geoconvention
Calgary • Canada • May 15-19 2017

Anti-leakage least-squares spectral analysis for data regularization

Ebrahim Ghaderpour, Wenyuan Liao*, Michael Lamoureux*, Da Li* and Spiros Pagiatakis***
**University of Calgary, **York University*

Summary

Spatial transformation of irregularly sampled data to regularly sampled data is a challenging problem in many areas such as seismology. The least-squares spectral analysis (LSSA) is an alternative to the classical Fourier analysis that can analyze irregularly sampled data (data series). Although the LSSA takes into account the correlation among the sinusoidal base functions on an irregularly spaced series, it still suffers from the problem of "spectral leakage" that is leaking energy from one spectral peak into another. We propose an iterative method called "anti-leakage LSSA" to attenuate the spectral leakage and consequently reconstruct the data on a regularly spaced series. In this method, we first search for a spectral peak with the highest energy, and then we remove (suppress) it from the original data. In the next step, we search for a new peak with the highest energy in the residual data and remove both, the new and the old components simultaneously from the original data using the least-squares method. We repeat this procedure until all significant spectral peaks are estimated and removed simultaneously from the original data. In each step, if the frequency corresponding to a new peak with the highest energy is sufficiently close to a previous estimated frequency, then the previous frequency will be removed from the set of estimated frequencies to be estimated more accurately in the next step. We demonstrate the robustness of our method on an irregularly sampled synthetic data.

Introduction

Regularization (a typical spectral interpolation) of irregularly sampled (unequally spaced) data is a crucial problem in seismology. There are a number of algorithms proposed already to solve this problem such as Minimum Weighted Norm Interpolation (MWNI) (Liu and Sacchi, 2004), Anti-Leakage Fourier Transform (ALFT) (Xu et al., 2005; Xu et al., 2010) and Arbitrarily Sampled Fourier Transform (ASFT) (Guo et al., 2015). These algorithms are established mainly on the basis of Fourier transform. For instance, the ALFT and ASFT estimate the Fourier coefficients of the data first by searching for the peak with maximum energy and subtracting its component from the data and then repeating the procedure again on the residual data until all the Fourier coefficients are estimated. This way the spectral leakages of Fourier coefficients emerged from the nonorthogonality of the global Fourier basis functions will be attenuated (experimentally).

However, these methods usually cannot find the correct location of a peak with maximum energy from a preselected set of frequencies because of the correlation between the sinusoids, and so they do not efficiently reduce the leakages. This shortcoming becomes more severe when data has more components. Moreover, the trends and the covariance matrix associated with the data (if they exist) are not considered in these algorithms.

The LSSA (Vaníček, 1969; Pagiatakis, 1999) improves these weaknesses. It basically estimates a frequency spectrum based on the least-squares fit of sinusoids to the data by accounting for measurement errors, trends and constituents of known forms. In the LSSA, since the selected frequencies are examined one at the time (out-of-context), the spectral leakages still appear which result in interpolation inaccuracy, although the nonorthogonality between the sine and cosine basis functions is taken into account for each frequency. In this contribution, we apply the idea of maximum energy in the LSSA to improve this shortcoming and show its outstanding performance in regularization.

Method

Let $\boldsymbol{f} = [f(x_\ell)]$ be the column vector of n *data* points, where the x_ℓ's are (inherently) irregularly spaced. Let $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_q]$ be the $n \times q$ matrix of constituents of known forms, and $\boldsymbol{\phi}_k = [\cos 2\pi k x_\ell,\ \sin 2\pi k x_\ell]$ be the $n \times 2$ matrix of a fixed frequency k, where k is in a preselected set of frequencies $\boldsymbol{K}$ that is typically the set of all positive integers less than or equal to $n/2$. Suppose that $\boldsymbol{C_f}$ is the covariance matrix associated with $\boldsymbol{f}$ (if it exists) and let $\boldsymbol{P} = \boldsymbol{C_f^{-1}}$. The algorithm for the anti-leakage LSSA is as follows:

