

Research Article

Optimal Nonparametric Covariance Function Estimation for Any Family of Nonstationary Random Processes

Johan Sandberg (EURASIP Member) and Maria Hansson-Sandsten (EURASIP Member)

Division of Mathematical Statistics, Centre for Mathematical Sciences, Lund University, 221 00 Lund, Sweden

Correspondence should be addressed to Johan Sandberg, sandberg@maths.lth.se

Received 28 June 2010; Revised 15 November 2010; Accepted 29 December 2010

Academic Editor: Antonio Napolitano

Copyright © 2011 J. Sandberg and M. Hansson-Sandsten. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A covariance function estimate of a zero-mean nonstationary 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, nonstationary AR-processes, and chirp processes, and their respective MSE optimal kernel functions.

1. Introduction

In several applications, including statistical time-frequency analysis [1–4], the covariance function of a nonstationary random process has to be estimated from one single observed realization. We assume that the complex-valued process, which we denote by $\{x(t), t \in \mathbb{Z}\}$, is in discrete time and has finite support: $x(t) = 0$ for all $t \notin T_n = \{1, \dots, n\}$. Most often, the mean of the process is assumed to be known or already estimated and, hereby, we can, without loss of generality, assume that the mean of the process is zero. An estimate of the covariance function defined and denoted by $r_x(s, t) = E[x(s)x(t)^*]$ is then accomplished by a weighted average of observations of $x(s+k)x(t+k)^*$ with different weights for different k , [5, 6], where $*$ denotes complex conjugate. Presumably, the weights, also known as the kernel function, are allowed to vary with the time-lag $\tau = s - t$. We denote and define this estimator by $R_{x;H} : T_n^2 \mapsto \mathbb{C}$:

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

where \mathcal{K}_τ is the set $\{-n+1+|\tau|, \dots, n-1-|\tau|\}$, and H is a kernel function which belongs to the set $\mathcal{H} = \{H : \mathcal{K}_\tau \times \mathcal{T} = \{-n+1, \dots, n-1\} \mapsto \mathbb{C}\}$ of all possible kernel functions, and

where we denote the cardinality of a set S by $|S|$. Some care has to be taken in order for this estimate to be nonnegative definite, but as this problem has appropriate solutions [7], we will not discuss it further. Naturally, one wishes to choose the kernel function H carefully so that the estimate does not suffer from large bias or variance.

If nothing except zero mean is assumed about the process, no other estimator of $r(s, t)$ than $x(s)x(t)^*$ can be justified. This is equivalent with $H(k, \tau) = 1$ for $k = 0$ and zero otherwise. We will assume that there is some a priori knowledge about the process. If, for example, the process is quasistationary in the sense that $r(s, t) \approx r(s+\delta, t+\delta)$ for integers $|\delta| < D$, then it may be wise to use a kernel function which is “bell-shaped” in the k -direction with a bandwidth proportional to D . And if, for example, it is known that the process decorrelates at large time lags, meaning that $r(s, s+\tau) \approx 0$ for $|\tau| > T$, then it makes sense to use a kernel function $H(k, \tau) \approx 0$ for $|\tau| > T$, [8].

The kernel function that gives the least mean square error (MSE) of the estimate (squared bias plus variance) for processes in continuous time was derived by Sayeed and Jones in the ambiguity domain as a function of the process characteristics, including second- and fourth-order moments up to amplitude scaling, frequency shift, and

time shift [9]. Their result has been used in, for example, [10, 11]. If the formula found by Sayeed and Jones is used in the discrete ambiguity domain, the resulting covariance estimator will not be MSE optimal, as discussed in [12]. However, for these processes the MSE optimal covariance function estimator can be computed in time-lag domain [12]. This can be used if one can, by prior knowledge, construct a random process for which the optimal kernel function is similar to the optimal kernel function for $\{x(t)\}$. In this paper, we prove that the MSE optimal kernel function for any parameterized family of random processes with an a priori parameter distribution can be computed by solving a system of linear equations. The solution is optimal in the sense that there is no other choice of kernel function H which gives a covariance function estimate $R_{x;H}$ with less expected MSE if the observed realization belongs to a process in the chosen family with the presumed parameter distribution.

The result derived in this paper is useful when so little is known about the random process that a nonparametric covariance function estimator of the form (1) ought to be used (rather than estimating parameters in a model). For such a situation, one has to decide which kernel function H to use. This choice of H must be guided by some prior knowledge about the random process. If this knowledge can be condensed into a parameterized family of processes where we can assign a probability distribution to the parameter space, then the optimal kernel function for the whole family of processes, computed as described in Section 2, can be used. It is important to stress that we do not need to assume that the realization we are about to observe is a realization from a process in this parameterized family. Rather, we believe that the realization comes from a process which has a suitable kernel function in common with some processes in the family.

