

A short note on phase and amplitude tracking for seismic event separation

Yunyue Elita Li and Laurent Demanet, Massachusetts Institute of Technology

SUMMARY

We propose a method for decomposing a seismic record into atomic events defined by a smooth phase and a smooth amplitude. The method uses an iterative refinement-expansion tracking scheme to minimize the highly nonconvex objective function. We demonstrate the proposed method on a noisy synthetic record from the shallow Marmousi model. Finally, we show an application of our method to low frequency extrapolation on the same record. This note is a short version of Li and Demanet (2015).

INTRODUCTION

We address the problem of decomposing a seismic record into elementary, or atomic components corresponding to individual wave arrivals. Letting t for time and x for receiver location, we seek to decompose a shot profile d into a small number r of atomic events v_j as

$$d(t, x) \simeq \sum_{j=1}^r v_j(t, x). \quad (1)$$

Each v_j should consist of a single wave front – narrow yet bandlimited in t , but coherent across different x – corresponding to an event of direct arrival, reflection, refraction, or corner diffraction.

In the simplest convolutional model, we would write $v_j(t, x) = a_j(x)w(t - \tau_j(x))$ for some wavelet w , amplitude $a_j(x)$, and time shift $\tau_j(x)$. In the Fourier domain, this model would read

$$\hat{v}_j(\omega, x) = \hat{w}(\omega)a_j(x)e^{i\omega\tau_j(x)}. \quad (2)$$

This model fails to capture frequency-dependent dispersion and attenuation effects, phase rotations, inaccurate knowledge of w , and other distortion effects resulting from near-resonances. To restore the flexibility to encode such effects without explicitly modeling them, we consider instead throughout this paper the expression

$$\hat{v}_j(\omega, x) = \hat{w}(\omega)a_j(\omega, x)e^{ib_j(\omega, x)}, \quad (3)$$

where the amplitudes a_j and the phases b_j are smooth in x and ω , and b_j deviates little from an affine (linear + constant) function of ω .

Finding physically meaningful, smooth functions a_j and b_j to fit a model such as Equation 1 and 3 is a hard optimization problem. Its nonconvexity is severe: it can be seen as a remnant, or cartoon, of the difficulty of full waveform inversion from high-frequency data. We are unaware that an authoritative solution to either problem has been proposed in the geophysical imaging community.

Many methods have been proposed to pick individual seismic events, such as AR filters (Leonard and Kennett, 1999) (close

in spirit to the matrix pencil method (Hua and Sarkar, 1990)), cross-correlations (Cansi, 1995), wavelets (Zhang et al., 2003), neural networks (Gentili and Michelini, 2006), etc. These papers mostly address the problem of picking isolated arrivals time, not parametrizing interfering events across traces. Some data processing methods operate by finding local slope events, such as plane-wave annihilation (Fomel, 2002). This idea has been used to construct prediction filters for localized wavelet-like expansion methods (Fomel and Liu, 2010), which in turn allow to solve problems such as trace interpolation in a convincing manner. It is plausible that concentration or clustering in an appropriate wavelet-like domain could be the basis for an algorithm of event separation. Separation of variables (SVD) in moveout-corrected coordinates has also been proposed to identify dipping events, such as in (Raoult, 1983) and (Blais, 2007). Smoothness criteria along reflection events have been proposed for separating them from diffraction events, such as in Fomel et al. (2007). These traditional methods fail for the event decomposition problem as previously stated:

- Because of cycle-skipping, gradient descent quickly converges to uninformative local minima.
- Because datasets don't often have useful low frequencies, multiscale sweeps cannot be seeded to guide gradient descent iterations toward the global minimum.
- Because the events are intertwined by possibly destructive interference, simple counter-examples show that greedy "event removal" methods like CLEAN or matching pursuit cannot be expected to succeed in general.
- Because wavefront shapes are not known in advance, linear transforms such as the slant stack (Radon), velocity scan, wavelets/curvelets, or any other kind of nonadaptive filters, don't suffice by themselves. The problem is intrinsically nonlinear.

The contribution of this paper is the observation that tracking in x and ω , in the form of careful growth of a trust region, can satisfactorily mitigate the nonconvexity of a simple nonlinear least-squares cost function, yielding favorable decomposition results on synthetic shot profiles. We have not been able to deal with the nonconvexity of this cost function in any other way.

