# A Signal-Dependent Time-Frequency Representation: Optimal Kernel Design

Richard G. Baraniuk and Douglas L. Jones, *Member, IEEE*

*Abstract*—**Time-frequency distributions (TFD's), which indicate the energy content of a signal as a function of both time and frequency, are powerful tools for time-varying signal analysis. The lack of a single distribution that is "best" for all applications has resulted in a proliferation of TFD's, each corresponding to a different, fixed mapping from signals to the time-frequency plane. A major drawback of all fixed mappings is that, for each mapping, the resulting time-frequency representation is satisfactory only for a limited class of signals. In this paper, we introduce a new TFD that adapts to each signal and so offers good performance for a large class of signals. The design of the signal-dependent TFD is formulated in Cohen's class as an optimization problem and results in a special linear program. Given a signal to be analyzed, the solution to the linear program yields the optimal kernel and, hence, the optimal time-frequency mapping for that signal. A fast algorithm has been developed for solving the linear program, allowing the computation of the signal-dependent TFD with a time complexity on the same order as a fixed-kernel distribution. Besides this computational efficiency, an attractive feature of the optimization-based approach is the ease with which the formulation can be customized to incorporate application-specific knowledge into the design process.**

## I. Introduction

TIME-frequency distributions (TFD's) are two-dimensional functions that indicate the joint time-frequency energy content of a signal. They have been utilized to study a wide range of signals, including speech, music, and other acoustical signals, biological signals, radar and sonar signals, and geophysical signals. Most TFD's of current interest are members of Cohen's bilinear (quadratic) class [1]–[3]. Examples include traditional favorites such as the Wigner distribution and the spectrogram (the squared magnitude of the short-time Fourier transform), as well as more recent introductions like the Choi–Williams distribution [4] and the cone-kernel distribution [5]. Each distribution in Cohen's class can be interpreted as the two-dimensional Fourier transform of a weighted version of the symmetric ambiguity function (AF) of the signal to be analyzed. That is, if $P(t, \omega)$ is a bilinear TFD, then

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

where

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

The weighting function $\Phi(\theta, \tau)$ is called the kernel of the distribution. The properties of a particular bilinear TFD are completely determined by its kernel function, via (1).

Since the AF is a bilinear function of the signal, it exhibits cross-components [6], which, if allowed to pass into the TFD, can reduce autocomponent resolution, obscure the true signal features, and make interpretation of the distribution difficult. Therefore, the kernel is often selected to weight the AF such that the autocomponents, which are centered at the origin of the $(\theta, \tau)$ ambiguity plane, are passed, while the cross-components, which are located away from the origin, are suppressed [7], [8]. That is, in order to suppress cross-components, $\Phi(\theta, \tau)$ should be the frequency response of a two-dimensional low-pass filter. When a low-pass kernel is employed, there is a tradeoff between cross-component suppression and autocomponent concentration: generally, as the passband region of the kernel is made smaller, the amount of cross-component suppression increases, but at the expense of autocomponent concentration.

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

Traditionally, TFD's have used fixed kernels, with fixed passband and stopband regions. However, specification of a fixed kernel limits the class of signals for which the corresponding time-frequency representation performs well. That is, given any fixed kernel, it is always possible to find signals with either significant cross-com-

Manuscript received December 24, 1991; revised June 15, 1992. This work was supported by the Sound Group of the Computer-Based Education Research Laboratory and the Joint Services Electronics Program under Grant N00014-90-J-1270.

R. G. Baraniuk is with the Laboratoire de Traitement du Signal, Ecole Normale Supérieure de Lyon, Lyon, France, on leave from the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892.

D. L. Jones is with the Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801.

IEEE Log Number 9206879.

ponent energy in the passband of the kernel or significant autocomponent energy in the stopband of the kernel.[1] By studying the shape of a kernel, we can predict both the class of signals for which the corresponding TFD performs well and the class of signals for which the TFD performs poorly.

To illustrate the limitations of fixed-kernel TFD's, consider two example signals. The first is the sum of two short Gaussian pulses:

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

with $a = 0.25$ and $T_0 = 5$. The second is the sum of two linear-FM, "chirp" signals:

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

with $a = 0.004$, $c = 0.025$, $T_0 = 3.5$, $\omega_1 = 1.72$, and $\omega_2 = 1.42$. 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 AF's of these signals are given in Figs. 1(c) and (d). Note the locations of the auto-components and cross-components.

The Wigner distribution has a kernel that performs no filtering of the AF,

$$\Phi_W(\theta, \tau) = 1 \tag{5}$$

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

The Choi–Williams distribution has an exponential kernel

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

that is large along both the $\theta$ and $\tau$ axes (see Fig. 3(a)). It performs well for signals whose AF autocomponents 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 (shown in Fig. 3(b)), the cross-components are suppressed at the expense of some smearing and shouldering of the autocomponents. The Choi–Williams distribution performs poorly, however, for signals whose AF autocomponents do not lie along the $\theta$ and $\tau$ axes. For example, signals with substantial frequency modulation are poorly represented (see Fig. 3(c)), because the kernel severely truncates the autocomponents of such signals.

The kernel generating the cone-kernel distribution is given by

$$\Phi_C(\theta, \tau) = \begin{cases} \dfrac{\sin(\theta|\tau|a^{-1})}{\theta} e^{-2\alpha\tau^2}, & |\tau| \le \beta,\ a \ge 2 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

[1]We note that the same limitation exists for the class of fixed-kernel, bilinear time-scale distributions, which includes the squared magnitude of the continuous wavelet transform [9].

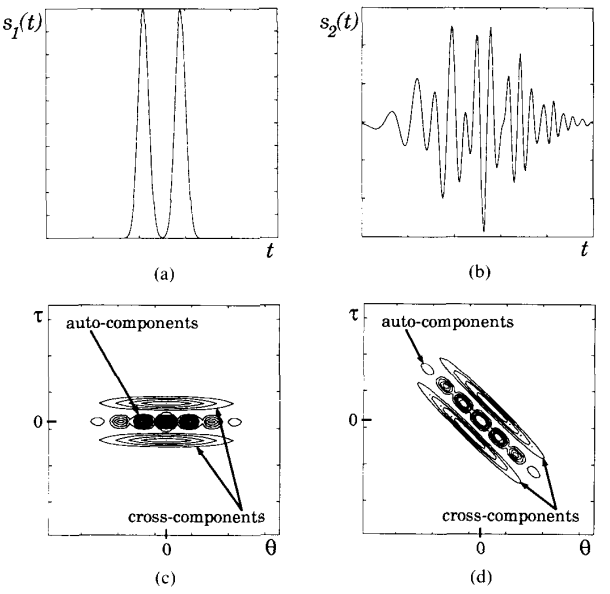



Fig. 1. Example signals for the TFD's of Figs. 2–5. (a) The signal given in (3), consisting of two short Gaussian pulses. (b) The real part of the signal given in (4), consisting of two parallel Gaussian chirps. (c) Equal-energy contour plot of the magnitude of the ambiguity function (AF) of the pulse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

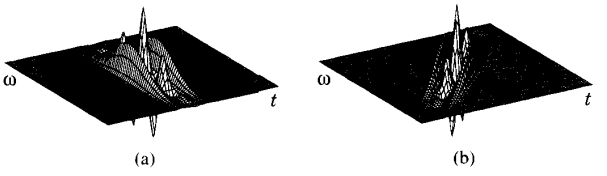



Fig. 2. Wigner distributions of (a) the pulse and (b) the chirp signals of Fig. 1. The Wigner kernel is $\Phi_W(\theta, \tau) = 1$, so the distribution is simply the two-dimensional Fourier transform of the AF. Note the large cross-components that appear midway between the autocomponents.

and is shown in Fig. 4(a). The time-frequency representation of the cone-kernel distribution is excellent for pulse-like signals whose AF autocomponents lie near the $\theta$ axis (see Fig. 4(b)), but 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 [10]

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

Results are excellent for signal components that resemble the window [7], but all mismatched components are distorted. Fig. 5 shows the spectrogram kernel and spectrograms 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 kernel, is matched to one of the signal components, the *matched-filter spectrogram* results. As Fig. 5(b) shows, the matched-filter technique can yield excellent

