TIME-UPDATING IAA-BASED SPECTRAL ESTIMATES WITH MISSING SAMPLES USING DATA INTERPOLATION

G. O. Glentis, A. Jakobsson†, and K. Angelopoulos**

*Dept. of Science and Technology of Telecommunications, University of Peloponnese, Greece
†Dept. of Mathematical Statistics, Lund University, Sweden

ABSTRACT

In this work, we propose a computationally efficient time-updating algorithm for estimating the spectral content of a signal with missing samples. The algorithm extends earlier work on the topic by formulating a data-interpolation scheme reducing the required complexity to a fraction of the earlier efficient implementation, without resulting in any noticeable loss of performance for even a quite large degree of missing samples.

Index Terms**—** Spectrum estimation theory and methods, Iterative Adaptive Approach (IAA), fast algorithms

1. INTRODUCTION

The problem of estimating the spectral content of signals with missing samples is commonly occurring in a wide range of applications, where samples may be lost or unmeasurable for a variety of reasons, such as sensor failure or due to different forms of disturbances, and the topic has attracted a steady interest over the recent decades (see, e.g. [1–6], and the references therein). The developed methods range from classical estimation methods such as the Lomb-Scargle periodogram, which is able to handle irregularly sampled data [7, 8], to high-resolution data-adaptive algorithms such as MAPES [5] and MIAA [6], where both the latter are formed under the assumption that the missing samples share the same spectral content as the given samples. Of these, the MIAA algorithm is based on the iterative adaptive approach (IAA) [9], which has been shown to be able to outperform conventional data-adaptive techniques, and, as a result, has since been expanded to a variety of topics (see, e.g., [10–13] and the references therein). Regrettably, the IAA method is computationally cumbersome, and notable efforts have been made in the recent literature to form computationally efficient batch as well as time-recursive (TR) implementations [11–17]. In this work, we examine the TR spectral estimation of data sets suffering from missing samples. In order to formulate the problem, let $y(n) \in \mathbb{C}$ represent a data sequence of observations for which one wishes to compute the spectral estimate, whereof subsets of length N may be assumed to be (reasonably) stationary, and let

$$\mathbf{y}_N(n) = \left[\begin{array}{ccc} y(n-N+1) & \ldots & y(n) \end{array} \right]^T \qquad (1)$$

where $(\cdot)^T$ denotes the transpose. In $\mathbf{y}_N(n)$, M_n and $\bar{M}_n$ samples are assumed to given and missing, at time n, respectively, such that the actually available measurements only forms an M_n-dimensional subset of $\mathbf{y}_N(n)$. In [6], a (batch) MIAA estimator of data on this form was presented, for which an efficient implementation was later introduced in [12]. In [16], a computationally efficient sliding-window TR magnitude squared coherence (MSC) spectral estimator was then developed for data of this form. We here term the resulting estimator, when used to form just the spectral estimate of a single signal, the fast TR-MIAA (FTR-MIAA) algorithm. In this work, we extends and improve on these works, forming a computationally efficient data interpolation formulation of the TR sliding window updating of the MIAA algorithm. This is done by exploiting interpolation of the already reconstructed and given samples in $\mathbf{y}_N(n)$ to reconstruct the most recent sample, $y(n)$, unless given, using the time domain interpolation approach introduced in [6] as a method for missing data reconstruction from the given data and the estimated MIAA spectrum. As compared to the FTR-MIAA algorithm, the presented data-interpolation algorithm allows for a notable reduction in the computational complexity, which is independent of the amount of missing data. The resulting algorithm has a complexity being proportional to $\mathcal{O}(N^2)$, which typically is significantly less than $\mathcal{O}\left(M_n^3\right)$, which is the complexity of the alternative available methods.

