OPTIMIZATION OF WEIGHTING FACTORS FOR MULTIPLE WINDOW TIME-FREQUENCY ANALYSIS

Maria Hansson-Sandsten and Johan Sandberg

Mathematical Statistics, Centre for Mathematical Sciences, Lund University.
Address: Box 118, SE-221 00 Lund, Sweden. Phone:+46 46 22 249 53. Fax: +46 46 22 246 23.
Email: sandsten@maths.lth.se,sandberg@maths.lth.se. Web: www.maths.lth.se/matstat

ABSTRACT

This paper concerns the optimal weighting factors for multiple window spectrogram estimation of different stationary and non-stationary processes. The choice of windows are of course important but the weighting factors in the average of the different spectrograms are as important. The criterion for optimization is the normalized mean square error where the normalization factor is the spectrogram estimate. This means that the unknown weighting factors will be present in the numerator as well as in the denominator. A quasi-Newton algorithm is used for the estimation. The optimization is compared for a number of well known sets of multiple windows and the results show that the number as well as the shape of the windows are important factors for a small mean square error.

1. INTRODUCTION

The idea of multiple windows or multitapering were introduced by Thomson, [1], and in the last decades the Thomson method has been used in many different application areas. It has been shown to outperform the Welch method, [2], in terms of leakage, resolution and variance for a stationary spectrally smooth process, [3]. For non-smooth spectra, however, the performance of the Thomson method degrades due to cross-correlation between spectra, [4]. Other appropriate choices are then e.g., [5, 6, 7]. A comparison of Hermite and Slepian functions (the Thomson method) has shown that in the case of time-varying signals and spectrogram estimation, Hermite functions are a better choice, [8].

The choice of windows are of course important but the weighting factors in the average of the different spectra/spectrograms are as important. In [9], the weighting factors are optimized for the Peak Matched Multiple Windows, [6]. A criterion is used where normalized bias, variance and mean square error is optimized for the predefined peaked spectrum. In the non-stationary case, different approaches to approximate a time-varying spectrum with a few multiple window spectrograms have been taken, e.g., [10, 11, 12, 13].

In this paper, the optimization method in [9] is used to find the weighting factors for non-stationary processes using predefined windows. We compare the Hermite functions, the Thomson windows, the Peak Matched Multiple Windows and the Welch windows and compute the performance with optimal weighting factors for different processes.

The paper is organized as follows: Section 2 presents the optimized weighting factors and in Section 3 the evaluation for different stationary and non-stationary processes are presented. Section 4 concludes the paper.

Thanks to the Swedish Research Council for funding.

2. OPTIMIZATION OF WEIGHTING FACTORS

The Multiple Window Spectrogram $\hat{S}_x(n,k)$ of the zero mean real valued random process $x(n)$, $n = 0,\dots,N_0-1$ is defined by

$$\hat{S}_x(n,k) = \sum_{j=1}^{I} \alpha_j \left| \sum_{m=0}^{N-1} x(m+nL) h_j(m) e^{-i2\pi \frac{k}{K} m} \right|^2, \quad (1)$$

for $k = 0\dots K-1$ and $0 \le n \le \frac{N_0-N}{L}$, where the assumption is that the data is stationary for the N samples $x(nL)\ \dots\ x(N-1+nL)$. Equation (1) is a weighted sum of spectrograms obtained by using the data windows $\mathbf{h}_j = [h_j(0)\ \dots\ h_j(N-1)]^T$, and the weighting factors α_j, $j = 1\dots I$. The parameter L is the step size and K the number of values in the DFT.

With one window, $I = 1$, the spectrogram has too large variance to be useful in the analysis of stochastic processes, as the variance is approximately $S_x(n,k)^2$.