The remainder of this paper is organized as follows. The MSE optimal solution is presented in Section 2. An example with a family of locally stationary processes is described in Section 3.1 and a family of nonstationary AR(1)-processes is described in Section 3.2. In Section 3.3, we compute the MSE optimal kernel function for a family of chirp processes which we use on a set of heart rate variability data. Conclusions and final remarks are given in the last section. Proofs are found in the Appendix.

2. The MSE Optimal Kernel Estimator for a Family of Processes

Let (Ω, \mathcal{F}, P) be a fixed probability space. Let $\{x_q(t), t \in \mathbb{Z}, q \in \mathcal{Q}\}$ be a family of random variables parameterized by $t \in \mathbb{Z}$ ("time") and $q \in \mathcal{Q}$, where \mathcal{Q} ("parameter space for the family of random processes") is a fixed subset of \mathbb{R}^L , for some fixed integer L . For a fixed $q \in \mathcal{Q}$, we think of $\{x_q(t), t \in \mathbb{Z}\}$ as a random process in discrete time, and we assume that this process has the following properties: (a) it has zero mean: $E[x_q(t)] = 0$, for all $t \in \mathbb{Z}$, (b) it has finite moments, and (c) it has finite support: $x_q(t) = 0$, for all $t \notin T_n = \{1, \dots, n\}$, where n is a fixed integer. We also assume that $x_q(t)$ and $x_p(s)$ are independent for $q \neq p$.

Now, let Q be a random element in \mathcal{Q} with distribution F_Q . Conditional on $Q = q$, we have observed a realization of the process $\{x_q(t), t \in \mathbb{Z}\}$, and we shall estimate the covariance function. Since Q is random, that is we do not know which process in the family we have an observation of, we would like to compute a kernel function which gives a low MSE for as many random processes within the family as possible. Also, we would like to take into account that we have a probability distribution of Q and, hence, it seems natural to let the optimization give a higher weight to members of the family with a high probability according to the distribution F_Q . In order to achieve this, we need a few definitions. The covariance function of $\{x_q(t), t \in \mathbb{Z}\}$, for a fixed $q \in \mathcal{Q}$, is denoted by $r_{x_q}(s, t)$, and, for a random element Q in \mathcal{Q} , we let $r_{x_Q}(s, t)$ denote the covariance function conditional on Q , that is,

$$r_{x_Q}(s, t) = E[x_Q(s)x_Q(t)^* | Q]. \quad (2)$$

Given a kernel function H , the estimator of $r_{x_Q}(s, t)$ is given by

$$R_{x_Q;H}(s, t) = \frac{1}{|\mathcal{K}_{s-t}|} \sum_{k \in \mathcal{K}_{s-t}} H(k, s-t)x_Q(s+k)x_Q(t+k)^*, \quad (3)$$

as in (1). We are now ready to define the kernel function that minimizes the expected error of $R_{x_Q;H}$.

Definition 1. The MSE optimal kernel function for a family $\{x_Q(t), t \in \mathbb{Z}\}$ of random processes is defined and denoted by

$$H_{x_Q-\text{opt}} = \arg \min_{H \in \mathcal{H}} E \left[\sum_{(s,t) \in T_n^2} |r_{x_Q}(s, t) - R_{x_Q;H}(s, t)|^2 \right]. \quad (4)$$

Remark 1. The definition can equivalently be written as

$$\begin{aligned} H_{x_Q-\text{opt}} &= \arg \min_{H \in \mathcal{H}} \int_{\mathcal{Q}} \sum_{(s,t) \in T_n^2} E \left[|r_{x_q}(s, t) - R_{x_q;H}(s, t)|^2 \right] dF_Q(q). \end{aligned} \quad (5)$$

As stated in the following theorem, the MSE optimal kernel function can be computed by solving a system of linear equations.

Theorem 1 (MSE optimal kernel for a family of random processes). *For a fixed $\tau \in \mathcal{T}$, the MSE optimal kernel function $H_{x_Q-\text{opt}}(k, \tau)$ for the family $\{x_Q(t), t \in \mathbb{Z}\}$, is found as the solution to the following system of linear equations:*

$$\begin{aligned} & \frac{1}{|\mathcal{K}_\tau|} \sum_{k \in \mathcal{K}_\tau} H_{x_Q-\text{opt}}(k, \tau) \sum_{t=\max(1, 1-\tau)}^{\min(n, n-\tau)} \int_{\mathcal{Q}} \rho_{x_q}(t+k, \tau, t+l, \tau) dF_Q(q) \\ &= \sum_{t=\max(1, 1-\tau)}^{\min(n, n-\tau)} \int_{\mathcal{Q}} r_{x_q}(t+\tau, t) r_{x_q}(t+\tau+l, t+l)^* dF_Q(q) \quad \forall l \in \mathcal{K}_\tau, \end{aligned} \quad (6)$$

where ρ_{x_q} is the fourth-order moment of $\{x_q(t)\}$ defined by $\rho_{x_q}(t_1, \tau_1, t_2, \tau_2) = E[x_q(t_1)x_q(t_1 + \tau_1)^*x_q(t_2)x_q(t_2 + \tau_2)]$.

