Efficient Capon-based approach exploiting temporal windowing for electric network frequency estimation

George Karantaidis and Constantine Kotropoulos

Department of Informatics,
Aristotle University of Thessaloniki, Greece
gkarantai@csd.auth.gr, costas@csd.auth.gr

1. Overview

- The proposed Electric Network Frequency (ENF) extraction framework:
  1. introduces an efficient approach based on the filter-bank Capon spectral estimator with **temporal windowing**;
  2. exploits the Toeplitz structure of the covariance matrix to obtain the **Gohberg-Semencul factorization** of the inverse covariance matrix by means of Krylov matrices for fast matrix inversion;
  3. employs the **Parzen window** at the stage preceding spectral estimation for the first time.
- Parzen window is also employed within the Short-Time Fourier Transform (STFT) method.
- The impact of different temporal windows is studied.
- The proposed approach employing the Parzen window is compared against state-of-the-art methods for ENF extraction.
- We study whether pairwise differences between the maximum correlation coefficient delivered by proposed fast Capon ENF estimation method employing temporal windows and that of other state-of-the-art ENF estimation methods are statistically significant.
- Experiments are conducted on two real-life datasets.

2. Window Selection and Estimation Procedure

- Window selection was not thoroughly investigated within ENF estimation, where the rectangular window has been used exclusively as a temporal window.
- Here, we employ temporal windowing. It is shown through extensive experiments that the selection of window function pays off.
- Proper window selection provides finer spectral resolution and boosts the accuracy of frequency estimation.
- The N-point Parzen window is defined as:

$$w(n) = \begin{cases} 1 - 6\left(\frac{|n|}{N/2}\right)^2 + 6\left(\frac{|n|}{N/2}\right)^3 & 0 \le |n| \le (N-1)/4 \\ 2\left(1 - \frac{|n|}{N/2}\right)^3 & (N-1)/4 \le |n| \le (N-1)/2. \end{cases} \quad (1)$$

- The general scheme for ENF estimation proposed in [1] was followed employing the parametrization suggested in [2].
  1. The raw signal is properly filtered. A sharp zero-phase band-pass FIR filter with $C_1 = 1001$ and $C_2 = 4801$ coefficients was applied around the 3rd harmonic of the signal recorded from power mains and around the 2nd harmonic of the speech signal, respectively.
  2. A tight band-pass frequency range of 0.1 Hz is employed. The frequency is estimated per frame.
  3. Between consecutive frames, there exists 1 sec shift, which is translated to 441 samples for the downsampled sampling frequency of $441\,Hz$.
  4. Each frame is then multiplied by a temporal window.
  5. The maximum periodogram value of each frame, which corresponds to an approximate ENF estimation $\omega_{q_{\max}}$ is obtained.
  6. A quadratic interpolation is employed and a quadratic model is fit to the logarithm of the estimated power spectrum [3].
- In order to evaluate the accuracy of the estimated frequencies, one employs the maximum correlation coefficient between the estimated ENF and the ground truth one.

3. Datasets

- The first dataset, denoted as Data 1, was recorded by connecting an electric outlet directly to the internal sound card of a desktop computer.
- The second one, denoted as Data 2, comprises of a speech recording captured by the internal microphone of a laptop computer.
- The original datasets were sampled at $44.1\,kHz$ using 16 bits per sample. Afterwards, the raw recordings were downsampled at $441\,Hz$, using proper anti-aliasing filtering.
- Apart from the fundamental frequency of $60\,Hz$, the second and the third harmonics were also maintained to perform experiments.
- Regarding the first dataset, only the third harmonic was used, because it provides the best results compared to the other two.
- In the speech recording, the second harmonic was used, because the other two suffered from extremely low SNR.

4. Proposed Method

- The periodogram can be interpreted as a filter bank approach, which uses a band-pass filter whose impulse response is given by the standard Fourier transform vector $\left[1, e^{-i\omega}, \dots, e^{-i(N-1)\omega}\right]^T$.
- The Capon method, is another filter bank approach based on a data-dependent filter. The Capon spectral estimate is given by:

$$\hat{\phi}(\omega) = \frac{m+1}{\mathbf{a}^*(\omega)\,\hat{\mathbf{R}}^{-1}\,\mathbf{a}(\omega)}. \quad (2)$$

$\hat{\mathbf{R}}$ is an estimate of the auto-covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N-m}\sum_{t=m+1}^{N} \begin{bmatrix} \tilde{y}(t) \\ \vdots \\ \tilde{y}(t-m) \end{bmatrix} \left[\tilde{y}^*(t), \dots, \tilde{y}^*(t-m)\right]. \quad (3)$$

where $\tilde{y}(t) = y(kF_s + t - 1)\, w(t - \frac{N+1}{2})$, $t = 1, 2, \dots, N$, F_s is the sampling frequency, and $k \in \mathbb{N}$. Here, $m = 10$ and $N = L\,F_s$, where L is the frame length in sec.

