

Time–Frequency-Based Detection Using Discrete-Time Discrete-Frequency Wigner Distributions

Cédric Richard, *Member, IEEE*

Abstract—During the last decade, a comprehensive theory for optimum time–frequency (TF)-based detection has been developed. This was originally proposed in the continuous-time continuous-frequency case. This paper deals with detectors operating on discrete-time discrete-frequency Wigner distributions (WDs). The purpose is to discuss some existing definitions of this distribution within the context of TF-based detection and selecting those that do not affect the performance of the decision device with which they are associated. This question is of interest since there exist several approaches for discretizing the WD, sometimes resulting in a loss of fundamental properties. First, the discrete-time discrete-frequency formulations of optimum detection are investigated. Next, the problem of the design of TF-based detectors from training data, keeping in mind severe effects of the curse of dimensionality, is considered.

I. INTRODUCTION

TIME-FREQUENCY (TF) representations have been extensively used for nonstationary signal processing since they extend the usual spectral analysis by making it time dependent. Among the myriad of solutions that have been proposed, the Wigner distribution (WD) is considered fundamental in a number of ways. Its usefulness derives from the fact that it satisfies many desired mathematical properties such as the correct marginal conditions and the weak correct-support conditions. This distribution is also a suitable candidate for TF-based detection since it is covariant to time shifts and frequency shifts, and it satisfies Moyal's condition [1], [2].

A. Discrete WDs

For a practical implementation, the WD must be formulated in a discrete-time discrete-frequency setting. The discrete WD is commonly presented in the form

$$W_{xy}^{(C)}[t, f] = \sum_{\tau=0}^{N-1} x[t + \tau] y^*[t - \tau] e^{-j(2\pi/N)f\tau}. \quad (1)$$

Note that it is a halfband TF representation, meaning that frequency components are computed in the normalized range $[-(1/4), 1/4]$. Recently, significant efforts have been made to define fullband representations. An overview of these results

may be found in [3]. In [4], Richman *et al.* have used a group theoretic approach to derive the definition

$$W_{xy}^{(R)}[t, f] = \frac{1}{N} \sum_{\tau=0}^{N-1} \sum_{\nu=0}^{N-1} \sum_{l=0}^{N-1} \rho_N[\nu, \tau] \cdot e^{-j(2\pi/N)(t\nu+f\tau)} e^{j(2\pi/N)\nu l} x[(l + \tau)_N] y^*[l]. \quad (2)$$

The above-mentioned function $\rho_N[\nu, \tau]$, whose expression is given in [4], depends on the parity of N , and $(a)_N = a \bmod N$. This distribution, which, here, is called R-WD, satisfies mathematical properties analogous to those satisfied by the continuous WD. Note that (2) has also been proposed by O'Neill *et al.* for odd-length signals [5]. The authors have used an axiomatic approach, which was extended in [6] to derive a discrete Cohen class. In [7] and [8], Stanković has derived a discrete distribution from the analysis of the WD defined in the frequency domain. This fullband distribution, which is called S-WD, uses the short-time Fourier transform (STFT) as a basic step

$$W_{xy}^{(S)}[t, f] = \frac{1}{N} \sum_{\nu=-B_f}^{B_f} \text{STFT}_x[t, f + \nu] \text{STFT}_y^*[t, f - \nu] \quad (3)$$

where the summation for each $[t, f]$ lasts until any of the conditions $(f + \nu \leq N - 1)$ or $(f - \nu \geq 0)$ is violated [8], i.e., $B_f = \min(f; N - f - 1)$, and

$$\text{STFT}_x[t, f] = \sum_{\tau} x[\tau] h^*[\tau - t] e^{-j(2\pi/N)f\tau}. \quad (4)$$

Note that Definition (4) differs from that in [8]. The first one is more often used in the literature, whereas the second one fully corresponds to the STFT numerical calculation by using the FFT routines. However, both lead to equivalent results. Since they do not require either analytic signal calculation or oversampling, the R-WD and the S-WD provide substantial advantages over the traditional way (1) of WD computing.

B. TF-Based Detection

By virtue of their rich structure, TF representations have been extensively used for detection in applications ranging from radar to machine fault diagnostics due to the need to deal with nonstationary signals. During the last decade, a comprehensive theory for optimum TF-based detection has been developed. In [2], Flandrin has characterized detection scenarios for which TF-based detectors are optimum. Marinovich has followed this approach in [9] for the case of Gaussian transient detection in the mix of colored and white Gaussian noise. In [1], Sayeed and

Manuscript received June 3, 2001; revised June 3, 2002. The associate editor coordinating the review of this paper and approving it for publication was Dr. Xiang-Gen Xia.

The author is with the Laboratoire de Modélisation et Sûreté des Systèmes (LM2S), Université de Technologie de Troyes (UTT), Troyes, France (e-mail: cedric.richard@utt.fr).

Publisher Item Identifier 10.1109/TSP.2002.801927.

Jones have proposed TF-based receivers that are optimum for detecting second-order nonstationary stochastic signals in the presence of Gaussian noise, with time and frequency shifts as unknown or random nuisance parameters. Except in [9], where implementation considerations are addressed through the use of (1), the TF detection problems mentioned previously were originally solved in the continuous-time continuous-frequency case. The resulting receivers then need to be discretized before they can be implemented as a computer program.

This paper deals with detectors operating on the discrete WD. The purpose is to evaluate the relevance of some definitions of this distribution within the context of TF-based detection, selecting those that do not affect the performance of the decision device with which they are associated. This question is of interest since there exist several approaches for discretizing the WD, sometimes resulting in a loss of fundamental properties. This paper is organized as follows. First, the discrete-time discrete-frequency formulations of optimum detection are investigated. Next, the problem of blind TF-based detectors design from training data is considered [10] and predictions made of the effects of the curse of dimensionality as a function of the discrete WD definition that has been selected. Finally, some concluding remarks are presented.

