Jointly filtering and regularizing seismic data using space-varying FIR filters

Apostolos Kontakis[†], Xander H. Campman[†], Geert J. T. Leus[*], Zijian Tang[†] and Mike Danilouchkine[†]

[*]Circuits and Systems group, Faculty of EEMCS, Delft University of Technology

Address: Mekelweg 4, 2628 CD, Delft, The Netherlands

Email: G.J.T.Leus@tudelft.nl

[†]Novel Geophysical Measurements (PTI/EN), Shell Global Solutions International B.V.

Address: Kessler Park 1, 2288 GS Rijswijk, The Netherlands

Email: A.Kontakis@student.tudelft.nl, {Xander.Campman, Zijian.Tang, Mike.Danilouchkine}@shell.com

Abstract—**Array forming in seismic data acquisition can be likened to FIR filtering. Misplacement of the receivers used to record seismic waves can lead to degraded performance with respect to the filtering characteristics of the array. We propose two methods for generating linear space-varying filters that take receiver misplacements into account and demonstrate their performance on synthetic data.**

I. INTRODUCTION

Variations in the sampling interval when sampling a signal are often difficult or impossible to prevent. This might make processing the sampled signal problematic, if the related tools can only handle uniformly sampled signals. The motivation for our work comes from the field of seismic data acquisition. A very common practice in this field, is to sum together the (possibly weighted) output signal of multiple seismic receivers. This process is known as array forming and is used to improve the signal-to-noise ratio and to reduce the amount of recorded data.

Being a weighted summation of the output of a finite number of receivers, array forming can be likened to FIR filtering. More specifically, we can view array forming as the application of a linear space-invariant (LSI) filter, since usually the same set of weights is used for the array elements of all arrays in a field.

A problem arises, however, when the receivers are misplaced due to e.g., terrain difficulties. Usually the array weights are designed for a specific geometrical layout of the array elements. When this geometrical layout is violated, it can prove detrimental for the filtering capabilities of the array, as shown in [1]. Fortunately, advances in acquisition hardware have enabled us a) to record the output of each individual receiver and b) to know with high (but limited) accuracy the actual location of each receiver. This makes more sophisticated techniques viable for array forming/filtering that can compensate for irregularities in sampling.

A number of solutions have been suggested for the problem of array forming/filtering nonuniformly sampled data, which can be roughly divided in three categories. Methods of the first category interpolate an FIR filter that is designed to filter uniformly sampled data. We shall call this FIR filter the prototype filter. An example is given in [2], where the prototype filter is interpolated to the actual locations of the receivers. These interpolated filter coefficients (or equivalently, the array weights) are then reweighted based on the sampling density at the area around each receiver. We shall refer to this method as geometry-compensating filtering (GCF). The second category involves methods that approximate the outputs of the prototype filter applied to the uniformly sampled data. An example is given in [3], which uses the projections onto convex sets (POCS) framework applied to nonuniformly sampled data. The third category involves reconstruction of the data at the regular sampling locations. The prototype filter can then be applied to the reconstructed data. The methods given in [4], [5] are examples of data reconstruction.

The goal of this work is to propose two methods that generate a linear space-varying (LSV) FIR filter suitable for filtering the nonuniformly sampled data. We will refer to these methods as **Method A** and **Method B**. The LSV FIR filter designed by any of these two methods generates the filter output at equi-spaced intervals. In this respect, both methods combine filtering and regularization in one operator. The difference is that

- **Method A** approximates the prototype filter in the spatial domain, which has been already designed,
- **Method B** approximates the *ideal* response of the prototype filter in the wavenumber domain[1]. In other words, **Method B** also skips the intermediate step of designing the prototype filter.

