Probab. Theory Relat. Fields 131, 479–492 (2005)
Digital Object Identifier (DOI) 10.1007/s00440-004-0371-x

Basilis Gidas · Alejandro Murua

Optimal transformations for prediction in continuous-time stochastic processes: finite past and future

Received: 7 December 2001 / Revised version: 30 March 2004 /
Published online: 10 February 2005 –

Abstract. In the classical Wiener-Kolmogorov linear prediction problem, one *fixes* a linear functional in the "future" of a stochastic process, and seeks its best predictor (in the L^2-sense). In this paper we treat a variant of the prediction problem, whereby we seek the *"most predictable"* non-trivial functional of the future *and* its best predictor; we refer to such a pair (if it exists) as an *optimal transformation for prediction.* In contrast to the Wiener-Kolmogorov problem, an optimal transformation for prediction may not exist, and if it exists, it may not be unique. We prove the existence of optimal transformations for finite "past" and "future" intervals, under appropriate conditions on the spectral density of a weakly stationary, continuous-time stochastic process. For rational spectral densities, we provide an explicit construction of the transformations via differential equations with boundary conditions and an associated eigenvalue problem of a finite matrix.

1. Introduction

Let $\{X_t : t \in \mathbb{R}\}$ be a real-valued stationary process on a probability space $(\Omega, \mathcal{F}, P)$. $L^2(\Omega, \mathcal{F}, P)$ will denote the Hilbert space of square-integrable, $\mathcal{F}$-measurable functions of $\{X_t\}$, and $\|\cdot\|_P$ its norm. For $I \subseteq \mathbb{R}$, $\mathcal{F}_I$ will denote the smallest σ-algebra generated by $\{X_t : t \in I\}$, and $L^2(\Omega, \mathcal{F}_I, P)$ the space of square-integrable, $\mathcal{F}_I$-measurable functions. Throughout this paper, the intervals $I_0 = [\Delta, \Delta+\tau], \Delta > 0, 0 < \tau \leq +\infty$, and $I_1 = [-T, 0], 0 < T \leq +\infty$, will play the role of "future" and "past", respectively.

Our work [8, 9] on a speech recognition problem lead to a variant (see (1.4) below) of the problem of whether the infimum

$$\inf\{\|\xi_0 - \xi_1\|_P^2 : \xi_i \in L^2(\Omega, \mathcal{F}_{I_i}, P), \quad E\xi_i = 0, \quad i = 0, 1, \quad \|\xi_0\|_P = 1\} \tag{1.1}$$

is attained. The linear version of this problem corresponds to restricting the ξ_i's to be linear functionals of $\{X_t\}$. More precisely, let $\{X_t : t \in \mathbb{R}\}$ be a weakly

B. Gidas: Division of Applied Mathematics, Brown University, Providence, R.I. 02912, USA. e-mail: gidas@dam.brown.edu

A. Murua: Department of Statistics, University of Washington, Seattle, WA 98195-4322, USA. e-mail: murua@stat.washington.edu

This research was partially supported by ARO (MURI grant) DAAH04-96-1-0445, NSF grant DMS-0074276, and CNPq grant 301179/00-0.

stationary, real-valued, mean zero, finite variance process, and let $\mathcal{H}_I$ denote the $\|\cdot\|_P$-closure of the span of $\{X_t : t \in I\}$. $\mathcal{H}$ will stand for $\mathcal{H}_{\mathbb{R}}$. Then the linear analogue of (1.1) is

$$\sigma^2 = \inf\{\|\xi_0 - \xi_1\|_P^2 : \xi_i \in \mathcal{H}_{I_i}, \quad i = 0, 1, \quad \|\xi_0\|_P = 1\}. \tag{1.2}$$

If $\{X_t\}$ is Gaussian, the two problems are equivalent [3, p.66], (see [10] for alternative proof).

Problems (1.1) and (1.2) may be viewed as a variant of the classical Wiener-Kolmogorov (WK) prediction problem. In the WK problem, one *fixes* a functional ξ_0 in the future, and seeks its *best predictor*. In problems (1.1) and (1.2) we seek the *most predictable, non-trivial* functional of the future and its best predictor. In contrast to the situation in the WK problem, problems (1.1) and (1.2) may not have a solution or may have multiple solutions. In analogy with the optimal transformations for regression [1], we refer to solutions of (1.1) and (1.2), if they exist, as (nonlinear/linear) *optimal transformations* for prediction.

The existence of optimal transformations for prediction is closely linked to properties of the projection operators onto the subspaces $L^2(\Omega, \mathcal{F}_{I_i}, P)$ for problem (1.1), and onto the subspaces $\mathcal{H}_{I_i}, i = 0, 1$, for problem (1.2). More precisely, for problem (1.2), let Q_i be the orthogonal projection from $\mathcal{H}$ onto $\mathcal{H}_{I_i}$ $(i = 0, 1)$, and $Q_{01} = Q_1|_{\mathcal{H}_{I_0}}$, $Q_{10} = Q_0|_{\mathcal{H}_{I_1}}$, $B = Q_{10}Q_{01}$. It can be shown [10] that problem (1.2) has a solution if and only if $\|B\|$ (operator norm) is an eigenvalue (not necessarily isolated) of B; moreover, the dimension of the manifold of solutions is equal to the multiplicity of $\|B\|$. Similar results hold for (1.1).

Hence the problem of existence of optimal transformations is reduced to the question: Under what conditions is $\|B\|$ an eigenvalue of B? If B is compact then necessarily $\|B\|$ is an eigenvalue. A natural question then is: Under what conditions is B compact? We have not been able to find a satisfactory solution to these questions for the nonlinear problem (1.1). For problem (1.2) with $T = \tau = +\infty$ and $\{X_t\}$ *linearly regular* [12, p.112], the operator B is related to Hankel operators [16, 15], and one can provide [15, 10] necessary and sufficient conditions (in terms of the spectral density $f(\omega)$ of $\{X_t\}$) for B to be compact.

But when $0 < T, \tau < +\infty$, the case treated in this paper, problem (1.2) is much subtler as we will explain below and in §2. Our main result (for $0 < T, \tau < +\infty$) is stated in Theorem 2.1 (§2). In brief, if $\{X_t\}$ has a spectral density $f(\omega)$ satisfying

$$\lim_{|\omega| \to +\infty} (1 + \omega^2)^n f(\omega) = +\infty, \tag{1.3}$$

with some integer $n \geq 1$, $\log f(\omega)$ is locally integrable, and the covariance function of the process satisfies certain mild differentiability and local integrability conditions, then B is compact. In the particular case when $f(\omega)$ is rational, we provide (§3) an explicit construction of the solutions. The rational case with $T = \tau = +\infty$ was treated in [20]; the method of [20] could in principle be extended in the case $T, \tau < +\infty$, but not in an effective way. Our construction is algorithmically very efficient – it provides an explicit representation of the solutions via a differential equation, and ultimately via a finite matrix eigenvalue problem; the construction extends a procedure [18, pp.135-142] for studying the classical *"forecasting"* problem.