Proof. See appendix. \square

We note that for a fixed $\tau \in \mathcal{T}$, the theorem gives $|\mathcal{K}_\tau| = 2n - 2|\tau| - 1$ linear equations, by which we can compute $H_{x_Q-\text{opt}}(k, \tau)$ for all $k \in \mathcal{K}_\tau$ using any standard routine available for solving linear equations. The bias of the estimator $R_{x;H_{x_Q-\text{opt}}}(s, t)$ is

$$\begin{aligned} & \text{bias}\left[R_{x;H_{x_Q-\text{opt}}}(s, t)\right] \\ &= \frac{1}{|\mathcal{K}_{s-t}|} \sum_{k \in \mathcal{K}_{s-t}} H_{x_Q-\text{opt}}(k, s-t)r_x(s+k, t+k) - r_x(s, t), \end{aligned} \quad (7)$$

and the variance is

$$\begin{aligned} & V\left[R_{x;H_{x_Q-\text{opt}}}(s, t)\right] \\ &= \frac{1}{|\mathcal{K}_{s-t}|^2} \sum_{k_1 \in \mathcal{K}_{s-t}} \sum_{k_2 \in \mathcal{K}_{s-t}} H_{x_Q-\text{opt}}(k_1, s-t)H_{x_Q-\text{opt}}(k_2, s-t) \\ & \quad \times C\left[x_q(s+k_1)x_q(t+k_1)^*, x_q(s+k_2)x_q(t+k_2)^*\right]. \end{aligned} \quad (8)$$

The MSE optimal kernel function is invariant to some manipulations of the family $\{x_Q(t)\}$, including amplitude scaling when the whole family is scaled by the same scalar. It is also invariant to frequency shift of each process in the family, in the following sense: let $\{y_q(t), t \in \mathbb{Z}, q \in \mathcal{Q}\}$ be defined by $y_q(t) = x_q(t)e^{-i2\pi f_q t}$, where $f_q \in \mathbb{R}$. Then $H_{y_Q-\text{opt}} = H_{x_Q-\text{opt}}$. This can be proved by inserting the relations $r_{y_q}(s, t) = e^{-i2\pi(s-t)f_q}r_{x_q}(s, t)$ and $R_{y_q;H}(s, t) = e^{-i2\pi f_q(s-t)}R_{x_q;H}(s, t)$ into (4). The MSE optimal kernel function is also approximately time-shift invariant in the following sense: let $\{z_q(t), t \in \mathbb{Z}, q \in \mathcal{Q}\}$, $z_q(t) = 0$ for all $t \notin T_n$ be defined by $z_q(t) = x_q(t - \delta_q)$, where $|\delta_q|$ is an integer much smaller than n . Then $H_{z_Q-\text{opt}}$ is given by

$$\begin{aligned} H_{z_Q-\text{opt}} &= \arg \min_{H \in \mathcal{H}} \int_{\mathcal{Q}} \sum_{s=1-\delta_q}^{n-\delta_q} \sum_{t=1-\delta_q}^{n-\delta_q} \\ & \quad \times E \left[\left| r_{x_q}(s, t) - \frac{1}{|\mathcal{K}_{s-t}|} \sum_{k \in \mathcal{K}_{s-t}} x_q(s+k) \right. \right. \\ & \quad \left. \left. \times x_q(t+k)^* H(k, s-t) \right|^2 \right] dF_Q(q) \\ &= \arg \min_{H \in \mathcal{H}} \int_{\mathcal{Q}} \sum_{s=1-\delta_q}^{n-\delta_q} \sum_{t=1-\delta_q}^{n-\delta_q} \\ & \quad \times E \left[\left| r_{x_q}(s, t) - R_{x_q;H}(s, t) \right|^2 \right] dF_Q(q). \end{aligned} \quad (10)$$

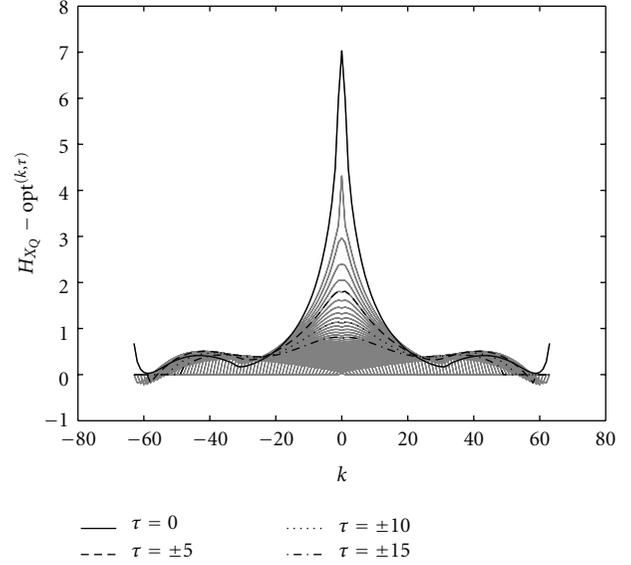


FIGURE 1: The MSE optimal kernel function for the family of locally stationary processes $\{x_{\alpha,\beta}(t), t \in \mathbb{Z}, (\alpha, \beta) \in \mathcal{Q}\}$.