II. DISCRETE TF FORMULATIONS OF OPTIMUM DETECTION

In this section, it is shown that the definition of the discrete WD has an influence on the performance of the decision device with which it is associated. The detection problem considered has been solved in the continuous-time continuous-frequency domain in [2]. Here, it is considered in the discrete case as

$$\left. \begin{aligned} H_0: x[t] &= w[t] \\ H_1: x[t] &= w[t] + s[t] \end{aligned} \right\}, \quad t \in \{0, \dots, N-1\} \quad (5)$$

where s is the Gaussian signal to be detected, characterized by the mean m and the covariance R_s . The additive noise w is zero-mean, Gaussian, and white with variance σ_0^2 .

A. Bayes-Optimum Detector

Let \mathcal{B} be an orthonormal basis of eigenvectors of R_s . Let ϕ_k be the k th vector of \mathcal{B} , and let λ_k be the corresponding eigenvalue. The solution of the classic problem (5) is

$$l_R + l_D \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \quad (6)$$

where γ is a given threshold, and

$$l_R = \frac{1}{\sigma_0^2} \sum_{k=0}^{N-1} \frac{\lambda_k}{\lambda_k + \sigma_0^2} |\tilde{x}[k]|^2 \quad (7)$$

$$l_D = 2 \sum_{k=0}^{N-1} \frac{1}{\lambda_k + \sigma_0^2} \operatorname{Re} \{ \tilde{x}[k] \tilde{m}^*[k] \} \quad (8)$$

with

$$\tilde{x}[k] = \sum_{t=0}^{N-1} x[t] \varphi_k^*[t], \quad \tilde{m}[k] = \sum_{t=0}^{N-1} m[t] \varphi_k^*[t]. \quad (9)$$

This result involves inner products in the discrete-time domain. It is very similar to the continuous-time case, except that (9) is

replaced by the expansion of x and m in a Karhunen-Loève basis of R_s . Let $W_{xy}^{(\cdot)}$ implicitly refer to $W_{xy}^{(C)}$, $W_{xy}^{(R)}$ and/or $W_{xy}^{(S)}$. As in [2], the expressions involving discrete WDs

$$l_R = \frac{1}{N\sigma_0^2} \sum_{t,f=0}^{N-1} W_x^{(\cdot)}[t, f] \left[\sum_{k=0}^{N-1} \frac{\lambda_k}{\lambda_k + \sigma_0^2} W_{\varphi_k}^{(\cdot)}[t, f] \right] \quad (10)$$

$$l_D = \frac{2}{N} \sum_{t,f=0}^{N-1} \operatorname{Re} \{ W_{xm}^{(\cdot)}[t, f] \} \left[\sum_{k=0}^{N-1} \frac{1}{\lambda_k + \sigma_0^2} W_{\varphi_k}^{(\cdot)}[t, f] \right] \quad (11)$$

are equivalent to (7) and (8) if $W_{xy}^{(\cdot)}$ satisfies the inner product conservation law

$$\begin{aligned} & \sum_{t,f=0}^{N-1} W_{x_1 y_1}^{(\cdot)}[t, f] \left(W_{x_2 y_2}^{(\cdot)}[t, f] \right)^* \\ &= N \left[\sum_{t=0}^{N-1} x_1[t] x_2^*[t] \right] \left[\sum_{t=0}^{N-1} y_1[t] y_2^*[t] \right]^*. \end{aligned} \quad (12)$$

This relationship holds only for the R-WD defined by (2). Then, the classic definition (1) and the S-WD given by (3) must be discarded for exact formulations (10) and (11), and the R-WD can be retained as a basic TF tool for optimum detection of second-order signals in the presence of Gaussian noise. Because the R-WD satisfies the property of covariance, it can also be used for composite hypothesis testing when the signal to be detected has time and frequency shifts as a couple of nuisance parameters [1].

Remark: As shown later, one can implement the optimum detector (6) using the classic definition (1) since it satisfies the correct time marginal density

$$\begin{aligned} l_R &= \frac{1}{N\sigma_0^2} \sum_{t,f=0}^{N-1} \frac{\lambda_t}{\lambda_t + \sigma_0^2} W_x^{(C)}[t, f] \\ l_D &= \frac{2}{N} \sum_{t,f=0}^{N-1} \frac{1}{\lambda_t + \sigma_0^2} \operatorname{Re} \{ W_{\tilde{x}m}^{(C)}[t, f] \}. \end{aligned}$$

Even if this solution is optimum, it must be discarded since it does not provide any meaningful interpretation.

B. Example

For detecting Gaussian signals in white Gaussian noise, the TF structure of (10) and (11) is optimum in the Bayes sense if $W_{xy}^{(\cdot)}$ satisfies the inner product conservation law (12). Not surprisingly, using this receiver with (1) or (3) can result in serious performance degradation, as shown in Fig. 1.

III. DATA-DRIVEN DESIGN OF DISCRETE TF DETECTORS

Optimal detection of signals embedded in noise requires the knowledge of certain underlying statistics. In most practical applications of current interest, these statistics are not available since phenomena are complex and poorly understood. However, one can often collect labeled signals resulting from experimental observation as for electrophysiological data [11] or mechanical systems data [12]–[14]. If the database is somehow representative of the unknown distribution, the following strategy based on deriving the detector directly from training data offers hope for optimal detection: 1) selecting a