Compared to existing works, **Method A** has a similar flavor as the approach of [2] since they both interpolate a prototype filter to the actual locations of the receivers. However, the interpolation in [2] is driven by the geometry of the receiver arrays while in this paper, the interpolation is based on the band-limited assumption on the received signal, which is also utilized in [4], [5]. In reality, such a band-limited assumption is often valid, and the corresponding interpolation process can yield a better approximation to the prototype filter response in the wavenumber domain as will be demonstrated in the paper.

---

[1]Wavenumber domain is also known as spatial frequency domain.

II. PROPOSED ALGORITHMS

For simplicity, we only consider filtering along one spatial dimension x. The continuous data is represented by $d(x)$; the data samples gathered at the N locations of the receivers x_j, $j = 0, \ldots, N-1$ are denoted by $d(x_j)$. The nominal locations of the receivers are defined on the grid $\bar{x}_n = n\Delta x$ while the actual locations of the receivers are assumed to lie on the denser grid $\bar{\bar{x}}_m = m\delta x = m(\Delta x/M)$ with M being a positive integer. This is not overly restrictive, since M can be large and the precise receiver locations are known with high, but limited, accuracy. We also define the indicator function $s(\bar{\bar{x}}_m)$ to take the value $s(\bar{\bar{x}}_m) = 1$ when a receiver is present at $\bar{\bar{x}}_m$ and the value zero otherwise. Suppose a prototype LSI FIR filter has already been defined on the nominal grid, whose ith tap is denoted as $h(\bar{x}_i)$. We assume that $h(\bar{x}_i) = 0$ if $\bar{x}_i < 0$ or $\bar{x}_i \geq L_f\Delta x$, where L_f is referred to as the spatial support of the FIR filter.

A. Method A

The filter to be designed in **Method A** is represented as a set of LSV FIR filters, whose filter taps $g_l(\bar{\bar{x}}_m)$ for $m = 0, 1, \ldots, NM-1$ are defined on the dense grid. The spatial support of $g_l(\bar{\bar{x}}_m)$ is the same as that of $h(\bar{x}_i)$, therefore $g_l(\bar{\bar{x}}_m) = 0$ if $\bar{\bar{x}}_m < 0$ or $\bar{\bar{x}}_m \geq L_f\Delta x$ (note that [4] imposes a different FIR constraint on the filter). We desire that the output of one such filter at output location $\check{x}_l$ should be identical as if the prototype filter were applied on uniformly sampled data. In other words,

$$\sum_{n=0}^{N-1} h(\check{x}_l - \bar{x}_n)d(\bar{x}_n) = \sum_{m=0}^{NM-1} g_l(\check{x}_l - \bar{\bar{x}}_m)s(\bar{\bar{x}}_m)d(\bar{\bar{x}}_m). \quad (1)$$

The output locations $\check{x}_l$ lie on the nominal grid, i.e., $\check{x}_l = l\Delta x = lM\delta x$. Utilizing the band-limited assumption on $d(\bar{\bar{x}}_m)$ it can be shown that

$$d(\bar{\bar{x}}_m) \approx \frac{1}{N} \sum_{p=-P}^{P} \left(\sum_{n=0}^{N-1} d(\bar{x}_n)e^{-j\frac{2\pi p}{N\Delta x}\bar{x}_n} \right) e^{j\frac{2\pi p}{N\Delta x}\bar{\bar{x}}_m}, \quad (2)$$

where $N = 2P+1$ (a similar expression can be derived when N is even). We can exchange the order of summation and arrive at

$$d(\bar{\bar{x}}_m) \approx \sum_{n=0}^{N-1} \left(\frac{1}{N} \sum_{p=-P}^{P} e^{j\frac{2\pi p}{N\Delta x}(\bar{\bar{x}}_m - \bar{x}_n)} \right) d(\bar{x}_n),$$

$$\approx \sum_{n=0}^{N-1} \underbrace{\frac{\sin(\frac{\pi}{\Delta x}(\bar{\bar{x}}_m - \bar{x}_n))}{N\sin(\frac{\pi}{N\Delta x}(\bar{\bar{x}}_m - \bar{x}_n))}}_{\text{sincd}(N;\bar{x}_n,\bar{\bar{x}}_m)} d(\bar{x}_n). \quad (3)$$

