PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 30, Fasc. 2 (2010), pp. 369–381

REGULARIZATION OF KERNELS FOR ESTIMATION OF THE WIGNER SPECTRUM OF GAUSSIAN STOCHASTIC PROCESSES

BY

PATRIK WAHLBERG (TURIN)

Abstract. We study estimation of the Wigner time-frequency spectrum of Gaussian stochastic processes. Assuming the covariance belongs to the Feichtinger algebra, we construct an estimation kernel that gives a mean square error arbitrarily close to the infimum over kernels in the Feichtinger algebra.

2000 AMS Mathematics Subject Classification: Primary: 60G15, 42B35, 60G35, 62M15, 94A12.

Key words and phrases: Time-frequency analysis, Gaussian stochastic processes, Wigner distribution, Wigner spectrum, minimum mean square error estimation, Cohen's class, the Feichtinger algebra.

1. INTRODUCTION

This paper treats time-frequency analysis of scalar-valued stochastic processes defined on $\mathbb{R}^d$. Time-frequency analysis is a relatively young field of study in the intersection of harmonic analysis, applied mathematics and signal processing ([4], [5], [7]), aiming at representation of mathematical objects (functions, operators, etc.) in the time and frequency variables *simultaneously*. If the domain of functions represents time, then $d = 1$, but $d > 1$ is also of interest (e.g. in image analysis). An important tool for representation of a signal f defined on $\mathbb{R}^d$ in the time-frequency domain is the Wigner (or Wigner–Ville) distribution ([4], [5], [7])

$$W(t, \xi) = \int_{\mathbb{R}^d} f(t + \tau/2)\overline{f(t - \tau/2)}e^{-i\tau\xi}d\tau,$$

which can be interpreted as a distribution of the signal's energy over the time-frequency domain $(t, \xi) \in \mathbb{R}^{2d}$. However, this interpretation is bothered by the fact that W is very seldomly non-negative everywhere, which can be seen as a consequence of the uncertainty principle [7]. As a remedy one can smooth W with a convolution kernel according to $W * \Phi$, which is called a *time-frequency representation in Cohen's class* determined by Φ ([4], [7]).

If we replace f by a stochastic process X, then W will be called the *Wigner process*. In this case the integral above is a stochastic mean square Riemann integral (see [17]). The *Wigner spectrum* ([4], [10], [11], [15]) is defined by

$$W_E(t,\xi) = \int_{\mathbb{R}^d} E\big(X(t+\tau/2)\overline{X(t-\tau/2)}\big)e^{-i\tau\xi}d\tau,$$

which equals the expectation of the Wigner distribution for certain processes [11]. The Wigner spectrum is a generalization of the spectral measure of a weakly stationary process to certain nonstationary processes. It contains "musical score" information of the process [7], i.e. it describes the time-variable frequency contents of the process.

In this paper we treat convolution kernels for estimation of the Wigner spectrum W_E of a zero mean Gaussian stochastic process X, using the Wigner process W. We use the following results from [17]. The Wigner process and the so-called ambiguity process A (the Fourier transform of W) of X are finite variance processes defined on $\mathbb{R}^{2d}$ provided the covariance function r belongs to the Feichtinger algebra $S_0(\mathbb{R}^{2d})$ (see [1] and [7], Chapters 11–12). *Cohen's class*, i.e. the stochastic convolution integral $W * \Phi$, where Φ is a deterministic kernel, is a finite variance process if Φ is continuous and bounded. Here $W * \Phi$ is a stochastic mean square Riemann integral. If moreover $\Phi \in S_0(\mathbb{R}^{2d})$, then $W * \Phi$ equals the stochastic Fourier integral of the process $A\phi$ which is the ambiguity process multiplied by the kernel $\phi = \mathcal{F}_1\mathcal{F}_2^{-1}\Phi$, where $\mathcal{F}_1$ and $\mathcal{F}_2$ denote the partial Fourier transforms of a function defined on $\mathbb{R}^{2d} = \mathbb{R}^d \oplus \mathbb{R}^d$ with respect to the first and the second $\mathbb{R}^d$ variable, respectively. Hence $\mathcal{F}_1$ is defined by

$$\mathcal{F}_1 f(\xi,y) = \int_{\mathbb{R}^d} f(x,y)e^{-ix\xi}dx, \qquad \xi,y \in \mathbb{R}^d,\ f \in L^1(\mathbb{R}^{2d}),$$

and $\mathcal{F}_2$ is defined analogously [13]. Thus we have a case of the well-known principle that convolution is equivalent to multiplication on the Fourier transform side.

We use $W * \Phi$ as an estimator of W_E. A natural criterion of performance of the estimator is the integral of the mean square error $E|W * \Phi(t,\xi) - W_E(t,\xi)|^2$ over $(t,\xi) \in \mathbb{R}^{2d}$. It is of interest to choose a kernel Φ such that the integral is minimized. If $r \in S_0(\mathbb{R}^{2d})$ and $\Phi \in S_0(\mathbb{R}^{2d})$, we can use a result of Parseval type from Wahlberg [17] saying that the integral equals the integral of the expression $E\big|(2\pi)^d A(\theta,\tau)\phi(\theta,\tau) - E\big(A(\theta,\tau)\big)\big|^2$ over $(\theta,\tau) \in \mathbb{R}^{2d}$. Due to the multiplicative action of the kernel in the ambiguity domain (θ,τ), it is easier to compute the optimal kernel there. It turns out to belong to $L^\infty(\mathbb{R}^{2d})$.