This note is organized as follows. We first define the objective function and derive its gradient. We point out that the objective function is severely nonconvex and explain an explicit initialization scheme with MUSIC and the expansion and refinement scheme for phase and amplitude tracking. Finally, we demonstrate our separation method on a noisy synthetic data record. We illustrate the potential of event separation for extrapolation to unobserved low frequencies.

Phase and amplitude tracking

METHOD

Cost function and its gradient

We consider the nonlinear least-squares optimization formulation with cost function

$$J(\{a_j, b_j\}) = \frac{1}{2} \|\widehat{u}(\omega, x) - \widehat{d}(\omega, x)\|_2^2 + \lambda \|\nabla_\omega^2 b_j(\omega, x)\|_2^2 \quad (4)$$

$$+ \mu \|\nabla_x b_j(\omega, x)\|_2^2 + \gamma \|\nabla_{\omega, x} a_j(\omega, x)\|_2^2,$$

where \widehat{d} is the measured data in the frequency domain, ∇_k and ∇_k^2 , with $k = \omega, x$, respectively denote first-order and second-order partial derivatives, and $\nabla_{\omega, x}$ denotes the full gradient. The prediction \widehat{u} is

$$\widehat{u}(\omega, x) = \sum_{j=1}^r \widehat{w}(\omega) a_j(\omega, x) e^{ib_j(\omega, x)},$$

with $\widehat{w}(\omega)$ assumed known. The constants λ , μ , and γ are chosen empirically.

It is important to regularize with $\nabla_\omega^2 b_j(\omega, x)$ rather than $\nabla_\omega b_j(\omega, x)$, so as to penalize departure from dispersion-free linear phases rather than penalize large traveltimes.

The cost function is minimized using a gradient descent within a growing trust region. The gradients of (4) with respect to a_j and b_j are computed as follows:

$$\frac{\partial J}{\partial a_j} = \frac{1}{2} \left(\frac{\partial u}{\partial a_j} \delta \widehat{u}^* + \delta \widehat{u} \frac{\partial u^*}{\partial a_j} \right) + 2\nu \nabla_{\omega, x}^2 a_j, \quad (5)$$

$$\frac{\partial J}{\partial b_j} = \frac{1}{2} \left(\frac{\partial u}{\partial b_j} \delta \widehat{u}^* + \delta \widehat{u} \frac{\partial u^*}{\partial b_j} \right) + 2\lambda \nabla_\omega^2 \cdot \nabla_\omega b_j + 2\mu \nabla_x^2 b_j,$$

where

$$\frac{\partial \widehat{u}}{\partial a_j} = e^{-ib_j}, \quad \frac{\partial \widehat{u}^*}{\partial a_j} = e^{ib_j} \quad (6)$$

$$\frac{\partial \widehat{u}}{\partial b_j} = -ia_j e^{-ib_j}, \quad \frac{\partial \widehat{u}^*}{\partial b_j} = ia_j e^{ib_j}$$

$$\delta \widehat{u} = \widehat{u} - \widehat{d}.$$

The notation $\nabla_{\omega, x}^2$ refers to the Laplacian in (ω, x) . All the derivatives in the regularization terms are discretized by centered second-order accurate finite differences.

Initialization

The objective function in equation (4) is highly nonconvex due to the oscillatory nature of seismic data. We initialize the iterations by making use of an explicit solution of the minimization problem in a very confined setting where

- we pick a single (most informative) seed trace x ,
- we pick a subset of (most informative) seed frequencies ω ,
- and we assume a simplified model where the amplitudes are constant, and the phases are linear in ω . In other words, we return to the convolutional model

$$\widehat{d}(\omega) \simeq \sum_{j=1}^r a_j e^{i\omega \tau_j} \quad (7)$$

of equation (2) in order to locally approximate equation (3).

In this situation, the problem reduces to a classical signal processing question of identification of sinusoids, i.e., identification of the traveltimes τ_j and amplitudes a_j . There exist at least two high-quality methods for this task: the matrix pencil method of Hua and Sarkar (1990), and the Multiple Signal Classification (MUSIC) algorithm (Schmidt, 1986). We choose the latter for its simplicity and robustness.

