16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

A TIME-FREQUENCY FORMULA FOR LMMSE FILTERS FOR NONSTATIONARY UNDERSPREAD CONTINUOUS-TIME STOCHASTIC PROCESSES

Patrik Wahlberg and Peter J. Schreier

School of Electrical Engineering & Computer Science, The University of Newcastle
Callaghan, NSW 2308, Australia
phone: +61 2 49216182, fax: +61 2 49216929, email: Firstname.Lastname@newcastle.edu.au

ABSTRACT

We study linear minimum mean square error (LMMSE) filters for estimating a nonstationary second-order continuous-time stochastic process from a noisy observation. The equation for the optimal filter is treated in the Weyl symbol domain, and the involved Weyl symbols are assumed to belong to certain modulation spaces. By discretizing this equation using a Gabor frame we transform it into a matrix equation and obtain a formula for the filter by matrix inversion. The inverse matrix has off-diagonal decay at a rate that increases the more underspread the process is.

1. INTRODUCTION

The problem of denoising continuous-time second-order complex-valued stochastic processes has been given considerable attention in the literature [6, 16, 18, 19]. The generality of the problem definition makes it interesting to a wide research community. A common assumption is that we observe a zero-mean process y, which is a noisy measurement of a zero-mean process of interest x (the "message"). The second-order statistics are given by the autocovariance functions $r_y(t,s) = E(y(t)y^*(s))$ and $r_x(t,s) = E(x(t)x^*(s))$ and the crosscovariance function $r_{xy}(t,s) = E(x(t)y^*(s))$. Often, the observation is assumed to be of the form

$$y(t) = x(t) + n(t), \tag{1}$$

where the additive noise n is uncorrelated with x.

A classical framework is to assume that the processes x and n are wide-sense stationary (WSS). There are two good reasons to study this restriction: (i) it is an accurate physical model in many engineering applications, and (ii) there exists a well developed mathematical theory for WSS processes. In particular, the Wiener filter, discovered independently by Wiener [19] and Kolmogorov [13], is the celebrated optimal solution for a causal linear time-invariant filter kernel h that minimizes the mean square error (MSE) $E|h * y(t) - x(t)|^2$. Here t is fixed arbitrary since the MSE is time-independent in the WSS case. If the requirement that the filter be causal is relaxed, the optimal filter, sometimes called the *non-causal Wiener filter*, is easier to compute. In the frequency domain it is

$$\mathscr{F}h(\xi) = \frac{\mathscr{F}r_{xy}(\xi)}{\mathscr{F}r_y(\xi)} = \frac{\mathscr{F}r_x(\xi)}{\mathscr{F}r_x(\xi) + \mathscr{F}r_n(\xi)}, \tag{2}$$

where $\mathscr{F}$ denotes Fourier transformation and $\mathscr{F}r_x$ is the nonnegative spectral density of x. The second equality in (2) holds in the signal-plus-noise case (1). It has an intuitively appealing interpretation: the filter attenuates frequencies where the noise is strong compared to the signal of interest x.

Although very useful, the WSS assumption has its limitations and much research efforts have been devoted to study and solve the optimal filtering problem in the case when x and/or y are nonstationary second-order processes [6, 16]. The optimal linear filter kernel is then time-varying and depends on two variables. A research goal has been to generalize the formula (2) to nonstationary processes. Then the frequency domain must be generalized to the time-frequency domain. The spectral density function $\mathscr{F}r_x$ of a WSS process has a natural generalization in the Wigner–Ville spectrum (WVS) [4]

$$\rho_x(t,\xi) = \int_{\mathbb{R}} r_x(t+\tau/2, t-\tau/2) e^{-j\tau\xi} d\tau, \tag{3}$$

and the cross–WVS ρ_{xy} is defined analogously. The corresponding definition for the filter kernel h is called the Weyl symbol,

$$\rho_h(t,\xi) = \int_{\mathbb{R}} h(t+\tau/2, t-\tau/2) e^{-j\tau\xi} d\tau. \tag{4}$$