However, there is no guarantee that the optimal kernel ϕ_{opt} in $L^\infty(\mathbb{R}^{2d})$ is a member of $S_0(\mathbb{R}^{2d})$. Since we need $\phi \in S_0(\mathbb{R}^{2d})$ to ensure that our formulas are true, we have to make an approximation procedure. Given $\epsilon > 0$ we show how to construct a kernel $\Phi \in \mathcal{S}(\mathbb{R}^{2d})$ which gives a criterion value within ϵ from the infimum over $\Phi \in S_0(\mathbb{R}^{2d})$. Here $\mathcal{S}(\mathbb{R}^{2d})$ denotes the Schwartz space of smooth functions such that a derivative of any order multiplied by any polynomial is uniformly bounded [13].

Many of the invariance properties of the Feichtinger algebra $S_0(\mathbb{R}^d)$ (see [1], [3], [7]) will be used. The function space $S_0(\mathbb{R}^d)$ is defined as follows. Given a *window* function $g \in \mathcal{S}(\mathbb{R}^d)$, the *short-time Fourier transform* of $f \in L^2(\mathbb{R}^d)$ is defined by

$$V_g f(t,\xi) = \int_{\mathbb{R}^d} f(x)\overline{g(x-t)}e^{-ix\xi}dx.$$

The window function g is often chosen to have a short duration, and then $V_g f$ gives a localized Fourier transform, i.e. a time-frequency representation, of f. Now $S_0(\mathbb{R}^d)$ is defined as all tempered distributions [13] f such that

$$\iint_{\mathbb{R}^{2d}} |V_g f(t,\xi)|dtd\xi < \infty.$$

The L^1-norm on $V_g f$ may be replaced by a mixed $L^{p,q}$-norm, $1 \leqslant p,q \leqslant \infty$, that is, by the requirement

$$\Big(\int_{\mathbb{R}^d}\big(\int_{\mathbb{R}^d} |V_g f(t,\xi)|^p dt\big)^{q/p} d\xi\Big)^{1/q} < \infty.$$

This is the definition of a family of spaces known as modulation spaces ([2], [7]) of which $S_0(\mathbb{R}^d)$ is the smallest space. The function space $S_0(\mathbb{R}^d)$ contains $\mathcal{S}(\mathbb{R}^d)$, is contained in $L^1(\mathbb{R}^d) \cap \mathcal{F}L^1(\mathbb{R}^d)$, and is invariant under: (i) conjugation, (ii) linear coordinate transformations, (iii) tensorization, (iv) restriction to subgroups of $\mathbb{R}^d$, and (v) the (partial) Fourier transform defined by the Lebesgue integration. Here $\mathcal{F}L^1(\mathbb{R}^d)$ is the space of functions such that the Fourier transform belongs to $L^1(\mathbb{R}^d)$. The Lebesgue integral in the (partial) Fourier transform can be replaced by a Riemann integral which is important in this paper. The space $S_0(\mathbb{R}^d)$ is moreover a Banach algebra, both under convolution and pointwise multiplication.

In this paper, the Fourier transform is defined by

$$\mathcal{F}f(\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}^d} f(t)e^{-it\xi}dt,$$

which means that the inverse Fourier transform is

$$\mathcal{F}^{-1}\widehat{f}(t) = (2\pi)^{-d}\int_{\mathbb{R}^d}\widehat{f}(\xi)e^{it\xi}d\xi.$$

We denote the partial Fourier transform with respect to $\mathbb{R}^d$ variables indexed by j,k by $\mathcal{F}_{j,k}f$ or $\widehat{f}_{j,k}$. The Lebesgue measure is denoted by μ, and $L^2(\Omega)$ is the set of finite variance $\mathbb{R}$- or $\mathbb{C}$-valued random variables on a probability space $(\Omega,\mathcal{B},P)$. We restrict ourselves to stochastic processes $X : \mathbb{R}^d \mapsto L^2(\Omega)$ that have zero mean and are Gaussian, i.e. $\{X(t_j)\}_{j=1}^N$ have multidimensional Gaussian probability density for any finite set $\{t_j\}_{j=1}^N \subset \mathbb{R}^d$ (see [9], [12]). The covariance function is denoted by $r(t,s) = E\big(X(t)\overline{X(s)}\big)$. In the case of $\mathbb{C}$-valued processes we require the following property.

DEFINITION 1.1. A *circularly symmetric* process X is zero mean $\mathbb{C}$-valued and such that the processes $\{e^{i\theta}X\}_{\theta\in[0,2\pi)}$ have identical probability distributions for all $\theta \in [0, 2\pi)$.

According to Grettenberg's theorem (see [6] and [12]) circular symmetry is equivalent to

$$E\big(X(t)X(s)\big) = 0 \quad \text{for all } (t,s) \in \mathbb{R}^{2d}. \tag{1.1}$$

The following moment function will be needed later.

DEFINITION 1.2. For $t_1, \tau_1, t_2, \tau_2 \in \mathbb{R}^d$ we set

$$c(t_1,\tau_1,t_2,\tau_2) = E\big(X(t_1+\tau_1/2)\overline{X(t_1-\tau_1/2)X(t_2+\tau_2/2)}X(t_2-\tau_2/2)\big).$$

If x_1, x_2, x_3, x_4 are zero mean $\mathbb{R}$-valued or $\mathbb{C}$-valued random variables, then Isserlis' theorem ([8], p. 244) states that

$$E(x_1x_2x_3x_4) = E(x_1x_2)E(x_3x_4) + E(x_1x_3)E(x_2x_4) + E(x_1x_4)E(x_2x_3). \tag{1.2}$$

For Gaussian $\mathbb{R}$-valued or $\mathbb{C}$-valued circularly symmetric processes it follows from (1.1), Definition 1.2, (1.2) and the properties of $S_0(\mathbb{R}^{2d})$ that

$$r \in S_0(\mathbb{R}^{2d}) \Longrightarrow c \in S_0(\mathbb{R}^{4d}) \Longleftrightarrow \widehat{c}_{1,3} \in S_0(\mathbb{R}^{4d}). \tag{1.3}$$

