

Reconstruction of irregularly sampled and aliased data with linear prediction filters

Mostafa Naghizadeh* and Mauricio Sacchi, University of Alberta

SUMMARY

Linear prediction filters in the f - x domain are often used to interpolate regularly sampled data. We study the problem of reconstructing irregularly sampled data via linear prediction filters methods. For this purpose, we propose a two-stage algorithm for the computation of missing data. First, we reconstruct the unaliased part of the data spectrum using a Fourier method (Minimum Weighted Norm Interpolation). Prediction filters are extracted from the reconstructed low frequency data components, and finally, they are used to reconstruct the aliased part of the spectrum. The applicability of the proposed method is examined using synthetic and field data examples.

INTRODUCTION

Reconstruction of seismic data using statistical approaches is an active area of research in exploration seismology. While many methods are based on statistical estimation theory, they also utilize information from the physics of wave propagation by taking into account relevant a priori information and assumptions. Methods proposed by Spitz (1991), Porsani (1999) and Gulunay (2003) successfully address removing alias from regularly sampled data. These methods utilize low frequency information to recover high frequency data components. Spitz (1991) computed prediction filters (autoregressive operators) from low frequencies to predict interpolated traces at high frequencies. This methodology is applicable only if the original seismic data are regularly sampled in space. Conversely, irregularly sampled data can be reconstructed using Fourier methods. In this case the Fourier coefficients of the irregularly sampled data are retrieved by inverting the inverse Fourier operator with a band limiting (Duijndam et al., 1999) and/or a sparsity constraint (Sacchi et al., 1998; Liu and Sacchi, 2004; Zwartjes and Gisolf, 2006).

We introduce a new strategy that combines the strengths of both prediction error methods and Fourier based methods to cope with the problem of reconstructing non-uniformly sampled, aliased data. The proposed algorithm involves the reconstruction of spatial data at low frequencies. Then, from the reconstructed low frequency portion of the data, a suite of prediction filters is extracted and used to reconstruct the aliased portion of the data in the f - x domain.

THEORY

Problem Definition

Consider a seismic gather containing a finite number of linear events. In addition, we assume that some traces in the gather are missing. By applying the Discrete Fourier Transform (DFT) with respect to time, the gather is transformed to the f - x domain. We let $\mathbf{x}(f)$ be the length- N vector of f - x data sampled on a regular grid $x_1(f), x_2(f), x_3(f), \dots, x_N(f)$, of which only M traces are available. Let the sets of integers $\mathcal{K} = \{k(1), k(2), k(3), \dots, k(M)\}$ and $\mathcal{U} = \{u(1), u(2), u(3), \dots, u(N-M)\}$ indicate the indices of available (known traces) and missing samples (unknown traces), respectively. The goal is to recover $\mathbf{x}_{\mathcal{U}}(f)$ from $\mathbf{x}_{\mathcal{K}}(f)$.

Minimum Weighted Norm Interpolation (MWNI) of unaliased data

Fourier reconstruction methods are well suited to reconstruct seismic data in the low frequency (unaliased) portion of the Fourier spectrum.

In addition, as it was shown by Duijndam et al. (1999), the reconstruction problem is well-conditioned at low frequencies where only a few wave numbers are required to honor the data. This makes the problem well-posed; therefore, it is quite easy to obtain a low frequency spatial reconstruction of the data.

With the previous reasoning in mind, we first proceed to restore the low frequency part of the data using a Fourier reconstruction method. In other words, we estimate the missing samples of $\mathbf{x}(f)$, that is $\mathbf{x}_{\mathcal{U}}(f)$ from $\mathbf{x}_{\mathcal{K}}(f)$ for temporal frequencies $f \in [f_{min_r}, f_{max_r}]$, where f_{min_r} and f_{max_r} denote the minimum and maximum (unaliased) frequencies in the data. In general, due to the band-limited nature of the seismic wavelet, we consider $f_{min_r} > 0$.