We see that the only difference between (10) and (5) is that the summation in (5) is made for $s = 1, \dots, n$ and $t = 1, \dots, n$ whereas in (10), the summation is made for $s = 1 - \delta_q, \dots, n - \delta_q$ and $t = 1 - \delta_q, \dots, n - \delta_q$. Since there is only a small shift in the area for which the minimization is performed, $H_{x_Q-\text{opt}}$ will be approximately as optimal as $H_{z_Q-\text{opt}}$ on the family $\{z_q(t), t \in \mathbb{Z}, q \in \mathcal{Q}\}$.

3. Examples

3.1. A Family of Locally Stationary Processes. We will now consider a family of processes which are approximately locally stationary. Let $\{x_{\alpha,\beta}(t), t \in \mathbb{Z}, (\alpha, \beta) \in \mathcal{Q}\}$, where $\mathcal{Q} = \{(\alpha, \beta) : 0 < \alpha \leq \beta \leq 1\}$ be a set of jointly Gaussian random variables such that $x_q(t)$ and $x_p(s)$, $q \neq p$, are independent, $E[x_{\alpha,\beta}(t)] = 0$, $x_{\alpha,\beta}(t) = 0$ for all $t \notin T_n = \{1, \dots, n\}$, $r_{\alpha,\beta}(s, t) = E[x_{\alpha,\beta}(s)x_{\alpha,\beta}(t)^*] = c_{\alpha,\beta}e^{-(s-t)^2/(an)^2}e^{-(s+t-n-1)^2/(\beta n)^2}$, where $c_{\alpha,\beta}$ is a normalization factor $c_{\alpha,\beta} = (\sum_{(s,t) \in T_n^2} e^{-(s-t)^2/(an)^2}e^{-(s+t-n-1)^2/(\beta n)^2})^{-1/2}$.

Each random process $\{x_{\alpha,\beta}(t), t \in \mathbb{Z}\}$ is approximately locally stationary in Silverman's sense [13, 14]. Such processes have been widely used in the literature, see for example [10, 15]. Now, let Q be a random element of \mathcal{Q} with uniform distribution on \mathcal{Q} , that is, the density function is 2 in \mathcal{Q} and 0 otherwise. The MSE optimal kernel function for this family, computed by the use of Theorem 1, is shown in Figure 1, where $n = 64$.

The optimal kernel function, $H_{x_Q-\text{opt}}$, for this family can be compared with the optimal kernel function for each member of the family. Figure 2 shows the ratio between the MSE when $H_{x_Q-\text{opt}}$ is used and the MSE when $H_{x_{\alpha,\beta}-\text{opt}}$ is used on realizations from $\{x_{\alpha,\beta}(t), t \in \mathbb{Z}\}$, where $H_{x_{\alpha,\beta}-\text{opt}}$ is the MSE optimal kernel function for the process $\{x_{\alpha,\beta}(t), t \in \mathbb{Z}\}$.

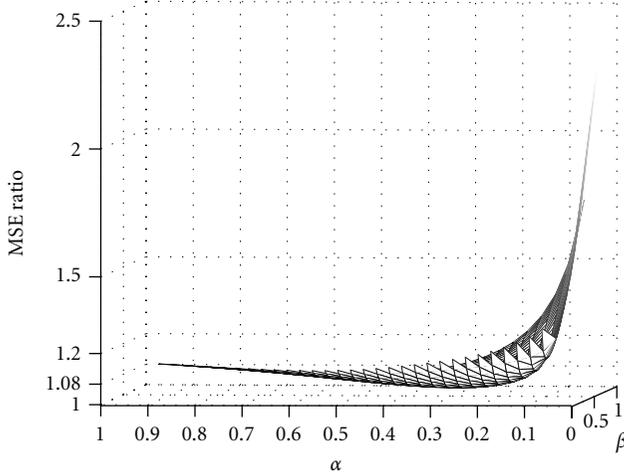


FIGURE 2: Ratio between MSE of the MSE optimal kernel for the family of locally stationary processes $\{x_Q(t)\}$ and the MSE for the kernel optimized for every $(\alpha, \beta) \in \mathcal{Q}$.

The kernel function $H_{x_Q-\text{opt}}$ works remarkably well for every member of the family, except when α is close to zero. In fact, for more than 50% of the members of this family, the use of the kernel function optimized for the whole family results in less than 8% larger MSE than the kernel optimized for each member.