In the WSS case ρ_x is independent of t and reduces to $\mathscr{F}r_x$ and ρ_h reduces to $\mathscr{F}h$. Hlawatsch et al. [10] have found that the approximation of the Weyl symbol of the optimal filter

$$\rho_h(t,\xi) \approx \frac{\rho_{xy}(t,\xi)}{\rho_y(t,\xi)} = \frac{\rho_x(t,\xi)}{\rho_x(t,\xi) + \rho_n(t,\xi)} \tag{5}$$

is reasonably accurate in the case of *underspread* processes x and y. Again, the second identity in (5) corresponds to the signal-plus-noise case. The concept of *underspreadness* is defined as a condition on the *expected ambiguity function* (EAF)

$$a_x(\nu,\tau) = \frac{1}{4\pi^2} \iint_{\mathbb{R}^2} \rho_x(t,\xi) e^{j(t\nu+\xi\tau)} dt d\xi. \tag{6}$$

When ρ is a symbol of a more general operator, i.e., not necessarily a covariance operator, then a is called *spreading function* [10]. A process x is said to be underspread if a_x is "effectively" compactly supported in a box centered at the origin of area $\ll 1$ [10, 15][1]. This means that the support may be larger than the box, but the fraction of energy outside

This work was supported by the Australian Research Council (ARC) under the Discovery Project DP0664365.

[1] Since we use a slightly different normalization of the Fourier transform and a_x, the value 1 in [10, 15] corresponds to 2π in our setup.

the box, or certain moments of the spreading function, are small.[2]

Since the EAF of an underspread process is concentrated around the origin, its Fourier transform, the WVS ρ_x, is smooth to some degree.

Our goal in this paper is to find a precise version of the formula (2) for nonstationary processes. Our requirements will involve a condition that resembles underspreadness, which will be discussed in Section 6. We will express the formula in terms of the Gabor coefficients for ρ_h, ρ_{xy} and ρ_y, which means that we work in the four-dimensional *phase (time-frequency) space of the time-frequency plane*.

2. THE LMMSE FILTER EQUATION

Let x and y be second-order zero-mean continuous-time, real-valued, or, if complex-valued, jointly proper[3] stochastic processes defined on $\mathbb{R}$. Let the auto-covariance functions be denoted by r_x and r_y, and the cross-covariance function by $r_{xy}(t,s) = E(x(t)y^*(s))$. Suppose $y(t)$ is an observation of a message $x(t)$. To recover $x(t)$ we linearly filter $y(t)$ using a kernel function h to get an estimate of $x(t)$, denoted $\widehat{x}(t)$, [6, 16, 18, 19]

$$\widehat{x}(t) = \int_{\mathbb{R}} h(t,s)y(s)ds. \tag{7}$$

Define the MSE at time t by $E|\widehat{x}(t) - x(t)|^2$. To minimize the MSE for a fixed arbitrary t we need to solve for h the integral equation [6, 16]

$$r_{xy}(t,s) = \int_{\mathbb{R}} h(t,u)r_y(u,s)du, \quad s,t \in \mathbb{R}. \tag{8}$$

This equation may be derived from the principle of orthogonality

$$x(t) - \widehat{x}(t) \perp y(s), \quad s \in \mathbb{R}, \tag{9}$$

in the Hilbert space of second-order stochastic variables.

If we treat h, r_{xy} and r_y as kernels for integral operators, then the equation (8) reads

$$r_{xy} = hr_y, \tag{10}$$

which means that the operator corresponding to the kernel r_{xy} is the composition of the operators corresponding to the kernels h and r_y.

3. THE WEYL SYMBOL, THE WEYL PRODUCT AND MODULATION SPACES

In this section we reformulate the equation (8) in the time-frequency domain.

Let ρ be a function of two variables. The formula

$$(\rho^w f)(x) = (2\pi)^{-1} \iint_{\mathbb{R}^2} \rho\left(\frac{x+y}{2}, \xi\right) e^{j(x-y)\xi} f(y)\, dy d\xi \tag{11}$$

defines a transformation of f, which has integral kernel

$$\begin{aligned} k(x,y) &= (2\pi)^{-1} \int_{\mathbb{R}} \rho\left(\frac{x+y}{2}, \xi\right) e^{j(x-y)\xi} d\xi \\ &= \mathscr{F}_2^{-1}\rho\left(\frac{x+y}{2}, x-y\right), \end{aligned} \tag{12}$$

where $\mathscr{F}_2^{-1}$ denotes inverse Fourier transformation in the second variable. The function ρ is called the Weyl symbol of the operator ρ^w. Since $k(x+y/2, x-y/2) = \mathscr{F}_2^{-1}\rho(x,y)$ according to (12), one can go backwards from operator kernel to symbol via (4). The Weyl calculus treats the correspondences between the symbol ρ and the operator ρ^w and has been developed to a very rich mathematical theory for partial differential equations and time-frequency analysis [5, 7, 12]. There are also other ways to create a map from symbol to operator, most notably the Kohn–Nirenberg correspondence. These kinds of operators, defined by a symbol function, are called pseudodifferential operators.