Step 1. Minimize the cost function $\psi_1(\boldsymbol{c}) = (\boldsymbol{f} - \boldsymbol{\phi c})^T \boldsymbol{P} (\boldsymbol{f} - \boldsymbol{\phi c})$ to compute the *residual data* $\boldsymbol{g} = \boldsymbol{f} - \boldsymbol{\phi}\,\hat{\boldsymbol{c}}$ and its *norm* $L = \boldsymbol{g}^T \boldsymbol{P} \boldsymbol{g}$, where $\hat{\boldsymbol{c}} = \boldsymbol{N}^{-1} \boldsymbol{\phi}^T \boldsymbol{P} \boldsymbol{f}$, $\boldsymbol{N} = \boldsymbol{\phi}^T \boldsymbol{P} \boldsymbol{\phi}$, and '$T$' indicates transpose.

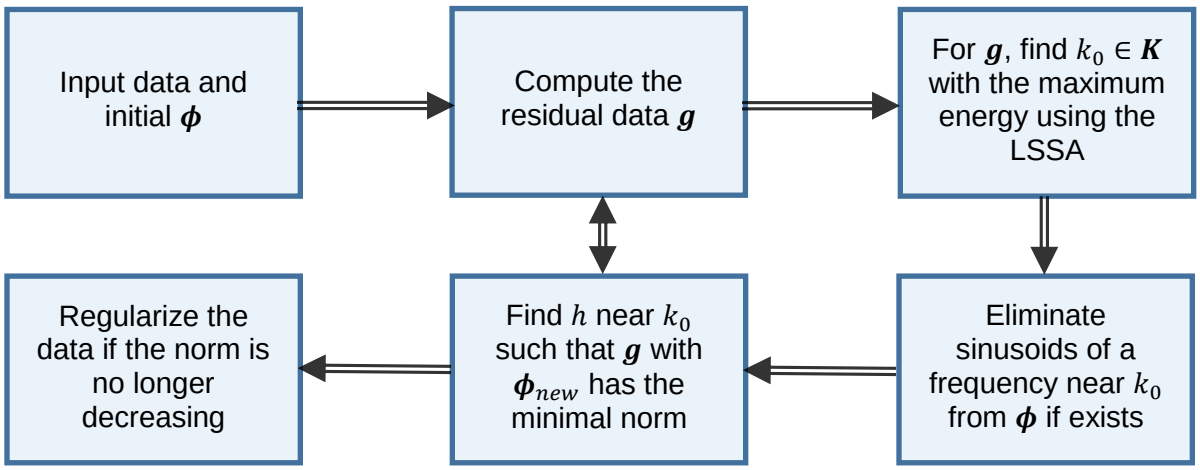
Step 2. Minimize the cost function $\psi_2(\boldsymbol{c}, \boldsymbol{c}_k) = (\boldsymbol{f} - \boldsymbol{\phi c} - \boldsymbol{\phi}_k \boldsymbol{c}_k)^T \boldsymbol{P} (\boldsymbol{f} - \boldsymbol{\phi c} - \boldsymbol{\phi}_k \boldsymbol{c}_k)$ to find k_0 in $\boldsymbol{K}$ with the maximum *energy* defined by $s(k) = (\boldsymbol{g}^T \boldsymbol{P} \boldsymbol{\phi}_k \widehat{\boldsymbol{c_k}})/L$, where

$$\widehat{\boldsymbol{c_k}} = \left(\boldsymbol{\phi}_k^T \boldsymbol{P} \boldsymbol{\phi}_k - \boldsymbol{\phi}_k^T \boldsymbol{P} \boldsymbol{\phi} \boldsymbol{N}^{-1} \boldsymbol{\phi}^T \boldsymbol{P} \boldsymbol{\phi}_k\right)^{-1} \boldsymbol{\phi}_k^T \boldsymbol{P} \boldsymbol{g}.$$

Step 3. Eliminate the cosine and sine basis functions of a frequency h in $\boldsymbol{\phi}$ such that $|h - k_0| < b$ if exists. Assuming that the difference between any two consecutive actual frequencies of the components in the data is greater than one, we may choose $b = 0.5$ to resolve the peaks and avoid singularity of $\boldsymbol{N}$.

