Signal-dependent time–frequency analysis using a radially Gaussian kernel

Richard G. Baraniuk¹

Laboratoire de Traitement du Signal, Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France

Douglas L. Jones

Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois, 1308 W. Main, Urbana, IL 61801, USA

Received 19 September 1991

Revised 30 April 1992

Abstract. Time–frequency distributions are two-dimensional functions that indicate the time-varying frequency content of one-dimensional signals. Each bilinear time–frequency distribution corresponds to a kernel function that controls its cross-component suppression properties. Current distributions rely on fixed kernels, which limit the class of signals for which a given distribution performs well. In this paper, we propose a *signal-dependent* kernel that changes shape for each signal to offer improved time–frequency representation for a large class of signals. The kernel design procedure is based on quantitative optimization criteria and two-dimensional functions that are Gaussian along radial profiles. We develop an efficient scheme based on Newton's algorithm for finding the optimal kernel; the cost of computing the signal-dependent time–frequency distribution is close to that of fixed-kernel methods. Examples using both synthetic and real-world multi-component signals demonstrate the effectiveness of the signal-dependent approach – even in the presence of substantial additive noise. An attractive feature of this technique is the ease with which application-specific knowledge can be incorporated into the kernel design procedure.

Zusammenfassung. Zeit–Frequenz-Verteilungen sind zweidimensionale Funktionen, die den zeitabhängigen Frequenzinhalt eindimensionaler Signale anzeigen. Jede bilineare Zeit–Frequenz-Verteilung enthält eine Kern-Funktion, die die Kreuz-Komponenten-Unterdrückungseigenschaften bestimmt. Übliche Verteilungen beruhen auf festen Kern-Funktionen, wodurch die Klasse von Signalen eingegrenzt wird, für die die jeweilige Verteilung günstig ist. In dieser Arbeit schlagen wir eine *signalabhängige* Kernfunktion vor, wodurch die Filterform sich für jedes Signal ändert, um verbesserte Zeit–Frequenz-Auflösungen für eine große Klasse von Signalen zu erhalten. Die Entwurfs-Prozedur für die Kerne basiert auf quantitativen Optimierungskriterien und zweidimensionalen Funktionen, die gaußisch längs radialer Profile sind. Wir entwickeln ein effizientes Verfahren, basierend auf dem Newton-Algorithmus, zur Berechnung der optimalen Kerne; der Aufwand zur Berechnung der signalabhängigen Zeit–Frequenz-Verteilung ist vergleichbar mit dem für feste Kern-Funktionen. Einige Beispiele, in denen sowohl synthetische als auch reale Multi-Komponenten-Signale benutzt werden, zeigen die Effizienz der signalabhängigen Lösung – selbst im Falle von additivem Rauschen. Eine attraktive Besonderheit dieser Technik ergibt sich dann, wenn anwendungsspezifische Eigenschaften in den Entwurf der Kern-Funktion eingebracht werden können.

Résumé. Les distributions temps–fréquence sont des fonctions bi-dimensionnelles qui indiquent le contenu fréquentiel variant dans le temps de signaux mono-dimensionnels. Toute distribution temps–fréquence bilinéaire correspond à une fonction noyau

Correspondence to: Dr. Richard G. Baraniuk, Laboratoire de Traitement du Signal, Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France.

¹ On leave from Department of Electrical and Computer Engineering, Rice University, P.O. Box 1892, Houston, Texas 77251-1892, USA.

qui contrôle les propriétés de suppression des composantes croisées. Les distributions actuelles reposent sur des noyaux fixes, ce qui limite la classe de signaux pour lesquels une distribution donnée remplit bien son rôle. Nous proposons dans cet article un noyau *dépendant du signal* dont la forme change pour chaque signal de manière à fournir une meilleure représentation temps-fréquence pour une large classe de signaux. La procédure de conception du noyau est basée sur des critères d'optimisation quantitatifs et des fonctions bi-dimensionnelles gaussiennes selon leurs profils radiaux. Nous développons une technique efficace de recherche du noyau optimale basée sur l'algorithme de Newton; le coût en calcul de la distribution temps-fréquence dépendant du signal est proche de celui des méthodes à noyau fixe. Des exemples utilisant à la fois des signaux synthétiques et des signaux expérimentaux multi-composants montrent l'efficacité de l'approche dépendant du signal, même en présence d'un bruit additif substantiel. Un des attraits de cette technique est l'aisance avec laquelle des connaissances sur l'application peuvent être incorporées dans la procédure de conception du noyau.

Keywords. Time-frequency analysis; bilinear time-frequency distributions; time-varying spectral analysis; Wigner distribution.

1. Introduction

Time-frequency distributions (TFDs) have proven useful for analyzing the time-varying frequency content of a wide variety of signals, including speech, music and other acoustical signals, biological signals, radar and sonar signals, and geophysical signals. Among the representations that have been developed are the TFDs of Cohen's class [9]. Examples include the Wigner distribution, the spectrogram (the squared-magnitude of the short-time Fourier transform), the Choi-Williams distribution [6], and the cone-kernel distribution [24]. A distribution $P(t, \omega)$ from Cohen's class can be interpreted as the two-dimensional Fourier transform of a weighted version of the ambiguity function (AF) of the signal

$$P(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\theta, \tau) \Phi(\theta, \tau) \times e^{-j\theta t - j\tau \omega} d\theta d\tau, \quad (1)$$

where $A(\theta, \tau)$ is the AF of the signal $s(t)$, given by

$$A(\theta, \tau) = \int_{-\infty}^{\infty} s^*\left(t - \frac{\tau}{2}\right) s\left(t + \frac{\tau}{2}\right) e^{j\theta t} dt. \quad (2)$$

The weighting function $\Phi(\theta, \tau)$ is called the *kernel*; it determines the specific properties of the distribution. The product $A(\theta, \tau)\Phi(\theta, \tau)$ is known as the *characteristic function*.

The AF is a bilinear functional of the signal, so it exhibits cross-components [13], which, if allowed

to pass into $P(t, \omega)$, can reduce auto-component resolution, obscure the true signal features, and make interpretation of the distribution difficult.² Therefore, the kernel is often selected to weight the AF such that the auto-components, which are centered at the origin of the (θ, τ) ambiguity plane, are passed, while the cross-components, which are located away from the origin, are suppressed [10, 13]. That is, in order to suppress cross-components, $\Phi(\theta, \tau)$ should be the frequency response of a two-dimensional lowpass filter. There is a tradeoff between cross-component suppression and auto-component concentration: generally, as the passband region of the kernel is made smaller, the amount of cross-component suppression increases, but at the expense of auto-component concentration.

For a given class of signals, a TFD will be said to offer good performance if, for each signal in the class, it achieves a high degree of both cross-component suppression and auto-component concentration and provides an accurate representation of the time-frequency content of the signal. An important practical and theoretical goal of time-frequency analysis is to define a TFD that attains good performance for a large class of different signals.

² Although cross-components are a great nuisance when the visual appearance of a TFD is important, they do provide important information on the relative phases of components in a multicomponent signal.

Traditionally, TFDs have employed fixed kernels, with fixed passband and stopband regions. However, specification of a fixed kernel limits the class of signals for which the time-frequency representation performs well. That is, given any fixed kernel, it is always possible to find signals with either significant cross-component energy in the passband of the kernel, or significant auto-component energy in the stopband of the kernel. The same limitation exists for the class of fixed-kernel, bilinear time-scale distributions, which includes the squared magnitude of the continuous wavelet transform [11].