When we compose two operators formulated as Weyl operators, we may translate the composition to the symbol level. That defines the Weyl product # by

$$\rho_1^w \rho_2^w = (\rho_1 \# \rho_2)^w.$$

On the Weyl symbol level, the equation for the filter (10) thus reads

$$\rho_{xy} = \rho_h \# \rho_y. \tag{13}$$

In order to discuss modulation spaces, we first define the short-time Fourier transform (STFT). Let f be a function defined on the time-frequency plane $\mathbb{R}^2$. A time-frequency variable is denoted by a capital letter like $X = (x_1, x_2) \in \mathbb{R}^2$. We denote the translation operator by $T_X f(Z) = f(Z - X)$, the modulation operator by $M_Y f(Z) = e^{2j\sigma(Y,Z)} f(Z)$, and the time-frequency shift operator, defined by $X, Y \in \mathbb{R}^2$, as

$$(\Pi(X,Y)f)(Z) = e^{2j\sigma(Y,Z)} f(Z - X) = M_Y T_X f(Z). \tag{14}$$

We use a slightly unusual form of the modulation operator, where $\sigma(X,Y) = \sigma((x_1,x_2),(y_1,y_2)) = y_1x_2 - y_2x_1$ denotes the symplectic form [5]. The STFT [5, 7] of f with respect to a window function g is defined by

$$V_g f(X,Y) = \langle f, \Pi(X,Y)g \rangle, \tag{15}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2(\mathbb{R}^2)$. It gives a description of f as a function of the time-frequency variable $(X,Y) \in \mathbb{R}^2 \oplus \mathbb{R}^2$. Note that (X,Y) is the time-frequency variable corresponding to the "time" variable $X \in \mathbb{R}^2$.

The weighted modulation spaces [2, 7] defined on $\mathbb{R}^2$ are defined by $M_m^{p,q}(\mathbb{R}^2) = \{f \in \mathscr{S}'(\mathbb{R}^2);\ V_g f \in L_m^{p,q}(\mathbb{R}^4)\}$, $p,q \in [1,\infty]$, where $L_m^{p,q}(\mathbb{R}^4)$ is the weighted mixed-norm space of all functions $h : \mathbb{R}^4 \mapsto \mathbb{C}$ such that

$$\|h\|_{L_m^{p,q}} = \left(\int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} |h(X,Y)m(X,Y)|^p dX \right)^{q/p} dY \right)^{1/q} < \infty.$$

Here the weight m is assumed to satisfy certain technical conditions [7]. In particular it is v-moderate, i.e. $m(X+Y) \le Cm(X)v(Y)$, $X, Y \in \mathbb{R}^4$, for another weight v. In this setup v

[2]Apart from this qualitative description there also exists an exact definition of underspread, which is mainly used for operators rather than processes. An operator whose spreading function has compact support in a box centered at the origin is *underspread* if the box has area not greater than 2π. It is *overspread* if the area is greater than 2π [14].

[3]This means that $E(x(t)x(s)) = E(y(t)y(s)) = E(x(t)y(s)) = 0$ for all $t, s \in \mathbb{R}$.

is a nondecreasing weight like $v(X) = (1+|X|^2)^{s/2}$, $s \geq 0$, which allows m to be either decreasing or increasing within certain boundaries defined by v. If $m \equiv 1$ then we denote $M^{p,q}(\mathbb{R}^2) = M_m^{p,q}(\mathbb{R}^2)$ and $M_m^p = M_m^{p,p}$.

The modulation spaces increase with the indices as

$$M_m^1(\mathbb{R}^2) \subseteq M_m^{p,q}(\mathbb{R}^2) \subseteq M_m^{r,s}(\mathbb{R}^2) \subseteq M_m^\infty(\mathbb{R}^2), \quad 1 \leq p \leq r, \quad 1 \leq q \leq s, \tag{16}$$

and they shrink when the weight increases,

$$m_1 \leq C m_2, \quad C > 0 \quad \Longrightarrow \quad M_{m_2}^{p,q}(\mathbb{R}^2) \subseteq M_{m_1}^{p,q}(\mathbb{R}^2). \tag{17}$$

The modulation spaces simultaneously quantify the asymptotic decay of a tempered distribution $f \in \mathscr{S}'(\mathbb{R}^2)$ in the time and frequency variables. The functions in the small modulation space $M^1(\mathbb{R}^2)$, also called Feichtinger's algebra [7], are continuous and well concentrated in both time and frequency, whereas the large space $M^\infty(\mathbb{R}^2)$ contains less regular objects such as periodized Dirac distributions.