2. TIME-UPDATING THE MIAA ALGORITHM USING DATA INTERPOLATION

By supposing that all but the most recent sample, $y(n)$, are either given or have been replaced by some form of reconstruction estimates, the thus reconstructed data vector, $\hat{\mathbf{y}}_N(n)$, can be expressed as

$$\hat{\mathbf{y}}_N(n) = \left[\begin{array}{cc} \hat{\mathbf{y}}_{N-1}^T(n-1) & y(n) \end{array} \right]^T \qquad (2)$$

This work was supported in part by the Swedish Research Council and Carl Trygger's foundation.

where $y(n)$ denotes the most recent either given or missing sample, and

$$\hat{\mathbf{y}}_{N-1}(n-1) = \mathbf{T}_{N-1,N}\hat{\mathbf{y}}_N(n-1) \tag{3}$$

with

$$\mathbf{T}_{N-1,N} = \left[\begin{array}{cc} \mathbf{0}_{N-1,1} & \mathbf{I}_{N-1} \end{array}\right] \tag{4}$$

where $\mathbf{0}_{k,1}$ and $\mathbf{I}_k$ denote a $k \times 1$ zero and $k \times k$ identity matrix, respectively. In the case that $y(n)$ is actually given, the time-recursive IAA estimate of $\hat{\mathbf{y}}_N(n)$ (which in this case have all given or reconstructed samples) may then be formed directly using the TR-IAA updating developed in [12], as

$$\alpha_n(\omega_k) = \frac{\mathbf{f}_N^H(\omega_k)\mathbf{R}_N^{-1}(n-1)\hat{\mathbf{y}}_N(n)}{\mathbf{f}_N^H(\omega_k)\mathbf{R}_N^{-1}(n-1)\mathbf{f}_N(\omega_k)} \tag{5}$$

$$\Phi_n(\omega_k) = |\alpha_n(\omega_k)|^2 \tag{6}$$

$$\mathbf{R}_N(n) = \sum_{k=0}^{K-1} \Phi_n(\omega_k)\mathbf{f}_N(\omega_k)\mathbf{f}_N^H(\omega_k) \tag{7}$$

where

$$\mathbf{f}_N(\omega_k) = \left[\begin{array}{cccc} 1 & e^{\jmath\omega_k} & \ldots & e^{\jmath\omega_k(N-1)} \end{array}\right]^T \tag{8}$$

Alternatively, when $y(n)$ is not available, the time-recursive IAA estimate may instead be formed on only the available $N-1$ dimensional time window. In this case, the resulting estimate may be obtained as

$$\alpha_n(\omega_k) = \frac{\mathbf{f}_{N-1}^H(\omega_k)\mathbf{R}_{N-1}^{-1}(n-1)\hat{\mathbf{y}}_{N-1}(n-1)}{\mathbf{f}_{N-1}^H(\omega_k)\mathbf{R}_{N-1}^{-1}(n-1)\mathbf{f}_{N-1}(\omega_k)} \tag{9}$$

$$\Phi_n(\omega_k) = |\alpha_n(\omega_k)|^2 \tag{10}$$

$$\mathbf{R}_{N-1}(n) = \sum_{k=0}^{K-1} \Phi_n(\omega_k)\mathbf{f}_{N-1}(\omega_k)\mathbf{f}_{N-1}^H(\omega_k) \tag{11}$$

The current, missing, sample, $y(n)$, should then also be interpolated from the previous estimates, which may be done using the time-domain *MIAA-t* missing data interpolation method proposed in [6], such that

$$\hat{y}(n) = \sum_{k=0}^{K-1} \Phi_n(\omega_k)\mathbf{f}_{N-1}^H(\omega_k)\mathbf{R}_{N-1}^{-1}(n)\hat{\mathbf{y}}_{N-1}(n)e^{\jmath\omega_k(N-1)} \tag{12}$$