Substituting (3) in (1) yields

$$\sum_{n=0}^{N-1} h(\check{x}_l - \bar{x}_n)d(\bar{x}_n) \approx$$
$$\sum_{m=0}^{NM-1} g_l(\check{x}_l - \bar{\bar{x}}_m)s(\bar{\bar{x}}_m) \left(\sum_{o=0}^{N-1} \text{sincd}(N;\bar{x}_o,\bar{\bar{x}}_m)d(\bar{x}_o) \right). \quad (4)$$

We rewrite (4) in matrix-vector form as

$$\mathbf{h}_l^H \mathbf{d} \approx \mathbf{g}_l^H \mathbf{SQd} \quad (5)$$

where

$\mathbf{h}_l$ is an $N \times 1$ vector with $h(\check{x}_l - \bar{x}_n)$ as its n-th element;
$\mathbf{g}_l$ is an $NM \times 1$ vector with $g_l(\check{x}_l - \bar{\bar{x}}_m)$ as its m-th element;
$\mathbf{S}$ is an $NM \times NM$ diagonal matrix with $s(\bar{\bar{x}}_m)$ as its m-th diagonal element;
$\mathbf{Q}$ is an $NM \times N$ matrix with $\text{sincd}(N;\bar{x}_o,\bar{\bar{x}}_m)$ as its (m,o)th element and
$\mathbf{d}$ is an $N \times 1$ vector with $d(\bar{x}_n)$ as its n-th element.

A sufficient condition for (5) to hold is

$$\mathbf{h}_l^H \approx \mathbf{g}_l^H \mathbf{SQ},$$

for which a suitable $\mathbf{g}_l$ can be found by solving the following least-squares problem

$$\min_{\mathbf{g}_l} \left\{ ||\mathbf{h}_l^H - \mathbf{g}_l^H \mathbf{SQ}||_2^2 \right\} \equiv \min_{\tilde{\mathbf{g}}_l} \left\{ ||\mathbf{h}_l^H - \tilde{\mathbf{g}}_l^H \tilde{\mathbf{Q}}||_2^2 \right\}, \quad (6)$$

where $\tilde{\mathbf{g}}_l^H$ is formed by removing its elements corresponding to the zeros of $\mathbf{S}$. Similarly, $\tilde{\mathbf{Q}}$ is constructed after removing the rows of $\mathbf{Q}$ corresponding to the zero columns in $\mathbf{S}$. This eliminates $\mathbf{S}$ from (6). In order to limit the spatial support of $\mathbf{g}_l$, we remove the elements of $\tilde{\mathbf{g}}_l$ that correspond to elements m of $\mathbf{g}_l$ for which $m < lM$ or $m \geq (l+L_f)M$ holds, thus forming $\tilde{\tilde{\mathbf{g}}}_l$. The corresponding rows of $\tilde{\mathbf{Q}}$ are removed as well, to form $\tilde{\tilde{\mathbf{Q}}}$. The problem has a closed-form solution given by

$$\tilde{\tilde{\mathbf{g}}}_l^H = \mathbf{h}_l^H \tilde{\tilde{\mathbf{Q}}}^H (\tilde{\tilde{\mathbf{Q}}}\tilde{\tilde{\mathbf{Q}}}^H)^{-1}. \quad (7)$$

A different FIR filter $\mathbf{g}_l$ has to be calculated for each output location $\check{x}_l$. The solution can be seen as a composition of two operations: the first operation is $\mathbf{h}_l^H \tilde{\tilde{\mathbf{Q}}}^H = (\tilde{\tilde{\mathbf{Q}}}\mathbf{h}_l)^H$, which can be interpreted as an interpolation of the filter to the actual locations of the receivers. The second operation $(\tilde{\tilde{\mathbf{Q}}}\tilde{\tilde{\mathbf{Q}}}^H)^{-1}$ deconvolves the effects of nonuniform sampling.