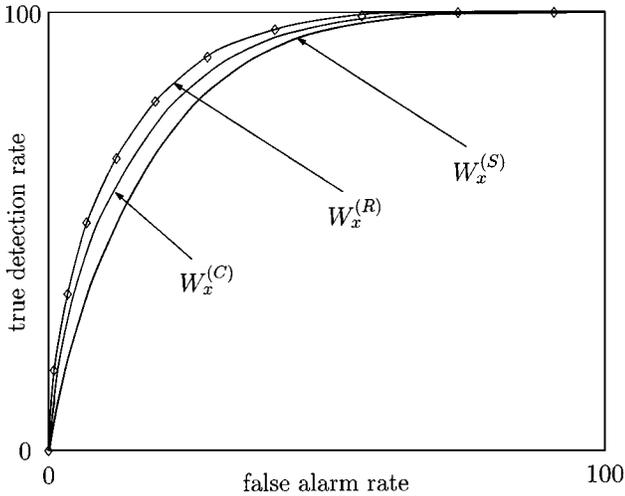


Fig. 1. Comparison of operating characteristics. For detecting Gaussian signals embedded in white Gaussian noise, the optimum receiver in the Bayes sense (\diamond) can be rewritten in time and frequency terms using the R-WD. However, making use of the classic definition or the S-WD can result in performance degradation.

class \mathcal{C} of detectors 2) and tempting to pick the detector of \mathcal{C} that minimizes, for example, an estimate of the error probability. This viewpoint will now be adapted within the context of TF-based detection. For practical reasons, the discussion will be restricted to linear detectors operating on discrete WD, which are of the form

$$\sum_{t,f=0}^{N-1} A[t, f] W_x^{(\cdot)}[t, f] \underset{H_0}{\overset{H_1}{>}} \gamma \quad (13)$$

where the $A[t, f]$ s and γ are parameters to be determined from available training data. Note that (13) has been widely used in the literature since it provides a flexible and meaningful quadratic decision function. Let $\mathcal{C}^{(\cdot)}$ be the class of linear detectors operating on $W_x^{(\cdot)}$. To further compare the $\mathcal{C}^{(\cdot)}$ s, one first has to study the redundancy of information carried by every representation $W_x^{(\cdot)}$. This result can then be used to evaluate the relevance of (1)–(3) and to predict their behavior when confronted with the curse of dimensionality.

A. Linear Signal Spaces and Induced Spaces

As a basis for subsequent development, this section provides a brief review of linear spaces, and the concept of induced spaces is introduced. A linear signal space \mathcal{S} over the field \mathbb{C} is a collection of signals that satisfies the following linearity property: If $x \in \mathcal{S}$ and $y \in \mathcal{S}$, then $(\alpha x + \beta y) \in \mathcal{S}$ for arbitrary complex coefficients α, β . Let $\{s_q\}$ be a nonempty set of elements of \mathcal{S} . One can say that it is a basis for \mathcal{S} if it is linearly independent and if it generates \mathcal{S} . The basis $\{s_q\}$ is orthonormal if $\langle s_q | s_{q'} \rangle = \sum_{t=0}^{N-1} s_q[t] s_{q'}^*[t] = \delta_{qq'}$, where $\delta_{qq'}$ is the Kronecker delta symbol. Then, every $x \in \mathcal{S}$ can be represented as $x = \sum_q \alpha_q s_q$, where $\alpha_q = \langle x | s_q \rangle$ is the inner product of x and s_q . In this paper, it is considered that \mathcal{S} is the linear space of N -sample complex-valued signals over \mathbb{C} .

Let $\mathbf{w}^{(\cdot)}$ be the map such that $W_{xy}^{(\cdot)} \triangleq \mathbf{w}^{(\cdot)}(x, y)$ for all x, y in \mathbb{C}^N . Let $\mathcal{W}^{(\cdot)}$ be the image of $\mathbf{w}^{(\cdot)}$ that is given by

$$\mathcal{W}^{(\cdot)} \triangleq \left\{ W_{xy}^{(\cdot)} : W_{xy}^{(\cdot)} = \mathbf{w}^{(\cdot)}(x, y) \text{ with } x, y \in \mathbb{C}^N \right\}. \quad (14)$$

Note that the space $\mathcal{W}^{(\cdot)}$ is not a linear space: Linear combinations of WD are not valid WD. The space $\mathcal{W}^{(\cdot)}$ can now be associated with the linear space $\tilde{\mathcal{W}}^{(\cdot)}$ over the field \mathbb{C} of all linear combinations of $W_{xy}^{(\cdot)}$ with $x, y \in \mathbb{C}^N$. In [15], $\tilde{\mathcal{W}}^{(\cdot)}$ is called “induced $W_{xy}^{(\cdot)}$ -domain space.” Here, the inner product is expressed as

$$\left\langle W_{x_1 y_1}^{(\cdot)} | W_{x_2 y_2}^{(\cdot)} \right\rangle = \sum_{t,f=0}^{N-1} W_{x_1 y_1}^{(\cdot)}[t, f] \left(W_{x_2 y_2}^{(\cdot)}[t, f] \right)^*$$

and $\dim(\tilde{\mathcal{W}}^{(\cdot)})$ can be defined as the cardinality of $\{\tilde{W}_q^{(\cdot)}\}$, where $\{\tilde{W}_q^{(\cdot)}\}$ is a basis for $\tilde{\mathcal{W}}^{(\cdot)}$.

B. $W_{xy}^{(C)}$ Versus $W_{xy}^{(R)}$

Following this brief presentation of induced spaces, the paper will now compare the classes of linear detectors operating on $W_x^{(C)}$ and $W_x^{(R)}$, which are denoted by $\mathcal{C}^{(C)}$ and $\mathcal{C}^{(R)}$. To further establish that $\mathcal{C}^{(C)} \subset \mathcal{C}^{(R)}$, one first has to determine $\dim(\tilde{\mathcal{W}}^{(R)})$ and $\dim(\tilde{\mathcal{W}}^{(C)})$.

