Frequency down-extrapolation with TV norm minimization

Rongrong Wang* and Felix J. Herrmann
Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia

SUMMARY

In this work, we present a total-variation (TV)-norm minimization method to perform model-free low-frequency data extrapolation for the purpose of assisting full-waveform inversion. Low-frequency extrapolation is the process of extending reliable frequency bands of the raw data towards the lower end of the spectrum. To this end, we propose to solve a TV-norm based convex optimization problem that has a global minimum and is equipped with a fast solver. The approach takes into account both the sparsity of the reflectivity series associated with a single trace, as well as the inter-trace correlations. A favorable byproduct of this approach is that it allows one to work with coarsely sampled trace along time (as coarse as 0.02 s per sample), hence substantially reduces the size of the proposed optimization problem. Last, we show the effectiveness of the method for frequency-domain FWI on the Marmousi model.

INTRODUCTION

We consider acoustic full-waveform inversion (FWI) for a typical situation where low-frequency information is missing. It is well known that the low wavenumber data can effectively reduce the number of local minima of FWI and therefore can be expected to improve the inversion when a bad initial guess of the model is provided. However, missing low-frequency situations are commonly observed in both marine and land acquisitions where data below 5 Hz is often too noisy to be useful.

Approaches tackling this problem consists of both data processing techniques and wave-equation based techniques. The former follows the general belief that the raw data has some special structure that can be exploited to extract missing low-frequency information from high frequencies. One common assumption is sparse reflectivity, i.e., the reflectivity series corresponding to a single trace is a sparse spike train. We comment that this is an idealized assumption that does not strictly hold even in a 2D constant medium, let alone when we consider possible frequency dependent dispersion and attenuation effects present in real dataset. As a result, it is important to design a robust extrapolation method, which produces acceptable results when the underlying assumptions are slightly violated and noise is added. Our point here is that we prefer a method that is arguably robust rather than relying on unrealistic assumptions to create what seems like superior but in practice infeasible results.

Many methods have been proposed for broadening the spectrum. In the work by Xie (2013), the author proposed to isolate single events by windowing in combination with making a linear phase assumption to extend the spectrum. Unfortunately, the requirement of manually selecting windows prevents this approach from effectively handling complex models where two

overlapping event cannot visually separated. (Taylor and McCoy, 1979), on the other hand, uses a convex optimization approach that minimizes the data misfit plus the ℓ_1-norm of the Green's function. In the light of recent results in super-resolution in the field of Compressed Sensing (Candes and Fernandez-Granda, 2014; Demanet and Nguyen, 2015), this approach is better understood. In particular, it is now known that this approach is sensitive to discretization errors and is also known to give erroneous results when impulsive arrivals (spikes) of opposite signs are too closely (closer than twice the shortest wavelength) located. More recently, Li and Demanet (2016) proposed a new method, which for the first time explores the inter-trace relationships. To arrive at their results, the authors use the Multiple Signal Classification (MUSIC) algorithm (Schmidt, 1986) to deconvolve traces with only a few events, followed by a lateral continuous extension of these events to the nearby traces while keeping feasibility of the data. Due to the use of the MUSIC algorithm, this method is immune to discretization errors and the benefits of continuation amongst traces in the lateral direction has been clearly demonstrated by their stylized examples.

The missing low-frequency treatment in the wave-equation based approaches, on the other hand, seek to modify the inversion procedure itself with the hope of increasing the convexity of the optimization problem, leading to variants of FWI (Wu et al., 2013; Hu, 2014; Sun and Symes, 2012; Van Leeuwen and Herrmman, 2013; Warner and Guasch, 2013; Biondi and Almomin, 2014). These methods still have the true model as the global minimum, hence more stable than the extrapolation based techniques. However, they are perhaps less intuitive and the performance may sometimes be hard to predict.

In this abstract, we revisit the convex optimization approach as a frequency extrapolation method in the data processing regime. We propose to use a TV norm regularizer that takes into account the spatial correlation among traces. Comparing to the original ℓ_1 regularizer (Taylor and McCoy, 1979), our method has the added benefit of increased stability and speed as well as robustness to the discretization. It should be noted that this method is only intended for extrapolating towards the lower end of the spectrum, not to the higher end nor intended to recovering the complete Green's function.