Assume for the moment that the number r of events is known, though we address its determination in the sequel. The MUSIC algorithm only needs data $\{\widehat{d}(\omega_k)\}$ at $m = 2r + 1$ frequencies in order to determine the arrival times and amplitudes for r different events. In practice, the number m of frequencies may be taken to be larger than $2r + 1$ if robustness to noise is a more important concern than the lack of linearity of the phase in ω . In either case, we sample the data $\widehat{d}(\omega_k)$ on a grid of spacing $\Delta\omega$ around the fundamental frequency ω_f of the source wavelet, where the signal-to-noise ratio is relatively high.

The variant of the MUSIC algorithm that we use in this paper requires building a m -by- m Toeplitz matrix by collecting translates of the data samples as

$$\mathbf{T}_{k\ell} = \widehat{d}(\omega_k - \omega_\ell + \omega_f), \quad (8)$$

where $k, \ell = 1, 2, \dots, m$. After a singular value decomposition of \mathbf{T} , we separate the components relative to the r largest singular values, from the others, to get

$$\mathbf{T} = \mathbf{U}_s \Sigma_s \mathbf{V}_s^T + \mathbf{U}_n \Sigma_n \mathbf{V}_n^T.$$

We interpret the range space of \mathbf{U}_s as the signal space, and the range space of \mathbf{U}_n as the noise space, hence the choice of indices. The orthogonal projector onto the noise subspace can be constructed from \mathbf{U}_n as

$$\mathbf{P}_n = \mathbf{U}_n \mathbf{U}_n^T. \quad (9)$$

We then consider a quantitative measure of the importance of any given arrival time $t \in [0, \frac{2\pi}{\Delta\omega}]$, via the estimator function

$$\alpha(t) = \frac{1}{\|\mathbf{P}_n e^{i\bar{\omega}t}\|}. \quad (10)$$

In the exponent, $\bar{\omega}$ is a vector with m consecutive frequencies on a grid of spacing $\Delta\omega$. In the noiseless case, the estimator function has r sharp peaks that indicate the r arrival times τ_j for the r events. In the noisy case, or in the case when the phases b_j are nonlinear in ω , the number of identifiable peaks is a reasonable estimator for r , and the locations of those peaks are reasonable estimators of τ_j , such that the signal contains r phases locally of the form $\omega \tau_j + \text{const}$. Once the traveltimes τ_j are found, the amplitudes a_j follow from solving the small, over-determined system in equation (7). Each complex amplitude is further factored into a positive amplitude and a phase rotation factor, and the latter is absorbed into the phase.

In practice, it is sometimes worthwhile to apply this procedure on a handful of nearby traces, as an initial guess (seed) for the next step.

Phase and amplitude tracking

Tracking by expansion and refinement

The phases and amplitudes generated by the initialization give a local seed that needs to be refined and expanded to the whole record:

- The refinement step is the minimization of (4) with the data misfit restricted to the current trust region Ω in (ω, x) space.
- The expansion step consists in growing the region Ω to include neighboring samples in ω and x .

These steps are nested rather than alternated: the inner refinement loop is run until the value of J levels off, before the algorithm returns to the outer expansion loop. The expansion loop is itself split into an outer loop for (slow) expansion in x , and an inner loop for (rapid) expansion in ω . The nested ordering of these steps is crucial for convergence to a meaningful minimizer.

A simple trick is used to speed up the minimization of J in the complement of Ω , where only the regularization terms are active: $a_j(\omega, x)$ is extended by a constant, while $b_j(\omega, x)$ is extended by a constant in x , and by a linear + constant in ω . These choices correspond to the exact minimizers of the Euclidean norms of the first, respectively second derivatives of a_j and b_j , with zero boundary conditions on the relevant derivatives at the endpoints. This trick may be called “preconditioning” the regularization terms.

At the conclusion of this main algorithm, the method returns r phases $b_j(\omega, x)$ that are approximately linear in ω and approximately constant in x ; and r amplitudes $a_j(\omega, x)$ that are approximately constant in both x and ω ; such that

$$\hat{d}(\omega, x) \simeq \sum_{j=1}^r \hat{w}(\omega) a_j(\omega, x) e^{ib_j(\omega, x)}. \quad (11)$$

