

# Registration-guided least-squares waveform inversion

Hyongsu Baek<sup>1</sup>, Henri Calandra<sup>2</sup>, Laurent Demanet<sup>1</sup>  
<sup>1</sup> MIT Mathematics department, <sup>2</sup> TOTAL S.A.

January 15 2013

## Abstract

Full waveform inversion with frequency sweeping cannot start from zero frequency because of the lack of low-frequency data, requiring a good starting model. We study a different iterative scheme where the notion of proximity of two traces is not the usual least-squares distance, but instead involves registration as in image processing. In order to create transported data, we introduce a nonconvex optimization problem and solve it in a multiscale fashion from low to high frequencies. This process requires defining low-frequency augmented signals in order to seed the frequency sweep at zero frequency. Successful registrations of noisy data, and application of the new method to model velocity estimation are demonstrated. In a crosshole seismic inversion example (transmission setting), we show that the new method decreases the model velocity error while conventional least-squares inversion converges to a spurious model.

## Introduction

Full waveform inversion (FWI) is traditionally based on minimizing a least-squares (LS) misfit criterion [?], but this choice generates spurious local minima or valleys that owe to the oscillatory nature of the data. This problem that high-frequency data depend in a very nonlinear way on the low-wavenumber components of the model velocity is called cycle-skipping. Frequency domain FWI addresses the resulting lack of convexity by performing the inversion in a stagewise fashion from low to high frequencies. However, the lack of low-frequency data in exploration seismology often hinders this frequency sweep/continuation approach. To restore the ability of high-frequency data to invert the smooth model velocity, we propose to replace the usual notion of data residual ( $d - u$ , with  $d$  the observed trace and  $u$  the predicted trace) by a more faithful, transport-based measure of proximity between two traces.

In order to generate transported data to better guide the optimization iterations, we propose to fit a parametrized warping between  $d$  and  $u$  by solving a nonconvex optimization similar to image registration. Matching (registration) between observed data and predicted data or between time-lapse images has been used in many different applications under different names: phase-shift [?], travelttime delay based on correlation [?], and alignment of seismic traces using dynamic programming for trace processing and interpretation [?]. The mappings we consider in this paper are piecewise polynomial, and seem to be a novel tool for model velocity estimation. We demonstrate on simple numerical experiments in the transmission setting that our method can avoid the cycle-skipping phenomenon, extending the basin of attraction of least-squares minimization.

## Method

FWI normally uses the least-squares misfit

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

where  $m(x)$  denotes a velocity model,  $u_s = \mathcal{F}[m]$ , and  $d_s$  are predicted and observed data at a shot  $s$  and at a receiver  $x_r$ , respectively. For notational simplicity, the subscripts  $s$  and  $r$  are omitted, whenever it does not cause confusion. The forward operator  $\mathcal{F}[m]$  relates a velocity model  $m$  to data  $u_s(x_r, t)$  through the acoustic wave equation,  $m\partial^2 u_s / \partial t^2 = \Delta u_s + f_s(x, t)$ , where  $f_s(x, t)$  is a source term. The adjoint-state method generates the gradient of  $J[m]$  via the imaging condition

$$\frac{\delta J}{\delta m} = \int q_s(x, t) \partial^2 u_s / \partial t^2(x, t) dt, \quad (2)$$

where the adjoint field  $q_s$  solves the adjoint wave equation with the data residual  $d_s - u_s$  in the right-hand side.

It is well known that  $J[m]$  is non-convex, particularly when the data are oscillatory. More specifically, when the time difference between corresponding arrivals in  $u_s$  and  $d_s$  is larger than a half wavelength, the steepest descent direction of the data misfit may result in increasing those time differences, consequently increasing the model error. In order to guide the optimization in the direction of the correct model update, we propose to change the residual  $d_s - u_s$  in the adjoint wave equation by replacing  $d_s$  by a version of  $u_s$  transported toward  $d_s$ . We denote these virtual, transported data by  $\tilde{d}_s$  and refer to them as *fractionally warped data*. Their construction is in the next section. The rationale behind  $\tilde{d}_s$  is that its arrivals can now be less than a quarter of a wavelength apart from those in  $u_s$ . The substitution of  $d_s$  by  $\tilde{d}_s$  is illustrated in Figure 1. We refer to our method as registration-guided least-squares (RGLS) method.