We shall use in particular the Sjöstrand space $M_w^{\infty,1}(\mathbb{R}^2)$ with a weight $w(X,Y) = w(Y)$ that depends on Y only and is nondecreasing. This space plays a special role in the theory of pseudodifferential operators [7, 9]. The functions in $M_w^{\infty,1}(\mathbb{R}^2)$ are bounded and continuous but generally lack higher-order smoothness.

We will need the following result concerning the Weyl product acting on modulation spaces defined on $\mathbb{R}^2$, which is a special case of results in [11]. Define the family of weight functions $v_s(X,Y) = \langle Y \rangle^s$ where $\langle X \rangle = (1+|X|^2)^{1/2}$ and $s \in \mathbb{R}$. If $s \geq 0$ and $|u| \leq s$ then

$$\|\rho_1 \# \rho_2\|_{M_{v_u}^{p,1}(\mathbb{R}^2)} \leq C \|\rho_1\|_{M_{v_u}^{p,1}(\mathbb{R}^2)} \|\rho_2\|_{M_{v_s}^{\infty,1}(\mathbb{R}^2)}, \quad C > 0, \tag{18}$$

i.e. if $\rho_2 \in M_{v_s}^{\infty,1}(\mathbb{R}^2)$ then Weyl multiplication from the right is a bounded transformation on $M_{v_u}^{p,1}(\mathbb{R}^2)$ for all $p \in [1,\infty]$ and all u such that $|u| \leq s$.

4. GABOR FRAMES FOR MODULATION SPACES

Let g be a window function defined on $\mathbb{R}^2$ and let $\Theta \subset \mathbb{R}^4$ be a lattice, i.e. a set of the form $\Theta = \{(an, bk)\}_{n,k \in \mathbb{Z}^2}$, determined by the positive real numbers a and b. We use the convention to denote elements in such a lattice by boldface Greek letters. Its components are denoted by the corresponding letter without boldface, with and without a prime symbol. For example, $\mathbf{\Lambda} = (\Lambda, \Lambda') \in \Theta$, $\Lambda = an$, $\Lambda' = bk$, $n \in \mathbb{Z}^2$, $k \in \mathbb{Z}^2$. The pair (g, Θ) gives rise to a *Gabor frame* for $L^2(\mathbb{R}^2)$ [1, 7], consisting of the collection of functions $\{\Pi(\mathbf{\Lambda})g\}_{\mathbf{\Lambda} \in \Theta}$, if there exists $0 < A \leq B < \infty$ such that

$$A\|f\|_{L^2}^2 \leq \sum_{\mathbf{\Lambda} \in \Theta} |\langle f, \Pi(\mathbf{\Lambda})g \rangle|^2 \leq B\|f\|_{L^2}^2, \quad f \in L^2(\mathbb{R}^2).$$

A frame does, in general, not consist of orthogonal functions. The Gabor frame operator $S = S(g, \Theta)$, is defined by

$$Sf = \sum_{\mathbf{\Lambda} \in \Theta} \langle f, \Pi(\mathbf{\Lambda})g \rangle \Pi(\mathbf{\Lambda})g.$$

Any $f \in L^2(\mathbb{R}^2)$ has a Gabor expansion

$$f = \sum_{\mathbf{\Lambda} \in \Theta} \langle f, \Pi(\mathbf{\Lambda})\widetilde{g} \rangle \Pi(\mathbf{\Lambda})g, \quad f \in L^2(\mathbb{R}^2) \tag{19}$$

with *non-unique* coefficients. Here $\widetilde{g} = S^{-1}g$ is the so-called *canonical dual window*, which depends on g, a and b. It is also possible to satisfy (19) with other choices of dual windows [7].

