Quadratic Time-Frequency Distributions

Selin Aviyente
Department of Electrical and Computer Engineering
Michigan State University

January 26, 2010
Introduction to Quadratic Time-Frequency Distributions

- STFT is a linear time-frequency distribution since it is computed by taking the inner product of the signal with a time-frequency localized window function.
- These types of representations have the advantage of computational efficiency and reconstruction.
- These advantages come at the expense of limited time-frequency resolution, limited by the window’s spread in time and frequency.
- Wigner-Ville distribution is an energy density computed by correlating the signal with a time and frequency translated version of itself.
- Wigner introduced this distribution in 1932 in the context of thermodynamics and Ville applied it to signal processing in 1948.
Wigner Distribution

- Wigner Distribution is defined as:
  \[ W(t, \omega) = \frac{1}{2\pi} \int x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})e^{-j\omega\tau} d\tau \]  
  \[ (1) \]

- It is the Fourier transform of the signal’s local autocorrelation function or time-varying correlation, 
  \[ R(t, \tau) = x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2}). \]

- This definition is analogous to Wiener-Khinchine Theorem which relates the power spectral density to autocorrelation function for stationary processes, 
  \[ S(\omega) = \int R(\tau)e^{-j\omega\tau} d\tau \]
  where \[ R(\tau) = E[x(t)x^*(t - \tau)]. \]

- Wigner Distribution can equivalently be expressed as:
  \[ W(t, \omega) = \frac{1}{2\pi} \int X^*(\omega + \frac{\theta}{2})X(\omega - \frac{\theta}{2})e^{-jt\theta} d\theta \]  
  \[ (2) \]
Properties of Wigner Distribution

- Real: Wigner Distribution is always real even if the signal is complex. \( R(t, \tau) = R^*(t, -\tau) \) (Hermitian symmetry)
- Bilinear
- Symmetric: If the signal is real \( W(t, \omega) = W(t, -\omega) \).
- Nonpositivity: Wigner distribution will always have regions of negative values for any signal.
- Time Support: For an infinite duration signal, Wigner distribution will be nonzero for all time. For a finite duration signal the Wigner distribution will be zero before the start of the signal and after the end. Therefore, \( W(t, \omega) = 0 \) for \( t \) outside \((t_1, t_2)\) if \( x(t) = 0 \) outside \((t_1, t_2)\). A similar argument can be made for frequency support.
Preserves Energy:

\[
\int \int W(t, \omega) dt d\omega = \frac{1}{2\pi} \int \int \int x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})e^{-j\omega \tau} d\tau dt d\omega
\]

\[
= \int \int x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})\delta(\tau) d\tau dt
\]

\[
= \int |x(t)|^2 dt = E
\]  

Time Marginal:

\[
\int W(t, \omega) d\omega = \int \int x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})e^{-j\omega \tau} d\tau d\omega
\]

\[
= \int x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})\delta(\tau) d\tau
\]

\[
= |x(t)|^2
\]  

Frequency Marginal: Similarly, \( \int W(t, \omega) dt = |X(\omega)|^2 \)
Time and Frequency Shift Invariant: If $x_{\text{new}}(t) = x(t - t_0)e^{j\omega_0 t}$ then $W_{\text{new}}(t, \omega) = W(t - t_0, \omega - \omega_0)$.

Scale Invariance: If $x_{\text{new}}(t) = \frac{1}{\sqrt{s}}x(t/s)$ then $W_{\text{new}}(t, \omega) = W\left(\frac{t}{s}, s\omega\right)$.

Phase shift does not modify the Wigner distribution, i.e. you cannot get phase information from Wigner distribution.

Moyal’s Formula: For any $f, g \in L_2(\mathbb{R})$:

$$\left| \int f(t)g^*(t)dt \right|^2 = \frac{1}{2\pi} \int \int W_f(t, \omega)W_g(t, \omega)dtd\omega$$  (5)
Instantaneous Frequency

- Treat Wigner distribution as a joint density function.
- The instantaneous frequency is the average frequency at a particular time $\mu_{\omega|t}$ similar to conditional expected value.
- For $x(t) = a(t) \exp(j\phi(t))$, the instantaneous frequency can be found as:

$$\phi'(t) = \frac{\int \omega W(t, \omega) d\omega}{\int W(t, \omega) d\omega}$$

(6)
Examples

- **Example 1:** If \( x(t) = e^{j\omega_0 t} \) then \( W(t, \omega) = \delta(\omega - \omega_0) \). Similarly, if \( x(t) = \sqrt{2\pi}\delta(t - t_0) \) then \( W(t, \omega) = \delta(t - t_0) \).