2.1 Mean Square Error Optimization

The mean square error (MSE) is computed in the time interval $[-T\cdot L \dots T\cdot L]$ and in the frequency interval $[-\frac{M}{K}\dots\frac{M}{K}]$ as the average of a number of $2M+1 \times 2T+1$ time-frequency values,

$$\xi = \frac{1}{(2M+1)(2T+1)} \sum_{n=-T}^{T} \sum_{k=-M}^{M} \frac{\varepsilon(n,k)}{E^2[\hat{S}_x(n,k)]}, \quad (2)$$

where the mean square error for each time and frequency value is defined

$$\varepsilon(n,k) = \mathrm{Variance}[\hat{S}_x(n,k)] + \mathrm{Bias}^2[\hat{S}_x(n,k)]. \quad (3)$$

The variance is

$$\mathrm{Variance}[\hat{S}_x(n,k)] =$$

$$\sum_{j=1}^{I}\sum_{g=1}^{I} \alpha_j \alpha_g \mathrm{cov}[\hat{S}_j(n,k), \hat{S}_g(n,k)] \approx$$

$$\sum_{j=1}^{I}\sum_{g=1}^{I} \alpha_j \alpha_g |\mathbf{h}_j^T \mathbf{\Phi}^H(k) \mathbf{R}_X^n \mathbf{\Phi}(k) \mathbf{h}_g|^2, \quad (4)$$

where the covariance matrix $\mathbf{R}_X^n = E[\mathbf{x}\mathbf{x}^T]$ with $\mathbf{x} = [x(nL)\dots x(nL+N-1)]^T$ and $\mathbf{\Phi}(k) =$

diag$[1\ e^{-i2\pi k/K}\ \ldots\ e^{-i2\pi(N-1)k/K}]$ and the superscript H denotes conjugate transpose, according to [4]. Reduction of the variance is established if the correlation between the windowed periodogram (subspectra),

$$S_j(n,k) = \left| \sum_{m=0}^{N-1} x(m+nL)h_j(m)e^{-i2\pi \frac{k}{K} m} \right|^2,$$

from the windows $\mathbf{h}_j$ and $\mathbf{h}_g$, $j \neq g$, is small for all frequency values k.

The bias is

$$\begin{aligned} &\text{Bias}[\hat{S}_x(n,k)] = E[\hat{S}_x(n,k)] - S_x(n,k) \\ &= \sum_{j=1}^{I} \alpha_j \mathbf{h}_j^T \mathbf{\Phi}^H(k) \mathbf{R}_X^n \mathbf{\Phi}(k) \mathbf{h}_j - S_x(n,k), \end{aligned} \quad (5)$$

where $S_x(n,k)$ is the Wigner-Ville spectrum. The optimization criterion of Eq. (2) includes the expressions of Eqs. (4,5) where $\mathbf{h}_j,\ j=1,\ \ldots\ ,I$ are known windows and $\mathbf{R}_X^n$ is the time-variable covariance matrix. The unknown variables are $\alpha_j,\ j=1,\ldots,I$ which appear both in the numerator and the denominator of Eq. (2). The minimization of the criterion is therefore done iteratively with a quasi-Newton algorithm, [14]. The criterion and its derivative are used in the algorithm. The algorithm is described in [9]. Using these weights in the multiple window spectrogram estimate is referred to as OPTWEI.

2.2 Averaging and Scale Optimization

Usually, the spectrograms from different windows are equally weighted and averaged in the final estimate, i.e.,

$$\alpha_j = \frac{1}{I}, \ \ j = 1 \ldots I. \quad (6)$$

Using equal weights according to Eq. (6) is referred to as EQWEI. The mean square error could be optimized using equal weights scaled with a constant factor, i.e.,

$$\alpha_j^c = \frac{c}{I} \ \ j = 1 \ldots I,$$

where a closed form expression for the factor c is found from

$$c = \frac{\sum_{n=-T}^{T} \sum_{k=-M}^{M} \frac{S_x^2(n,k)}{E^2[\hat{S}_x(n,k)]}}{\sum_{n=-T}^{T} \sum_{k=-M}^{M} \frac{S_x(n,k)}{E[\hat{S}_x(n,k)]}},$$

which is referred to as SCWEI.