B. Method B

In comparison to **Method A**, **Method B** does not rely on the knowledge of the prototype filter $\mathbf{h}_l$, which needs to be pre-designed. To this end, we will try to approximate the behavior of the prototype filter in the wavenumber domain. As a first step, let us use a circular convolution operator to describe the target LSV FIR filter as well as the prototype filter. Accordingly, (1) should be adapted to the form

$$\sum_{n=0}^{N-1} h(\bar{x}_{(l-n)\text{mod}(N)})d(\bar{x}_n) =$$
$$\sum_{m=0}^{NM-1} g_l'(\bar{\bar{x}}_{(lM-m)\text{mod}(NM)})s(\bar{\bar{x}}_m)d(\bar{\bar{x}}_m), \quad (8)$$

where $g_l'(\bar{\bar{x}}_m)$ stands for the LSV FIR filter from **Method B**. In deriving the above, we have used the assumption that the output locations $\check{x}_l$ lie on the nominal grid, i.e., $\check{x}_l = l\Delta x = lM\delta x$. Just like **Method A**, we require that $g_l'(\bar{\bar{x}}_m)$ have a

spatial support in the interval $[0, \; L_f \Delta x)$. In other words, $g_l'(\bar{\bar{x}}_m) = 0$ if $\bar{\bar{x}}_m < 0$ or $\bar{\bar{x}}_m \geq L_f \Delta x$.

To get rid of the dependence on $h(\bar{x}_i)$ in (8), we resort to the wavenumber domain. Suppose the wavenumber response of the prototype filter is known as $h_w(k)$ with $-\pi \leq k \leq \pi$. If we use $\mathbf{H}_w$ to denote an $N \times N$ diagonal matrix whose nth diagonal element is given by the sample $h_w(k)$ at $k = -\pi + n\frac{2\pi}{N}$ for $n = 0, 1, \ldots, N-1$, then the wavenumber-domain counterpart of (8) can be expressed as

$$\mathbf{H}_w(\mathbf{Fd}) = \mathbf{FG'SQF}^H(\mathbf{Fd}), \qquad (9)$$

where we have coined an $N \times NM$ matrix $\mathbf{G'}$ to represent the circular convolution of the filter with the (l, m)th element of $\mathbf{G'}$ given by $g_l'(\bar{\bar{x}}_{(lM-m)\text{mod}(NM)})$; $\mathbf{F}$ stands for the N-point DFT matrix, and $\mathbf{S}$ and $\mathbf{Q}$ are defined in (5). A sufficient condition for (9) to hold is

$$\mathbf{H}_w \approx \mathbf{FG'SQF}^H. \qquad (10)$$

The right-hand side of (10) can be broken down to three parts:

- $\mathbf{G'}$ is the LSV filter. Its output is defined on the nominal grid and its input on the dense grid.
- $\mathbf{F}$ acts on the columns of $\mathbf{G'}$ to produce the DFT of the LSV filter $g_l(\check{x}_l - \bar{\bar{x}}_m)$ with respect to $\check{x}_l$.
- $\mathbf{SQF}^H$ acts of the rows of $\mathbf{G'}$ to produce the nonuniform DFT of $g_l(\check{x}_l - \bar{\bar{x}}_m)$ with respect to $\bar{\bar{x}}_m$.

The product of $\mathbf{FG'SQF}^H$ is a wavenumber connection matrix[2] [6], which can be viewed as the wavenumber response of the LSV FIR filter. Note that the off-diagonal elements in $\mathbf{FG'SQF}^H$ are in most cases non-zero, which means that the relation between the spectra of the data before and after filtering in the wavenumber domain is not simply an element-wise multiplication for nonuniform sampling.