- **Example 2:** For a chirp signal, \( x(t) = e^{j(\beta t^2/2 + \omega_0 t)} \) then \( W(t, \omega) = \delta(\omega - \beta t - \omega_0) \), i.e. the Wigner distribution is concentrated around the instantaneous frequency.

- **Example 3:** Chirp with a Gaussian envelope, \( x(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \), \( W(t, \omega) = \frac{1}{\pi} e^{-\alpha t^2 - (\omega - \beta t - \omega_0)^2/\alpha} \). This is the most general case for which the Wigner distribution is positive.
Wigner Distribution of the Sum of Two Signals: Cross-Terms

Suppose \( x(t) = x_1(t) + x_2(t) \), then
\[
W(t, \omega) = W_{11}(t, \omega) + W_{22}(t, \omega) + W_{12}(t, \omega) + W_{21}(t, \omega).
\]

\( W_{12}(t, \omega) = \frac{1}{2\pi} \int x_1^*(t - \frac{\tau}{2})x_2(t + \frac{\tau}{2})e^{-j\omega \tau} d\tau \) is called the cross-term or the cross Wigner distribution. \( W_{12} = W_{21}^* \).

The cross Wigner distribution is complex. Therefore, \( W(t, \omega) = W_{11}(t, \omega) + W_{22}(t, \omega) + 2Re(W_{12}(t, \omega)) \).
Examples of Cross-terms

Let \( x(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \).

\[
W(t, \omega) = \frac{1}{2\pi} \int (A_1 e^{j\omega(t+\tau/2)} + A_2 e^{j\omega_2(t+\tau/2)})
\]

\[
= \int [A_1^2 e^{j\omega_1 \tau} + A_1 A_2 e^{j(\omega_2-\omega_1) t} e^{j(\omega_1+\omega_2) \frac{\tau}{2}}] e^{-j\omega \tau} d\tau
\]

\[
+ A_2^2 e^{j\omega_2 \tau} + A_2 A_1 e^{j(\omega_1-\omega_2) t} e^{j(\omega_1+\omega_2) \frac{\tau}{2}} e^{-j\omega \tau} d\tau
\]

\[
= \int A_1^2 e^{j(\omega_1-\omega) \tau} d\tau + \int A_2^2 e^{j(\omega_2-\omega) \tau} d\tau
\]

\[
+ 2A_1 A_2 \cos((\omega_1 - \omega_2) t) \int e^{j(\omega_1+\omega_2) \frac{\tau}{2}} e^{-j\omega \tau} d\tau
\]

\[
= A_1^2 \delta(\omega - \omega_1) + A_2^2 \delta(\omega - \omega_2)
\]

\[
+ 2A_1 A_2 \cos((\omega_1 - \omega_2) t) \delta(\omega - \frac{\omega_1 + \omega_2}{2})
\]
Observations about Cross-Terms

- The cross-terms are always located at center time and frequency location between two components.
- Cross-terms usually take on negative values since they oscillate.
- The farther away the two components are the less the oscillations and negative terms.
- Frequency of oscillation depends on the difference in time and frequency centers, $\sqrt{(\Delta t)^2 + (\Delta \omega)^2}$. 

\[ \sqrt{(\Delta t)^2 + (\Delta \omega)^2} \]
Since any signal can be broken up into a sum of parts in an arbitrary way, the cross terms can be neither bad nor good since they are not uniquely defined.

A signal is called monocomponent if it is well delineated in a region, otherwise it is a multicomponent signal.

Interference terms also exist in a real single between positive and negative frequencies. Therefore, use the analytic signal.
What can we do about the interference terms?

- Apply filtering/smoothing operators in time and frequency:

\[ C_\theta(t, \omega) = \int \int W(t', \omega') \theta(t, t', \omega, \omega') dt' d\omega' \]  (7)

- One can choose \( \theta \) (the kernel function) such that \( C_\theta(t, \omega) \geq 0 \).

