MULTI TAPER WIGNER DISTRIBUTION WITH PREDETERMINED DOPPLER-LAG BANDWIDTH AND SIDELOBE SUPPRESSION

Maria Hansson-Sandsten

Mathematical Statistics, Centre for Mathematical Sciences, Lund University.
Box 118, SE-221 00 Lund, Sweden.
phone: + (46) 46 22 249 53, fax: + (46) 46 22 246 23, email: sandsten@maths.lth.se, web: www.maths.lth.se/matstat

ABSTRACT

Multi taper spectrogram decomposition of the time-lag kernel of a time-frequency distribution might result in computationally efficient calculations if the number of multi tapers to be considered in the spectrogram are small. In this paper, penalty functions are designed and used in the computation of multi tapers corresponding to the Wigner distribution time-lag kernel. The resulting multi taper spectrogram will approximately fulfill the concentration of the Wigner distribution but will also suppress the usual Wigner distribution cross-terms outside a predetermined doppler-lag bandwidth. The level of the cross-term suppression is determined by a parameter of the penalty function. The proposed method uses a limited number of multi tapers which is determined by the decided bandwidth. The time-frequency concentration of the proposed method is compared to other well-known distributions. The performance for white noise disturbances are also evaluated.

1. INTRODUCTION

The area of time-frequency analysis is well covered in the signal processing literature and a large number of time-frequency distributions have been proposed for various types of applications. From time-frequency concentration viewpoint, the Wigner distribution is the optimal choice with high time-frequency resolution and today a large number of time-frequency kernels exist with different ability to suppress the resulting cross-terms from the Wigner distribution, e.g., [1, 2, 3, 4]. Another important aspect is also when the signal components to be resolved are disturbed by additive noise. The theoretical results for computing the bias and variance of the Wigner distribution for the case of additive noise are given in [5] and a minimum-variance kernel is obtained in [6].

Computationally efficient algorithms can be found using the eigenvalues and eigenvectors of the rotated time-lag kernel, where the resulting multi taper spectrogram is the smoothed Wigner-Ville estimate, [4]. The phrase multi taper was originally introduced by Thomson, [7], for the case of stationary processes with smooth spectra. Multi taper decomposition of time-lag kernels have been analyzed from several aspects, for existing kernels, e.g., in [8, 9], and new multi taper techniques for non-stationary signal analysis have also been proposed, e.g., in [10, 11, 12, 13]. One of the advantages of the Thomson multi tapers is the strong sidelobe suppression outside a predetermined frequency interval. Other methods have been proposed for the multi taper spectrum estimate of stationary processes where the tapers also fulfil the criterion of strong sidelobe suppression, [14, 15]. The aspect of time-frequency localization and orthogonality in the time-frequency domain (in contrast to only considering the frequency domain) was noted by [16] and made the Hermite functions to become often used as multi tapers for spectrogram estimation of non-stationary processes.

The Wigner distribution gives the best time-frequency concentration and the aim is to find the multi tapers corresponding to this kernel, with the corresponding time-frequency kernel limited to a predefined doppler-lag bandwidth to prevent cross-terms. In this paper we use the idea of [15] to suppress the sidelobes of the multi taper spectrogram outside a pre-determined ambiguity domain area. An advantage of the approach is that the number of windowed spectrograms to be averaged are limited. The idea was initially presented in [17] using another penalty function and limiting the Thomson multi taper kernel.

2. SPECTROGRAM DECOMPOSITION OF TIME-FREQUENCY KERNELS

The connection between a multi taper spectrogram and a smoothed Wigner distribution is found using the following approach. The multi taper spectrogram is defined as

$$S_x(t,f) = \sum_{k=1}^{K} \alpha_k \left| \int_{-\infty}^{\infty} h_k^*(t-t_1)x(t_1)e^{-i2\pi f t_1} dt_1 \right|^2. \quad (1)$$

With $t_1 = t' + \frac{\tau}{2}$ and $t_2 = t' - \frac{\tau}{2}$,

$$S_x(t,f) = \sum_{k=1}^{K} \alpha_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t'+\frac{\tau}{2})x^*(t'-\frac{\tau}{2}) \cdot h_k(t-t'-\frac{\tau}{2})h_k^*(t-t'+\frac{\tau}{2})e^{-i2\pi f\tau} d\tau dt'. \quad (2)$$