3. RESULTS

3.1 Band-limited white noise process

The evaluation is done for different stationary and non-stationary processes. The band-limited white noise process with the covariance function

$$r_w(n-m) = B\,\frac{\sin(\pi B(n-m))}{\pi B(n-m)}, \quad (7)$$

generates a Toeplitz covariance matrix $R_{stat}(n,m)$, which is shown in Figure 1a) as Case 1 for $B = (8+3)/128 \approx 0.08$. The locally stationary process approach, [15, 16], where the covariance function of a non-stationary process is defined by

$$r_{ns}(n,m) = r_w(n-m)e^{-((n-m)/(F_s))^2}e^{-((n+m)/(F_s))^2}, \quad (8)$$

gives a time-variable band-limited spectrum where the time-variable power of the bandlimited white-noise process changes with a Gaussian envelope. Two examples are seen in Figure 1b) and c) as a more slowly varying process, Case 2, ($F_s = 120$), and a faster varying process, Case 3, ($F_s = 60$).

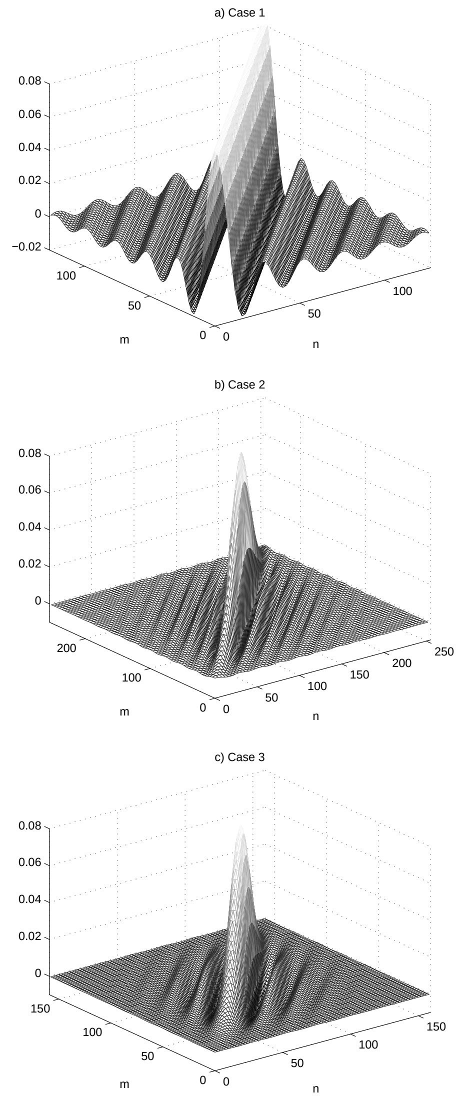

Figure 1: The three different test covariance matrices for band-limited white noise processes, a) Stationary process, $B = 0.08$, b) Slowly time-varying non-stationary process, $F_s = 120$, c) Time-varying non-stationary process, $F_s = 60$.

The weighting factors are optimized using four different sets of multiple windows. The Thomson multiple windows (TH), [1], give uncorrelated subspectra and thereby low variance for a stationary white noise process. The TH windows $\mathbf{h}_j$ are given by the solution of an eigenvalue problem and the number of windows relates to the bandwidth as

$$I \approx B \cdot N - 3. \tag{9}$$

The Peak Matched multiple windows, (PM), [6], are designed to give small correlation between subspectra when the spectrum of the stationary process includes peaks and notches. The windows are given by the solution of the generalized eigenvalue problem where the number of windows satisfies Eq. (9). Other parameter choices are peak height $C = 20$ dB and sidelobe suppression $K = 30$ dB, [6].

The Welch method (WO), [2], utilizes time-shifted equal windows. In this paper we use a Hanning window of appropriate length so that the number of windows, I, is fitted into the total window length N with 50 % overlap.

A set of Hermite functions (HE) is computed as

$$\begin{aligned} h_1(t) &= e^{-t^2/2}, \\ h_2(t) &= 2t\, e^{-t^2/2}, \\ h_j(t) &= 2t\, h_{j-1}(t) - 2(j-2)h_{j-2}(t), \quad j = 3 \ldots I, \end{aligned}$$

with $t = \frac{n}{F_s^H}$ for $n = -N/2 \ldots N/2-1$. The parameter F_s^H is chosen so that the first Hermite function is approximately equal to the first Slepian function of the Thomson method in each case (similar approach as in [8]).

