OPTIMAL WIGNER CROSS-SPECTRUM ESTIMATION

Johan Sandberg and Maria Hansson-Sandsten

Mathematical Statistics, Centre for Mathematical Sciences, Lund University.
Address: Box 118, SE-221 00 Lund, Sweden. Phone: + 46 46 22 285 50. Fax: +46 46 22 246 23.
Email: sandberg@maths.lth.se, sandsten@maths.lth.se. Web: www.maths.lth.se/matstat

ABSTRACT

An expression for the mean-square error (MSE) optimal Wigner cross-spectrum estimator is derived within a wide class of estimators. Using this expression and a simple model we are able to construct a new estimator which turns out to be superior compared to standard estimators such as Thomson multitapers and Welch method on simulated data.

1. INTRODUCTION

Time-frequency analysis of non-stationary processes has been approached from different view-points and with different assumptions of the non-stationarity, e.g., the evolutionary spectrogram, [1], that assumes oscillatory processes and Wigner spectral estimation of harmonizable processes, [2]. The mean square error optimal kernel for the class of Gaussian harmonizable processes has been obtained by Sayeed and Jones, [3, 4], and other optimization criteria are minimization of variance, [5, 6] and entropy, [7].

The ordinary cross-spectral density, or cross-spectrum, between two jointly stationary zero-meaned stochastic processes $\{x(t), t \in \mathbb{R}\}$ and $\{y(t), t \in \mathbb{R}\}$ is defined as the Fourier transform of the cross-covariance function: $R_{xy}(f) = \int_{-\infty}^{\infty} \mathsf{E}\left[x(t)y(t+\tau)^*\right] \mathrm{e}^{-i2\pi f\tau} d\tau$, where E denotes expectation, * denotes complex conjugate, f shall be thought of as frequency and τ as time-lag. The cross-spectrum tells us at which frequencies the signals possess a high amount of linearly synchronized energy and it also gives information about phase coupling. The cross-spectrum does not evolve with time, as the processes are jointly stationary. Since the class of stationary processes is insufficient to model most phenomenon in nature, there have been attempts to generalize the cross-spectrum to the class of non-stationary processes. In this case, the cross-spectrum will be a function of both time and frequency. It is not trivial to make an appropriate extension of the definition of cross-spectrum suitable to the large class of non-stationary processes [8]. In this paper we will adopt the widely used Wigner cross-spectrum, $W_{xy}(t,f)$, which is defined by:

$$W_{xy}(t,f) = \int_{-\infty}^{\infty} \mathsf{E}\left[x\left(t+\frac{\tau}{2}\right) y\left(t-\frac{\tau}{2}\right)^*\right] \mathrm{e}^{-i2\pi f\tau} d\tau. \quad (1)$$

The calculation of the two-dimensional convolution between the kernel and the Wigner distribution of a process realization can be simplified using kernel decomposition and calculating multiple window spectrograms, [9, 10, 4]. The time-lag estimation kernel is rotated and the corresponding eigenvectors and eigenvalues are calculated. The estimate of the Wigner spectrum is given as the weighted sum of the spectrograms of the data with the different eigenvectors as sliding windows and the eigenvalues as weights, [11].

In this paper we address the question of how to estimate $W_{xy}(t,f)$ when a realization of $\{x(t)\}$ and $\{y(t)\}$ has been observed. In Section 2 we derive a formula for the MSE optimal estimator within a wide class of estimators. In Section 3 we will use this formula to construct an estimator optimal for a simple model. This estimator can also be computed using multiple windows as described in Section 4. In Section 5 we show that the estimator works well compared to other common estimators, such as Welch and Thomson, when applied to simulated data. Section 6 concludes the paper.