- Eq. (2) is computed for dense frequency samples $\omega_q = \frac{2\pi q}{Q}$, $q = 0, 1, \dots, Q-1$ with $Q = 4N$ every sec. ENF is estimated by the angular frequency sample $\omega_{q_{\max}}$ where the Power Spectral Density (PSD) of Eq. (2) attains a maximum for $q \in [0, \frac{Q}{2} - 1]$ and $f_k = \frac{\omega_{q_{\max}}}{2\pi} F_s$. The Capon method has been found to be able to resolve fine details of PSD, making it a superior alternative of periodogram-based methods for ENF estimation.
- The proposed approach exploits the Toeplitz structure of the covariance matrix $\mathbf{R}_N$ in order to reduce the computational complexity of the inversions included in the process of Capon spectral estimation using GS factorization [4].
- It is not limited to Toeplitz structures, but is expanded to low displacement rank matrices, where the displacement representation of any square matrix $\mathbf{A}$ is defined as [5]:

$$\nabla_{\mathbf{D}_N, \mathbf{D}_N^T}\mathbf{A} = \mathbf{A} - \mathbf{D}_N \mathbf{A} \mathbf{D}_N^T \quad (4)$$

with $\mathbf{D}_N$ being a lower triangular matrix.

6. Experimental Evaluation

- The proposed approach was compared against state-of-the-art approaches.
- A systematic study was also carried out in order to examine the impact of the window on both datasets. Four different windows along with four different frame lengths were employed.
- A correlation coefficient of 0.9990 is obtained for the proposed approach, when a frame length of 20 sec is employed. This value exceeds the state-of-the-art linear prediction estimation.
- When shorter frame lengths of 5 and 10 sec are employed, the accuracy gets higher and reaches 0.9991, while for the other methods accuracy drops as the frame length gets shorter.

Table 1: Correlation coefficient for various frame lengths - Data 1

Frame length (in sec)	1	5	10	20
Proposed with Parzen window	0.9990	0.9991	0.9991	0.9990
ML [6]	0.8826	0.9852	0.9953	0.9977
Linear Prediction [7]	0.9651	0.9959	0.9976	0.9984
Welch[2]	0.9847	0.9989	0.9989	0.9983
Weighted Spectrogram [7]	0.8255	0.9873	0.9944	0.9966

- The Parzen window yields the highest accuracy among approaches and is not affected by the frame length at all.

Table 2: Correlation coefficient for various windows - Data 1

Frame length (in sec)	1	5	10	20
Parzen	0.9990	0.9991	0.9991	0.9990
Hamming	0.9989	0.9991	0.9990	0.9988
Kaizer	0.0086	0.0495	0.0438	0.9976
Rectangular	0.0047	0.0798	0.0689	0.9975

- Our approach outperforms the existing ML approach and the high resolution MUSIC method.
- It is lagging behind the pure linear prediction method without the additional denoising procedure with respect to the correlation coefficient for about 0.0015.

Table 3: Correlation coefficient for various frame lengths - Data 2

Frame length (in sec)	10	33
Proposed with rectangular window	0.8663	0.9351
ML [6]	0.9059	0.9319
Linear Prediction [7]	0.9213	0.9366
MUSIC [2]	0.9087	0.9318
Weighted Spectrogram [7]	0.8787	0.9125

Table 4: Correlation coefficient for various windows - Data 2

Frame length (in sec)	5	10	20	33
Parzen	0.7063	0.7773	0.8413	0.8785
Hamming	0.7453	0.8128	0.8703	0.8987
Kaizer	0.8081	0.8663	0.9035	0.9228
Rectangular	0.8092	0.8663	0.9036	0.9351

- The accuracy obtained using the STFT for various windows is presented in Fig. 1. Parzen window yields highly accurate results even when an 1 sec frame length is employed. This accuracy approaches 0.9990, which means that the STFT approach with proper temporal window can outperform state-of-the-art methods.

