Discrete Stochastic
Time-Frequency Analysis
and Cepstrum Estimation

Johan Sandberg
A Comparison Between Different Discrete Ambiguity Domain Definitions in Stochastic Time-Frequency Analysis

Johan Sandberg and Maria Hansson-Sandsten, Member, IEEE

Abstract—The ambiguity domain plays a central role in estimating the time-varying spectrum and in estimating the covariance function of nonstationary random processes in continuous time. For processes in discrete time, there exist different definitions of the ambiguity domain, but it is well known that neither of these definitions perfectly resembles the usefulness of the continuous ambiguity domain. In this paper, we present some of the most frequently used definitions of the ambiguity domain in discrete time: the Claesen-Mecklenbräucker, the Jeong-Williams, and the Nuttall definitions. For the first time, we prove their equivalence within some necessary conditions and we present theorems that justify their usage.

Index Terms—Ambiguity domain, Claesen-Mecklenbräucker, covariance function estimation, discrete-time discrete-frequency, Jeong-Williams, nonstationary random processes, Nuttall, time-frequency analysis.

I. INTRODUCTION

RESEARCH within the field of stochastic time-frequency analysis is directed towards two different objectives. The first is to define the time-varying spectrum of a random process in terms of its distribution, a question that despite decades of work, remains largely open. This leads to ambiguity domain definitions that can be well estimated, see for example [8]. After the mean function has been removed, \( \{ x(t), t \in \mathbb{Z} \} \) is a zero-meaned random process in discrete time with finite moments and finite support: \( x(t) = 0 \quad \forall t \notin \{1, \ldots, n\} \). From a statistical point of view, the most reasonable covariance function estimator is

\[
R_H(s,t) = \sum_{k=1}^{\max(s,t)} H(k, s-t) x(s+k) x(t+k)^* \tag{1}
\]

where \( H(k, r) \in \mathbb{C}, k \in \{ -n+1 \pm |r|, \ldots, n-1-|r| \} \), \( r \in \{ -n+1, \ldots, n-1 \} \) is a, usually bell-shaped, kernel function that has to be suitably chosen. [9]. Some care has to be taken in order for this estimate to be nonnegative definite, but as this problem has appropriate solutions, [10], we will not discuss it further. The covariance function estimator, \( R_H(s,t) \), possesses great flexibility and one may argue that any sound nonparametric covariance function estimator should be of this form. The continuous analog \( R_{H_{C}}(s,t) \) to (1), where summation has been replaced by integration, for a process \( x_{cont} \) in continuous time, can be written in the ambiguity domain

\[
R_{H_{C}}(s,t) = \int \psi(v, s-t) x_{cont}(v + (s-t)/2)
\]
The ambiguity processes in continuous time

\[ A_x(\nu, \tau) = \int_{-\infty}^{\infty} x(t + \frac{\tau}{2}) x(t - \frac{\tau}{2})^* \ e^{-i2\pi t\nu} \ dt \]
The Claasen-Mecklenbräuker ambiguity process

\[ AC(\nu, \tau) = \sum_{t=-\infty}^{\infty} x(t + \frac{\tau}{2}) x(t - \frac{\tau}{2})^* e^{-i2\pi t\nu}, \tau \text{ even} \]
The Nuttall ambiguity process

\[ A_N(\nu, \tau) = \sum_{u \in [-0.5 + |\nu| : 1/2n : 0.5 - 1/2n - |\nu|]} X(u + \nu)X(u - \nu)^* e^{i2\pi u \tau} \]

\[ X(f) = \sum_{t=1}^{n} x(t)e^{-i2\pi ft} \]
The Jeong-Williams ambiguity process

\[ A_J(\nu, \tau) = \begin{cases} \sum_{t=1,\ldots,n} x(t + \frac{\tau}{2}) x(t - \frac{\tau}{2})^* & \tau \text{ even} \\ \sum_{t=0.5,\ldots,n-0.5} x(t + \frac{\tau}{2}) x(t - \frac{\tau}{2})^* & \tau \text{ odd} \end{cases} \]
The ambiguity processes in continuous time

\[ A_x(\nu, \tau) = \int_{-\infty}^{\infty} x(t + \frac{\tau}{2}) x(t - \frac{\tau}{2})^* \ e^{-i2\pi t \nu} \ dt \]

A property of the ambiguity process

\[ \mathcal{F}_\phi(\nu, \tau)A_x(\nu, \tau) = \int_{-\infty}^{\infty} H(k, s - t)x(s + k)x(t + k)dk \]
Optimal stochastic discrete time–frequency analysis in the ambiguity and time-lag domain

Johan Sandberg*, Maria Hansson-Sandsten
Lund University, Centre for Mathematical Sciences, Box 118, SE-221 00 Lund, Sweden

