

On Linearly Precompressed Non-Parametric Spectrum Estimation

Jens P. Elsner, Martin Braun, Leonid Chaichenets and Friedrich K. Jondral
 Universität Karlsruhe (TH), Germany, {elsner, braun, chaichenets, fj}@int.uni-karlsruhe.de

Abstract—Extending the derivation of a maximum likelihood spectrum estimator by Stoica and Sundin, non-parametric spectrum estimation with linear precompression is introduced. For real-time applications, linear precompression allows for a scalable trade-off between data rate and accuracy of the estimate. Precompression is based on linear projection with a compression matrix formed from sequences with perfect periodic autocorrelation. This basis has the property of preserving the power spectrum. The derived estimator is verified with numerical simulations.

I. INTRODUCTION

Traditionally, digital spectrum estimation is based on equidistant sampling of a time-continuous signal above the Nyquist frequency. In certain applications such as broadband real-time spectrum estimation for spectrum analyzers or cognitive radios [1] sampling frequency or computational resources are limited, but one would still like to estimate the frequency content of the signal in a very wide frequency band. Random subsampling is one approach to solve this problem, but spectrum reconstruction uses iterative algorithms, which are computationally involved - see e.g. [2], [3]. A major advantage of these algorithms is that they deal with missing data, which can occur in numerous applications. However, if the focus lies on data reduction by deliberately discarding data prior to analysis, linear precompression with a circulant matrix is an efficient alternative to reduce the sampling rate¹ at the expense of estimator variance. Linear precompression in itself is simple and well-known, but, to the best of the authors' knowledge, its application to spectrum estimation has not yet been reported in literature.

The paper is structured as follows. In Section II non-parametric spectrum estimation with linear precompression is introduced as a modified approximate maximum likelihood estimate, restating and elaborating on results from Stoica and Sundin [5]. Section III briefly describes perfect sequences and gives a simple construction for compression matrices. Two families of sequences, ternary and polyphase sequences, are compared with respect to suitability for precompression. Simulation results are presented in Section IV. Section V concludes.

II. LINEARLY PRECOMPRESSED NON-PARAMETRIC SPECTRUM ESTIMATION

In contrast to parametric approaches, a non-parametric spectrum estimator assumes no prior knowledge of the underlying

¹This approach bears similarities with the recently established research field Compressed Sensing [4]. In contrast, the focus here does not lie on detection or reconstruction of the original signal, but on spectrum estimation.

model structure and estimates the power spectral density (PSD) function directly. In [5], Stoica and Sundin derive a non-parametric spectrum estimator based on a small set of assumptions. This derivation² is extended to linear precompression.

A. Maximum Likelihood Estimation

Let $y(t)|_{t=T,2T,\dots}$ denote a complex-valued stationary band-limited random process with equidistant time instances, and let $\Phi(\omega)$ denote its frequency-continuous PSD function. The random observation vector $\mathbf{y} \in \mathbb{C}^N$ with N observations is $\mathbf{y} = (y(1), \dots, y(N))^T$. The PSD maximum likelihood estimator is derived under the following assumptions:

- 1) The data vector \mathbf{y} has a circular Gaussian distribution with zero mean and covariance matrix \mathbf{R} .
- 2) The PSD function $\Phi(\omega)$ is piecewise constant on M equidistant intervals ϕ_k and strictly positive.
- 3) M is a divisor of N : for any given M , there is an integer L such that $LM = N$.
- 4) The data vector \mathbf{y} of dimension N is precompressed into the compressed data vector $\tilde{\mathbf{y}}$ of dimension C via the projection matrix $\mathbf{A} \in \mathbb{C}^{C \times N}$: $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{y}$.
- 5) The precompression matrix \mathbf{A} is row-orthogonal, i.e., $\mathbf{A}\mathbf{A}^* = \mathbf{I}$ holds.
- 6) \mathbf{A} is also approximately column-orthogonal, i.e., $\mathbf{A}^*\mathbf{A} \approx \mathbf{I}$ holds.