2. THE WIGNER SPECTRUM AND THE WIGNER PROCESS

As stated in the Introduction the Wigner spectrum of a process X with covariance r is defined by

$$W_E(t,\xi) = \int_{\mathbb{R}^d} r(t+\tau/2, t-\tau/2)e^{-i\tau\xi}d\tau = \mathcal{F}_2(r\circ\kappa)(t,\xi), \quad t,\xi \in \mathbb{R}^d,$$

where $\kappa(t,\tau) = (t+\tau/2, t-\tau/2)$ for $t, \tau \in \mathbb{R}^d$ ([4], [10], [11], [15]). The *expected ambiguity function* of the process is defined by

$$A_E(\theta,\tau) = \int_{\mathbb{R}^d} r(t+\tau/2, t-\tau/2)e^{-it\theta}dt, \quad \theta,\tau \in \mathbb{R}^d, \tag{2.1}$$

and $A_E = \mathcal{F}_1\mathcal{F}_2^{-1}W_E$. If $r \in S_0(\mathbb{R}^{2d})$, then $W_E, A_E \in S_0(\mathbb{R}^{2d})$ due to the invariance properties of S_0 (see [1]).

We shall use the following results from [10], [11], [15], [17]. Let X be a zero mean $\mathbb{R}$-valued or $\mathbb{C}$-valued circularly symmetric, Gaussian process, and let $r \in S_0(\mathbb{R}^{2d})$. Then the following stochastic Riemann integrals [9] are well defined:

$$\begin{aligned} W(t,\xi) &= \int_{\mathbb{R}^d} X(t+\tau/2)\overline{X(t-\tau/2)}e^{-i\tau\xi}d\tau \in L^2(\Omega) \quad \text{for all } (t,\xi)\in\mathbb{R}^{2d}, \\ A(\theta,\tau) &= \int_{\mathbb{R}^d} X(t+\tau/2)\overline{X(t-\tau/2)}e^{-it\theta}dt \in L^2(\Omega) \quad \text{for all } (\theta,\tau)\in\mathbb{R}^{2d}. \end{aligned} \tag{2.2}$$

Thus W and A are second order stochastic processes on $\mathbb{R}^{2d}$. We call them the *Wigner process* and the *ambiguity process* of X, respectively. We have

$$
\begin{aligned}
E\big(W(t,\xi)\big) &= W_E(t,\xi) \quad \text{for all } (t,\xi) \in \mathbb{R}^{2d}, \\
E\big(A(\theta,\tau)\big) &= A_E(\theta,\tau) \quad \text{for all } (\theta,\tau) \in \mathbb{R}^{2d},
\end{aligned} \tag{2.3}
$$

and for the covariance of A:

$$
E\big(A(\theta_1,\tau_1)\overline{A(\theta_2,\tau_2)}\big) = \widehat{c}_{1,3}(\theta_1,\tau_1,-\theta_2,\tau_2) \in S_0(\mathbb{R}^{4d}) \tag{2.4}
$$

by (1.3). If $\Phi \in S_0(\mathbb{R}^{2d})$, then the convolution stochastic Riemann integral

$$
W * \Phi(t,\xi) = \iint_{\mathbb{R}^{2d}} W(t-s,\xi-\eta)\Phi(s,\eta)dsd\eta \in L^2(\Omega) \quad \text{for all } (t,\xi) \in \mathbb{R}^{2d} \tag{2.5}
$$

is a second order stochastic process on $\mathbb{R}^{2d}$. The convolution $W * \Phi$ is said to belong to *Cohen's class* defined by Φ (see [4]). The process $W * \Phi$ can be represented by the stochastic Riemann integral

$$
W * \Phi(t,\xi) = \iint_{\mathbb{R}^{2d}} A(\theta,\tau)\phi(\theta,\tau)e^{i(\theta t-\tau\xi)}d\theta d\tau, \quad t,\xi \in \mathbb{R}^d, \tag{2.6}
$$

where $\phi = \mathcal{F}_1\mathcal{F}_2^{-1}\Phi \in S_0(\mathbb{R}^{2d})$. Finally, we have the following result.

Suppose W_1 and A_1 are two second order stochastic processes on $\mathbb{R}^{2d}$ related through the Fourier stochastic Riemann integral

$$
W_1(t,\xi) = (2\pi)^{-d} \iint_{\mathbb{R}^{2d}} A_1(\theta,\tau)e^{i(\theta t-\tau\xi)}d\theta d\tau \quad \text{for all } (t,\xi) \in \mathbb{R}^{2d} \tag{2.7}
$$

and suppose that

$$
g(\theta_1,\tau_1,\theta_2,\tau_2) = E\big(A_1(\theta_1,\tau_1)\overline{A_1(\theta_2,\tau_2)}\big) \in S_0(\mathbb{R}^{4d}). \tag{2.8}
$$

Then we have the Riemann integral identity

$$
\iint_{\mathbb{R}^{2d}} E|W_1(t,\xi)|^2 dtd\xi = \iint_{\mathbb{R}^{2d}} E|A_1(t,\xi)|^2 d\theta d\tau < \infty. \tag{2.9}
$$

3. ESTIMATION OF W_E FROM W USING COHEN'S CLASS

The convolution formula (2.5) can be used for estimation of W_E from realizations of the process W. This is of interest in many applications since W_E gives a time-frequency picture of the process. An optimal kernel for this problem was first determined by Sayeed and Jones [15]. The paper [17] and the present paper is an attempt to refine their results.

In fact, first we give sufficient conditions for the optimal ambiguity domain kernel to exist. Moreover, since the optimal kernel in general needs not to belong to the space $S_0(\mathbb{R}^{2d})$, which justifies the transformation to the ambiguity domain, we give a regularization scheme in order to approximate the optimal kernel with a kernel in $S_0(\mathbb{R}^{2d})$. The result is the following.

