STOCHASTIC LINEARIZATION OF NONLINEAR POINT DISSIPATIVE SYSTEMS

JAMES A. RENEKE

Stochastic linearization produces a linear system with the same covariance kernel as the original nonlinear system. The method passes from factorization of finite-dimensional covariance kernels through convergence results to the final input/output operator representation of the linear system.

2000 Mathematics Subject Classification: 34K23, 93B18.

1. Introduction. Linearization and hence stochastic linearization of a nonlinear system is about local behavior of the system in time and space. Since the system functions of monitoring and control are concerned with local behavior, they are usually based on linearizations of the underlying nonlinear system [19]. The nonparametric methods of linearization which are the subject of this investigation are based on the covariances of the input and output processes for the system. The data typically looks like Figure 1.1. Existence of the covariance is implied by the assumption that the underlying nonlinear system is point dissipative [3], that is, there is a compact set which each trajectory of the system without stochastic excitation enters and remains within.

Nonparametric methods of linearization which only require observations of inputs and outputs rather than models of the nonlinear system are potentially useful in two situations [5]: first, when the system is evolving in time or is frequently reconfigured and model updates are difficult or expensive to obtain; second, when the monitoring or control functions are to be exercised at a low level by smart devices without the high level logic required for choosing or changing the system model.

The covariance function R of a zero-mean output process determines a reproducing kernel Hilbert (RKH) space with kernel R. This RKH space is said to represent the output process [21] and has been exploited in signal analysis [28]. In a reasonable sense, the RKH space representation of the process contains all of the information on the process available from observations. Starting with a known linear system excited by a Wiener process [24] provides an explicit representation of the RKH space as a space of Hellinger integrable functions. Further, the linear input/output operator for the system provides a factorization of the nonnegative Hermitian operator on the space of Hellinger integrable functions with matrix representation R.

When the underlying system is nonlinear, we show, in Section 3, that factoring discrete versions of R yields in the limit the matrix representation of a linearization of the nonlinear system. This stochastic linearization is the best possible in that when excited
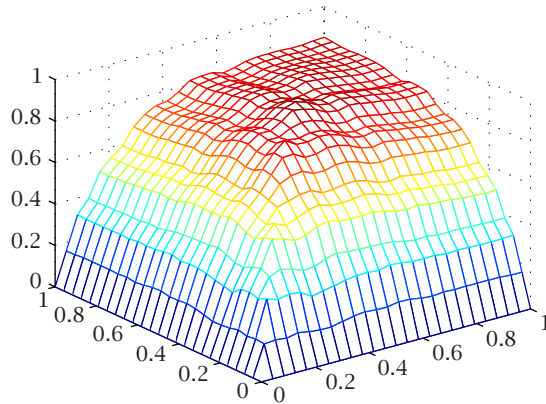
FIGURE 1.1. Sample covariance kernel, where $0 \leq x \leq 1$, $0 \leq y \leq 1$, and $z = R(x, y)$.

by a Wiener process the linearization yields a process with covariance R, that is, one which is indistinguishable from the original process. A natural way to proceed, building on our experience with linear systems, is to seek a factorization of R in terms of limits of the Cholesky factors [11] of discrete approximations R_t of R. The next section provides the background for a reasonable notion of convergence required to make this approach feasible.

2. Background. The RKH space approach to linear system modeling [24] provides discrete nonparametric model representations in terms of factorizations of the discrete covariances of the input and output processes for the system. Thus the representations are in terms of data, avoiding the dimension or order problem associated with parametric approaches. The RKH space method eliminates decisions about the form of the model, such as the number of terms to be included, which require a high level logic.

Let R^d denote the space of d-tuples of real numbers with the usual inner product $\langle \cdot, \cdot \rangle$ and norm $|\cdot|$. Let G denote the class of continuous functions f from $[0, \infty)$ into R^d such that $f(0) = 0$. We define a family of pseudonorms $\{N_x, x \geq 0\}$ on G by $N_x(f) = \sup_{z \leq x} |f(z)|$ for each f in G and $x \geq 0$. Two classes of linear operators defined initially on G are introduced. These operators are used to describe the systems of interest. Let $\mathcal{B}$ denote the set of linear operators on G to which B belongs only when

(1) $[Bf](0) = 0$ for each f in G,
(2) for each $T > 0$ there is a number b such that

$$\left| [Bf](t) - [Bf](s) \right| \leq b \int_s^t N_x(f) dx \tag{2.1}$$

for each f in G and $0 \leq s \leq t \leq T$.

Let $\mathcal{A}$ denote the set of linear operators on G to which A belongs only in case $A - I$ is in $\mathcal{B}$. If B is in $\mathcal{B}$, then $I - B$ is an invertible operator from G onto G and $(I - B)^{-1}$ is in $\mathcal{A}$. If A is in $\mathcal{A}$, then A is an invertible operator from G onto G and $I - A^{-1}$ is in $\mathcal{B}$.

The classes of operators $\mathcal{A}$ and $\mathcal{B}$ can serve to describe linear systems, but what is their relation with nonlinear systems? In particular, how do we relate these operators to the available observation process $\{Y(t),\ 0 \leq t\}$? The next three theorems provide answers to these questions.

In general, the covariance function R of a stochastic process $\{Y(t),\ 0 \leq t\}$ defined by

$$R(s,t) = E\big((Y(s) - E(Y(s)))(Y(t) - E(Y(t)))^T\big), \tag{2.2}$$

where E denotes the expectation operator, is nonnegative, that is,

$$\sum_{p,q=0}^{n} \langle R(t_p, t_q)x_q, x_p \rangle \geq 0 \tag{2.3}$$

for each sequence $\{t_p\}_0^n$ in $[0,\infty)$ and each sequence $\{x_p\}_0^n$ in R^d [18, 21].

From this point we will reserve R to denote the covariance function of some observation process $\{Y(t),\ 0 \leq t\}$. In order to see the structure of the problem, modeling the process $\{Y(t),\ 0 \leq t\}$ from partial information, we begin by assuming complete information, that is, $R(s,t)$ is known exactly for $0 \leq s,t$.

THEOREM 2.1 [1, 18]. *For each covariance function R from $[0,\infty) \times [0,\infty)$ into the $d \times d$ matrices, there is a Hilbert space $\{G_R, Q_R\}$ of functions from $[0,\infty)$ into R^d with reproducing kernel R, that is,*
  (1) $R(\cdot,t)x$ *is in G_R for each t in $[0,\infty)$ and x in R^d,*
  (2) $Q_R(f, R(\cdot,t)x) = \langle f(t), x \rangle$ *for each f in G_R, x in R^d, and t in $[0,\infty)$.*

The theorem only asserts the existence of the RKH space $\{G_R, Q_R\}$ with kernel R.

When our observation process is of the form $Y = AW$, where A is an invertible operator in $\mathcal{A} \cup \mathcal{B}$, W is the standard d-dimensional Wiener process, and $R(s,t) = EY(s)Y(t)^T$, we can obtain more, namely, an explicit representation of $\{G_R, Q_R\}$. (See [21] for an alternative representation.)