The number of windows are chosen as $I = 8$ for all different methods and the window length is in all cases $N = 128$. For TH and PM the number of windows are then chosen in concordance with the recommendation of Eq. (9). For Case 1 (stationary process), the MSE is computed and optimized for the frequency interval $\pm M/K = \pm 8/256$ ($2 * M + 1 = 17$ values) and for $T = 0$. For the non-stationary cases, $M = 8$ and $T \cdot L = 64$ with $T = 8$, (17×17 values) for Case 2 and $T \cdot L = 16$ with $T = 8$, (17×17 values) for Case 3.

The MSE for Case 1 are shown in Figure 2a), for the different window setups (methods), where the MSE from EQWEI is shown with circles, the SCWEI with pluses and OPTWEI with stars. The TH and HE are optimal for the stationary bandlimited white-noise process using the EQWEI and thereby the optimization of the weighting factors (SCWEI and OPTWEI) do not give any improvement of the MSE. The PM, does not give a small error using EQWEI but using SCWEI and also OPTWEI, the MSE decreases. The overall smallest error however, is given by the TH and HE as expected, as these two window setups are optimal for a stationary bandlimited process.

In Case 2, Figure 2b), the slowly varying non-stationary process shows the importance of using SCWEI (pluses) and OPTWEI (stars) compared to the EQWEI (circles). The difference of these two weightings are, however, not that large. It could also be noted that the HE is now slightly better than the TH, which is in concordance with the study in [8]. The weighting factors, α_j are depicted in Figure 4a) for the Hermite functions. The level of EQWEI (circles) is increased for SCWEI (pluses) and for OPTWEI the further change is not that severe which also is the reason for the small change in the mean square error estimate.

In Case 3, Figure 2c), using EQWEI on the faster varying non-stationary process gives a very large error. Using SCWEI and OPTWEI gives a much lower MSE. The weighting factors, α_j are depicted in Figure 4b) for the Hermite functions and the difference between EQWEI and SCWEI is larger in than for Case 2.

3.2 Band-limited peaked spectrum process

Instead of using a bandlimited white noise process the stationary covariance function in Eq. (7) is replaced with the covariance function $r_x(n-m)$ of a band-limited peaked spectrum according to,

$$S_x(f) = \begin{cases} e^{\frac{-2C|f|}{10Blog_{10}(e)}} & |f| \leq B/2 \\ 0 & |f| > B/2. \end{cases} \tag{10}$$

In Eq. (10), $S_x(f)$ is a peaked spectrum with $S_x(0) = 1$, $B = 0.08$ and $S_x(B/2) = -C$ dB, where $C = 20$ dB throughout this paper. The non-stationary processes are found from Eq. (8) with $r_w(n-m)$ replaced with $r_x(n-m)$.

The results from these processes are presented in Figure 3. In Case 1, Figure 3a), for the stationary peaked spectrum process, the optimized weighting factors of the PM gives the smallest MSE, which is concordance with [6, 9], where these windows and optimized weighting factors are shown to be optimal for this process. In Case 2, Figure 3b), for the slowly time-varying process, the benefit in using these windows and optimizing the weighting factors clearly shows up as the lowest MSE is given from the combination PM and OPTWEI.

In Case 3, we also see that this is the same result is given for the more time-varying process. The different weighting factors are depicted in Figure 5 for Case 2 and Case 3 where we see that the optimized weighting factors (stars) are very different compared to the usual averaging (circles) and scaled equal weights (pluses).

4. CONCLUSION

We compare the Hermite functions, the Thomson windows, the Peak Matched Multiple Windows and the Welch windows and compute the performance with optimal weighting factors for different stationary and non-stationary processes. A quasi-Newton algorithm is used for the estimation. The results show that the number as well as the shape of the windows are important factors for a small mean square error. It is also shown that a scaling optimization of the usual averaging could give almost as small mean square error as an optimization of the individual weighting factors in case of a smooth spectrum but for a peaked spectrum, a reduction of the mean square error is achieved using individual optimization of the weights.