Given any $\epsilon > 0$, we can compute a convolution kernel which gives an error within ϵ from the infimum over kernels in $S_0(\mathbb{R}^{2d})$.

Suppose $\Phi \in S_0(\mathbb{R}^{2d})$. We define the mean square error integral

$$J(\Phi) = \iint\limits_{\mathbb{R}^{2d}} E|W * \Phi(t,\xi) - W_E(t,\xi)|^2 dt d\xi$$

to be minimized by the kernel Φ, and put

$$J_o = \inf_{\Phi \in S_0(\mathbb{R}^{2d})} J(\Phi).$$

In the following proposition we compute a kernel $\Phi_{\delta_n,\gamma} \in \mathcal{S}(\mathbb{R}^{2d})$, indexed by a natural number n and a positive real γ, such that $J(\Phi_{\delta_n,\gamma})$ is arbitrarily close to J_o for sufficiently large n and sufficiently small γ. As preparation we choose a standard mollifier $\psi \in C_c^\infty(\mathbb{R}^{2d})$ which denotes the set of smooth functions of compact support. Let ψ have support in the unit ball, $\int_{\mathbb{R}^{2d}} \psi(x)dx = 1$ and $\psi \geqslant 0$. We define $\psi_\gamma(x) = \gamma^{-2d}\psi(x/\gamma)$ for $\gamma > 0$. Let $B(\delta^{-1}) \subset \mathbb{R}^{2d}$ denote the closed ball of radius δ^{-1}, define (subscript r for *restriction*)

$$c_r(\theta,\tau) := \widehat{c}_{1,3}(\theta,\tau,-\theta,\tau). \tag{3.1}$$

Then $c_r \in S_0(\mathbb{R}^{2d})$ since S_0 is invariant under restrictions to subgroups [1]. Let

$$U(\delta) := \{(\theta,\tau) \in \mathbb{R}^{2d};\ c_r(\theta,\tau) \geqslant \delta\} \subset \{(\theta,\tau) \in \mathbb{R}^{2d};\ c_r(\theta,\tau) > 0\} := U, \tag{3.2}$$

where $U(\delta)$, $\delta > 0$, is a closed subset of the open set U since $\widehat{c}_{1,3} \in S_0(\mathbb{R}^{4d}) \subset C(\mathbb{R}^{4d})$, which means the continuous and bounded functions [1]. Let χ_B denote the indicator function of the set B and define

$$\phi_o(\theta,\tau) = \frac{|A_E(\theta,\tau)|^2}{(2\pi)^d c_r(\theta,\tau)} \chi_U(\theta,\tau). \tag{3.3}$$

The function ϕ_o restricted to U is continuous. From (2.3), (2.4), (3.1) and the Cauchy–Schwarz inequality we have $0 \leqslant \phi_o(\theta,\tau) \leqslant (2\pi)^{-d}$. Thus $\phi_o \in \mathcal{S}'(\mathbb{R}^{2d})$, which means the set of tempered distributions [13], and we can define

$$\Phi_o := \mathcal{F}_1^{-1}\mathcal{F}_2\phi_o \in \mathcal{S}'(\mathbb{R}^{2d}).$$

THEOREM 3.1. *Suppose X is a zero mean Gaussian $\mathbb{R}$-valued or $\mathbb{C}$-valued circularly symmetric process whose covariance function $r \in S_0(\mathbb{R}^{2d})$. Let $\delta, \gamma > 0$. Define*

$$\begin{aligned} \phi_1 &= \phi_{1,\delta} = \phi_o \chi_{B(\delta^{-1}) \cap U(\delta)}, \\ \phi_2 &= \phi_{2,\delta,\gamma} = \phi_{1,\delta} * \psi_\gamma \end{aligned} \tag{3.4}$$

and

$$\Phi_{\delta,\gamma} = \mathcal{F}_1^{-1} \mathcal{F}_2 \phi_{2,\delta,\gamma},$$

where the notation $\phi_1 = \phi_{1,\delta}$ indicates that ϕ_1 depends on δ, etc. Then $\phi_2 \in C_c^\infty(\mathbb{R}^{2d})$, and hence $\Phi_{\delta,\gamma} \in \mathcal{S}(\mathbb{R}^{2d})$. There exists a sequence $\{\delta_n\}_{n=1}^\infty$, $\delta_n \downarrow 0$, with the following property:

Given any $\epsilon > 0$, for n sufficiently large and γ sufficiently small we have

$$J(\Phi_{\delta_n,\gamma}) - J_o < \epsilon. \tag{3.5}$$

Furthermore, we have the limit in $\mathcal{S}'(\mathbb{R}^{2d})$:

$$\lim_{n \to \infty,\ \gamma \to 0} \Phi_{\delta_n,\gamma} = \Phi_o. \tag{3.6}$$

Proof. Suppose $\Phi \in S_0(\mathbb{R}^{2d})$, and so also $\phi \in S_0(\mathbb{R}^{2d})$. Moreover, since we also have $r \in S_0(\mathbb{R}^{2d})$, formula (2.6) and

$$W_E(t,\xi) = (2\pi)^{-d} \iint_{\mathbb{R}^{2d}} A_E(\theta,\tau) e^{i(\theta t - \tau\xi)} d\theta d\tau \tag{3.7}$$

are true. As explained in [17] an ordinary Riemann integral can be considered as a stochastic Riemann integral. Hence we can combine (2.6) and (3.7) into the stochastic Riemann integral

$$\begin{aligned} &W * \Phi(t,\xi) - W_E(t,\xi) \\ &\qquad = \frac{1}{(2\pi)^d} \iint_{\mathbb{R}^{2d}} \big((2\pi)^d A(\theta,\tau)\phi(\theta,\tau) - A_E(\theta,\tau)\big) e^{i(\theta t - \tau\xi)} d\theta d\tau. \end{aligned}$$