In order to accomplish this we will introduce another RKH space, this time associated with the input process. Let k denote an increasing scalar function with $k(0) = 0$. Let G_K denote the subspace of functions in G which are Hellinger integrable with respect to k, that is, f is in G_K only in case there is a number M such that

$$\sum_{p=1}^{n} \frac{|f(t_p) - f(t_{p-1})|^2}{(k(t_p) - k(t_{p-1}))} = \sum_{t} \frac{|df|^2}{dk} \leq M \tag{2.4}$$

for each increasing sequence $\{t_p\}_0^n$ in $[0,\infty)$. The least such number M is denoted by $\int_0^\infty |df|^2/dk$. Finally, let Q_K denote the inner product for G_K defined by $Q_K(f,g) = \int_0^\infty df\, dg/dk$, the limit through refinement of sums $\sum_t df\, dg/dk$. We will use the short notation $dk(x,y)$ for the difference $k(y) - k(x)$.

The space $\{G_K, Q_K\}$ is an RKH space with kernel given by $K(s,t) = k(\min(s,t))I$, where I is the $d \times d$ identity matrix [18]. Elements of $\mathcal{B}$ map G into G_K and elements of $\mathcal{A}$ map G_K onto G_K. From now on, we will be concerned primarily with the restrictions of elements of $\mathcal{A}$ and $\mathcal{B}$ to G_K.

Let L denote the function from $\mathcal{A} \cup \mathcal{B}$ into the space of $d \times d$ matrix-valued functions on $[0, \infty) \times [0, \infty)$ defined by $LD(s,t) = [D^*K(\cdot, t)](s) = [DK(\cdot, s)](t)^T$, for D in $\mathcal{A} \cup \mathcal{B}$ and (s,t) in $[0, \infty) \times [0, \infty)$. We use D^* to denote the adjoint in $\{G_K, Q_K\}$ of the restriction of D to G_K and call LD the matrix representation of D. Notice that $\langle [Df](t), x \rangle = Q_K(f, LD(\cdot, t)x)$, for each f in G_K and x in $[0, \infty)$.

In general, the variance of a scalar input process is an increasing function k. When the input process is the standard scalar Wiener process, we can use the special case $k(t) = t$. Note that, when $d = 1$, K is the covariance function of the Wiener process.

THEOREM 2.2 [24]. *Suppose R is the covariance of the process $Y = AW$, where A is an invertible operator in $\mathcal{A} \cup \mathcal{B}$, and W is the d-dimensional Wiener process. For $0 \leq s, t$, $R(s,t) = [AA^*K(\cdot, t)](s)$, where A^* is the adjoint of A in $\{G_K, Q_K\}$ and K is the reproducing kernel of $\{G_K, Q_K\}$, that is, R is the matrix representation of AA^*.*

Let LA denote the matrix representation of the assumed operator A. We can write

$$R(s,t) = Q_K(LA(\cdot, t), LA(\cdot, s)) \tag{2.5}$$

for $0 \leq s, t$, that is, we can use A to obtain a representation of R.

The following example illustrates this last observation. Note that a state space formulation of the model would have to be infinite dimensional; however, the input and output processes are scalar.

EXAMPLE 2.3. Suppose W is the standard scalar Wiener process and

$$Y(t) = [AW](t) = \int_0^t \frac{1}{t - u + 1} dW(u). \tag{2.6}$$

Direct calculation for $s \leq t$ yields

$$R(s,t) = EY(s)Y(t) = \int_0^s \frac{1}{s - u + 1} \frac{1}{t - u + 1} du$$

$$= \begin{cases} \dfrac{1}{t - s} \ln \left[\dfrac{(s+1)(t-s+1)}{t+1} \right] & \text{if } 0 \leq s < t, \\ \dfrac{s}{s + 1} & \text{if } s = t. \end{cases} \tag{2.7}$$

If $s \leq t$, then

$$LA(s,t) = [AK(\ , s)](t) = \int_0^t \frac{1}{t - u + 1} dK(u, s)$$

$$= \int_0^s \frac{1}{t - u + 1} du = \ln(t + 1) - \ln(t - s + 1). \tag{2.8}$$

If $t \leq s$, then $LA(s,t) = LA(t,t)$. Assuming $s \leq t$, Theorem 2.2 yields

$$R(s,t) = Q_K(LA(, s), LA(, t)) = \int_0^s \frac{1}{s - u + 1} \frac{1}{t - u + 1} du \tag{2.9}$$

which agrees with the direct calculation.

Theorem 2.4 [24]. *Given that A in $\mathcal{A} \cup \mathcal{B}$ is invertible and R is represented in terms of A (see (2.5)), the RKH space with kernel R is given by $G_R = G_K$ and $Q_R(f,g) = Q_K(f,(AA^*)^{-1}g)$, for each f and g in G_R.*

For our problem, that is, R associated with a general observation process $\{Y(t), 0 \leq t\}$, the underlying system might be nonlinear and the linear operator A assumed in Theorems 2.2 and 2.4 unavailable. We seek a linearization in $\mathcal{A} \cup \mathcal{B}$ of the underlying system, which will play the role of A, through a factorization of the covariance function R. Since R can be factored in many different ways, we will have to justify our choice in the end. The method returns an element of $\mathcal{A} \cup \mathcal{B}$, which we will denote by A, with matrix representation LA.

Finite-dimensional approximations. For calculations, the matrix representations of the operators have to be projected down to finite-dimensional spaces. A more detailed explication appears in [24].

A class of polygonal functions, the *K-polygonal functions*, arises naturally in RKH spaces and can be used along with projection methods to develop finite-dimensional approximations to system operators. Any function f on $[0,\infty)$ of the form

$$f(s) = \sum_{p=0}^{n} K(s,t_p)x_p, \tag{2.10}$$

where $t = \{t_p\}_0^n$ is an increasing sequence in $[0,\infty)$ and $\{x_p\}_0^n$ is a sequence in R^d, is called a *K-polygonal function*. The subspace of all K-polygonal functions based on a fixed increasing finite sequence t in $[0,\infty)$ is a closed linear subspace of G_K. We let Π_t denote the orthogonal projection of G_K onto this subspace. Also, let P_t denote the projection on G_K defined by

$$[P_t f](s) = \begin{cases} f(s) & \text{if } s \leq t, \\ f(t) & \text{if } t < s, \end{cases} \tag{2.11}$$

for each f in G_K and $0 \leq s, t$.

Theorem 2.5 [1, 18]. *For each positive number T, the union of the finite-dimensional subspaces $\Pi_t G_K$, t a partition of $[0,T]$, is dense in $P_T G_K$ with respect to the inner product norm $N_K(f) = Q_K(f,f)^{1/2}$.*

For convenience and clarity we restrict our attention in the rest of the paper to the case $d = 1$ (observations and inputs are both scalar). This, of course, does not restrict the underlying dynamical system to be one-dimensional. (See Example 2.3.) Furthermore, we assume k is an increasing function on $[0,\infty)$ with $k(0) = 0$. Recall that $K(s,t) = k(\min(s,t))$.