Assumptions 1), 2) and 3) are taken directly from [5]; 4) and 5) specify linear precompression. Assumptions 5) and 6) allow for the approximation

$$(\mathbf{A}\mathbf{X}\mathbf{A}^*)^{-1} \approx \mathbf{A}\mathbf{X}^{-1}\mathbf{A}^*, \quad (1)$$

where \mathbf{X} is any invertible $(N \times N)$ -matrix. (1) will be used as an identity in the following; it is approximately fulfilled for the precompression matrices constructed in Section III³.

Gaussian statistics and stationarity are preserved under the linear transform \mathbf{A} . Let the covariance matrix after transformation be denoted by $\tilde{\mathbf{R}}$. The relationship between the covariance matrices \mathbf{R} and $\tilde{\mathbf{R}}$ can be expressed as

$$\tilde{\mathbf{R}} = E\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^*\} = E\{(\mathbf{A}\mathbf{y})(\mathbf{A}\mathbf{y})^*\} = \mathbf{A}\mathbf{R}\mathbf{A}^*. \quad (2)$$

²In their paper, Stoica and Sundin derive popular spectrum estimators, e.g., the Thomson multitaper method, from a maximum likelihood viewpoint. The extension to linear precompression is straightforward, so some steps of the derivation are omitted here for brevity.

³An error bound depending on \mathbf{A} is subject to research.

The log-likelihood function of the data vector $\tilde{\mathbf{y}}$ has to be maximized, which is equivalent to minimizing

$$\begin{aligned} f &= \log |\tilde{\mathbf{R}}| + \tilde{\mathbf{y}}^* \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{y}} \\ &= \log |\mathbf{A}\mathbf{R}\mathbf{A}^*| + \tilde{\mathbf{y}}^* (\mathbf{A}\mathbf{R}\mathbf{A}^*)^{-1} \tilde{\mathbf{y}} \\ &\stackrel{(1)}{=} \log |\mathbf{A}\mathbf{R}\mathbf{A}^*| + \tilde{\mathbf{y}}^* \mathbf{A}\mathbf{R}^{-1} \mathbf{A}^* \tilde{\mathbf{y}}. \end{aligned} \quad (3)$$

The relationship between \mathbf{R} and $\Phi(\omega)$ can be expressed as [5]

$$\mathbf{R} = E\{\mathbf{y}\mathbf{y}^*\} = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{a}(\omega)\mathbf{a}(\omega)^* \Phi(\omega) d\omega \quad (4)$$

with $\mathbf{a}(\omega) = (e^{j\omega}, \dots, e^{jN\omega})^T$. Using assumption 2) and $\beta = 1/M$, this can be rewritten as

$$\begin{aligned} \mathbf{R} &= \sum_{k=1}^M \phi_k \frac{1}{2\pi} \int_{2\pi(k-1)\beta}^{2\pi k\beta} \mathbf{a}(\omega)\mathbf{a}(\omega)^* d\omega \\ &= \sum_{k=1}^M \phi_k \mathbf{D}_k \mathbf{\Gamma} \mathbf{D}_k^* \end{aligned} \quad (5)$$

with

$$\mathbf{\Gamma} = \frac{1}{2\pi} \int_{-\pi\beta}^{\pi\beta} \mathbf{a}(\omega)\mathbf{a}(\omega)^* d\omega \quad (6)$$

and

$$\mathbf{D}_k = \text{diag}(e^{j2\pi\beta(k-1/2)}, \dots, e^{jN2\pi\beta(k-1/2)}) \quad (7)$$

where $\mathbf{D} = \text{diag}(\mathbf{x})$ with $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ denotes a diagonal matrix with entries $\mathbf{D}_{ii} = x_i$.

Minimizing the objective function (3) yields the maximum likelihood estimator under the given assumptions. To make estimation computationally tractable, two approximations will be introduced in the following.