This is a representation according to (2.7). The argument showing that (2.8) holds for $A_1 = (2\pi)^d A\phi - A_E$ is as follows. Using (2.3) and (2.4) we obtain

$$\begin{aligned} g(\theta_1,\tau_1,\theta_2,\tau_2) &= E\big(A_1(\theta_1,\tau_1)\overline{A_1(\theta_2,\tau_2)}\big) \\ &= (2\pi)^{2d} \widehat{c}_{1,3}(\theta_1,\tau_1,-\theta_2,\tau_2)\phi(\theta_1,\tau_1)\overline{\phi(\theta_2,\tau_2)} \\ &\quad - (2\pi)^d A_E(\theta_1,\tau_1)\overline{A_E(\theta_2,\tau_2)}\big(\phi(\theta_1,\tau_1) + \overline{\phi(\theta_2,\tau_2)}\big) \\ &\quad + A_E(\theta_1,\tau_1)\overline{A_E(\theta_2,\tau_2)}. \end{aligned}$$

Now $g \in S_0(\mathbb{R}^{4d})$ follows from the facts that $\widehat{c}_{1,3} \in S_0(\mathbb{R}^{4d})$, $\phi, A_E \in S_0(\mathbb{R}^{2d})$, and the invariance properties of S_0. Hence (2.9) is true and we have

$$J(\Phi) = \iint_{\mathbb{R}^{2d}} E|(2\pi)^d A(\theta,\tau)\phi(\theta,\tau) - A_E(\theta,\tau)|^2 d\theta d\tau. \tag{3.8}$$

Let us consider $J(\Phi) = J(\mathcal{F}_1^{-1}\mathcal{F}_2\phi) := G(\phi)$ as a function of ϕ instead of Φ. Moreover, let us relax the condition $\phi \in S_0(\mathbb{R}^{2d})$ and let ϕ be an arbitrary function in $L^\infty(\mathbb{R}^{2d})$. Then it is clear that the integral (3.8) is minimized by a function which mimimizes the integrand pointwise, i.e.

$$\phi_{opt}(\theta,\tau) = \arg\min_{\phi(\theta,\tau)} E|(2\pi)^d A(\theta,\tau)\phi(\theta,\tau) - A_E(\theta,\tau)|^2.$$

By the principle of orthogonality in the Hilbert space $L^2(\Omega)$ (see [13]) we conclude that there is a unique $\phi_{opt}(\theta,\tau)$ such that

$$E\Big(\big((2\pi)^d A(\theta,\tau)\phi_{opt}(\theta,\tau) - A_E(\theta,\tau)\big)\overline{A(\theta,\tau)}\Big) = 0$$
$$\Longleftrightarrow (2\pi)^d E|A(\theta,\tau)|^2\phi_{opt}(\theta,\tau) = |A_E(\theta,\tau)|^2.$$

For $(\theta,\tau) \in \mathbb{R}^{2d}$ such that $E|A(\theta,\tau)|^2 = 0$, $\phi_{opt}(\theta,\tau)$ can be defined arbitrarily. We set it to zero, and using (2.4), (3.1) and (3.2) we thus obtain $\phi_{opt} = \phi_o$ defined by (3.3). By (3.8) and $S_0(\mathbb{R}^{2d}) \subset L^\infty(\mathbb{R}^{2d})$,

$$\infty > J_o = \inf_{\Phi \in S_0(\mathbb{R}^{2d})} J(\Phi) = \inf_{\phi \in S_0(\mathbb{R}^{2d})} G(\phi) \geqslant G(\phi_o). \tag{3.9}$$

There is however no guarantee that ϕ_o defined by (3.3) belongs to the space $S_0(\mathbb{R}^{2d})$ as required in the derivation of (3.8). Therefore we design an approximation procedure. We have $c_r \in S_0(\mathbb{R}^{2d}) \subset L^1(\mathbb{R}^{2d})$ and $c_r \geqslant 0$. Let $A \subset \mathbb{R}$. In what follows we denote the Lebesgue measure of the set $\{(\theta,\tau);\ c_r(\theta,\tau) \in A\} \subset \mathbb{R}^{2d}$ by $\mu(c_r \in A)$. We observe that the set $\{t > 0;\ \mu(c_r = t) > 0\}$ is countable. In fact, the function $t \mapsto \mu(c_r > t)$ defined for $t > 0$ is finite everywhere, nonnegative, continuous from the right and monotonically nonincreasing [14]. Thus it has at most countably many discontinuity points t_j, i.e. points t_j such that $0 < \lim_{t\uparrow t_j} \mu(c_r > t) - \mu(c_r > t_j) = \lim_{t\uparrow t_j} \mu(t < c_r \leqslant t_j) = \mu(c_r = t_j)$. Hence we can choose a sequence $\{\delta_n\}_{n=1}^\infty$ such that $\delta_n \downarrow 0$ and

$$\mu(c_r = \delta_n) = 0 \quad \text{for all } n \geqslant 1. \tag{3.10}$$

This property of the choice of a sequence $\{\delta_n\}_{n=1}^\infty$ will be used at the end of the proof. We apply the elements of the sequence $\{\delta_n\}_{n=1}^\infty$ in definition (3.4), i.e. we set

$$\phi_{1n} := \phi_o \chi_{B(\delta_n^{-1}) \cap U(\delta_n)},$$

and insert ϕ_{1n} as ϕ into (3.8), which gives the Riemann integral

$$G(\phi_{1n}) - G(\phi_o) = \iint_U \frac{|A_E(\theta,\tau)|^4}{c_r(\theta,\tau)}(1-\chi_{B(\delta_n^{-1})\cap U(\delta_n)})d\theta d\tau$$
$$\leqslant \iint_U c_r(\theta,\tau)(1-\chi_{B(\delta_n^{-1})\cap U(\delta_n)})d\theta d\tau$$