Recently, two Fourier-based reconstruction methods were introduced: Band Limited Fourier Reconstruction (BLFR) (Duijndam et al., 1999; Schonewille et al., 2003) and Minimum Weighted Norm Interpolation (MWNI) (Liu and Sacchi, 2004; Sacchi and Liu, 2005; Liu et al., 2004). In our implementation, we have adopted MWNI. It is important to mention, however, that similar results were obtained using BLFR. These methods can retrieve the complex Fourier coefficients of the reconstructed data directly from the observations by inverting the inverse Fourier operator. The non-uniqueness of the reconstruction problem (Sacchi et al., 1998) is circumvented by the incorporation of a constraint in the form of a spectral norm. Details pertaining the MWNI method are discussed in Appendix A.

Multi-step estimation of prediction filters and high-frequency data reconstruction

Let us consider reconstructed data in the band $f \in [f_{min_r}, f_{max_r}]$. In appendix B we show that linear events in the f - x domain can be predicted using Multi-Step Autoregressive (MSAR) operators of the form,

$$x_k(f) = \sum_{j=1}^L P_j(\alpha f) x_{k-\alpha j}(f), \quad k = \alpha L + 1, \dots, N, \quad (1)$$

$$x_k^*(f) = \sum_{j=1}^L P_j(\alpha f) x_{k+\alpha j}^*(f), \quad k = 1, \dots, \alpha N - L. \quad (2)$$

These equations corresponds to a special type of autoregressive (AR) model where forward (equation 1) and backward (equation 2) autoregressive equations are computed by "jumping" α steps at the time. The length of the AR operator is L and $P_j(f)$ is the prediction filter. The parameter $\alpha = 1, 2, \dots, \alpha_{max}$ is the step factor used to extract the prediction filter for frequency αf from frequency f . Since the step factor is a positive integer it is clear that low frequencies provide vital information for our data reconstruction algorithm.

The parameter α_{max} is the upper limit of the step factor in equations 1 and 2. The later depends on the number of traces N , and the length of prediction filter L . This parameter is given by

$$\alpha_{max} = \lfloor \frac{N - \frac{L+1}{2}}{L} \rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the integer part.

Equations (1) and (2) are also considered as a generalization of the prediction filters used by Spitz (1991), in order to incorporate all possible prediction filters into the reconstruction scheme. Multiple prediction filters can be extracted for a given high frequency $f' = \alpha f$. In other words, all possible combinations of α and f leading to the product $f' = \alpha f$ will deliver a prediction filter that can be used to reconstruct data at frequency component f' .

MSAR Reconstruction

To continue with our analysis, a few comments are in order. The MSAR strategy requires us to find prediction filters in the reconstructed band $[f_{min_r}, f_{max_r}]$, where min_r and max_r are discrete frequency indices. We might encounter the case where for some frequency f' , the frequency f'/α may not fall in the reconstructed interval $[f_{min_r}, f_{max_r}]$; and then, the MSAR method will not be applicable. This situation can be solved by extrapolating prediction filters using the method proposed in Appendix B of Spitz (1991). It should be mentioned that this situation can be avoided by properly choosing the values of f_{min_r} and f_{max_r} in such a way that $max_r \geq 2(min_r - 1)$.

So far we have outlined a method to extract, from reconstructed low frequency data components $\mathbf{x}(f)$, prediction filters for high frequency data component $\mathbf{x}(f') = \mathbf{x}(\alpha f)$.

Following the procedure proposed by Wiggins and Miller (1972) and Spitz (1991), one can compute the missing samples from known data and prediction filter coefficients. In this case, the forward and backward autoregressive equations for $\alpha = 1$,

$$x_k(f) = \sum_{j=1}^L P_j(f)x_{k-j}(f), \quad k = L+1, \dots, N, \quad (3)$$

$$x_k^*(f) = \sum_{j=1}^L P_j(f)x_{k+j}^*(f), \quad k = 1, \dots, N-L, \quad (4)$$

are used to isolate the unknown data components. This is done by expanding the last two equations and rewriting them, after a few mathematical manipulations, in matrix form as follows

$$\tilde{\mathbf{A}}(\mathbf{P}(f)) \begin{bmatrix} x_{u(1)}(f) \\ x_{u(2)}(f) \\ \vdots \\ x_{u(N-M)}(f) \end{bmatrix} = \tilde{\mathbf{B}}(\mathbf{P}(f)) \begin{bmatrix} x_{k(1)}(f) \\ x_{k(2)}(f) \\ \vdots \\ x_{k(M)}(f) \end{bmatrix}. \quad (5)$$