$\mathbf{\Gamma}$ can be well approximated by eigenvectors associated with the first L largest eigenvalues (which are approximately of equal magnitude) [5] and hence be written as

$$\mathbf{\Gamma} \approx \mathbf{U}\mathbf{U}^* \quad (8)$$

for an appropriately chosen matrix $\mathbf{U} \in \mathbb{C}^{N \times L}$. With this approximation and (3), (5) can be rewritten as a matrix multiplication:

$$\begin{aligned} \mathbf{R} &= \sum_{k=1}^M \phi_k \mathbf{D}_k \mathbf{U}\mathbf{U}^* \mathbf{D}_k^* \\ &= \mathbf{W}\mathbf{\Phi}\mathbf{W}^* \end{aligned} \quad (9)$$

with

$$\mathbf{\Phi} = \text{diag}(\underbrace{\phi_1, \dots, \phi_1}_{L \text{ times}}, \dots, \underbrace{\phi_M, \dots, \phi_M}_{L \text{ times}}) \quad (10)$$

and the block matrix

$$\mathbf{W} = (\mathbf{D}_1 \mathbf{U} \cdots \mathbf{D}_M \mathbf{U}). \quad (11)$$

Note that approximately [5]

$$\mathbf{W}\mathbf{W}^* = \mathbf{I}. \quad (12)$$

Using (9) the objective function (3) is modified to

$$\begin{aligned} f &= \log |\mathbf{A}(\mathbf{W}\mathbf{\Phi}\mathbf{W}^*)\mathbf{A}^*| + \tilde{\mathbf{y}}^* \mathbf{A}(\mathbf{W}\mathbf{\Phi}\mathbf{W}^*)^{-1} \mathbf{A}^* \tilde{\mathbf{y}} \\ &= \log |(\mathbf{A}\mathbf{W})\mathbf{\Phi}(\mathbf{A}\mathbf{W})^*| + ((\mathbf{A}\mathbf{W})^* \tilde{\mathbf{y}})^* \mathbf{\Phi}^{-1} (\mathbf{A}\mathbf{W})^* \tilde{\mathbf{y}}. \end{aligned} \quad (13)$$

To minimize, (13) is differentiated with respect to ϕ_k . Using the identity $\frac{d}{d\phi_k} \log |\mathbf{X}(\phi_k)| = \text{tr}(\mathbf{X}^{-1} \frac{d}{d\phi_k} \mathbf{X})$:

$$\begin{aligned} 0 &= \text{tr}(((\mathbf{A}\mathbf{W})\mathbf{\Phi}(\mathbf{A}\mathbf{W})^*)^{-1} (\mathbf{A}\mathbf{W}) \mathbf{I}_k (\mathbf{A}\mathbf{W})^*) + \\ &\quad \frac{-1}{\phi_k^2} (\mathbf{W}^* \mathbf{A}^* \tilde{\mathbf{y}})^* \mathbf{I}_k (\mathbf{W}^* \mathbf{A}^* \tilde{\mathbf{y}}). \end{aligned} \quad (14)$$

with

$$\mathbf{I}_k = \text{diag}(\underbrace{0, 0, \dots, 0}_{(k-1)L \text{ times}}, \underbrace{1, 1, \dots, 1}_{L \text{ times}}, \underbrace{0, 0, \dots, 0}_{(M-k)L \text{ times}}). \quad (15)$$

Solving (14) for ϕ_k is achieved by applying (1), which transforms the inverse part of the trace into $((\mathbf{A}\mathbf{W})\mathbf{\Phi}(\mathbf{A}\mathbf{W})^*)^{-1} = (\mathbf{A}\mathbf{W})\mathbf{\Phi}^{-1}(\mathbf{A}\mathbf{W})^*$ and applying assumption 2) and (12). The compressed ML estimator is thus

$$\hat{\phi}_k = \frac{1}{L_k} (\mathbf{W}^* \mathbf{A}^* \tilde{\mathbf{y}})^* \mathbf{I}_k (\mathbf{W}^* \mathbf{A}^* \tilde{\mathbf{y}}) \quad (16)$$

where

$$L_k = \text{tr}((\mathbf{A}\mathbf{W})\mathbf{I}_k(\mathbf{A}\mathbf{W})^*). \quad (17)$$

The values of L_k are constant and can be calculated a-priori.