Our study uses the well-known equivalence [3, p. 73] between the spaces $\mathcal{H}_{I_i}$, and certain function spaces $\mathcal{L}_{I_i}(f)$. When $T = \tau = +\infty$, the spaces $\mathcal{L}_{I_i}(f)$ can be represented [3, p. 96] in terms of the Hardy spaces on the half-plane, and the operator B can be written in terms of Hankel operators. On the other hand, for $T, \tau < +\infty$, the $\mathcal{L}_{I_i}(f)$'s are [3, p. 108] spaces of entire functions of exponential type, and can be represented in terms of Krein spaces [3, pp. 220–322] whose structure is more complicated than that of Hardy spaces; this makes the relevant operator B more complex and its analysis more delicate. Under the above conditions on $f(\omega)$, we will show that the spaces $\mathcal{L}_{I_i}(f)$ are subspaces of certain Sobolev spaces – a property heavily exploited in our analysis. We do not know of any other work related to the existence of optimal transformations for problem (1.2) with $T, \tau < +\infty$. The tools of Hardy spaces and Hankel operators are not the appropriate tools for the case $T, \tau < +\infty$, although a weak result can be obtained (see Remark at the end of §2) using these tools. An interesting open problem both for $T = \tau = +\infty$ and $0 < T, \tau < +\infty$ is to provide conditions under which $\|B\|$ is an eigenvalue of B but B is *not* compact. In [7], we study the statistical estimation of optimal transformations (if they exist), on the basis of a finite discrete sample, and establish the consistency of the estimators as the sampling rate goes to zero and the sample size goes to infinity. The existence and statistical estimation of optimal transformations, have been useful [8, 9] in the modeling and optimal sampling of the *acoustic signal* in speech recognition.

For discrete-time processes, the analogue of problem (1.2) with $T = \tau = +\infty$ was first posed by Helson and Szegö [11] in connection with a Functional Analysis problem in trigonometric series; their work stimulated a great deal of mathematical research in the theory of bounded analytic, BMO (Bounded Mean Oscillation) functions, and other problems (see [14, pp. 249-287], [4, pp. 144, 254], and references therein). The operator B was first introduced in [6], and has been used (see [12, pp. 191–223]) extensively for studying the *regularity* of stationary random processes. All the above studies are fundamentally different from our study here.

We end this introduction with an open problem which is intermediate between problems (1.1) and (1.2), and which was motivated by our studies [8, 9] of a speech problem: The problem is specified by replacing the spaces $L^2(\Omega, \mathcal{F}_{I_i}, P), i = 0, 1$, in problem (1.1), by spaces whose elements ξ_i are *formally* represented by

$$\xi = \int_I u(X_t, t)\, dt, \qquad I = I_0,\ I_1, \tag{1.4}$$

with $u(x, t)$ in a class of (non-linear) functions so that $E\{\xi_i^2\} < +\infty$, and $E\xi_i = 0$. The linear problem (1.2) corresponds to $u(x, t) = \alpha(t)\, x$; in this case the space of "functions" $\alpha(t)$ for which $E\{\xi^2\} < +\infty$ can easily be characterized using the spectral representation [18, 3] of $\{X_t\}$. We have not been able to find a convenient characterization of the space of functions $u(x, t)$ for which $E\{\xi^2\} < +\infty$. The existence of optimal transformations in this case is an interesting open problem.

The organization of this paper is as follows. Section 2 contains the existence of solutions for (1.2), and Section 3 the construction of the solutions when the spectral density is rational.

2. Existence of Optimal Linear Transformations

Let $I_i, \mathcal{H}, \mathcal{H}_{I_i}$ $(i = 0, 1)$, B, and σ^2 be as in §1. The main result of this section is:

Theorem 2.1. *Let* $\{X_t : t \in \mathbb{R}\}$ *be a weakly stationary, mean-zero stochastic process with spectral density* $f(\omega)$ *satisfying (1.3), and assume that* $\log f(\omega)$ *is locally integrable. Assume also that* $0 < T, \tau < +\infty$, *and that the covariance function* $R(t)$ *of* $\{X_t\}$ *has* $2n$ *derivatives in the interval* $[\Delta,\ T + \Delta + \tau]$ *satisfying*

$$\int_{-T}^{0} \int_{\Delta}^{\Delta+\tau} \left| \left(1 - \frac{d^2}{dt^2}\right)^n R(t' - s) \right|^2 dt'ds < +\infty. \tag{2.1}$$

Then B *is compact. In particular* σ^2 *is attained.*

The proof of Theorem 2.1 uses the well-known [12, pp.16–20],[3, p.108] equivalence of $\mathcal{H}_I$, $I \subseteq \mathbb{R}$, to the space $\mathcal{L}_I(f)$ defined as follows. Let $\mathcal{L}(f) = L^2(\mathbb{R}, f)$ and $\|\cdot\|_f$ its norm. The space $\mathcal{L}_I(f)$ is the $\|\cdot\|_f$-closure of the linear span of $\{e^{i\omega t} : t \in I\}$, $\omega \in \mathbb{R}$. The one-to-one *isometric* correspondence between $\mathcal{H}_I$ and $\mathcal{L}_I(F)$ is given [12, pp.16-20] [18, pp.14-18] by the spectral representation of $\{X_t\}$. If $\phi_i \in \mathcal{L}_{I_i}(f)$ $(i = 0, 1)$ correspond to the ξ_i's of (1.2), then the infimum in (1.2) is given by $\sigma^2 = \inf\{\|\phi_0 - \phi_1\|_f^2 : \phi_i \in \mathcal{L}_{I_i}(f), i = 0, 1, \|\phi_0\|_f = 1\}$. Now let P_i denotes the orthogonal projection of $\mathcal{L}(f)$ onto $\mathcal{L}_{I_i}(f)$, and $P_{01} = P_1|_{\mathcal{L}_{I_0}}$, $P_{10} = P_0|_{\mathcal{L}_{I_1}}$, $A = P_{10}P_{01}$; these correspond to Q_i, Q_{01}, Q_{10}, and B, respectively. Clearly B is compact if and only if A is compact, and σ^2 is attained if and only if $\|A\|$ is an eigenvalue of A. It can be shown [10] that $1 - \sigma^2 = \|B\| = \|A\| = \lambda_1$ where $\sqrt{\lambda_1}$ is the *maximal linear correlation* between the system of random variables $\{\xi_0 \in \mathcal{H}_{I_0}\}$ and $\{\xi_1 \in \mathcal{H}_{I_1}\}$ (equivalently, the cosine of the angle between the Hilbert spaces $\mathcal{L}_{I_i}$).