Let Δ_{t_0} be the pulse defined by $\Delta_{t_0}[t] = 1$ if $t = t_0$ and 0 otherwise. By construction, any R-WD can be represented as

$$W_{xy}^{(R)} = \sum_{t,\tau=0}^{N-1} x[(t+\tau)_N] y^*[t] W_{\Delta_{(t+\tau)_N} \Delta_t}^{(R)}. \quad (15)$$

This implies that $\{W_{\Delta_{(t+\tau)_N} \Delta_t}^{(R)}\}_{0 \leq t, \tau \leq N-1}$ generates the space $\tilde{\mathcal{W}}^{(R)}$. Furthermore, a simple calculus shows that our family is an orthonormal basis for $\tilde{\mathcal{W}}^{(R)}$ since

$$\left\langle W_{\Delta_{(t+\tau)_N} \Delta_t}^{(R)} | W_{\Delta_{(t'+\tau')_N} \Delta_{t'}}^{(R)} \right\rangle = \delta_{tt'} \delta_{\tau\tau'} \quad (16)$$

where δ is the Kronecker delta symbol. It directly follows that

$$\dim(\tilde{\mathcal{W}}^{(R)}) = \text{card}\left\{ W_{\Delta_{(t+\tau)_N} \Delta_t}^{(R)} \right\}_{0 \leq t, \tau \leq N-1} = N^2 \quad (17)$$

from which one can conclude that the N^2 samples of $W_x^{(C)}$ are linearly independent. Note that this result is applicable to the auto distribution $W_x^{(C)}$ since taking $x = y$ does not modify the above proof. This implies that $\mathcal{C}^{(R)} \equiv \mathcal{Q}$, where \mathcal{Q} is the class of quadratic detectors, which is defined as

$$\sum_{t,\tau=0}^{N-1} x[t] Q[t, \tau] x^*[\tau] \underset{H_0}{\overset{H_1}{>}} \gamma. \quad (18)$$

Here, Q denotes a $N \times N$ Hermitian matrix to ensure realness of the detection statistic.

Attention is now focused on $W_{xy}^{(C)}$. A recently proposed algebraic study has shown that this distribution encodes information in a redundant fashion since linear relations connect the $W_{xy}^{(C)}[t, f]$ s [16]. One key point of this work was that the cross WD of any complex-valued signals x and y can be expanded in the form

$$W_{xy}^{(C)} = \sum_{(t,\tau) \in \mathcal{J}} x[t+\tau] y^*[t-\tau] W_{\Delta_{(t+\tau)} \Delta_{(t-\tau)}}^{(C)} \quad (19)$$

where \mathcal{J} is the collection of pairs (t, τ) such that both $t+\tau$ and $t-\tau$ are elements of $\{0, \dots, N-1\}$. Simple calculus shows that

$$\left\langle W_{\Delta(t+\tau)\Delta(t-\tau)}^{(C)} \middle| W_{\Delta(t'+\tau')\Delta(t'-\tau')}^{(C)} \right\rangle = \delta_{tt'}\delta_{\tau\tau'} \quad (20)$$

which means that $\{W_{\Delta(t+\tau)\Delta(t-\tau)}^{(C)} : (t, \tau) \in \mathcal{J}\}$ is an orthonormal basis for $\tilde{\mathcal{W}}^{(C)}$. It follows that $\dim(\tilde{\mathcal{W}}^{(C)})$ is equal to the cardinality of \mathcal{J} . Combining $0 \leq t+\tau \leq N-1$ and $0 \leq t-\tau \leq N-1$ gives $-t \leq \tau \leq t$ if $0 \leq t \leq \lfloor (N-1)/2 \rfloor$, $-(N-t-1) \leq \tau \leq (N-t-1)$, otherwise, where $\lfloor (N-1)/2 \rfloor$ denotes integer part of $(N-1)/2$. Then, one has

$$\begin{aligned} \dim(\tilde{\mathcal{W}}^{(C)}) &= \sum_{t=0}^{\lfloor (N-1)/2 \rfloor} [2t+1] \\ &+ \sum_{t=\lfloor (N-1)/2 \rfloor+1}^{N-1} [2(N-t-1)+1] \quad (21) \end{aligned}$$

from which it is finally obtained that $\dim(\tilde{\mathcal{W}}^{(C)})$ equals $\lfloor (N^2+1)/2 \rfloor$. Because $\dim(\tilde{\mathcal{W}}^{(C)}) < N^2$, linear relations then connect the N^2 samples $W_{xy}^{(C)}[t, f]$ of any representation $W_{xy}^{(C)} \in \tilde{\mathcal{W}}^{(C)}$. As a conclusion, the R-WD potentially leads to a larger class of detection structures than the classic definition (1) since the $W_{xy}^{(R)}[t, f]$ s are linearly independent, whereas the $W_{xy}^{(C)}[t, f]$ s are not. In particular, one has $\mathcal{C}^{(C)} \subset \mathcal{C}^{(R)} \triangleq \mathcal{Q}$, which suggests that $\mathcal{C}^{(R)}$ can provide detectors that (at least) perform better than those of the class $\mathcal{C}^{(C)}$.

C. Influence of the Curse of Dimensionality

As introduced previously, there are several strategies for solving a detection problem, depending on *a priori* knowledge on competing hypotheses. In this section, it is supposed that our knowledge is restricted to a training set, which is used to adjust the free parameters of a given detection structure. To achieve good performance, it is well known in pattern recognition that one needs to match the complexity of the detector to the amount of training data [17]–[19]. As illustrated in Fig. 2, if the detector is too complex, it is likely to learn the training data, but it will probably not generalize properly. In contrast, if it is not complex enough, it might not be able to extract all the discriminant information available in the training set. This experimental evidence, which is known as the curse of dimensionality, has been studied theoretically by Vapnik and Chervonenkis [17]. In particular, these authors have formally defined the complexity of a detector, which is called dimension of Vapnik–Chervonenkis (or VC-dimension). This parameter, hereafter denoted V_c , can be used to compute a confidence interval for the error probability of any detector d designed from training data. The following inequality holds with a probability of $(1-\epsilon)$:

$$|P_e(d) - P_{emp}(d, \mathcal{A}_M)| \leq E(M, V_c, \epsilon) \quad (22)$$

where

$$E(M, V_c, \epsilon) = \sqrt{\frac{V_c}{M} \left(1 + \log \frac{2M}{V_c}\right)} - \frac{1}{M} \log \frac{\epsilon}{4} \quad (23)$$