- For example, spectrogram is a smoothed version of Wigner distribution that is positive for all time and frequency and does not have any interference terms. However, this comes at the expense of loss of high resolution in Wigner.
In particular:

\[ P_S(t, \omega) = \int \int W_s(t', \omega') W_g(t - t', \omega - \omega') \, dt' \, d\omega' \]  \hspace{1cm} (8)

where \( W_g \) is the Wigner distribution of the window function. The spectrogram is the 2D convolution of Wigner distribution of the signal with the Wigner distribution of the window function.

- Spectrogram is smoothed Wigner distribution since \( W_g(t, \omega) \) acts like a lowpass filter.

- There is no window function that would produce a spectrogram with the desirable properties of Wigner distribution, e.g. instantaneous frequency.
Inversion of Wigner Distribution

- Wigner distribution is written as:
  \[
  W(t, \omega) = \frac{1}{2\pi} \int x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})e^{-j\omega\tau} d\tau
  \] (9)

- Therefore, \(s^*(t - \frac{\tau}{2})s(t + \frac{\tau}{2}) = \int W(t, \omega)e^{j\tau\omega} d\omega\).

- Let \(t = \frac{\tau}{2}\) and then setting \(\tau = t\), we have

  \[
  s(t) = \frac{1}{s^*(0)} \int W(t/2, \omega)e^{it\omega} d\omega
  \] (10)

- The signal can be recovered from the Wigner distribution up to a constant. The constant can be obtained from the normalization condition up to an arbitrary phase factor.
Representability

- Not every function of time and frequency is a proper Wigner distribution because there may not exist a signal that will generate it.

- For example, most positive two dimensional functions are not proper Wigner distributions.

- Generally, the inversion formula can be used to find the signal and then calculate the Wigner distribution from the derived signal. If the two functions agree then it is a representable Wigner distribution.
Let $v = \tau/2$, then $W(t, \omega) = 2 \int s(t + v)s^*(t - v)e^{-j2v\omega}dv$.

Next discretize $v$:
\[ W(t, \omega) = 2\Delta \sum_n s(t + n\Delta)s^*(t - n\Delta)e^{-j2n\Delta\omega} \quad (11) \]

Now sample $t$, $\Delta = \Delta t$:
\[ W(m\Delta t, \omega) = 2\Delta t \sum_n s((m + n)\Delta t)s^*((m - n)\Delta t)e^{-j2\omega n\Delta t} \quad (12) \]

This means that Wigner distribution is periodic with $\pi/\Delta t$:
\[ W(m\Delta t, \omega + \frac{\pi}{\Delta t}) = W(m\Delta t, \omega). \]
According to Shannon’s sampling theorem, the spectrum of a sampled signal is periodic with $\frac{2\pi}{\Delta t}$. For Wigner distribution, the highest frequency component must be less than or equal to $\frac{\pi}{2\Delta t}$, otherwise aliasing will occur.

To obtain alias free Wigner distribution, the sampling rate should be doubled, you can use interpolation.

Interpolate $s[m]$ and pass through a lowpass filter to obtain $y[m]$.

If the original sample interval is $\Delta t$, interpolated sample interval is $0.5\Delta t$.

$$W(m\frac{\Delta u}{2}, \omega) = \frac{2\Delta u}{2} \sum y[(m+n)\frac{\Delta u}{2}]y^*[((m-n)\frac{\Delta u}{2})e^{-j2\omega n\frac{\Delta u}{2}}}$$ (13)

Let $\theta = \frac{\omega\Delta u}{2}$:

$$W(m, \theta) = 2 \sum y[m + n]y^*[m - n] \exp(-j2\theta n)$$ (14)
In practice, we don’t have infinite duration signals, use pseudo Wigner distribution.