The limitations of fixed-kernel TFDs can be illustrated with two example signals. The first is the sum of two short Gaussian pulses,

$$s_1(t) = e^{-a(t-T_0)^2} + e^{-a(t+T_0)^2}, \quad (3)$$

with $a = 0.25 \text{ sec}^{-2}$ and $T_0 = 5 \text{ sec}$. The second is the sum of two linear-FM 'chirp' signals,

$$s_2(t) = e^{-a(t+T_0)^2 + jc(t+T_0)^2 - j\omega_1(t+T_0)} + e^{-a(t-T_0)^2 + jc(t-T_0)^2 - j\omega_2(t-T_0)}, \quad (4)$$

with $a = 0.004 \text{ sec}^{-2}$, $c = 0.025 \text{ rad sec}^{-2}$, $T_0 = 3.5 \text{ sec}$, $\omega_1 = 1.72 \text{ rad sec}^{-1}$ and $\omega_2 = 1.42 \text{ rad sec}^{-1}$. Both signals are sampled once per second for computational purposes. Signal $s_1(t)$ is shown in Fig. 1(a); the real part of $s_2(t)$ is shown in Fig. 1(b). The AFs of these signals are given in Figs. 1(c) and 1(d). Note the locations of the auto- and cross-components.

The kernel of the Wigner distribution is constant

$$\Phi_w(\theta, \tau) = 1, \quad (5)$$

so both auto- and cross-components are passed from the AF into the TFD without attenuation. The Wigner distributions of the example signals are shown in Fig. 2; although the auto-components are highly concentrated, there are large cross-components.

The exponential kernel of the Choi-Williams distribution,

$$\Phi_{cw}(\theta, \tau) = e^{-\theta^2 \tau^2 / \sigma}, \quad (6)$$

Fig. 1. Example signals for the TFDs of Figs. 2-5. (a) The signal given in (3), consisting of two short Gaussian pulses. (b) The real part of the signal given in (4), consisting of two parallel Gaussian chirps. (c) Equal-energy contour plot of the magnitude of the ambiguity function (AF) of the pulse signal. (d) The AF of the chirp signal.

is large along both the θ and τ axes (see Fig. 3(a)). Thus, it performs well for signals whose AF auto-components lie along these axes. This is the case with, for example, the pulse signal of Fig. 1(a). In the Choi-Williams distribution of this signal (see Fig. 3(b)), the cross-components are suppressed at the expense of some smearing and shouldering of the auto-components. The Choi-Williams distribution performs poorly, however, for signals whose AF auto-components do not lie along the θ and τ axes. For example, signals with substantial frequency modulation are poorly represented (see Fig. 3(c)), because the kernel severely truncates the auto-components of such signals.

The cone-kernel is given by

$$\Phi_c(\theta, \tau) = \begin{cases} e^{-2a\tau^2} \sin(\theta|\tau|a^{-1})/\theta, & |\tau| \leq \beta, a \geq 2, \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

and is shown in Fig. 4(a). The time-frequency representation of the cone-kernel distribution is excellent for pulse-like signals whose AF auto-components lie near the θ axis (see Fig. 4(b)), but

Fig. 2. Wigner distributions of (a) the pulse and (b) the chirp signals of Fig. 1. The Wigner kernel is $\Phi_w(\theta, \tau) = 1$, so the distribution is simply the two-dimensional Fourier transform of the AF. Note the large cross-components that appear midway between the auto-components.

Fig. 3. (a) Equal-energy contour plot of the Choi-Williams kernel for $\sigma = 10$. (b) Choi-Williams distribution of the pulse signal of Fig. 1(a). Since the AF auto-components of this signal lie along the θ axis and the cross-components lie off the θ axis (see Fig. 1(c)), the auto-components are passed into the TFD, while the cross-components are suppressed. (c) Choi-Williams distribution of the chirp signal of Fig. 1(b). The AF auto-components line up with neither the θ nor the τ axis (see 1(d)) and so are truncated.

poor for some chirp signals, as evidenced by Fig. 4(c).

The spectrogram kernel is related to the AF of the analysis window $w(t)$, by [7]

$$\Phi_S(\theta, \tau) = \int_{-\infty}^{\infty} w^*\left(t - \frac{\tau}{2}\right) w\left(t + \frac{\tau}{2}\right) e^{-j\theta t} dt. \quad (8)$$

Signal Processing

The time-frequency representation is excellent for signal components that resemble the window [10], but all mismatched components are distorted. Figure 5 shows the spectrogram kernel and spectrograms of the example signals computed using a Gaussian window of length similar to the effective length of the signal components of $s_1(t)$. If the analysis window of the spectrogram, and hence the

Fig. 4. (a) The cone-kernel for $a=2$, $\alpha=0.036 \text{ sec}^{-2}$, and $\beta=8 \text{ sec}$. (b) Cone-kernel distribution of the pulse signal of Fig. 1(a). Since the AF auto-components line up with the θ axis, an excellent time-frequency representation results. (c) Cone-kernel distribution of the chirp signal of Fig. 1(b). Since the AF auto-components do not lie near the axes, the representation is poor.

kernel, is matched to one of the signal components, the result is called the *matched-filter spectrogram*. As Fig. 5(b) shows, the matched-filter technique can yield excellent results. However, the technique works for only one type of signal component, plus requires a priori knowledge of the form of the component. Mismatch between the window and other signal components can result in severe distortion of the representation, as in Fig. 5(c).

The two example signals differ only in their orientation in time-frequency (compare Figs. 2(a) and 2(b)), yet none of the aforementioned TFDs perform adequately for both. The fundamental limitation of fixed kernel TFDs should now be clear: a fixed kernel yields good performance only for certain configurations of AF auto- and cross-components, and thus only for a limited class of signals. Since the locations of the auto- and cross-components depend on the signal to be analyzed, we expect to obtain good performance for a broad

class of signals only by using a *signal-dependent* kernel. A signal-dependent kernel can provide a good time-frequency representation by adjusting its shape to pass auto-components and suppress cross-components, regardless of their shape and orientation.

Several previous authors have recognized the need for signal-dependent kernels. The adaptive spectrogram representation for speech signals developed by Glinski adapts the window based on a segmentation (provided by the user) of the signal into pitch periods [12]. Nuttall designs a kernel composed of Gaussian components based on information that the user provides after viewing the Wigner distribution [21]. Jones and Boashash adapt the modulation rate of a fixed window to match an estimate of the signal's instantaneous frequency [15]. Kadambe, Boudreaux-Bartels and Duvaut utilize an adaptive filtering technique coupled with AR modeling and clustering to design

Fig. 5. (a) The kernel generating the spectrogram (the squared-magnitude of the short-time Fourier transform) when the analysis window is matched to the signal components of Fig. 1(a). (b) (Matched-filter) spectrogram of the pulse signal of Fig. 1(a). (c) Spectrogram of the chirp signal of Fig. 1(b) computed using the same window.

kernels [19]. Optimal smoothing kernels are considered by Andrieux et al., but only for simple signals of the form $s(t) = e^{jv(t)}$, and only for the restrictive class of Gaussian kernels [1]. Jones and Parks develop a technique using Gaussian kernels which vary with time and frequency to maximize a local measure of signal-energy concentration [16, 17]. It often works well but is quite computationally expensive. These efforts constitute the present state of the art of signal-dependent time-frequency analysis. While meritorious for demonstrating the potential of signal dependency, each of these approaches suffers from one or more substantial drawbacks; each either requires human intervention, is ad hoc, excessively restricts the class of allowable kernels, or is computationally expensive.

We propose a new procedure for selecting a signal-dependent kernel. Given a signal, the

method automatically designs a kernel that is optimal with respect to a set of performance criteria that attempt to capture, mathematically, the kernel properties that lead to good performance. Since we consider a large class of kernels, good performance is expected for a wide range of signals. Moreover, the computational complexity of the procedure is close to that of fixed-kernel techniques.

This paper is organized as follows. In the next section, we present an optimization-based design procedure for a signal-dependent kernel. In Section 3 we derive an efficient algorithm to solve the optimization problem. Examples that demonstrate the effectiveness of the signal-dependent approach are given in Section 4. In Section 5 we propose extensions to the optimization formulation to incorporate application-specific knowledge about the signal being analyzed. A discussion and conclusion are offered in the final section.