For f in G_K and $\{t_p\}_0^n$ an increasing sequence in $[0,\infty)$, let $f_t = (f(t_0), f(t_1), \ldots, f(t_n))^T$. Similarly, let K_t denote the $(n+1) \times (n+1)$ matrix whose (p,q) element is

given by $K_t(p,q) = K(t_{p-1}, t_{q-1})$ for $1 \le p, q \le n+1$. If

$$[\Pi_t f](s) = \sum_{p=0}^{n} K(s, t_p) x_p \tag{2.12}$$

and $x = (x_0, x_1, \ldots, x_n)^T$, then $x = (K_t)^{-1} f_t$ [24]. Here $(K_t)^{-1}$ is a pseudoinverse of K_t, that is, $(K_t)^{-1} = M$ is an $(n+1) \times (n+1)$ matrix such that $M(1,q) = M(p,1) = 0$ for $1 \le p, q \le n+1$, and $M(2:n+1, 2:n+1)$ is the inverse of $K_t(2:n+1, 2:n+1)$.

Suppose A is in $\mathcal{A} \cup \mathcal{B}$ with matrix representation LA. Let LA_t denote the $(n+1) \times (n+1)$ matrix whose (p,q) element is given by $LA_t(p,q) = LA(t_{p-1}, t_{q-1})$, where $\{t_p\}_0^n$ is an increasing sequence in $[0, \infty)$. We will use the same notation for various functions without comment. For instance, $R_t(p,q) = R(t_{p-1}, t_{q-1})$ for $1 \le p, q \le n+1$. With this understanding, for s in $[0, \infty)$,

$$[A\Pi_t f](s) = \sum_{q=0}^{n} [AK(\cdot, t_q)](s) x_q = \sum_{q=0}^{n} LA(t_q, s)^T x_q \tag{2.13}$$

and, for $p = 1, 2, \ldots, n+1$,

$$[A\Pi_t f](t_{p-1}) = [(LA_t)^T (K_t)^{-1} f_t](p). \tag{2.14}$$

Thus if $h = \Pi_t A \Pi_t f$, then $h_t = (LA_t)^T (K_t)^{-1} f_t$. Note that the finite-dimensional approximations converge [24] but covergence is not a sequential convergence but rather a net convergence, that is, through refinements of partitions.

THEOREM 2.6. *Suppose A is in $\mathcal{A} \cup \mathcal{B}$ and R is the matrix representation of AA^*. For each pair of positive numbers c and T, there is a partition s of $[0, T]$ such that if $\{t_q\}_0^n$ refines s, then*

$$\left| R_t(p,q) - Q_K(\Pi_t A^* K(\cdot, t_{p-1}), \Pi_t A^* K(\cdot, t_{q-1})) \right| < c \tag{2.15}$$

for $p, q = 1, 2, \ldots, n+1$.

INDICATION OF PROOF. Let $\{r_p\}_0^m$ be a partition of $[0, T]$ such that

$$(dk(r_{p-1}, r_p))^{1/2} < \frac{c}{8k(T)(N_K(P_T A^*))^2} \tag{2.16}$$

for $p = 1, 2, \ldots, m$. If $r_{p-1} \le u < v \le r_p$, then

$$N_K(K(\cdot, v) - K(\cdot, u)) = \left(\int_u^v \frac{|dk|^2}{dk} \right)^{1/2} = (dk(u, v))^{1/2} < \frac{c}{8k(T)(N_K(P_T A^*))^2}. \tag{2.17}$$

Let s be a partition of $[0,T]$ refining r such that if $\{t_p\}_0^n$ refines s, then

$$N_K(P_T A^* K(\cdot, r_p) - \Pi_t A^* K(\cdot, r_p)) < \frac{c}{4k(T)N_K(P_T A^*)} \tag{2.18}$$

for $p = 0, 1, 2, \ldots, m$. If $r_{p-1} < t_q \le r_p$, then

$$
\begin{aligned}
& N_K(P_T A^* K(\cdot, t_q) - \Pi_t A^* K(\cdot, t_q)) \\
& \quad \le N_K(P_T A^* K(\cdot, t_q) - P_t A^* K(\cdot, r_p)) \\
& \qquad + N_K(P_T A^* K(\cdot, r_p) - \Pi_t A^* K(\cdot, r_p)) + N_K(\Pi_t A^* K(\cdot, r_p) - \Pi_t A^* K(\cdot, t_q)) \\
& \quad < 2N_T(P_T A^*) dk(t_q, r_p) + \frac{c}{4k(T)N_K(P_T A^*)} \\
& \quad < \frac{c}{2k(T)N_K(P_T A^*)}.
\end{aligned} \tag{2.19}
$$

Hence

$$
\begin{aligned}
& \left| R_t(p,q) - Q_K(\Pi_t A^* K(\cdot, t_{p-1}), \Pi_t A^* K(\cdot, t_{q-1})) \right| \\
& \quad = \big| Q_K(P_T A^* K(\cdot, t_{p-1}), P_T A^* K(\cdot, t_{q-1})) \\
& \qquad - Q_K(\Pi_t A^* K(\cdot, t_{p-1}), \Pi_t A^* K(\cdot, t_{q-1})) \big| \\
& \quad \le \big| Q_K(P_T A^* K(\cdot, t_{p-1}) - \Pi_t A^* K(\cdot, t_{p-1}), P_T A^* K(\cdot, t_{q-1})) \big| \\
& \qquad + \big| Q_K(\Pi_t A^* K(\cdot, t_{p-1}), P_T A^* K(\cdot, t_{q-1}) - \Pi_t A^* K(\cdot, t_{q-1})) \big| \\
& \quad \le N_K(P_T A^* K(\cdot, t_{p-1}) - \Pi_t A^* K(\cdot, t_{p-1})) N_K(P_T A^* K(\cdot, t_{q-1})) \\
& \qquad + N_K(\Pi_t A^* K(\cdot, t_{p-1})) N_K(P_T A^* K(\cdot, t_{q-1}) - \Pi_t A^* K(\cdot, t_{q-1})) \\
& \quad < c. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square
\end{aligned} \tag{2.20}
$$

CHOLESKY FACTORIZATIONS. The upper Cholesky factor of a nonnegative symmetric matrix S is an upper triangular matrix S^u with nonnegative diagonals such that $(S^u)^T S^u = S$.

We can tie R to $\{G_K, Q_K\}$ without supposing the existence of a continuous linear transformation A by assuming in the rest of the paper that for each positive number T there is a positive number c such that if $[x, y]$ and $[u, v]$ are subintervals of $[0, T]$, then

$$\left| R(y,v) - R(y,u) - R(x,v) + R(x,u) \right| \le c \, dk(x,y) dk(u,v). \tag{2.21}$$

With this assumption R is the matrix representation of a nonnegative Hermitian member of $\mathcal{A} \cup \mathcal{B}$ [24] which we will denote by H in the rest of the paper. What happens when $H = AA^*$ and A is time-invariant?