METHODOLOGY

Under the no dispersion and no attenuation assumption, a single seismic trace d_j with m events can be written as

$$d_j(t) = q(t) * G_j(t), \qquad j = 1, ..., nr,$$
$$G_j(t) = \sum_{i=1}^{m} a_i^j \chi(t - t_i^j), \quad i = 1, ..., m, \qquad (1)$$

where q is the time-signature of the source wavelet, $G_j(t)$ is the "reflectivity" series at the jth receiver, nr is the total number of

receivers, $\chi(t-t_i)$ represents a single spike corresponding to a seismic event at time t_i and a_i^j is the reflection coefficient. Let us assume that we are working with the discretized versions so G_j, d_j, and q are all vectors in $\mathbf{R}^N$. Due to the band-limited nature of the source wavelet and the uneven spectral distribution of the wavelet's energy, we can only assume that data has an acceptable SNR within a middle frequency band, say $\Omega = [\omega_l, \omega_l + 1, ..., \omega_h]$—i.e.,

$$\hat{d}_j(\omega) = \hat{q}(\omega)\hat{G}_j(\omega), \quad \omega \in \Omega.$$

As it is often desired in FWI, the goal of this paper is to extend the reliable band from Ω to include a range of lower frequencies, say Ω_0. To simplify the problem, we assume that the source wavelet is known and its spectrum covers Ω.

The classical way to invert the reflectivity series is via sparsity promoting minimization

$$G_{\text{est}}^j = \arg\min_{G} \|G\|_1 \text{ subject to } \hat{G}(\omega) = \frac{\hat{d}_j(\omega)}{\hat{q}(\omega)}, \quad \omega \in \Omega. \quad (2)$$

Once G_{est}^j is found, the extrapolated data $\hat{d}_j(\Omega_0)$ is simply obtained by

$$\hat{d}_j(\omega) = \hat{q}(\omega)\hat{G}_{\text{est}}^j(\omega), \quad \omega \in \Omega_0.$$

Since both q and d are known, we essentially look for $\hat{G}(\omega), \omega \in \Omega_0$ given the knowledge of $\hat{G}(\omega), \omega \in \Omega$. This is a cousin of the well-known super-resolution problem in image processing except that our focus now are the low frequencies.

This method is appealing for its convex nature (which mean it has no local minimum) and robustness to noise, as well as the availability of fast algorithms from the Compressed Sensing community. However, its current form of this approach has some major issues that prevent it from accurately recovering perfect spike trains:

1. **the dilemma of discretization:** when taking time samples of traces, coarse sampling may result in the true spike locations in the reflectivity series to go off the grid causing large disretization errors, while fine discretization leads to unfeasible demands on storage and computation;

2. **deviation from perfect spikes**: for 2D media, the reflectivity series, which is the Green's functions restricted to the receiver locations, is not consistent with perfect spikes (the wave equation only permits Dirac in its solution in odd dimensions);

3. **limited resolution:** even with full knowledge of the low frequencies, i.e., data is known in $[0, \omega_h]$, the sparse signal estimation is subject to resolution issues (Candes and Fernandez-Granda, 2014): e.g. if two spikes with opposite signs are closer than the smallest wavelength presented in the data, they cannot be resolved by ℓ_1-norm minimization. (Donoho et al., 1992) showed that spikes with the same sign are immune to this issue. Unfortunately, we cannot assume positivity for the reflectivity.

While these issues introduce errors in the low-frequency extrapolation, they rarely make the extrapolation completely fail (most

errors are local where two spikes are too close). Therefore, we seek to reduce the error by using regularizers that explore similarities between neighboring traces in either a shot gather or a seismic line, which we refer to as "spatial similarity". Recall the reflectivity series associated with the jth trace is denoted by G_j. For simplicity of notation, let us define the 2D reflectivity gather G as a matrix formed by putting G_j, $j = 1, ..., m$ in its columns, i.e., $G := [G_1, ..., G_m]$. By definition, G has the same size as a shot gather. As a 2D image, the y-axis of G represents time and x-axis represents offset. Data constraints on G_j in Equation 2 become data constraints on G.