2. Optimal kernel design

In order to find the bilinear TFD that provides the 'best' time-frequency representation for a given signal, we formulate the signal-dependent kernel design procedure as an optimization problem. The problem formulation requires a class of two-dimensional kernel functions from which the optimal kernel is chosen, and a performance index that measures the quality of the time-frequency representation with respect to criteria deemed important by the designer. The kernel maximizing the value of the performance measure is selected as the optimal kernel for the signal.

The class of kernels will be specified by a set of constraints. We must ensure that it contains a great enough variety of kernel shapes to provide good performance for all signals of interest in a given application. Worth considering are constraints that force the kernel to be smooth, suppress cross-components [10], satisfy the time and frequency marginal distributions [6], preserve the time or frequency support of the signal [24], or satisfy Moyal's formula [9]. The performance measure will be chosen to yield a tractable optimization problem that can be solved efficiently. One example of a useful performance index is a measure of the signal-energy concentration of the TFD [16, 17].

2.1. The 1/0 optimal kernel method

An example of a method for optimal signal-dependent kernel design is discussed in [2, 5]. In those papers, the constraints force the kernel to be bounded, radially nonincreasing, and of finite volume – a two-dimensional lowpass filter – while

the performance index measures the energy passed from the AF into the TFD. The optimal kernel yields excellent results for a variety of signals; the optimal-kernel TFDs for the signals discussed in the Introduction are given in Fig. 6. A fundamental property of this technique is that the optimal kernel is essentially a 1/0 mask, and hence can introduce ringing into the TFD. In applications where ringing is objectionable, suboptimal tapering of the kernel is necessary (the TFDs of Fig. 6 were prepared with such tapering). In this paper, we formulate a new kernel design procedure using radially Gaussian functions [3]. This class of kernels is inherently tapered and may potentially yield more desirable results for many signals.

2.2. Radially Gaussian kernels

An attractive alternative to the 1/0 kernel should possess the following properties: it should be low-pass to suppress cross-components and noise in the TFD, it should be smooth to reduce ringing artifacts in the TFD, and it should take a functional form for which an optimization problem can be easily solved (for example, a gradient ascent optimization procedure requires existence of the first partial derivatives of the kernel).

A functional form that satisfies all of the above requirements is a *radially Gaussian kernel*. A radially Gaussian kernel is a two-dimensional function that is Gaussian along any radial profile

$$\Phi(\theta, \tau) = e^{-(\theta^2 + \tau^2)/2\sigma^2(\psi)}. \quad (9)$$

The function $\sigma(\psi)$ controls the 'spread' of the Gaussian at radial angle ψ ; we will call $\sigma(\psi)$ the *spread function*. The angle ψ is measured between

Fig. 6. 1/0 optimal kernel TFDs of the example signals of Fig. 1. Details of the computation of this signal-dependent time-frequency representation are given in [2, 5].

Fig. 7. (a) A spread function $\sigma(\psi)$ and its corresponding radially Gaussian kernel $\Phi(\theta, \tau) = e^{-(\theta^2 + \tau^2)/2\sigma^2(\psi)}$ in (b) equal-energy contours and (c) three dimensions. The spread function controls the spread of the kernel at angle $\psi = \arctan \tau/\theta$.

the radial line through the point (θ, τ) and the θ axis,

$$\psi = \arctan \frac{\tau}{\theta}. \quad (10)$$

The spread function determines the basic shape of the equal-energy contours of its corresponding kernel. Clearly, if $\sigma(\psi)$ is smooth, then $\Phi(\theta, \tau)$ is also smooth. A radially Gaussian kernel is a generalization of a two-dimensional lowpass Gaussian kernel. However, while the basic contour shape of a Gaussian must be elliptical, the basic contour shape of a radially Gaussian kernel is arbitrary. An example of a spread function and its corresponding radially Gaussian kernel is given in Fig. 7.

Signal Processing

It is natural to express radially Gaussian kernels in polar coordinates³ using $r = \sqrt{\theta^2 + \tau^2}$ as the radius variable,

$$\Phi(r, \psi) = e^{-r^2/2\sigma^2(\psi)}. \quad (11)$$

2.3. Continuous-time optimization formulation

Since the shape of a radially Gaussian kernel is completely parameterized by the one-dimensional function $\sigma(\psi)$, finding the optimal radially Gaussian kernel for a signal is equivalent to finding the optimal spread function $\sigma_{\text{opt}}(\psi)$ for the signal. Therefore, given a signal, we define the optimal

³ In the sequel, (θ, τ) will represent the ambiguity plane in rectangular coordinates, while (r, ψ) will represent polar coordinates.

kernel Φ_{opt} as the radially Gaussian function whose spread function $\sigma_{\text{opt}}(\psi)$ solves the following optimization problem:

$$\max_{\psi} \int_0^{2\pi} \int_0^{\infty} |A(r, \psi) \Phi(r, \psi)|^2 r \, dr \, d\psi, \quad (12)$$

subject to

$$\Phi(r, \psi) = e^{-r^2/2\sigma^2(\psi)}, \quad (13)$$

$$\frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} |\Phi(r, \psi)|^2 r \, dr \, d\psi \leq \alpha, \quad (14)$$

$$\alpha \geq 0,$$

where $A(r, \psi)$ is the AF of the signal in polar coordinates. The constraint (13) limits the scope of the optimization to the class of radially Gaussian kernels; the constraint (14) limits the volume of the optimal kernel (we derive guidelines for the choice of the volume parameter α in the next section). The special structure of the radially Gaussian kernels permits a simplification of (14) to an equivalent volume constraint,

$$\frac{1}{2\pi} \int_0^{\pi} \sigma^2(\psi) \, d\psi \leq \alpha. \quad (15)$$

Also, since the AF is symmetric about the origin, that is,

$$|A(r, \psi)|^2 = |A(r, \psi + \pi)|^2, \quad (16)$$

$\sigma_{\text{opt}}(\psi)$ must be determined only on the interval $0 \leq \psi < \pi$.

The constraints (13)–(15) and performance measure (12) are formulated so that the optimal kernel passes auto-components and suppresses cross-components. The constraints force the optimal kernel to be a lowpass filter of fixed volume α . As discussed in the Introduction, lowpass kernels are desirable, because the AF auto-components are centered at the origin of the ambiguity plane, while the cross-components tend to lie away from the origin [10, 13].

The constraints do not dictate the exact shape of the passband of the optimal radially Gaussian

kernel; its shape is determined by maximizing the performance measure. Clearly, in order to maximize the performance measure, $\Phi(r, \psi)$ should be large where $A(r, \psi)$ is large, regardless of whether the peaks correspond to auto- or cross-components. However, assuming that the auto- and cross-components are somewhat separated in the ambiguity plane, the radially Gaussian constraint imposes a penalty on kernels whose passbands extend over cross-components. Kernels having large spread in the direction of cross-components must waste precious kernel volume over the regions *between* the auto- and cross-components, where $|A(r, \psi)|^2$ and thus also the $|A(r, \psi) \Phi(r, \psi)|^2$ contribution to the performance measure is small. Hence, the optimization formulation favors kernels that pass the components concentrated at the origin – precisely the auto-components.

By the same reasoning, we infer that the performance of this kernel design technique deteriorates for signals whose AF auto- and cross-components overlap considerably. However, for this class of signals all current methods of time-frequency analysis fail without a fortuitous conjunction of auto-components and kernel passband.

Note that the performance measure and constraints are radially symmetric and, hence, insensitive to the orientation angle of the signal components in the ambiguity or time-frequency planes. Furthermore, it is straightforward to show that the formulation is insensitive to the time scale of the signal. We show this in Appendix A.