We identify the instantaneous autocorrelation function as

$$r_x(t,\tau) = x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2}), \quad (3)$$

and the time-lag kernel

$$\rho(t,\tau) = \sum_{k=1}^{K} \alpha_k h_k(t+\frac{\tau}{2})h_k^*(t-\frac{\tau}{2}), \quad (4)$$

giving the quadratic class of time-frequency distributions, [18], as

$$C_x(t,f) = S_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_x(t',\tau)\rho(t-t',\tau)^* e^{-i2\pi f\tau} dt' d\tau. \quad (5)$$

Thanks to the Swedish Research Council for funding.

 ISSN 2076-1465

Defining

$$\rho^{rot}(t_1,t_2) = \rho(\frac{t_1+t_2}{2}, t_1-t_2), \tag{6}$$

and if the kernel $\rho^{rot}(t_1,t_2)$ satisfies the Hermitian property

$$\rho^{rot}(t_1,t_2) = (\rho^{rot}(t_2,t_1))^*,$$

then solving the integral

$$\int \rho^{rot}(t_1,t_2)q(t_1)dt_1 = \lambda q(t_2)$$

results in eigenvalues λ_k and eigenfunctions $q_k(t)$ which form a complete set and can be used as weights, α_k, and multi tapers, $h_k(t) = q_k(t)$, $k = 1 \ldots K$, in Eq. (1).

Quadratic time-frequency distributions can be formulated as a multiplication of the Doppler-lag (or ambiguity domain) kernel

$$\phi(\nu,\tau) = \int_{-\infty}^{\infty} \rho(t,\tau)e^{-i2\pi\nu t}dt, \tag{7}$$

and the ambiguity function

$$A_x(\nu,\tau) = \int_{-\infty}^{\infty} r_x(t,\tau)e^{-i2\pi\nu t}dt, \tag{8}$$

as

$$A_x^C(\nu,\tau) = A_x(\nu,\tau)\cdot\phi(\nu,\tau).$$

The Wigner distribution has the simple doppler-lag kernel $\phi(\nu,\tau) = 1$ for all values of ν and τ and by using penalty functions we attempt to limit this function with the aim to reduce cross-terms.

3. PENALTY FUNCTIONS

The time-lag kernel of the Wigner distribution is defined as

$$\rho(t,\tau) = \delta(t). \tag{9}$$

In the discrete-time case the corresponding rotated time-lag kernel $\rho^{rot}(t_1,t_2) = \delta(\frac{t_1+t_2}{2})$, is sampled giving a rotated kernel matrix

$$\mathbf{R} = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ & \cdots & & \\ \vdots & 1 & 0 & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}. \tag{10}$$

In [15], a frequency penalty function was used to suppress the sidelobes of the multi tapers suitable for the stationary case of a peaked spectrum. The multi tapers and weights were given as the solution of a generalized eigenvalue problem,

$$\mathbf{R}\mathbf{q}_k = \lambda_k \mathbf{W}\mathbf{q}_k, \quad k = 1\ldots N, \tag{11}$$

where the covariance matrix $\mathbf{R}$, $N \times N$, corresponded to a peaked spectrum and the matrix $\mathbf{W}$, $N \times N$, was the corresponding covariance matrix of a penalty spectrum with a pre-defined bandwidth and side-lobe suppression. A similar idea is used here where the set of multi tapers $\mathbf{h}_k = \mathbf{q}_k = [q_k(0)\ q_k(1)\ldots q_k(N-1)]^T$ and the weights $\alpha_k = \lambda_k$, $k = 1\ldots K$, are found from the solution of Eq. (11) with $\mathbf{R}$ given from Eq. (10) and $\mathbf{W}$ will be specified below. The multi tapers and weights are then used to compute the corresponding discrete-time case of the spectrogram in Eq. (1).