Next, let $I = [a,b]$ be any bounded interval. If $\mathcal{L}_I(f) \neq \mathcal{L}(f)$, then [3, p.108] any $\phi \in \mathcal{L}_{[a,b]}(f)$ has an analytic continuation to an entire function $\phi(z)$, $z \in \mathbb{C}$, of exponential type with exponent less or equal to $\max(|a|, |b|)$. If $f(\omega)$ satisfies (1.3) and $\log f(\omega)$ is locally integrable, then we can say a bit more. First, we note that the two conditions together imply

$$\int_{\mathbb{R}} \frac{\log f(\omega)}{1 + \omega^2}\, d\omega > -\infty \tag{2.2}$$

In turn, this implies [3, p.112] that $\mathcal{L}_I(f) \neq \mathcal{L}(f)$ and that an element $\phi \in \mathcal{L}_I(f)$ has an entire analytic continuation $\phi(z)$, $z \in \mathbb{C}$, which satisfies for any $\epsilon > 0$

$$|\phi(z)| \leq \|\phi\|_f \exp\{M_\epsilon + max(|a|, |b|)|\Im z| + \epsilon|z|\}. \tag{2.3}$$

with some constant M_ϵ independent of ϕ. Now let $f_n(\omega) = (1 + \omega^2)^{-n}$. If $\phi \in \mathcal{L}_I(f)$, then condition (1.3) and the bound (2.3) yield $\|\phi\|_{f_n} \leq C\|\phi\|_f$ with a constant C independent of ϕ. Hence, if $\{\phi_N\}$ is a sequence from the linear span of $\{e^{i\omega t} : t \in I\}$ that converges to $\phi \in \mathcal{L}_I(f)$ (in the $\|\cdot\|_f$ norm), then it also converges to ϕ in the $\|\cdot\|_{f_n}$ norm, and so $\phi \in \mathcal{L}_I(f_n)$. Thus we have proven

Lemma 2.1. *Let I and $f_n(\omega)$ be as above. If f satisfies (1.3) and* $\log f(\omega)$ *is locally integrable, then $\mathcal{L}_I(f) \subset \mathcal{L}_I(f_n)$, and the injection of $\mathcal{L}_I(f)$ into $\mathcal{L}_I(f_n)$ is continuous.*

Next, it is well-known [12, pp.28–32] that the elements of $\mathcal{L}_I(f_n)$, $I = [a,b]$, have the following representation

$$\phi(\omega) = e^{i\omega a}\sum_{k=0}^{n-1}\alpha_k(1-i\omega)^k + (1-i\omega)^n\int_a^b e^{i\omega t}\, v(t)\, dt \tag{2.4a}$$

$$= e^{i\omega b}\sum_{k=0}^{n-1}\beta_k(1-i\omega)^k + (1-i\omega)^n\int_a^b e^{i\omega t}\, u(t)\, dt. \tag{2.4b}$$

with u, $v \in L^2[a,b]$, and $\{\alpha_k\}$, $\{\beta_k\}$ constants. Hence if the conditions of Lemma 2.1 hold, then the elements of $\mathcal{L}_I(f)$ are also represented by (2.4).

Remark. The spaces $\mathcal{L}(f_n)$, $\mathcal{L}_I(f_n)$ are the well-known Sobolev spaces; their elements have a well-defined inverse Fourier transform in the sense of tempered distributions [17]. Moreover, the inverse Fourier transform of the elements of $\mathcal{L}_I(f_n)$, and hence those of $\mathcal{L}_I(f)$, are tempered distributions with *compact support* in $[a,b]$. Using this, a generalization of the classical Paley-Wiener theorem [17, p. 334], and the continuity of the injection in Lemma 2.2, one can improve (2.3) to $|\phi(z)| \le \|\phi\|_f C_n(1+|z|)^n \exp\{max(|a|,|b|)|\Im z|\}$. This bound or bound (2.3) and the representation (2.4) are key in proving the next lemma, which, in turn, is key in proving Theorem 2.1.

Lemma 2.2. *Let $I = [a,b]$ be bounded and $f_n(\omega) = (1+\omega^2)^{-n}$. Assume that f satisfies* (1.3)*, and that $\log f(\omega)$ is locally integrable. Then (a) The injection of $\mathcal{L}_{[a,b]}(f)$ into $\mathcal{L}_{[a,b]}(f_n)$ is compact. (b) If $\{\phi_N\}_{N\ge 0}$ is a sequence in $\mathcal{L}_{[a,b]}(f)$ converging weakly to some $\phi \in \mathcal{L}_{[a,b]}(f)$, and ϕ_N has the representation* (2.4b) *with parameters $\{\beta_k^{(N)}\}_{k=0}^{n-1}$ and $u_N(t) \in L^2(a,b)$, and ϕ has the representation* (2.4b)*, then*

$$\beta_k^{(N)} \longrightarrow \beta_k, \quad ||u_N - u||_{L^2(a,b)} \longrightarrow 0 \;\; as\; N \to +\infty$$

Proof. (a) It suffices to show that the unit ball B in $\mathcal{L}_{[a,b]}(f)$ is relatively compact in $\mathcal{L}_{[a,b]}(f_n)$. Let $\{\phi_N\}$ be a sequence in B; since B is weakly compact, we can extract a subsequence also denoted by $\{\phi_N\}$ so that ϕ_N converges weakly to some $\phi \in B$. Now, $\phi_N(\omega)$ has an entire continuation $\phi_N(z)$, $z \in \mathbb{C}$, and by (2.3) $\phi_N(z)$ is uniformly bounded on every compact subset of $\mathbb{C}$; hence $\{\phi_N(z)\}$ form a normal family [19, p. 282], and therefore one can extract a subsequence to be denoted again by $\{\phi_N(z)\}$ that converges uniformly on every compact subset of $\mathbb{C}$ to some $\psi(z)$. In particular $\phi_N(\omega) \to \psi(\omega)$, for all $\omega \in \mathbb{R}$, and hence $\phi(\omega) = \psi(\omega)$, $\omega \in \mathbb{R}$.

Now, for every $R > 0$

$$||\phi_N - \phi||_{f_n}^2 = \int_{|\omega|\geq R} \frac{|\phi_N(\omega) - \phi(\omega)|^2}{(1+\omega^2)^n}\, d\omega + \int_{|\omega|<R} \frac{|\phi_N(\omega) - \phi(\omega)|^2}{(1+\omega^2)^n}\, d\omega$$
$$\leq \frac{||\phi_N - \phi||_f^2}{\inf_{|\omega|\geq R}(1+\omega^2)^n f(\omega)} + C \sup_{|\omega|<R} |\phi_N(\omega) - \phi(\omega)|^2 \qquad (*)$$

with some $C > 0$; using (1.3) and the uniform convergence of ϕ_N to ϕ on compact subsets of $\mathbb{R}$, we see that the right-hand-side of $(*)$ becomes arbitrarily small for sufficiently large R and N, which proves (a). To prove (b) we write

$$e^{i\omega b} \sum_{k=0}^{n-1} \beta_k^{(N)} (1 - i\omega)^k = (1 - i\omega)^n \int_b^{+\infty} e^{i\omega t} w_N(t)\, dt$$

with $w_N(t) = e^{-t} \sum_{k=0}^{n-1} \lambda_k^{(N)} t^k$, where the coefficients $\{\lambda_k^{(N)}\}_{k=0}^{n-1}$ are linearly related to $\{\beta_k^{(N)}\}_{k=0}^{n-1}$. Similarly, we write

$$e^{i\omega b} \sum_{k=0}^{n-1} \beta_k (1 - i\omega)^k = (1 - i\omega)^n \int_b^{+\infty} e^{i\omega t} w(t)\, dt$$

with $w(t) = e^{-t} \sum_{k=0}^{n-1} \lambda_k t^k$.