\[ W(m, \theta) = 2 \sum_{n=-\infty}^{\infty} w(n)y[m + n]y^*[m - n] \exp(-j2\theta n) \]

\[ = 2 \sum_{n=-(2L-1)}^{2L-1} y[m + n]y^*[m - n] \exp(-j2\theta n) \]

\[ w[n] = \begin{cases} 1, & |n| < 2L \\ 0, & otherwise \end{cases} \]

\[ W(m, \theta) = 2 \sum_{n=-(2L-1)}^{0} y[m + n]y^*[m - n] \exp(-j2\theta n) \]

\[ + 2 \sum_{n=0}^{2L-1} y[m + n]y^*[m - n] \exp(-j2\theta n) - 2y[m]y^*[m] \]

\[ = 4Re\left( \sum_{n=0}^{2L-1} y[m + n]y^*[m - n] \exp(-j2\theta n) \right) - 2y[m]y^*[m] \]
Discretize frequency

\[
\begin{align*}
DW(m, k) &= 4 \text{Re} \left( \sum_{n=0}^{2L-1} y[m+n]y^*[m-n] \exp(-j \frac{4\pi kn}{2L}) \right) \\
&- 2y[m]y^*[m], \quad 0 \leq k < 2L
\end{align*}
\] (15)

\[DW(m, k) = DW(m, k + iL), \quad i = \ldots, -2, -1, 0, 1, 2, \ldots\]

Half the output is redundant.

\[
\begin{align*}
DW(m, k) &= 4 \text{Re} \left( \sum_{n=0}^{L} y[m+n]y^*[m-n] \exp(-j \frac{2\pi kn}{L}) \right) \\
&- 2y[m]y^*[m]
\end{align*}
\] (16)
Computational Complexity

- For each m, \( L \) multiplications, \( L \) point FFT.
- Computational complexity is \( O(L^2 \log L) \).
General Class of Cohen’s Class of Distributions

- Goal: Retain desirable properties of Wigner distribution such as high resolution, energy preservation, marginals and reduce the effect of cross-terms simultaneously.
- General class of quadratic time-frequency distributions:

\[
C(t, \omega) = \frac{1}{4\pi^2} \int \int \int s(u + \frac{\tau}{2})s^*(u - \frac{\tau}{2})\phi(\theta, \tau)e^{-j\theta t}e^{-j\omega \tau}e^{j\theta u} du d\tau d\theta
\]

(17)

- \(\phi(\theta, \tau)\) is the kernel function that determines the properties of the distribution. For Wigner distribution, \(\phi(\theta, \tau) = 1\).
Another way to look at Cohen’s class is through the ambiguity function.

\[ A(\theta, \tau) = \int s(u + \frac{\tau}{2})s^*(u - \frac{\tau}{2})e^{j\theta u} du \]  

- Fourier transform of local autocorrelation function with respect to time.
- \( \theta \) is the spectral lag/Doppler frequency parameter. Ambiguity function measures the energy concentration in time and frequency.
Wigner distribution is the double Fourier transform of ambiguity function

\[ W(t, \omega) = \frac{1}{4\pi^2} \int \int A(\theta, \tau) e^{-j\theta t} e^{-j\tau \omega} d\theta d\tau \] (19)

General Cohen’s class can be represented as:

\[ C(t, \omega) = \frac{1}{4\pi^2} \int \int A(\theta, \tau) \phi(\theta, \tau) e^{-j\theta t} e^{-j\tau \omega} d\theta d\tau \] (20)

The kernel acts like a two-dimensional filter. \( M(\theta, \tau) = A(\theta, \tau) \phi(\theta, \tau) \) is sometimes referred to as the characteristic function of the time-frequency distribution.
Alternatively, Cohen’s class of TFDs can be written in the autocorrelation domain.

Let $\psi(t, \tau) = \frac{1}{2\pi} \int \phi(\theta, \tau) e^{-j\theta t} d\theta$ be the kernel function in the time/time-lag domain.

Then we can write $C(t, \omega)$:

$$C(t, \omega) = \frac{1}{2\pi} \int \int \psi(t-u, \tau) s(u + \frac{\tau}{2}) s^*(u - \frac{\tau}{2}) e^{-j\omega \tau} d\tau du$$  (21)

First convolve $R(t, \tau)$ with the kernel $\psi(t, \tau)$ along time, then take Fourier Transform with respect to $\tau$. 
Basic Properties Related to the Kernel
Energy Preservation: For the total energy of the signal to be preserved, $\phi(0, 0) = 1$.