Doppler penalty function-DP

A doppler penalty (DP) function is defined in [15],

$$S_{W_\nu}(\nu) = \begin{cases} P & \text{if } \frac{\Delta\nu_p}{2} \leq |\ \nu\ | < 0.5 \\ 1 & \text{if } |\ \nu\ | < \frac{\Delta\nu_p}{2} \end{cases} \tag{12}$$

to decrease the leakage from the sidelobes outside the doppler interval of width $\Delta\nu_p$. A corresponding penalty Toeplitz covariance matrix, $\mathbf{W}_\nu$, is found as

$$\mathbf{W}_\nu = \begin{bmatrix} r_{W_\nu}(0) & r_{W_\nu}(1) & \ldots & r_{W_\nu}(N-1) \\ r_{W_\nu}(1) & r_{W_\nu}(0) & \ldots & r_{W_\nu}(N-2) \\ \vdots & & \ddots & \vdots \\ r_{W_\nu}(N-1) & r_{W_\nu}(N-2) & \ldots & r_{W_\nu}(0) \end{bmatrix}, \tag{13}$$

where $r_{W_\nu}(n)$, $n = 0\ldots N-1$, is the time-discrete covariance function corresponding to $S_{W_\nu}(\nu)$.

Lag penalty function-LP

A simple lag penalty (LP) function is a diagonal suppression matrix $\mathbf{W}_\tau$, $(N \times N)$, where values outside a certain time interval are suppressed. The diagonal elements $W_\tau(n,n) = w_\tau(n)$, is defined by

$$w_\tau(n) = \begin{cases} P & \text{if } |N/2 - n + 1| \geq \frac{\Delta\tau_p}{2} \\ 1 & \text{otherwise} \end{cases} \tag{14}$$

Doppler-lag penalty function-DLP

The combination to a penalty function that influences lag- as well as doppler domain in a proper way is not straight forward and could be done in many different ways, e.g. as in, [17]. In this paper we choose the definition of a doppler-lag penalty function (DLP) as the matrix $\mathbf{W}_{\nu\tau}$ with elements as

$$W_{\nu\tau}(n,m) = \begin{cases} P \cdot W_\nu(n,m) & \text{if } |N/2 - n + 1| \geq \frac{\Delta\tau_p}{2} \\ & \text{AND } |N/2 - m + 1| \geq \frac{\Delta\tau_p}{2} \\ W_\nu(n,m) & \text{otherwise,} \end{cases} \tag{15}$$

which almost corresponds to just suppressing the taper functions 10 dB outside the time interval $\Delta\tau_p$. The resulting tapers from such a suppression is however not orthogonal to the matrix $\mathbf{R}$ which is the case for the tapers found from the solution of the generalized eigenvalue problem using the penalty matrix $\mathbf{W}_{\nu\tau}$. The orthogonality property of the tapers is always considered to be important from variance reduction aspects.

It is notable that using a $\mathbf{R}$-matrix corresponding to the time-lag kernel of the Wigner distribution, the solution of the generalized eigenvalue problem will be a rotation of the inverse penalty matrix. The interpretation of using the Wigner distribution time-lag kernel and limiting with use of penalty functions is however intuitively nice.

The corresponding multi taper spectrograms are named according to the penalty function used for the matrices ($\mathbf{W}_\nu$, $\mathbf{W}_\tau$ and $\mathbf{W}_{\nu\tau}$) as **Doppler Penalty Wigner Distribution-DPWD**, **Lag Penalty Wigner Distribution -LPWD** and **Doppler Lag Penalty Wigner Distribution-DLPWD** respectively.

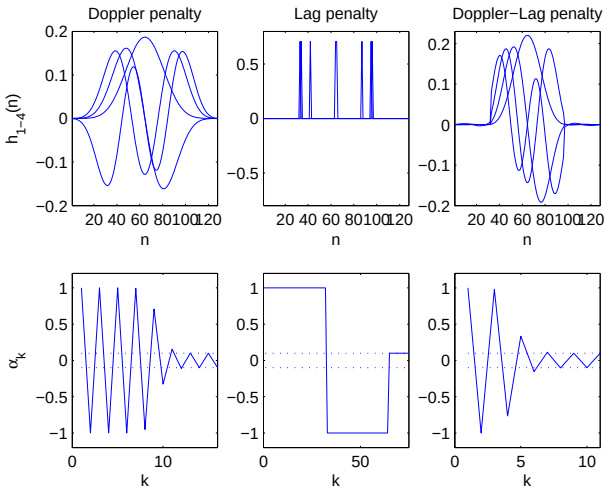

Figure 1: The four first multi tapers (upper row) and weighting factors (lower row) using different penalty functions with parameters $N = 128$, $P = 10$ dB, $\Delta\tau_p = 64$ and $\Delta\nu_p = 0.08$.