Since $\phi_N \longrightarrow \phi$ weakly in $\mathcal{L}_{[a,b]}(f)$, part (a) implies $||\phi_N - \phi||_{f_n} \longrightarrow 0$. But

$$||\phi_N - \phi||_{f_n}^2 = ||w_N - w||_{L^2(b,+\infty)}^2 + ||u_N - u||_{L^2(a,b)}^2$$

Hence $||u_N - u||_{L^2(a,b)} \longrightarrow 0$ and $||w_N - w||_{L^2(b,+\infty)} \longrightarrow 0$ as $N \to +\infty$. But it is easily seen that $||w_N - w||_{L^2(b,+\infty)} \longrightarrow 0$ if and only if $\lambda_k^{(N)} \longrightarrow \lambda_k$. Since $\{\lambda_k^{(N)}\}$ and $\{\lambda_k\}$ are linearly related to $\{\beta_k^{(N)}\}$ and $\{\beta_k\}$), respectively, we have $\beta_k^{(N)} \longrightarrow \beta_k$, $k = 0, 1, \ldots, n-1$. □

Proof of Theorem 2.1. It suffices to prove that A is compact. Since $A = P_{10}P_{01}$ and P_{10} is bounded, it suffices to prove that P_{01} is compact. To this end, let $\{\phi_N\}$, $\phi_N \in \mathcal{L}_{I_0}(f)$ be a sequence converging weakly to zero. Then $\{P_{01}\phi_N\}$ also converges weakly to zero. We need to show that $||P_{01}\phi_N||_f \longrightarrow 0$ as $N \to +\infty$. Since $\phi_N \in \mathcal{L}_{I_0}(f)$ and $P_{01}\phi_N \in \mathcal{L}_{I_1}(f)$, they have a representation of the form (2.4), say

$$\phi_N(\omega) = e^{i\omega(\Delta+\tau)} \sum_{k=0}^{n-1} \beta_k^{(N)} (i\omega)^k + (1 - i\omega)^n \int_\Delta^{\Delta+\tau} e^{i\omega t} u_N(t)\, dt$$

$$(P_{01}\phi_N)(\omega) = e^{-i\omega T} \sum_{k=0}^{n-1} \alpha_k^{(N)} (i\omega)^k + (1 - i\omega)^n \int_{-T}^0 e^{i\omega t} v_N(t)\, dt$$

A straightforward computation, whose justification is given in the Appendix, yields

$$\begin{aligned}||P_{01}\phi_N||_f^2 &= \sum_{k,\ell=0}^{n-1} \beta_k^{(N)}\overline{\alpha_\ell^{(N)}}(-1)^\ell R^{(k+\ell)}(\Delta+\tau+T) \\ &+ \sum_{k=0}^{n-1} \beta_k^{(N)}(-1)^k \int_{-T}^{0} \overline{v_N(s)} \left(1-\frac{d}{dt}\right)^n R^{(k)}(\Delta+\tau-s)\,ds \\ &+ \sum_{k=0}^{n-1} \overline{\alpha_k^{(N)}}(-1)^k \int_{\Delta}^{\Delta+\tau} u_N(t) \left(1-\frac{d}{dt}\right)^n R^{(k)}(T+t)\,dt \\ &+ \int_{-T}^{0}\int_{\Delta}^{\Delta+\tau} \left[\left(1-\frac{d^2}{dt^2}\right)^n R(t'-s)\right] u_N(t')\overline{v_N(s)}\,dt'ds \qquad (2.5)\end{aligned}$$

This together with (2.1) and Lemma 2.2 (b) quickly yields $||P_{01}\phi_N||_f \longrightarrow 0$ as $N \to +\infty$. □

Remark. As we mentioned in the Introduction, if (2.2) holds then there are necessary and sufficient conditions for the operator A in the case $T = \tau = +\infty$ to be compact. This compactness can be used to induce a compactness result for A in the case $T, \tau < +\infty$: let $P_{(-\infty,0]}$ be the orthogonal projection from $\mathcal{L}(f)$ onto $\mathcal{L}_{(-\infty,0]}(f)$, and $P_\Delta = P_{(-\infty,0]}|_{\mathcal{L}_{[\Delta,+\infty)}(f)}$. Then it is easily seen that $P_{01} = P_1 P_\Delta|_{\mathcal{L}_{I_0}(f)}$ where P_{01} and P_1 are the operators for the finite intervals I_i (i.e. for $T, \tau < +\infty$). Hence if P_Δ is compact, then so is P_{01} (and hence A). This result is relatively weak, as the compactness of P_Δ sees only the gap between the intervals $I_0 = [-T, 0]$ and $I_1 = [\Delta, \Delta+\tau]$, and not the intervals themselves.

3. Rational Spectral Density

In this section we consider the case when the spectral density $f(\omega)$ is rational, i.e.

$$f(\omega) = \left|\frac{P(i\omega)}{Q(i\omega)}\right|^2, \quad \omega \in \mathbb{R}, \qquad (3.1)$$

where $P,\ Q$ are polynomials of $i\omega$ of degree p and q, respectively, with the properties: (i) the coefficients of $P,\ Q$ are real, (ii) $q - p \geq 1$, and (iii) $P(i\omega)$ and $Q(i\omega)$ have all its roots in the lower open half-plane $\{x + iy : y < 0\}$. Existence of optimal transformations can easily be deduced from Theorem 2.1, as well as from the remark at the end of §2 (P_Δ is [3, pp.100–101] of finite rank, and hence compact).

The main goal of this section is to give an explicit construction of the transformations in terms of $\lambda_1 = \|A\|$ and the coefficients of P and Q, and a characterization of $\lambda_1 = \|A\|$ as the largest eigenvalue of an explicit finite matrix. Our construction works verbatim for any positive eigenvalue $\lambda > 0$ of A (see remark at the end of this section). As we mentioned in the Introduction, the case $T = \tau = +\infty$ was treated in [20] by a different method.