Time Marginal:

$$\int C(t, \omega) d\omega = \frac{1}{4\pi^2} \int \int \int \int s(u + \frac{\tau}{2}) s^*(u - \frac{\tau}{2}) \phi(\theta, \tau) e^{j\theta(u-t)} e^{-j\tau\omega} du d\tau d\theta d\omega$$

$$= \frac{1}{2\pi} \int \int \int \delta(\tau) s(u + \frac{\tau}{2}) s^*(u - \frac{\tau}{2}) \phi(\theta, \tau) e^{j\theta(u-t)} du d\tau d\theta$$

$$= \frac{1}{2\pi} \int \int |s(u)|^2 \phi(\theta, 0) e^{j\theta(u-t)} du d\theta$$

(22)

For this integral to equal $|s(t)|^2$,

$$\frac{1}{2\pi} \int \phi(\theta, 0) e^{j\theta(u-t)} d\theta = \delta(t - u).$$
Therefore, $\phi(\theta, 0) = 1$. 
- **Frequency Marginal:** For the frequency marginal to be satisfied, $\phi(0, \tau) = 1$.

- **Reality:** For a distribution to be real the characteristic function should be Hermitian symmetric, i.e.
  $$A(\theta, \tau)\phi(\theta, \tau) = A^*(\theta, \tau)\phi^*(-\theta, -\tau).$$
  Therefore,
  $$\phi(\theta, \tau) = \phi^*(-\theta, -\tau).$$

- **Time and Frequency Shift Invariance:** The kernel is independent of $t$ and $\omega$.

- **Scale Invariance:** A distribution is scale invariant if a signal is linearly scaled then the spectrum is inversely scaled.
  $$C_{sc}(t, \omega) = C(at, \omega/a)$$
  for $s_{sc}(t) = \sqrt{as}(at)$. For scale invariance,
  $$\phi(\theta, \tau) = \phi(\theta \tau).$$
Inversion

- Ambiguity function can be recovered by taking the inverse Fourier transform of \( \frac{M(\theta, \tau)}{\phi(\theta, \tau)}. \)

\[
s^*(u - \frac{\tau}{2})s(u + \frac{\tau}{2}) = \frac{1}{2\pi} \int \frac{M(\theta, \tau)}{\phi(\theta, \tau)} e^{-j\theta u} d\theta \tag{23}
\]

- Let \( t = u + \frac{\tau}{2} \) and \( t' = u - \frac{\tau}{2} \):

\[
s^*(t')s(t) = \frac{1}{2\pi} \int \frac{M(\theta, t - t')}{\phi(\theta, t - t')} e^{-j\theta(t+t')/2} d\theta \tag{24}
\]

- Let \( t' = 0 \):

\[
s(t) = \frac{1}{2\pi s^*(0)} \int \frac{M(\theta, t)}{\phi(\theta, t)} e^{-j\theta t/2} d\theta
\]

\[
= \frac{1}{2\pi s^*(0)} \int \int \int \frac{C(t', \omega)}{\phi(\theta, t)} e^{j\omega t + j\theta(t' - t)/2} dt' d\omega d\theta
\]

- The signal can be recovered if the division by the kernel is uniquely defined for all \( \theta, \tau \).
Instantaneous Frequency

- Instantaneous frequency can be found from the distribution,
  \[ \frac{\int \omega C(t,\omega) d\omega}{\int C(t,\omega) d\omega}. \]

- For this to be equal to the actual instantaneous frequency, i.e. the derivative of the phase, two conditions need to be satisfied. The time marginal should be satisfied →
  \[ \phi(\theta, 0) = 1 \quad \text{and} \quad \frac{\partial \phi(\theta, \tau)}{\partial \tau} \bigg|_{\tau=0} = 0. \]
Consider the ambiguity function of the sum of two sinusoids, \( A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \):

\[
A(\theta, \tau) = \int |A_1|^2 e^{j\omega_1 (u+\tau/2)} e^{-j\omega_1 (u-\tau/2)} e^{j\theta u} \, du
\]
\[
+ \int |A_2|^2 e^{j\omega_2 (u+\tau/2)} e^{-j\omega_2 (u-\tau/2)} e^{j\theta u} \, du
\]
\[
+ \int A_1 A_2 e^{j\omega_1 (u+\tau/2)} e^{-j\omega_2 (u-\tau/2)} e^{j\theta u} \, du
\]
\[
+ \int A_2 A_1 e^{j\omega_2 (u+\tau/2)} e^{-j\omega_1 (u-\tau/2)} e^{j\theta u} \, du
\]