ABSTRACT

In stochastic time–frequency analysis, the covariance function is often estimated from only one observed realization with the use of a kernel function. For processes in continuous time, this can equivalently be done in the ambiguity domain, with the advantage that the mean square error optimal ambiguity kernel can be computed. For processes in discrete time, several ambiguity domain definitions have been proposed. It has previously been reported that in the Jeong–Williams ambiguity domain, in contrast to the Nutall and the Claassen–McDerrmott ambiguity domain, any smoothing covariance function estimator can be represented as an ambiguity kernel function. In this paper, we show that the Jeong–Williams ambiguity domain cannot be used to compute the mean square error (MSE) optimal covariance function estimate for processes in discrete time. We also prove that the MSE optimal estimator can be computed without the use of the ambiguity domain, as the solution to a system of linear equations. Some properties of the optimal estimator are derived.

© 2010 Elsevier B.V. All rights reserved.
Covariance function estimator in continuous time

\[ R_H(s, t) = \int_{-\infty}^{\infty} H(k, s - t)x(s + k)x(t + k)^* \, dk \]
Covariance function estimator in continuous time

\[ R_H(s, t) = \int_{-\infty}^{\infty} H(k, s-t)x(s+k)x(t+k)^*dk \]

Mean square error optimal kernel

\[ H_{\text{MSE-opt}} = \mathcal{F} \left| \mathbb{E} [A_x(v, s-t)] \right|^2 \frac{\mathbb{E} \left[ |A_x(v, s-t)|^2 \right]}{\mathbb{E} \left[ |A_x(v, s-t)|^2 \right]} \]
Mean square error optimal kernel in discrete time

\[
\frac{1}{|\mathcal{K}_\tau|} \sum_{k \in \mathcal{K}_\tau} \sum_{t = \max(1, 1 - \tau)}^{\min(n, n - \tau)} \rho_x(t + l, \tau, t + k, \tau) H_{x_{opt}}(k, \tau) \\
= \sum_{t = 1}^{n} r_x(\tau + t, t)r_x(\tau + t + l, t + l)^* \quad \forall l \in \mathcal{K}_\tau
\]

\[\mathcal{K}_\tau = \{-n + 1 + |\tau|, \ldots, n - 1 - |\tau|\}\]
Optimal Non-Parametric Covariance Function Estimation for any Family of Random Processes

Johan Sandberg and Maria Hansson-Sandsten

Lund University, Sweden

Abstract

A covariance function estimate of a zero-mean non-stationary random process in discrete time is accomplished from one observed realization by weighting observations with a kernel function. Several kernel functions have been proposed in the literature. In this paper we prove that the mean square error (MSE) optimal kernel function for any parameterized family of random processes can be computed as the solution to a system of linear equations. Even though the resulting kernel is optimized for members of the chosen family, it seems to be robust in the sense that it is often close to optimal for many other random processes as well. We also investigate a few examples of families, including a family of locally stationary processes, non-stationary AR-processes and chirp-processes, and kernels.
Mean square error optimal kernel for a family of processes in discrete time

\[
R_{x; H}(s, t) = \frac{1}{|\mathcal{K}_\tau|} \sum_{k \in \mathcal{K}_{s-t}} H(k, s-t)x(s+k)x(t+k)^*
\]

\[
H_{x Q-\text{opt}} = \arg \min_{H \in \mathcal{H}} \int_{Q} \sum_{(s,t) \in T_n^2} \mathbb{E} \left[ |r_{x_q}(s, t) - R_{x_q; H(s, t)}|^2 \right] dF_Q(q)
\]
Multitaper Estimation of Frequency-Warped Cepstra With Application to Speaker Verification

Johan Sandberg, Maria Hansson-Sandsten, Member, IEEE, Timo Kinnunen, Rahim Saeidi, Student Member, IEEE, Patrick Flandrin, Fellow, IEEE, and Pierre Borgnat, Member, IEEE

Abstract—Usually the mel-frequency cepstral coefficients are estimated either from a periodogram or from a windowed periodogram. We state a general estimator which also includes multitaper estimators. We propose approximations of the variance and bias of the estimate of each coefficient. By using Monte Carlo computations, we demonstrate that the approximations are accurate. Using the proposed formulas, the peak matched multitaper estimator is shown to have low mean square error (variance) on speech-like processes. It is also shown to perform slightly better in the NIST 2006 speaker verification task as compared to the Hamming window conventionally used in this context.

Index Terms—Cepstral analysis, MFCC, multiple windows, multitapers, speaker verification, speech analysis.