2. THE MEAN SQUARE ERROR OPTIMAL SOLUTION

The object of this paper is to estimate the Wigner cross-spectrum, $W_{xy}(t,f)$, defined in (1), using one observed realization of the non-stationary stochastic processe $\{x(t),\, t \in \mathbb{R}\}$ and one realization of $\{y(t),\, t \in \mathbb{R}\}$. A quite general non-parametric estimator is given by:

$$\widehat{W}_{xy,\Phi}(t,f) = \int\int \Phi(t-t_1, f-f_1) \int x\left(t_1+\frac{\tau}{2}\right) y\left(t_1-\frac{\tau}{2}\right) \times \mathrm{e}^{-i2\pi f_1\tau} d\tau dt_1 df_1,$$

where Φ is a smoothing kernel function which has to be appropriate chosen. All common non-parametric Wigner cross-spectrum estimators (e.g. Short Time Fourier Transform, Welch-method, all multi-taper methods, etc) can be written on this form, with different smoothing kernel functions. The optimal smoothing kernel function to use is unfortunately problem dependent. We will now derive a relation between the MSE-optimal kernel function Φ_{opt} and some properties of the random process. That is, we would like to solve the following optimization problem:

$$\Phi_{\mathrm{opt}} = \arg \min_{\Phi:\mathbb{R}^2 \mapsto \mathbb{R}^2} \int\int \mathsf{E}\left|W_{xy}(t,f) - \widehat{W}_{xy;\Phi}(t,f)\right|^2 dt df.$$

To simplify the above relation, we introduce the following notations:

$$A_{xy}(\nu,\tau) = \int\int W_{xy} \mathrm{e}^{i2\pi(f\tau - t\nu)} df dt$$

$$\widehat{A}_{xy}(\nu,\tau) = \int x\left(t+\frac{\tau}{2}\right) y\left(t-\frac{\tau}{2}\right) \mathrm{e}^{-i2\pi t\nu} dt$$

$$\phi(\nu,\tau) = \int\int \Phi(t,f) \mathrm{e}^{i2\pi(f\tau - t\nu)} df dt$$

$$\phi_{\mathrm{opt}}(\nu,\tau) = \int\int \Phi_{\mathrm{opt}}(t,f) \mathrm{e}^{i2\pi(f\tau - t\nu)} df dt,$$

This work was supported by the Swedish Research Council.

where ϕ shall be thought of as the ambiguity kernel corresponding to the smoothing kernel Φ. Using Parsevals formula we see that

$$\int\int \mathsf{E}\left|W_{xy}(t,f)-\widehat{W}_{xy;\Phi}(t,f)\right|^2 dt\,df$$

$$= \int\int \mathsf{E}\left|A_{xy}(\nu,\tau)-\phi(\nu,\tau)\widehat{A}_{xy}(\nu,\tau)\right|^2 d\nu\, d\tau.$$

And hence

$$\phi_{\text{opt}} = \arg\min_{\phi:\mathbb{R}^2\mapsto\mathbb{R}^2} \int\int \mathsf{E}\left|A_{xy}(\nu,\tau)-\phi(\nu,\tau)\widehat{A}_{xy}(\nu,\tau)\right|^2 d\nu\, d\tau.$$

We see that the solution to this minimisation problem is given by:

$$\phi_{\text{opt}}(\nu,\tau) = \frac{\left|\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right]\right|^2}{\left|\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right]\right|^2+\mathsf{V}\left[\widehat{A}_{xy}(\nu,\tau)\right]}.$$

The smoothing kernel is then given as

$$\Phi_{\text{opt}}(\nu,\tau) = \int\int \frac{\left|\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right]\right|^2}{\left|\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right]\right|^2+\mathsf{V}\left[\widehat{A}_{xy}(\nu,\tau)\right]} \times \mathrm{e}^{i2\pi(\nu t - f\tau)} df\, dt. \quad (2)$$