REFERENCES

[1] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept 1982.

[2] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. on Audio Electroacoustics*, vol. AU-15, no. 2, pp. 70–73, June 1967.

[3] T. P. Bronez, "On the performance advantage of multitaper spectral analysis," *IEEE Trans. on Signal Processing*, vol. 40, no. 12, pp. 2941–2946, Dec 1992.

[4] A. T. Walden, E. McCoy, and D. B. Percival, "The variance of multitaper spectrum estimates for real gaussian processes," *IEEE Trans. on Signal Processing*, vol. 42, no. 2, pp. 479–482, Feb 1994.

[5] K. S. Riedel, "Minimum bias multiple taper spectral estimation," *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 188–195, January 1995.

[6] M. Hansson and G. Salomonsson, "A multiple window method for estimation of peaked spectra," *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 778–781, March 1997.

[7] B. Farhang-Boroujeny, "Prolate filters for nonadaptive multitaper spectral estimators with high spectral dynamic range," *IEEE Signal Processing Letters*, vol. 15, pp. 457–460, Sept 2008.

[8] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency analysis," in *Proc. of Int. Symposium of Time-Frequency and Time-Scale Analysis*, June 1996, pp. 511–514.

[9] M. Hansson, "Optimized weighted averaging of peak matched multiple window spectrum estimates," *IEEE Trans. on Signal Processing*, vol. 47, no. 4, pp. 1141–1146, April 1999.

[10] F. Cakrak and P. J. Loughlin, "Multiple window time-varying spectral analysis," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 448–453, 2001.

[11] J. Xiao and P. Flandrin, "Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement," *IEEE Trans. on Signal Processing*, vol. 55, no. 6, pp. 2851–2860, 2007.

[12] W. J. Williams and S. Aviyente, "Spectrogram decompositions of time-frequency distributions," in *Proc. of the ISSPA*. IEEE, 2001, pp. 587–590.

[13] L. L. Scharf and B. Friedlander, "Toeplitz and Hankel kernels for estimating time-varying spectra of discrete-time random processes," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 179–189, 2001.

[14] R. Fletcher, *Practical Methods for Optimization*, John Wiley & Sons, sec. edition, 1987.

[15] R. A. Silverman, "Locally stationary random processes," *IRE Trans. Inf. Theory*, vol. 3, pp. 182–187, 1957.

[16] P. Wahlberg and M. Hansson, "Kernels and multiple windows for estimation of the Wigner-Ville spectrum of gaussian locally stationary processes.," *IEEE Trans. on Signal Processing.*, vol. 55, no. 1, pp. 73–84, January 2007.