For the uncompressed case $C = N$, this reduces to the result of Stoica and Sundin [5], since then $L_k = L$ for all k :

$$\hat{\phi}_k = \frac{1}{L} (\mathbf{W}^* \mathbf{y})^* \mathbf{I}_k (\mathbf{W}^* \mathbf{y}). \quad (18)$$

B. Cramér-Rao Bound

Since the introduction of the compression matrix will likely lead to an increased estimation variance, the calculation of the Cramér-Rao bound (CRB) in [5] shall also be generalized for the compressed case. From the Slepian-Bangs formula [5], the inverse of the $(M \times M)$ CRB matrix \mathbf{P}_{cr} is given element-wise by

$$\mathbf{P}_{\text{cr}}^{-1}_{i,j} = \text{tr}(\tilde{\mathbf{R}}^{-1} (\frac{\partial}{\partial \phi_i} \tilde{\mathbf{R}}) \tilde{\mathbf{R}}^{-1} (\frac{\partial}{\partial \phi_j} \tilde{\mathbf{R}})). \quad (19)$$

Using (1), (2), (9) and (12),

$$\tilde{\mathbf{R}}^{-1} = \mathbf{A}\mathbf{W}\mathbf{\Phi}^{-1}(\mathbf{A}\mathbf{W})^* \quad (20)$$

and

$$\frac{\partial}{\partial \phi_k} \tilde{\mathbf{R}} = \frac{\partial}{\partial \phi_k} \mathbf{A}\mathbf{R}\mathbf{A}^* = \mathbf{A}\mathbf{W}\mathbf{I}_k\mathbf{W}^*\mathbf{A}^*. \quad (21)$$

From approximation (12) and assumption 5) follows $\mathbf{A}\mathbf{W}(\mathbf{A}\mathbf{W})^* = \mathbf{I}$, which simplifies (19) after inserting (20) and (21) to

$$\mathbf{P}_{\text{cr}}^{-1}_{i,j} = \text{tr}((\mathbf{A}\mathbf{W})\mathbf{\Phi}^{-2}\mathbf{I}_i(\mathbf{A}\mathbf{W})^*) \delta_{ij}. \quad (22)$$

The CRB for each of the estimations $\hat{\phi}_k$ is thus

$$\text{CRB}(\hat{\phi}_k) = \frac{\phi_k^2}{L_k}. \quad (23)$$

As above, for the uncompressed case $C = N$, $L_k = L$ and (23) reduces to the result of Stoica and Sundin [5]. Since in general $L_k < L$ for the compressed case, the minimum possible estimator variance increases.

III. CONSTRUCTION OF COMPRESSION MATRICES FROM SEQUENCES WITH PERFECT PERIODIC AUTOCORRELATION

For linear precompression assumptions 5) and 6) need to be fulfilled. This is the case for compression matrices formed from sequences with perfect periodic autocorrelation.

A. Sequences with perfect periodic autocorrelation

A sequence $\{a_n\}$ with perfect periodic autocorrelation and period N , normalized to unit energy, satisfies [6]

$$R_a(\tau) = \sum_{n=0}^{N-1} a_n a_{n+\tau}^* = \begin{cases} 1 & : \tau = 0 \\ 0 & : \tau \neq 0 \end{cases} \quad (24)$$

where the index is to be interpreted modulo N . Such a perfect sequence has a constant discrete periodic spectrum, which follows directly from the Wiener-Khinchin relationship between the periodic autocorrelation and its Fourier transform [6] and (24):

$$\text{DFT}(R_a(\tau)) = |\text{DFT}(a_n)|^2 = 1. \quad (25)$$

The construction of perfect sequences is non-trivial. Two constructions of sequences suited for implementation of pre-compression, ternary and polyphase sequences⁴, are cited in the following. For further constructions, refer to, e.g., surveys by Fan and Darnell [6] or Lüke et al. [7] and references therein.

Perfect Ternary Sequences: Perfect ternary sequences are perfect sequences with a ternary alphabet. Of special interest are sequences with $a_n \in \{-1, 0, 1\}$. Ipatov constructs such perfect ternary sequences from a (not necessarily binary) maximum length sequence (m-sequence) $\{b_n\}$ with $b_n \in \mathbb{F}_q$ [6]. The period of such an m-sequence is $q^m - 1$. Any nonzero element of \mathbb{F}_q can be expressed as a power of a primitive element α . Let m be odd. Furthermore, let $q = p^s$, where p is an odd prime and s is an integer. Then

$$a_n = \begin{cases} 0 & : b_n = 0 \\ \frac{1}{\sqrt{E}}(-1)^{u+n} & : b_n = \alpha^u \end{cases} \quad (26)$$

is a perfect ternary sequence with period $N = \frac{q^m - 1}{q - 1}$, normalized to its resulting energy E .