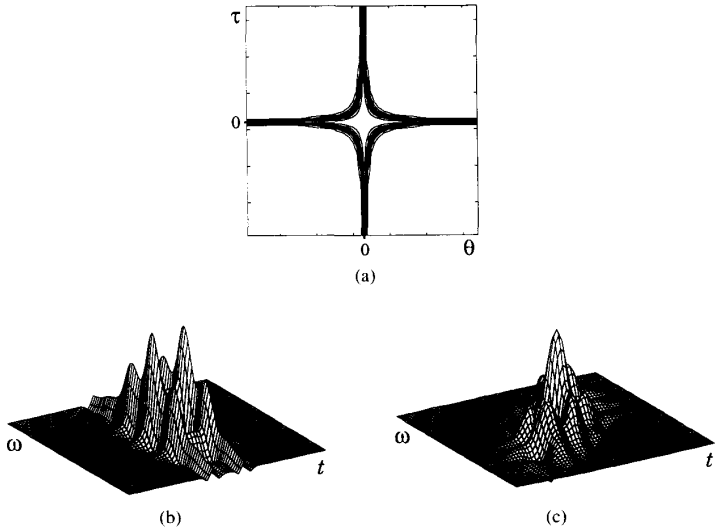



Fig. 3. (a) Equal-energy contour plot of the Choi–Williams kernel for $\sigma = 10$. (b) Choi–Williams distribution of the pulse signal of Fig. 1(a). Since the AF autocomponents of this signal lie along the $\theta$ axis and the cross-components lie off the $\theta$ axis (see Fig. 1(c)), the autocomponents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

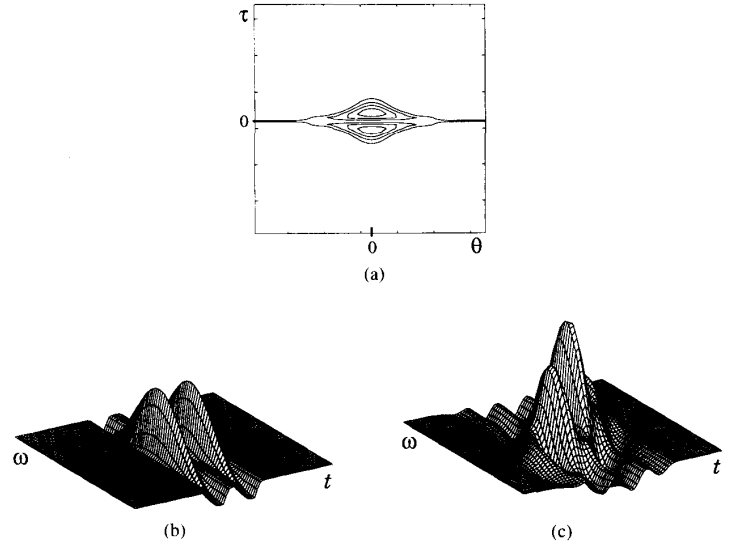



Fig. 4. (a) The cone kernel for $a = 2$, $\alpha = 0.036$, and $\beta = 8$. (b) Cone-kernel distribution of the pulse signal of Fig. 1(a). Since the AF autocomponents line up with the $\theta$ axis, an excellent representation results. (c) Cone-kernel distribution of the chirp signal of Fig. 1(b). Since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

results. However, the technique works for only one type of signal component and, in addition, 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).

Although the two example signals differ only in their orientation in time-frequency (compare Figs. 2(a) and (b)), none of the aforementioned TFD's perform adequately for both. This is indicative of a fundamental limitation of TFD's employing fixed kernels:

*A fixed kernel results in good performance only for certain configurations of AF autocomponents and cross-components, and thus only for a limited class of signals.*

Since the locations of the autocomponents 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 provides a good time-frequency representation by adjusting its shape in order to pass autocomponents and suppress cross-components, regardless of their location and orientation.

Signal-dependent kernels have been proposed by sev-

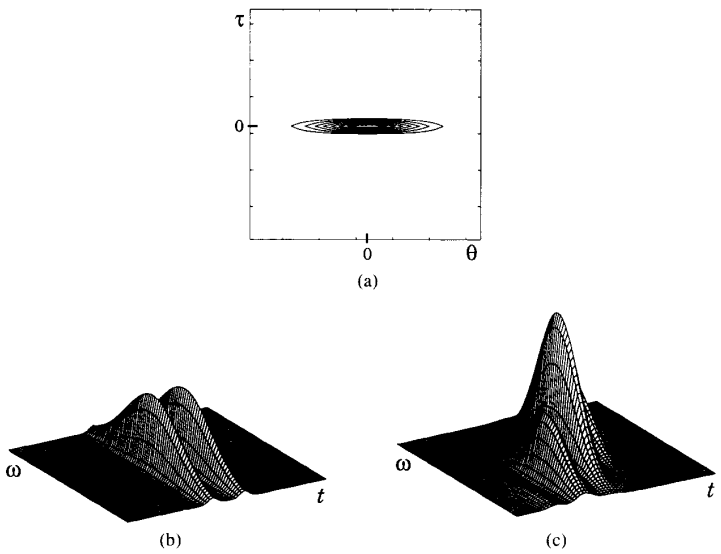



Fig. 5. (a) The kernel generating the spectrogram (the squared magnitude of the short-time Fourier transform) when the analysis window is matched to the signal components of Fig. 1(a). (b) (Matched-filter) spectrogram of the pulse signal of Fig. 1(a). (c) Spectrogram of the chirp signal of Fig. 1(b)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

eral authors. 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 [11]. Nuttall designs a kernel composed of Gaussian components based on information that the user provides after viewing the Wigner distribution [12]. Jones and Boashash adapt the modulation rate of a fixed window to match an estimate of the signal's instantaneous frequency [13]. In [14], Cohen designs a signal-dependent kernel that totally concentrates the distribution of a constant amplitude monocomponent signal along its instantaneous frequency. Kadambe *et al.* utilize an adaptive filtering technique coupled with AR modeling and clustering to design kernels [15]. Optimal smoothing kernels are considered by Andrieux *et al.*, but only for simple signals of the form $s(t) = e^{j\nu(t)}$, and only for the restrictive class of Gaussian kernels [16]. Jones and Parks develop a technique using Gaussian kernels whose shape and orientation vary with time and frequency to maximize a local measure of signal-energy concentration [17], [18]. It often works well but is computationally expensive. An approach similar to that in [13] is adopted by Mihovilovic and Bracewell, who derive an iterative algorithm for estimating the aspect ratio and orientation angle of a time-varying Gaussian kernel [19]. Recently, Loughlin *et al.* [20] have devised a technique for selecting a nonbilinear distribution from the positive class [21] that resembles a spectrogram, yet satisfies the marginal distributions.

These efforts constitute the present state-of-the-art of signal-dependent time-frequency analysis. While meritorious for demonstrating the potential of signal dependency, most of these approaches suffer 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 on the same order as 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. Section III concerns the solution of the design problem; we will see that the optimization formulation results in a linear program with a simple solution. Examples that demonstrate the effectiveness of the signal-dependent approach are given in Section IV. In Section V 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. A companion paper [22] focuses on a fast algorithm to solve the linear program that defines the optimal kernel.

## II. Optimal Kernel Design

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 [23]. 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 index is selected as the optimal kernel for the signal.

The class of kernels will be specified by a set of con-

straints, and must be large enough to allow for good performance for all signals of interest in a given application. Constraints applied in current fixed-kernel distributions include, for example, constraints that force the kernel to suppress cross-components [7], satisfy the time and frequency marginal distributions [4], preserve the time or frequency support of the signal [5], or satisfy Moyal's formula [2]. The performance measure must be chosen to yield a tractable optimization problem that can be solved efficiently. An example of a useful performance index is a measure of the signal-energy concentration of the distribution [17], [18].

### A. Continuous-Time Optimization Formulation

The optimization formulation we propose for bilinear TFD kernel design relies on the AF of the signal and the characterization of the TFD indicated by (1). We propose optimality criteria based on the AF for three reasons. First, the multiplicative operation of the kernel on the AF is conceptually and mathematically simple, which simplifies the construction of a quality measure. Second, as discussed in the Introduction, the AF serves to separate the autocomponents and cross-components. Third, the AF leads to efficient computation of the signal-dependent TFD, via (1). We will refer to the TFD obtained using the optimal, signal-dependent kernel as the optimal-kernel (OK) distribution $P_{\text{opt}}(t, \omega)$.