The optimization problem (12)–(15) is formulated strictly to find the kernel that optimally passes auto-components and suppresses cross-components. Other constraints that result in additional kernel properties are considered in Section 5.

Conceptually, the computation of the optimal kernel TFD involves first finding the AF of the signal in polar coordinates. The optimization problem (12)–(15) is then solved for the optimal spread function $\sigma_{\text{opt}}(\psi)$, which determines the optimal radially Gaussian kernel $\Phi_{\text{opt}}(\theta, \tau)$. The optimal kernel TFD $P_{\text{opt}}(t, \omega)$ is then computed via (1).

2.4. Kernel volume selection

By controlling the volume under the optimal kernel, the parameter α controls the tradeoff between cross-component suppression and smearing of the auto-components. If α is too small, the kernel will induce excessive smearing of the auto-components. If α is too large, extra kernel volume will be available to extend over the cross-components, and little cross-component suppression will result. While the exact value of α is application-dependent, we offer some guidelines for its selection in this section.

A reasonable lower bound for α is the volume of a spectrogram kernel, since there appear to be few benefits gained from more smoothing [18]. Since the spectrogram kernel is the AF of the analysis window, its volume is easily calculated. Assuming that the window has unit energy yields

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Phi_s(\theta, \tau)|^2 d\theta d\tau \\ = |\Phi_s(0, 0)|^2 = 1. \end{aligned} \quad (17)$$

A lower bound for the kernel volume parameter is, therefore,

$$\alpha \geq 1. \quad (18)$$

We can derive a heuristic upper bound for α by computing the amount of smearing in the TFD for a simple Gaussian mono-component signal

$$s(t) = \frac{1}{\sqrt{4\pi}} e^{-t^2/2}. \quad (19)$$

The AF and optimal kernel of volume α for this signal are given by

$$\begin{aligned} A(\theta, \tau) &= e^{-(\theta^2 + \tau^2)/4}, \\ \Phi_{\text{opt}}(\theta, \tau) &= e^{-(\theta^2 + \tau^2)/4\alpha}. \end{aligned} \quad (20)$$

Using (1), the optimal-kernel TFD for this signal is

$$P(t, \omega) = \frac{2\alpha}{\alpha + 1} e^{-(t^2 + \omega^2)(\alpha/(\alpha + 1))}. \quad (21)$$

We quantify the amount of smearing in the TFD by computing the radius r_1 in the (t, ω) plane at which this auto-component decays to e^{-1} , or 37%, of its peak height

$$r_1 = \sqrt{\frac{\alpha + 1}{\alpha}} = \sqrt{1 + \frac{1}{\alpha}}. \quad (22)$$

Since $r_1 = 1$ when no smoothing is performed ($\alpha \rightarrow \infty$), this expression represents a 'smearing index' for any value of α . For $\alpha \gg 1$, we find that essentially no smearing occurs and that there is essentially no benefit to increasing α further. However, for multicomponent signals, kernels with large values of α are more likely to pass extra cross-component energy into the TFD. Therefore, the value of α for which $r_1 = 1.1$ (the 'knee point' in the r_1 versus α curve) seems a reasonable upper bound

$$\alpha \leq 5. \quad (23)$$

Note that these bounds should be used only as a guide to prevent over- and under-smoothing of the optimal-kernel TFD. The actual value selected for α will depend on the amount of smoothing required for a specific application. In some applications, the signal cross-components are located far out in the ambiguity plane, and values of α considerably larger than 5 can still result in adequate cross-component suppression.

3. Optimal kernel solution

In practice, TFDs are computed at discrete time and frequency locations, so we now reformulate the optimization problem by discretizing (12)–(15). With suitably dense sampling, the discrete kernel solution converges weakly to the continuous-time kernel solution. Before presenting the discretized optimization problem and an algorithm to solve it, we first discuss several discretization issues.

3.1. Discretization issues

Efficient computation of the optimal-kernel TFD requires both rectangular and polar coordinate systems. Computation of the optimal kernel will require samples of the AF on a polar grid, while efficient computation of the signal-dependent TFD via a two-dimensional fast Fourier transform (FFT) will require samples of the characteristic function – the AF-kernel product – on a rectangular grid. We now briefly summarize the results of [4], which contains details regarding the discretization grids.

We define the $M \times N$ rectangular sampled AF $A_d(m, n)$ as a uniformly sampled version of the rectangular coordinate continuous AF,

$$\begin{aligned} A_d(m, n) &= A(\theta, \tau)|_{\theta=m\Delta\theta, \tau=n\Delta\tau,} \\ m &= -\frac{1}{2}M+1, \dots, \frac{1}{2}M, \\ n &= -\frac{1}{2}N+1, \dots, \frac{1}{2}N. \end{aligned} \quad (24)$$

The rectangularly sampled kernel and characteristic function are defined similarly. Sampling the continuous AF, kernel, and characteristic function on too coarse a grid could result in loss of important signal information. In [4], we show that if a signal is bandlimited to a frequency $(\pi/2T)$ rad sec^{-1} and sampled every T seconds to create a discrete signal L samples long, then setting

$$M = N = L \quad (25)$$

and

$$\Delta_\theta = \frac{2\pi}{LT}, \quad \Delta_\tau = 2T \quad (26)$$

ensures that no information in the AF is lost upon discretization. Performing this discretization on (2) yields a formula for the direct calculation of the rectangularly sampled AF,

$$\begin{aligned} A_d(m, n) &= T \sum_{k=0}^{L-1} s^*(kT-nT) \\ &\quad \times s(kT+nT) e^{j2\pi mk/L}. \end{aligned} \quad (27)$$

This can be efficiently implemented using a series of L -point FFTs.

Also in [4], we define the $P \times Q$ polar sampled AF $A_p(p, q)$ as the set of uniformly spaced samples of a scaled version $A'(r, \psi)$ of the polar AF of the signal to be analyzed,

$$\begin{aligned} A_p(p, q) &= A'(r, \psi)|_{r=p\Delta_r, \psi=q\Delta_\psi,} \\ p &= 0, 1, \dots, P-1, \\ q &= 0, 1, \dots, Q-1. \end{aligned} \quad (28)$$

The AF A' is scaled to have an aspect ratio⁴ of unity, which ensures the most even distribution of uniformly spaced polar samples in the ambiguity plane. In (28), p and q represent uniformly spaced samples in radius and angle, respectively, while Δ_r and Δ_ψ represent the sample step sizes. Reasonable values for the sampling parameters are

$$P = \frac{L}{\sqrt{2}}, \quad Q = \pi L \quad (29)$$

and

$$\Delta_r = 2\sqrt{\frac{\pi}{L}}, \quad \Delta_\psi = \frac{1}{L}. \quad (30)$$

Note that AF samples are required only in the upper half of the ambiguity plane, due to the symmetry relation (16). The polar sampled AF can be computed either directly or using two-dimensional interpolation from a rectangularly sampled AF.

Applying the same sampling conditions to (13) yields the polar sampled radially Gaussian kernel

$$\Phi_p(p, q) = e^{-(p\Delta_r)^2/2\sigma_q^2}, \quad (31)$$

which is parameterized by a positive spread vector $\sigma \in \mathbb{R}^Q$, whose elements correspond to samples of the spread function,

$$\sigma_q = \sigma(q\Delta_\psi). \quad (32)$$

Given a spread vector, the corresponding rectangularly sampled radially Gaussian kernel can be computed by converting each rectangular sample

⁴ The aspect ratio of the AF is defined as its height to width ratio and is given by $2LT(2\pi/T)^{-1} = LT^2/\pi$.

point (m, n) to equivalent polar coordinates $(r(m, n), \psi(m, n))$:

$$\Phi_d(m, n) = e^{-r^2(m, n)/2\sigma^2(\psi(m, n))}. \quad (33)$$

The radius $r(m, n)$ can be calculated exactly, while the spread function value at angle $\psi(m, n)$ should be interpolated from the elements of σ to ensure that the kernel remains smooth [4]. The interpolation need not be elaborate – cubic polynomial interpolation works fine in practice.