## Seismogram registration

In the sequel we will assume that the traces  $d(t)$  and  $u(t)$  are comparable waveforms, ideally with the same number of wave arrivals. We propose to find piecewise polynomials  $p(t)$  and  $A(t)$  so as to have a good match of the warped prediction with the observed data,  $d(t) \approx A(t)u(p(t))$ , then define fractionally warped data as

$$\tilde{d}(t) = A(t)u((1 - \alpha)t + \alpha p(t)), \quad (3)$$

for some small  $0 < \alpha < 1$ . The warping  $p(t)$ , the amplitude  $A(t)$ , and the value of  $\alpha$  are chosen so that fractionally warped data  $\tilde{d}(t)$  have a similar shape as the prediction  $u(t)$  but phase shifts less than a quarter of a wave period apart with respect to the predicted data  $u(t)$ . We use piecewise cubic Hermite polynomials for  $p(t)$  and  $A(t)$ .

In order to find  $A(t)$  and  $p(t)$  we propose to solve the following least-squares minimization problem for each trace, using frequency sweeps:

$$J_{warp}(p, A) = \frac{1}{2} \int |D(t) - A(t)U(p(t))|^2 dt + \frac{1}{2} \int |p(t) - t|^2 dt, \quad (4)$$

where  $D(t)$  and  $U(t)$  are what we call *low-frequency augmented* (LFA) signals for  $d(t)$  and  $u(t)$ , respectively. This minimization problem can be viewed as a form of seismogram registration. Due

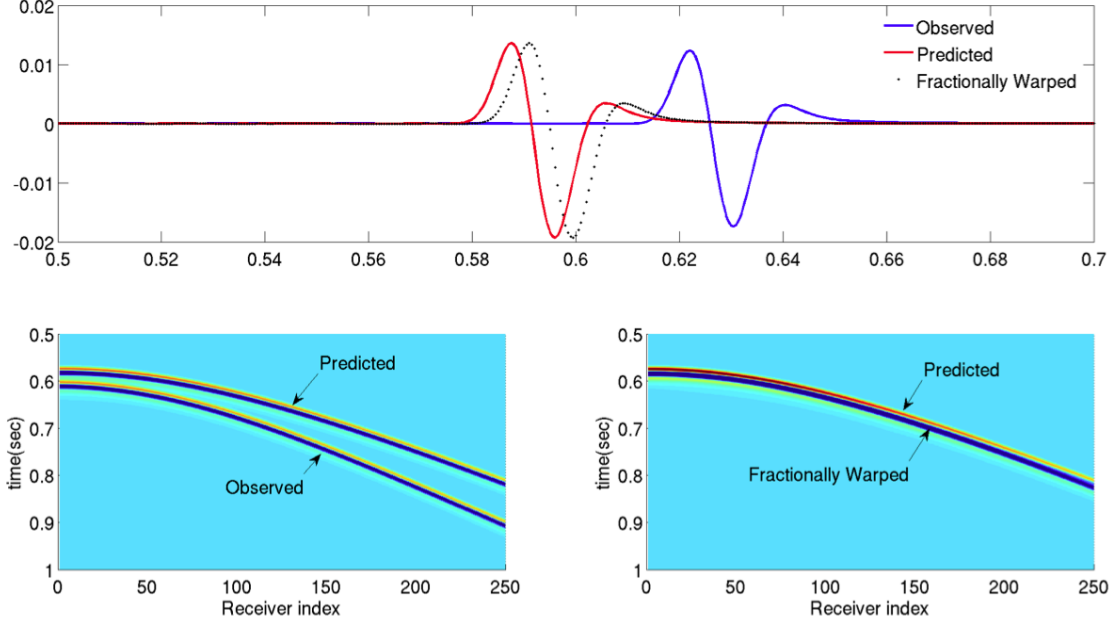


Figure 1: Replacement of observed data with fractionally warped data, which are a mapped (transported slightly) version of the given predicted data towards the observed data.

to the oscillatory nature of the predicted and observed data,  $J_{warp}(p, A)$  is non-convex and its minimization would suffer from the same cycle-skipping phenomenon as in least-squares FWI. To remedy this problem, we solve the registration problem in a multiscale fashion from zero to higher frequencies. The reason for considering  $D(t)$  and  $U(t)$  rather than  $d(t)$  and  $u(t)$  is that the original traces do not contain frequency components around zero Hz, hence would not allow to properly seed the frequency continuation procedure. LFA signals are generated to overcome the lack of low frequencies in the original traces. We propose three nonlinear transformations to get LFA signals,  $U(t)$ , from the band-limited signal  $u(t)$ : (1)  $U_h = u(t) + |u(t) + i\mathcal{H}[u(t)]|$ ; (2)  $U_s = u^2(t)$ ; and (3)  $U_a = |u(t)|$ , where  $\mathcal{H}[u(t)]$  is the Hilbert transform of  $u(t)$ . Among the three proposed LFA signals, we found that  $U_h$  is the most appropriate in terms of frequency content, convergence of frequency sweeping, and registration errors. In the sequel, we only report numerical experiments involving  $U_h$ . An example of matching noisy traces by registration is shown in Figure 2.