Given a signal and its AF, $A(\theta, \tau)$, we define the optimal kernel as the real, nonnegative function[2] $\Phi_{\text{opt}}(\theta, \tau)$ that solves the following optimization problem [23]:

$$\max_{\Phi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(\theta, \tau)\Phi(\theta, \tau)|^2 \, d\theta \, d\tau \tag{9}$$

subject to

$$\Phi(0, 0) = 1 \tag{10}$$

$$\Phi(\theta, \tau) \quad \text{is radially nonincreasing} \tag{11}$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Phi(\theta, \tau)|^2 \, d\theta \, d\tau \leq \alpha, \qquad \alpha \geq 0. \tag{12}$$

The radially nonincreasing constraint (11) can be expressed explicitly as

$$\Phi(r_1, \psi) \geq \Phi(r_2, \psi) \quad \forall \, r_1 < r_2, \quad \forall \, \psi \tag{13}$$

where $\Phi(r, \psi)$ is the kernel function represented in polar coordinates ($r$ and $\psi$ correspond to radius and angle, respectively).

The constraints (10)–(12) and performance measure (9) are formulated so that the optimal kernel passes autocomponents and suppresses cross-components. The constraints force the optimal kernel to be a low-pass filter of fixed volume $\alpha$. As discussed in the Introduction, low-pass kernels are desirable, because AF autocomponents are centered at the origin of the $(\theta, \tau)$ plane, while cross-components tend to lie away from the origin.

The constraints do not dictate the exact shape of the passband of the kernel; the shape is determined by maximizing the performance measure. Clearly, without the monotonicity constraint (11), the optimal kernel would be large, wherever $|A(\theta, \tau)|^2$ is large, regardless of whether the peaks correspond to autocomponents or cross-components. However, assuming that the autocomponents and cross-components are somewhat separated in the $(\theta, \tau)$ plane, the monotonicity constraint imposes a penalty on kernels whose passbands extend over cross-components. These kernels must waste precious kernel volume over regions between the autocomponents and cross-components, where $|A(\theta, \tau)|^2$—and thus also the $|A(\theta, \tau)\Phi(\theta, \tau)|^2$ contribution to the performance measure—is small. Hence, the optimization formulation favors kernels that pass the components concentrated at the origin—precisely the autocomponents.

By the same reasoning, we infer that the performance of this kernel design technique deteriorates for signals whose AF autocomponents and cross-components overlap considerably. However, for this class of signals all current methods of time-frequency analysis fail without a fortuitous conjunction of autocomponents 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 $(\theta, \tau)$ plane. Furthermore, it is straightforward to show that the formulation is insensitive to the time scale of the signal.

The goal of the optimization problem (9)–(12) is strictly to find the kernel that optimally passes autocomponents and suppresses cross-components. Other constraints that encourage additional kernel properties are considered in Section V. Conceptually, the computation of the OK TFD involves first finding the AF of the signal and then solving the optimization problem (9)–(12) for the optimal kernel. The OK TFD $P_{\text{opt}}(t, \omega)$ is then computed using (1).

### B. Selecting the Kernel Volume Parameter $\alpha$

By controlling the volume under the optimal kernel, the parameter $\alpha$ controls the tradeoff between cross-component suppression and smearing of the autocomponents. If $\alpha$ is too small, the resulting TFD will be smeared excessively. If $\alpha$ 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 kernel volume is application-dependent, we offer some guidelines for its choice in this section.

Our approach is to first determine the value of $\alpha$ such that the optimal kernel passes a given amount of energy from the AF into the OK distribution for a "concentrated" signal. Then, bounds on energy transfer will suggest bounds on $\alpha$. The signal most concentrated in $(\theta, \tau)$

[2]Since the signal-dependent kernel optimization is defined only in terms of the squared magnitude of the kernel, $|\Phi_{\text{opt}}(\theta, \tau)|^2$, an additional constraint on the phase of the kernel is necessary to ensure a unique solution to the problem. Here, we assume that the phase is zero, implying that the optimal kernel is a real, nonnegative function.

space is the Gaussian [24]

$$s(t) = \frac{1}{\sqrt[4]{\pi}} e^{-(t^2/4)}, \tag{14}$$

which has an AF with circular equal-energy contours[3]

$$A(\theta, \tau) = e^{-(1/4)(\theta^2 + \tau^2)}. \tag{15}$$

Letting $\zeta$ represent the fraction of energy passed from the AF into the TFD by the optimal kernel, we have (using Parseval's formula for Fourier transform pairs)

$$\zeta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |P_{\text{opt}}(t, \omega)|^2 \, dt \, d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(\theta, \tau) \Phi_{\text{opt}}(\theta, \tau)|^2 \, d\theta \, d\tau \tag{16}$$

The optimal kernel for this Gaussian signal takes on the value 1 on a circular disk and 0 elsewhere in the $(\theta, \tau)$ plane. Computing both the kernel volume, using (12), and the energy transfer measure, using (16), in terms of the radius of the disk yields the relationship between $\alpha$ and $\zeta$ for this signal

$$\alpha = -\ln (1 - \zeta). \tag{17}$$

Small values of $\alpha$ will result in excessive smoothing of the OK distribution. Since there appears to be little benefit gained from smoothing more than the spectrogram [25], the optimal kernel volume that transfers the same amount of energy in the OK TFD as a spectrogram kernel is a reasonable lower bound for $\alpha$. It is straightforward to show, using (16), that the maximum possible energy transfer for a spectrogram is $\zeta = 1/2$. Using this value in (17) yields a lower bound for the optimal kernel volume

$$\alpha \geq 0.69. \tag{18}$$

To minimize autocomponent distortion due to smearing, we must maximize the energy transferred by the optimal kernel. However, as $\zeta$ is increased, a point of diminishing returns is reached: for large $\zeta$, increasing $\zeta$ further requires an exponential increase in kernel volume, making it more likely that significant cross-component energy will be passed into the OK TFD. Upon examination of (17), a reasonable upper bound for the kernel volume (the "knee point" in the $\zeta$ versus $\alpha$ curve) is

$$\alpha \leq 3.0. \tag{19}$$

Note that these bounds should be used only as a guide to prevent oversmoothing and undersmoothing of the OK distribution. The actual value selected for $\alpha$ 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 $(\theta, \tau)$ plane, and values of $\alpha$ considerably larger than 3.0 can still result in adequate cross-component suppression.

[3]More general Gaussian signals corresponding to AF's with noncircular contours lead to a more complicated analysis, but identical results.

## III. Optimal Kernel Solution

### A. Discrete-Time Optimization Formulation

In practice, TFD's are computed at discrete time and frequency locations, so we reformulate the kernel optimization problem by discretizing (9)–(12) on a rectangular grid of samples. With suitably dense sampling, the discrete optimal kernel converges weakly to the continuous-time optimal kernel. Performing the discretization, we define the optimal discrete kernel to be the real, nonnegative function $\Phi_d^{\text{opt}}(m, n)$ that solves

$$\max_{\Phi_d} \sum_{m=-(M/2)}^{(M/2)-1} \sum_{n=-(N/2)}^{(N/2)-1} |A_d(m, n) \Phi_d(m, n)|^2 \tag{20}$$

subject to

$$\Phi_d(0, 0) = 1 \tag{21}$$

$$\Phi_d(m, n) \quad \text{is radially nonincreasing} \tag{22}$$

$$\frac{\Delta_\theta \Delta_\tau}{2\pi} \sum_{m=-(M/2)}^{(M/2)-1} \sum_{n=-(N/2)}^{(N/2)-1} |\Phi_d(m, n)|^2 \leq \alpha, \qquad \alpha \geq 0 \tag{23}$$

where $A_d(m, n)$ and $\Phi_d(m, n)$ represent the discrete AF and kernel, respectively.