Gabor theory has been generalized from the Hilbert space L^2 to modulation spaces by Feichtinger, Gröchenig and Leinert [3, 8]. If $\Theta = \{(an, bk)\}_{n,k \in \mathbb{Z}^2}$, $g \in M_v^1(\mathbb{R}^2)$ and $\{\Pi(\mathbf{\Lambda})g\}_{\mathbf{\Lambda} \in \Theta}$ is a Gabor frame for $L^2(\mathbb{R}^2)$, then also $\widetilde{g} \in M_v^1(\mathbb{R}^2)$. We have the norm equivalence

$$\begin{aligned} & C^{-1}\|f\|_{M_m^{p,q}(\mathbb{R}^2)} \\ & \leq \Big(\sum_{k \in \mathbb{Z}^2} \Big(\sum_{n \in \mathbb{Z}^2} |\langle f, \Pi(an, bk)\widetilde{g} \rangle|^p m(an, bk)^p \Big)^{q/p} \Big)^{1/q} \\ & \leq C\|f\|_{M_m^{p,q}(\mathbb{R}^2)}, \quad C > 0, \end{aligned} \tag{20}$$

and the reconstruction formula (19) holds for the whole scale $1 \leq p, q \leq \infty$ of modulation spaces [7]. We denote the weighted discrete $l^{p,q}(\Theta)$ norm by $\|\cdot\|_{l_m^{p,q}(\Theta)}$, where it is understood that the weight function m is sampled on the lattice Θ, as in (20). Thus, using (15), we may draw the following conclusion from (20). We have $f \in M_m^{p,q}(\mathbb{R}^2)$ if and only if its Gabor coefficients $V_{\widetilde{g}} f \in l_m^{p,q}(\Theta)$ [7]. Thus modulation spaces admits time-frequency discretization without loss of information.

5. RESULTS FOR THE FILTERING PROBLEM

The Weyl product is bilinear and has the continuity property (18) when it acts on modulation spaces. Now we discretize the equation (13) using a Gabor frame. For fixed ρ_y, the Gabor coefficients for ρ_{xy} will then depend linearly on the Gabor coefficients for ρ_h. The linear dependence may be described by a matrix that depends on the Gabor coefficients of ρ_y. Thanks to (18) and (20) the matrix acts continuously on Gabor coefficient sequence spaces. By inversion of this matrix we obtain an expression for the Gabor coefficients of the Weyl symbol ρ_h of the optimal filter.

Let the window function be the Gaussian $\Phi(X) = 2\pi^{1/2}\exp(-|X|^2)$. Then $\{\Pi(an, bk)\Phi\}_{n,k \in \mathbb{Z}^2}$ is a Gabor frame for $L^2(\mathbb{R}^2)$, and thus also for the modulation spaces $M_m^{p,q}(\mathbb{R}^2)$, if the lattice parameters satisfy $0 < ab < \pi$ [7]. Let the symbols ρ_{xy}, ρ_h and ρ_y be members of modulation spaces so that they have the Gabor expansions

$$\begin{aligned} \rho_{xy} &= \sum_{\mathbf{\Lambda} \in \Theta} c_{xy}(\mathbf{\Lambda}) \Pi(\mathbf{\Lambda}) \Phi, \quad c_{xy}(\mathbf{\Lambda}) = \langle \rho_{xy}, \Pi(\mathbf{\Lambda}) \widetilde{\Phi} \rangle, \\ \rho_h &= \sum_{\mathbf{\Lambda} \in \Theta} c_h(\mathbf{\Lambda}) \Pi(\mathbf{\Lambda}) \Phi, \quad c_h(\mathbf{\Lambda}) = \langle \rho_h, \Pi(\mathbf{\Lambda}) \widetilde{\Phi} \rangle, \\ \rho_y &= \sum_{\mathbf{\Lambda} \in \Theta} c_y(\mathbf{\Lambda}) \Pi(\mathbf{\Lambda}) \Phi, \quad c_y(\mathbf{\Lambda}) = \langle \rho_y, \Pi(\mathbf{\Lambda}) \widetilde{\Phi} \rangle. \end{aligned} \tag{21}$$