## Waveform Inversion

The scalar acoustic wave equation is discretized with a 4<sup>th</sup> order accuracy finite difference scheme in space. For the time discretization, the explicit 2<sup>nd</sup> order leapfrog scheme is used. Perfectly matched layers (PML) surround the computational domain. The methods of updating the model velocity were explained earlier. The step size for the update is fixed for both RGLS and conventional LS optimization. For RGLS optimization, 100-130 iterations suffice to significantly reduce the model rms error in the example shown below. For the conventional LS method, model root-mean-square (rms) errors increases, and we stop the iterations after 100-130 steps as well. Seismogram registration is performed every iteration since the predicted data  $u(t)$  change as the velocity model is updated. In the following example, the reference (true) velocity model for observed data is

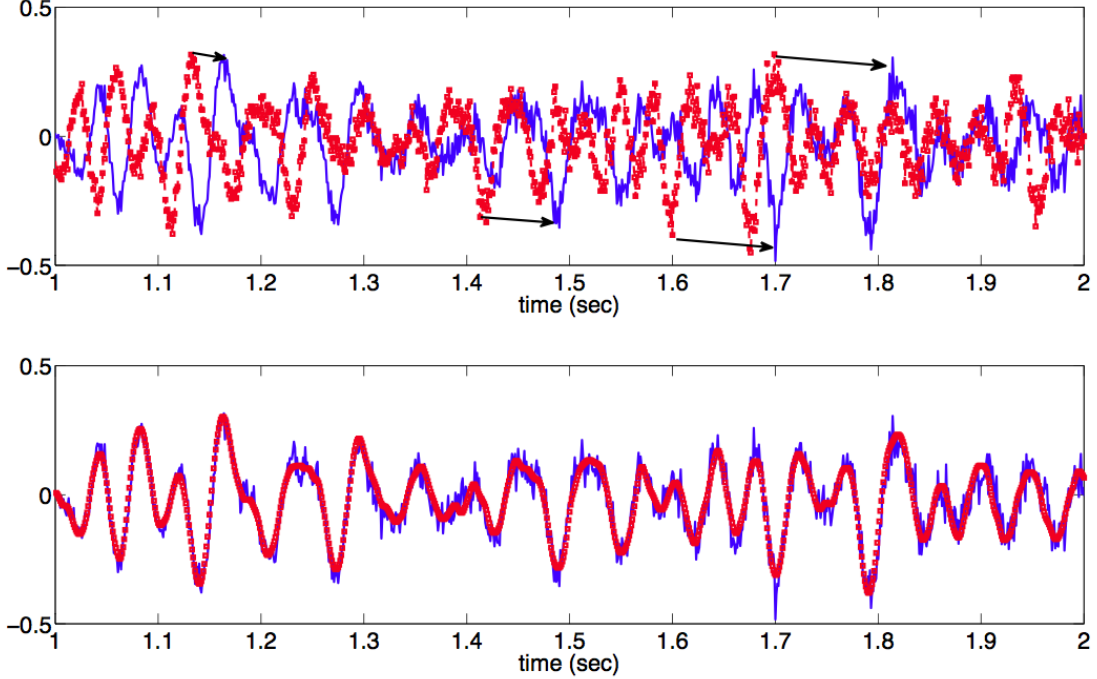


Figure 2: Matching noisy synthetic traces. Synthetic data are generated using the Marmousi velocity model and the other trace is created by applying a known mapping  $p(t)$  to it. The black arrows in the top figure connect corresponding peaks. The LFA signal  $U_h$  is used for frequency sweeping.