The resulting combined algorithm, hereafter termed the *Time Recursive Interpolation MIAA* (TRI-MIAA) algorithm is constructed by the TR updating of (5)–(7) in the case when the current sample is known, or by the TR updating of (9)–(12), in the case when the sample is missing, noting that initialization should be performed by a complete MIAA-t step being applied for spectral estimation and missing data interpolation. In contrast to the FTR-MIAA algorithm introduced in [16], where $M_n \leq N$ given samples are used at each time instant, requiring the manipulation of non-structured matrices, and thus resulting in a TR updating scheme with a complexity of $\mathcal{O}\left(M_n^3\right)$ operations per update, the proposed TRI-MIAA time updating algorithm requires the estimation and the inversion of (well-structured) Toeplitz matrices, formed using (7) and (11), respectively. As a result, the proposed TRI-MIAA spectral estimation algorithm allows for a fast implementation along the lines of [11, 12, 15] where, thanks to the Toeplitz structure of the involved data covariance matrices, the matrix inverse products as well as trigonometric polynomials associate to these matrix inverses may all be efficiently computed exploiting the Gohburg-Semencul (GS) representation of the involved Toeplitz matrices using the celebrated Levinson-Durbin algorithm. However, the updating may be formed even more efficiently by only exploiting and propagating the displacement of $\mathbf{R}_N(n)$ for both the representation of $\mathbf{R}_N(n)$ and $\mathbf{R}_{N-1}(n)$. By keeping the required displacement generator vectors at a fixed size equal to $N \times 1$, one avoids the monitoring of the vector sizes and the then required up- or down-dating for the case when the sample is given or missing, respectively, as well as allows for a lower complexity approximative gradient-based or preconditioning conjugate gradient (PCG) based updating of the displacement generator vectors reminiscent to the ones developed in [14, 15]. With this observation, we proceed to detail the resulting efficient implementation.

3. EFFICIENT IMPLEMENTATION OF THE TRI-MIAA ALGORITHM

The initialization of the proposed implementation exploits the standard MIAA algorithm, which maybe implemented efficiently as detailed in [12,13] at a cost of $m\left(M_0^3 + K\log_2(K)\right)$ operations, where M_0 is the number of the given data at the first data block, $\hat{\mathbf{y}}_N(0)$, and m is the number of iterations required for convergence (typically, $m = 10 - 15$ iterations are adequate for the convergence of the MIAA algorithm). In the case when the most recent sample, $y(n)$, is given, the spectral estimate and the updating of the covariance matrix may be performed along the lines of [11,12,15], where it was shown that $\mathbf{R}_N(n)$ is a $N \times N$ Toeplitz matrix which can be extracted from an increased dimension circulant matrix as

$$\sum_{k=0}^{K-1} \Phi_n(\omega_k)\mathbf{f}_K(\omega_k)\mathbf{f}_K^H(\omega_k) = \begin{bmatrix} \mathbf{R}_N(n) & \times \\ \times & \times \end{bmatrix} \tag{13}$$

where $\times$ denotes terms of no relevance. Given the first column of $\mathbf{R}_N(n)$, $\mathbf{R}_N^{-1}(n)$ may be represented using a GS representation of the form

$$\begin{aligned} \mathbf{R}_N^{-1}(n) &= \mathcal{L}\left(\mathbf{t}_N^1(n)\right)\mathcal{L}^H\left(\mathbf{t}_N^1(n)\right) \\ &- \mathcal{L}\left(\mathbf{t}_N^2(n)\right)\mathcal{L}^H\left(\mathbf{t}_N^2(n)\right) \end{aligned} \tag{14}$$

where $\mathcal{L}\left(\mathbf{x}_N\right) \triangleq \left[\mathbf{x}_N \; \mathbf{Z}_N\mathbf{x}_N \; \ldots \; \mathbf{Z}_N^{N-1}\mathbf{x}_N\right]$ is a Krylov matrix, and $\mathbf{Z}_N$ is the down shifting operator with ones on the diagonal below the main diagonal, and zeros elsewhere, and where $\mathbf{t}_N^1(n)$ is the power normalized forward predictor defined as

$$\mathbf{R}_N(n)\mathbf{t}_N^1(n) = \mathbf{e}_N^1\sqrt{\mathbf{e}_N^{1T}\mathbf{t}_N^1(n)} \qquad (15)$$

with $\mathbf{e}_N^1$ denoting a $N \times 1$ vector with one in the first element and zeros elsewhere, and where

$$\mathbf{t}_N^2(n) = \mathbf{Z}_N\left(\mathbf{J}_N\mathbf{t}_N^1(n)\right)^* \qquad (16)$$