Conceptually, the $M \times N$ discrete AF $A_d(m, n)$ is obtained by sampling the continuous-time AF on a rectangular grid

$$A_d(m, n) = A(\theta, \tau)|_{\theta = \Delta_\theta m, \tau = \Delta_\tau n}$$

$$-\frac{M}{2} \leq m \leq \frac{M}{2} - 1,$$

$$-\frac{N}{2} \leq n \leq \frac{N}{2} - 1. \tag{24}$$

Discrete kernels are defined similarly. Sampling of the continuous-time AF on too coarse of a grid ($\Delta_\theta$ or $\Delta_\tau$ too large) could result in a loss of important signal information. Appropriate values for the sampling parameters to ensure that this does not occur are derived in [26], [27] based on the assumption that no information is lost if the continuous-time AF (as an arbitrary two-dimensional function) can be exactly reconstructed from its samples $A_d(m, n)$. To summarize the results, suppose that $L$ samples of a discrete-time signal are obtained by sampling a continuous-time signal that is bandlimited to $1/4T$ Hz

$$s_d(k) = s(kT), \qquad k = 0, 1, \cdots, L - 1. \tag{25}$$

Then, the critical values of the sampling parameters are

$$\Delta_\theta = \frac{2\pi}{LT}, \qquad \Delta_\tau = 2T, \qquad M = N = L. \tag{26}$$

Discretizing (2) using these parameters yields a formula for the direct calculation of the discrete AF

$$A_d(m, n) = T \sum_{k=0}^{L-1} s^*(k - n) s(k + n) e^{j(2\pi/M)mk} \tag{27}$$

which can be efficiently implemented using a series of $L$-point fast Fourier transforms (FFT's).

In discretizing (9)–(12) on a rectangular grid of samples, we have ignored the fact that the kernel nonincreasing constraint (11) implicitly assumes a polar coordinate representation of the kernel, whose exact analog can be implemented only on a polar grid of samples. An alternate discretization would see (9)–(12) discretized on a polar grid. However, computation of the AF and OK distribution would then require the execution of either a polar Fourier transform, for which no fast algorithm currently exists, or costly interpolation from a rectangular grid. Therefore, we abide by our original discretization and approximate the polar grid required by the nonincreasing constraint by a set of paths on the rectangular grid. Fig. 6 illustrates a tree structure [28] that approximates the radial dependencies of the kernel for the upper half-plane of a $32 \times 32$ rectangular grid. The discretized kernel nonincreasing constraint is implemented using a set of constraints of the form

$$\Phi_d(m, n) \leq \Phi_d[p(m, n)] \tag{28}$$

where $p(m, n)$ represents the first sample point encountered when moving along the branch of the tree from sample $(m, n)$ towards the origin. The branches of the tree are constructed to minimize the deviation between the branch that connects a sample to the origin and the radial line through that sample. This deviation is small, so as the sampling density is increased, the tree converges to a set of truly radial lines. Further details regarding the construction of the tree are given in the Appendix.

Since the performance measure (20) and constraints (21)–(23) are linear in $|\Phi_d|^2$, the optimal kernel can be found by applying linear programming [29] to solve for the $MN$ unknowns $\Phi_d(m, n)$. (Since $\Phi_d$ is assumed to be real and nonnegative, knowing $|\Phi_d|^2$ is equivalent to knowing $\Phi_d$.) This suggests a simple procedure for determining samples of the OK distribution. First, the discrete AF of the signal to be analyzed is computed. Next, the linear program is solved for the optimal kernel, which is then multiplied by the AF. The two-dimensional FFT of the product is the discrete OK distribution.

### B. Fast Algorithm

A solution for the optimal kernel using standard linear programming methods may be simple, but it is also computationally expensive. Use of the standard simplex algorithm would result in the optimal kernel solution dominating the total cost of computing the OK distribution. However, by exploiting the simple structure of this particular linear program, we have developed an extremely efficient procedure to construct the optimal kernel, described in a companion article [22]. For an $L \times L$ discrete AF and kernel, the algorithm solves the linear program in $O(L^2 \log L)$ operations. Therefore, the total cost to compute the OK TFD is also $O(L^2 \log L)$, assuming that an $L \times L$ two-dimensional FFT is employed to produce the distribution from the AF-kernel product. This is the same order of computation as required to compute fixed-kernel TFD's such as the Wigner distribution, spectrogram, etc.

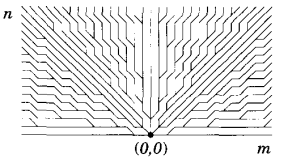



Fig. 6. Minimum-junction tree for the upper half-plane of a $32 \times 32$ rectangular sampling grid. The kernel nonincreasing constraint (22) is implemented by forcing the kernel to be nonincreasing along the approximately radial branches of the tree. Construction details for this tree are given in Appendix A.

Note that since the magnitude of the AF is symmetric

$$|A_d(m, n)|^2 = |A_d(-m, -n)|^2 \tag{29}$$

the optimal kernel is also symmetric, and thus only needs to be computed at sample points in the upper half of the ambiguity plane. This reduces the number of variables in the linear program by approximately a factor of two. Kernel samples in the lower half-plane can be obtained using a symmetry relation similar to (29).

## IV. Examples

In order to compare the performance of the signal-dependent TFD with the fixed-kernel TFD's described in the Introduction, the optimal kernels and OK TFD's were computed for the same pulse and chirp signals given by (3) and (4). Sixty-four samples of each signal were analyzed using a kernel volume of $\alpha = 1.2$ and a minimum-junction tree similar to that in Fig. 6 for the nonincreasing constraint. Also, as discussed in Section V-A, Gaussian tapering with smoothness factor $\sigma = 1.5$ was applied to each optimal kernel prior to computing the OK TFD. The optimal kernels and TFD's are shown in Figs. 7 and 8. The cross-components visible in the fixed-kernel TFD's of Figs. 2–5 are virtually eliminated, yet the autocomponents remain quite concentrated. For monocomponent signals, the resulting OK TFD very closely resembles the Wigner distribution, even for fairly modest kernel volumes.

Low-pass kernels have proven useful for reducing the high noise sensitivity of the bilinear TFD's [30]. To investigate the performance of the design procedure under noisy conditions, we corrupted the chirp signal with additive, white Gaussian noise. The signal-to-noise ratio, measured as the ratio of the total signal power to the total noise power, was 0 dB. The real part of the noisy signal is shown in Fig. 9(a); Figs. 9(b), (c)–(e) illustrate the Wigner, Choi–Williams, and cone-kernel distributions, and the matched-filter spectrogram. The optimal kernel was computed using the same parameters as above and was found to be identical to that shown in Fig. 8(a). The cross-component and noise suppression of the OK distribution, shown in Fig. 9(f), are excellent, indicating that the signal-dependent kernel design procedure is robust in

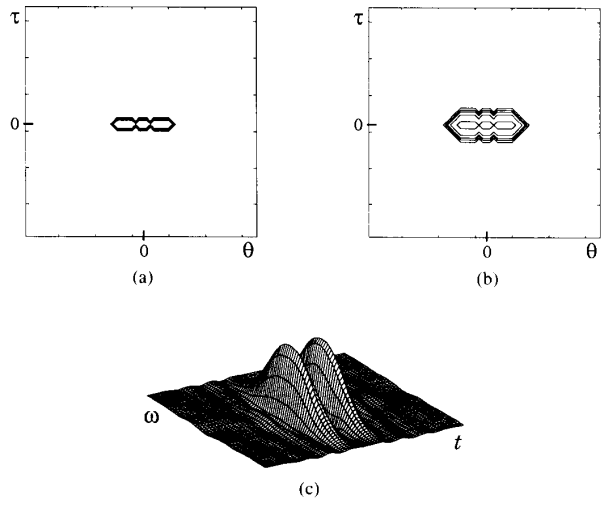



Fig. 7. Optimal kernel and OK TFD for the pulse signal of Fig. 1(a). (a) Optimal kernel of volume $\alpha = 1.2$. Note the close match to the AF autocomponents of Fig. 1(c). (b) Optimal kernel tapered using $\sigma = 1.5$ (see Section V-A for a discussion of kernel tapering). (c) TFD computed using the tapered optimal kernel.