Define the matrix

$$M(c_y)(\mathbf{\Lambda}, \mathbf{\Omega}) = \sum_{\mathbf{\Gamma} \in \Theta} \mathscr{M}(\mathbf{\Omega}, \mathbf{\Gamma}, \mathbf{\Lambda}) c_y(\mathbf{\Gamma}), \quad \mathbf{\Lambda}, \mathbf{\Omega} \in \Theta, \tag{22}$$

where

$$\mathscr{M}(\mathbf{\Omega},\mathbf{\Gamma},\mathbf{\Lambda})$$
$$= \pi^{1/2}\exp\left(j\left[\sigma(\Omega+\Omega'+\Gamma-\Gamma',\Lambda')+\sigma(\Omega'+\Gamma',\Omega+\Gamma)\right]\right)$$
$$\times\exp\left(-\frac{1}{4}\left|\Omega-\Omega'-\Gamma-\Gamma'\right|^2\right)$$
$$\times V_{\widetilde{\Phi}}\Phi\left(\Lambda-\frac{\Omega+\Omega'+\Gamma-\Gamma'}{2},\Lambda'-\frac{\Omega+\Omega'-\Gamma+\Gamma'}{2}\right) \tag{23}$$

depends on the Gaussian window Φ and its canonical dual window $\widetilde{\Phi}$. The matrix multiplication of $M(c_y)$ and a sequence c defined on Θ is defined by

$$(M(c_y)\cdot c)(\mathbf{\Lambda}) = \sum_{\mathbf{\Omega}\in\Theta} M(c_y)(\mathbf{\Lambda},\mathbf{\Omega})c(\mathbf{\Omega}), \quad \mathbf{\Lambda}\in\Theta.$$

We have the following results for the concepts we have introduced. They are special cases of results proved in [17].

Theorem 1 *Suppose $s\geq 0$, $|u|\leq s$, $1\leq p\leq\infty$, $\rho_{xy}\in M^{p,1}_{v_u}(\mathbb{R}^2)$, $\rho_y\in M^{\infty,1}_{v_s}(\mathbb{R}^2)$, and ρ_y^w is an invertible operator on $L^2(\mathbb{R}^2)$. Then the matrix $M(c_y)$ is invertible on $l^{p,1}_{v_u}(\Theta)$, and the Gabor coefficients of the Weyl symbol for the optimal filter can be computed by the matrix inverse product*

$$c_h = M(c_y)^{-1}\cdot c_{xy}, \tag{24}$$
$$\|c_h\|_{l^{p,1}_{v_u}(\Theta)} \leq C\|c_{xy}\|_{l^{p,1}_{v_u}(\Theta)}, \quad C>0. \tag{25}$$

Theorem 2 *Suppose the assumptions of Theorem 1 are satisfied, and moreover $s>2$. Then the matrices $M(c_y)$ and $M(c_y)^{-1}$ have polynomial off-diagonal decay according to*

$$|M(c_y)(\mathbf{\Lambda},\mathbf{\Omega})| \leq C\langle\mathbf{\Lambda}-\mathbf{\Omega}\rangle^{-t}, \quad \mathbf{\Lambda},\mathbf{\Omega}\in\Theta, \quad C>0,$$
$$|M(c_y)^{-1}(\mathbf{\Lambda},\mathbf{\Omega})| \leq C\langle\mathbf{\Lambda}-\mathbf{\Omega}\rangle^{-t}, \quad \mathbf{\Lambda},\mathbf{\Omega}\in\Theta, \quad C>0,$$

for all $t<s/2-1$.

Theorem 3 *Suppose the assumptions of Theorem 1 are satisfied, except that $\rho_y\in M^{\infty,1}_w(\mathbb{R}^2)$ where the weight $w(X,Y)=e^{2\alpha|Y|}$ where $\alpha=\alpha(a,b)$ is a certain constant that depends on a and b [17]. Then the matrices $M(c_y)$ and $M(c_y)^{-1}$ have exponential off-diagonal decay according to*

$$|M(c_y)(\mathbf{\Lambda},\mathbf{\Omega})| \leq C\exp(-\alpha|\mathbf{\Lambda}-\mathbf{\Omega}|), \quad \mathbf{\Lambda},\mathbf{\Omega}\in\Theta, \quad C>0,$$
$$|M(c_y)^{-1}(\mathbf{\Lambda},\mathbf{\Omega})| \leq C\exp(-\alpha'|\mathbf{\Lambda}-\mathbf{\Omega}|), \quad \mathbf{\Lambda},\mathbf{\Omega}\in\Theta, \quad C>0,$$

where $\alpha'\in(0,\alpha)$.

Theorem 1 gives a time-frequency formula for the optimal filter which may be seen as a nonstationary generalization of (2). More precisely, (24) is a formula for the Gabor coefficients of the optimal filter's Weyl symbol. The inequality (25) says that the Gabor coefficients for the optimal filter's Weyl symbol c_h has at least the same summability (decay at infinity) property as the given sequence c_{xy}.