This result is analogous to the optimal kernel solution for non-stationary spectral estimation [3]. For any given ambiguity kernel function the MSE of the Wigner cross-spectrum can be computed by:

$$\varepsilon_\phi = \int\int \mathsf{E}\left[\left|W_{xy}(t,f)-\widehat{W}_{xy,\Phi}(t,f)\right|^2\right] dt\, df$$

$$= \int\int \left(|1-\phi(\nu,\tau)|^2\left|\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right]\right|^2 + |\phi(\nu,\tau)|^2\mathsf{V}\left[\widehat{A}_{xy}(\nu,\tau)\right]\right) d\nu\, d\tau.$$

The expectation, $\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right]$, and the variance, $\mathsf{V}\left[\widehat{A}_{xy}(\nu,\tau)\right]$, can be expressed in terms of the auto covariance functions, $r_x(s,t)$ and $r_y(s,t)$, and the cross-covariance function $r_{xy}(s,t)$:

$$\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right] = \int r_{xy}\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right)\mathrm{e}^{-i2\pi t\nu} dt,$$

and $\mathsf{V}\left[\widehat{A}_{xy}(\nu,\tau)\right] =$

$$-\left|\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right]\right|^2 + \int\int \mathsf{E}\left[x\left(t_1+\frac{\tau}{2}\right)x\left(t_2+\frac{\tau}{2}\right)^* y\left(t_1-\frac{\tau}{2}\right)y\left(t_2-\frac{\tau}{2}\right)^*\right]\mathrm{e}^{-i2\pi(t_1-t_2)\nu} dt_2\, dt_1$$

$$= \int\int \left(r_x\left(t_1+\frac{\tau}{2}, t_2+\frac{\tau}{2}\right) r_y\left(t_1-\frac{\tau}{2}, t_2-\frac{\tau}{2}\right) + r_{xy}\left(t_1+\frac{\tau}{2}, t_2-\frac{\tau}{2}\right) r_{xy}\left(t_2+\frac{\tau}{2}, t_1-\frac{\tau}{2}\right)\right)\mathrm{e}^{-i2\pi(t_1-t_2)\nu} dt_2\, dt_1,$$

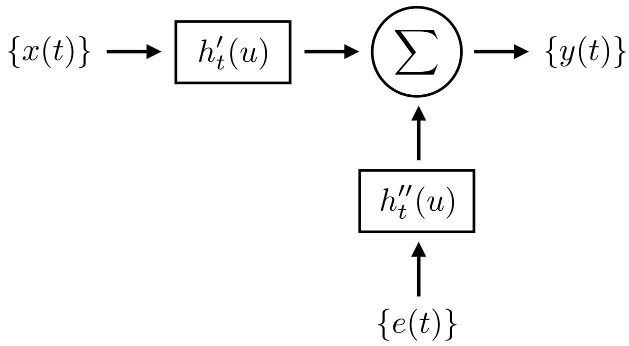

Figure 1: A time-varying filter model.

where the last equality holds for zero-meaned real-valued Gaussian processes. In next section we will introduce a quite general model. Under some simplifications we will be able to compute the optimal smoothing kernel for Wigner cross-spectrum estimation for this model.

3. A TIME-VARYING LINEAR FILTER MODEL

3.1 The model

A quite general linear filter model with time varying impulse response functions, $h'_t(u)$, and $h''_t(u)$ is described in Figure 1. In this paper we study the optimal kernel for time-frequency spectral estimation in a special case of this model. Let $\{x(t), t\in\mathbb{Z}\}$ be a white Gaussian process with variance σ_x^2 and let $\{y(t)\}$ be defined by:

$$z(t) = az(t-1)+x(t) \quad , \quad |a|<1 \quad (3)$$

$$y(t) = h(t)z(t)+g(t)e(t) \quad (4)$$

where $\{e(t), t\in\mathbb{Z}\}$ is a white Gaussian process with variance σ_e^2 and independent of $\{x(t)\}$, h and g are different Hanning windows. This simple model is well chosen to allow us to study optimal kernels for a wide range of different processes. The width of the Hanning envelopes h and g determines the non-stationarity of the process. The parameter a plays the important role of determining the width of the spectral content, as shown by Figure 2. The optimal kernel is indifferent to frequency shifts, which means that the fact that the spectral content in our model reaches its maximum at either 0 or 0.5 does not limit the models generality. This is easily seen in the Wigner domain, where the kernel is to be 2D convoluted in the time-frequency plane.