Step 4. Repeat Step 1 to find h in $\boldsymbol{I} = [k_0 - b, k_0 + b]$ such that $\boldsymbol{g}$ with $\boldsymbol{\phi}_{new} = [\boldsymbol{\phi}, \cos 2\pi h x_\ell,\ \sin 2\pi h x_\ell]$ has the lowest norm L and then go to Step 2. Finding h up to a chosen decimal place can be done by appropriate partitioning of $\boldsymbol{I}$. Terminate the process if L no longer decreases. Use the frequencies of all sinusoids listed in the final $\boldsymbol{\phi}$ and their corresponding estimated amplitudes to regularize the data.

In Steps 1 and 2, the method of least-squares has been used for the minimization. The frequencies of the constituents are real numbers. Removing the constituent of a particular frequency with maximum energy from the data reduces the leakages. On the other hand, the elimination of the basis functions of a frequency h in Step 3 is crucial in the anti-leakage LSSA because considering the correlation between the sinusoids of different frequencies results in a more accurate h in the next step. This correlation will be taken into account as the column dimension of $\boldsymbol{\phi}$ increases in the process. The residual data $\boldsymbol{g}$ approaches zero rapidly because the frequencies are accurately estimated recursively that also prevents increasing the column dimension of $\boldsymbol{\phi}$ in many applications. We summarize the anti-leakage LSSA in the flowchart below.



Example

We use the anti-leakage LSSA to regularize the data of size 128 given by Equation (13) (Xu et al. 2005) with three additional constituents:

$$f(x_\ell) = 5\sin(25.6\,x_\ell) + 2.5\,\sin(128\,x_\ell + 1) + \sqrt{3}\,\sin(140\,x_\ell) + \sqrt{2} + \pi x_\ell,$$

where x_ℓ (ℓ = 1,2,3,…,128) is a random number in [0,1] generated by the MATLAB command "*rand*". All the x_ℓ's are sorted in ascending order. We choose the initial set of frequncies as $\boldsymbol{K}$ = {1,2,3,…,64}, and we estimate the frequencies up to 4 decimal places. Also, we select the initial $\boldsymbol{\phi}$ as $\boldsymbol{\phi} = [\mathbf{1}, \boldsymbol{x}]$ to account for the linear trend present in the data, where '$\mathbf{1}$' and '$\boldsymbol{x}$' are the column vectors of all ones and the 128 random numbers, respectively. Since there is no covariance matrix associated with the data, we do not consider the square weight matrix $\boldsymbol{P}$ in the calculation.

Note that the data has three frequency components with irrational frequencies 4.0743665, 20.3718327 and 22.2816920, rounded to 7 decimal places. Table 1 shows the result of the anti-leakage LSSA. In the first iteration, the frequency 4.0384 is estimated that is approximately 0.036 different from its actual value 4.0743665. This is a shortcoming of the existing methods such as the LSSA and ASFT caused by the presence of other constituents in the data. In the second and third iterations, the other two frequencies are estimated that by removing their corresponding components from the data simultaneously, the first frequency is better approximated and so forth (see the highlighted numbers in Table 1). The norm of the final residual data is 0.0036 which means a very high accuracy in regularization (see Figure 1).