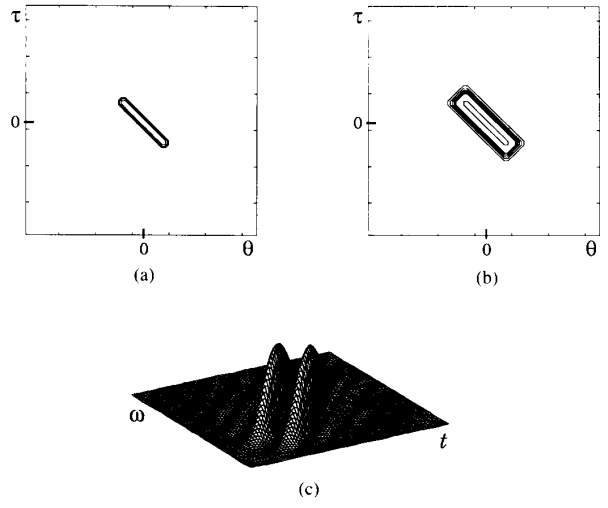



Fig. 8. Optimal kernel and OK TFD for the chirp signal of Fig. 1(b). (a) Optimal kernel of volume $\alpha = 1.2$. Note the close match to the AF autocomponents of Fig. 1(d). (b) Optimal kernel tapered using $\sigma = 1.5$. (c) TFD computed using the tapered optimal kernel.

the presence of significant additive noise. When comparing the OK distribution with the matched-filter spectrogram, remember that the optimal kernel is computed "blind," using no *a priori* information regarding the signal, whereas the matched filter spectrogram requires accurate estimates of the signal parameters (chirp rate, for example). This suggests that the OK TFD may prove useful for the automatic detection of signals in noise.

Fig. 10 illustrates the analysis of 2.5 ms of an echolocation pulse emitted by the large brown bat, *Eptesicus fuscus*. The signal was sampled every 7 $\mu$s, yielding approximately 350 data samples. While the Wigner distri-

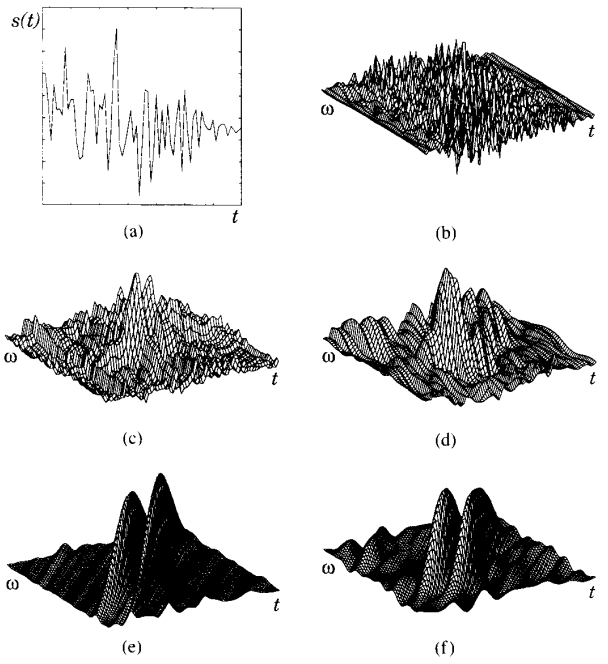



Fig. 9. Time-frequency distributions of the chirp signal of Fig. 1(b) corrupted by additive white Gaussian noise. The SNR, measured as the ratio of the total signal power to the total noise power, is 0 dB. (a) Real part of the time signal. (b) Wigner distribution. (c) Choi-Williams distribution ($\sigma = 5$). (d) Cone-kernel distribution ($a = 2$, $\alpha = 0.036$, and $\beta = 8$). (e) Matched-filter spectrogram. (f) TFD computed from the tapered optimal kernel of volume $\alpha = 1.2$.

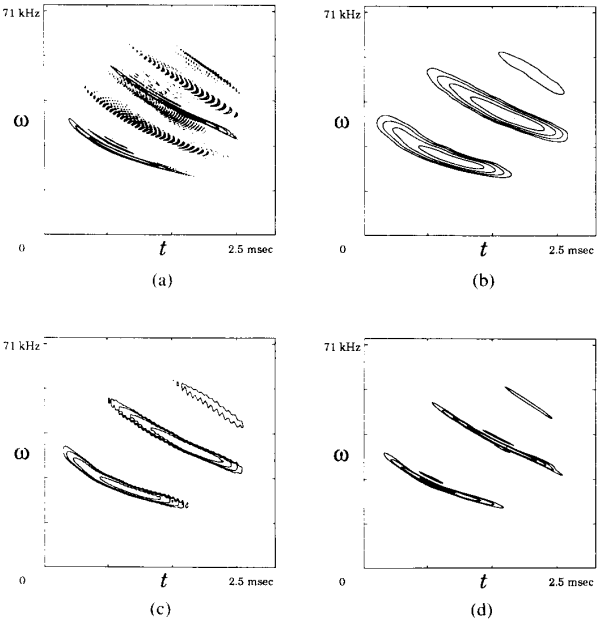



Fig. 10. Time-frequency distributions of 2.5 ms of an echolocation pulse emitted by the large brown bat, *Eptesicus fuscus*. (a) Equal-energy contour plot of the Wigner distribution. (b) Spectrogram computed using a Gaussian window with an $e^{-1}$ time width of 160 $\mu$s. (c) Cone-kernel distribution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

bution of Fig. 10(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 $\mu$s yields a smeared TFD (see Fig. 10(b)). The cone-kernel distribution of Fig. 10(c) is similar. In the OK TFD of Fig. 10(c), computed using a kernel volume of $\alpha = 4$, the fundamental and first two harmonics are sharply resolved, while the cross-components have been suppressed.

As demonstrated in Fig. 10, the OK TFD performs well with slightly hyperbolic-FM signals. However, as the FM law becomes highly nonlinear, the AF autocomponents become less concentrated, and performance deteriorates somewhat. For these types of signals, a technique allowing time and frequency variation of the optimal kernel would produce better results. See [17], [18] for a description of related techniques.

## V. Kernel Design with Additional Constraints

The optimization formulation of Sections II-A and III-A is a general tool 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 formulation is the ease with which it can be customized to incorporate application-specific knowledge into the design process. *A priori* information or requirements are included either by adding additional constraints, modifying the performance measure, or postprocessing the optimal kernel.

### A. Smoothness Constraints

An important consequence of the linear program formulation is that the optimal, signal-dependent kernel takes on essentially only the values 1 and 0 [22]. Although such a kernel is optimal according to the criteria of Section III-A, its sharp cutoff may introduce ringing in the OK distribution, especially for small values of the kernel volume parameter $\alpha$. Smooth kernels can be created either by post-processing the optimal kernel or by adding explicit smoothness constraints to the optimization formulation.

A simple postprocessing approach to smoothing creates a transition band surrounding the region where $\Phi_d = 1$ by tapering the kernel. Let $g(k)$ be a tapering function such that $g(0) = 1$ and $g(k)$ smoothly decreases to zero as $k$ increases. Examples of useful tapering functions include the Gaussian

$$g(k) = e^{-(k^2/\sigma)}, \qquad k \geq 0 \tag{30}$$

and the raised-cosine functions. We taper the kernel using the following procedure (repeated for $k = 0, 1, \cdots$): at each sample that is outside the region of support of $\Phi_d$ but adjacent (within one unit horizontally or vertically) to a sample where $\Phi_d = g(k)$, set $\Phi_d = g(k + 1)$. Note that because it increases the kernel volume, this scheme introduces a tradeoff between the amount of tapering and the amount of cross-component suppression: more gradual tapering results in less ringing, but also potentially lets more cross-component energy through into the TFD. The optimal kernels given in Figs. 7(a) and 8(a) are shown tapered in Figs. 7(b) and 8(b). Alternatively, the kernel tapering can be balanced so that the kernel volume remains constant.

An example will demonstrate the desirability of kernel tapering when the volume parameter is small. Fig. 11 shows the OK distribution for the chirp signal of Fig. 1(b) which results from an untapered kernel. Compare this TFD to Fig. 8(c) and note the substantial ringing that occurs without tapering.