3.2. Discrete optimization formulation

Performing the discretization of (12), (13) and (15), we define the optimal discrete kernel as the radially Gaussian function whose spread vector σ_{opt} solves the following optimization problem, which we write in standard optimization notation:

$$\max_{\sigma \geq 0} f(\sigma), \quad (34)$$

subject to

$$\|\sigma\|^2 \leq \gamma^2, \quad (35)$$

where

$$f(\sigma) = \sum_{q=0}^{Q-1} \sum_{p=1}^{P-1} p |A_p(p, q)|^2 e^{-(p\Delta_r)^2/\sigma_q^2}, \quad (36)$$

$$\|\sigma\|^2 = \sum_{q=0}^{Q-1} \sigma_q^2, \quad (37)$$

and

$$\gamma = \sqrt{\frac{2\pi\alpha}{\Delta_\psi}}. \quad (38)$$

The expression $\sigma \geq 0$ should be interpreted as a component-wise inequality. In the next section we propose an efficient algorithm to solve this constrained nonlinear programming problem in variables σ_q .

Solving this problem determines the optimal radially Gaussian kernel Φ_{opt} . Computation of the kernel on a rectangular grid and multiplication by the rectangularly sampled AF yields the characteristic function, whose two-dimensional FFT is the discrete optimal-kernel TFD.

3.3. The step-project ascent algorithm

Define the *feasible set* of the optimization (34)–(35) as

$$\mathcal{F} = \{\sigma \in \mathbb{R}^Q : \|\sigma\|^2 \leq \gamma^2, \sigma \geq 0\}. \quad (39)$$

Stated in words, the goal of the optimization is to find the vector $\sigma_{\text{opt}} \in \mathcal{F}$ that maximizes $f(\sigma)$.

The step-project algorithm is motivated by the concept of steepest ascent [20], a well known iterative technique for solving *unconstrained* optimization problems. Denoting the estimate of the solution to (34)–(35) at iteration k by $\sigma(k)$, a steepest ascent step takes the form

$$\text{Step: } \sigma(k+1) = \sigma(k) + \mu(k) \nabla f(k), \quad (40)$$

where $\mu(k) > 0$ is the step size and $\nabla f(k)$ is the gradient of f evaluated at the point $\sigma(k)$

$$\nabla f(k) = \left[\frac{\partial f}{\partial \sigma_1(k)}, \dots, \frac{\partial f}{\partial \sigma_{Q-1}(k)} \right]^T, \quad (41)$$

$$\frac{\partial f}{\partial \sigma_q(k)} = \frac{2\Delta_r^2}{\sigma_q^3(k)} \sum_{p=1}^{P-1} p^3 |A_p(p, q)|^2 e^{-(p\Delta_r)^2/\sigma_q^2(k)}. \quad (42)$$

The iteration (40) does not lead to a solution of the constrained problem (34)–(35). Since $\nabla f(k) \geq 0$ (see (42) and recall that $\sigma \geq 0$), the value $\|\sigma(k)\|^2$ will increase each iteration, eventually violating the norm constraint (35). However, if $\sigma(k)$ is rescaled every iteration by projecting it back onto $\mathcal{F}$,

$$\text{Project: } \sigma(k+1) \leftarrow \sigma(k+1) \frac{\gamma}{\|\sigma(k+1)\|}, \quad (43)$$

a convergent algorithm results. This rescaling changes the volume, but not the basic shape, of the kernel corresponding to $\sigma(k+1)$. Readers familiar with the gradient-projection algorithm [23] will notice a strong similarity to that method.

Contained in [4] is a procedure for selecting the iteration step size $\mu(k)$. The procedure ensures that the progression of iterations satisfies a set of technical conditions similar to the well-known Goldstein–Armijo principle [20]. These conditions

are used in [4] to prove that the step-project algorithm is *globally stable*; that is, it is guaranteed to converge to a local maximum of $f(\sigma)$ from any feasible initial starting point $\sigma(0)$. It is important to note that global stability does not guarantee that the algorithm will find the global maximum of $f(\sigma)$. The nonlinear performance surface defined by (36) has been demonstrated to possess local maxima for some signals. While potentially this is a problem, in practice this method has always converged to good solutions.

Each iteration of the step-project algorithm involves $O(PQ)$ computations, so the total cost of finding the optimal spread vector is $O(PQi)$, where i is the number of iterations required for adequate convergence. For all of the examples we have considered, the value of i has been bounded approximately by $5 \leq i \leq 30$.

3.4. Enhanced algorithm using Newton iteration

Like all steepest-ascent algorithms, the convergence order of the step-project algorithm is unity. Application of Newton's algorithm would increase the convergence order to two, but could destroy the global stability property – Newton's algorithm is guaranteed to converge to a solution σ_{opt} only when started 'close enough' to σ_{opt} . To increase the speed of convergence without sacrificing stability, we employ a hybrid algorithm: the step-project algorithm is run for a fixed number of iterations (say 5–10) to get 'close' to a solution, then Newton's algorithm is applied to 'zero in' quickly.

Newton's algorithm is usually derived for unconstrained optimization problems. To incorporate the constraint⁵ (35), we introduce a Lagrange multiplier λ , define the Lagrangian as

$$l(\sigma, \lambda) = f(\sigma) + \lambda (\|\sigma\|^2 - \gamma^2), \quad (44)$$

and solve

$$\max_{\sigma, \lambda} l(\sigma, \lambda). \quad (45)$$

⁵ The volume constraint can be assumed to be an equality constraint, since it is easily shown that the solution to (34)–(35) lies on the surface of $\mathcal{F}$.

Defining

$$h(\sigma) = \|\sigma\|^2 - \gamma^2, \quad (46)$$

Newton's algorithm solves (45) by recursively solving a linearized version of the first-order necessary conditions for a maximum [20, p. 431],

$$\nabla l(\sigma, \lambda) = 0, \quad h(\sigma) = 0. \quad (47)$$

Letting $\sigma(k)$ and $\lambda(k)$ represent estimates of the solution to (47) at iteration k , we obtain new estimates using

$$\begin{aligned} \sigma(k+1) &= \sigma(k) + d(k), \\ \lambda(k+1) &= \lambda(k) + y(k). \end{aligned} \quad (48)$$

The steps $d(k)$ and $y(k)$ are computed by solving the following matrix equation, obtained by applying Taylor's theorem to (47):

$$\begin{bmatrix} L(k) & \nabla h(\sigma(k)) \\ \nabla h(\sigma(k))^T & 0 \end{bmatrix} \begin{bmatrix} d(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} -\nabla l(\sigma(k), \lambda(k)) \\ -h(\sigma(k)) \end{bmatrix}. \quad (49)$$

Here $L(k)$ is the Hessian matrix of second partial derivatives of the Lagrangian (44) at iteration k , and is given by

$$L(k) = F(k) + 2\lambda(k)I, \quad (50)$$

with $F(k)$ being the Hessian of the performance index (36) and I being the identity matrix. The elements $F_{ij}(k)$ of $F(k)$ are obtained by differentiating (42),

$$\begin{aligned} F_{ij}(k) &= \frac{\partial^2 f}{\partial \sigma_i(k) \partial \sigma_j(k)} \\ &= \begin{cases} \frac{2\Delta_r^2}{\sigma_i^4(k)} \sum_{p=1}^{P-1} \left(2 \left(\frac{p\Delta_r}{\sigma_i(k)} \right)^2 - 3 \right) \\ \quad \times p^3 |A_p(p, q)|^2 e^{-(p\Delta_r)^2 / \sigma_i^2(k)}, & i=j, \\ 0, & i \neq j. \end{cases} \end{aligned} \quad (51)$$

The solution to the matrix equation (49) is well defined when the matrix on the left-hand side is negative definite. This is the case when the estimates $\sigma(k)$ and $\lambda(k)$ are 'close enough' to a solution.