by (2.3), (2.4), (3.1) and the Cauchy–Schwarz inequality. Since $c_r \in S_0(\mathbb{R}^{2d}) \subset L^1(\mathbb{R}^{2d}) \cap C(\mathbb{R}^{2d})$, the integral of c_r is independent of whether it is defined in the Lebesgue or Riemann sense. This is also true for the integral over the compact set $B(\delta_n^{-1}) \cap U(\delta_n)$. Since $B(\delta_n^{-1}) \cap U(\delta_n) \uparrow U$ as $n \to \infty$, i.e. $\delta_n \downarrow 0$, we have, by continuity of the Lebesgue measure,

$$\lim_{n\to\infty} G(\phi_{1n}) - G(\phi_o) = 0. \tag{3.11}$$

Defining $\phi_2 = \phi_{1n} * \psi_\gamma$ we have $\phi_2 \in C_c^\infty(\mathbb{R}^{2d})$ since both ϕ_{1n} and ψ_γ have compact support and ψ_γ is smooth. Hence $\Phi_{\delta_n,\gamma} = \mathcal{F}_1^{-1}\mathcal{F}_2\phi_2 \in \mathcal{S}(\mathbb{R}^{2d})$ [13]. Since $\mathcal{S}(\mathbb{R}^{2d}) \subset S_0(\mathbb{R}^{2d})$, we have $J(\Phi_{\delta_n,\gamma}) = G(\phi_2)$. The bound $\|\phi_{1n}\|_{L^\infty} \leqslant (2\pi)^{-d}$ implies $\|\phi_2\|_{L^\infty} \leqslant (2\pi)^{-d}$. Using $\sup(c_r) < \infty$ and $\sup(A_E) < \infty$, from (3.8) we obtain

$$G(\phi_2) - G(\phi_{1n}) = (2\pi)^{2d} \iint_{\mathbb{R}^{2d}} c_r(\theta,\tau)\big(\phi_2^2(\theta,\tau) - \phi_{1n}^2(\theta,\tau)\big)d\theta d\tau$$
$$- 2(2\pi)^d \iint_{\mathbb{R}^{2d}} |A_E(\theta,\tau)|^2\big(\phi_2(\theta,\tau) - \phi_{1n}(\theta,\tau)\big)d\theta d\tau$$
$$\leqslant C\|\phi_2 - \phi_{1n}\|_{L^1(\mathbb{R}^{2d})},$$

where $C > 0$. Thus, if we show

$$\lim_{\gamma\to 0} \|\phi_2 - \phi_{1n}\|_{L^1(\mathbb{R}^{2d})} = 0, \tag{3.12}$$

then, using (3.9) and (3.11), we will have, for any $\epsilon > 0$,

$$J(\Phi_{\delta_n,\gamma}) - J_o \leqslant G(\phi_2) - G(\phi_{1n}) + G(\phi_{1n}) - G(\phi_o) < \epsilon$$

for n sufficiently large and γ sufficiently small, which proves (3.5).

Now we are going to prove (3.12). We define the compact set

$$K_n := B(\delta_n^{-1}) \cap U(\delta_n) \subset U \subset \mathbb{R}^{2d}.$$

Since $\operatorname{supp}(\phi_{1n}) \subset K_n$ and $\operatorname{supp}(\psi_\gamma) \subset B(\gamma)$, we have

$$\|\phi_2 - \phi_{1n}\|_{L^1(\mathbb{R}^{2d})} \leqslant \iint_{\substack{x\in K_n+B(\gamma)\\ y\in B(\gamma)}} |\phi_{1n}(x-y) - \phi_{1n}(x)|\psi_\gamma(y)dxdy. \tag{3.13}$$

We define a subset of K_n:

$$K_{n\gamma} := \{x \in K_n;\ x + B(\gamma) \subset K_n\}.$$

Then for n sufficiently large and γ sufficiently small $K_{n\gamma}$ is nonempty. In fact, if N is sufficiently large, there exists $x \in \mathbb{R}^{2d}$ such that $c_r(x) \geqslant 2\delta_n$, $n \geqslant N$. (Here we have assumed that c_r is not identically zero.) Thus $x \in U(\delta_n)$, $n \geqslant N$, and $x \in B(\delta_n^{-1})$, $n \geqslant M$ for a certain M, i.e. $x \in K_n$, $n \geqslant \max(N, M)$. Since c_r is continuous, there is a ball $B \ni x$ such that $c_r(x + B) \geqslant \delta_n$, $n \geqslant N$. Thus $x + B \subset K_n$ for n sufficiently large.

The set $K_{n\gamma}$ is bounded. It is also closed, and hence compact, for the following reason. Let $x_k \in K_{n\gamma}$ be a sequence such that $\lim_{k\to\infty} x_k = x$ and let $y \in B(\gamma)$. Then $x_k + y \in K_n$ for all k and $x + y \in K_n$ since K_n is closed. Since $y \in B(\gamma)$ is arbitrary, $x \in K_{n\gamma}$. Thus $K_{n\gamma}$ is a closed set. We split the integral over x in (3.13), and employ $\|\phi_{1n}\|_{L^\infty} \leqslant (2\pi)^{-d}$, according to

