A Parameter-free Kernel Design Based on Cumulative Distribution Function for Correntropy

Jongmin Lee, Pingping Zhu, and Jose C. Principe

Abstract—This paper proposes a parameter-free kernel that is translation invariant and positive definite. The new kernel is based on the data cumulative distribution function (CDF) that provides all the statistical information about the observed samples. Without an explicit kernel size parameter, this novel kernel is used to define the autocorrentropy function, which is a generalized similarity measure, and spectral density estimator. Numerical examples show that the proposed method provides comparable performance to the existing Gaussian kernel with optimized kernel size.

I. INTRODUCTION

Given a time series, the autocorrelation function is generally considered as a similarity measure between the data set and a shifted copy of the data. If the data is periodic, the autocorrelation function is also periodic. However, in many practical applications where non-Gaussianities and nonlinearities exist, there may be a need to use alternative measures of similarity.

In [1] [2] [3], a generalized correlation was proposed, which is named correntropy and includes a sum of even-order moments of the data when the Gaussian kernel is utilized. By analogy with the Wiener-Khinchin theorem [4] the Fourier transform of the autocorrentropy also exists yielding a new spectral estimator that is not restricted to second-order moments. But a free parameter, i.e., kernel size of the Gaussian kernel, should be carefully selected.

In this paper we propose a new parameter-free kernel for the autocorrentropy function. The proposed kernel utilizes a translation invariant function that is based on the cumulative distribution function (CDF) of data. This allows the kernel to adjust its size automatically to the data statistics and to reflect the full statistical information of the given data. It is proved that the proposed kernel is translation invariant and that the correntropy spectral density with this kernel can be estimated by the Fourier transform. Numerical examples for the detection of a sine wave in Gaussian noise show the autocorrentropy is superior to the autocorrelation function in non-Gaussian noise close to it in Gaussian noise. Note that correntropy with this new kernel does not involve any free parameter, just like the autocorrelation.

The rest of this paper is organized as follows. Section II describes related work, namely autocorrentropy with its advances and limitations. In Section III, we design a new parameter-free kernel using CDF and convolution theorem. This kernel can be used for spectral density estimation. In Section IV, numerical examples are introduced to demonstrate the performance of the new kernel. Finally we conclude this paper in Section V.

II. RELATED WORK

Given a real strictly stationary random process \( \{x(t), t \in T\} \) with \( t \) denoting time and \( T \) being an index set of interest, the generalized autocorrelation as a similarity measure, called autocorrentropy function, is defined by

\[
V_x(t, s; \sigma) = E[\kappa_\sigma(x_t, x_s)]
\]

where \( \kappa_\sigma(\cdot, \cdot) \) is a positive definite kernel and \( \sigma \) is a kernel size parameter that is selected by the user. And the centered autocorrentropy function, which is the generalized autocovariance function, is defined by

\[
U_x(t, s; \sigma) = E_{x_t, x_s}[\kappa_\sigma(x_t, x_s)] - E_{x_t} E_{x_s}[\kappa_\sigma(x_t, x_s)].
\]

When a translation invariant kernel such as Gaussian is utilized

\[
\kappa_\sigma(x_t, x_s) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_t - x_s)^2}{2\sigma^2}\right),
\]

the centered correntropy function \( U_x(t, s; \sigma) \) is a function of the lag \( \tau = t - s \). By analogy with the Wiener-Khinchin theorem the correntropy spectral density is defined [3] as

\[
S(\omega; \sigma) = \int_{-\infty}^{\infty} U_x(\tau; \sigma) e^{-j\omega \tau} d\tau.
\]

The Fourier transform of \( U_x(\tau; \sigma) \) projects the generalized energy of periodic patterns onto the corresponding eigenfunction \( e^{-j\omega \tau} \) with frequency \( \omega \). Note that the spectral representation is built from using all even moments of the random variable [3].

The correntropy with Gaussian kernel may extract the periodic structure from the given data much better than autocorrelation, when corrupted by non-Gaussian noise. However, a free parameter - the kernel size \( \sigma \), which is inherent in most translation invariant kernel methods, should be carefully selected. In fact, correntropy with the Gaussian kernel is a function of two arguments: the lag and the kernel size that plays the role of a scale parameter for the similarity [3]. In [5] a triangular kernel that is parameter-free was introduced. The kernel normalizes the observed samples by the extreme values. This kernel is translation invariant and positive definite but it may not reflect the full statistical information of data.
The negative side of $g(x, x')$ is illustrated in the upper panel of Fig. 1 for various distributions (the positive side is symmetric). The function $g(x, x')$ in (5) only depends on the difference between $x$ and $x'$:

$$g(x, x') = g(x - x').$$

(6)

This is referred as a translation invariant function since $g$ is independent of the absolute position of $x$ but instead depend only on $x - x'$ [6].