with $\mathbf{J}_N$ denoting the exchange matrix, and $(\cdot)^*$ the complex conjugate. Then, the TR spectral estimate may be computed in terms of trigonometric polynomials as [11, 12, 15]

$$\alpha_n(\omega_k) = \frac{\psi_n(\omega_k)}{\varphi_n(\omega_k)} \qquad (17)$$

$$\psi_n(\omega_k) \triangleq \mathbf{f}_N^H(\omega_k)\mathbf{d}_N(n) \qquad (18)$$

$$\varphi_n(\omega_k) \triangleq \mathbf{f}_N^H(\omega_k)\mathbf{R}_N^{-1}(n-1)\mathbf{f}_N(\omega_k) \qquad (19)$$

where $\mathbf{d}_N(n) \triangleq \mathbf{R}_N^{-1}(n-1)\hat{\mathbf{y}}_N(n)$ is computed using (14), whereas the coefficients of the trigonometric polynomial $\varphi_n(\omega_k) = \sum_{\ell=-N+1}^{N-1} c_n^\ell e^{\jmath\ell\omega_k}$ are computed in terms of $\mathbf{t}_N^1(n-1)$ and $\mathbf{t}_N^2(n-1)$ as proposed in [18], without the need of forming the matrix $\mathbf{R}_N^{-1}(n-1)$ explicitly, using triangular Toeplitz matrix products of the form

$$\begin{bmatrix} c_n^{N-1} \\ \vdots \\ c_n^0 \end{bmatrix} = \mathcal{L}(\mathbf{t}_N^1(n-1))\mathbf{D}_N\mathbf{t}_N^{1*}(n-1) - \mathcal{L}(\mathbf{t}_N^2(n-1))\mathbf{D}_N\mathbf{t}_N^{2*}(n-1) \qquad (20)$$

noting that $c_n^{-\ell} = \left(c_n^\ell\right)^*$, and where $\mathbf{D}_N$ is a $N \times N$ anti-diagonal matrix of the form

$$\mathbf{D}_N = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 2 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ N & 0 & \dots & 0 \end{bmatrix}. \qquad (21)$$

In the alternative situation, when $y(n)$ is missing, the efficient implementation should instead be formed using a time window of size $N-1$. In order to avoid complicating the updating by propagating both $\mathbf{t}_N^1(n)$ and $\mathbf{t}_{N-1}^1(n)$, as indicated by (5)-(7) and (9)-(12), we here only make use of the former, in conjunction with the rearrangement of the computation tasks required in the missing data case, by noting that

$$\begin{bmatrix} \mathbf{R}_{N-1}^{-1}(n) & \mathbf{0}_{N-1} \\ \mathbf{0}_{N-1}^T & 0 \end{bmatrix} = \mathcal{L}\left(\mathbf{t}_N^1(n)\right)\mathcal{L}^H\left(\mathbf{t}_N^1(n)\right) - \mathcal{L}\left(\hat{\mathbf{t}}_N^2(n)\right)\mathcal{L}^H\left(\hat{\mathbf{t}}_N^2(n)\right) \qquad (22)$$

where $\hat{\mathbf{t}}_N^2(n) \triangleq \left(\mathbf{J}_N\mathbf{t}_N^1(n)\right)^*$ is used instead of (16), implying that

$$\mathbf{d}_{N-1}(n) \triangleq \mathbf{R}_{N-1}^{-1}(n-1)\hat{\mathbf{y}}_{N-1}(n-1) \qquad (23)$$

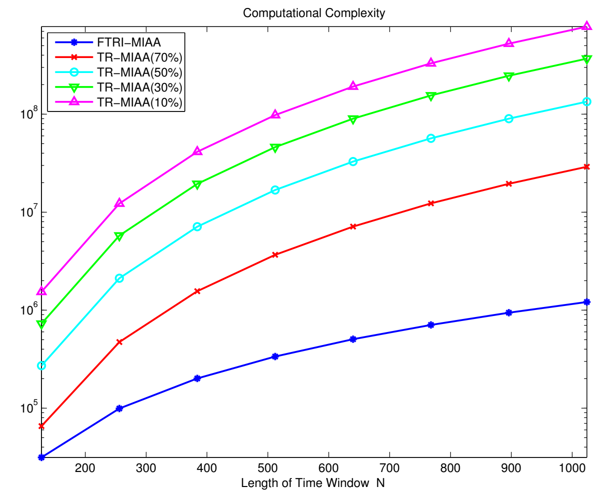

Fig. 1. Computational complexity per updating for the proposed FTRI-MIAA algorithm and the TR-MIAA algorithm implemented using [12, 13], for 10% up to 70% missing data, for a time window length varying from $N = 128$ up to $N = 1024$, with the number of frequency points set equal to $K = 5N$.