Table 1. The result of frequency estimation of the constituents of the irregularly sampled data after each iteration

Iteration number	1st frequency	2nd frequency	3rd frequency	Norm of residual
First	4.0384			23.5040
Second	4.0384	20.3057		13.3170
Third	4.0384	20.3057	22.2947	3.5079
Fourth	**4.0748**	20.3057	22.2947	2.2501
Fifth	4.0748	**20.3708**	22.2947	0.3068
Sixth	4.0748	20.3708	**22.2818**	0.0468
Seventh	4.0748	**20.3718**	22.2818	0.0319
Eighth (final)	**4.0744**	20.3718	22.2818	0.0036

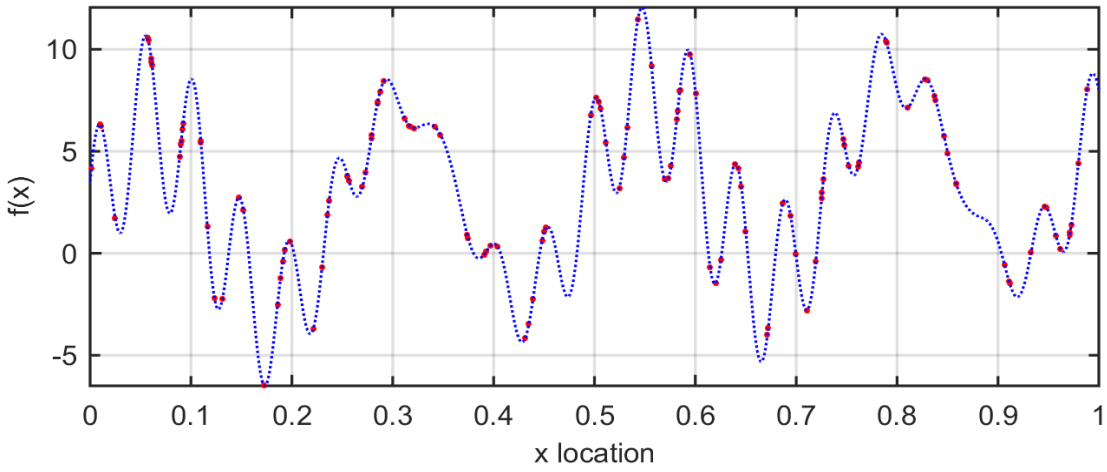


Figure 1. The irregularly sampled data (128 red dots) and its regularization using the anti-leakage LSSA (blue dots).

Conclusions

We proposed a new algorithm to regularize irregularly sampled data by taking into account the covariance matrix associated with the data and constituents of known forms. This method applies the LSSA to search for frequency components with maximum energy and uses an iterative algorithm to estimate the actual frequencies of the components in the data and consequently reconstructs the data on a regularly spaced series. The synthetic example in this work showed the robustness and effectiveness of this method in regularization. In future work, we extend the method to higher dimension cases and apply it to irregularly sampled seismic data and will demonstrate its outstanding performance for trace interpolation.

Acknowledgements

This research has been financially supported by Pacific Institute for the Mathematical Sciences (PIMS), Natural Sciences and Engineering Research Council (NSERC) of Canada and the postdoctoral program at the Department of Mathematics and Statistics, University of Calgary.

References

Guo, Z., Y. Zheng and W. Liao, 2015. Arbitrarily Sampled Fourier Transform for 5D Interpolation: GeoConvention, Calgary, Canada

Liu, B., M. D. Sacchi, 2004. Minimum weighted norm interpolation of seismic records: *Geophysics*, **69**(6), 1560-1568.

Pagiatakis, S., 1999. Stochastic significance of peaks in the least-squares spectrum: J of Geodesy **73**, 67-78.

Vaníček P., 1969. Approximate spectral analysis by least-squares fit: Astrophysics and Space Science **4**, 387-391

Xu, S., Y. Zhang, and G. Lambaré, 2010. Antileakage Fourier transform for seismic data regularization in higher dimensions: Geophysics, **75**, no. 6, WB113–WB120.

Xu, S., Y. Zhang, D. Pham, and G. Lambaré, 2005. Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, no. 4, V87–V95.