3.2. A Family of Nonstationary AR(1)-Processes. Let $e(t)$ be a stationary AR(1)-process: $e_{\theta_1}(t) = \theta_1 e_{\theta_1}(t-1) + \epsilon(t)$, $|\theta_1| < 1$, where $\{\epsilon(t), t \in \mathbb{Z}\}$ is a white Gaussian noise process with variance $(1 - |\theta_1|)^{1.5}$. This process is enveloped in order to get a nonstationary random process: $x_{\theta_1, \theta_2}(t) = e_{\theta_1}(t) e^{-(t-n/2-0.5)^2 / (\theta_2 n)^2}$. As seen, the process x_{θ_1, θ_2} is described by two parameters, θ_1 and θ_2 . In this example, we will compare two different kernels. The first one, $H_{\text{rect-opt}}$, is optimized using Theorem 1 for the rectangular parameter set $0.5 \leq \theta_1 \leq 0.9$ and $0.5 \leq \theta_2 \leq 1$. More formally, we apply Theorem 1 on the family $\{x_{\theta_1, \theta_2}(t), t \in \mathbb{Z}, (\theta_1, \theta_2) \in \mathcal{Q}\}$, where $\mathcal{Q} = [0.5, 0.9] \times [0.5, 1]$ and where Q is a random element of \mathcal{Q} with uniform distribution. The uniform distribution is approximated with an equidistant grid of point masses in order to simplify the integral expression. The second kernel is a separable kernel function, that is, a kernel function that can be separated into one function dependent on k and one function dependent on τ . Such kernels are well suited for covariance function estimation of the random processes that we consider in this example. We choose the separable kernel $H_{\text{sep-opt}}(k, \tau) = h_1(k)h_2(\tau)$, where h_1 and h_2 are Hanning windows, each with a length and amplitude that have been numerically MSE optimized for $(\theta_1, \theta_2) = (0.7, 0.75)$.

Thus, we now have two kernels, the first one optimized on the rectangular space $0.5 \leq \theta_1 \leq 0.9$ and $0.5 \leq \theta_2 \leq 1$ using Theorem 1, and the second one numerically optimized in its four parameters (length of the two Hanning windows and their amplitudes). We will now compare these for all processes $0 < \theta_1 < 1.5$, $0 < \theta_2 < 1$. Note that we

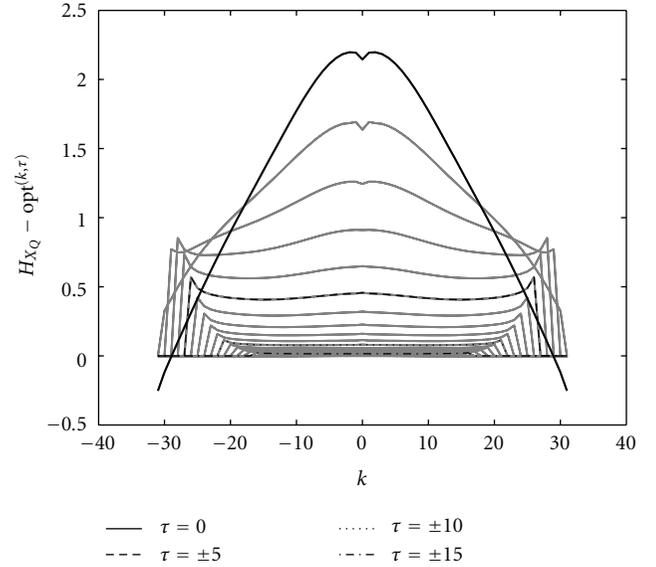


FIGURE 3: The MSE optimal kernel function for the family of nonstationary AR(1)-processes.

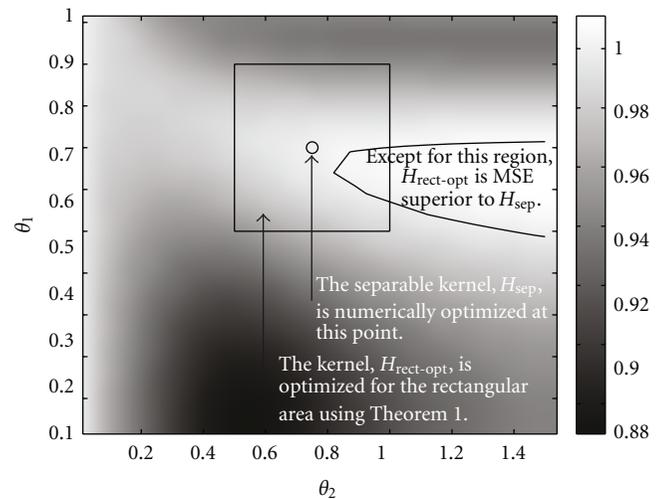


FIGURE 4: The ratio between the MSE of the optimal kernel function for the family of nonstationary AR(1)-processes and the MSE of a separable kernel function. The first kernel has been optimized, as given by Theorem 1, for the processes with parameters inside the rectangle. The lengths and the amplitudes of the two Hanning windows of the separable kernel have been optimized for the parameter values at the circle. The black contour shows the border where the two kernels give equally MSE. Outside this region the kernel optimized as described in this paper is MSE superior to the separable kernel.

include processes outside the rectangular, where $H_{\text{rect-opt}}$ is optimized. The ratio between the MSE of the kernels is given in Figure 4 as a function of the parameter space. We see that the first kernel, $H_{\text{rect-opt}}$, is better than the separable kernel nearly everywhere.