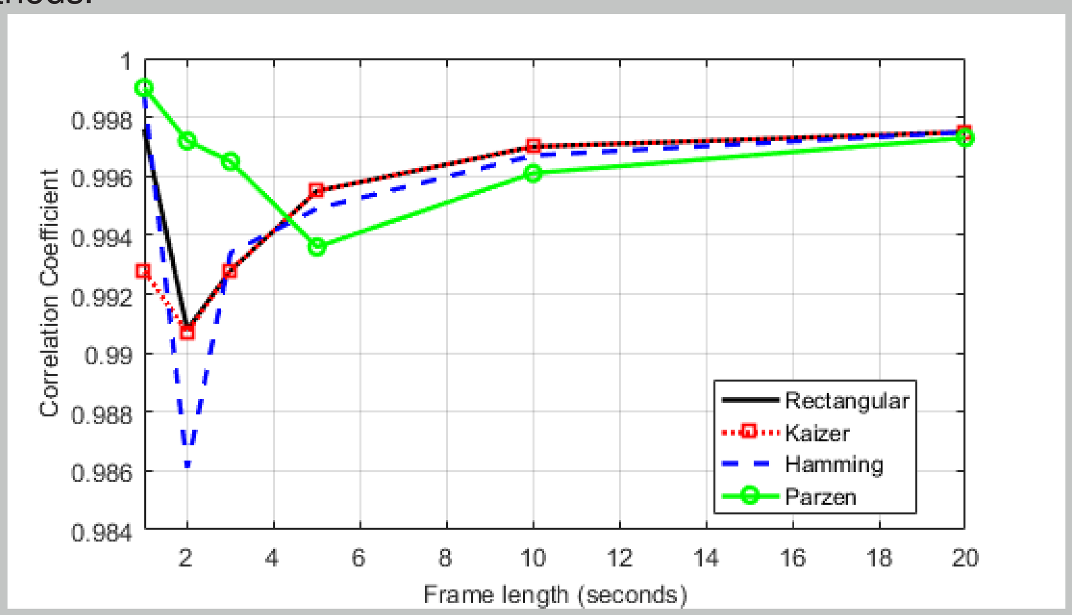

Figure 1: STFT using different windows for ENF estimation of Data 1.

- In order to determine whether the correlation coefficient of the proposed method is significantly different from that of other methods (H_1: $c_1 \neq c_2$), hypothesis testing was applied.
- Fisher transformation, $z = 0.5 \ln \frac{1+c}{1-c}$ was employed for each pair of correlation coefficients under examination.
- For significance level 95%, the test statistic $\theta = \sqrt{K-3}\,(z_1 - z_2)$ was outside the region $-1.96 < \theta < 1.96$, where $K = 1800$.
- The null hypothesis was rejected for every pair of comparisons. The differences between the correlation coefficients were significant at confidence level of 95%.

6. Conclusions

- Resorting to the Toeplitz structure of the covariance matrices and exploiting Krylov matrices, a fast and efficient he ENF estimation has been developed, which yields higher accuracy compared to the state-of-the-art methods in power recordings.
- A trade-off between speed and accuracy is observed, which is of crucial importance in forensic applications.
- The proposed approach provides state-of-the-art results, even when very short frame lengths are employed.
- The choice of the window is not a trivial task. Even a STFT achieves very good results for properly chosen window of very short length.

References

[1] A. J. Cooper, "An automated approach to the electric network frequency (ENF) criterion - theory and practice," *Int. J. of Speech Language and the Law*, vol. 16, no. 2, 2009.

[2] G. Karantaidis and C. Kotropoulos, "Assessing spectral estimation methods for electric network frequency extraction," in *Proc. 22nd Panhellenic Conf. Informatics*, pp. 202–207, 2018.

[3] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, "ENF extraction from digital recordings using adaptive techniques and frequency tracking," *IEEE Trans. Inf. Forensics and Security*, vol. 7, no. 4, pp. 1330–1338, 2012.

[4] G. Glentis and A. Jakobsson, "Efficient implementation of iterative adaptive approach spectral estimation techniques," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4154–4167, 2011.

[5] M. Xue, L. Xu, and J. Li, "IAA spectral estimation: Fast implementation using the Gohberg-Semencul factorization," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3251–3261, 2011.

[6] D. Bykhovsky and A. Cohen, "Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model," *IEEE Trans. Inf. Forensics and Security*, vol. 8, no. 5, pp. 774–753, 2013.

[7] X. Lin and X. Kang, "Robust electric network frequency estimation with rank reduction and linear prediction," *ACM Trans. Multimedia Com. Commun. and Appl.*, vol. 14, no. 4, pp. 1–13, 2018.