There are numerous ways to define spatial similarity. Here we explore one that is related to the finite-difference operator, which in turn is related to the TV norm (the absolute sum of gradient vectors). More rigorously, we use the following function to characterize the spatial similarity of G,

$$\mathscr{S}(G) = \frac{\sum\limits_{j=1}^{m-1} \|G_j - G_{j+1}\|_2^2}{\sum\limits_{j=1}^{m-1} (\|G_j\|_2^2 + \|G_{j+1}\|_2^2)}.$$

The smaller this quantity is, the more similarities consecutive reflectivity series share laterally. In case where G_j and G_{j+1} are not correlated, we have $\langle G_j, G_{j+1} \rangle = 0$ and $\mathscr{S}(G) = 1$. Hence in general we consider G to have no spatial similarity when $\mathscr{S}(G) \geq 1$.

A pitfall here is that no matter how continuous the traces in a reflectivity gather G visually appear, they can have no spatial similarity in terms of the value of $\mathscr{S}(G)$. This can be seen from a simple example when G contains only 1 dipping event. In that case, G as an image contains a single line. Assume the equation of line is $y = ax + b$, where y is the time index and x is the receiver or offset index. Suppose the slope $a = -2$. Then it is easy to verify the 0 correlation between neighbouring traces, i.e., $\langle G_j, G_{j+1} \rangle = 0$, and that $\mathscr{S}(G) = 1$, despite the fact that the line is visually continuous.

Instead of directly trying to construct the low frequency component of G from its high frequency component, we propose to first recover some "good" gather G_m in the sense that it is associated with G through some mathematical relations but with greater spatial similarity. Once G_m is obtained, the low frequencies of G can be recovered from that of G_m through their relations. To this end, we define $G_m = f * G$, where

$$f = (\underbrace{1, ..., 1}_{r}, 0, ..., 0)^T$$

(the choice of r is to be discussed later), and where the notation $f * G$ means f is being convolved with each column in G. The benefit of these convolutions is that they make the events (lines) in G effectively thicker so nearby reflectivity series are more likely to be correlated. In fact, one can check for the previous example that, the value of $\mathscr{S}$ decreases from $\mathscr{S}(G) = 1$ to $\mathscr{S}(G_m) = 2/r$ provided that $r > 2$.

In order to utilize both the spatial similarity and the sparsity of G, we proposed a TV norm regularizer that replaces the ℓ_2

norm in the definition of $\mathscr{S}(\cdot)$ by an ℓ_1 norm and adds a vertical TV regularizer. The resulting problem is

$$G_{m,\text{est}} = \underset{G_m}{\arg\min} \|G_m\|_{TV},$$

$$\text{subject to } \hat{G}_m(\omega) = \hat{G}(\omega)\hat{f}(\omega) = \frac{\hat{d}(\omega)\hat{f}(\omega)}{\hat{q}(\omega)}, \omega \in \Omega,$$

(3)

where for a given matrix M, the 2D TV-norm is defined as $\|M\|_{TV} := \sum_{i,j} \|\nabla M(i,j))\|_{\ell_1}$ with

$$\nabla M(i,j) = \begin{bmatrix} M(i,j) - M(i,j+1) \\ M(i,j) - M(i+1,j) \end{bmatrix}.$$

The low frequencies of G are obtained through the relation

$$\hat{G}(\omega) = \hat{G}_m(\omega)/\hat{f}(\omega), \omega \in \Omega_0.$$

Let us illustrate the effectiveness of our approach on a simple synthetic example. The ground truth G (Figure 1a) contains three events, each corresponding to a line in the reflectivity gather. Red dots in Figure 1a represent the value 1, blue are -1, and blanks are zeros. The crossing points of the events are those where ℓ_1 minimization (Equation 2) has trouble to recover (Figure 1b). Here we have set $\Omega = [30,50]$ and $\Omega_0 = [0,30]$. The auxiliary image G_m defined by $G_m = f * G$ with parameter $r = 2$ is plotted in Figure 1c, where one can see that the events are thickened. After performing the TV norm minimization, the inverted G_m in Figure 1d has much milder inaccuracy at the crossing points in the boxed area than those in Figure 1b.