Application: frequency extrapolation

Frequency extrapolation, a.k.a. bandwidth extension, is a tantalizing test of the quality of a representation such as (11). A least-squares fit is first performed to find the best constant approximations $a_j(\omega, x) \simeq \alpha_j(x)$, and the best affine approximations $b_j(\omega, x) \simeq \omega\beta_j(x) + \phi_j(x)$, from values of ω within a useful frequency band. These phase and amplitude approximations can be evaluated at values of $\omega \in [\omega_0 \ \omega_1]$ outside this band, to yield synthetic flat-spectrum atomic events of the form

$$\hat{v}_j^e(\omega, x) = \alpha_j(x) e^{i(\omega\beta_j(x) + \phi_j(x))}, \text{ for } \omega \in [\omega_0 \ \omega_1], \quad (12)$$

where the superscript e denotes the extrapolation. These synthetic events can be further multiplied by a high-pass or low-pass wavelet, and summed up, to create a synthetic dataset. This operation is the seismic equivalent of changing the pitch of a speech signal without speeding it up or slowing it down.

Extrapolation to high frequencies should benefit high resolution imaging, whereas extrapolation to low frequencies should help avoid the cycle-skipping problem that full waveform inversion encounters when the low frequencies are missing from the data.

Notice that extrapolation to zero frequency is almost never accurate using such a simple procedure: more physical information is required to accurately predict zero-frequency wave propagation.

NUMERICAL EXAMPLES

We test our method on a noisy shot gather. The shot gather is generated from a shallow part of the Marmousi model using finite difference modeling. The finite difference scheme is second-order accurate in time and forth-order accurate in space. We use a 20 Hz Ricker wavelet as the source wavelet. Receiver spacing is 40 m.

As preprocessing, we remove the direct arrival from the shot record, because it has the strongest amplitude that would overwhelm the record. We apply an automatic gain control (AGC) to the remaining events so that the amplitudes on the record are more balanced. We mute the later arrivals in the data so that we resolve a limited number of events at a time. We transform the data to the frequency domain, bandlimit the data between 7 and 40 Hz, and finally add white Gaussian noise to the data. The noise has zero phase and its maximum amplitude is set to 30% of the maximum amplitude of the clean data.

Figure 1(a) shows the noisy shot record with 6 major events. We choose three seed traces at the center of the record where the events are well separated. In order to identify six different atomic events, we apply the MUSIC algorithm using 13 frequencies around the dominant frequency. To our inversion method, the data record in Figure 1(a) contains two types of noise. Besides the additive random noise, the scattering events that do not fit into our *six* atomic events model also pose a challenge to our inversion algorithm.

Figure 1(b) shows the inverted shot profile using the proposed tracking method. The inversion has clearly separated the six strong reflection events from the record, ignored the scattering events with smaller amplitudes and cleaned up the random noise from the record significantly. This demonstrate the robustness of our method. Figure 2 shows the corresponding well-separated atomic reflection events.

For each atomic events j , we estimate the two parameters $\beta_j(x)$ and $\phi_j(x)$ at each trace x within the data bandwidth. We then extrapolate the data to frequencies within $[1, 90]$ Hz using Equation 12. The waveforms are more compact (not shown). The high resolution data could be used for broadband high resolution seismic imaging.

In order to evaluate the accuracy of low frequency extrapolation, we model the seismic profile using a broadband source wavelet, whose amplitude spectrum is mostly flat between 1 and 7 Hz. We then compare the modeled data with the low frequency data obtained by frequency extrapolation within the same bandwidth. Figure 3 compares the two low frequency shot records. Note that the amplitudes differ quite substantially between the two records, while the phase information is rather similar.

There are (at least) two reasons why the extrapolated low-

Phase and amplitude tracking

frequency data do not exactly match the modeled low frequency data (even in phase). First, the unmodelled scattering events contribute to the low frequencies and overlap with the low frequency signal from the modeled reflection events. Second, There are numerical dispersion effects on the modeled low frequency data, whereas the extrapolated record is constructed using a nondispersive assumption. The second order finite differencing scheme we use here has a relatively lower phase speed at low frequencies than at high frequencies (Tong, 1994). The simulated low frequency data arrives later in the shot record (compare Figures 1a and 3a). Consequently, the phase differences due to numerical dispersion effects are more significant at far offsets than near offsets.