B. Translation Invariant Positive Definite Kernel

According to Bochner’s theorem and its Fourier criterion, the translation invariant function $g(x, x') = g(x - x')$ is a positive definite kernel if the Fourier transform is nonnegative [7]. Moreover the convolution in the original space becomes the product in the Fourier domain. $g(x - x')$ is not positive definite, therefore we need to design a new kernel which is a positive definite function:

$$\kappa(x, x') = \kappa(x - x') = (g \otimes g)(x - x')$$

where $\otimes$ denotes convolution operator and $g(x - x')$ is defined in (5). Let $F$ denote the Fourier transform. $F[g(x - x')]$ is real because $g(x - x')$ is real and even. And $F[\kappa(x, x')]$ is nonnegative:

$$F[\kappa(x, x')] = F[(g \otimes g)(x - x')] = F[g(x - x')] \cdot F[g(x - x')] \geq 0. \quad (8)$$

Therefore, $\kappa(x, x')$ in (7) is a translation invariant, positive definite, and parameter-free kernel.

C. Correntropy Spectral Density Estimation

Using the parameter-free CDF kernel of (7) with lag index $\tau = t - s$ of a random process $x(t)$, the autocorrentropy function in (1) is redefined as

$$V_x(\tau) = E[\kappa(x_{t+s})] = E[(g \otimes g)(x_{t+s})] \quad (9)$$

where $g(x_{t+s}) = F_Z(|x_{t+s}|)$. Also, the centered autocorrentropy function in (2) becomes

$$U_x(\tau) = E_{x_{t+s}}[(g \otimes g)(x_{t+s})] - E_{x_{t+s}}E_{x_{t+s}}[(g \otimes g)(x_{t+s})]. \quad (10)$$

The correntropy spectral density still given by (4) remains the same except for $\sigma$:

$$S(\omega) = \int_{-\infty}^{\infty} U_x(\tau) e^{-j\omega \tau} d\tau. \quad (11)$$

D. Existence of Correntropy and its Spectral Density

We further show that for any strictly stationary random process the correntropy exists and so does its spectral density by the Fourier transform. The absolute value of the correntropy with the probability density function $f$ can be bounded.
as follows
\[ |V_0(x)| = \left| \int_{-\infty}^{\infty} \kappa(x) f(x) dx \right| \]
\[ \leq \int_{-\infty}^{\infty} |\kappa(x)| |f(x)| dx \leq \|\kappa(x)\| \|f(x)\| \]
\[ (12) \]

where \( \kappa(x) \) is a translation invariant positive definite kernel and \( x \) is an arbitrary random variable of the stationary random process at lag \( \tau \). Both of Gaussian and CDF kernels are finite. When \( \kappa \) is the Gaussian kernel, the correntropy always exists because the Gaussian function is absolutely integrable \((L^1)\) so that correntropy values are finite. We now show that the CDF kernel we built also exists in the \( L^1 \) sense.

\[ \|\kappa(x)\|_1 = \int_{-\infty}^{\infty} |(g \otimes g)(x)| dx \]
\[ = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} g(u) g(x-u) du \right] dx \]
\[ = \int_{-\infty}^{\infty} g(u) \left[ \int_{-\infty}^{\infty} g(x-u) dx \right] du \]
\[ = \int_{-\infty}^{\infty} g(u) du \int_{-\infty}^{\infty} g(x) dx \]
\[ (13) \]

where \( g \) is defined in (5), the integral of \( g \) in (13) is

\[ \int_{-\infty}^{\infty} g(x) dx = 2 \int_{-\infty}^{0} F_Z(-x) dx \]
\[ = 2 \int_{-\infty}^{0} F_Z(z) dz \]
\[ (14) \]

where \( Z = -|X| \) and \(-\infty < Z \leq 0\). Since the integral of \( F_Z(z) \) is finite, the CDF kernel is absolutely integrable \((L^1)\). And thus the correntropy with the CDF kernel also always exists.

When one is interested in spectral estimation, another issue is the integrability of correntropy over the lags, which is a sufficient condition for the Fourier transform to exist. Practically this case is not important since we always use finite windows, so the integral always exists. Our experience shows that for large lags the correntropy decays faster than the autocorrelation function, what can be expected because of the kernel that decreases on average when the two values of the argument are different. However we have not been able to prove this fact mathematically. The only step that is mandatory for correntropy is to subtract its mean value that is always positive because of the positive nature of the kernel, i.e., we should always work with the centered correntropy for spectral estimation.

**IV. Numerical Tests**

We present numerical examples to demonstrate the performance of the CDF kernel. The observed data \( x(t) \), from which we estimate the spectral density, is a sum of a real sinusoidal \( u(t) = a_1 \sin(\omega t) \) and zero-mean symmetric white noise \( n(t) \) that can be either Gaussian or \( \alpha \)-stable distributed noise. The amplitude \( a_1 \) depends on the signal-to-noise level.