Two comments are in order. A natural question is that whether we can increase spatial similarity by making the receivers denser. The answer is yes, but in doing so, we also increased the total number reflectivity series. In the end, the normalized error for each G_j is not decreased. Due to the limitation of space, we omit details of the argument. The second comment is about the choice of r. While mostly heuristic, we highlight some rough guidance. In principle, we want to choose r such that the derivative of the convolution kernel f are resolvable under ℓ_1 minimization. From the discussion in Candes and Fernandez-Granda (2014), the resolvability at least requires $r \geq N/(T\omega_h)$ where ω_h is the highest frequency in Ω, T is the sampling duration and N is the length of trace. On the other hand, we do not want r to be too large. Otherwise f becomes a powerful low-pass filter making the events in G_m lose resolution.

Acceleration

A typical seismic trace is sampled at 1,2,4 or 8 milliseconds time interval with a sampling duration about $4-20\,\text{s}$. This means that the trace length of a shot gather can be as long as 20000 samples. Performing the ℓ_1 or TV norm minimization, for example, on a 20000×300 shot gather can be quite expensive, especially if we want to raise one more dimension of G to include reflectivity series underlying a whole seismic line so as to obtain a 2 dimensional spatial continuation. As a consequence, it is desirable to use coarser time grid. Since the reflectivity series in G is assumed to be consist of spikes, which means it has a broad spectrum, then a downsampling may cause significant aliasing. On the other hand, our auxiliary variable $G_m = f * G$ has relatively low energy in high frequency
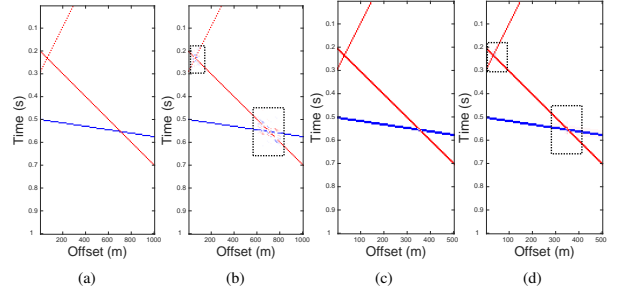

Figure 1: The effect of spatial continuation. a) Ground truth reflectivity gather G; b) Inverted reflectivity gather using $30-50\,\text{Hz}$ data via ℓ_1 minimization; c) Fattened reflectivity gather G_m defined as $G_m = f * G$, $f = [1,1,0,\ldots,0]^T$; d) Inverted G_m using $30-50\,\text{Hz}$ data via TV-norm minimization. The reconstruction error of ℓ_1 minimization in the boxed areas is greatly improved by the use of TV norm minimization and fattening.

bands because the convolution kernel f, besides delivering the already mentioned thickening benefit, also acts as a low-pass filter. This means traces corresponding to G_m, i.e. $d * f$, can be subsampled without much loss. The final extrapolation result is also barely affected because we only care about obtaining the low frequency information.

NUMERICAL EXPERIMENTS

We perform inversion experiments to demonstrate the effectiveness of extrapolation results of the proposed algorithm. We use the Marmousi model with about 300 meters water layer added on top of it to generate the data. Assuming that the data is only available in $5-20\,\text{Hz}$ frequency bands, which we generated using frequency domain finite difference modeling code. It is also reasonable to assume the availability of broadband $(1-20\,\text{Hz})$ data for direct waves. Because we can easily obtain this data by doing forward modeling on a constant water velocity model. To perform extrapolation, we remove the direct waves from $5-20\,\text{Hz}$ synthetic data since they are too strong compared to the reflections and the refractions. We carry out the proposed algorithm twice to extrapolate from $\Omega = \{5,5.1,...,20\}\,\text{Hz}$ to $\Omega_0 = \{1,1.1,...,4.9\}\,\text{Hz}$. First time on the reflectivity gathers (implemented in paralell for each shot) and the next time on the reflectivity series associated with a seismic line. We compare these two results. While implementing on a seismic line, we define the TV norm for all three directions (along time, sources, and receivers). The direct output of the extrapolation algorithm is $\hat{G}(\omega), \omega \in \Omega_0$. We use it to construct the final extrapolated data by the formula $\hat{d}(\omega) = \hat{G}(\omega)\hat{q}(\omega) + \hat{d}_0(\omega)$, where $\hat{d}_0(\omega)$ is the data for direct waves at frequency ω and $\hat{q}(\omega)$ is the Fourier coefficient for the source wavelet. The inversion is performed using frequency continuation under inverse crime.