$$(3.14)\quad \|\phi_2 - \phi_{1n}\|_{L^1(\mathbb{R}^{2d})} \leqslant \iint_{\substack{x\in K_{n\gamma}\\ y\in B(\gamma)}} |\phi_{1n}(x-y) - \phi_{1n}(x)|\psi_\gamma(y)dxdy$$
$$+ \iint_{\substack{x\in (K_n+B(\gamma))\setminus K_{n\gamma}\\ y\in B(\gamma)}} |\phi_{1n}(x-y) - \phi_{1n}(x)|\psi_\gamma(y)dxdy$$
$$\leqslant \iint_{\substack{x\in K_{n\gamma}\\ y\in B(\gamma)}} |\phi_{1n}(x-y) - \phi_{1n}(x)|\psi_\gamma(y)dxdy + 2(2\pi)^{-d}\mu\Big(\big(K_n + B(\gamma)\big) \setminus K_{n\gamma}\Big),$$

where $\big(K_n + B(\gamma)\big) \setminus K_{n\gamma} = \big(K_n + B(\gamma)\big) \cap K_{n\gamma}^c$ and $K_{n\gamma}^c$ denotes the complement. Since K_n is compact and ϕ_o is continuous on $K_n \subset U$, the restriction of $\phi_{1n} = \phi_o \cdot \chi_{K_n}$ to K_n is uniformly continuous. In the integral on the right-hand side we have both $x - y \in K_n$ and $x \in K_n$. Hence, for any $\alpha > 0$, $|\phi_{1n}(x-y) - \phi_{1n}(x)| \leqslant \alpha/\mu(K_n)$ for all $x \in K_{n\gamma}$ and $y \in B(\gamma)$ if γ is sufficiently small, and thus the integral is less than or equal to α. Therefore, it tends to zero as $\gamma \to 0$. Thus (3.12) will be proved if we show

$$(3.15)\qquad \lim_{\gamma\to 0} \mu\Big(\big(K_n + B(\gamma)\big) \setminus K_{n\gamma}\Big) = 0.$$

We can write

$$(3.16)\quad \mu\Big(\big(K_n + B(\gamma)\big) \setminus K_{n\gamma}\Big) = \mu\Big(\big(K_n + B(\gamma)\big) \setminus K_n\Big) + \mu(K_n \setminus K_{n\gamma}).$$

As $\gamma \downarrow 0$, $K_n + B(\gamma)$ is shrinking towards K_n. Since the Lebesgue measure is continuous, we have

$$(3.17)\qquad \lim_{\gamma\downarrow 0} \mu\Big(\big(K_n + B(\gamma)\big) \setminus K_n\Big) = 0.$$

Let us look at the second term on the right-hand side of (3.16). $K_{n\gamma}$ increases as $\gamma \downarrow 0$ towards

$$\bigcup_{\gamma>0} K_{n\gamma} = K_n^o, \tag{3.18}$$

where $K_n^o = \overline{K_n^c}^c$ denotes the interior of K_n. Here $\overline{K_n}$ means closure. To prove the $\supset$ inclusion of (3.18) let $x \in K_n^o$. Since K_n^o is open, there exists a $\gamma > 0$ such that $x + B(\gamma) \subset K_n^o \subset K_n$, i.e. $x \in K_{n\gamma}$. For the opposite inclusion, let $x \notin K_n^o$ i.e. $x \in \overline{K_n^c}$. Then for any $\gamma > 0$ we have $x + B(\gamma) \cap K_n^c \neq \emptyset$, i.e. $x \notin K_{n\gamma}$. Thus (3.18) holds. Since K_n is closed, we have a disjoint union $K_n = \partial K_n \cup K_n^o$, where $\partial K_n = \overline{K_n} \cap \overline{K_n^c}$ denotes the boundary ([16], Chapter 2) of K_n. Now it follows again from continuity of the Lebesgue measure that

$$\lim_{\gamma\downarrow 0} \mu(K_n \setminus K_{n\gamma}) = \mu(\partial K_n). \tag{3.19}$$

We shall now finally show that $\mu(\partial K_n) = 0$ for all n, which together with (3.17) proves

$$\lim_{\gamma\downarrow 0} \mu\Big(\big(K_n + B(\gamma)\big) \setminus K_{n\gamma}\Big) = 0.$$

This proves (3.15), since $\big(K_n + B(\gamma)\big) \setminus K_{n\gamma}$ decreases as γ decreases, and hence we have shown (3.12). The claim $\mu(\partial K_n) = 0$ follows from

$$\begin{aligned}\partial K_n &= B(\delta_n^{-1}) \cap U(\delta_n) \cap \big(\overline{B(\delta_n^{-1})^c} \cup \overline{U(\delta_n)^c}\big)\\ &= \big(\partial B(\delta_n^{-1}) \cap U(\delta_n)\big) \cup \big(B(\delta_n^{-1}) \cap U(\delta_n) \cap \overline{U(\delta_n)^c}\big).\end{aligned}$$

In fact, since the first set is a subset of the boundary of a sphere, it has Lebesgue measure zero. For a point x in the second set we have $c_r(x) \geqslant \delta_n$ by the definition of $U(\delta_n)$, and the existence of a sequence $x_k \in U(\delta_n)^c$, i.e. $c_r(x_k) < \delta_n$, such that $\lim_{k\to\infty} x_k = x$. By continuity of c_r, we have $c_r(x) = \lim_{k\to\infty} c_r(x_k) \leqslant \delta_n$. So $c_r(x) = \delta_n$, and the second set has Lebesgue measure zero due to (3.10). Hence $\mu(\partial K_n) = 0$ for all n, which proves (3.15), and thereby also (3.12). Thus (3.5) has been proved.