It can be shown [10] that a pair $\phi_i \in \mathcal{L}_{I_i}(f)$ $(i = 0, 1)$, $\|\phi_0\|_f = 1$, attains $\sigma^2 = 1 - \lambda_1$, if and only if the pair satisfies

$$\int_{\mathbb{R}} e^{-i\omega t}[\lambda_1\phi_0(\omega) - \phi_1(\omega)]f(\omega)\,d\omega = 0, \quad \text{for all } t \in [\Delta, \Delta + \tau] \tag{3.2a}$$

$$\int_{\mathbb{R}} e^{-i\omega t}[\phi_0(\omega) - \phi_1(\omega)]f(\omega)\,d\omega = 0, \quad \text{for all } t \in [-T, 0] \tag{3.2b}$$

subject to $\|\phi_0\|_f = 1$. Following [18] we define

$$x(t) = \int_{\mathbb{R}} e^{-i\omega t}\frac{\phi_0(\omega)}{|Q(i\omega)|^2}\,d\omega, \quad y(t) = \int_{\mathbb{R}} e^{-i\omega t}\frac{\phi_1(\omega)}{|Q(i\omega)|^2}\,d\omega \tag{3.3}$$

For *any* $\phi_i \in \mathcal{L}_{I_i}(f)$, $x(t)$ and $y(t)$ have [18, p.140] the following properties

Lemma 3.1. *$x(t)$ and $y(t)$ have continuous derivatives up to order $p+q-1$, and satisfy the "boundary conditions" ($k = 0, 1, \ldots, p-1$)*

$$Q(\frac{d}{dt})x^{(k)}(t) = 0 \text{ for } t \le \Delta, \quad Q(-\frac{d}{dt})x^{(k)}(t) = 0 \text{ for } t \ge \Delta + \tau \tag{3.4a}$$

$$Q(\frac{d}{dt})y^{(k)}(t) = 0 \text{ for } t \le -T,\ Q(-\frac{d}{dt})y^{(k)}(t) = 0 \text{ for } t \ge 0 \tag{3.4b}$$

For $k = p$, the above equations hold only in the corresponding open intervals.

Since $2q \le p + q - 1$, Lemma 3.1 implies that equations (3.2) are equivalent to

$$P(\frac{d}{dt})P(-\frac{d}{dt})(\lambda_1 x(t) - y(t)) = 0 \quad \text{for } \Delta \le t \le \Delta + \tau \tag{3.5a}$$

$$P(\frac{d}{dt})P(-\frac{d}{dt})(x(t) - y(t)) = 0 \quad \text{for } -T \le t \le 0. \tag{3.5b}$$

The solutions to these equations subject to the "boundary conditions" (3.4) are given in the next lemma whose proof is evident by inspection. We treat only the case $p \ge 1$; the simpler case $p = 0$ can be treated the same way.

Lemma 3.2. *If $\phi_i \in \mathcal{L}_{I_i}(f)$ ($i = 0, 1$) satisfy* (3.2)*, then $x(t)$ and $y(t)$ are given by*

$$x(t) = \begin{cases} w_+(t) \quad \text{for } t < \Delta, \quad w_-(t) \quad \text{for } \Delta + \tau < t \\ \frac{1}{\lambda_1}(v_-(t) - u_1(t)) \quad \text{for } \Delta \le t \le \Delta + \tau \end{cases} \tag{3.6a}$$

$$y(t) = \begin{cases} v_+(t) \quad \text{for } t < -T, \quad v_-(t) \quad \text{for } 0 < t \\ (w_+(t) - u_2(t)) \quad \text{for } -T \le t \le 0 \end{cases} \tag{3.6b}$$

where $w_\pm \in Ker\ Q(\pm\frac{d}{dt})$, $v_\pm \in Ker\ Q(\pm\frac{d}{dt})$, $u_i \in Ker\ P(\frac{d}{dt})P(-\frac{d}{dt})$, ($i = 1, 2$), satisfy the boundary conditions with $\ell = 0, \ldots, q-1$, $k = 0, \ldots, p-1$

$$\left.\begin{array}{l}\lambda_1 w_+^{(\ell)}(\Delta) = v_-^{(\ell)}(\Delta) - u_1^{(\ell)}(\Delta),\ \lambda_1 w_-^{(\ell)}(\Delta+\tau) = v_-^{(\ell)}(\Delta+\tau) - u_1^{(\ell)}(\Delta+\tau)\\ Q(\frac{d}{dt})[v_-^{(k)} - u_1^{(k)}](\Delta) = 0,\ Q(-\frac{d}{dt})u_1^{(k)}(\Delta+\tau) = 0\end{array}\right\} \tag{3.7a}$$

$$\left.\begin{array}{l}v_+^{(\ell)}(-T) = w_+^{(\ell)}(-T) - u_2^{(\ell)}(-T), \quad v_-^{(\ell)}(0) = w_+^{(\ell)}(0) - u_2^{(\ell)}(0)\\ Q(\frac{d}{dt})u_2^{(k)}(-T) = 0, \quad Q(-\frac{d}{dt})[w_+^{(k)} - u_2^{(k)}](0) = 0\end{array}\right\} \tag{3.7b}$$

Next let $s_\pm(t) \in \mathbb{R}^q$, $r(t) \in \mathbb{R}^{2p}$ be bases for Ker $Q(\pm\frac{d}{dt})$ and Ker $P(\frac{d}{dt})$ $P(-\frac{d}{dt})$, respectively. Then we may write $w_\pm(t) = a_\pm^\top s_\pm(t)$, $v_\pm(t) = a_\pm^\top s_\pm(t)$, $u_i(t) = c_i^\top r(t)$ $(i = 1, 2)$, with some $a_+,\ a_-,\ b_+,\ b_- \in \mathbb{R}^q$, and $c_1,\ c_2 \in \mathbb{R}^{2p}$ (here $\top$ denotes transposition). Then equations (3.7) can be written as

$$\mathbf{M}(-T, 0, \Delta, \Delta+\tau)\big(a_+\ a_-\ b_+\ b_-\ c_1\ c_2\big)^\top = 0 \tag{3.8}$$

where $\mathbf{M}(-T, 0, \Delta, \Delta+\tau)$ is a matrix that can be written explicitly in terms of the following matrices: let $D = (1, \frac{d}{dt}, \ldots, \frac{d^{q-1}}{dt^{q-1}})^\top$, and define the matrices $S_\pm(t) = (Ds_\pm^\top(t))^\top$, $R(t) = (Dr^\top(t))^\top$, $S_\pm^Q(t) = Q(\mp\frac{d}{dt})S_\pm(t)$, $R_\pm^Q(t) = Q(\mp\frac{d}{dt})R(t)$, $L(t, t') = (R_-^Q(t), R_+^Q(t'))^\top$, $O_+(t) = (0_{p\times q}, S_+^Q)^\top$, $O_-(t) = (S_-^Q(t), 0_{p\times q})^\top$, where $0_{p\times q}$ is the $p \times q$ null matrix.