The tapering procedure is simple, but yields a kernel that is formally suboptimal. To obtain smooth optimal kernels, additional smoothness constraints can be added to the kernel optimization formulation (9)–(12). For example, in [26], [27], [31] the kernel is constrained to be Gaussian along any radial profile

$$\Phi(\theta, \tau) = e^{-[\theta^2 + \tau^2/2\sigma^2(\psi)]}. \tag{31}$$

The term $\sigma^2(\psi)$ represents the dependence of the Gaussian "spread" on radial angle $\psi$. Note that any kernel of the form (31) is bounded and radially nonincreasing. Also, if $\sigma(\psi)$ is smooth, then clearly $\Phi(\theta, \tau)$ is also smooth. Unfortunately, while the addition of this constraint results in smooth kernels, the efficient algorithm derived in [22] can no longer be applied to solve for the optimal kernel, because the system (20)–(23), (31) is nonlinear. A gradient ascent algorithm that solves this system is given in [26], [27], [31]. Note that 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.

### B. Marginal Distributions

A TFD satisfies the time and frequency marginal distributions if its kernel takes the value 1 along the $\theta$ and $\tau$ axes [4], [32]. If the following constraints are added to the linear program (20)–(23):

$$\Phi_d(m, 0) = \Phi_d(0, n) = 1 \quad \forall\, n, m \tag{32}$$

the signal-dependent kernel that results will be optimal with respect to the more restricted class of low-pass kernels corresponding to TFDs satisfying the time and frequency marginal distributions. The augmented linear program can be solved using the fast algorithm discussed in [22], because the new constraint does not disturb the special structure that makes efficient solution possible.

The tapered kernel and marginal-satisfying OK distribution computed for the chirp signal are shown in Fig. 12. Comparison with Fig. 3(c) suggests that this customized OK distribution is an attractive alternative to the Choi–Williams distribution in applications where the marginals must be satisfied.

Another technique for finding TFD's that satisfy the marginal distributions has been developed recently by Loughlin, *et al.* In [20] they find the nonbilinear TFD


Fig. 11. OK TFD of the chirp signal of Fig. 1(b) computed using the untapered optimal kernel. Compare to Fig. 8(c), and note the substantial ringing that occurs without tapering.

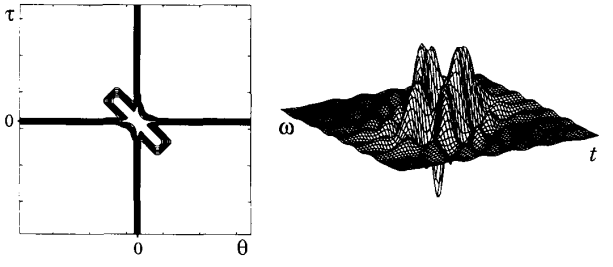



Fig. 12. (a) The optimal marginal-satisfying kernel (shown tapered) and (b) TFD computed for the chirp signal of Fig. 1 (b). The kernel is similar to that of Fig. 8(b), except along the $\theta$ and $\tau$ axes, where it takes the value 1.

from the positive class of marginal-satisfying TFD's [21] that is closest to a composite spectrogram prototype in the sense of minimum cross-entropy. A potential drawback to this approach is that the TFD design is based on the spectrogram, which, as discussed in the Introduction, is biased by the choice of window function. Application of the minimum cross-entropy technique to the OK TFD could result in a positive, marginal satisfying TFD with reduced bias, especially for chirp signals.

### C. Time or Frequency Support

A TFD preserves the convex hull of the time support of a signal if its kernel satisfies a "cone constraint" [5]

$$\int_{-\infty}^{\infty} \Phi(\theta, \tau) e^{-j\theta t}\, d\theta = 0 \qquad \forall\ |t| > 2|\tau|. \tag{33}$$

A similar constraint preserves embedded intervals of zero energy. Since (33) is a linear constraint on the kernel, it would appear that a discretized version could be added to the original constraints of the linear program (20)–(23). However, the resulting optimization formulation has two substantial drawbacks. First, the linear program loses the special structure that permits the application of the fast algorithm derived in [22]. Second, it is unknown whether there exist two-dimensional functions that simultaneously satisfy (33) and (22), implying that the augmented linear program may be ill posed.

In this situation, postprocessing of the kernel is useful. To construct a signal-dependent kernel that preserves time support, we first solve the original linear program for the optimal kernel $\Phi_{\text{opt}}$, then find another kernel $\tilde{\Phi}$ that both preserves time support and is "close" in some sense to $\Phi_{\text{opt}}$. If a least squares distance criterion is used, $\tilde{\Phi}$ is given by the orthogonal projection of $\Phi_{\text{opt}}$ onto the set of support-preserving kernels [33]. In continuous variables, we have

$$\tilde{\Phi}(\theta, \tau) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \Phi_{\text{opt}}(\nu, \tau) e^{-j\nu t}\, d\nu\ \text{rect}_{2|\tau|}(t) \right] e^{j\theta t}\, dt \tag{34}$$

where

$$\text{rect}_y(x) = \begin{cases} 1, & |x| \le y \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

In practice, $\tilde{\Phi}_d(m, n)$ is obtained by taking one-dimensional FFT's 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 FFT's.

While the post-processed kernel is no longer optimal with respect to the original optimization formulation, its passband retains the same general shape and orientation as the optimal kernel. The postprocessed kernel and TFD found using this technique for the chirp signal discussed in the Introduction are shown in Fig. 13. Comparison with Fig. 4(c) suggests that the customized OK 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 [33].

### D. Generalized Performance Measures

Consider a performance measure that is a modified version of (20)

$$\max_{\Phi_d} \sum_{m=-(M/2)}^{(M/2)-1} \sum_{n=-(N/2)}^{(N/2)-1} f(|A_d(m, n)|^2)\ |\Phi_d(m, n)|^2 \tag{36}$$

where $f$: $\mathbb{R}_+ \to \mathbb{R}_+$. Given any such positive transformation $f$, the optimization problem remains linear in the variables $|\Phi_d(m, n)|^2$, so the linear program (36), (21)–(23) can be solved using the same fast algorithm as discussed above, but using the transformed data

$$|\hat{A}_d(m, n)|^2 = f(|A_d(m, n)|^2). \tag{37}$$

(Note, the OK TFD is still computed as the Fourier transform of $A_d\Phi_d$ regardless of the transformation $f$.)

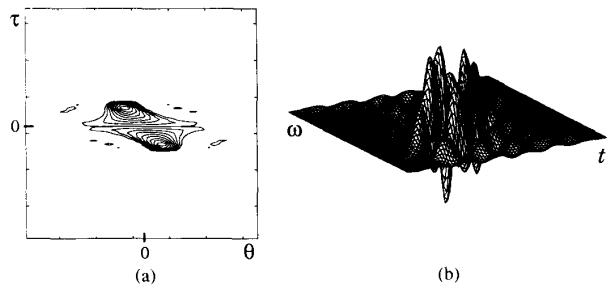
For example, if the autocomponents of interest in a signal are known to lie in a region $\mathcal{R}$ of the ambiguity plane, then setting

$$\Lambda_d(m, n) = \begin{cases} 1, & (m, n) \in \mathcal{R} \\ 0, & \text{otherwise} \end{cases} \tag{38}$$

and

$$f(|A_d(m, n)|^2) = \Lambda_d(m, n)\, |A_d(m, n)|^2 \tag{39}$$

will discourage the optimal kernel passband from lying outside of $\mathcal{R}$. Thus, autocomponents and cross-components lying outside of $\mathcal{R}$ will be suppressed. This pass-




(a) (b)

Fig. 13. (a) The optimal time support-preserving kernel and (b) TFD computed for the signal of Fig. 1(b). (The optimal kernel was not tapered prior to the time-support postprocessing.) The postprocessing changes the shape of the optimal kernel (compare to Fig. 8(a)), but not its orientation angle.

band constraint is potentially useful for analyzing transient signals in the presence of sinusoidal interference signals. In this case, some of the interference can be rejected by setting $\Lambda_d(m, n) = 1$ everywhere except in a region along the $n$ axis, where autocomponents corresponding to sinusoids lie in the ambiguity plane. In general, the transformation (39) is valid for any real, nonnegative, symmetric (as in (29)) $\Lambda_d(m, n)$.

As a second example, consider a signal where a component of interest is of low energy relative to other signal components. Since the function $|A_d(m, n)|^2$ involves fourth powers of the signal, the amplitude of the AF autocomponent of interest may be so small 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 that grows 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, the OK design procedure appears quite robust at low signal-to-noise ratios (as demonstrated in Fig. 9).