$v_{true}(x, z) = 5200 + 900 \exp(-|(x, z) - (1250, 1250)|^2/10^6)$  with a peak 6100 m/s at the center. This model has a high-velocity zone located at the center. The initial model is constant at  $v_{init}(x, z) = 5100$  m/s, i.e., without any a priori knowledge about the true model. Moreover, the initial model is chosen to create large discrepancies in traveltime between the predicted and the observed data. The grid size is 501 by 501 with a distance of 5 m between grid points along both directions. A Ricker wavelet with center frequency 50 Hz is used as an acoustic source. We now compare convergence of the LS and RGLS methods. Figure 3 shows the contour plots of velocity models in a crosshole inversion test (sources on the left edge, receivers on the right edge). The RGLS method converges to a velocity model with a qualitatively similar feature to the true model. Conventional LS method converges to a wrong model with a reduced velocity around 4000 m/s. The converged model is expected to not exactly replicate the true model, since the data are largely incomplete. In the case of full aperture data, RGLS recovers the model faithfully in a pointwise sense (not shown). The failure of LS minimization can be understood heuristically by noticing that the prediction  $u$  has been slowed down in the new model, hence has a lower  $l_2$  norm than in the original model, while its support is disjoint from that of  $d$ . The enlarged basin of attraction can be appreciated from convergence plots over the iterations of both methods. In Figure 4, we plot data misfit vs. iteration for both LS and RGLS, and rms values of  $m_T - m_k$ , where  $m_T$  is the true velocity model and  $m_k$  is the  $k^{th}$  step velocity model. We point out that data misfit in the convergence plot is calculated with the given observed data as in conventional LS misfit. Our proposed RGLS decreases the model rms error, regardless of the data misfit changes. LS decreases the data misfit but increases the model error at all iteration steps.

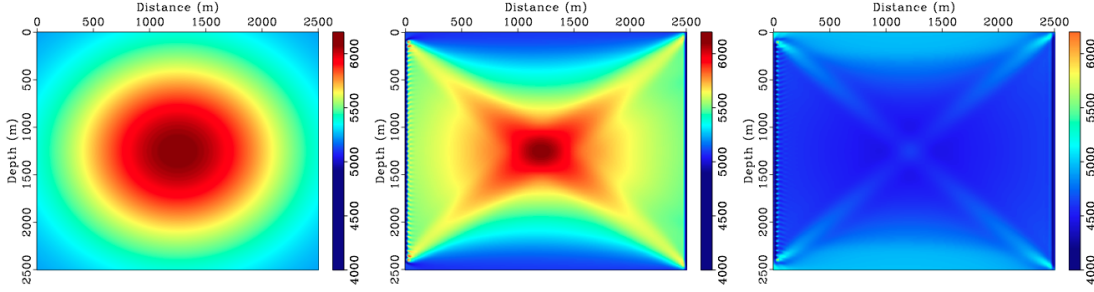


Figure 3: This is a color figure. Contour plots of velocity models: (left) reference/true model, (center) converged model from RGLS optimization, (right) converged model from LS optimization. The model obtained from LS optimization has 4500 m/s over most parts of the domain.

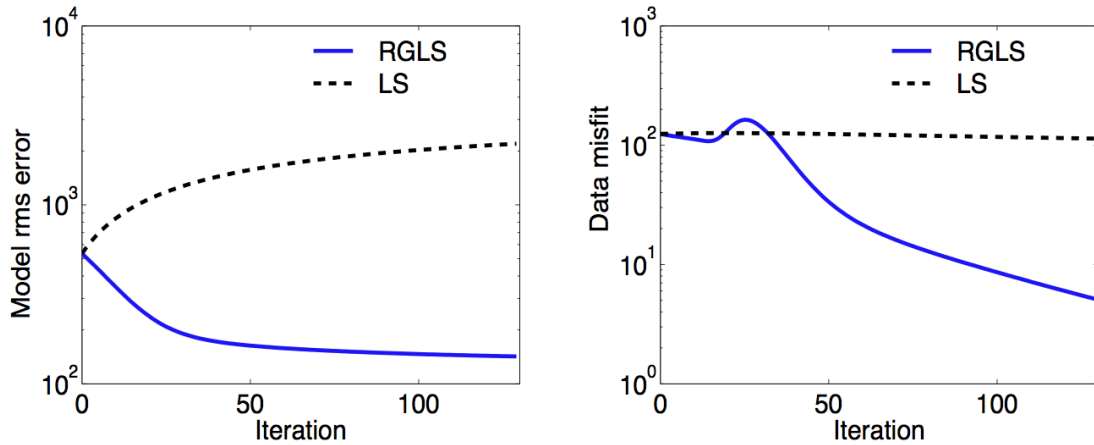


Figure 4: Convergence plot of the model rms error  $m_T - m_k$  (left) and data misfit  $J$  (right)

## Conclusions

We present a new method (RGLS) to overcome the cycle-skipping problem in FWI. The successful application of the method to a transmission seismic inversion problem is demonstrated, where conventional least-squares method converges to a wrong model. The proposed method substitutes a transported version of predicted data for the observed data in the conventional least-squares residual. In order to generate such transported data, mappings in the form of piecewise polynomials are found through a non-convex optimization formulation. The limitation of the RGLS method is that the predicted and observed data waveforms need to be sufficiently comparable.

## Acknowledgements

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