![Normalized spectra estimates in various noises. GSNR=3dB. ω = 0.1. Due to different scales of kernels, each spectral density is normalized by the highest peak. The ensemble average over 50 estimates are presented. The product kernel indicates autocorrelation. The other kernels are for autocorrentropy. Note that the estimates of the CDF kernel do not change much in various noise environments while the others varies a lot.](image-url)
in the tests and \( \omega \) is known. A simple binary hypothesis testing is considered to see the presence of the signal \( u(t) \) that is buried in impulsive noise with low signal-to-noise ratio. We define the generalized SNR (GSNR) due to \( \alpha \)-stable noise with \( \alpha < 2 \): \( \text{GSNR} = P_u / \gamma^2 \) where \( P_u \) and \( \gamma \) are respectively the power of \( u(t) \) and scale parameter (or dispersion ratio) of \( \alpha \)-stable noise. When \( \alpha = 2.0 \), the \( \alpha \)-stable noise is equivalent with Gaussian noise whose variance is \( 2\gamma^2 \).

Spectral densities are estimated for the normalized frequency \( \omega = 0.1 \) in four different noises: \( (\alpha = 1.5, \gamma = 1), (\alpha = 1.5, \gamma = 5), (\alpha = 2, \gamma = 1), \) and \( (\alpha = 2, \gamma = 5) \). Fig. 2 compares the corretnropy spectral densities of the CDF kernel with those of other kernels. We can compare the relative height between the peak and noise floor level. When the symmetric \( \alpha \)-stable noise is \( (\alpha = 1.5, \gamma = 1), \) the CDF kernel and the Gaussian kernel with \( \sigma = 2.6 \) can estimate lower noise floor than others. However, for \( \gamma = 5 \) the kernel with \( \sigma = 2.6 \) is no longer the best but \( \sigma = 10 \) is better. When the noise is Gaussian distribution \( (\alpha = 2.0) \) in Fig. 2(b), the product kernel is the best for both of \( \gamma = 1 \) and \( \gamma = 5 \). While the Gaussian kernel with larger kernel size is close to the best one, the kernel with \( \sigma = 2.6 \) for \( \gamma = 5 \) leads to poor performance. But the CDF kernel automatically adjust the kernel size so that it can be quite close to those best kernels no matter the observed data distribution or scale is. Here the kernel sizes \( \sigma = 2.6 \) and 10 were considered after scanning over a range as shown in the upper panel of Fig. 4.

Fig. 3 shows the results of a simple statistical test that is identifying the presence of a sine wave in noise. The binary hypothesis is considered: \( H_0 : x(t) = n(t) \) and \( H_1 : x(t) = u(t) + n(t) \). The presence of \( u(t) \) at frequency \( \omega \) is detected at low GSNR. The decision statistic \( \zeta_0 \) is obtained from the peak value of the spectrum density at \( \omega \). The probability of detection \( P_d(\omega) = Pr(\zeta_0 > \epsilon_0 | H_1) \) and the probability of false alarm \( P_{fa}(\omega) = Pr(\zeta_0 > \epsilon_0 | H_0) \) are compared for the likelihood ratio test with a certain threshold \( \epsilon_0 \) that is determined by \( P_{fa} \). For the sine wave with normalized frequency \( \omega = 0.1 \) and the different noises, 10,000 Monte Carlo experiments were performed. The presence of the signal at the desired frequency \( \omega \) may not be detected with the autocorrelation when noise power spectrum is too high. Especially, in Fig. 3(a) the ROC curves show the autocorrelation function could not detect well because of impulsiveness and higher noise floor. One can relate these results with spectral estimation in Fig. 2(a). The corretnropy with Gaussian kernel achieves better detection probability if we select the appropriate kernel sizes: \( \sigma = 2.6 \) for \( (\alpha = 1.5, \gamma = 1) \); about \( \sigma = 10 \) for \( (\alpha = 1.5, \gamma = 5) \). However, the CDF kernel eliminates such a free parameter while it is close to the highest detection probability regardless of the distribution of the observed data. Even when the noise is Gaussian, the CDF kernel shows comparable performance as illustrated in Fig. 3(b). In order to provide more insight about the performance of CDF kernel, the upper panel of

*Fig. 3. ROC curves. GSNR = -12dB. Triangular kernel, which is also parameter-free and briefly described in Section II, is compared.*
Fig. 4. Detection probability vs. kernel size for Gaussian kernel in three different noise distributions when $P_{fa} = 0.1$ (up). To compare with Fig. 3(a), the ROCs for the case of $(\alpha = 1.3, \gamma = 3)$ are also presented (down).

Fig. 4 illustrates the optimized kernel sizes for the detection, and the lower one compares the CDF kernel with others in another non-Gaussian noise environment $(\alpha = 1.3, \gamma = 3)$.

V. CONCLUSION

We have designed a new CDF-based kernel. The CDF contains all the statistical information about the given data so that it can allow the new kernel to automatically adjust the kernel size parameter to the data statistics. It has shown that the kernel is translation invariant, positive definite, and parameter free. The autocorrentropy function based on this kernel can estimate spectral densities with comparable performance to the existing methods such as the autocorrelation and the autocorrentropy that adopts the Gaussian kernel with optimized kernel size. Further theoretical work is needed to study the conditions under which the correntropy spectral estimation exists, and if the CDF kernel includes the sum of moments of the data statistics. But overall this methodology is ready to be applied for nonparametric spectral estimation applications. Recall that correntropy spectral estimation is not limited by the use of second-order moments as the conventional power spectrum.

REFERENCES