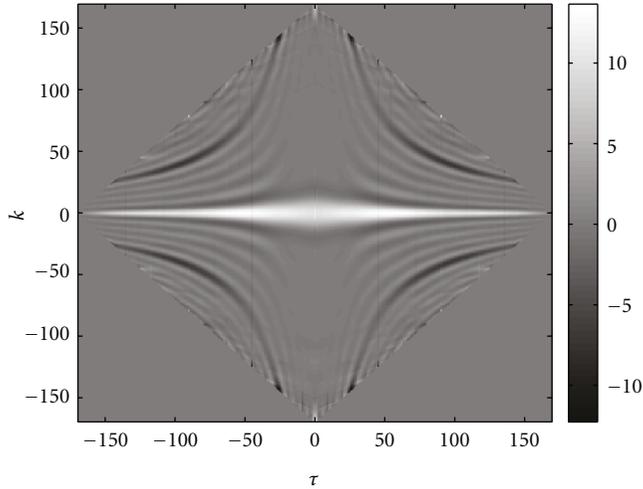


FIGURE 5: The MSE optimal kernel function for the family of enveloped chirp processes.

3.3. A Family of Chirp Processes. In this example, we will study measurements of heart rate variability (HRV), [16]. Such measurements are often modeled as an observed realization from a nonstationary random process with stationary mean. The second-order moments are considered to be of greatest value from a medical perspective, [17]. Our HRV measurements can be expected to have an increasing frequency as the recording is made during an experiment with increasing respiratory rate.

Our data consist of $n = 170$ HRV measurements with sampling rate 2 Hz. After the mean of the data has been removed, we consider it to be an observation of a nonstationary zero-meaned random process. In order to estimate the covariance function of this process, we use an estimator of the form (1), where we compute the kernel function to be MSE optimal to the following family of jointly Gaussian distributed enveloped chirps: let $\{x_{(\alpha,\beta,\gamma)}(t), t \in \mathbb{Z}, (\alpha,\beta,\gamma) \in \mathcal{Q}\}$, where $x_{(\alpha,\beta,\gamma)}(t) = Aw_{\gamma}(t) \sin(2\pi\nu_{(\alpha,\beta)}(t)t + \nu_0)$ for all $t \in T_n$ and 0 otherwise, $\nu_{(\alpha,\beta)}(t) = \alpha t/n + \beta$, $w_{\gamma}(t) = 1/\gamma e^{-(t-n/2-0.5)^2/(\gamma n)^2}$, ν_0 is a random variable uniformly distributed in $[0, 2\pi)$, and A is a Rayleigh-distributed random variable independent of ν_0 . The parameters α and β can be thought of as the raise and starting point of the chirp frequency and γ as the width of the envelope. We choose the a priori distribution of the parameters to be uniform on $-0.1 \leq \alpha \leq 0.1$, $-0.25 \leq \beta \leq 0.25$, $0.1 \leq \gamma \leq 1$, but in order to simplify the computations we approximate this distribution with a uniform point distribution. The MSE optimal kernel function is computed as described in Theorem 1 and can be seen in Figure 5. As mentioned in the introduction, non-parametric covariance function estimators are not guaranteed to be non-negative definite, [7]. We make the resulting covariance matrix estimate non-negative definite by writing the estimate as an eigenvalue decomposition and removing the negative eigenvalues and their respective eigenvectors. The corresponding Wigner spectrum is computed using Jeong and Williams discrete Wigner representation, [4, 18].

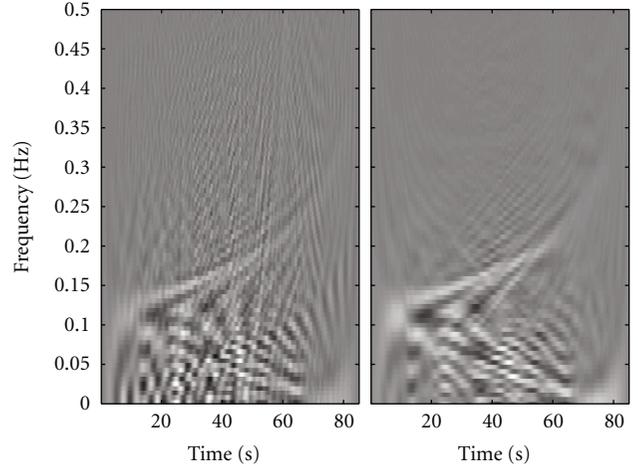


FIGURE 6: Left: Wigner distribution of HRV data. Right: Wigner spectrum of the estimated covariance function.