THEOREM 2.7 [24]. *Suppose A is a time-invariant operator in $\mathcal{A} \cup \mathcal{B}$. There is a continuous function M on $[0,\infty)$ such that for each positive number T and partition t of $[0,T]$,*

$$(K_t^u)^{-T} L A_t$$

$$= \begin{bmatrix} \dfrac{\int_0^{t_1} M(t_1 - u)\,du}{\sqrt{k_t(2)}} & \dfrac{\int_0^{t_1} M(t_2 - u)\,du}{\sqrt{k_t(2)}} & \dfrac{\int_0^{t_1} M(t_3 - u)\,du}{\sqrt{k_t(2)}} & \dfrac{\int_0^{t_1} M(t_4 - u)\,du}{\sqrt{k_t(2)}} & \cdots \\[4mm] 0 & \dfrac{\int_{t_1}^{t_2} M(t_2 - u)\,du}{\sqrt{dk_t(2,3)}} & \dfrac{\int_{t_1}^{t_2} M(t_3 - u)\,du}{\sqrt{dk_t(2,3)}} & \dfrac{\int_{t_1}^{t_2} M(t_4 - u)\,du}{\sqrt{dk_t(2,3)}} & \cdots \\[4mm] 0 & 0 & \dfrac{\int_{t_2}^{t_3} M(t_3 - u)\,du}{\sqrt{dk_t(3,4)}} & \dfrac{\int_{t_2}^{t_3} M(t_4 - u)\,du}{\sqrt{dk_t(3,4)}} & \cdots \\[4mm] 0 & 0 & 0 & \dfrac{\int_{t_3}^{t_4} M(t_4 - u)\,du}{\sqrt{dk_t(4,5)}} & \cdots \\[4mm] \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{2.22}$$

Therefore, assuming equally spaced partition points, the diagonal elements of $(K_t^u)^{-T} L A_t$ all have the same sign. Hence $\pm (K_t^u)^{-T} L A_t$ is an upper Cholesky factor. Further,

$$((K_t^u)^{-T} L A_t)^T (K_t^u)^{-T} L A_t \approx R_t. \tag{2.23}$$

In our problem the underlying system is nonlinear and we do not start with a factorization of H. Furthermore, the factorization we seek is not necessarily in terms of time-invariant linear operators. Even so, this result suggests that we seek a factorization of R as a limit in some sense of the upper Cholesky factors R_t^u of R_t. We want convergence in terms of finite-dimensional linear operators associated with the upper Cholesky factors R_t^u in the following way.

For each positive number T and partition t of $[0,T]$, let A_t denote the linear transformation of G_K defined for each f in G_K by

$$[A_t f](u) = \frac{k(t_p) - k(u)}{k(t_p) - k(t_{p-1})} [(R_t^u)^T (K_t^u)^{-T} f_t](p) \\ + \frac{k(u) - k(t_{p-1})}{k(t_p) - k(t_{p-1})} [(R_t^u)^T (K_t^u)^{-T} f_t](p+1), \tag{2.24}$$

where $t_{p-1} \le u < t_p$ for some $p = 1,2,\ldots,t^{-1}(T)$. If $T \le u$, then $[A_t f](u) = [A_t f](T)$.

THEOREM 2.8. *For each positive number T and partition t of $[0,T]$,*

$$[(A_t)^* f](u) = \frac{k(t_p) - k(u)}{k(t_p) - k(t_{p-1})} [(K_t^u)^T R_t^u (K_t^u)^{-1} f_t](p) \\ + \frac{k(u) - k(t_{p-1})}{k(t_p) - k(t_{p-1})} [(K_t^u)^T R_t^u (K_t^u)^{-1} f_t](p+1) \tag{2.25}$$

for each f in G_K and $0 \leq t_{p-1} \leq u < t_p \leq T$. If $T \leq u$, then

$$[(A_t)^* f](u) = [(K_t^u)^T R_t^u (K_t^u)^{-1} f_t](t^{-1}(T) + 1). \tag{2.26}$$

INDICATION OF PROOF.

$$\begin{aligned}
Q_K((A_t)^* f, g) &= Q_K(f, A_t g) \\
&= (f_t)^T (K_t)^{-1} ((K_t^u)^T R_t^u)^T (K_t)^{-1} g_t \qquad (2.27) \\
&= ((K_t^u)^T R_t^u (K_t)^{-1} f_t)^T (K_t)^{-1} g_t.
\end{aligned}$$

Hence the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

SUMMARY OF STANDING ASSUMPTIONS AND NOTATION FOR THE REST OF THE PA-PER. (1) Assume $d = 1$ and k is an increasing function on $[0, \infty)$ with $k(0) = 0$. Recall that $K(s, t) = k(\min(s, t))$ is the reproducing kernel of $\{G_K, Q_K\}$.

(2) Let R denote the covariance function for a scalar observation process $\{Y(t), 0 \leq t\}$. R is nonnegative (see (2.3)). Assume that R satisfies inequalities (2.21) and is the matrix representation of nonnegative Hermitian operator H in $\mathcal{A} \cup \mathcal{B}$.

(3) P_t and Π_t are projections given by (2.11) and (2.12), respectively.

(4) K_t and R_t are discretizations of K and R, respectively. For instance, $R_t(p, q) = R(t_{p-1}, t_{q-1})$.

(5) K_t^u and R_t^u are upper Cholesky factors of K_t and R_t, respectively. For instance, K_t^u is an upper triangular matrix with nonnegative diagonal such that $(K_t^u)^T K_t^u = K_t$.

(6) $\{A_t\}$ is a family of continuous linear transformations of $\{G_K, Q_K\}$ given by (2.24).

(7) $dk(x, y)$ is the difference $k(y) - k(x)$.

3. Main results. The objective is linearization of an unknown underlying nonlinear system generating the observation process $\{Y(t), 0 \leq t\}$ from data which we interpret as $\{R_t\}$. The quality of the linearization should be measurable in terms of the sampling rates and statistics of the observations. The first part of this objective can be achieved by establishing convergence in some reasonable sense of the finite-dimensional operators $\{A_t\}$. Conditions which imply convergence should be restricted to conditions on the data $\{R_t\}$ as opposed to conditions on the underlying system.

Given that A is in $\mathcal{A} \cup \mathcal{B}$, T is a positive number, and $0 \leq x \leq T$, we will say that the net $\{((A_t)^* - \Pi_t A^*) K(\cdot, x), t \text{ a partition of } [0, T] \text{ refining } \{0, x, T\}\}$ has limit 0 provided that for each positive number c there is a partition r of $[0, T]$ refining $\{0, x, T\}$ such that if t refines r, then $N_K(((A_t)^* - \Pi_t A^*) K(\cdot, x)) < c$. Further, given that T is a positive number and $0 \leq x \leq T$, we will say that the net $\{(A_t)^* K(\cdot, x), t \text{ a partition of } [0, T]$ refining $\{0, x, T\}\}$ is Cauchy provided there is a partition r of $[0, T]$ refining $\{0, x, T\}$ such that if s refines r and t refines s, then $N_K(((A_t)^* - (A_s)^*) K(\cdot, x)) < c$.

THEOREM 3.1. (1) *If A is in $\mathcal{A} \cup \mathcal{B}$, and for each positive number T and $0 \leq x \leq T$ the net*

$$\{((A_t)^* - \Pi_t A^*) K(\cdot, x), \ t \text{ a partition of } [0, T] \text{ refining } \{0, x, T\}\} \tag{3.1}$$

has limit 0, *then for each positive number* T *and* $0 \leq x \leq T$ *the net*

$$\{(A_t)^* K(\cdot, x), \ t \text{ a partition of } [0, T] \text{ refining } \{0, x, T\}\} \tag{3.2}$$

is Cauchy.

(2) *If, for each positive number* T *and* $0 \leq x \leq T$, *the net (3.2) is Cauchy, then there is a linear operator* A *in* $\mathcal{A} \cup \mathcal{B}$ *such that for each positive number* T *and* $0 \leq x \leq T$, *the net (3.1) has limit* 0.