In the extrapolation algorithm, we set the time duration and the total number of samples in each reflectivity series to be 10 s and 400 respectively (i.e. sampling in every 0.025 s) . We set the parameter $r = 10$ in the definition of f, and solve the TV norm minimization problem (Eq 3) approximately using the NESTA

algorithm (Becker et al., 2011).

Figure 2 shows a comparison between the extrapolated data and the true data for three frequencies. Only the real part is plotted. Among the 12 sub-figures, the ones in the first column are frequency slices of the true data, those in the second column are extrapolated data with ℓ_1 minimization on reflectivity gathers, in the third column are extrapolated data with TV norm minimization on reflectivity gathers, and in last column are TV norm minimization on the seismic line. The first, second and third row corresponds to 4 Hz, 3 Hz and 2 Hz frequency slices, respectively. We can see that the use of spatial information has greatly enhanced and stabilized the result.
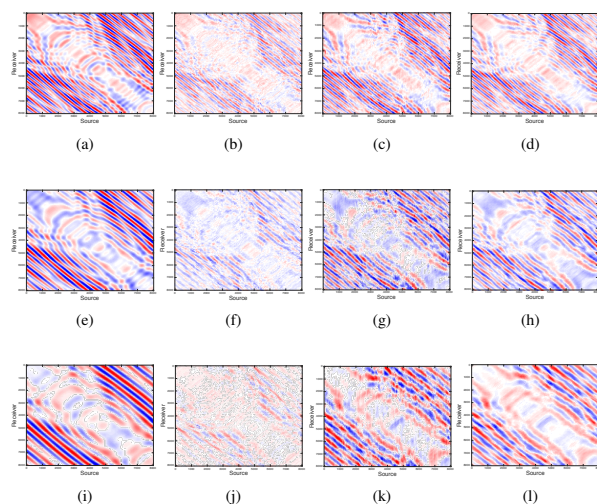

Figure 2: First row: 4 Hz data slices. Second row: 3 Hz data slices. Third row: 2 Hz data slices. First column: true data slices. Second column: extrapolated data with ℓ_1 minimization. Third column: extrapolated data with TV norm minimzation performed on a shot gather. Forth column: extrapolated data with TV norm minimization performed on a seismic line.

We perform Full Waveform Inversion using frequency continuation with overlapping frequency batches in the form of $..., \{5\,\text{Hz}, 5.5\,\text{Hz}\}, \{5.5\,\text{Hz}, 6\,\text{Hz}\}, ...$. For each frequency batch, we use 20 L-BFGS iterations. 67 uniformly spaced sources and 134 uniformly spaced receivers at depth of 24 m from the top surface are used for this experiment. We do not allow water velocity above the sources and receivers to be updated, i.e., setting the gradient of each update to 0 for the grid points between $0 - 24\,\text{m}$ in depth. This is to make the inversion more robust to extrapolation errors. We use a 1D linear initial guess, whose velocity profile is plotted in Figure 4. Figure 3a shows the true velocity model. Figure 3b-d show the inversion results. Figure 3b shows the inverted model with narrowband $5 - 15\,\text{Hz}$ true data, 3c shows the inverted model with broadband $1 - 15\,\text{Hz}$ true data, and 3d shows the inverted model with $1 - 5\,\text{Hz}$ extrapolated data and 5-15Hz true data, where the extroplation is obtained by applying the proposed method to a seismic line. While the extrapolation is not entirely accurate, we demonstrated that if used carefully, it can help the FWI to correct large wavelength velocity errors in the initial model and
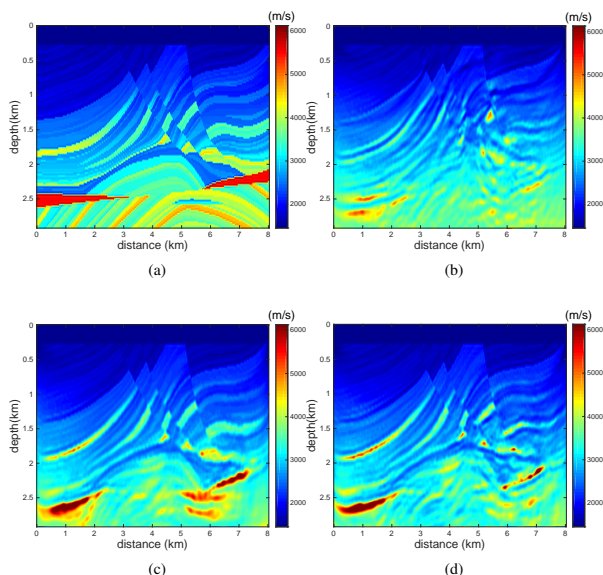
accelerate the convergence.