required by the spectral estimation recursions in the missing data case can be computed as

$$\begin{bmatrix} \mathbf{d}_{N-1}(n) \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{N-1}^{-1}(n-1) & \mathbf{0}_{N-1} \\ \mathbf{0}_{N-1}^T & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{y}}_{N-1}(n-1) \\ \times \end{bmatrix}$$

followed by a similar treatment for

$$\tilde{\mathbf{d}}_{N-1}(n) \triangleq \mathbf{R}_{N-1}^{-1}(n)\hat{\mathbf{y}}_{N-1}(n-1)$$

required by (12) for the reconstruction of missing sample. Finally, the coefficients of the trigonometric polynomial associated to $\mathbf{R}_{N-1}^{-1}(n)$, required to form the spectral estimator part in the missing data case,

$$\tilde{\varphi}_n(\omega_k) \triangleq \mathbf{f}_N^H(\omega_k)\mathbf{R}_{N-1}^{-1}(n-1)\mathbf{f}_N(\omega_k) = \sum_{\ell=-N+2}^{N-2} \tilde{c}_n^\ell e^{\jmath\ell\omega_k}$$

are estimated as

$$\begin{bmatrix} \tilde{c}_n^{N-1} \\ \vdots \\ \tilde{c}_n^0 \end{bmatrix} = \mathcal{L}(\mathbf{t}_N^1(n-1))\mathbf{D}_N\mathbf{t}_N^{1*}(n-1) - \mathcal{L}(\hat{\mathbf{t}}_N^2(n-1))\mathbf{D}_N\hat{\mathbf{t}}_N^{2*}(n-1) \qquad (24)$$

where it is noted that $\hat{\mathbf{t}}_N^2(n-1)$ is used instead of $\mathbf{t}_N^2(n-1)$, and that by construction the highest degree coefficient is algebraically $c_n^{N-1} = 0$. The computational complexity of the resulting fast TRI-MIAA (FTRI-MIAA) algorithm is $\mathcal{O}(N^2)$ per updating. Figure 1 illustrates the resulting computational speed-up as compared to the FTR-MIAA algorithm for various levels of missing samples.