A potential drawback of Newton's algorithm is high computational cost, since, in general, each iteration requires the inversion of $L(k)$. Fortunately, from (50) and (51) it is easily seen that $L(k)$ is diagonal, and hence trivial to invert. Therefore, the overall cost of the combined algorithm employing both step-project and Newton's algorithm iterations remains $O(PQi)$ (typically, i lies between 5 and 30). Further details of the Newton's algorithm solution to this problem are given in [4].

4. Examples

In order to compare the optimal-kernel TFD with the fixed-kernel TFDs discussed in the Introduction, the optimal radially Gaussian kernel was computed using the techniques of the previous section for the pulse and chirp signals of Fig. 1. Sixty-four samples of each signal were analyzed; the volume parameter in each case was set to $\alpha = 1.4$. The iterative kernel solution algorithm converged in 11 iterations for both signals. Equal-energy contour plots of the optimal kernels are given in Figs. 8(a) and 9(a); the optimal-kernel

TFDs are shown in Figs. 8(b) and 9(b). The cross-components visible in the fixed-kernel TFDs of Figs. 2-5 are virtually eliminated, yet the auto-components are still quite concentrated. The main difference between the present results and the tapered 1/0 optimal-kernel TFDs of Fig. 6 is the absence of ringing.

Lowpass kernels have proven useful for reducing the high noise sensitivity of some bilinear TFDs [8]. The performance of the optimal kernel design procedure under noisy conditions was tested by embedding a signal composed of the sum of two chirp signals that cross in time-frequency in 3 dB SNR (measured as the ratio of the total signal power to the total noise power) additive white Gaussian noise. Sixty-four signal samples were analyzed; the kernel volume parameter was set to $\alpha = 2$. Convergence of the iterative algorithm was reached in 16 iterations. Figures 10(a), (b), (c) and (d) illustrate the Wigner, Choi-Williams and cone-kernel distributions and the matched-filter spectrogram (obtained using the noise-free signal as a window) of the noisy signal. The cross-component and noise suppression of the optimal-kernel TFD shown in Fig. 10(e) are excellent, indicating that the kernel design procedure is robust in the presence of significant additive noise. When comparing the optimal-kernel TFD with the matched-filter spectrogram, recall that (unlike the matched-filter spectrogram) the optimal kernel and TFD are computed using no a priori information regarding

Fig. 8. (a) Optimal radially Gaussian kernel of volume $\alpha = 1.4$ computed for the pulse signal of Fig. 1(a). (b) Optimal radially Gaussian kernel TFD.

Fig. 9. (a) Optimal radially Gaussian kernel of volume $\alpha = 1.4$ computed for the chirp signal of Fig. 1(b). (b) Optimal radially Gaussian kernel TFD.

the signal. Figure 10(f) shows the optimal kernel for this signal.

In Fig. 11 we have analyzed 2.5 msec of an echolocation pulse emitted by the large brown bat, *Eptesicus fuscus*. The signal was sampled every 7 μ sec, yielding approximately 350 data samples. While the Wigner distribution of Fig. 11(a) is highly concentrated, it has large cross-components. A spectrogram analysis conducted using a Gaussian window with an e^{-1} time width of 160 μ sec yields a smeared TFD (see Fig. 11(b)). The cone-kernel distribution of Fig. 11(c) is similar. In the optimal radially Gaussian kernel TFD of Fig. 11(d), the fundamental and first two harmonics are sharply resolved, and the cross-components have been suppressed. Using a kernel volume of $\alpha = 4$, the kernel solution algorithm converged in 10 steps.

Several other examples of radially Gaussian kernel TFDs are given in [14].

5. Customizing the optimization formulation

The optimization formulations of Sections 2 and 3 are general tools for signal-dependent kernel design. In certain applications, properties in addition to cross-component suppression are required of the kernel. An attractive feature of the optimal kernel design procedure is the ease with which it can be customized to incorporate application-specific knowledge. A priori information or

requirements are included either by adding constraints, post-processing the optimal kernel, or modifying the performance measure.

5.1. Marginal distributions

A TFD satisfies the time and frequency marginal distributions if its kernel takes the value 1 along the θ and τ axes [6]. If the following constraints are added to the nonlinear program (34)–(35):

$$\Phi_p(p, 0) = \Phi_p\left(p, \frac{Q}{2}\right) = 1 \quad \forall p, \quad (52)$$

the resulting signal-dependent kernel will be optimal with respect to the more restrictive class of radially Gaussian kernels whose TFDs satisfy the marginals. However, (35) must be modified in this case, since kernels satisfying (52) require $\sigma_0 = \sigma_{Q/2} = \infty$. Redefining the kernel volume constraint as

$$\frac{\Delta_\psi}{2\pi} \sum_{q \neq 0, Q/2} \sigma_q^2 \leq \alpha, \quad \alpha \geq 0, \quad (53)$$

the nonlinear program (34), (52)–(53) can be solved using the techniques discussed in Section 3.

The optimal marginal-satisfying kernel and TFD computed for the chirp signal discussed in the Introduction are shown in Fig. 12. Compared to

Fig. 10. Time-frequency distributions of a signal – two chirps that cross in time-frequency – corrupted by additive white Gaussian noise. The SNR, measured as the ratio of the total signal power to the total noise power, is 3 dB. (a) Wigner distribution. (b) Choi-Williams distribution ($\sigma = 5$). (c) Cone-kernel distribution ($a = 2$, $\alpha = 0.009$ and $\beta = 16$). (d) Matched-filter spectrogram. (e) Optimal radially Gaussian kernel TFD. (f) Optimal radially Gaussian kernel.

Fig. 3(c), this customized optimal-kernel TFD is an attractive alternative to the Choi-Williams distribution in applications where the marginals must be satisfied.

5.2. Time or frequency support

A TFD preserves the convex hull of the time support of a signal if its kernel satisfies a ‘cone constraint’ [24]

Signal Processing

$$\int_{-\infty}^{\infty} \Phi(\theta, \tau) e^{-j\theta t} d\theta = 0 \quad \forall |t| > 2|\tau|. \quad (54)$$

A similar constraint preserves embedded intervals of zero signal energy. To construct a signal-dependent kernel that preserves time support, post-processing of the optimal radially Gaussian kernel is useful. First, we compute the optimal kernel Φ_{opt} for the signal, then find another kernel $\tilde{\Phi}$ that both

Fig. 11. Time-frequency distributions of 2.5 msec of an echolocation pulse emitted by the large brown bat, *Eptesicus fuscus*. (a) Equal-energy contour plot of the Wigner distribution. (b) Spectrogram computed using a Gaussian window with an e^{-1} time width of 160 μ sec. (c) Cone-kernel distribution ($a=2$, $\alpha=\beta=5 \times 10^{-4}$). (d) TFD computed from the optimal radially Gaussian kernel of volume $\alpha=4$. The fundamental and first two harmonics of the chirp are sharply resolved in this signal-dependent TFD.

preserves time support and is 'close' in some sense to Φ_{opt} . If a least-squares distance criterion is used, $\tilde{\Phi}$ is given by the orthogonal projection of Φ_{opt} onto the set of support-preserving kernels [22]. In continuous variables, we have

$$\begin{aligned} \tilde{\Phi}(\theta, \tau) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \Phi_{\text{opt}}(\nu, \tau) e^{-j\nu t} d\nu \text{rect}_{\tau}(t) \right] \\ &\quad \times e^{j\theta t} dt, \end{aligned} \quad (55)$$

where

$$\text{rect}_y(x) = \begin{cases} 1, & |x| \leq y, \\ 0, & \text{otherwise.} \end{cases} \quad (56)$$

In practice, $\tilde{\Phi}_d(m, n)$ is obtained by taking one-dimensional FFTs of the optimal kernel $\Phi_d^{\text{opt}}(m, n)$ along the m coordinate, zeroing the result outside of a cone, and then taking inverse FFTs.