The notation $\tilde{\mathbf{A}}(\mathbf{P}(f))$ and $\tilde{\mathbf{B}}(\mathbf{P}(f))$ reflects the fact that these two matrices only depend on the prediction filters. Hence, the missing samples are computed using:

$$\mathbf{x}_{\mathcal{U}}(f) = [\tilde{\mathbf{A}}^*(\mathbf{P}(f))\tilde{\mathbf{A}}(\mathbf{P}(f))]^{-1}\tilde{\mathbf{A}}^*(\mathbf{P}(f))\tilde{\mathbf{B}}(\mathbf{P}(f))\mathbf{x}_{\mathcal{K}}(f), \quad (6)$$

where $\mathbf{x}_{\mathcal{U}}(f)$ and $\mathbf{x}_{\mathcal{K}}(f)$ indicate the vectors of unknown and known data samples, respectively. In addition, $\tilde{\mathbf{A}}^*$ stands for the transpose and complex conjugate of $\tilde{\mathbf{A}}$.

TESTS

Synthetic example

In order to examine the performance of the MSAR reconstruction technique we construct a synthetic data example. The data consist of three linear events. Two of them are severely aliased. In addition we have randomly removed 60% of the traces. Figures 1a and 1b show the original complete data and the data with missing traces, respectively. The data with missing traces were reconstructed using the MWNI method for normalized frequencies in the range 0.035 to 0.5. The result is portrayed in Figure 1c. In addition, we have used MWNI to reconstruct only the low frequency portion of the data from normalized frequencies in the range 0.035 to 0.075. The remaining frequency components were reconstructed with the MSAR technique (0.075 to 0.5), the final result is provided in Figure 1d. Both methods were capable of reconstructing the data. However, the high frequencies were better restored via the combined application of MWNI and MSAR. This is emphasized by Figures 2 where we show the f - k panels of the data portrayed in Figure 1. In particular, one observes that all the high frequency artifacts generated by MWNI (Figure 2b) were attenuated (Figure 2d).

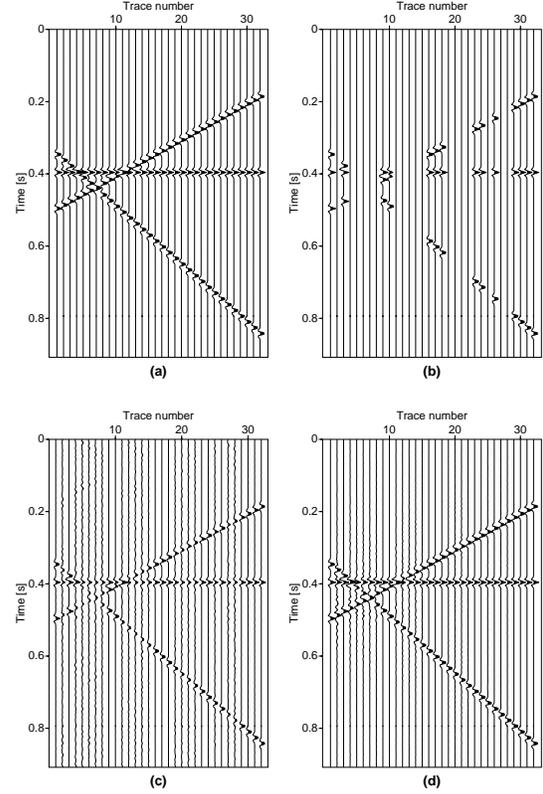


Figure 1: Synthetic example in t - x domain. a) Original data. b) The data with missing traces. c) Reconstructed section using MWNI. d) Reconstructed section using MWNI plus MSAR where MWNI was used to reconstruct data in the normalized frequency band 0.035 – 0.075, the remaining part of the band (0.075 – 0.5) was reconstructed via MSAR.