It is shown in Figure 6 together with the Wigner distribution of the data [2, 19].

4. Conclusions and Final Remarks

A non-parametric estimate of the covariance function of a random process is often obtained by the use of a kernel function. Different kernel functions have been proposed, [8]. In order to favor one kernel over another, some prior knowledge about the random process is needed. In this paper, we have proved that the MSE optimal kernel function for any parameterizable family of random processes can be computed. In a few examples, we have demonstrated that the resulting kernel can be close to optimal for all members of the family. Moreover, the resulting kernels are often robust in the sense that they also work well for nonmembers of the family.

Appendix

Proof of Theorem 1

We would like to solve the following minimization problem:

$$H_{x_Q\text{-opt}} = \arg \min_{H \in \mathcal{H}} \mathbb{E} \left[\sum_{(s,t) \in T_n^2} |r_{x_Q}(s,t) - R_{x_Q;H}(s,t)|^2 \right]$$

With $\tau = s - t$:

$$= \arg \min_{H \in \mathcal{H}} \mathbb{E} \left[\sum_{\tau=-n+1}^{n-1} \sum_{t=\max(1,1-\tau)}^{\min(n,n-\tau)} |r_{x_Q}(t+\tau,t) - R_{x_Q;H}(t+\tau,t)|^2 \right]$$

$$\begin{aligned}
&= \arg \min_{H \in \mathcal{H}} \int_{\mathcal{Q}} \sum_{\tau=-n+1}^{n-1} \sum_{t=\max(1,1-\tau)}^{\min(n,n-\tau)} \\
&\quad \times \mathbb{E} \left[\left| r_{x_q}(t+\tau, t) - R_{x_q; H}(t+\tau, t) \right|^2 \right] dF_{\mathcal{Q}}(q) \\
&= \arg \min_{H \in \mathcal{H}} \int_{\mathcal{Q}} \sum_{\tau=-n+1}^{n-1} \sum_{t=\max(1,1-\tau)}^{\min(n,n-\tau)} \\
&\quad \times \mathbb{E} \left[\left| r_{x_q}(t+\tau, t) - \frac{1}{|\mathcal{K}_{\tau}|} \sum_{k \in \mathcal{K}_{\tau}} H(k, \tau) \right. \right. \\
&\quad \quad \left. \left. \times x_q(t+\tau+k) x_q(t+k)^* \right|^2 \right] dF_{\mathcal{Q}}(q) \\
&= \arg \min_{H \in \mathcal{H}} \int_{\mathcal{Q}} \sum_{\tau=-n+1}^{n-1} \sum_{t=\max(1,1-\tau)}^{\min(n,n-\tau)} \\
&\quad \times \mathbb{E} \left[-r_{x_q}(t+\tau, t) \frac{1}{|\mathcal{K}_{\tau}|} \sum_{k \in \mathcal{K}_{\tau}} H(k, \tau)^* x_q(t+\tau+k)^* x_q(t+k) \right. \\
&\quad \quad - r_{x_q}(t+\tau, t)^* \frac{1}{|\mathcal{K}_{\tau}|} \sum_{k \in \mathcal{K}_{\tau}} H(k, \tau) x_q(t+\tau+k) x_q(t+k)^* \\
&\quad \quad + \frac{1}{|\mathcal{K}_{\tau}|^2} \sum_{k_1 \in \mathcal{K}_{\tau}} \sum_{k_2 \in \mathcal{K}_{\tau}} H(k_1, \tau)^* H(k_2, \tau) x_q(t+\tau+k_1)^* \\
&\quad \quad \quad \left. \left. \times x_q(t+k_1) x_q(t+\tau+k_2) x_q(t+k_2)^* \right] dF_{\mathcal{Q}}(q) \right. \\
&= \arg \min_{H \in \mathcal{H}} \int_{\mathcal{Q}} \sum_{\tau=-n+1}^{n-1} \sum_{t=\max(1,1-\tau)}^{\min(n,n-\tau)} \\
&\quad \times \left(-r_{x_q}(t+\tau, t) \frac{1}{|\mathcal{K}_{\tau}|} \sum_{k \in \mathcal{K}_{\tau}} H(k, \tau)^* r_{x_q}(t+\tau+k, t+k) \right. \\
&\quad \quad - r_{x_q}(t+\tau, t)^* \frac{1}{|\mathcal{K}_{\tau}|} \sum_{k \in \mathcal{K}_{\tau}} H(k, \tau) r_{x_q}(t+\tau+k, t+k)^* \\
&\quad \quad \left. + \frac{1}{|\mathcal{K}_{\tau}|^2} \sum_{k_1 \in \mathcal{K}_{\tau}} \sum_{k_2 \in \mathcal{K}_{\tau}} H(k_1, \tau) H(k_2, \tau)^* \right. \\
&\quad \quad \quad \left. \left. \times \rho_{x_q}(t+k_1, \tau, t+k_2, \tau) \right) dF_{\mathcal{Q}}(q) \right. \\
&\quad \quad \quad \left. \right) dF_{\mathcal{Q}}(q). \tag{A.1}
\end{aligned}$$