The matrices $S_+(t)$ and $S_-(t)$ are non-singular for all $t \in \mathbb{R}$, as is the matrix $R_\mp(t, t')$. This and a straightforward calculation show that (3.8) has a non-trivial solution if and only if λ_1 is an eigenvalue of the $q \times q$ matrix

$$(S_+(\Delta))^{-1}\{S_-(\Delta) - R(\Delta)(L_(\Delta, \Delta+\tau))^{-1}O_-(\Delta)\}(S_-(0))^{-1} \\ \times \{S_+(0) - R(0)(L(-T, 0))^{-1}O_+(0)\} \tag{3.9}$$

with a_+ the corresponding eigenvector. Moreover, the vectors a_-, b_+, b_-, c_1, c_2 that solve (3.8) can be computed directly in terms of a_+ and λ_1; they are linear in a_+.

Let $K = K(-T, 0, \Delta, \Delta+\tau)$ be the $q \times q$ matrix given by (3.9). The above observations can be used to construct the functions $w_\pm(t), v_\pm(t), u_i(t)(t), i = 1, 2$ (and hence $x(t)$ and $y(t)$) as follows: let $\widetilde{a}_+$ be a normalized ($|\widetilde{a}_+| = 1$) eigenvector of K with eigenvalue λ_1, and $\widetilde{a}_-$, $\widetilde{b}_\pm$, $\widetilde{c}_i$ the corresponding solutions of (3.8), computed in terms of $\widetilde{a}_+$ and λ_1. From these, we compute functions $\widetilde{w}_\pm(t)$, $\widetilde{v}_\pm(t)$, $\widetilde{u}_i(t)$, which in turn determine $\widetilde{x}(t)$, $\widetilde{y}(t)$ by (3.6). Now the a_+ that corresponds to the original solution $x(t)$, $y(t)$, may be written as $a_+ = |a_+|\widetilde{a}_+$ for some normalized eigenvector $\widetilde{a}_+$ of K. The length $\mu = |a_+|$ may be determined from $\widetilde{x}(t)$ (and hence from $\widetilde{a}_+$ and λ_1) by noting that $x(t) = \mu\widetilde{x}(t)$ and that $x(t)$ satisfies (by (3.3))

$$\mu Q(\frac{d}{dt})P(-\frac{d}{dt})\widetilde{x}(t) = \int_{\mathbb{R}} e^{-i\omega t}\frac{P(i\omega)}{Q(i\omega)}\phi_0(\omega)\,d\omega$$

Since $\phi_0 \in \mathcal{L}_{I_0}(f)$, we have $Q(\frac{d}{dt})P(-\frac{d}{dt})x(t) \in L^2(\mathbb{R})$. Then μ is determined from $\widetilde{x}(t)$ by the requirement that $\|\phi_0\|_f = 1$. These arguments prove part of the next theorem.

Theorem 3.1. *Let $f(\omega)$ be given by (3.1) with $p \geq 1$. Assume that $0 < T, \tau < +\infty$. Then every pair ϕ_0, ϕ_1 of optimal transformations arises from a normalized eigenvector $\widetilde{a}_+$ of the matrix K (given by (3.9)) with eigenvalue $\lambda_1 = \|A\|$. Let $x(t)$, $y(t)$ be the functions constructed by the above procedure. Then the corresponding ϕ_0, ϕ_1 are given by*

$$\begin{aligned}\phi_0(\omega) = {} & \frac{\mu}{2\pi} e^{i\omega\Delta} \sum_{j=0}^{q-p-1} (-i\omega)^j \sum_{k=j+p+1}^{q} \beta_k \sigma_{k-j-1}(\Delta) \\ & + \frac{\mu}{2\pi} e^{i\omega(\Delta+\tau)} \sum_{j=0}^{q-p-1} (-i\omega)^j \sum_{k=j+p+1}^{q} \beta_k \sigma_{k-j-1}(\Delta+\tau) \\ & + \frac{\mu}{2\pi} \int_{\Delta}^{\Delta+\tau} e^{i\omega t} \left(Q(-\frac{d}{dt}) Q(\frac{d}{dt}) x(t) \right) dt \end{aligned} \tag{3.10a}$$

$$\begin{aligned}\phi_1(\omega) = {} & \frac{\mu}{2\pi} e^{-i\omega T} \sum_{j=0}^{q-p-1} (-i\omega)^j \sum_{k=j+p+1}^{q} \beta_k \xi_{k-j-1}(-T) \\ & + \frac{\mu}{2\pi} \sum_{j=0}^{q-p-1} (-i\omega)^j \sum_{k=j+p+1}^{q} \beta_k \xi_{k-j-1}(0) \\ & + \frac{\mu}{2\pi} \int_{-T}^{0} e^{i\omega t} \left(Q(-\frac{d}{dt}) Q(\frac{d}{dt}) y(t) \right) dt \end{aligned} \tag{3.10b}$$

where for $p \leq k \leq q-1$

$$\begin{aligned} \sigma_k(\Delta) &= Q(\frac{d}{dt}) x^{(k)}(\Delta + 0), \quad \xi_k(0) = Q(\frac{d}{dt}) y^{(k)}(0-0) \\ \sigma_k(\Delta+\tau) &= Q(\frac{d}{dt}) x^{(k)}(\Delta+\tau+0) - Q(\frac{d}{dt}) x^{(k)}(\Delta+\tau-0) \\ \xi_k(-T) &= Q(\frac{d}{dt}) y^{(k)}(-T+0) - Q(\frac{d}{dt}) y^{(k)}(-T-0) \end{aligned}$$

with the notation $g(t_0+0) = \lim_{t\downarrow t_0} g(t)$, and $g(t_0-0) = \lim_{t\uparrow t_0} g(t)$, for a function $g(t)$ with left and right limits at $t = t_0$.

Proof. Treating $x(t)$ as a generalized function, we obtain from (3.3)

$$Q(-\frac{d}{dt}) Q(\frac{d}{dt}) x(t) = \int_{\mathbb{R}} e^{-i\omega t} \phi_0(\omega)\, d\omega, \quad t \in \mathbb{R}$$

i.e. $\phi_0(\omega)$ is the Fourier transform (in the sense of distributions) of $Q(-\frac{d}{dt})Q(\frac{d}{dt})$ $x(t)$. To recover ϕ_0, let $z(t) = Q(\frac{d}{dt})x(t)$ and note that $z^{(k)}(t) = Q(\frac{d}{dt})w_+^{(k)}(t) = 0$ for $t < \Delta$, $k \geq 0$ (in particular for $k = p, \ldots, q-1$). Now, let $\psi(t)$ be a test function in $C^\infty(\mathbb{R})$ with compact support. The above properties of $z^{(k)}(t)$, Lemma 3.2 and a straightforward calculation yield

$$\int \left(Q(-\frac{d}{dt})z(t)\right) \psi(t)\,dt = -\sum_{k=1}^{q} \beta_k \sum_{j=0}^{k-1} (-1)^j$$
$$\{\sigma_{k-j-1}(\Delta)\psi^{(j)}(\Delta)+\sigma_{k-j-1}(\Delta+\tau)\psi^{(j)}(\Delta+\tau)\}$$
$$+\sum_{k=0}^{q} \beta_k (-1)^k \int_{\Delta}^{+\infty} z(t)\psi^{(k)}(t)\,dt$$