INDICATION OF PROOF. Assume that the hypothesis of (1) holds, c and T are positive numbers, and $0 \leq x \leq T$. There is a partition r of $[0, T]$ refining $\{0, x, T\}$ such that if s refines r and t refines s, then $N_K(\Pi_t A^* K(\cdot, x) - \Pi_s A^* K(\cdot, x)) < c/3$. In addition, we may assume that $N_K((A_s)^* K(\cdot, x) - \Pi_s A^* K(\cdot, x)) < c/3$. Hence

$$
\begin{aligned}
N_K\big(&(A_t)^* K(\cdot, x) - (A_s)^* K(\cdot, x)\big) \\
&\leq N_K\big((A_t)^* K(\cdot, x) - \Pi_t A^* K(\cdot, x)\big) + N_K\big(\Pi_t A^* K(\cdot, x) - \Pi_s A^* K(\cdot, x)\big) \\
&\quad + N_K\big(\Pi_s A^* K(\cdot, x) - (A_s)^* K(\cdot, x)\big) \\
&< c,
\end{aligned}
\tag{3.3}
$$

that is, the net (3.2) is Cauchy.

Assume the hypothesis of (2) holds. For each $x \geq 0$ choose $T > x$ and let $LA(\cdot, x)$ denote the limit of (3.2). Note that $LA(\cdot, x)$ is in G_K and if $x = t_{q-1} < u$, then

$$
\begin{aligned}
\big[(A_t)^* K(\cdot, x)\big](u) &= [A_t K(\cdot, u)](x) \\
&= \big[(R_t^u)^T (K_t^u)^{-T} (K(\cdot, u))_t\big](q) \\
&= \sum_{i=1}^{q} R_t(i, q)(K_t^u)^{-T}(i, \cdot)(K(\cdot, u))_t \\
&= \sum_{i=1}^{q} R_t(i, q) \sum_{j=1}^{i} (K_t^u)^{-T} k_t(j) \\
&= [A_t K(\cdot, x)](x) \\
&= \big[(A_t)^* K(\cdot, x)\big](x),
\end{aligned}
\tag{3.4}
$$

that is, $LA(u, x) = LA(x, x)$. Therefore we may define a linear, causal function A on G_K by $[Af](x) = Q_K(f, LA(\cdot, x))$ for each f in G_K and $x \geq 0$.

Further, note that

$$
\begin{aligned}
N_K\big((A_t)^* f\big)^2 &= (f_t)^T (K_t)^{-1} (R_t^u)^T K_t^u (K_t)^{-1} (K_t^u)^T R_t^u (K_t)^{-1} f_t \\
&= (f_t)^T (K_t)^{-1} R_t (K_t)^{-1} f_t \\
&= Q_K(\Pi_t A A^* \Pi_t f, f) \\
&\leq N_K(A A^*) N_K^2(f),
\end{aligned}
\tag{3.5}
$$

that is, $N_K((A_t)^*) \leq N_K^{1/2}(A A^*)$.

Therefore, for each positive number T, pair of subintervals $[x, y]$ and $[u, v]$ of $[0, T]$, and positive number c there is a partition t of $[0, T]$ refining both $\{0, x, y, T\}$ and $\{0, u, v, T\}$ such that

$$
\begin{aligned}
& \big| LA(y, v) - LA(x, v) - LA(y, u) + LA(x, u) \big| \\
& \quad \leq c + \big| \big[(A_t)^* (K(\cdot, v) - K(\cdot, u)) \big](y) \\
& \qquad\quad - \big[(A_t)^* (K(\cdot, v) - K(\cdot, u)) \big](x) \big| \\
& \quad \leq c + N_K \big((A_t)^* (K(\cdot, v) - K(\cdot, u)) \big) (dk(x, y))^{1/2} \\
& \quad \leq c + N_K \big((A_t)^* \big) (dk(u, v))^{1/2} (dk(x, y))^{1/2} \\
& \quad \leq c + N_K^{1/2} (AA^*) (dk(u, v))^{1/2} (dk(x, y))^{1/2},
\end{aligned}
\tag{3.6}
$$

that is, the linear operator defined in terms of LA is in $\mathcal{A} \cup \mathcal{B}$ [24]. Further, R is the matrix representation of AA^*.

For each pair of positive numbers c and T and $0 \leq x \leq T$, there is a partition r of $[0, T]$ refining $\{0, x, T\}$ such that if t refines r, then $N_K(((A_t)^* - A^*)K(\cdot, x)) < c/2$, $N_K((A^* - \Pi_t A^*)K(\cdot, x)) < c/2$ and thus $N_K(((A_t)^* - \Pi_t A^*)K(\cdot, x)) < c$. Therefore (3.1) has limit 0.

Thus showing that $\{(A_t)^* K(\cdot, x)\}$ is Cauchy is more basic since we do not need to assume a factorization AA^* of the operator with matrix representation R. Our search then is for conditions on the finite-dimensional covariances $\{R_t\}$ which allow us to conclude that $\{(A_t)^* K(\cdot, x)\}$ is Cauchy. $\qquad\square$

THEOREM 3.2. *The following are equivalent.*

(1) *For each positive number T and $0 \leq x \leq T$, the net (3.2) is Cauchy.*

(2) *For each pair of positive numbers c and T and $0 \leq x \leq T$, there is a partition r of $[0, T]$ refining $\{0, x, T\}$ such that if s refines r and t refines s, then*

$$
\big| R(x, x) - Q_K((A_t)^* K(\cdot, x), (A_s)^* K(\cdot, x)) \big| < c.
\tag{3.7}
$$

INDICATION OF PROOF. Assume (1) and let A be the linear operator defined in the proof of Theorem 3.2(2). For each positive number T, $0 \leq x \leq T$, and partition t of $[0, T]$ refining $\{0, x, T\}$,

$$
\begin{aligned}
& \big| R(x, x) - Q_K((A_t)^* K(\cdot, x), (A_s)^* K(\cdot, x)) \big| \\
& \quad \leq \big| R(x, x) - Q_K(A^* K(\cdot, x), (A_s)^* K(\cdot, x)) \big| + \big| Q_K((A^* - (A_t)^*) K(\cdot, x), (A_s)^* K(\cdot, x)) \big| \\
& \quad = \big| Q_K(A^* K(\cdot, x), (A^* - (A_s)^*) K(\cdot, x)) \big| + \big| Q_K((A^* - (A_t)^*) K(\cdot, x), (A_s)^* K(\cdot, x)) \big|.
\end{aligned}
\tag{3.8}
$$

And hence (2) follows.