From (10), we formulate the following least-squares problem

$$\min_{\mathbf{G'}} \{ ||\mathbf{W} \odot (\mathbf{H}_w - \mathbf{FG'SQF}^H)||_F^2 \qquad (11)$$

where $\mathbf{W}$ is a weighting matrix that can apply an individual weight to each element of $\mathbf{H}_w - \mathbf{FG'SQF}^H$ with the symbol $\odot$ denoting the Hadamard (element-wise) product. Including such a weighting matrix is beneficial, for instance, to enable a better trade-off between different approximation errors in the passband, stopband as well as the "do-not-care" (transition) zones.

As in **Method A**, the matrix $\mathbf{S}$ can be eliminated by removing all the columns of $\mathbf{G'}$ and rows of $\mathbf{Q}$ with the same index as the elements of the diagonal of $\mathbf{S}$ that have the value zero. Let $\tilde{\mathbf{G}}'$ and $\tilde{\mathbf{Q}}$ respectively be the reduced form of $\mathbf{G'}$ and $\mathbf{Q}$ resulting from this column- and row-removal. Then (11) can be rewritten as

$$\min_{\mathbf{G'}} \{ ||\mathbf{W} \odot (\mathbf{H}_w - \mathbf{F}\tilde{\mathbf{G}}'\tilde{\mathbf{Q}}\mathbf{F}^H)||\}_F^2 \qquad (12)$$

---

[2]In [6] its continuous counterpart is called "frequency connection function". We use the term "wavenumber connection matrix" for consistency with the rest of the terminology in this paper.

This problem can be restated in its vectorized form as

$$\min_{vec(\mathbf{G'})} \{ || \, vec(\mathbf{W} \odot (\mathbf{H}_w - \mathbf{F}\tilde{\mathbf{G}}'\tilde{\mathbf{Q}}\mathbf{F}^H)||\}_2^2 \equiv$$
$$\min_{vec(\mathbf{G'})} \{ || \, diag(vec(\mathbf{W})) \, vec(\mathbf{H}_w - \mathbf{F}\tilde{\mathbf{G}}'\tilde{\mathbf{Q}}\mathbf{F}^H)||\}_2^2, \quad (13)$$

Here $vec(\mathbf{A})$ returns a column-vector that stacks the columns of the matrix $\mathbf{A}$. The function $diag(\mathbf{v})$ returns a diagonal matrix with the vector $\mathbf{v}$ on its main diagonal.

Using the identity $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \, vec(\mathbf{B})$, where $\otimes$ denotes the Kronecker product, (13) can be rewritten as

$$\min_{vec(\mathbf{G'})} \{ || \, diag(vec(\mathbf{W})) \, vec(\mathbf{H}_w) -$$
$$diag(vec(\mathbf{W}))(\mathbf{F}^*\tilde{\mathbf{Q}}^T \otimes \mathbf{F}) \, vec(\tilde{\mathbf{G}}))||\}_2^2$$

where * denotes the complex conjugate of a matrix. Let

$$\tilde{\mathbf{U}} = diag(vec(\mathbf{W}))(\mathbf{F}^*\tilde{\mathbf{Q}}^T \otimes \mathbf{F})$$
$$\tilde{\mathbf{g}}' = vec(\tilde{\mathbf{G}}').$$

The elements of $\tilde{\mathbf{g}}'$ and rows of $\tilde{\mathbf{U}}$ that should be removed due to the limited spatial support of the filter $g_l'(\bar{x}_n)$ are given by the indexes of those elements of $\tilde{\mathbf{g}}'$ that correspond to the zero elements of $g_l'(\bar{\bar{x}}_m)$. If we call $\tilde{\tilde{\mathbf{g}}}'$ and $\tilde{\tilde{\mathbf{U}}}$ the results after the corresponding row and element removal, the solution is given by

$$\tilde{\tilde{\mathbf{g}}}' = (\tilde{\tilde{\mathbf{U}}}^H\tilde{\tilde{\mathbf{U}}})^{-1}\tilde{\tilde{\mathbf{U}}}^H \, diag(vec(\mathbf{W})) \, vec(\mathbf{H}_w) \qquad (14)$$

$\mathbf{G'}$ can be constructed from the elements of $\tilde{\tilde{\mathbf{g}}}'$ and can be applied to the nonuniformly sampled data.