This together with the fact (see Lemma 3.1) that $z(t)$ has $p-1$ continuous derivatives and $Q(-\frac{d}{dt})z(t) = 0$ for t outside the interval $[\Delta, \Delta+\tau]$, imply

$$\int_{\mathbb{R}} \left(Q(-\frac{d}{dt})z(t)\right) \psi(t)\,dt = \int_{\mathbb{R}} [\sum_{j=0}^{q-p-1} \delta^{(j)}(t-\Delta) \sum_{k=j+p+1}^{q} \beta_k \sigma_{k-j-1}(\Delta)]\psi(t)\,dt$$
$$+\int_{\mathbb{R}} [\sum_{j=0}^{q-p-1} \delta^{(j)}(t-\Delta-\tau) \sum_{k=j+p+1}^{q} \beta_k \sigma_{k-j-1}(\Delta+\tau)]$$
$$\psi(t)\,dt + \int_{\Delta}^{\Delta+\tau} [Q(-\frac{d}{dt})Q(\frac{d}{dt})x(t)]\psi(t)\,dt$$

This quickly yields (3.10a). The proof of (3.10b) is identical. □

Remark. 1. It can be shown that if $\tau = +\infty$, $T = +\infty$, then formulas (3.10) hold without the second term in (3.10a), and without the first term in (3.10b).
2. If $\lambda > 0$ is an eigenvalue of A and $\phi_0^{(\lambda)}$ is a corresponding normalized eigenvector, and we define $\phi_1^{(\lambda)} = P_1\phi_0^{(\lambda)}$, then it is easily seen that the pair $(\phi_0^{(\lambda)}, \phi_1^{(\lambda)})$ satisfies (3.4) with λ_1 replaced by λ. The procedure for constructing the optimal transformations can be used to construct $\phi_0^{(\lambda)}$ and $\phi_1^{(\lambda)}$, and establish that λ is also an eigenvalue of K (given by (3.9). Conversely, if λ is an eigenvalue of K, then reversing the steps in our construction one obtains that λ is an eigenvalue of A. In particular, K is a positive-definite matrix and its largest eigenvalue is exactly equal to $\lambda_1 = \|A\|$.

Appendix

Poof of Equation (2.5): The representation of $\phi_N(\omega)$ and $P_{01}\phi_N$ given above (2.5), yield

$$\phi_N(\omega)\overline{(P_{01}\phi_N)(\omega)} =$$
$$\sum_{k,\ell=0}^{n-1} \beta_k^{(N)} \overline{\alpha_\ell^{(N)}} (-1)^\ell (i\omega)^{k+\ell} e^{i\omega(\Delta+\tau+T)}$$
$$+\sum_{k=0}^{n-1} \beta_k^{(N)} \int_{-T}^{0} \overline{v_N(s)} (1+i\omega)^n (i\omega)^k e^{i\omega(\Delta+\tau-s)}\,ds$$

$$+\sum_{k=0}^{n-1}\overline{\alpha_k^{(N)}}\int_{\Delta}^{\Delta+\tau} u_N(t)(1-i\omega)^n(-1)^k(i\omega)^k e^{i\omega(T+t)}\,dt$$
$$+\int_{-T}^{0}\int_{\Delta}^{\Delta+\tau}(1+\omega^2)^n e^{i\omega(t-s)}u_N(t)\overline{v_N(s)}\,dtds \tag{A.1}$$

This will yield (2.5) via a standard regularization procedure [2, pp.132–134]. Let $\lambda(\omega)$ be a real-valued, C^∞ function with compact support, satisfying $\lambda(0)=1$. We multiply both sides of (A.1) by $f(\omega)\lambda(\omega/\rho)$, $\rho>0$, and integrate with respect to ω. The left hand side (LHS) of the resulting equation converges, as $\rho\to+\infty$, to the LHS of (2.5) by using $\lambda(\omega/\rho)\to\lambda(0)=1$, the boundedness of $\lambda(\omega)$, and the Lebesgue dominated convergence theorem. The convergence of the right hand side of the integrated form of (A.1) to that of (2.5) is easily deduced from the following Lemma.

Lemma. *Let $\lambda(\omega)$ be as above, $\rho>0$, and define $R_\rho^{(k)}(t)=\int_{\mathbb{R}}(i\omega)^k f(\omega)\lambda(\omega/\rho)$ $e^{i\omega t}\,d\omega$. Under the assumptions on the covariance function in Theorem 2.1, we have*

$$\lim_{\rho\uparrow+\infty} R_\rho^{(k)}(t)=R^{(k)}(t) \tag{A.2}$$

where the convergence is pointwise in $I=[\Delta,\Delta+\tau+T]$ for $k<2n$, and dt-a.e. in I for $k=2n$. Moreover, (A.2) also holds in the $L^2[I,dt]$ sense for $k\le 2n$.

Proof. The proof follows closely that of Proposition 3.2.2 of [2]. Let $\Lambda(t)$ be the inverse Fourier transform of $\lambda(\omega)$. The assumptions on $\lambda(\omega)$ imply [5, p.155] that $\Lambda(t)$ is a slowly decreasing function. Thus for every pair of integers $m>0,\ k\ge 0$, there exist a constant $C=C_m^{(k)}$ such that $|s|^m|\chi^{(k)}(s)|\le C$. To prove the first part of the lemma, we note that the square integrability of $R^{(k)}$ ($k\le 2n$) in I, implies that almost every point in I is a Lebesgue point of $R^{(k)}$ [19, p.138], i.e. for every $\epsilon>0$, there exists $\delta=\delta(t,\epsilon)$ such that

$$\int_{-s}^{s}|R^{(k)}(t-s')-R^{(k)}(t)|\,ds'\le\epsilon\,\delta\qquad\text{for all } 0\le s\le\delta. \tag{A.3}$$

This together with the slowly decreasing property of $\Lambda(t)$ will yield that (A.2) holds for every Lebesgue point of $R^{(k)}$, $k\le 2n$ (hence everywhere for $k<2n$). We write ($k\le 2n$)

$$R_\rho^{(k)}(t)-R^{(k)}(t)=\Big(\int_{-\delta}^{\delta}+\int_{|s|>\delta}\Big)[\rho^{k+1}R(t-s)\Lambda^{(k)}(\rho s)$$
$$-\rho R^{(k)}(t)\Lambda(\rho s)]\,ds \tag{A.4}$$