Assume that (2) holds, T is a positive number, and $0 \le x \le T$. If s is a partition of $[0,T]$ refining $\{0,x,T\}$ and t refines s, then

$$
\begin{aligned}
Q_K&\big((A_s)^*K(\cdot,x),(A_s)^*K(\cdot,x)\big)\\
&= \big((K_s^u)^T R_s^u (K_s)^{-1} K_s(\cdot,s^{-1}(x)+1)\big)^T \cdot (K_s)^{-1}(K_s^u)^T R_s^u (K_s)^{-1} K_s(\cdot,s^{-1}(x)+1)\\
&= \big(R_s^u(\cdot,s^{-1}(x)+1)\big)^T K_s^u (K_s)^{-1} R_s^u(\cdot,s^{-1}(x)+1)\\
&= \big(R_s^u(\cdot,s^{-1}(x)+1)\big)^T R_s^u(\cdot,s^{-1}(x)+1)\\
&= R(x,x).
\end{aligned}
\tag{3.9}
$$

Hence

$$
\begin{aligned}
N_K&\big(((A_t)^* - (A_s)^*)K(\cdot,x)\big)^2\\
&= N_K\big((A_t)^*K(\cdot,x)\big)^2 - 2Q_K\big((A_t)^*K(\cdot,x),(A_s)^*K(\cdot,x)\big)\\
&\quad + N_K\big((A_s)^*K(\cdot,x)\big)^2\\
&= 2R(x,x) - 2Q_K\big((A_t)^*K(\cdot,x),(A_s)^*K(\cdot,x)\big)
\end{aligned}
\tag{3.10}
$$

from which (1) follows. $\square$

THEOREM 3.3. *If T is a positive number, $\{s_p\}_{p=0}^n$ is a partition of $[0,T]$, t refines s, and $\{u_p\}_{p=0}^n$ is an increasing integer-valued sequence such that $s = t[u]$, then, for $1 \le p \le n+1$ and $1 \le q \le u(n)+1$,*

$$
\begin{aligned}
(K_s^u)^{-T}&\big(K_t^u(\cdot,u(I-1)+1)\big)^T(p,q)\\
&= \begin{cases}
0 & \text{if } p=1 \text{ or } q=1,\\
0 & \text{if } p \ge 2,\ q \le u(p-2)+1,\\
\dfrac{\sqrt{dk_t(q-1,q)}}{\sqrt{dk_s(p-1,p)}} & \text{if } p \ge 2,\ u(p-2)+1 < q \le u(p-1)+1,\\
0 & \text{if } p \ge 2,\ u(p-1)+1 < q.
\end{cases}
\end{aligned}
\tag{3.11}
$$

Furthermore,

$$
\begin{aligned}
\big((K_s^u)^T R_s^u(\cdot,p)\big)^T &(K_s)^{-1}\big(K_t^u(\cdot,u(I-1)+1)\big)^T R_t^u(\cdot,u(q-1)+1)\\
&= \sum_{i=2}^{p} \frac{R_s^u(i,p)}{\sqrt{dk_s(i-1,i)}} \sum_{j=u(i-2)+1}^{u(i-1)} \sqrt{dk_t(j,j+1)}\, R_t^u(j+1,u(q-1)+1).
\end{aligned}
\tag{3.12}
$$

Indication of proof.

$$
K_t^u(i,q) = \begin{cases} 0, & i = 1 \text{ or } q = 1, \\ \sqrt{dk_t(i-1,i)} & \text{if } 2 \le i \le q, \\ 0 & \text{if } q < i, \end{cases}
$$

$$
(K_t^u)^{-1}(i,q) = \begin{cases} 0, & i = 1 \text{ or } q = 1, \\ \dfrac{-1}{\sqrt{dk_t(q-1,q)}} & \text{if } 2 \le i = q-1, \\ \dfrac{1}{\sqrt{dk_t(q-1,q)}} & \text{if } 2 \le i = q, \\ 0 & \text{otherwise.} \end{cases}
$$

(3.13)

Hence

$$
\begin{aligned}
K_t^u&(q, u(I-1)+1)(K_s^u)^{-1}(\cdot, p) \\
&= K_t^u(q, u(p-1)+1)(K_s^u)^{-1}(p-1, p) + K_t^u(q, u(p-1)+1)(K_s^u)^{-1}(p, p) \\
&= \begin{cases} 0, & p = 1, \\ 0 & \text{if } q \le u(p-1)+1, \\ \dfrac{\sqrt{dk_t(q-1,q)}}{\sqrt{dk_s(p-1,p)}} & \text{if } p \ge 2, \ u(p-1)+1 < q \le u(p)+1, \\ 0 & \text{if } p \ge 2, \ u(p)+1 < q. \end{cases}
\end{aligned}
$$

(3.14)

Furthermore,

$$
\begin{aligned}
((K_s^u)^T & R_s^u(\cdot, p))^T (K_s)^{-1} (K_t^u(\cdot, u(I-1)+1))^T R_t^u(\cdot, u(q-1)+1) \\
&= (R_s^u(\cdot, p))^T (K_s^u)^{-T} (K_t^u(\cdot, u(I-1)+1))^T R_t^u(\cdot, u(q-1)+1) \\
&= \sum_{i=2}^{p} R_s^u(i,p) [(K_s^u)^{-T} (K_t^u(\cdot, u(I-1)+1))^T R_t^u(\cdot, u(q-1)+1)](i) \\
&= \sum_{i=2}^{p} \frac{R_s^u(i,p)}{\sqrt{dk_s(i-1,i)}} \sum_{j=u(i-2)+1}^{u(i-1)} \sqrt{dk_t(j,j+1)} R_t^u(j+1, u(q-1)+1).
\end{aligned}
$$

(3.15)

$\square$

COROLLARY 3.4. *For each positive number T, $0 \leq x \leq T$, and partition s of $[0,T]$ refining $\{0,x,T\}$, if t refines s ($s = t[u]$), then*

$$
\begin{aligned}
& \left| R(x,x) - Q_K\big((A_t)^*K(\cdot,x),(A_s)^*K(\cdot,x)\big) \right| \\
&= \left| \sum_{i=2}^{s^{-1}(x)+1} \frac{R_s^u(i,s^{-1}(x)+1)}{\sqrt{dk_s(i-1,i)}} \left\{ \sqrt{dk_s(i-1,i)}\, R_s^u(i,s^{-1}(x)+1) \right. \right. \\
&\qquad\qquad\qquad\qquad \left.\left. - \sum_{j=u(i-2)+1}^{u(i-1)} \sqrt{dk_t(j,j+1)}\, R_t^u(j+1,t^{-1}(x)+1) \right\} \right|.
\end{aligned}
$$
(3.16)

THEOREM 3.5. *Assume that for each positive number T and $0 \leq x \leq T$ there is a positive number M such that if $\{s_p\}_0^n$ is a partition of $[0,T]$ refining $\{0,x,T\}$, then*

$$
0 \leq R_s^u(i,s^{-1}(x)+1) \leq M\sqrt{dk_s(i-1,i)}
$$
(3.17)

for $i = 2,3,\ldots,s^{-1}(x)+1$. Furthermore, assume that if t refines s ($s = t[u]$), then

$$
\sum_{j=u(i-2)+1}^{u(i-1)} \sqrt{dk_t(j,j+1)}\, R_t^u(j+1,t^{-1}(x)+1) \leq \sqrt{dk_s(i-1,i)}\, R_s^u(i,s^{-1}(x)+1) \quad (3.18)
$$

for $i = 2,3,\ldots,s^{-1}(x)+1$.

Then for each positive number T and $0 \leq x \leq T$, the net (3.2) is Cauchy.