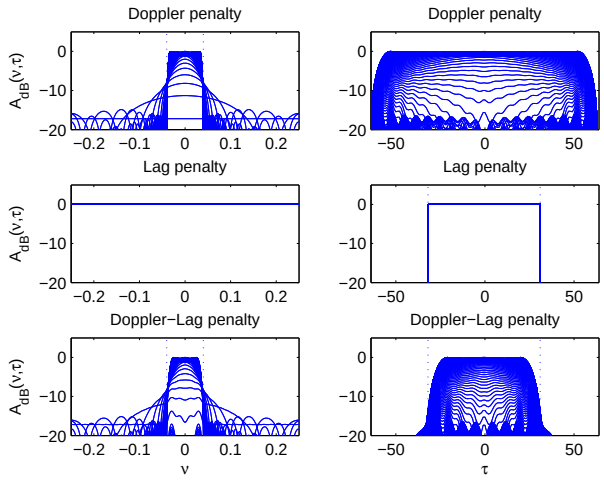

Figure 2: The ambiguity domain kernels using the different penalty functions with parameters $N = 128$, $P = 10$ dB, $\Delta\tau_p = 64$ and $\Delta\nu_p = 0.08$.

4. MULTI TAPERS AND WEIGHTS

An example from the solution of the resulting eigenvectors and eigenvalues from applying the different penalty functions of Eq. (11) are shown in Figure 1. The length of the tapers are $N = 128$ and the parameters of the penalty functions are $P = 10$ dB, $\Delta\nu_p = 0.08$ and $\Delta\tau_p = 64$. The four first eigenvectors for each case are depicted in the upper figures and the eigenvalues in decreasing order are plotted in the lower figures. A number of eigenvalues, determined by the frequency bandwidth $\Delta\nu_p$, have absolute values close to one and the remaining are close to the chosen suppression level $P = 10$ dB (0.1). The number of averaged spectrograms in Eq. (1), determined by K is chosen as the number of eigenvalues with absolute values significantly larger than 0.1. For the DP function, $K = 11$.

The resulting eigenvectors from the LPWD are not very useful from the aspect of multi taper spectrogram calculation as each taper just cover one or two samples. This is of course a result of the diagonal penalty matrix combined with the anti-diagonal $\mathbf{R}$-matrix from the Wigner distribution. If the LP function should have been combined with some other distribution kernel, the tapers would have been more useful. We use these windows for comparison although the results from the estimation will be similar of using a Doppler-independent kernel with a lag-window of constant level one and 0.1 outside a certain lag bandwidth. The number of eigenvalues that have absolute values larger than 0.1, determined by the lag bandwidth $\Delta\tau_p = 64$, are in this case, i.e., $K = \Delta\tau_p = 64$.

In the third case, where the DLP function are combined according to Eq. (15), the eigenvectors show a lag limitation corresponding to $\Delta\tau_p = 64$ samples. This can be compared with the multi tapers of the DPWD, which cover all time samples. The shapes of the tapers are otherwise similar to the shapes of the tapers of the DPWD although a more restricted number of eigenvalues, $K = 6$, now are found to be larger than 0.1, which gives a reasonable number of spectrograms to be averaged.

The ambiguity domain kernels of the different distributions in Eq. (7), corresponding to the multi tapers and weighting factors of Figure 1 are depicted in Figure 2. We see that the DP function causes a suppression of 10 dB outside the predefined bandwidth $\Delta\nu_p = 0.08$. For the LPWD the ambiguity kernel is very accurate, which is caused by the large degrees of freedom of the large number of tapers as each window is only one or two samples long. For the DLPWD the suppression is 10 dB outside the time limit of $\Delta\tau_p = 64$ as well as the frequency limit $\Delta\nu_p = 0.08$.