Conceivably, a series of constraints like the above marginal and cone constraints could be simultaneously appended to the optimal kernel design formulation in the hope that the resulting optimal-kernel TFD would suppress cross-components yet still preserve every one of the Wigner distribution's desirable properties. Unfortunately, some of the constraints required are mutually incompatible. For example, in order for Moyal's formula [2] to hold in the TFD, the kernel must be all-pass, that is, $|\Phi(\theta, \tau)| = 1$ for all $\theta$ and $\tau$. This constraint is clearly incompatible with the low-pass constraints (11), (12) that are necessary for cross-component suppression. Thus, it appears that (at least within the framework of Cohen's bilinear class) only a subset of the desirable properties of the Wigner distribution can be obtained simultaneously with cross-component suppression.

## VI. Conclusion

In this paper, we have proposed a new signal-dependent time-frequency representation. Signal-dependence is motivated by the fact that a fixed kernel limits the class of signals for which its corresponding TFD performs well. Only by using a kernel that adapts to each signal can a TFD analyze a large class of signals effectively. The simple examples presented here demonstrate both the necessity of signal dependence as well as the excellent performance of the proposed design procedure.

Several additional attractive properties distinguish the optimal kernel design procedure from other techniques proposed in the literature. First, it is based on quantitative optimality criteria. Second and third, it is automatic and relies neither on *a priori* knowledge about the signal nor good fortune to yield a high quality time-frequency representation. Fourth, 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 good even in the presence of substantial additive noise, suggesting that the technique may prove useful for the automatic detection of signals in noise.

Employing the fast algorithm derived in the companion paper [22], the computational cost of finding the optimal kernel TFD is $O(L^2 \log L)$, with $L$ being the number of signal samples to analyze. Since this is the same order as the cost to compute any fixed-kernel TFD, the signal-dependent technique proposed here is practically useful, and not merely of theoretical interest.

This paper by no means exhausts the possibilities for signal-dependent time-frequency analysis. As demonstrated in Section V, 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 TFD's other than bilinear. The optimal radially Gaussian kernel formulation developed in [27], [31] is one example. While the choice of class and performance measure is crucial to success, once a satisfactory class and measure are found, the design of a signal-dependent time-frequency representation reduces to simply solving an optimization problem. Given both 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
## Radial Tree Approximation

In the context of this paper, a *tree* is a set of connections between sample points that lie on a rectangular grid. Trees find application in the kernel nonincreasing constraint (22), (28) of Section III-A. A connection between two samples corresponds to a constraint on the kernel values at those samples, according to (28). The chains of connections that emanate from the origin or root of the tree to the outer edges are called *branches*. If the branches are arranged to fall along approximately radial lines, then

the tree approximates the radially nonincreasing constraint on the kernel. This Appendix describes two recursive methods for constructing approximately radial trees.

### A. Minimum-Norm Trees

A minimum-norm tree is constructed to minimize the deviation between the branch of the tree that connects a sample to the origin and the radial line through that sample. An example is given in Fig. 14. Recall from Section III-B that the tree need be defined only in the upper half-plane of samples. The pattern of branches repeats, so the tree is constructed in two steps. First, the branches for samples located in the triangularly shaped region

$$\mathfrak{U} = \{(m, n): 0 \le m \le \frac{M}{2}, 0 \le n \le m\} \quad (40)$$

are specified. Then the pattern is copied throughout the rest of the tree. Since step two is straightforward, the remainder of this appendix deals exclusively with step one. The columns of samples in $\mathfrak{U}$ can be enumerated by their $m$ index[4]; we denote that $k$th column by $C_k$

$$C_k = \{(k, n): 0 \le n \le k\}, \qquad 0 \le k \le \frac{M}{2}. \quad (41)$$

Note that the height of the columns increases with increasing $k$.

In the region $\mathfrak{U}$, the tree is grown recursively, from the origin outwards. At step $k$ of the recursion, the branches are extended to the samples in $C_k$ from the samples in $C_{k-1}$. The iterations are continued until the tree is the required size. Since all radial lines through samples in $\mathfrak{U}$ lie at angles between 0° and 45° from the $m$ axis, samples in $C_k$ are linked back to samples $C_{k-1}$ either with horizontal or diagonal connections. For example, let $s_n$ be a sample in $C_k$, $s_h$ the horizontally adjacent sample in $C_{k-1}$, and $s_d$ the diagonally adjacent sample in $C_{k-1}$. If $s_n$ corresponds to sample $(m, n)$, then $s_h$ and $s_d$ correspond to samples $(m-1, n)$ and $(m-1, n-1)$, respectively.

The minimum-norm tree is locally optimal in the sense that it connects $s_n$ to the sample in $C_{k-1}$ (either $s_h$ or $s_d$) whose path to the origin deviates least from the radial line through $s_n$. The procedure for choosing the correct connection is simple. First, compute the distances from the radial line through $s_n$ to each sample that lies on the branch that connects $s_h$ to the origin, and place them in a deviation vector $\boldsymbol{\delta}_h$. Likewise, for the branch from $s_d$, compute a deviation vector $\boldsymbol{\delta}_d$. Next, calculate the norms, $\epsilon_h$ and $\epsilon_d$, of the deviation vectors

$$\epsilon_h = \|\boldsymbol{\delta}_h\|, \qquad \epsilon_d = \|\boldsymbol{\delta}_d\|. \quad (42)$$

We have found the $l_\infty$ and $l_2$ norms useful for this purpose as each results in a slightly different tree. Finally, choose the connection corresponding to the smaller of $\epsilon_h$ and $\epsilon_d$.

[4]The samples $A_d(m, n)$ form a rectangular matrix. We assume that $n$, $-N/2 \le n \le N/2 - 1$, is the row index and $m$, $-M/2 \le m \le M/2 - 1$, is the column index.

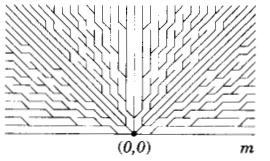



Fig. 14. Minimum-norm tree for the upper half-plane of a 32 × 32 rectangular sampling grid, constructed to minimize the deviation of any branch from a radial line passing through the origin. Note the large number of junctions and dead branches.

If $\epsilon_h = \epsilon_d$, locate the two samples closest to the origin on the branches from $s_h$ and $s_d$ to the origin where the distances to the radial line differ. Then, connect $s_n$ to the branch that minimizes this distance. The tree of Fig. 14 is an $l_\infty$ minimum-norm tree for the upper half-plane of a 32 × 32 sampling grid.

For the $l_\infty$ minimum-norm tree, we believe that the minimax deviation between the branch connecting any sample to the origin and the radial line through that sample is bounded above by twice the intersample spacing. Therefore, as the sampling density of the tree is increased, the branches of the tree converge to truly radial lines. Another important property of the minimum-norm tree is the presence of *dead branches*, paths that do not reach the outside edges of the tree. Dead branches are formed whenever there is more than one junction in a column of samples, a junction being a sample where two branches merge on their paths to the origin. In an $M \times N$ minimum norm tree, there are $O(MN)$ junctions and $O(MN)$ dead branches.

### B. The Minimum-Junction Tree

In some applications, the numerous dead branches of the minimum-norm tree may be undesirable, if only for aesthetic reasons. To construct a tree with no dead branches, we must limit each column in the region $\mathfrak{U}$ to a single junction, hence the name minimum-junction tree. One junction per column limits the total number of junctions in this tree to $O(M)$.

Consider again two adjacent columns $C_{k-1}$ and $C_k$ in $\mathfrak{U}$. Column $C_{k-1}$ must have at least one junction, since $C_k$ contains one more sample than $C_{k-1}$. Since the junction in $C_{k-1}$ connects to two samples in $C_k$ (one with a horizontal link and one with a diagonal link), to prevent a dead branch from forming in $C_{k-1}$, each remaining sample in $C_k$ must connect to a unique sample in $C_{k-1}$. Therefore, samples in $C_{k-1}$ that lie above the junction must connect to $C_k$ with diagonal links, while samples that lie below must connect with horizontal links. Hence, once the junction sample in $C_{k-1}$ is specified, all connections between $C_{k-1}$ and $C_k$ are fixed.