3.2 The optimal kernel

The expectation and variance of $\widehat{A}_{xy}(\nu,\tau)$ for this system is given by:

$$\mathsf{E}\left[\widehat{A}_{xy}(\nu,\tau)\right] = \begin{cases} 0 & \tau<0 \\ a^\tau\sigma_x^2 H(\nu)\mathrm{e}^{i\pi\tau\nu} & \tau\geq 0 \end{cases} \quad (5)$$

and $\mathsf{V}\left[\widehat{A}_{xy}(\nu,\tau)\right] =$

$$\begin{cases} \frac{\sigma_x^4}{1-a^2}\sum_t h(t)^2 + \sigma_e^2\sigma_x^2\sum_t g(t)^2 & \tau<0 \\ \sigma_e^2\sigma_x^2\sum_t g(t)^2 + \sigma_x^4 H(\nu)^2 * \left(\frac{1}{1-a^2}+a^{2\tau}\frac{\sin(2\pi\tau\nu)}{2\pi\nu}\right) & \tau\geq 0 \end{cases} \quad (6)$$

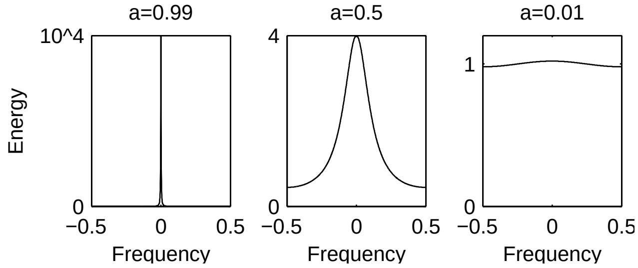

Figure 2: The model parameter a determines the frequency width of $\{z(t)\}$. Since kernel optimization is indifferent to frequency shifts, the shape of the frequency content, but not its location on the frequency axis, will determine the optimal kernel.

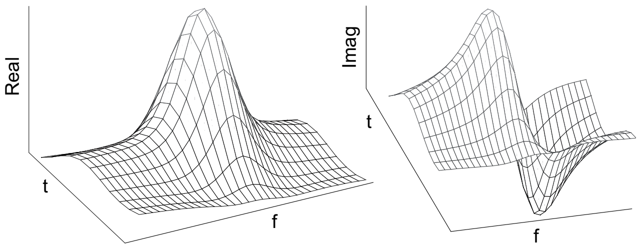

Figure 3: Example of the optimal kernel for our model.

where $*$ denotes convolution and H is the Fourier transform of h. An example of the optimal kernel is shown in Figure 3. As expected, signals with narrow spectral content correspond to an optimal kernel which is narrow along the frequency axis in Wigner domain. This is illustrated in Figure 4. A similar relation holds between the widths of the Hanning windows h and g and the width of the optimal kernel along the time axis.

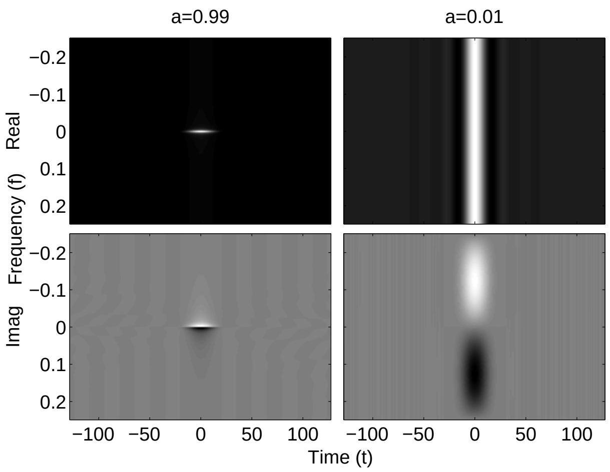

Figure 4: Optimal kernels for different values on a. Here, h is a Hanning window of length 32.