Figure 3: (a)True model; (b) Inverted model with 5-15 Hz true data; (c) Inverted model with 1-15Hz true data; (d) Inverted model with 5-15 Hz true data and the 1-5 Hz extrapolated data on seismic line.
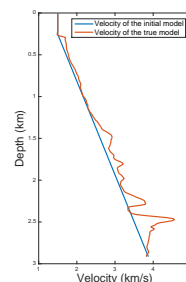

Figure 4: Mean velocity of the initial model compared to that of the true model at various depths.

CONCLUSION

In this paper, we proposed a method for low-frequency data extrapolation. The method inherits the stability of the traditional ℓ_1 minimization approach, while having both increased accuracy and speed. Synthetic inversion test shows that the artificial low frequency data is accurate enough to correct most of the large wavelength kinetic error in the initial model.

ACKNOWLEDGEMENT

REFERENCES

Becker, S., J. Bobin, and E. J. Candès, 2011, Nesta: a fast and accurate first-order method for sparse recovery: SIAM Journal on Imaging Sciences, **4**, 1–39.

Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Candes, E. J., and C. Fernandez-Granda, 2014, Towards a mathematical theory of super-resolution: Communications on Pure and Applied Mathematics, **67**, 906–956.

Demanet, L., and N. Nguyen, 2015, The recoverability limit for super-resolution via sparsity: arXiv preprint arXiv:1502.01385.

Donoho, D. L., I. M. Johnstone, J. C. Hoch, and A. S. Stern, 1992, Maximum entropy and the nearly black object: J. Roy. Stat. Soc. B Met., 41–81.

Hu, W., 2014, Fwi without low frequency data - beat tone inversion: Presented at the SEG Technical Program Expanded Abstracts, SEG.

Li, Y. E., and L. Demanet, 2016, Full waveform inversion with extrapolated low frequency data: arxiv:1601.05232.

Schmidt, R., 1986, Multiple emitter location and signal parameter estimation: IEEE Trans. Antennas Propagation, **AP-34**, 276 – 280.

Sun, D., and W. W. Symes, 2012, Waveform inversion via nonlinear differential semblance optimization: Presented at the 2012 SEG Annual Meeting, Society of Exploration Geophysicists, SEG.

Taylor, H. L., B. S., and J. McCoy, 1979, Deconvolution with the ℓ_l norm: Geophysics, **49**, 39–52.

Van Leeuwen, T., and F. J. Herrmman, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International.

Warner, M., and L. Guasch, 2013, Adaptive waveform inversion: Presented at the SEG Technical Program Expanded Abstracts, SEG.

Wu, R.-S., J. Luo, and B. Wu, 2013, Ultra-low-frequency information in seismic data and envelope inversion: Presented at the 2013 SEG Annual Meeting, Society of Exploration Geophysicists, SEG.

Xie, X.-B., 2013, Recover certain low-frequency information for full waveform inversion: Presented at the SEG Technical Program Expanded Abstracts, SEG.