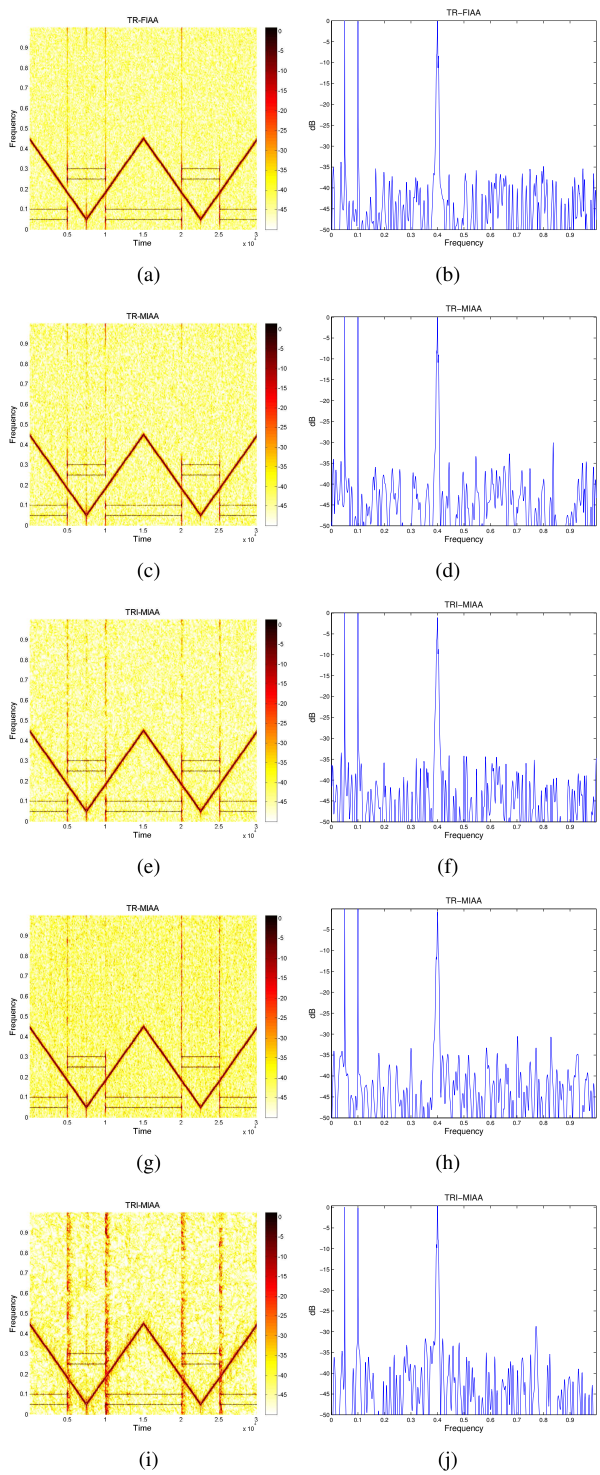

Fig. 2. Time-recursive spectrograms are shown in the first column as (a) TR-IAA on the complete data, (c) FTR-MIAA, and (e) FTRI-MIAA, when 30% of the data are missing, and (g) FTR-MIAA, and (i) FTRI-MIAA, when 60% of the data are missing. Snapshots at time $n = 14000$ are shown in the second column.

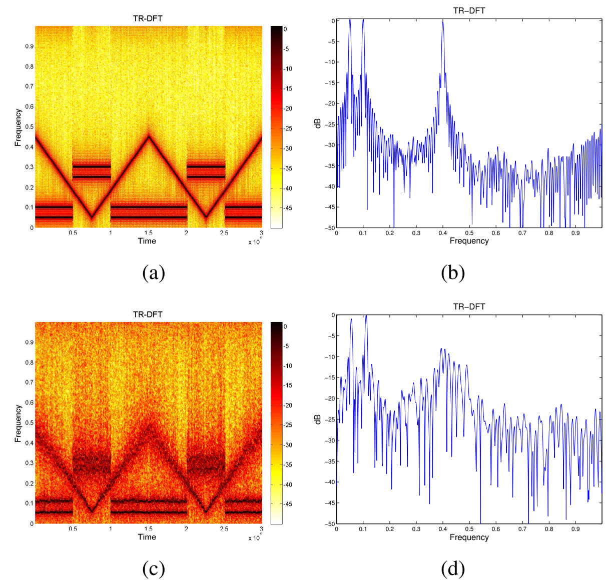

Fig. 3. Time recursive spectral estimation using the DFT. Spectrograms are shown in the first column as (a) TR-DFT on the complete data, and (c) TR-DFT when 10% of the data are missing. Snapshots at time $n = 14000$ are shown in the second column.

4. SIMULATIONS

To evaluate the performance of the proposed algorithm, the spectrum of a time-varying signal consisting of a mixtures of two complex sinusoids with abruptly changing frequencies and a complex-valued linear chirp with descending/ascending linear frequency variations, being corrupted by an additive zero-mean complex Gaussian noise (see [15] for a detailed description of the signal) is computed using the DFT spectrogram, the FTR-MIAA algorithm [16], and the proposed FTRI-MIAA algorithm. Figures 2 and 3 show the resulting performance, with 2(a) giving the reference spectrogram as obtained from the TR-IAA algorithm [15], given the complete data set. Figures 2(c) and (e) show the FTR-MIAA and FTRI-MIAA estimates, respectively, when 30% of the samples have randomly (with a uniform distribution) been omitted from the signal. Similarly, Figures 2(g) and (i) show the corresponding estimates when 60% of the samples are missing. Both cases clearly illustrate the similar performance of the FTR-MIAA and the proposed algorithm. The computational complexities of the proposed method is here about 3% and 12% of the complexity of the FTR-MIAA algorithm, respectively. As a comparison, the DFT spectrogram is shown in Figures 3(a) and (c) for the complete signal, and for the case when 10% of the samples are missing, clearly illustrating the preferable performance of the IAA-based algorithms. In the figures, the corresponding snapshots at time $n = 14000$ are presented in the second column.