INDICATION OF PROOF. The net

$$
\left\{ \sum_{i=2}^{s^{-1}(x)+1} \sqrt{dk_s(i-1,i)}\, R_s^u(i,s^{-1}(x)+1), \ t \text{ a partition of } [0,T] \text{ refining } \{0,x,T\} \right\}
$$
(3.19)

is nonincreasing but bounded below. Let $L(x)$ denote the limit. Then

$$
\begin{aligned}
& \left| R(x,x) - Q_K\big((A_t)^*K(\cdot,x),(A_s)^*K(\cdot,x)\big) \right| \\
&= \left| \sum_{i=2}^{s^{-1}(x)+1} \frac{R_s^u(i,s^{-1}(x)+1)}{\sqrt{dk_s(i-1,i)}} \left\{ \sqrt{dk_s(i-1,i)}\, R_s^u(i,s^{-1}(x)+1) \right. \right. \\
&\qquad\qquad\qquad\qquad \left.\left. - \sum_{j=u(i-2)+1}^{u(i-1)} \sqrt{dk_t(j,j+1)}\, R_t^u(j+1,t^{-1}(x)+1) \right\} \right| \\
&\leq M\left(\sum_{i=2}^{s^{-1}(x)+1} \sqrt{dk_s(i-1,i)}\, R_s^u(i,s^{-1}(x)+1) \right. \\
&\qquad\qquad \left. - \sum_{i=2}^{t^{-1}(x)+1} \sqrt{dk_t(i-1,i)}\, R_t^u(i,t^{-1}(x)+1) \right) \\
&\leq M\left(\sum_{i=2}^{s^{-1}(x)+1} \sqrt{dk_s(i-1,i)}\, R_s^u(i,s^{-1}(x)+1) - L(x) \right).
\end{aligned}
$$
(3.20)

Hence for each pair of positive numbers c and T and $0 \leq x \leq T$, there is a partition r of $[0,T]$ refining $\{0,x,T\}$ such that if s refines r and t refines s, then

$$\left| R(x,x) - Q_K\big((A_t)^* K(\cdot,x),(A_s)^* K(\cdot,x)\big) \right| < c. \tag{3.21}$$

The conclusion follows from Theorem 3.2. □

COROLLARY 3.6. *If m is an increasing function on $[0,\infty)$ which is absolutely continuous with respect to k and $R(x,y) = m(\min(x,y))$ for $0 \leq x,y$, then for each positive number T and $0 \leq x \leq T$ the net (3.2) is Cauchy.*

THEOREM 3.7. *If $R(a,s)R(b,t) = R(b,s)R(a,t)$ for $0 \leq a,b \leq s,t$ and for each positive number T and $0 \leq x \leq T$ there is a positive number M such that if $\{s_p\}_0^n$ is a partition of $[0,T]$ refining $\{0,x,T\}$, then*

$$0 \leq R_s^u\big(i,s^{-1}(x)+1\big) \leq M\sqrt{dk_s(i-1,i)} \tag{3.22}$$

for $i = 2,3,\dots,n+1$, then for each positive number T and $0 \leq x \leq T$ the net (3.2) is Cauchy.

LEMMA 3.8. *For each integer $q \geq 2$,*
(1)

$$R_t^u(q,q)^2 = R_t(q,q) - \frac{R_t(q-1,q)^2}{R_t(q-1,q-1)}, \tag{3.23}$$

(2)

$$R_t^u(q,k) = \frac{R_t(q,k)}{R_t(q,q)} R_t^u(q,q), \tag{3.24}$$

for $k > q$.

PROOF OF LEMMA 3.8. We will proceed by induction. Assume for convenience that $R_t(1,1) = 1$. Note that $R_t^u(1,1) = 1$, $R_t^u(1,q) = R_t(1,q)$, and $R_t^u(2,2)^2 = R_t(2,2) - R_t^u(1,2)^2 = R_t(2,2) - R_t(1,2)^2$. Further,

$$
\begin{aligned}
R_t^u(2,k) &= \frac{R_t(2,k) - R_t^u(1,2)R_t(1,k)}{R_t^u(2,2)} \\
&= \frac{R_t(2,k)}{R_t(2,2)}\big(R_t(2,2) - R_t(1,2)^2\big)^{1/2} \\
&= \frac{R_t(2,k)}{R_t(2,2)} R_t^u(2,2).
\end{aligned}
\tag{3.25}
$$

Assume (1) and (2) for $q \geq 2$ and $k > q$.

$$R_t^u(q+1,q+1)^2 = R_t(q+1,q+1) - R_t^u(1,q+1)^2 - \cdots - R_t^u(q+1,q+1)^2 \qquad (3.26)$$

$$= R_t(q+1,q+1) - R_t^u(1,q+1)^2$$
$$- \frac{R_t(2,q+1)^2}{R_t(2,2)^2}(R_t(2,2) - R_t(1,2)^2) \qquad (3.27)$$
$$- \cdots - \frac{R_t(q,q+1)^2}{R_t(q,q)^2}\left(R_t(q,q) - \frac{R_t(q-1,q)^2}{R_t(q-1,q-1)}\right)$$

$$= R_t(q+1,q+1) - R_t(1,q+1)^2 + R_t(1,q+1)^2 - \frac{R_t(2,q+1)^2}{R_t(2,2)}$$
$$+ \frac{R_t(2,q+1)^2}{R_t(2,2)} - \frac{R_t(3,q+1)^2}{R_t(2,2)} \qquad (3.28)$$
$$+ \cdots + \frac{R_t(q-1,q+1)^2}{R_t(q-1,q-1)} - \frac{R_t(q,q+1)^2}{R_t(q,q)}$$

$$= R_t(q+1,q+1) - \frac{R_t(q,q+1)^2}{R_t(q,q)}. \qquad (3.29)$$

Also, for $k > q+1$,

$$R_t^u(q+1,k)$$
$$= (R_t(q+1,k) - R_t^u(1,q+1)R_t^u(1,k) \qquad (3.30)$$
$$- \cdots - R_t^u(q,q+1)R_t^u(q,k))(R_t^u(q+1,q+1))^{-1}$$

$$= \left(R_t(q+1,k) - R_t^u(1,q+1)R_t^u(1,k)\right.$$
$$- \frac{R_t(2,q+1)}{R_t(2,2)}R_t^u(2,2)^2\frac{R_t(2,k)}{R_t(2,2)} \qquad (3.31)$$
$$\left.- \cdots - \frac{R_t(q,q+1)}{R_t(q,q)}R_t^u(q,q)^2\frac{R_t(q,k)}{R_t(q,q)}\right)(R_t^u(q+1,q+1))^{-1}$$

$$= \left(R_t(q+1,k) - \frac{R_t(1,q+1)^2R_t(q+1,k)}{R_t(q+1,q+1)}\right.$$
$$- \frac{R_t(2,q+1)^2R_t(q+1,k)R_t^u(2,2)^2}{R_t(2,2)^2R_t(q+1,q+1)} \qquad (3.32)$$
$$\left.- \cdots - \frac{R_t(q,q+1)^2R_t(q+1,k)R_t^u(q,q)^2}{R_t(q,q)^2R_t(q+1,q+1)}\right)(R_t^u(q+1,q+1))^{-1}$$