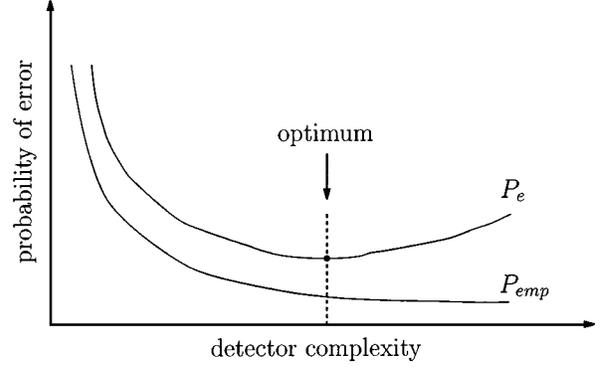


Fig. 2. Illustration of the behavior of generalization error P_e and empirical error P_{emp} during a typical training stage as a function of the detector complexity. Note that P_{emp} , which is an estimate of P_e based on training data, is also called training error.

where

\mathcal{A}_M	M -sample training set,
$P_e(d)$	probability of error of d ;
$P_{emp}(d, \mathcal{A}_M)$	estimate of this probability based on \mathcal{A}_M .

The cardinality M of the training set is generally fixed so that one needs to carefully control V_c in order to reach a low $P_e(d)$.

For linear detectors, it can be shown that V_c is equal to $L+1$, where L is the dimension of the space spanned by learning samples [19], or by their TF representations within the context of TF-based detection. With (17) and (21), this directly implies that $V_c = N^2 + 1$ for linear detectors operating on $W_x^{(R)}$, and $V_c = \lfloor (N^2+1)/2 \rfloor + 1$ if one considers the case of linear detectors associated with $W_x^{(C)}$. In the following, these VC-dimensions are denoted by $V_c^{(R)}$ and $V_c^{(C)}$, respectively. Since $V_c^{(R)} > V_c^{(C)}$, detectors of $\mathcal{C}^{(C)}$ are less sensitive to the curse of dimensionality than those of $\mathcal{C}^{(R)}$. Thus, they can sometimes lead to improved performance for a given data set, even though information is being discarded. Let us now concentrate on computer simulations to illustrate this phenomenon. The problem chosen to be addressed deals with detecting the presence or absence of the 16-length signal $s[t]\exp(j\phi_0)$ embedded in additive noise $w[t]$ (SNR = -6 dB). Here, $s[t]$ is a deterministic signal, and ϕ_0 denotes a random phase uniformly distributed over $[-\pi, \pi]$. The noise $w[t]$ is white non-Gaussian with probability density function given by $f_w \triangleq (1-\epsilon)\mathcal{N}(0, \sigma^2) + \epsilon\mathcal{N}(0, K^2\sigma^2)$ with $\epsilon = 0.5$ and $K = 5$, where $\mathcal{N}(0, \sigma^2)$ is the normal probability density function with mean 0 and variance σ^2 . First, detectors of the form (13) operating on $W_x^{(R)}$ and $W_x^{(C)}$ were designed as in [11] using 6000 realizations of the signal plus noise and noise only. The comparison of their performance, which is depicted in Fig. 3, conforms with the theoretical result presented in Section III-B: $\mathcal{C}^{(R)}$ provides detectors that (at least) perform better than those of $\mathcal{C}^{(C)}$ if the effect of the curse of dimensionality is negligible, i.e., enough training samples are available. Second, the same experiment was conducted with only 100 realizations each of signal present and signal absent. Fig. 4 shows that the detector associated with $W_x^{(C)}$ yields higher performance than the detector operating on $W_x^{(R)}$, even though $\mathcal{C}^{(C)} \subset \mathcal{C}^{(R)}$. As explained previously, the influence of the curse of dimensionality due to the small number of available training samples justifies this surprising result.

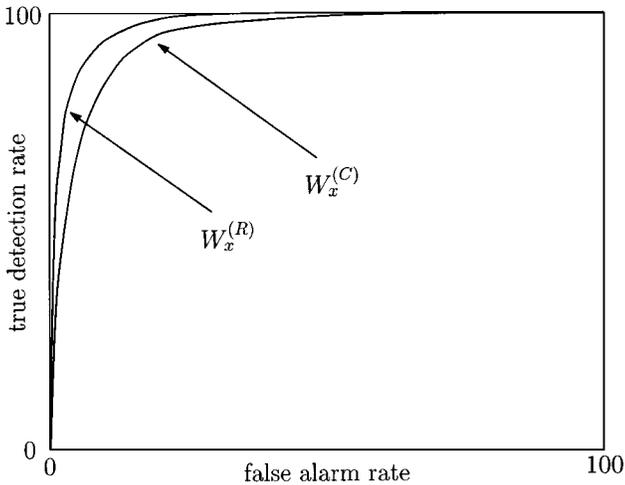


Fig. 3. Comparison of operating characteristics, which shows that the detector of $C^{(R)}$ performs better than the detector provided by $C^{(C)}$. Here, the influence of the curse of dimensionality is negligible since the training process was conducted with 6000 16-length signals.