We denote the target of minimization with $F : \mathcal{H} \mapsto \mathbb{R}$, associate H with $\mathbb{R}^{2(2n^2-2n+1)}$, and we find the minima by

setting its derivative with respect to $H(k, \tau)^*$ to zero:

$$\begin{aligned}
&\frac{\partial F}{\partial H(k, \tau)^*} \\
&= \int_{\mathcal{Q}} \sum_{t=\max(1,1-\tau)}^{\min(n,n-\tau)} \left(-r_{x_q}(t+\tau, t) \frac{1}{|\mathcal{K}_{\tau}|} r_{x_q}(t+\tau+k, t+k) \right. \\
&\quad \left. + \frac{1}{|\mathcal{K}_{\tau}|^2} \sum_{k_1 \in \mathcal{K}_{\tau}} H(k_1, \tau) \right. \\
&\quad \quad \left. \times \rho_{x_q}(t+k_1, \tau, t+k, \tau) \right) dF_{\mathcal{Q}}(q) = 0, \tag{A.2}
\end{aligned}$$

which concludes the proof.

Acknowledgments

This work was supported by the Swedish Research Council. The first author would like to thank Johannes Sivén at Lund University for stimulating and valuable discussions.

References

- [1] J. Sandberg and M. Hansson-Sandsten, "A comparison between different discrete ambiguity domain definitions in stochastic time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 868–877, 2009.
- [2] G. Matz and F. Hlawatsch, "Wigner distributions (nearly) everywhere: time-frequency analysis of signals, systems, random processes, signal spaces, and frames," *Signal Processing*, vol. 83, no. 7, pp. 1355–1378, 2003.
- [3] M. B. Priestley, "Evolutionary spectra and non-stationary processes," *Journal of the Royal Statistical Society B*, vol. 27, no. 3, pp. 204–237, 1965.
- [4] W. Martin, "Time-frequency analysis of random signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '82)*, vol. 7, pp. 1325–1328, 1982.
- [5] R. Hyndman and M. Wand, "Nonparametric autocovariance function estimation," *Australian Journal of Statistics*, vol. 39, pp. 313–325, 1997.
- [6] D. Ruppert, M. P. Wand, U. Holst, and O. Hössjer, "Local polynomial variance-function estimation," *Technometrics*, vol. 39, no. 3, pp. 262–273, 1997.
- [7] P. Hall, N. I. Fisher, and B. Hoffmann, "On the nonparametric estimation of covariance functions," *The Annals of Statistics*, vol. 22, no. 4, pp. 2115–2134, 1994.
- [8] M. G. Amin, "Spectral smoothing and recursion based on the nonstationarity of the autocorrelation function," *IEEE Transactions on Signal Processing*, vol. 39, no. 1, pp. 183–185, 1991.
- [9] A. M. Sayeed and D. L. Jones, "Optimal kernels for non-stationary spectral estimation," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 478–491, 1995.
- [10] P. Wahlberg and M. Hansson, "Kernels and multiple windows for estimation of the Wigner-Ville spectrum of Gaussian locally stationary processes," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 73–84, 2007.

- [11] P. Wahlberg and M. Hansson, "Optimal time-frequency kernels for spectral estimation of locally stationary processes," in *Proceedings of the IEEE Workshop on Statistical Signal Processing*, pp. 250–253, 2003.
- [12] J. Sandberg and M. Hansson-Sandsten, "Optimal stochastic discrete time-frequency analysis in the ambiguity and time-lag domain," *Signal Processing*, vol. 90, no. 7, pp. 2203–2211, 2010.
- [13] R. A. Silverman, "Locally stationary random processes," *IRE Transactions on Information Theory*, vol. 3, pp. 182–187, 1957.
- [14] R. A. Silverman, "A matching theorem for locally stationary random processes," *Communications on Pure and Applied Mathematics*, vol. 12, pp. 373–383, 1959.
- [15] P. Flandrin, *Time-Frequency/Time Scale Analysis*, Academic Press, New York, NY, USA, 1999.
- [16] E. Kristal-Boneh, M. Raifel, P. Froom, and J. Ribak, "Heart rate variability in health and disease," *Scandinavian Journal of Work, Environment and Health*, vol. 21, no. 2, pp. 85–95, 1995.
- [17] M. Hansson-Sandsten and P. Jönsson, "Multiple window correlation analysis of HRV power and respiratory frequency," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1770–1779, 2007.
- [18] J. Jeong and W. J. Williams, "Alias-free generalized discrete-time time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 40, no. 11, pp. 2757–2765, 1992.
- [19] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution—a tool for time-frequency signal analysis. Part II: discrete-time signals," *Philips Journal of Research*, vol. 35, no. 4-5, pp. 276–300, 1980.