5. EVALUATION

Complex sinusoids with Gaussian envelopes at different time-frequency locations are often used to evaluate the resolution performance of different algorithms. We define one such sinusoidal component as

$$s(n) = e^{2\pi i f_0 n} \cdot h(n - \frac{N_s}{2}) \quad -\frac{N_s}{2} \le n < \frac{N_s}{2}, \qquad (16)$$

where $h(n) = e^{-\frac{1}{4N_s}n^2}$ and $N_s = 64$ and $f_0 = 0.05$.

5.1 Resolution and suppression of cross-terms

To actually investigate the advantages and drawbacks in terms of time- and frequency resolution, different combinations of time-shifted, $s(n - \delta_t)$ and frequency-shifted, $s(n)e^{2\pi i\delta_f}$ signals with δ_t and δ_f as parameters, are evaluated. The following four different cases are investigated:

$$\begin{aligned}
&1) \quad x_1(n) = \quad s(n) + s(n-\delta_t) \quad \text{time-shift} \qquad (17)\\
&2) \quad x_2(n) = \quad s(n) + s(n)e^{2\pi i\delta_f} \quad \text{frequency-shift}\\
&3) \quad x_3(n) = \quad s(n) + s(n-\delta_t)e^{2\pi i\delta_f} \quad \text{time-frequency-shift}\\
&4) \quad x_4(n) = \quad s(n) + s(n-\delta_t) + s(n)e^{2\pi i\delta_f} + s(n-\delta_t)e^{2\pi i\delta_f}\\
&\qquad\qquad \text{time-, frequency- and time-frequency-shifts}
\end{aligned}$$

As the Wigner distribution of a single component is well known to have the best time-frequency resolution, the sum

of the Wigner distributions of the single components is chosen as the ideal performance, e.g. for Case 1,

$$S_{ideal}(t,f) = S(t,f) + S(t-\delta_t, f),$$

where $S(t,f)$ is the Wigner distribution of $s(n)$. The usual cross-terms from the Wigner distribution of multi-component signals are not included in the ideal performance function. The performance of the penalty function algorithms is compared to the performance of a number of well-known algorithms/kernels using the measure,

$$MSE_{norm} = \frac{\int_t \int_f (S_{algorithm}(t,f) - S_{ideal}(t,f))^2}{\int_t \int_f (S_{ideal}(t,f))^2}, \qquad (18)$$

where we use a grid of 256 samples in the time domain and 256 samples in the frequency domain corresponding to 0 to 0.25 of normalized frequency. It can be noted that the Wigner distribution of $x_i(n)$ for $i = 1,2 \ldots 3$ of Eq. (18), (2-component signals) give $MSE_{norm} = 1$. The reason is of course that the sum of the power of the cross-terms is equal to the sum of the power of the two components, [4].

The parameter choices for all algorithms evaluated are in each case optimized to give the smallest possible MSE_{norm} for $\delta_t = 64$ and $\delta_f = 0.08$. The methods compared are the proposed three different multi taper cases of Section IV, the single Hanning window spectrogram (Hann), the Choi-Williams kernel (CW-kernel), [3], the lag-independent kernel (LI-kernel) and the doppler-independent kernel (DI-kernel), [4]. The CW-kernel is defined as

$$\phi(\nu,\tau) = e^{-\frac{\nu^2\cdot\tau^2}{\sigma}},$$

the LI-kernel as

$$\phi(\nu,\tau) = e^{-\frac{\nu^2}{\sigma}},$$

and the DI-kernel as

$$\phi(\nu,\tau) = e^{-\frac{\tau^2}{\sigma}}.$$

For Case 1 we compare the results of Hann, DI-kernel and CW-kernel with the results from the methods with different penalty functions. The optimal parameters chosen are $\sigma = 1300$ for the DI-kernel and $\sigma = 0.08$ for the CW-kernel. For Hann, the window length is $N = 56$. The lag penalty function parameter is chosen as $\Delta\tau_p = 32$ for LPWD and for the DPWD as $\Delta\nu_p = 0.04$. For the DLPWD, $\Delta\tau_p = 48$ and $\Delta\nu_p = 0.06$. The results of Figure 3a) show that the LPWD, DLPWD and the DI-kernel have similar performances when the distance between the two time-shifted components is increased. For $\delta_t > 40$ Hann as well as the CW-kernel gives a much larger MSE_{norm}. The DPWD (not shown) gives a $MSE_{norm} \approx 1$ which is similar to the Wigner distribution (not shown) as the frequency penalty function does not effect the cross-terms between two time-shifted components.