While the post-processed kernel is no longer optimal with respect to the original optimization

Fig. 12. (a) The optimal marginal-satisfying kernel and (b) TFD computed for the chirp signal of Fig. 1(b). The kernel is similar to that of Fig. 9(a), except along the θ and τ axes, where it takes the value 1.

formulation (in fact, it is no longer radially Gaussian), its passband retains the same general shape and orientation as the optimal kernel. The post-processed kernel and TFD found using this technique for the chirp signal discussed in the Introduction are shown in Fig. 13. Compared to Fig. 4(c), the customized optimal-kernel TFD can offer a considerable performance improvement over the cone-kernel distribution in applications where the time support of the signal must be preserved. Similar methods can be applied to create kernels that preserve the frequency support of the signal [22].

5.3. Generalized performance measure

Consider a performance measure that is a generalized version of (12),

$$\max_{\Phi} \int_0^{2\pi} \int_0^{\infty} f(|A(r, \psi)|^2) |\Phi(r, \psi)|^2 r \, dr \, d\psi, \quad (57)$$

where $f: \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Given any such transformation f , the optimization problem (57), (13)–(15) can be solved using the techniques discussed in Section 3, but with the transformed data

$$|\hat{A}(r, \psi)|^2 = f(|A(r, \psi)|^2). \quad (58)$$

(Note that the optimal-kernel TFD is computed as the Fourier transform of $A(\theta, \tau)\Phi(\theta, \tau)$ regardless of the transformation f .)

A transformation that is useful for constraining the locations of the passband and stopband of the optimal kernel is

$$f(|A(r, \psi)|^2) = \Gamma(r, \psi) |A(r, \psi)|^2, \quad (59)$$

Fig. 13. (a) The optimal time support-preserving kernel and (b) TFD computed for the signal of Fig. 1(b). The postprocessing changes the shape of the optimal kernel (compare to Fig. 9(a)), but not its orientation angle.

where $\Gamma(r, \psi)$ is any real, non-negative, symmetric (as in (16)) weighting function. As an example, consider the use of $\Gamma(r, \psi)$ to constrain the passband of the optimal kernel. If the auto-components of interest in a signal are known to lie in a region $\mathcal{R}$ of the ambiguity plane, then setting

$$\Gamma(r, \psi) = \begin{cases} 1, & (r, \psi) \in \mathcal{R}, \\ 0, & \text{otherwise} \end{cases} \quad (60)$$

discourages the optimal kernel passband from lying outside of $\mathcal{R}$. Thus, auto- and cross-components lying outside of $\mathcal{R}$ will be suppressed. This passband constraint is useful, for example, for analyzing transient signals in the presence of sinusoidal interference signals. In this case, some of the interference can be rejected by setting $\Gamma(r, \psi) = 1$ everywhere except in a region along the τ axis ($\psi = \pi/2$), where AF auto-components corresponding to sinusoids lie. In practice, $\Gamma(r, \psi)$ should be chosen to be reasonably smooth. Otherwise, an unsmooth optimal kernel could result, leading to ringing in the optimal-kernel TFD.

As another example of a useful transformation, consider a signal in which a component of interest is of low energy relative to other signal components. Since the function $|A(r, \psi)|^2$ involves fourth powers of the signal, the amplitude of the AF auto-component of interest may be so small compared to other components that the optimal kernel is not attracted to the region where it resides. In this case, a nonlinear transformation is useful for reducing the dynamic range of the ambiguity surface. Any continuous function growing more slowly than linearly is appropriate, such as $f(x) = \log(x+1)$ or $f(x) = \sqrt{x}$. While dynamic range reduction of the AF has the undesirable effect of amplifying any noise present in the signal, as demonstrated in Section 4, the optimal kernel design procedure appears quite robust at low signal-to-noise ratios.

6. Conclusion

This paper has presented a technique for designing a signal-dependent time-frequency representation. The signal-dependent approach is

motivated by the fact that a fixed kernel limits the class of signals for which its corresponding TFD performs well; that is, only by using a kernel that adapts to each signal can a TFD analyze a large class of signals effectively. The simple examples we have presented demonstrate both the necessity of signal dependence as well as the excellent performance of the proposed design procedure.

The design procedure possesses several additional attractive properties. First, it is based on quantitative optimality criteria. Second, it is automatic and relies neither on a priori knowledge about the signal nor good fortune to yield a high-quality time-frequency representation. Third, the optimization criteria are formulated so that the optimal-kernel TFD is insensitive to the time-scale and orientation of the signal in time-frequency. Finally, performance remains excellent even in the presence of substantial additive noise, which suggests that the technique may prove useful for the automatic detection of signals in noise.

The computational cost of finding the optimal-kernel TFD is $O(L^2 i + L^2 \log L)$, with L being the number of signal samples and i the number of iterations required by the kernel solution algorithm. In experiments, the number of iterations required has typically been less than 30, so the cost of this technique is only a few times greater than both fixed-kernel methods and the 1/0 optimal-kernel method proposed in [2, 5].

The proposed optimization formulation (12)–(15) is reminiscent of the 1/0 optimal-kernel method, with the additional constraint that the kernel be a radially Gaussian function. This new constraint results in an optimal-kernel TFD with less ringing artifacts, and in most cases we prefer the appearance of the new TFD. However, the inherent smoothness of the radially Gaussian kernel could result in less cross-component suppression for signal components that are closely spaced in time-frequency.

This paper by no means exhausts the possibilities for signal-dependent time-frequency analysis. As demonstrated, additional constraints based on a priori or application-specific knowledge are often

easily incorporated into the formulation discussed here. Furthermore, totally new optimization formulations may be developed by utilizing other constraints, other performance measures, or classes of TFDs other than the bilinear. While the choice of class and performance measure is critical to success, once a satisfactory class and measure are found, the design of a signal-dependent time-frequency representation reduces to solving an optimization problem. Given the generality of the approach and the promise shown by the specific methods described in this paper, it seems likely that adaptive time-frequency representations will emerge as powerful tools for time-varying spectral analysis.

Appendix A. Scale-invariance of the optimal kernel

This appendix is devoted to proving the following fundamental result.

THEOREM 1. *The optimal radially Gaussian kernel is covariant to time-scale changes in the signal. That is, if the optimal kernel for a signal $s(t)$ is $\Phi_{\text{opt}}(\theta, \tau)$, then the optimal kernel for the time-scaled signal $a^{-1/2}s(at)$, $a \neq 0$, is $\Phi_{\text{opt}}(\theta/a, a\tau)$.*

PROOF. To prove the result, we show that $\Phi_{\text{opt}}(\theta/a, a\tau)$ both satisfies the constraints (13)–(14) and maximizes the performance measure (12) for the signal $a^{-1/2}s(at)$ and its AF. Note that we use rectangular, rather than polar, coordinates in the proof.

We first show that the kernel $\Phi_{\text{opt}}(\theta/a, a\tau)$ is radially Gaussian by evaluating it along a radial line lying at an arbitrary angle ψ_0 in the (θ, τ) ambiguity plane. Denoting this line by Ψ_0 , we have

$$\Phi_{\text{opt}}(\theta/a, a\tau)|_{\Psi_0} = e^{-(\theta^2/a^2 + a^2\tau^2)/2\sigma^2(\psi_0)}. \quad (61)$$

Using the fact that θ and τ are related to the radius r along Ψ_0 from the origin by

$$\theta = r \cos \psi_0, \quad \tau = r \sin \psi_0 \quad (\text{along } \Psi_0) \quad (62)$$

yields

$$\begin{aligned} \Phi_{\text{opt}}(\theta/a, a\tau)|_{\Psi_0} \\ = e^{-r^2[(\cos \psi_0/a)^2 + (a \sin \psi_0)^2]/2\sigma^2(\psi_0)}, \end{aligned} \quad (63)$$

which shows that $\Phi_{\text{opt}}(\theta/a, a\tau)$ is Gaussian along an arbitrary radial slice in the (θ, τ) plane.