III. RESULTS

The performance of **Method A** and **Method B** was examined using synthetically created seismic data. A portion of the spectral content of the synthesized data can be found at the lower part of Fig. 1. The reflections from the Earth's subsurface appear mostly on the lower part of the wavenumber spectrum in the range $-0.02m^{-1} \leq k' \leq 0.02m^{-1}$. This region appears highlighted in all figures. All the plots have been smoothed by a 5-tap moving average filter. The peaks found at $k' = \pm 0.055m^{-1}$ are due to waves propagating along the Earth's surface. Their presence is often undesired and should be removed before further processing of the seismic data. Notice that when the data are not uniformly sampled, the wavenumber content appears smeared. This is a well-known side-effect introduced by nonuniform sampling [7].

We generated 100 different realizations of the receiver locations and filtered the nonuniformly sampled data using **Methods A** and **B**. **Method A** approximates the prototype filter whose wavenumber response is given in the upper part of Figure 1. In **Method B**, the prototype filter is not given but is supposed to have a similar passband and stopband region in the wavenumber domain as in **Method A**. The LSV FIR filters resulting form **Method A** and **B** have the same spatial support as that of the prototype filter. The average spectrum of the 100
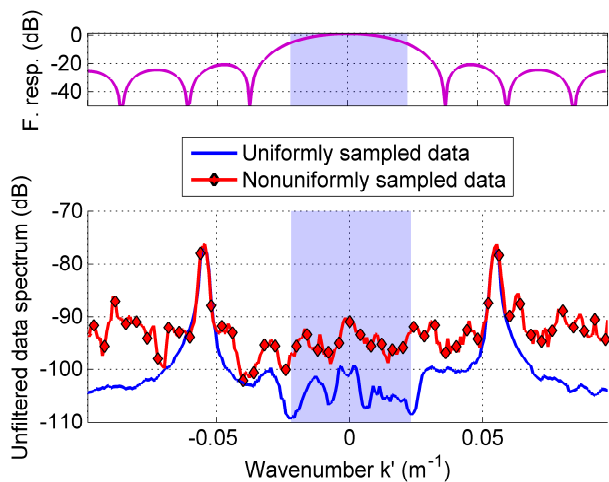
Fig. 1. Upper part: the wavenumber response of the prototype filter. Lower part: the wavenumber content for the data.
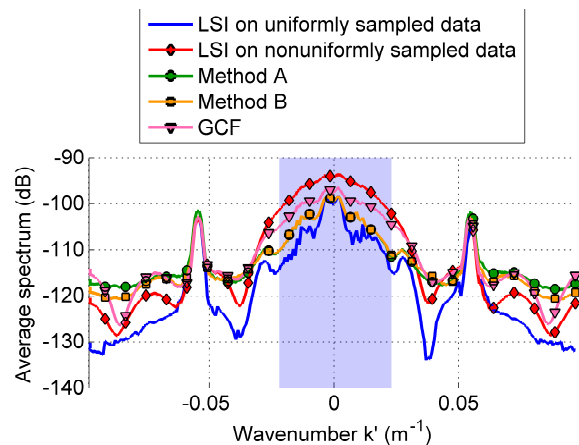

Fig. 2. Average spectrum of the filtered data (over 100 realizations).

Fig. 3. Standard deviation from the average spectrum (over 100 realizations).

realizations can be seen in Fig. 2. In addition to **Methods A** and **B**, an adapted version of the GCF was implemented[3].