In Case 2, the LI-kernel with $\sigma = 0.002$ is used instead of the DI-kernel as the LI-kernel has a better suppression of cross-terms between frequency-shifted components. All the other methods have the same parameter choices as in Case 1. From the results in Figure 3b), the LI-kernel, the DPWD and the DLPWD have the smallest error. The CW-kernel and Hann have larger MSE_{norm} for $\delta_f > 0.05$. In this case the

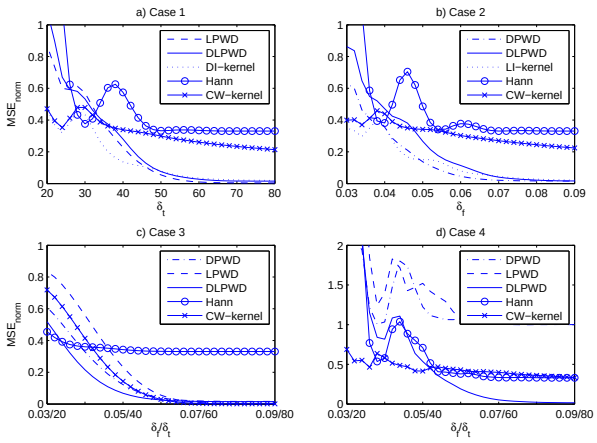

Figure 3: The normalized mean square error for different algorithms and different cases: a) Case 1 - Time-shift, b) Case 2 - Frequency-shift , c) Case 3 - Time-frequency-shift, d) Case 4 - Time-, frequency- and time-frequency-shift.

LPWD (not shown) give a result close to the Wigner distribution (not shown), ($MSE_{norm} \approx 1$), as time domain suppression has no effect in the frequency domain.

For Case 3, Figure 3c), the two components are shifted both in time and frequency and the smallest errors are now given from LPWD, DPWD, DLPWD as well as from the CW-kernel. The parameter choices are the same as for the cases above for all methods except for the CW-kernel where $\sigma = 2.5$ now gives the smallest error. A closer study shows that the DLPWD gives the smallest error of all algorithms for $\delta_f > 0.06$ and $\delta_t > 48$ (the limits of the penalty function) but that the CW-kernel is lower for larger values of δ_f and δ_t.

Including components with shifts in all directions, the 4 component case of Case 4, give cross-terms between all components in all directions, and the CW-kernel now using optimal parameter choice $\sigma = 0.03$ give cross-term errors in the directions of the time- and frequency- axis, Figure 3d). The result is very similar to Hann for $\delta_t > 40$ and $\delta_f > 0.05$. With optimal parameter settings, same as above, the performance of the LPWD and DPWD are bad as they just suppress cross-terms in either time- or frequency direction. The DLPWD however, gives a very nice performance for components with larger distances than the penalty function bandwidth.

5.2 Noise sensitivity

The performance for one signal component $s(n)$ of Eq. (16) is investigated for different amounts of additive noise where the disturbance is complex-valued circularly symmetric Gaussian white noise with varying standard deviation β. The number of simulations in each case are 500 and the performance is measured as the estimated normalized mean square error, MSE_{norm}, Eq. (18), from these measurements compared to the Wigner distribution $S(t,f)$ of $s(n)$. The algorithm parameters are for Hann $N = 56$, the CW-kernel $\sigma = 2.5$, the LI-kernel $\sigma = 0.002$ and the DI-kernel $\sigma = 1300$. For the penalty function algorithms the parameter choices are $\Delta\tau_p = 32$ for LPWD, $\Delta\nu_p = 0.04$ for the DPWD and $\Delta\tau_p = 48$, $\Delta\nu_p = 0.06$ for the DLPWD.