MSAR Reconstruction

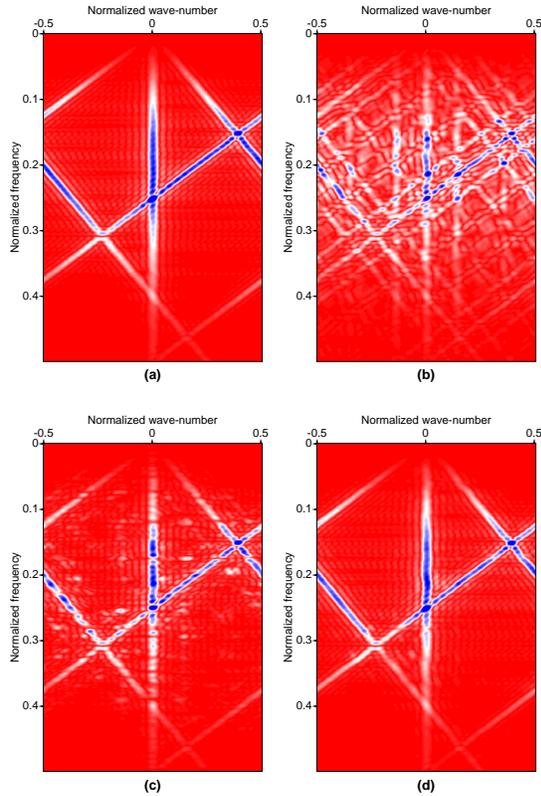


Figure 2: The f - k representation of Figure 1. a) Original data. b) Data with missing traces. c) Reconstructed data via MWNI. d) Reconstructed data via MWNI and MSAR.

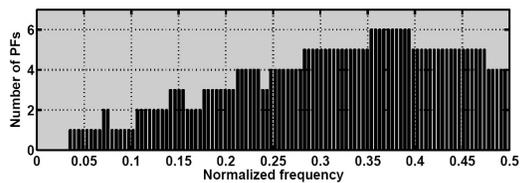


Figure 3: The number of prediction filters contributing to each frequency components in the example in Figure 1d. The average filter for any given frequency is used in the reconstruction stage of the algorithm (equation (6)).

Figure 3 shows the number of prediction filter extracted for each frequency using the MSAR method. Due to lack of information, the very low frequencies were excluded from the reconstruction. Prediction filters for the low frequency end of the data could have been estimated by extrapolation of prediction filters as suggested in Spitz (1991, Appendix B).

Real data example

In order to test the performance of the MSAR reconstruction on a real data set, we apply the technique to the reconstruction of a near offset section from a marine data set from the Gulf of Mexico. Events arising from diffractions on a salt body make the reconstruction difficult for the MWNI method. About 40 % of the traces were removed from the original section (Figure 4a) to simulate a section with missing traces (Figure 4b). The section of missing traces is reconstructed using MWNI (Figure 4c) and MWNI (low frequencies) plus MSAR (high frequencies) (Figure 4d). From a comparison of these figures it is easy to see that the combined application of MWNI and MSAR produce a result where the steeply dipping events are better preserved during reconstruction.

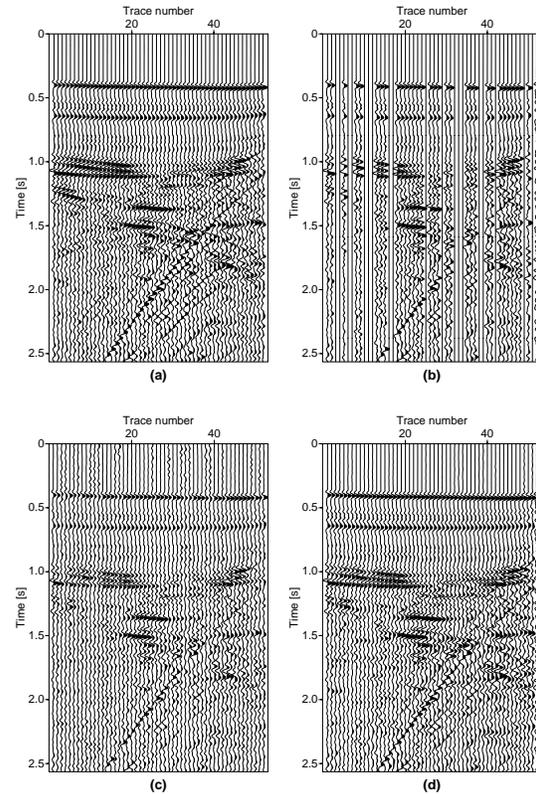


Figure 4: Reconstruction of a near offset section. a) Original section. b) Section with missing traces. c) Reconstructed section using MWNI. d) Reconstructed section using MSAR.