In addition, we also compared the proposed methods against 1) the ideal case where the prototype filter is applied on the uniformly sampled and 2) the case where the prototype filter is applied directly on nonuniform data. In the latter case, the spectral leakage induced by the irregularities is most obvious in the passband. In Fig. 2 we see that using the LSI filter on nonuniformly sampled data gives an output that differs almost 15dB at the edges of the passband. In contrast, **Methods A** and **B** give a filtered output that, on average, is closer to the ideal case in the low wavenumbers, exhibiting less than 5dB maximum difference from the uniformly sampled data case (Fig. 2) in the passband region. This is due to the fact that **Methods A** and **B** compensate for irregularities in sampling. The attenuation in the stopband, however, is less when using **Methods A** and **B**. This is because the FIR filters generated for each individual output will not, in general, have zeros at the same locations of their wavenumber responses. This leads to a more flat response in the stopband. GCF is somewhere in the middle, as it interpolates the filter to the locations of the receivers and compensates for sampling density, but does not deconvolve the effects of nonuniform sampling.

The standard deviation of the output spectrum can be seen in Fig. 3. **Methods A** and **B** exhibit in this case a significantly smaller standard deviation in the lower wavenumbers, for example, in Fig. 3, around 10dB lower than using the LSI filter. This means that **Methods A** and **B** may provide an output that is, generally, stable.

IV. CONCLUSION

We proposed two methods for generating LSV FIR filters suitable for filtering nonuniformly sampled data. The resulting filters yield a more accurate output in the passband than simply

---

[3]The method in [2] works on data having two spatial dimensions and is adapted here to the single spatial dimension case.
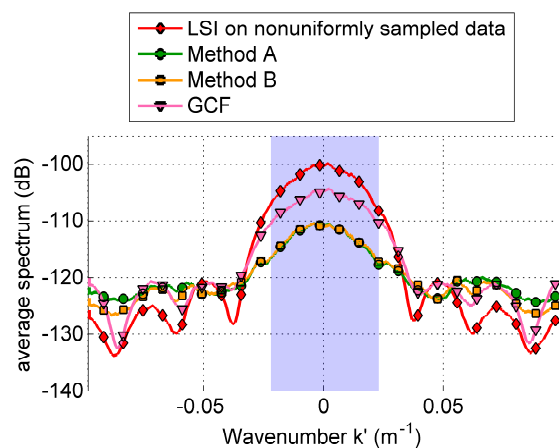
applying an LSI filter directly to nonuniformly sampled data. The output is also more stable, as it varies less for different realizations of the receiver locations.

REFERENCES

[1] P. Newman and J. Mahoney, "Patterns - with a pinch of salt," *Geophysical Prospecting*, vol. 21, no. 2, pp. 197–219, 1973.
[2] A. Özbek and R. Ferber, "Multi-dimensional filtering of seismic data sampled on an irregular grid," in *67th EAGE Conference & Exhibition*. EAGE, 2005.
[3] K. Kose and A. Cetin, "Low-pass filtering of irregularly sampled signals using a set theoretic framework [lecture notes]," *Signal Processing Magazine, IEEE*, vol. 28, no. 4, pp. 117–121, 2011.
[4] H. Johansson and P. Löwenborg, "Reconstruction of nonuniformly sampled bandlimited signals by means of time-varying discrete-time fir filters," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, 2006.
[5] A. Duijndam, M. Schonewille, and C. Hindriks, "Reconstruction of band-limited signals, irregularly sampled along one spatial direction," *Geophysics*, vol. 64, no. 2, pp. 524–538, 1999.
[6] G. Margrave, "Theory of nonstationary linear filtering in the fourier domain," in *1997 SEG Annual Meeting*, 1997.
[7] G. Blacquière and L. Ongkiehong, "Single sensor recording: Anti-alias filtering, perturbations and dynamic range," in *2000 SEG Annual Meeting*, 2000.