Next, the simple change of integration variables $\theta' = \theta/a$, $\tau' = a\tau$ shows that $\Phi_{\text{opt}}(\theta/a, a\tau)$ satisfies the volume constraint (14):

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Phi_{\text{opt}}(\theta/a, a\tau)|^2 d\theta d\tau \\ = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Phi_{\text{opt}}(\theta', \tau')|^2 d\theta' d\tau' \leq \alpha. \end{aligned} \quad (64)$$

Finally, we show that $\Phi_{\text{opt}}(\theta/a, a\tau)$ maximizes the value of the performance measure (12) for the signal $a^{1/2}s(at)$. An important fact, which is easily proved by direct substitution into the definition of the AF (2), is that if the AF for the signal $s(t)$ is $A(\theta, \tau)$, then the AF for the scaled signal $a^{1/2}s(at)$, $a \neq 0$, is $A(\theta/a, a\tau)$. Proceeding with the proof, suppose to the contrary that $\Phi_{\text{opt}}(\theta/a, a\tau)$ does not maximize the value of the performance measure for the signal $a^{1/2}s(at)$; that is, suppose that there exists a kernel $\hat{\Phi}(\theta, \tau)$ such that

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(\theta/a, a\tau) \hat{\Phi}(\theta, \tau)|^2 d\theta d\tau \\ > \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(\theta/a, a\tau) \Phi_{\text{opt}}(\theta/a, a\tau)|^2 d\theta d\tau. \end{aligned} \quad (65)$$

Then, performing the same change of integration variables in the above as in (64) suggests that there exists a kernel $\hat{\Phi}(a\theta', \tau'/a)$ such that

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(\theta', \tau') \hat{\Phi}(a\theta', \tau'/a)|^2 d\theta' d\tau' \\ > \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(\theta', \tau') \Phi_{\text{opt}}(\theta', \tau')|^2 d\theta' d\tau'. \end{aligned} \quad (66)$$

Since this contradicts the optimality of the original kernel $\Phi_{\text{opt}}(\theta, \tau)$ for the signal $s(t)$, $\Phi_{\text{opt}}(\theta/a, a\tau)$ must maximize the value of the performance measure for the signal $a^{1/2}s(at)$.

Therefore, since $\Phi_{\text{opt}}(\theta/a, a\tau)$ both satisfies the constraints and maximizes the value of the performance measure, it is the optimal kernel for the signal $a^{1/2}s(at)$. $\square$

Acknowledgments

The authors wish to thank Prof. Lee Potter for the optimization insight and Curtis Condon, Ken White and Prof. Ai Feng for the bat data and the permission to use it in this paper. This work was supported by the Sound Group of the Computer-based Education Research Laboratory at the University of Illinois, the Joint Services Electronics Program, Grant No. N00014-90-J-1270, and the National Science Foundation, Grant No. MIP 90-12747.

References

- [1] J.C. Andrieux, M.R. Felix, G. Mourgues, P. Bertrand, B. Izrar and V.T. Nguyen, "Optimum smoothing of the Wigner-Ville distribution", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 35, No. 6, June 1987, pp. 764-769.
- [2] R.G. Baraniuk and D.L. Jones, "Optimal kernels for time-frequency analysis", in: F.T. Luk, ed., *Advanced Signal Processing Algorithms, Architectures and Implementations*, Proc. SPIE 1348, 1990, pp. 181-187.
- [3] R.G. Baraniuk and D.L. Jones, "A radially Gaussian, signal-dependent time-frequency representation", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Toronto, 1991, pp. 3181-3184.
- [4] R.G. Baraniuk, Shear madness: Signal-dependent and metaplectic time frequency representations, Technical Report No. UILU-ENG-92-2226, Coordinated Science Laboratory, University of Illinois, Urbana, 1992.
- [5] R.G. Baraniuk and D.L. Jones, "A signal-dependent time-frequency representation: Optimal kernel design", *IEEE Trans. Signal Process.*, Vol. 41, No. 4, April 1993.
- [6] H.-I. Choi and W.J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 37, No. 6, June 1989, pp. 862-871.
- [7] T.A.C.M. Claassen and W.F.G. Mecklenbräuker, "The Wigner distribution - A tool for time-frequency signal analysis; Part III: Relations with other time frequency signal transformations", *Phillips J. Res.*, Vol. 35, 1980, pp. 372-389.
- [8] L. Cohen, "On a fundamental property of the Wigner distribution", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 35, 1987, pp. 559-561.
- [9] L. Cohen, "Time-frequency distributions - A review", *Proc. IEEE*, Vol. 77, No. 7, July 1989, pp. 941-981.
- [10] P. Flandrin, "Some features of time-frequency representations of multicomponent signals", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, 1984, pp. 41.B.4.1-41.B.4.4.
- [11] P. Flandrin and O. Rioul, "Affine smoothing of the Wigner-Ville distribution", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Albuquerque, 1990, pp. 2455-2458.
- [12] S.C. Glinski, Diphone speech synthesis based on a pitch-adaptive short-time Fourier transform, Ph.D. Thesis, University of Illinois, Urbana, 1981.
- [13] F. Hlawatsch, "Interference terms in the Wigner distribution", *Proc. Internat. Conf. Digital Signal Processing*, Florence, Italy, 1984.
- [14] F. Hlawatsch and G.F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations", *IEEE Signal Process. Mag.*, Vol. 9, No. 2, April 1992, pp. 21-67.
- [15] G. Jones and B. Boashash, "Instantaneous frequency, instantaneous bandwidth and the analysis of multicomponent signals", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, 1990, pp. 2467-2470.
- [16] D.L. Jones and T.W. Parks, "A high resolution data-adaptive time-frequency representation", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Dallas, April 1987, pp. 681-684.
- [17] D.L. Jones and T.W. Parks, "A high resolution data-adaptive time-frequency representation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 38, No. 12, December 1990, pp. 2127-2135.
- [18] D.L. Jones and T.W. Parks, "A resolution comparison of several time-frequency representations", *IEEE Trans. Signal Process.*, Vol. 40, No. 2, February 1992.
- [19] S. Kadambe, G.F. Boudreaux-Bartels and P. Duvaut, "Window length selection for smoothing the Wigner distribution by applying an adaptive filter technique", *IEEE Internat. Conf. Acoust. Speech Signal Process.*, 1989, pp. 2226-2229.
- [20] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1973.
- [21] A.H. Nuttall, "Wigner distribution function: Relation to short-term spectral estimation, smoothing, and performance in noise", *NUSC Technical Report 8225*, 16 February 1988.
- [22] S. Oh, R.J. Marks and J.W. Pitton, "Kernel synthesis for generalized time-frequency distributions using the method of projection onto convex sets", in: F.T. Luk, ed., *Advanced Signal Processing Algorithms, Architectures and Implementations*, Proc. SPIE 1348, 1990, pp. 197-207.

- [23] J.B. Rosen, "The gradient projection method for nonlinear programming, Part II, Nonlinear constraints", *J. Soc. Indust. Appl. Math.*, Vol. 9, No. 4, December 1961, pp. 514-532.
- [24] Y. Zhao, L.E. Atlas and R.J. Marks, "The use of cone-shape kernels for generalized time-frequency representation of nonstationary signals", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 38, July 1990, pp. 1084-1091.