(25)
\[ A(\theta, \tau) = |A_1|^2 e^{i\omega_1 \tau} \delta(\theta) + |A_2|^2 e^{i\omega_2 \tau} \delta(\theta) + A_1 A_2 e^{i(\omega_1 + \omega_2) \tau/2} \delta(\theta + \omega_1 - \omega_2) + A_2 A_1 e^{i(\omega_1 + \omega_2) \tau/2} \delta(\theta + \omega_2 - \omega_1) \]

(26)

- The auto-terms are concentrated around \( \theta \) and \( \tau \) axes.
- The cross-terms are concentrated around the difference of the frequencies, off the origin in the ambiguity domain.
Similarly, for \( s(t) = A_1 \delta(t - t_1) + A_2 \delta(t - t_2) \):

\[
A(\theta, \tau) = A_1^2 \int \delta(t + \frac{\tau}{2} - t_1) \delta(t - \frac{\tau}{2} - t_1) e^{i\theta t} dt
+ A_2^2 \int \delta(t + \frac{\tau}{2} - t_2) \delta(t - \frac{\tau}{2} - t_2) e^{i\theta t} dt
+ A_1 A_2 \int \delta(t + \frac{\tau}{2} - t_1) \delta(t - \frac{\tau}{2} - t_2) e^{i\theta t} dt
+ A_2 A_1 \int \delta(t + \frac{\tau}{2} - t_2) \delta(t - \frac{\tau}{2} - t_1) e^{i\theta t} dt
= A_1^2 e^{i\theta t_1} \delta(\tau) + A_2^2 e^{i\theta t_2} \delta(\tau)
+ A_1 A_2 e^{i\theta(t_1 + t_2)/2} \delta(\tau - (t_1 - t_2)) + A_1 A_2 e^{i\theta(t_1 + t_2)/2} \delta(\tau + (t_1 - t_2))
\]
Kernel Design

- $\phi(\theta, \tau)$ should be small away from the $\theta - \tau$ axis.
- $\phi(\theta, \tau) \ll 1$ for $\theta \tau \gg 0$.
- We can use product kernels which are characterized by functions of one variable.
- Let $x = \theta \tau$, then $\phi(x) \ll 1$ for $x \gg 0$. Therefore, kernel design reduces to designing a one-dimensional window function, $h(t)$, $\phi(\theta \tau) = \int h(t)e^{-j\theta \tau t}dt$.
- Marginals: $\phi(0) = 1 \rightarrow \int h(t)dt = 1$.
- Cross-term Minimization: $\phi(x) \ll 1$, $h(t)$ should be a lowpass window.
- Real: $\phi^*(x) = \phi(-x), h(t) = h(-t)$. 
Some Common Kernels

- Choi-Williams Kernel: $\phi(\theta, \tau) = \exp\left(-\frac{\theta^2 \tau^2}{\sigma}\right)$. Satisfies the marginals and preserves energy. $\sigma$ controls the trade-off between resolution and the amount of cross-terms. As $\sigma \to \infty$, it gets close to the Wigner distribution.

- Sinc kernel (Born Jordan Kernel): $\frac{\sin(a\theta \tau)}{a\theta \tau}$. Satisfies the marginals and preserves energy. $a$ controls the mainlobe width, larger the $a$ the narrower the mainlobe.

- Zhao-Atlas-Marks: $\phi(\theta, \tau) = g(\tau) |\tau| \frac{\sin(a\theta \tau)}{a\theta \tau}, g(\tau) = 1, a = 1/2$. The cross-terms are under the auto-terms.
For $A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t}$, the cross-terms are:

\[
C_{12} = \frac{1}{4\pi^2} \int \int \phi(\theta, \tau) A_1 A_2 e^{i(\omega_1 + \omega_2)/2\tau} 2\pi \delta(\theta + \omega_1 - \omega_2) e^{-j\theta t} e^{-j\tau \omega} d\theta d\tau \\
= \frac{A_1 A_2}{2\pi} \int \phi(\omega_2 - \omega_1, \tau) e^{i(\omega_1 + \omega_2)/2\tau} e^{-j(\omega_2 - \omega_1) t} e^{-j\tau \omega} d\tau \\
= \frac{A_1 A_2}{2\pi} e^{-j(\omega_2 - \omega_1) t} \int \phi(\omega_2 - \omega_1, \tau) e^{-j\tau (\omega - \omega_1)} d\tau
\]