4. MULTIPLE WINDOWS

Instead of calculating the time-frequency estimate using the kernel, it is possible to simplify the calculations using a multiple window spectrogram, [11]. The Cohen's class is written

$$\begin{aligned}\widehat{W}_{xy}(t,f) &= \int\int A_{xy}(\nu,\tau)\phi(\nu,\tau)\mathrm{e}^{-i2\pi(\tau f-t\nu)}d\tau d\nu \\ &= \int\int r_{xy}(u,\tau)\rho(t-u,\tau)\mathrm{e}^{-i2\pi f\tau}du d\tau \\ &= \int\int x(u+\frac{\tau}{2})y^*(u-\frac{\tau}{2})\rho(t-u,\tau)e^{-i2\pi f\tau}du d\tau.\end{aligned}$$

Using a change of variables $u=(t_1+t_2)/2$ and $\tau=t_1-t_2$ gives,

$$\begin{aligned}W_{xy}^{Q}(t,f) &= \int\int x(t_1)y^*(t_2)\rho\left(t-\frac{t_1+t_2}{2},t_1-t_2\right) \\ &\quad\times\mathrm{e}^{-i2\pi f(t_1-t_2)}dt_1dt_2 \\ &= \int\int x(t_1)y^*(t_2)\rho^{rot}(t-t_1,t-t_2) \\ &\quad\times\mathrm{e}^{-i2\pi ft_1}\mathrm{e}^{i2\pi ft_2}dt_1dt_2 \qquad (7)\end{aligned}$$

where

$$\rho^{rot}(t_1,t_2)=\rho\left(\frac{t_1+t_2}{2},t_1-t_2\right).$$

The general kernel can be expressed as

$$\rho^{rot}(t_1,t_2)=\sum_{k=1}^{\infty}\lambda_k u_k(t_1)v_k^*(t_2),$$

using singular value decomposition resulting in singular values λ_k and different complete sets u_k and v_k. Using the singular values and singular vectors, Eq. (7) is rewritten as a weighted sum of spectrograms,

$$\begin{aligned}W_{xy}^{Q}(t,f) &= \sum_{k=1}^{\infty}\lambda_k\int\int x(t_1)y^*(t_2)\mathrm{e}^{-i2\pi ft_1}\mathrm{e}^{i2\pi ft_2} \\ &\quad\times u_k(t-t_1)v_k^*(t-t_2)dt_1dt_2. \\ &= \sum_{k=1}^{\infty}\lambda_k\left(\int x(t_1)\mathrm{e}^{-i2\pi ft_1}u_k^*(t-t_1)dt_1\right) \\ &\quad\times\left(\int y(t_2)\mathrm{e}^{-i2\pi ft_2}v_k(t-t_2)dt_2\right)^*.\end{aligned}$$

Depending on the different λ_k the number of spectrograms that are averaged could be just a few or an infinite number. With just a few λ_k that differs from zero the multiple window spectrogram solution is an effective solution from implementation aspects.