It remains to prove (3.6), which is equivalent to the limit in $\mathcal{S}'(\mathbb{R}^{2d})$:

$$\lim_{\substack{n\to\infty\\ \gamma\to 0}} \phi_2 = \phi_o. \tag{3.20}$$

Choose an arbitrary $\varphi \in \mathcal{S}(\mathbb{R}^{2d})$ and denote by $(f,\varphi)_{\mathcal{S}',\mathcal{S}}$ the action of $f \in \mathcal{S}'(\mathbb{R}^{2d})$ on $\varphi \in \mathcal{S}(\mathbb{R}^{2d})$. Since $\|\phi_o\|_{L^\infty} \leqslant (2\pi)^{-d}$, we have

$$\begin{aligned}&|(\phi_o - \phi_{1n}, \varphi)_{\mathcal{S}',\mathcal{S}}|\\ &\qquad\leqslant \int_{U\setminus K_n} \phi_o(x)|\varphi(x)|dx \leqslant (2\pi)^{-d} \int_{U\setminus K_n} |\varphi(x)|dx \to 0\end{aligned} \tag{3.21}$$

as $n \to \infty$, since $K_n \uparrow U$ and $\varphi \in L^1(\mathbb{R}^{2d})$. Moreover, by splitting the integral as is done in (3.14) and by $\|\phi_{1n}\|_{L^\infty} \leqslant (2\pi)^{-d}$, we obtain

$$
\begin{aligned}
(3.22)\quad |(\phi_2 - \phi_{1n}, \varphi)_{\mathcal{S}',\mathcal{S}}| &\leqslant \iint_{\substack{x \in K_n + B(\gamma) \\ y \in B(\gamma)}} \psi_\gamma(y)|\varphi(x)||\phi_{1n}(x-y) - \phi_{1n}(x)|dxdy \\
&\leqslant \iint_{\substack{x \in K_{n\gamma} \\ y \in B(\gamma)}} \psi_\gamma(y)|\varphi(x)||\phi_{1n}(x-y) - \phi_{1n}(x)|dxdy \\
&\quad + 2(2\pi)^{-d}\|\varphi\|_{L^\infty(\mathbb{R}^{2d})}\mu\Big(\big(K_n + B(\gamma)\big) \setminus K_{n\gamma}\Big) \\
&\to 0 \quad \text{as } \gamma \to 0.
\end{aligned}
$$

Indeed, let $\alpha > 0$; then we have for the integrand on the right-hand side of (3.22), as in (3.14), $|\phi_{1n}(x-y) - \phi_{1n}(x)| \leqslant \alpha/\|\varphi\|_{L^1(\mathbb{R}^{2d})}$ when $x \in K_{n\gamma}$ and $y \in B(\gamma)$ if γ is sufficiently small. Thus the integral on the right-hand side of (3.22) is less than or equal to α, i.e. it tends to zero as $\gamma \to 0$. The second term tends to zero by (3.15). Combination of (3.21) and (3.22) now gives

$$
\lim_{\substack{n \to \infty \\ \gamma \to 0}} (\phi_2 - \phi_o, \varphi)_{\mathcal{S}',\mathcal{S}} = 0 \quad \text{for all } \varphi \in \mathcal{S}(\mathbb{R}^{2d}),
$$

i.e. (3.20) has been proved. Hence, (3.6) is shown, and the proof of the theorem is complete. ■

REFERENCES

[1] H. G. Feichtinger, *On a new Segal algebra*, Monatsh. Math. 92 (1981), pp. 269–289.

[2] H. G. Feichtinger, *Modulation spaces on locally compact abelian groups*, Technical Report, University of Vienna, 1983; also in: *Wavelets and Their Applications*, M. Krishna, R. Radha and S. Thangavelu (Eds.), Allied Publishers Private Limited, New Dehli Mumbai Kolkata Chennai Hagpur Ahmedabad Bangalore Hyderbad Lucknow, 2003, pp. 99–140.

[3] H. G. Feichtinger and G. Zimmermann, *A Banach space of test functions for Gabor analysis*, in: *Gabor Analysis and Algorithms – Theory and Applications*, H. G. Feichtinger and T. Strohmer (Eds.), Birkhäuser, 1998, pp. 123–170.

[4] P. Flandrin, *Time-frequency/Time-scale Analysis*, Academic Press, 1999.

[5] G. B. Folland, *Harmonic Analysis in Phase Space*, Princeton University Press, 1989.

[6] T. L. Grettenberg, *A representation theorem for complex normal processes*, IEEE Trans. Inform. Theory 11 (1965), pp. 305–306.

[7] K. Gröchenig, *Foundations of Time-Frequency Analysis*, Birkhäuser, 2001.

[8] C. W. Helstrom, *Probability and Stochastic Processes for Engineers*, Maxwell Macmillan, 1991.

[9] M. Loève, *Probability Theory*, 2nd edition, Van Nostrand, 1963.

[10] W. Martin, *Time-frequency analysis of random signals*, Proc. ICASSP-82 (1982), pp. 1325–1328.

[11] W. Martin and P. Flandrin, *Wigner–Ville spectral analysis of non-stationary processes*, IEEE Trans. ASSP 33 (1985), pp. 1461–1470.

[12] K. S. Miller, *Complex Stochastic Processes*, Addison-Wesley, 1974.

[13] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I*, Wiley, 1975.
[14] W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, 1987.
[15] A. M. Sayeed and D. L. Jones, *Optimal kernels for nonstationary spectral estimation*, IEEE Trans. Signal Process. 43 (1995), pp. 478–491.
[16] G. F. Simmons, *Topology and Modern Analysis*, McGraw-Hill, 1963.
[17] P. Wahlberg, *The random Wigner distribution of Gaussian stochastic processes with covariance in* $S_0(\mathbb{R}^{2d})$, J. Funct. Spaces Appl. 3 (2) (2005), pp. 163–181.

Dipartimento di Matematica, Università di Torino
Via Carlo Alberto 10, 10123 Turin (TO), Italy
E-mail: patrik.wahlberg@unito.it

Received on 12.8.2009;
revised version on 28.9.2010