⁴Unfortunately perfect sequences with small phase alphabet are rare. The longest known perfect binary sequence is $\{a_n\} = \{1, 1, 1, -1\}$, which means $N = 4$ [7]. The longest known quaternary perfect sequence yields $N = 16$ [7]. This is obviously too short for precompression.

Perfect Polyphase Sequences: Perfect polyphase sequences are perfect sequences with $a_n = e^{j\beta_n}$. The following construction is due to Zadoff and Chu [6]. Let M be an integer coprime to N . Then, with $0 \leq n < N$,

$$a_n = \begin{cases} \frac{1}{\sqrt{N}} e^{j\frac{\pi M}{N} n^2} & : N \text{ even} \\ \frac{1}{\sqrt{N}} e^{j\frac{\pi M}{N} (n+1)n} & : N \text{ odd} \end{cases} \quad (27)$$

is a normalized perfect sequence of length N .

B. Compression matrices for spectrum estimation and implementation notes

Let a_n be a normalized perfect sequence. Construct \mathbf{A} as follows:

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{N-1} \\ a_{N-1} & a_0 & a_1 & \cdots & a_{N-2} \\ a_{N-2} & a_{N-1} & a_0 & \cdots & a_{N-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{N-(C-1)} & \cdots & \cdots & a_{N-C+1} & a_{N-C} \end{pmatrix}. \quad (28)$$

Such a circulant matrix \mathbf{A} with rows of time-shifted perfect sequences fulfills assumption 5), $\mathbf{A}\mathbf{A}^* = \mathbf{I}$, as the rows of \mathbf{A} are orthogonal.

Another important property of the constructed compression matrix \mathbf{A} is preservation of the PSD. Any circulant matrix is diagonalized by the Fourier matrix \mathbf{F} , and, more specifically [8]

$$\mathbf{A} = \mathbf{F} \text{diag}(\mathbf{F}\mathbf{A}) \mathbf{F}^{-1} \quad (29)$$

holds. The eigenvectors of the projection are hence given by the columns of the Fourier matrix and the eigenvalues are given by the DFT of the first row of \mathbf{A} . They have, due to (25), constant amplitude. Hence, the projection introduces only a phase shift, the PSD is preserved.

Precompression can be implemented efficiently in hardware as time shifted preintegration as is discussed in [1]. Perfect ternary sequences are a good choice due to the small number of phases and ease of implementation in analog hardware. Using digital signal processing, other sequences can also easily be implemented - perfect polyphase sequences are an obvious choice as they are available for arbitrary N .

IV. SIMULATION RESULTS

The results from a compressed spectrum estimation can be seen in Figure 1. For better comparison with uncompressed estimators, the same input signal as in [5] was used. Compression is based on an Ipatov sequence of length $N = 1057$. The uncompressed case $C = N$ is also shown (dotted line); for a non-parametric estimator this can be considered the best possible estimation. For $C = 300$ (solid line), which indicates a compression to 28%, the estimated spectrum is close enough to the uncompressed case to use it for applications such as energy detection. Both results are averages over 10000 trials, the dashed lines indicate one standard deviation from the compressed estimator.

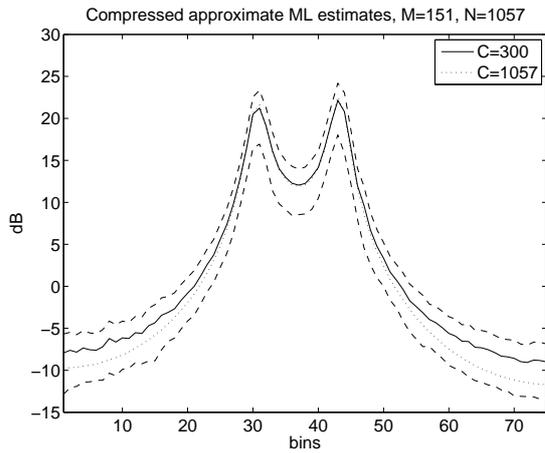


Fig. 1. Comparison between compressed and uncompressed spectrum estimates. Spectrum is generated from an ARMA (4,4) process with $A(q^{-1}) = 1 - 0.2q^{-1} + 1.61q^{-2} - 0.19q^{-3} + 0.8556q^{-4}$ and $C(q^{-1}) = 1 - 0.21q^{-2} + 0.25q^{-4}$.