Since the matrix $M(c_y)$ is determined by ρ_y as specified by (21), (22) and (23), the result (24) is conceptually similar to (5), where the matrix inversion in (24) is replaced by a pointwise inversion by ρ_y. The solution (24) is more complicated than the approximation (5), because: (i) it is formulated in the time-frequency (Gabor coefficient) domain of the Weyl symbols instead of the Weyl symbols directly, and (ii) the inversion is not pointwise but involves the inversion of a matrix.

However, if we make stronger assumptions on the weight w in the requirement $\rho_y\in M^{\infty,1}_w(\mathbb{R}^2)$ as in Theorems 2 and 3, meaning that the modulation space $M^{\infty,1}_w(\mathbb{R}^2)$ shrinks due to (17), we may conclude that the matrix inversion is close to a pointwise product, since the matrices have rapid off-diagonal decay. So under these assumptions the formula (24) is almost a pointwise inversion, which increases its resemblance to (5).

Nevertheless, there remains a qualitative difference between the formulas, namely their domain: the formula (5) concerns quantities defined on the time-frequency plane $\mathbb{R}^2$ whereas (24) concerns sequences defined on $\Theta\subset\mathbb{R}^4$, which is a discretization of the phase space of the time-frequency plane $\mathbb{R}^2$.

6. A PROPERTY SIMILAR TO UNDERSPREADNESS

For the results of Theorems 1 – 3 to work, we need to assume that $\rho_y\in M^{\infty,1}_w(\mathbb{R}^2)$ for certain weights w and that the operator ρ_y^w is $L^2(\mathbb{R}^2)$-invertible. The further assumption $\rho_{xy}\in M^{p,1}_{v_u}(\mathbb{R}^2)$ then implies that $\rho_h\in M^{p,1}_{v_u}(\mathbb{R}^2)$ according to (25) and (20). We would like to understand what the condition $\rho_y\in M^{\infty,1}_w(\mathbb{R}^2)$ means for the EAF a_y.

Set $a'(x_1,x_2)=4\pi a(2x_2,-2x_1)$ where the EAF a is defined by (6) and a symbol $\rho\in M^{\infty,1}_w(\mathbb{R}^2)$. Then $\rho=\mathscr{F}_\sigma a'$ where the *symplectic* Fourier transform [5] is defined by

$$(\mathscr{F}_\sigma f)(X) := \pi^{-1}\int_{\mathbb{R}^2} f(Y)e^{2j\sigma(X,Y)}\,dY.$$

Then, since $\mathscr{F}_\sigma\Pi(X,Y)f = e^{2i\sigma(Y,X)}\Pi(-Y,-X)\mathscr{F}_\sigma f$ and $\mathscr{F}_\sigma\Phi=\Phi$, we have, using the definition of the STFT (15) and Parseval's formula $\langle\mathscr{F}_\sigma f,g\rangle=\langle f,\mathscr{F}_\sigma g\rangle$,

$$\begin{aligned} V_\Phi\rho(X,Y) &= \langle\mathscr{F}_\sigma a',\Pi(X,Y)\Phi\rangle \\ &= e^{2j\sigma(X,Y)}\langle a',\Pi(-Y,-X)\Phi\rangle \\ &= \pi e^{2j\sigma(X,Y)}\mathscr{F}_\sigma(a'T_{-Y}\Phi)(X). \end{aligned}$$

Hence, if the weight depends on the second variable only $w(X,Y)=w(Y)$ and is even,

$$\begin{aligned} \|\rho\|_{M^{\infty,1}_w(\mathbb{R}^2)} &= \int_{\mathbb{R}^2}\sup_X|V_\Phi\rho(X,Y)|w(Y)dY \\ &= \pi\int_{\mathbb{R}^2}\sup_X|\mathscr{F}_\sigma(a'T_Y\Phi)(X)|w(Y)dY \\ &= \pi\int_{\mathbb{R}^2}\|a'T_Y\Phi\|_{\mathscr{F}_\sigma L^\infty}w(Y)dY. \end{aligned}$$