Let $I_1=I_1(t;\rho,\delta)$, $I_2=I_2(t;\rho,\delta)$ be the first and second integral, respectively. To bound I_1 we integrate by parts to obtain $|I_1|\le|\sum_{\ell=0}^{k-1}\sum_{l=0}^{1}(-1)^l\rho^{k-\ell}$ $R^{(\ell)}(t-(-1)^l\delta)\Lambda^{(k-\ell-1)}((-1)^l\rho\delta)|+|\int_{-\delta}^{\delta}[R^{(k)}(t-s)-R^{(k)}(t)]\rho\Lambda(\rho s)\,ds|$. The last term, to be denoted by $I_{1,2}$, is bounded as follows. For any Lebesgue point t, the function $u(s) = \int_0^s |R^{(k)}(t - s') - R^{(k)}(t)|\, ds'$ satisfies $|u(s)| \leq \epsilon\, \delta$, for all $|s| \leq \delta$. If $v(s) = (1 + s^2)^{-1}$, then by the slowly decreasing property of $\Lambda(t)$ we have with some constant C

$$\begin{aligned}|I_{1,2}| &\leq C \int_{-\delta}^{\delta} |R^{(k)}(t - s) - R^{(k)}(t)|\ \rho v(\rho s)\, ds \\ &\leq C([u(\delta) - u(-\delta)]\ \rho v(\rho\delta) + \rho \int_{-\delta}^{\delta} u(s)\, d(-v(\rho s))) \leq \epsilon\, C\, (1 + \pi)\end{aligned} \tag{A.5}$$

Now we bound I_2. Using the property $|R(s)| \leq R(0)$, and the slowly decreasing property of $\Lambda(t)$ we obtain (with some constant C) $|I_2| \leq \left| \int_{|s|>\delta} \rho^{k+1} R(t - s)\Lambda^{(k)}(\rho s) ds \right| + |R^{(k)}(t)| \int_{|s|>\delta} \rho|\Lambda(\rho s)|ds \leq C R(0)\rho^r \int_{|s|>\rho\delta} |s|^{-(k+2)} ds + |R^{(k)}(t)| \int_{|s|>\rho\delta} |\Lambda(s)|\, ds$, which quickly implies $\lim_{\rho\uparrow+\infty} I_2 = 0$.

Next we show that (A.2) holds in the $L^2(\mathrm{I}, dt)$ sense. Let $||\cdot||_2$ denote the norm of $L^2(\mathrm{I}, dt)$. Note that for any $\epsilon > 0$ there exist a $\delta > 0$ for which [2, p. 140] $||R^{(k)}(\cdot - s) - R^{(k)}(\cdot)||_2 \leq \epsilon$, for all $|s| \leq \delta$. This L^2 property replaces (A.3) in the present case. With I_1 and I_2 defined below (A.4), it suffices to show that $\lim_{\rho\uparrow+\infty}(||I_1||_2 + ||I_2||_2) = 0$. Preceeding as in the proof of the first part of the lemma, we get $||I_2||_2 \leq ||R^{(k)}||_2 \int_{|s|>\rho\delta} |\Lambda(s)|\, ds + 2\ \ R(0) \times K/(\delta^{k+1}\rho) \times \sqrt{\tau + T}$, which yields $\lim_{\rho\uparrow+\infty} ||I_2||_2 = 0$. Similarly, the above L^2 property yields $||I_1||_2 \leq \sum_{\ell=0,\, l=0}^{k-1,1} \rho^{k-\ell} ||R^{(\ell)}(\cdot - (-1)^l\delta)||_2\ \ |\Lambda^{(k-\ell-1)}((-1)^l\rho\delta)| + \epsilon\, \delta||\Lambda||_1$. Then the slowly decreasing property of $\Lambda(s)$ yield $\lim_{\rho\uparrow+\infty} ||I_1||_2 \leq \epsilon\, \delta||\Lambda||_1$. This completes the proof of the Lemma. □

Acknowledgements. We thank one of the referees for bringing Yaglom's paper to our attention, and the other referee for his constructive remarks. We also thank the referee whose question on equation (2.5) led to the Appendix.

References

[1] Breiman, L., Friedman, J. H.: Estimating optimal transformations for multiple regression and correlation (with discussion). J. Am. Statis. Association **80**, 580–619 (1985)
[2] Butzer, P. L., Nessel, R. J.: Fourier Analysis and Approximation. New York: Academic Press, 1971
[3] Dym, H., McKean, H. P.: Gaussian Processes, Function Theory, and the Inverse Spectral Theorem. Academic Press, Inc., New York, 1976
[4] Garnett, J.: Bounded Analytic Functions. Academic Press, New York, 1981
[5] Gelfand, I. M., Shilov, G. E.: Generalized Functions. Volume I. (Translated by Eugene Saletan). Academic Press, New York, 1964
[6] Gelfand, I. M., Yaglom, A. M.: Calculation of the amount of information about a random function contained in another such function. Uspekhi Mat. Nauk **12** (73), 3–52 (1957)
[7] Gidas, B., Murua, A.: Estimation and consistency for certain linear functionals of continuous-time processes from discrete data. Technical Report 473, Department of Statistics, University of Washington, Seattle, Washington, USA, 2004

[8] Gidas, B., Murua, A.: Classification and clustering of stop consonants via nonparametric transformations and wavelets. In Int. Conf. Acoustics, Speech, and Signal Processing **1**, 872–875 (1995)

[9] Gidas, B., Murua, A.: Stop consonants discrimination and clustering using nonlinear transformations and wavelets. In: (eds). Levinson, S. E., Shepp, L. editors. Image Models (And Their Speech Model Cousins), v. 80 of *IMA*, pp. 13–62. Springer-Verlag, 1996

[10] Gidas, B., Murua, A.: Optimal transformations for prediction in continuous time stochastic processes. In: (eds). Rajput, B., Karatzas, I., Taqqu, M. Stochastic Processes and Related Topics: In Memory of Stamatis Cambanis, pp. 167–183. Birkhäuser, 1998

[11] Helson, H., Szegö, G.: A problem in prediction theory. Ann. Mat. Pura Appl. **51**, 107–138 (1960)

[12] Ibragimov, I. A., Rozanov, Y. A.: Gaussian Random Processes. Springer-Verlag, New York, 1978

[13] Koosis, P.: The Logarithmic Integral I. Cambridge University Press, 1988

[14] Koosis, P.: The Logarithmic Integral II. Cambridge University Press, 1992

[15] Partington, J. R.: An Introduction to Hankel Operators. Cambridge University Press, 1988

[16] Peller, V. V., Krushchev, S. V.: Hankel Operators, Best Approximations, and Stationary Gaussian Processes. Russian Math. Surveys **37** (1), 61–144 (1982)

[17] Reed, M., Simon, B.: Methods of Modern Mathematical Physics. **2**, Academic Press, New York, 1974

[18] Rozanov, Y. A.: Stationary Random Processes. Holden-Day, San Francisco, 1967

[19] Rudin, W.: Real and Complex Analysis. McGraw-Hill, Inc., third edition, 1987

[20] Yaglom, A. M.: Stationary Gaussian Processes Satisfying the Strong Mixing Condition and Best Predictable Functionals. Bernoulli-Bayes-Laplace Anniversary Vol., Proceedings of International Research Seminar, Stat. Lab., Univ. of California, Berkeley, Springer-Verlag 1965, pp. 241–252