If $\phi(\theta, \tau) = f(\theta, \tau) \sin(a\theta\tau)$:

\[
\frac{A_1 A_2}{4j\pi} \int f(\omega_2 - \omega_1, \tau)[e^{-j\tau (\omega + (a-0.5)\omega_1 - (a+0.5)\omega_2)} - e^{-j\tau (\omega - (a+0.5)\omega_1 + (a-0.5)\omega_2)}] d\tau
\]

(27)

Crossterms will be at $\omega = (a + 0.5)\omega_2 - (a - 0.5)\omega_1$ and $\omega = -(a - 0.5)\omega_2 + (a + 0.5)\omega_1$. 
Some Other Distributions

- Rihaczek’s Complex Energy Density: The energy of a complex deterministic signal over finite ranges of $t$ and $\omega$, $C(t, \omega) = \frac{1}{\sqrt{2\pi}} s(t) S^*(\omega) e^{-j\omega t}$. The kernel associated with this distribution is $\phi(\theta, \tau) = e^{jr\tau/2}$. It satisfies the energy preservation and marginals.

- Running Energy Distribution: Defines a running spectrum. Define $S^-(t, \omega) = |\int_{-\infty}^{t} s(\tau) e^{-j\omega \tau} d\tau|^2$.

- Page Distribution: Differentiate Running Energy Distribution, $|S^-(t, \omega)|^2$, with respect to time. $P(t, \omega) = 2Re(s^*(t) S^-(t, \omega) e^{j\omega t})$. Its kernel is $\phi(\theta, \tau) = e^{jr|\tau|/2}$. It satisfies the marginals. The future does not affect the past and hence the longer a frequency exists, the larger the intensity of that frequency as time increases.
There is a lack of a single TFD that is ’best’ for all signals. Therefore, perform optimization with respect to a cost function.

There are different adaptive representations.

Adaptive STFT: Spectrogram performs best when the duration of the analysis window matches the local signal components. An adaptive-window STFT is defined as

$$\int x(\tau) w_{t,\omega}(\tau - t) e^{-j\omega \tau} d\tau$$

where

$$w_{t,\omega}(\tau) = (-2Re(c_{t,\omega})/\pi)^{1/4} \exp(c_{t,\omega}(\tau - t)^2) e^{-j\omega \tau},$$

where the real part of $c_{t,\omega}$ controls the time duration and the imaginary part determines the chirp rate. This parameter is adapted at each time-frequency location to maximize local concentration.
Adaptive Quadratic Representations: Given a signal and its ambiguity function, find the optimal kernel such that:

\[
\max_{\phi} \int \int |A(\theta, \tau)\phi(\theta, \tau)|^2 d\theta d\tau
\]

subject to \(\phi(0,0) = 1\)

\(\phi(\theta, \tau)\) is radially nonincreasing

\[
\int \int |\phi(\theta, \tau)|^2 d\theta d\tau \leq \alpha, \alpha \geq 0
\]

The optimal kernel passes auto-components and suppresses cross-terms. The constraints force the kernel to be a lowpass filter of fixed volume, \(\alpha\). Usually, \(1 \leq \alpha \leq 5\).

This idea can be generalized to different shape kernels such as radially Gaussian kernel.
Discrete-Time Implementation

Similar to discrete-time Wigner distribution:

$$C(n, \omega) = \sum_{m} R(n, m) \ast_{n} \psi(n, m) e^{-j\omega m}$$  \hspace{1cm} (29)$$

Convolve the local autocorrelation function with the kernel $\psi(n, m)$ along $n$ and take the DTFT with respect to time lag, $m$.

The time lag index $m$ takes the values,

$$\ldots, -2, -1, 0, 1, 2, \ldots$$

and the time shift index $n$ takes the values $$\ldots, -1, -0.5, 0, 0.5, 1, \ldots.$$  

A commonly used RID kernel is the binomial kernel

$$\psi(n, m) = \frac{1}{2^{|m|}} \binom{|m|}{k} \delta(n - |m| + k).$$