The assumption $\rho\in M^{\infty,1}_w(\mathbb{R}^2)$ is thus equivalent to $a\in W(\mathscr{F}_\sigma L^\infty,L^1_w)$, which denotes a so-called *Wiener amalgam space* [2] whose local norm is $\mathscr{F}_\sigma L^\infty(\mathbb{R}^2)$ and whose global norm is the weighted integral norm $L^1_w(\mathbb{R}^2)$. Wiener amalgam spaces are a family of spaces that control local and global behavior using two different norms. A function to be measured is cut off with a translated window $T_Y\Phi$ and measured with the local norm, and then a global norm is computed with respect to the translation parameter Y. In our case the local norm is $\mathscr{F}_\sigma L^\infty$, which means that a may be very "rough" (not smooth) locally. For instance $\delta_0 \in \mathscr{F}_\sigma L^\infty$ since $\mathscr{F}_\sigma \delta_0 = \pi^{-1} \in L^\infty$. The global norm $L^1_w(\mathbb{R}^2)$ being finite means that a decays rapidly, at a rate which increases with the growth of w. Hence the requirement $\|\rho\|_{M_w^{\infty,1}(\mathbb{R}^2)} < \infty$ says that the EAF a can be very irregular locally, but decays as an L^1_w function globally. This is rather similar, in a qualitative sense, to the description of underspreadness in [10,15] discussed in Section 1.

In conclusion we may say that $\rho_y \in M_w^{\infty,1}(\mathbb{R}^2)$, which is part of the requirements for Theorems 1 – 3, may be interpreted as the condition that y is underspread (in a loose sense). The degree of underspreadness increases when the weight w becomes more rapidly increasing.

7. CONCLUSIONS

In this paper we have obtained a time-frequency formula for the LMMSE filter for nonstationary second-order continuous-time stochastic processes. The formula expresses the Gabor coefficients of the Weyl symbol of the filter in terms of the corresponding quantities for the autocovariance operator of the measured signal and the cross-covariance operator between the desired and the measured signal. This formula may be regarded as a time-frequency generalization of the classical frequency domain formula for the WSS case. The matrix in our formula has rapid off-diagonal decay to a degree which increases with the degree of underspreadness of the involved processes.

REFERENCES

[1] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, 2003.

[2] H. G. Feichtinger, "Modulation spaces on locally compact Abelian groups", *Technical Report*, University of Vienna, 1983; also in: M. Krishna, R. Radha, S. Thangavelu, eds., *Wavelets and Their Applications*, Allied Publishers Private Limited, 99–140, 2003.

[3] H. G. Feichtinger and K. Gröchenig, "Gabor frames and time-frequency analysis of distributions", *J. Funct. Anal.* **146**(2), 464–495, 1997.

[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] V. Fomin, *Optimal Filtering. Volume 1: Filtering of Stochastic Processes*, Kluwer Academic Publishers, 1999.

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

[8] K. Gröchenig and M. Leinert, "Wiener's lemma for twisted convolution and Gabor frames", *J. Amer. Math. Soc.* **17** (1), 1–18, 2004.

[9] K. Gröchenig, "Time-frequency analysis of Sjöstrand's class", *Rev. Mat. Iberoamericana*, **22** (2), 703–724, 2006.

[10] F. Hlawatsch, G. Matz, H. Kirchauer and W. Kozek, "Time-frequency formulation, design, and implementation of time-varying optimal filters for signal estimation," *IEEE Trans. Signal Process.* **48** (5), 1417–32, 2000.

[11] A. Holst, J. Toft and P. Wahlberg, "Weyl product algebras and modulation spaces," *J. Funct. Anal.*, **251**, 463–491, 2007.

[12] L. Hörmander, *The Analysis of Linear Partial Differential Operators III*, Springer-Verlag, 1985.

[13] A. Kolmogorov, "Interpolation and extrapolation", *Bull. Acad. Sci., USSR, Ser. Math.*, **5**, 3–14, 1941.

[14] W. Kozek and G. E. Pfander, "Identification of operators with bandlimited symbols", *SIAM J. Math. Anal.*, **37** (3) 867–888, 2006.

[15] G. Matz and F. Hlawatsch, "Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations", *IEEE Trans. Inf. Theory*, **52** (3), 1067–1086, 2006.

[16] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley, 1967.

[17] P. Wahlberg and P. J. Schreier, "Gabor discretization of the Weyl product for modulation spaces and filtering of nonstationary stochastic processes", *Appl. Comput. Harm. Anal.*, accepted for publication, 2008.

[18] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise*, Prentice-Hall, 1962.

[19] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, MIT Press, 1949.