Like the minimum-norm tree, the minimum-junction tree is also constructed recursively. The $k$th iteration proceeds as follows. For each sample in $C_k$, compute the error norms $\epsilon_h$ and $\epsilon_d$ as before. Then find the sample in $C_k$ (call it $s_{\max}$) having the largest error norm and set the connection from $s_{\max}$ to $C_{k-1}$ to minimize this error. If a di-

TABLE I
COMPARISON OF MAXIMUM RADIAL DEVIATIONS IN $M \times M$ MINIMUM-NORM AND MINIMUM-JUNCTION TREES FOR SEVERAL VALUES OF $M$. A DEVIATION OF 1.0 CORRESPONDS TO AN ERROR OF ONE SAMPLE INCREMENT

| Tree Size ($M$) | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|
| Minimum-norm tree | 1.029 | 1.272 | 1.435 | 1.607 | 1.729 | 1.877 |
| Minimum-junction tree | 0.902 | 1.191 | 1.457 | 1.715 | 1.999 | 2.218 |

agonal connection is chosen, set all connections to samples in $C_k$ above $s_{\max}$ to diagonal; if a horizontal connection is chosen, set all connections to samples in $C_k$ below $s_{\max}$ to horizontal. Now repeat the procedure for the samples in $C_k$ whose connections remain unspecified until all samples in $C_k$ have been assigned a connection to $C_{k-1}$. An $l_\infty$ minimum-junction tree was shown in Fig. 6.

Unfortunately, we have been unable to analytically bound the radial deviation for the minimum-junction tree. Table I gives the results of an empirical study that compares the maximum errors for various sizes of minimum-norm and minimum-junction trees. At present, it is not clear which of the two trees is preferred. We have employed both in applications and found the results satisfying. Asymptotically, the minimum-norm tree has the smallest radial deviation, but also many dead branches.

## ACKNOWLEDGMENT

The authors wish to thank L. Potter, S. Meyn, and G. Cybenko for discussing the optimization problems, T. Brotherton and T. Posch for suggesting some useful constraints, and C. Condon, K. White, and A. Feng for the bat data and for permission to use it in this paper.

## REFERENCES

[1] L. Cohen, "Generalized phase space distribution functions," *J. Math. Phys.*, vol. 7, pp. 781-786, 1966.
[2] L. Cohen, "Time-frequency distributions—a review," *Proc. IEEE*, vol. 77, pp. 941-981, July 1989.
[3] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency representations," *IEEE Signal Processing Mag.*, vol. 9, pp. 21-67, Apr. 1992.
[4] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 862-871, June 1989.
[5] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone-shape kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1084-1091, July 1990.
[6] F. Hlawatsch, "Interference terms in the Wigner distribution," in *Proc. Int. Conf. Digital Signal Processing* (Florence, Italy), 1984.
[7] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing—ICASSP '84*, 1984, pp. 41.B.4.1-41.B.4.4.
[8] L. Cohen and T. E. Posch, "Generalized ambiguity functions," in *Proc. IEEE Inter. Conf. Acoust., Speech, Signal Processing—ICASSP '85*, 1985, pp. 27.6.1-27.6.4.
[9] P. Flandrin and O. Rioul, "Affine smoothing of the Wigner-Ville distribution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing—ICASSP '90*, 1990, pp. 2455-2458.
[10] T. A. C. M. Claasen 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, no. 6, pp. 372-389, 1980.
[11] S. C. Glinski, "Diphone speech synthesis based on a pitch-adaptive short-time Fourier transform," Ph.D. dissertation, Univ. Illinois at Urbana-Champaign, Dep. Elec. Comput. Eng., 1981.
[12] A. H. Nuttall, "Wigner distribution function: Relation to short-term spectral estimation, smoothing, and performance in noise," Tech. Rep. NUSC 8225, Naval Underwater Syst. Center, Feb. 1988.
[13] G. Jones and B. Boashash, "Instantaneous frequency, instantaneous bandwidth and the analysis of multicomponent signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing—ICASSP '90*, 1990, pp. 2467-2470.
[14] L. Cohen, "Distributions concentrated along the instantaneous frequency," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1348, pp. 149-157, 1990.
[15] S. Kadambe, G. F. Boudreaux-Bartels, and P. Duvaut, "Window length selection for smoothing the Wigner distribution by applying an adaptive filter technique," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing—ICASSP '89*, 1989, pp. 2226-2229.
[16] 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 Processing*, vol. 35, pp. 764-769, June 1987.
[17] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing—ICASSP '87*, 1987, pp. 681-684.
[18] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2127-2135, Dec. 1990.
[19] D. Mihovilovic and R. N. Bracewell, "Adaptive chirplet representation of signals on time-frequency plane," *Electron. Lett.*, vol. 27, June 20, 1991.
[20] P. J. Loughlin, J. W. Pitton, and L. E. Atlas, "An information theoretic approach to positive time-frequency distributions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing—ICASSP '92*, 1992, pp. V125-V128.
[21] L. Cohen and T. E. Posch, "Positive time-frequency distribution functions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 31-37, Feb. 1983.
[22] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: Fast algorithm for kernel design," *IEEE Trans. Signal Processing*, to be published, Nov. 1993.
[23] R. G. Baraniuk and D. L. Jones, "Optimal kernels for time-frequency analysis," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1348, pp. 181-187, 1990.
[24] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.*, vol. 93, pp. 429-457, 1946.
[25] D. L. Jones, and T. W. Parks, "A resolution comparison of several time-frequency representations," *IEEE Trans. Signal Processing*, vol. 40, pp. 413-420, Feb. 1992.
[26] R. G. Baraniuk and D. L. Jones, "Shear madness: Signal-dependent and metaplectic time frequency representations," Tech. Rep. UILU-ENG-92-2226, Coordinated Sci. Lab., Univ. Illinois, Urbana-Champaign, 1992.
[27] R. G. Baraniuk and D. L. Jones, "A radially Gaussian, signal-dependent time-frequency representation," *Signal Processing*, vol. 32, no. 2, May 1993.
[28] H. R. Lewis and L. Denenberg, *Data Structures and Their Algorithms*. New York: Harper Collins, 1991.
[29] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1973.
[30] L. Cohen, "On a fundamental property of the Wigner distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 559-561, Apr. 1987.
[31] R. G. Baraniuk and D. L. Jones, "A radially Gaussian, signal-dependent time-frequency representation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing—ICASSP '91*, 1991, pp. 3181-3184.
[32] B. Boashash, "Time-frequency signal analysis," in *Advances in Spectrum Estimation and Array Processing*, vol. 1, S. Haykin, Ed. Englewood Cliffs, NJ:Prentice-Hall, 1991, pp. 418-517.
[33] S. Oh, R. J. Marks, and J. W. Pitton, "Kernel synthesis for generalized time-frequency distributions using the method of projection onto convex sets," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1348, pp. 197-207, 1990.

**Richard G. Baraniuk** received the B.Sc. degree in 1987 from the University of Manitoba, Canada, the M.Sc. degree in 1988 from the University of Wisconsin-Madison, and the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, all in electrical and computer engineering.

While at the University of Illinois, he held a joint appointment with the CERL Sound Group and the Coordinated Science Laboratory. During 1992–1993 he was with the Signal Processing Laboratory of Ecole Normale Supérieure de Lyon, Lyon, France, supported by a Canadian NSERC NATO postdoctoral fellowship. Currently, he is an Assistant Professor of Electrical and Computer Engineering at Rice University in Houston, TX. His research interests lie in the area of signal and image processing and include wavelet theory and time-frequency analysis.

**Douglas L. Jones** (S'82–M'83) received the B.S.E.E. degree in 1983, the M.S.E.E. degree in 1986, and the Ph.D. degree in 1987, all from Rice University.

During the 1987–1988 academic year, he was at the University of Erlangen-Nuremberg in Germany on a Fulbright postdoctoral fellowship. Since 1988, he has been an Assistant Professor of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. He is an author of the laboratory textbook *A Digital Signal Processing Laboratory Using the TMS32010*. His research interests are in digital signal processing, including time-frequency and time-varying signal analysis, efficient algorithms for VLSI implementation, and various applications.