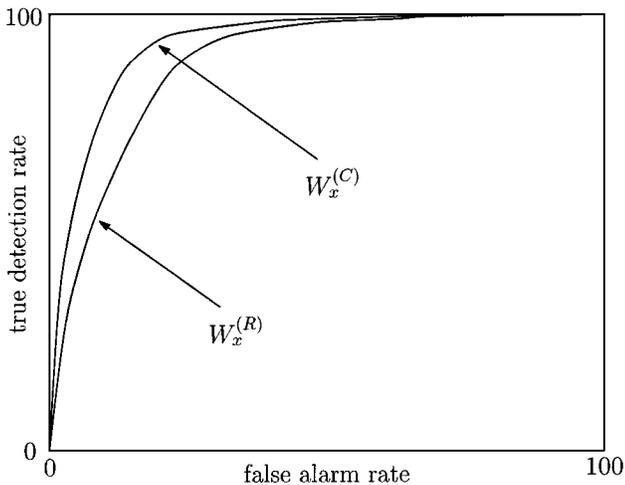


Fig. 4. Illustration of the curse of the dimensionality and its effects. The detector of $C^{(R)}$ has lower performance than the detector provided by $C^{(C)}$, even though $C^{(C)} \subset C^{(R)}$. Both detection structures were designed from 200 16-length signals.

D. Versatile Alternative: $W_{xy}^{(S)}$

For a fixed size of training set, it has been shown that it is important to control the level of complexity of a detection structure in order to optimize its generalization performance. There are, in fact, many different ways in which to achieve an appropriate complexity. For example, regularization techniques involve adding to the error function an extra penalty term, which is designed to influence the form of the solution [20]. Another way of dealing with the curse of dimensionality is to preprocess the data to change the dimension of the input space. In the context of TF-based detection, this involves how smoothing TF representations can lead to significant improvement in generalization performance of receivers, as has been shown in [21] with spectrograms of variable windows lengths. Recently, the signal-dependent kernel proposed in [22] and the adaptive diffusion process presented in [23] have also been used for the purpose of detector complexity control [24], [25]. The goal of

this subsection is to show that the S-WD provides a substantial advantage over the R-WD and the traditional definition (1) for designing detectors with improved generalization performance. For practical reasons, the discussion, as before, will be limited to linear detectors of the form (13).

Thus far, the case of detectors operating on $W_x^{(C)}$ and $W_{xy}^{(R)}$ has been considered, and it has been shown that $V_c^{(C)} = \lfloor (N^2 + 1)/2 \rfloor + 1$ and $V_c^{(R)} = N^2 + 1$, where N is the length of signals on stage. In practice, it is then necessary to apply postprocessing transformations to the input TF representations in order to match the complexity of such detection structures to the amount of training data. In the case of detectors operating on the S-WD, it is clear that one can take advantage of the window $h[t]$ in (4) to control the dimension of the induced $W_{xy}^{(S)}$ -domain space, which is denoted $\tilde{\mathcal{W}}^{(S)}$. The dimension of $\tilde{\mathcal{W}}^{(S)}$ will not be evaluated further as it is a tedious calculus exercise involving the transcription of $W_{xy}^{(S)}$ in the discrete ambiguity domain by means of the two-dimensional Fourier transform. However, it is noted that $\dim(\tilde{\mathcal{W}}^{(S)})$ satisfies the following inequalities:

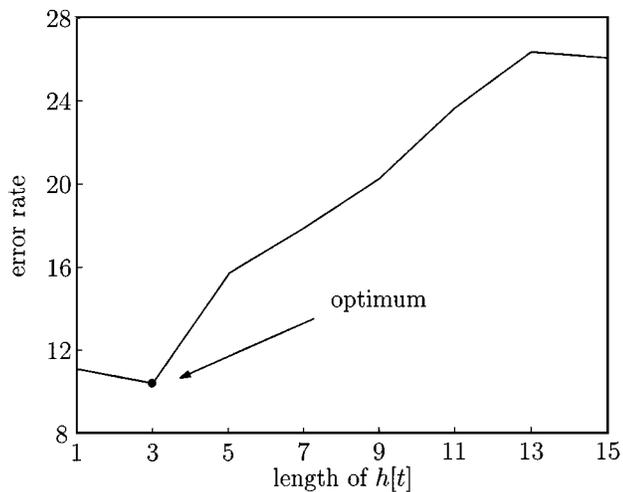
- 1) $\dim(\tilde{\mathcal{W}}^{(S)}) = N$ if $L = 1$;
- 2) $\dim(\tilde{\mathcal{W}}^{(S)}) \leq N^2$;

where L is the length of the smoothing window $h[t]$. Making use of $W_{xy}^{(S)}[t, f] \propto x[t] y^*[t]$ when $L = 1$ leads to 1). To show 2), it suffices to note that the linear space $\tilde{\mathcal{W}}^{(S)}$ contains N^2 -sample elements. This finally implies that $C^{(S)} \subseteq C^{(R)} \triangleq \mathcal{Q}$, where $C^{(S)}$ is the class of linear detectors operating on the S-WD.

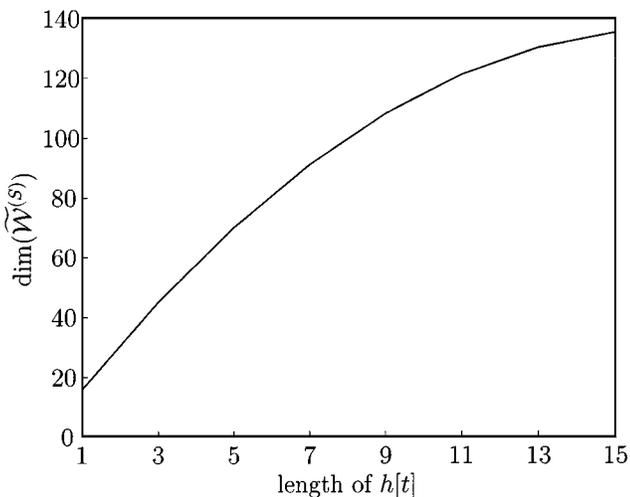
In [7] and [8], Stanković has shown that crossterms in the S-WD can be avoided while improving concentration of autoterms by setting the parameter B_f in (3) to a value lower than $\min(f; N - f - 1)$. Given any fixed smoothing window $h[t]$, the influence of B_f on detection performance will now be investigated. One has $\text{STFT}_x[t, f + \nu] = \text{STFT}_{x+\nu}[t, f]$ with $x_{+\nu}[t] = x[t] \exp(-j(2\pi/N)t\nu)$. By construction, every sample $W_{xy}^{(S)}[t, f]$ with $(t, f) \in \{0, \dots, N-1\}^2$ can then be expressed as