5. A COMPARISON WITH OTHER METHODS

We can use the MSE to compare different estimation methods. Figure 7 shows the MSE, normalized with respect to the signal energy, as a function of the model parameter a. We have compared three kernels that are optimal for $a=0.01$, $a=0.5$ and $a=0.99$ with the Welch method (window length=32 and number of windows=8), and Thomson multiple windows (window length=16 and number of windows=8), [12, 13]. Each kernel or set of windows is, for

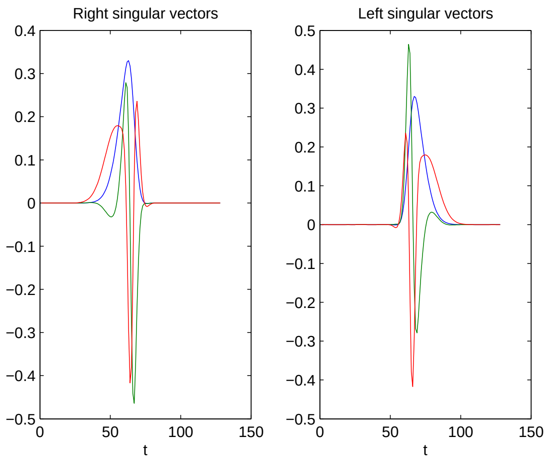

Figure 5: The singular vectors corresponding to the kernel depicted in Figure 3.

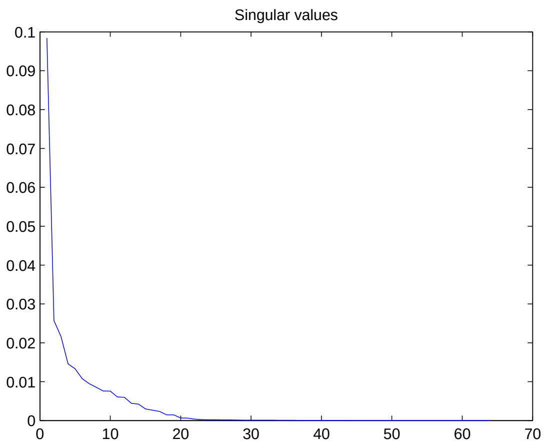

Figure 6: The singular values corresponding to the kernel depicted in Figure 3.

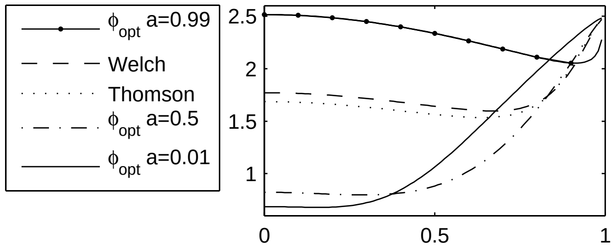

Figure 7: Logarithm of the MSE (normalized with the signal energy) for different estimating methods as a function of the model parameter a.

each a, multiplied with the mean square error optimal scalar. We see that the optimal kernels are not extremely sensitive to the model parameter a. The optimal kernel for $a = 0.5$ performs better than both Thomson multiple windows and Welch method in most cases.

6. CONCLUSIONS

We have derived the optimal Cohen class estimator of Wigner cross-spectrum. For a simple model we have compared the optimal estimator with Thomson multiple windows and the Welch method. We found that Thomson multiple windows and the Welch method works well in some cases, but they are often far from optimal.

REFERENCES

[1] M. B. Priestley, "Evolutionary spectra and non-stationary processes," *J. Roy. Statistics Soc. ser. B*, vol. 27, pp. 204–229, 1965.

[2] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1461–1470, 1985.

[3] A. M. Sayeed and D. L. Jones, "Optimal kernels for nonstationary spectral estimation," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 478–491, Feb. 1995.

[4] 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.

[5] 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.

[6] M. G. Amin, "Minimum variance time-frequency distribution kernels for signals in additive noise," *IEEE Trans. on Signal Processing*, vol. 44, no. 9, pp. 2352–2356, Sept 1996.

[7] S. Aviyente and W. J. Williams, "Minimum entropy time-frequency distributions," *IEEE Signal Processing Letters*, vol. 12, pp. 37–40, January 2005.

[8] Langford B. White and Boualem Boashash, "Cross spectral analysis of nonstationary processes," *IEEE*

Transactions of Information Theory, vol. 36, no. 4, pp. 830–835, July 1990.

[9] 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.

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

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

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

[13] 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.