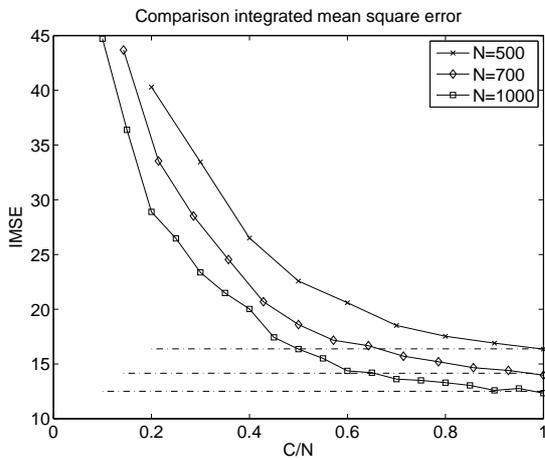


Fig. 2. Mean performance of compression factors C/N for different numbers of observation N .

To quantify the effects of the compression, the integrated normalized mean square error (INMSE) is introduced:

$$\text{INMSE} = \frac{1}{M} \sum_k E \left[\frac{(\hat{\phi}_k - \phi_k)^2}{\phi_k^2} \right]. \quad (30)$$

This measure compares the estimation to the real spectrum. Figure 2 shows the INMSE for different compression rates and different block lengths. Here, Zadoff-Chu sequences were used to facilitate the scaling of N . As seen in the figure, the INMSE converges towards the respective uncompressed error (dashed line). It is worth noting that even for high compression rates, high and low energy regions can be distinguished with high reliability.

V. CONCLUSION

Based on the maximum likelihood approach of Stoica and Sundin, it was shown that linear precompression with

compression matrices formed from perfect sequences offers a simple trade-off between data rate and accuracy in spectrum estimation. Key properties of the proposed compression matrix are row-orthogonality and constant amplitude of eigenvalues. An estimator with linear precompression can be useful in real-time applications to reduce the amount of data before estimating the PSD. This gives a new degree of freedom when analysing the approximate frequency content of non-stationary signals. The time window length can be chosen independently of the numbers of samples used for spectrum estimation; in other words the time-window length can be changed without changing the estimator.

REFERENCES

- [1] J. P. Elsner, M. Braun, H. Jäkel, and F. K. Jondral, "Compressed Spectrum Estimation for Cognitive Radios," in *Proceedings of 19th Virginia Tech Symposium on Wireless Communications*, June 2009.
- [2] Y. Wang, P. Stoica, and T. L. Marzetta, "Nonparametric spectral analysis with missing data via the em-algorithm," *Elsevier Digital Signal Processing*, vol. 15, 2004.
- [3] E. Gudmundson, A. Jakobsson, and S. Somasundaram, "On the reconstruction of gapped sinusoidal data," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2008.
- [4] E. J. Candès and M. B. Wakin, "An Introduction To Compressive Sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, March 2008.
- [5] P. Stoica and T. Sundin, "On Nonparametric Spectral Estimation," *Circuits, Systems and Signal processing*, Springer, vol. 18, no. 2, 1999.
- [6] P. Z. Fan and M. Darnell, "The synthesis of perfect sequences," *Lecture notes in Computer Science, Cryptography and Coding*, Springer, 1995.
- [7] H. Luke, H. Schotten, and H. Hadinejad-Mahram, "Binary and quadriphase sequences with optimal autocorrelation properties: a survey," *IEEE Transactions on Information Theory*, vol. 49, no. 12, Dec. 2003.
- [8] R. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, 2006.