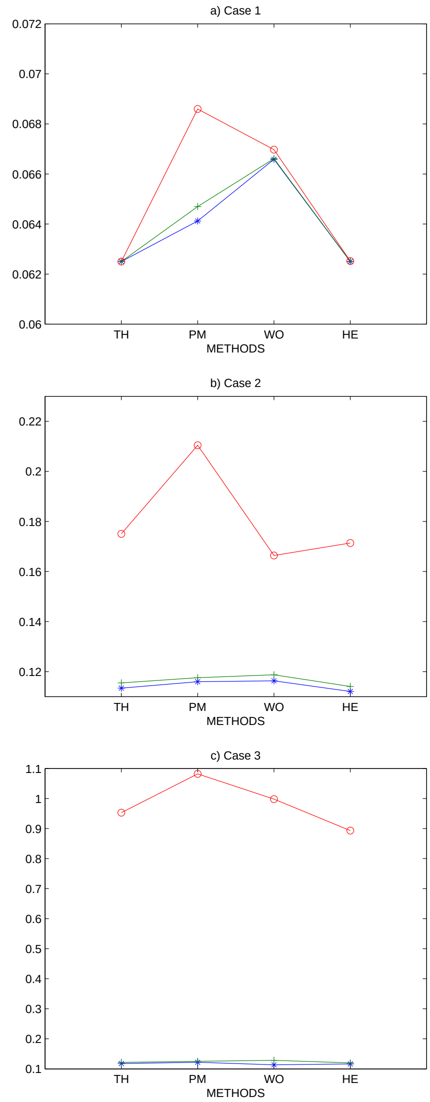

Figure 2: The mean square error for the three bandlimited white noise processes with EQWEI (circles), SCWEI (pluses) and OPTWEI (stars) for different window sets of TH, PM, WO and HE, a) Stationary process, b) Slowly time-varying non-stationary process, c) Time-varying non-stationary process.

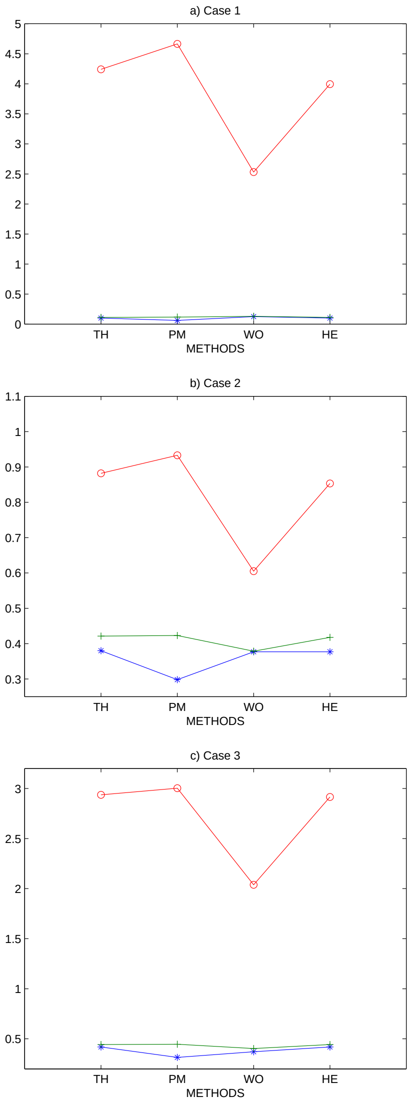

Figure 3: The mean square error for the three band-limited peaked spectrum processes with EQWEI (circles), SCWEI (pluses) and OPTWEI (stars) for different window sets TH, PM, WO and HE, a) Stationary process, b) Slowly time-varying non-stationary process, c) Time-varying non-stationary process.

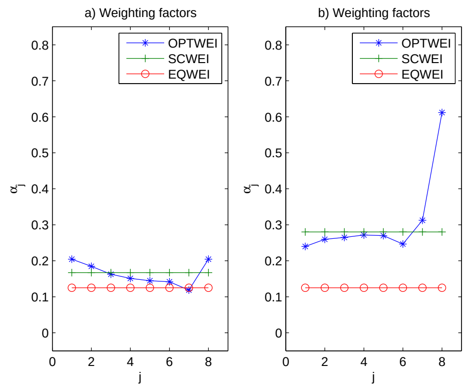

Figure 4: The different weighting factors for the Hermite functions in a) Case 2, b) Case 3, for the band-limited white noise process.

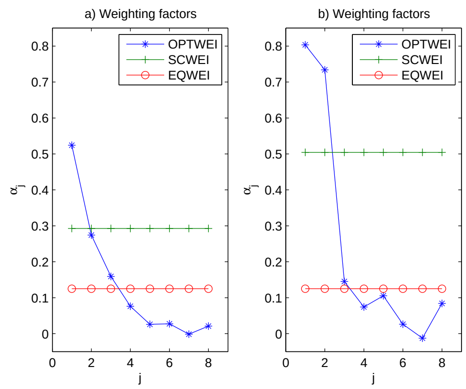

Figure 5: The different weighting factors for the Peak Matched Multiple Windows in a) Case 2, b) Case 3 for the band-limited peaked spectrum process.