In Figure 4, the resulting MSE_{norm} for all the different

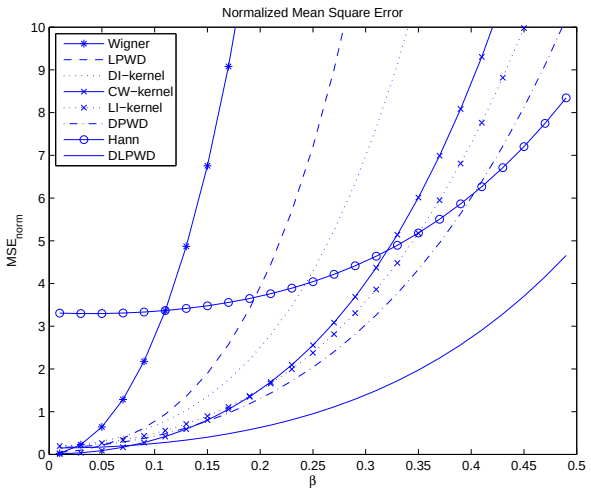

Figure 4: The normalized mean square error of different algorithm estimates for a Gaussian windowed complex sinusoid disturbed by Gaussian noise of varying standard deviation β.

algorithms are depicted. The robustness against large disturbances is clearly best for the DLPWD. For very small noise disturbances, the CW-kernel and Wigner spectrum naturally have a better performance followed by the LPWD. The reason is of course that for the noise-free single component, both the Wigner distribution and the CW-kernel perform very nicely. However, the resolution for the DLPWD, DPWD, the DI-kernel and the LI-kernel are about the same. The Hann has a much larger value.

6. CONCLUSION

Different penalty functions are proposed to suppress the Wigner distribution ambiguity function outside a predefined doppler-lag bandwidth. Using a multi taper spectrogram with a reasonable number of spectrogram averages gives the solution where the multi tapers and weights are computed from a generalized eigenvalue problem including the Wigner distribution time-lag matrix and a penalty matrix. The results show that the proposed method give a better result than the Choi-Williams kernel, the doppler-independent and lag-independent kernels as well as the single Hanning spectrogram from the aspects of time-frequency resolution and cross-term suppression. The proposed method does also give a robust result with low mean square error for signals disturbed by white noise.

REFERENCES

[1] 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.

[2] G. R. Arce and S. R. Hasan, "Elimination of interference terms of the discrete Wigner distribution using nonlinear filtering," *IEEE Trans. on Signal Processing*, vol. 48, 2000.

[3] H.-I. Choi and W. J. Williams, "Improved time-frequency representation of multi-component signals using exponential kernels," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, 1989.

[4] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, 1995.

[5] L. Stanković and S. Stanković, "Wigner distribution of noisy signals," *IEEE Trans. on Signal Processing*, vol. 41, 1993.

[6] S. B. Hearon and M. G. Amin, "Minimum-variance time-frequency distribution kernels," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1258–1262, May 1995.

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

[8] G. S. Cunningham and W. J. Williams, "Kernel decomposition of time-frequency distributions," *IEEE Trans. on Signal Processing*, vol. 42, pp. 1425–1442, June 1994.

[9] M. G. Amin, "Spectral decomposition of time-frequency distribution kernels," *IEEE Trans. on Signal Processing*, vol. 42, no. 5, pp. 1156–1165, May 1994.

[10] S. Aviyente and W. J. Williams, "Multitaper reduced interference distribution," in *Proc. of the tenth IEEE Workshop on Statistical Signal and Array Processing*. IEEE, 2000, pp. 569–573.

[11] 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.

[12] F. Cakrak and P. J. Loughlin, "Multiwindow time-varying spectrum with instantaneous bandwidth and frequency constraints," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1656–1666, 2001.

[13] 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. 73–87, January 2007.

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

[15] M. Hansson, "Optimization of weighting factors in the peak matched multiple window method," in *Proc. of the ICASSP*, Munich, Germany, April 21-24 1997, IEEE, pp. 3973–3976.

[16] 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. 173–176.

[17] M. Hansson, "Multiple window decomposition of time-frequency kernels using a penalty function for suppressed sidelobes," in *Proc. of the ICASSP*, Toulouse, France, May 14-19 2006, IEEE.

[18] B. Boashash, "Theory of quadratic TFDs," in *Time Frequency Signal Analysis and Processing; A Comprehensive Reference*, B. Boashash, Ed., chapter 3. Elsevier, 2003.