CONCLUSIONS

A method for spatial reconstruction of high frequency data components was introduced. The method involves the cooperative application of a Fourier-based technique (MWNI) to reconstruct the non-aliased part of the data and a Multi-Step Autoregressive (MSAR) algorithm to reconstruct the high frequency and potentially aliased part

MSAR Reconstruction

of the data. The MSAR algorithm relies on extracting information from low frequencies (reconstructed via MWNI) to reconstruct high frequencies. Since the prediction filter of a given high frequency can be computed from more than one reconstructed low frequency, more than one prediction filter can be extracted for a given high frequency. In this case, an average of prediction filters is used to reconstruct the data.

The results of synthetic data reconstruction, and also the real data example show that MSAR is capable of eliminating high frequencies artifacts often encountered when MWNI is used to reconstruct the complete seismic band. The proposed methodology can be applied to regularly sampled sections as well, and used for data interpolation.

Modifying MSAR for multi-dimensional data reconstruction is straightforward. In higher dimensions, the reconstruction of low frequencies using MWNI should be more stable, and as a result the prediction filters can be calculated with high precision, leading to a better reconstruction of high frequencies. Wang (2002) introduced a way to use 2D AR operators to interpolate the regularly sampled data in the f - x - y domain. The same style of AR operators can be used to extend the MSAR method to the 2D case.

ACKNOWLEDGMENTS

This research has been supported by the sponsors of the Signal Analysis and Imaging Group, at the University of Alberta and the National Sciences and Engineering Research Council of Canada via a Discovery Grant to MDS. MN would also like to acknowledge financial support in the form of a scholarship from Institute of Petroleum Engineering at the University of Tehran. We also wish to thank Sam Kaplan who reviewed this material.

APPENDIX A

MINIMUM WEIGHTED NORM INTERPOLATION (MWNI)

Interpolation of band-limited data with missing samples can be summarized in the following inversion scheme:

$$\text{Minimize } \|\mathbf{x}\|_{\mathcal{W}}^2 \quad \text{Subject to } \mathbf{G}\mathbf{x} = \mathbf{y} \quad (\text{A-1})$$

where $\|\cdot\|_{\mathcal{W}}^2$ indicates a specific weighted norm and \mathbf{G} is the sampling matrix which maps all data samples to available samples. Its transpose, \mathbf{G}^T , fills the position of missing samples with zeros. A regularization norm can be selected in the wave-number domain as follows:

$$\|\mathbf{x}\|_{\mathcal{W}}^2 = \sum_{k \in \mathcal{K}} \frac{X_k^* X_k}{W_k^2}. \quad (\text{A-2})$$

Here, X_k indicates the coefficients of the Fourier transform of the vector of spatial data \mathbf{x} . The values of W_k determine the type of interpolation. For Band-limited Minimum Weighted Norm Interpolation a diagonal matrix is defined as:

$$\Upsilon_k = \begin{cases} W_k^2 & k \in \mathcal{K} \\ 0 & k \notin \mathcal{K} \end{cases}, \quad (\text{A-3})$$

where \mathcal{K} indicates the region of support of the Fourier transform. The pseudoinverse of Υ is defined as:

$$\Upsilon_k^{-1} = \begin{cases} W_k^{-2} & k \in \mathcal{K} \\ 0 & k \notin \mathcal{K} \end{cases} \quad (\text{A-4})$$

For Band-limited Minimum Norm Interpolation, the values of W_k are equal to one, while for Minimum Weighted Norm Interpolation, their

values must be iteratively updated to find an optimal reconstruction. The minimizer of the cost function (A-1) is given by

$$\hat{\mathbf{x}} = \mathbf{F}^H \Upsilon \mathbf{F} \mathbf{G}^T (\mathbf{G} \mathbf{F}^H \Upsilon \mathbf{F} \mathbf{G}^T + \alpha \mathbf{I})^{-1} \mathbf{y} \quad (\text{A-5})$$

where \mathbf{F} is the Fourier matrix, α is trade-off parameter, \mathbf{I} is the identity matrix, while T and H stand for transpose and Hermitian operators, respectively. For further details see Liu (2004) and Liu and Sacchi (2004).