I. INTRODUCTION

The cepstrum was introduced by Bogert, Healy and Tukey in the early 1960s [1]. It is defined as the inverse Fourier transform of the log-spectrum of a stationary random process [2]. The cepstrum has become a fundamental tool in many applications. One example is the estimation of the periodic spectrum of speech. A well-known estimator of this spectrum is the windowed periodogram [4]. The windowed periodogram has low bias in general, but it still suffers from high variance. Therefore, one may consider using a so-called multitaper spectral estimator instead. A multitaper spectral estimator is an average of windowed periodograms using different orthogonal windows (aka tapers), e.g., the Thomson [5], the sine [6], and the peak matched multitapers [7]. The multitaper spectrum estimator is known to have low variance, but has not gained much attention in MFCC estimation [8]. One reason may be that the statistical properties of the multitaper MFCC estimator have previously not been investigated. It is our purpose to address this issue in this letter.

In Section II of this letter we state the general form of an MFCC estimator, which includes the MFCC computed from the periodogram, the windowed periodogram, the Bartlett and the Welch method, as well as multitaper spectrum estimators. The statistical properties of the cepstrum computed from the periodogram are well known [2], [9], [10]. However, the bias and variance of the cepstrum or of the MFCCs computed from the multitaper spectrum estimator have not been, to the best of our knowledge, studied so far. In Section III-A, we therefore
A recorded sound signal
The cepstrum is the Fourier transform of the log-spectrum.

- 20 ms Frame
- log-Spectrum
- Cepstrum
The cepstrum is the Fourier transform of the log-spectrum.
The multitaper spectrum estimator

Periodogram
Periodogram
Periodogram
Periodogram
Periodogram
Average
The multitaper cepstrum estimator

\[ \hat{c}_M = \frac{1}{m} \Phi^H \log(M\hat{s}) \]

\[ \hat{s} = [\hat{s}(0) \hat{s}(1) \ldots \hat{s}(n - 1)]^T \]

\[ \hat{s}(p) = \sum_{j=1}^{k} \lambda(j) |w_j^T \Psi_p x|^2, \ p = 0, \ldots, n - 1 \]
Bias and variance of the multitaper cepstrum estimator

\[
\text{bias } [\hat{c}_M] \approx \frac{1}{m} \Phi^H \left( \log \left( \frac{\text{ME } [\hat{s}]}{\text{Ms}} \right) - \frac{\text{diag } (\text{MV } [\hat{s}] \text{M}^T)}{2 (\text{ME } [\hat{s}])^2} \right)
\]

\[
\text{V } [\hat{c}_M] \approx \frac{1}{m^2} \Phi^H \frac{\text{MV } [\hat{s}] \text{M}^T}{\text{ME } [\hat{s}] \text{E } [\hat{s}]^T \text{M}^T} \Phi
\]
Bias

Variance
Mean square error

Coeficient

Rectangular
Hanning
Thomson, k=4
Thomson, k=8
Thomson, k=12

Mel−Cepstrum coefficient number
Speaker verification result

![Graph showing the relationship between false alarm probability and miss probability for different methods. The graph includes a line for Hamming and a dashed line for Multitapers. The y-axis represents miss probability in percent, ranging from 0 to 40, and the x-axis represents false alarm probability in percent, ranging from 0 to 40. The graph illustrates the performance comparison of the two methods.]
Optimal Cepstrum Smoothing

Johan Sandberg and Maria Hansson-Sandsten

Lund University, Sweden

Abstract

The cepstrum of a random process has proved to be a useful tool in a wide range of applications. The common cepstrum estimator based on the periodogram suffers from large variance, and, to a smaller degree, from bias. The variance can be reduced by smoothing. However, the smoothing may be performed in four different domains: the covariance, the spectral, the log-spectral and the cepstral domain. We present the mean square error (MSE) optimal smoothing kernels in each domain for estimation of the cepstrum. The lower MSE bound of each of the four families of estimators are compared. We also demonstrate how the four MSE optimal estimators differ in robustness.
Definition of the cepstrum

\[ c(q) = \frac{1}{N} \sum_{p=-n+1}^{n-1} \log(S(p)) e^{i2\pi \frac{p}{N} q} \]

\[ S(p) = \sum_{\tau=-n+1}^{n-1} r(\tau) e^{-i2\pi \frac{p}{N} \tau} \]

\[ r(\tau) = \mathbb{E} [x(t)x(t+\tau)] \]
Un-smoothed estimator

\[ \hat{c}(q) = \frac{1}{N} \sum_{p=-n+1}^{n-1} \log \left( \hat{S}(p) \right) e^{i2\pi \frac{p}{N} q} \]

\[ \hat{S}(p) = \sum_{\tau=-n+1}^{n-1} \hat{r}(\tau) e^{-i2\pi \tau \frac{p}{N}} \]

\[ \hat{r}(\tau) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-|\tau|} x(t)x(t + |\tau|) & |\tau| < n \\ 0 & \text{otherwise} \end{cases} \]
THANK YOU