$$W_{xy}^{(S)}[t, f] = \sum_{\nu} \xi[\nu, f] S_{x+\nu y-\nu}[t, f] \quad (24)$$

where $S_{x+\nu y-\nu}[t, f]$ denotes the spectrogram sample, which is defined as $(1/N) \text{STFT}_{x+\nu}[t, f] \text{STFT}_{y-\nu}^*[t, f]$, and $\xi[\nu, f] = 1$ if $|\nu| \leq B_f$, and 0 otherwise. Note that $\nu = 0$ produces the spectrogram. In the case of the traditional way (1) of WD computing, it has been shown in [16] that sets of TF generators $\{W_{xy}^{(C)}[t, f]: (t, f) \in \mathcal{L}^{(C)}\}$, which are denoted $\mathcal{G}_{\mathcal{L}^{(C)}}^{(C)}$, exist so that every $W_{xy}^{(C)}[t, f]$ with $(t, f) \in \{0, \dots, N-1\}^2$ can be expressed as a linear combination of the elements of $\mathcal{G}_{\mathcal{L}^{(C)}}^{(C)}$ for all $x, y \in \mathbb{C}^N$. Each set is completely determined by a family of TF locations $\mathcal{L}^{(C)}$, depending only on the definition of the TF distribution under consideration and the linear signal space fixed beforehand, which is \mathbb{C}^N here. The cardinality of $\mathcal{L}^{(C)}$ equals the dimension of the induced $W_{xy}^{(C)}$ -domain space. Obviously, this framework can be used to study any other discrete TF distribution such as the S-WD and the spectrogram. With (24), it follows that any set of generators for the $S_{xy}[t, f]$ s is also a set of generators for the $W_{xy}^{(S)}[t, f]$ s for all B_f in $\{0, \dots, \min(f; N - f - 1)\}$ and for all x, y in \mathbb{C}^N . This implies



(a)



(b)

Fig. 5. Complexity control of the linear detector operating on the S-WD. These plots show the evolution of (a) an estimate of the generalization error and (b) the dimension of the induced S-WD domain space as a function of the length L of the smoothing window $h[t]$. The training process was conducted with the same learning set as in Fig. 4.

that given any B_f , the two classes of linear detectors operating on the S-WD and the spectrogram coincide, from which one can conclude that B_f is not an influencing factor in detection performance. This parameter can thus be used to improve the readability of TF representations during the detection process.

As an illustration of the use of $W_{xy}^{(S)}$ to control the complexity of the detector (13), the example presented in Section III-C is used. Fig. 5(a) and (b) shows the behavior of an estimate of the generalization error and the dimension of $\tilde{\mathcal{W}}^{(S)}$, as a function of the length L of the rectangular window $h[t]$. It is noted that the optimization of the error rate with respect to the complexity of the detector results in a minimum for $L = 3$. As indicated in Fig. 6, this detection structure yields higher performance than detectors operating on $W_x^{(C)}$ and $W_x^{(R)}$. This confirms that making use of $W_x^{(S)}$ allows the control of the VC-dimension of the receiver (13) in order to reach a low error rate.

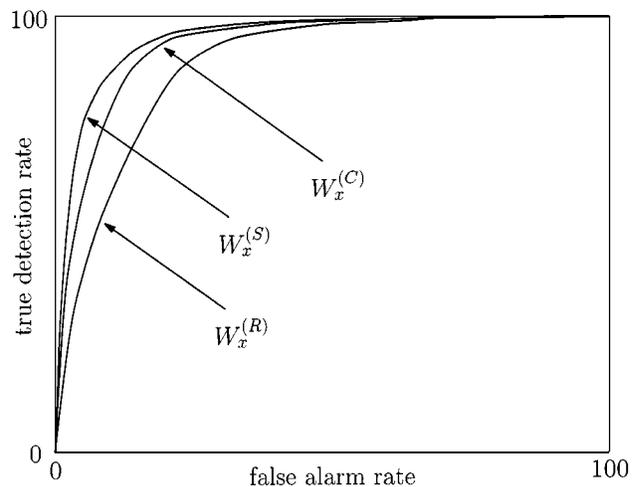


Fig. 6. Comparison of operating characteristics, which shows that the detector of $\mathcal{C}^{(S)}$ with optimized complexity ($L = 3$) performs better than detectors provided by $\mathcal{C}^{(C)}$ and $\mathcal{C}^{(R)}$. The training process was conducted with the same learning set as in Fig. 4.

IV. CONCLUSION

In this paper, the problems that can arise in attempts to design detectors operating in the discrete-time discrete-frequency domain have been discussed, as well as the potential improvement in detection performance, which can be achieved by using the R-WD, the S-WD, or the traditional way of WD computing. The discussion was restricted to the theory of optimum TF-based detection developed in the continuous-time continuous-frequency case during the last decade. Discrete formulations have been investigated, and it has been shown that the R-WD can be retained as a basic tool for optimum detection of second-order signals in the presence of Gaussian noise. However, designing such detection structures requires the knowledge of conditional probability density functions. Because this statistical information is not available in most applications of current interest, a technique for deriving linear detectors directly from labeled training data was considered. Calculating the dimension of the linear space induced by the R-WD, it was shown that this discrete WD leads to the largest class of linear detection structures operating in the TF domain. Nevertheless, these detectors have a severe disadvantage in that they tend to have significantly lower performance than detectors operating on the traditional discrete WD, particularly when the number of training samples is relatively small. This phenomenon, which is termed the curse of dimensionality, has been extensively addressed in the literature. It arises in attempts to perform pattern recognition in high-dimensional space, and its worst effects can be alleviated by a preprocessing stage, leading to a reduction in the dimensionality of input data. An advantage of the S-WD over the other discrete WD is that it involves a smoothing procedure. As has been shown, this can be used as a technique for dimensionality control, leading to a decision device with better generalization properties. Thus far, attention has been restricted to three particular forms of the discrete WD. Obviously, the framework described throughout this paper can be used to study other discrete distributions from a detection viewpoint.