APPENDIX B

MULTI-STEP AUTOREGRESSIVE OPERATOR

In this appendix we provide a proof for the MSAR theory. Interested readers can find further details about prediction filters and their properties in Spitz (1991). A seismic section with linear events can be represented in the f - x domain as:

$$S(m\Delta x, n\Delta f) = \sum_{k=1}^L A_k e^{-i2\pi(n\Delta f)(m\Delta x) \cdot p_k}, \quad (\text{B-1})$$

where Δf and Δx are frequency and spatial sampling intervals, respectively. In addition, p_k and A_k are the slope and amplitude of each linear event, respectively. This means that each linear event, for a monochromatic frequency component f , can be represented as complex harmonic in the f - x domain. Now consider the case with $\Delta x' = \alpha \Delta x$ and $\Delta f' = \frac{\Delta f}{\alpha}$. In this case it is easy to show that

$$S(m\Delta x', n\Delta f') = S(m\alpha\Delta x, n\frac{\Delta f}{\alpha}) \quad (\text{B-2})$$

In addition, one can show that a superposition of L harmonics can be represented by an autoregressive (AR) model of the form:

$$S(m\Delta x, n\Delta f) = \sum_{j=1}^L P(j, n\Delta f) S((m-j)\Delta x, n\Delta f) \quad (\text{B-3})$$

Similarly if we use $\Delta f'$ and $\Delta x'$, we obtain:

$$S(m\Delta x', n\Delta f') = \sum_{j=1}^L P'(j, n\frac{\Delta f}{\alpha}) S((m-j)\alpha\Delta x, n\frac{\Delta f}{\alpha}). \quad (\text{B-4})$$

A comparison of expressions (B-2), (B-3) and (B-4) leads to the following expression

$$P'(j, n\frac{\Delta f}{\alpha}) = P(j, n\Delta f), \quad j = 1, 2, \dots, L. \quad (\text{B-5})$$

which is the basis for the MSAR reconstruction method.

It can also be shown that there exist predictability properties for each component of the prediction filter on the frequency axis. This means that if the prediction filters are known for some frequencies, one can find the prediction filter for other frequencies by applying prediction operators to the prediction filter components. More succinctly, one can find the prediction filters of prediction filters. For further details see Spitz (1991, Appendix B).

MSAR Reconstruction

REFERENCES

- Duijndam, A., M. Schonewille, and C. Hindricks, 1999, Reconstruction of band-limited signals, irregularly sampled along one spatial direction: *Geophysics*, **64**, 524–538.
- Gulunay, N., 2003, Seismic trace interpolation in the fourier transform domain: *Geophysics*, **68**, 355–369.
- Liu, B., 2004, Multi-dimensional reconstruction of seismic data: PhD thesis, University of Alberta.
- Liu, B. and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: *Geophysics*, **69**, 1560–1568.
- Liu, B., M. D. Sacchi, and D. Trad, 2004, Simultaneous interpolation of 4 spatial dimensions: 74th Annual International Meeting, Soc. of Expl. Geophys., 2009–2012.
- Porsani, M., 1999, Seismic trace interpolation using half-step prediction filters: *Geophysics*, **64**, 1461–1467.
- Sacchi, M. and B. Liu, 2005, Minimum weighted norm wavefield reconstruction for ava imaging: *Geophysical Prospecting*, **53**, 787–801.
- Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete fourier transform: *IEEE Transaction on Signal Processing*, **46**, 31–38.
- Schonewille, M., R. Romijn, A. Duijndam, and L. Ongkiehong, 2003, A general reconstruction scheme for dominant azimuth 3d seismic data: *Geophysics*, **68**, 2092–2105.
- Spitz, S., 1991, Seismic trace interpolation in the F - X domain: *Geophysics*, **56**, 785–796.
- Wang, Y., 2002, Seismic trace interpolation in the f - x - y domain: *Geophysics*, **67**, 1232–1239.
- Wiggins, R. and S. Miller, 1972, New noise-reduction technique applied to long-period oscillations from the alaska earthquake: *Bull. Seism. Soc. Am.*, **62**, 471–479.
- Zwartjes, P. and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: *Geophysics*, **71**, V171–V186.