5. REFERENCES

[1] B. Porat and B. Friedlander, "ARMA spectral estimation of time series with missing observations," *IEEE Trans. Inform. Theory*, vol. 30, no. 4, pp. 601–602, July 1986.

[2] F. A. Marvasti, Ed., *Nonuniform Sampling: Theory and Practice*, Kluwer Academic/Plenum Publishers, 2001.

[3] P. Broersen, S. de Waele, and R. Bos, "Estimation of autoregressive spectra with rarandom missing data," in *Proc. 20th IEEE Intrument. Measure. Technol. Conf.*, Vail, CO, May 2003, vol. 2, pp. 1154–1159.

[4] Y. Wang, J. Li, and P. Stoica, *Spectral Analysis of Signals - The Missing Data Case*, Morgan & Claypool, 2005.

[5] Y. Wang, P. Stoica, J. Li, and T. L. Marzetta, "Nonparametric spectral analysis with missing data via the EM algorithm," *Digital Signal Processing*, vol. 15, no. 2, pp. 191–206, 2005.

[6] P. Stoica, J. Li, and J. Ling, "Missing Data Recovery via a Nonparametric Iterative Adaptive Approach," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 241–244, April 2009.

[7] N. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophys. Space Sci.*, pp. 447–462, 1976.

[8] J. D. Scargle, "Studies in astronomical time series analysis II: Statistical aspects of spectral analysis of unevenly spaced data," *Astrophys. J.*, vol. 263, pp. 835–853, Dec. 1982.

[9] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source Localization and Sensing: A Nonparametric Iterative Approach Based on Weighted Least Squares," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 425–443, January 2010.

[10] E. Gudmundson, A. Jakobsson, J. A. Jensen, and P. Stoica, "Blood Velocity Estimation Using Ultrasound and Spectral Iterative Adaptive Approaches," *Signal Process.*, vol. 91, no. 5, pp. 1275–1283, May 2011.

[11] M. Xue, L. Xu, and J. Li, "IAA Spectral Estimation: Fast Implementation using the Gohberg-Semencul Factorization," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3251 – 3261, July 2011.

[12] G.-O. Glentis and A. Jakobsson, "Efficient Implementation of Iterative Adaptive Approach Spectral Estimation Techniques," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4154–4167, Sept. 2011.

[13] D. Vu, L. Xu, M. Xue, and J. Li, "Nonparametric Missing Sample Spectral Analysis and Its Applications to Interrupted SAR," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 1–14, Feb. 2012.

[14] G.-O. Glentis and A. Jakobsson, "Superfast Approximative Implementation of the IAA Spectral Estimate," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 472–478, Jan. 2012.

[15] G.-O. Glentis and A. Jakobsson, "Time-Recursive IAA Spectral Estimation," *IEEE Signal Processing Letters*, vol. 18, no. 2, pp. 111–114, Feb. 2011.

[16] K. Angelopoulos, G. O. Glentis, and A. Jakobsson, "Efficient Time Recursive Coherence Spectrum Estimation," in *European Signal Processing Conference*, Bucharest, August 27-31 2012.

[17] K. Angelopoulos, G. O. Glentis, and A. Jakobsson, "Computationally Efficient Sparsity-Inducing Coherence Spectrum Estimation of Complete and Non-Complete Data Sets," *Elsevier Signal Processing*, to appear.

[18] G.-O. Glentis, "A Fast Algorithm for APES and Capon Spectral Estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4207–4220, Sept. 2008.