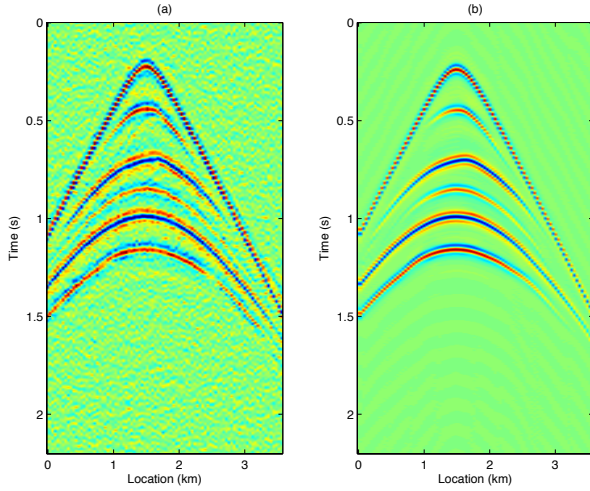


Figure 1: Comparison of the original shot record in (a) and the inverted shot record in (b).

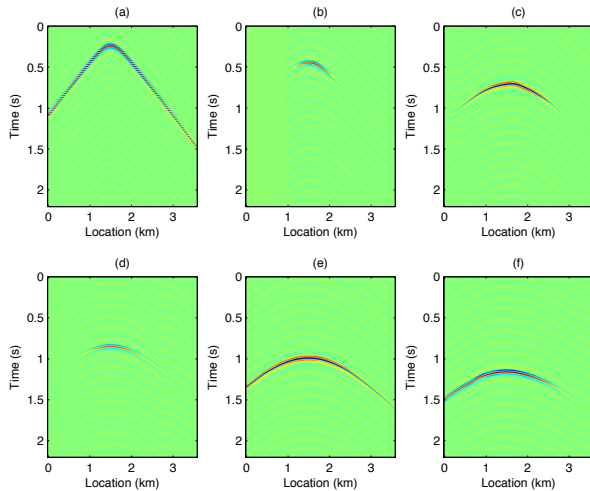


Figure 2: The six well-separated atomic event.

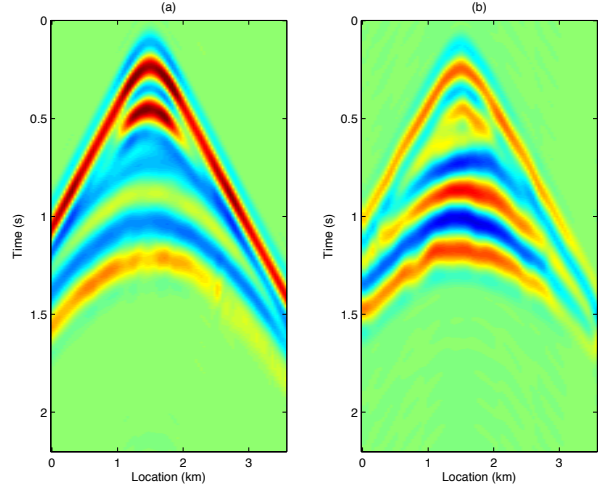


Figure 3: Comparison of the shot record modeled with a low frequency broadband ([1 7] Hz) wavelet in (a) and the shot record after frequency extrapolation to [1 7] Hz in (b).

DISCUSSIONS AND CONCLUSIONS

We propose a data-driven method for decomposing seismic records into individual, atomic events that correspond to isolated arrivals. The only piece of physical information needed to define these events is the fact that the acoustic or elastic wave equations are mostly dispersion-free, i.e., that they give rise to solutions with mostly linear phases in the frequency domain. The method is otherwise robust to random noise and unmodeled physical effects that create a departure from this model.

The success of the proposed iterative method for the nonconvex optimization formulation hinges on a delicate interplay between refinement of phases and amplitudes within a trust region, vs. slowly growing this trust region. If the region is grown too fast, the iterations will converge to an undesirable local minimum.

The most important limitation of the tracking method is the deterioration of the accuracy in phase (hence traveltime) estimation in the presence of unmodeled or splitting events. Our solution for the determination of the number of events further depends on the availability of a trace where the MUSIC estimator correctly categorizes them. Extrapolation to low frequencies is also limited by the inability of the phase-amplitude model to capture the physics of “zero-frequency radiation” due to the generic presence of a scattering pole in the Green’s function at $\omega = 0$. Finally, we observe a mild ill-conditioning of low-frequency extrapolation when comparing the wobbles of events in Figure 3(b), to those in Figure 1(b).

ACKNOWLEDGEMENTS

The authors thank Total SA for support. LD is also supported by AFOSR, ONR, and NSF.