REFERENCES

- [1] A. M. Sayeed and D. L. Jones, "Optimal detection using bilinear time-frequency and time-scale representations," *IEEE Trans. Signal Processing*, vol. 43, pp. 2872–2883, Dec. 1995.
- [2] P. Flandrin, "A time-frequency formulation of optimum detection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1377–1384, Sept. 1988.
- [3] A. H. Costa and G. F. Boudreaux-Bartels, "An overview of aliasing errors in discrete-time formulations of time-frequency representations," *IEEE Trans. Signal Processing*, vol. 47, pp. 1463–1474, May 1999.
- [4] M. S. Richman, T. W. Parks, and R. G. Shenoy, "Discrete-time, discrete-frequency, time-frequency analysis," *IEEE Trans. Signal Processing*, vol. 46, pp. 1517–1527, June 1998.
- [5] J. C. O'Neill, P. Flandrin, and W. J. Williams, "On the existence of discrete Wigner distributions," *IEEE Signal Processing Lett.*, vol. 6, pp. 304–306, Dec. 1999.
- [6] J. C. O'Neill and W. J. Williams, "Shift covariant time-frequency distributions of discrete signals," *IEEE Trans. Signal Processing*, vol. 47, pp. 789–799, Mar. 1999.
- [7] L. Stanković, "A method for time-frequency signal analysis," *IEEE Trans. Signal Processing*, vol. 42, pp. 225–229, Jan. 1994.
- [8] L. Stanković and I. Djurović, "A note on 'An overview of aliasing errors in discrete-time formulations of time-frequency representations'," *IEEE Trans. Signal Processing*, vol. 49, pp. 257–259, Jan. 2001.
- [9] N. Marinovich, "The singular value decomposition of the Wigner distribution and its applications," in *The Wigner Distribution: Theory and Applications in Signal Processing*, W. Mecklenbräuker and F. Hlawatsch, Eds. Amsterdam, The Netherlands: Elsevier, 1997, pp. 319–373.
- [10] D. L. Jones and A. M. Sayeed, "Blind quadratic and time-frequency based detectors from training data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1995, pp. 1033–1036.
- [11] C. Richard and R. Lengellé, "Data-driven design and complexity control of time-frequency detectors," *Signal Process.*, vol. 77, pp. 37–48, 1999.
- [12] B. Samimy, G. Rizzoni, A. M. Sayeed, and D. L. Jones, "Design of training data-based quadratic detectors with application to mechanical systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1996, pp. 1767–1770.
- [13] G. Matz and F. Hlawatsch, "Time-frequency methods for detection with application to the detection of knock in car engines," in *Proc. IEEE-SP Workshop Statist. Signal Array Process.*, 1998, pp. 196–199.
- [14] B. W. Gillespie and L. E. Atlas, "Optimizing time-frequency kernels for classification," *IEEE Trans. Signal Processing*, vol. 49, pp. 485–496, Mar. 2001.
- [15] F. Hlawatsch and W. Krattenthaler, "Bilinear signal synthesis," *IEEE Trans. Signal Processing*, vol. 40, pp. 352–363, Feb. 1992.
- [16] C. Richard, "Linear redundancy of information carried by the discrete Wigner distribution," *IEEE Trans. Signal Processing*, vol. 49, pp. 2536–2544, Nov. 2001.
- [17] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory Probab. Appl.*, vol. 16, pp. 264–280, 1971.
- [18] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1999.
- [19] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. New York: Springer-Verlag, 1996.
- [20] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
- [21] C. Richard and R. Lengellé, "Sur le contrôle de la complexité des détecteurs opérant dans le domaine temps-fréquence par le biais de la fonction de paramétrization," in *Proc. XVII Colloque Gretsi*, Gretsi, France, 1999, pp. 905–908.
- [22] D. L. Jones and R. G. Baraniuk, "An adaptive optimal kernel time-frequency representation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2361–2371, Oct. 1995.
- [23] P. Gonçalves and E. Payot, "Adaptive diffusion equation for time-frequency representations," in *Proc. IEEE Digital Signal Process. Workshop*, 1998.
- [24] M. Davy, C. Doncarli, and G. F. Boudreaux-Bartels, "Improved optimization of time-frequency based signal classifiers," *IEEE Signal Processing Lett.*, vol. 8, pp. 52–57, Feb. 2001.
- [25] J. Gosme, P. Gonçalves, C. Richard, and R. Lengellé, "Adaptive diffusion and discriminant analysis for complexity control of time-frequency detectors," in *Proc. XI Eur. Signal Process. Conf.*, 2002.



Cédric Richard (M'01) was born in Sarrebourg, France, on January 24, 1970. He received the Dipl.-Ing. and the M.S. degrees in 1994 and the Ph.D. degree in 1998 from Compiègne University of Technology, Compiègne, France, all in electrical and computer engineering.

Since 1995, he has been with the Systems Modeling and Dependability Laboratory, Troyes University of Technology, Troyes, France, where he is currently an Associate Professor within the Department of Information Systems Engineering. His current re-

search interests involve time-frequency analysis, statistical estimation and decision theories, and pattern recognition.