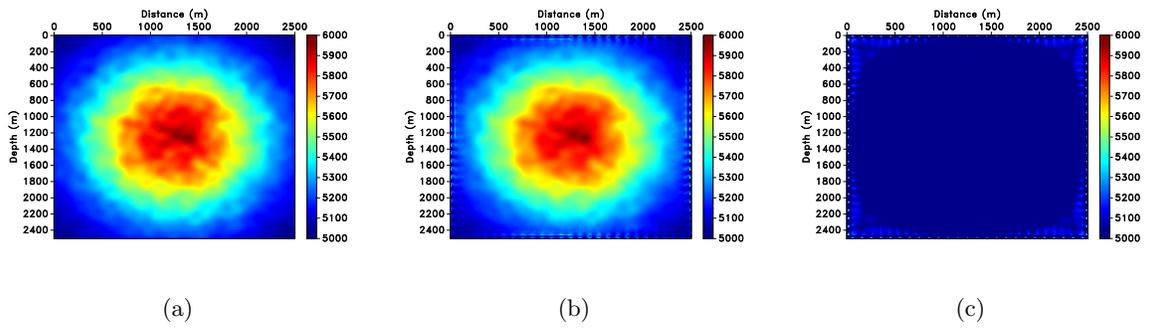
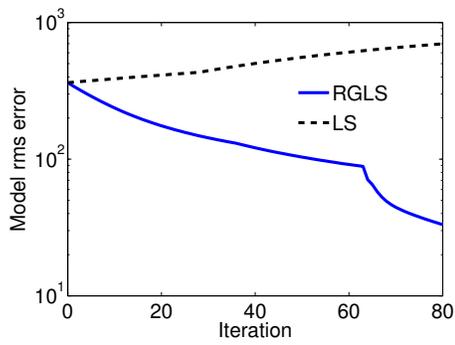
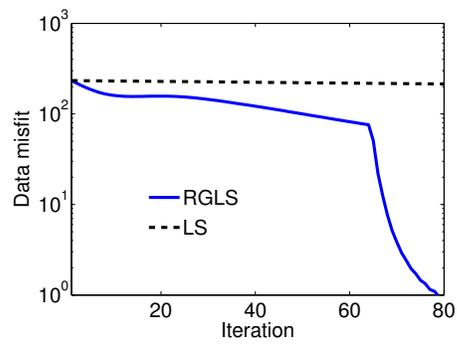


Figure 8: Inversion example 2: Plots of velocity models of (a) true model, (b) of converged model of RGLS optimization, and (c) of converged model of LS optimization.

Baek, Calandra, & Demanet –



(a)



(b)

Figure 9: Inversion example 2: convergence of model (velocity) rms error $V_k - V_{true}$ (a) and data misfit J (b).

Baek, Calandra, & Demanet –

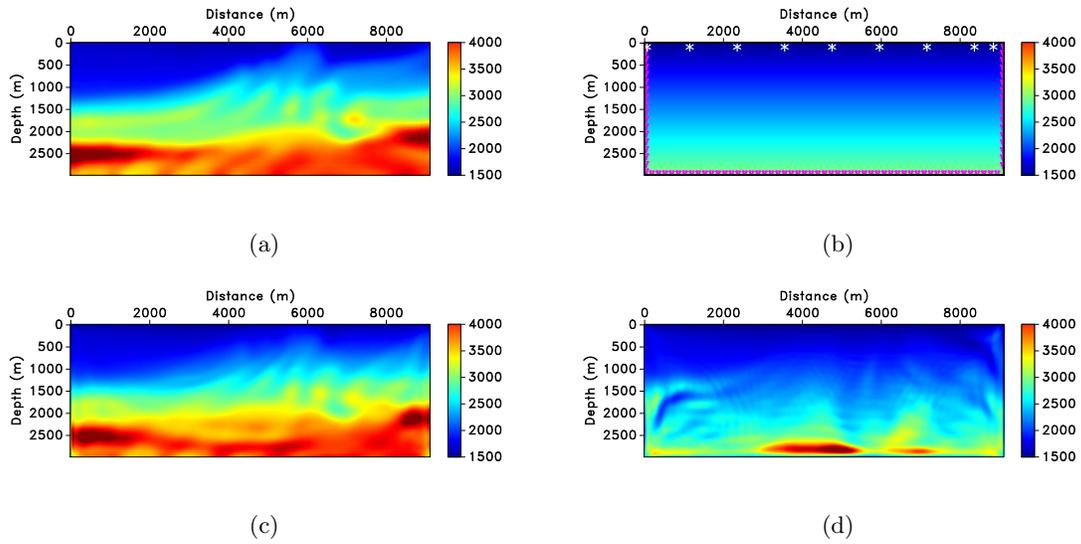


Figure 10: Inversion example 3: Plots of velocity models of (a) true model, (b) of initial model with sources and receivers marked in white and magenta, (c) of converged model of RGLS optimization, and (d) of converged model of LS optimization. The white asterisks and magenta inverted triangles in (b) indicate sources on the top and receivers at three sides, respectively.

Baek, Calandra, & Demanet –

Example	Reference model	Initial model
Example 1	V_H	5100 m/s
	V_L	6000 m/s
Example 2	$V_R + \text{Noise}$	5100 m/s
Example 3	Marmousi	$V_{init} = 1500 + 0.5z$ m/s

Table 1: Velocity models and data types for inversion examples. Reference velocity models V_H , V_L and V_R are $V_H(x, z) = 5200 + 900 \exp(-|(x, z) - (1250, 1250)|^2/10^6)$ and $V_L(x, z) = 5500 - 900 \exp(-|(x, z) - (1250, 1250)|^2/10^6)$, and $V_R(x, z) = 5000 + 900 \exp(-|(x, z) - (1250, 1250)|^2/10^6)$, respectively.