$$= \frac{R_t(q+1,k)}{R_t(q+1,q+1)}\left(R_t(q+1,q+1) - R_t(0,q+1)^2\right.$$
$$- \frac{R_t(2,q+1)^2R_t^u(2,2)^2}{R_t(2,2)} \qquad (3.33)$$
$$\left.- \cdots - \frac{R_t(q,q+1)^2R_t^u(q,q)^2}{R_t(q,q)^2}\right)(R_t^u(q+1,q+1))^{-1}$$

$$
\begin{aligned}
&= \frac{R_t(q+1,k)}{R_t(q+1,q+1)} \big(R_t(q+1,q+1) - R_t(1,q+1)^2 - R_t^u(2,q+1)^2 \\
&\qquad\qquad - \cdots - R_t^u(q+1,q+1)^2 \big) \big(R_t^u(q+1,q+1) \big)^{-1}
\end{aligned}
\tag{3.34}
$$

$$
= \frac{R_t(q+1,k)}{R_t(q+1,q+1)} R_t^u(q+1,q+1).
\tag{3.35}
$$

Thus (1) and (2) hold for all $q \geq 2$ and $k > q$. $\qquad\square$

LEMMA 3.9. *If $T > 0$, $\{s_p\}_0^n$ is a partition of $[0,T]$ with refinement $\{t_q\}$, $1 \leq p \leq n-2$, $t_i = s_i$ for $0 \leq i \leq p$, $s_p < t_{p+1} < s_{p+1}$, and $t_{i+1} = s_i$ for $p+1 \leq i \leq n$, then*
  (1) $R_t^u(1:p+1, 1:p+1) = R_s^u(1:p+1, 1:p+1)$,
  (2) $R_t^u(1:p+1, p+3:n+2) = R_s^u(1:p+1, p+2:n+1)$,
  (3) $(R_t^u(p+2,q))^2 + (R_t^u(p+3,q))^2 = (R_s^u(p+2,q-1))^2$, $p+3 \leq q \leq n+1$,
  (4) $R_t^u(p+3:n+2, p+3:n+2) = R_s^u(p+2:n+1, p+2:n+1)$.
Alternately, if $t_i = s_i$ for $0 \leq i \leq n-1$, $s_{n-1} < t_n < s_n$, and $t_{n+1} = s_n$, then
  (5) $R_t^u(1:n, 1:n) = R_s^u(1:n, 1:n)$,
  (6) $R_t^u(1:n, n+2) = R_s^u(1:n, n+1)$,
  (7) $(R_t^u(n+1,n+2))^2 + (R_t^u(n+2,n+2))^2 = (R_s^u(n+1,n+1))^2$.

PROOF OF LEMMA 3.9. The first two parts are immediate. Note that

$$
\sum_{k=1}^{p+3} R_t^u(k,p+3)^2 = R_t(p+3,p+3) = R_s(p+2,p+2) = \sum_{k=1}^{p+2} R_s^u(k,p+2)^2.
\tag{3.36}
$$

Hence

$$
R_t^u(p+2,p+3)^2 + R_t^u(p+3,p+3)^2 = R_s^u(p+2,p+2)^2.
\tag{3.37}
$$

Suppose $q > p+3$. Then

$$
\begin{aligned}
R_t^u&(p+2,q)^2 + R_t^u(p+3,q)^2 \\
&= \frac{R_t(p+2,q)^2}{R_t(p+2,p+2)^2} R_t^u(p+2,p+2)^2 \\
&\quad + \frac{R_t(p+3,q)^2}{R_t(p+3,p+3)^2} R_t^u(p+3,p+3)^2 \\
&= \frac{R_t(p+3,q)^2}{R_t(p+3,p+3)^2} \left[\frac{R_t(p+2,p+3)^2}{R_t(p+2,p+2)^2} R_t^u(p+2,p+2)^2 \right. \\
&\qquad\qquad \left. + R_t^u(p+3,p+3)^2 \right] \\
&= \frac{R_s(p+2,q-1)^2}{R_s(p+2,p+2)^2} \big[R_t^u(p+2,p+3)^2 + R_t^u(p+3,p+3)^2 \big] \\
&= \frac{R_s(p+2,q-1)^2}{R_s(p+2,p+2)^2} R_s^u(p+2,p+2)^2 \\
&= R_s^u(p+2,q-1)^2.
\end{aligned}
\tag{3.38}
$$

Parts (4), (5), (6), and (7) follow in a similar way. $\qquad\square$

LEMMA 3.10. *If $T > 0$, $0 \le x \le T$, $\{s_p\}_0^n$ is a partition of $[0,T]$ refining $\{0,x,T\}$, and t is a refinement of s ($s = t[u]$), then*

$$\sum_{j=u(i-1)+1}^{u(i)} \sqrt{dk_t(j,j+1)} R_t^u(j+1,t^{-1}(x)+1) \le \sqrt{dk_s(i,i+1)} R_s^u(i+1,s^{-1}(x)+1)$$

(3.39)

for $i = 1,2,\dots,s^{-1}(x)$.

PROOF OF THEOREM 3.7. The theorem follows from Lemma 3.10 and Theorem 3.5.
□

THEOREM 3.11. *If m is an increasing function on $[0,\infty)$ which is absolutely continuous with respect to k, $m(0) = 1$, and $R(x,y) = \exp(-c|x-y|)m(\min(x,y))$ for $0 \le x, y$, where c is a nonnegative number, then, for each positive number T and $0 \le x \le T$, the net (3.2) is Cauchy.*

PROOF OF THEOREM 3.11. If $0 \le a,b \le s,t$, then

$$\begin{aligned}
R(a,s)R(b,t) &= e^{-c(s-a)}m(a)e^{-c(t-b)}m(b) \\
&= e^{-c(s-b)}m(b)e^{-c(t-a)}m(a) \\
&= R(b,s)R(a,t).
\end{aligned}$$

(3.40)

□

We will show that R_t is positive definite (see (2.3)) by developing a Cholesky factorization of R_t. We accomplish this by first defining R_t^u inductively, a row at a time. Assuming $R_t^u(1,1) = R_t(1,1)^{1/2} = 1$, $R_t^u(1,k) = R_t(1,k)/R_t^u(1,1) = e^{t(k)}$ for $k > 0$, and for $q = 2,3,\dots$,

(1)

$$\begin{aligned}
R_t^u(q,q)^2 &= R_t(q,q) - \frac{R_t(q-1,q)^2}{R_t(q-1,q-1)} \\
&= m(t_q) - e^{2c(t(q)-t(q-1))}m(t_{q-1}) > 0,
\end{aligned}$$

(3.41)

(2)

$$R_t^u(q,k) = \frac{R_t(q,k)}{R_t(q,q)} R_t^u(q,q),$$

(3.42)

for $k > q$.

We conclude that $(R_t^u)^T R_t^u = R_t$ by reading (2.5), (2.11), (2.12), (2.21) and (2.24), (3.26), (3.27), (3.28),(3.29), (3.30) in reverse order. Thus R_t is positive definite and satisfies the hypothesis of Theorem 3.7 and we draw the conclusion that the net is Cauchy.

Theorem 3.11 provides a rich class of examples and we can begin to examine relations between members of the class. For instance, we can think of c as a damping coefficient and investigate the effects of c on the linearized system's behavior. Figure 3.1 shows a collection of examples with the graph of m in the upper left corner and, in a clockwise direction, the graph for R_t for $c = 4,2,0$. We look at the question of system behavior in the next section.
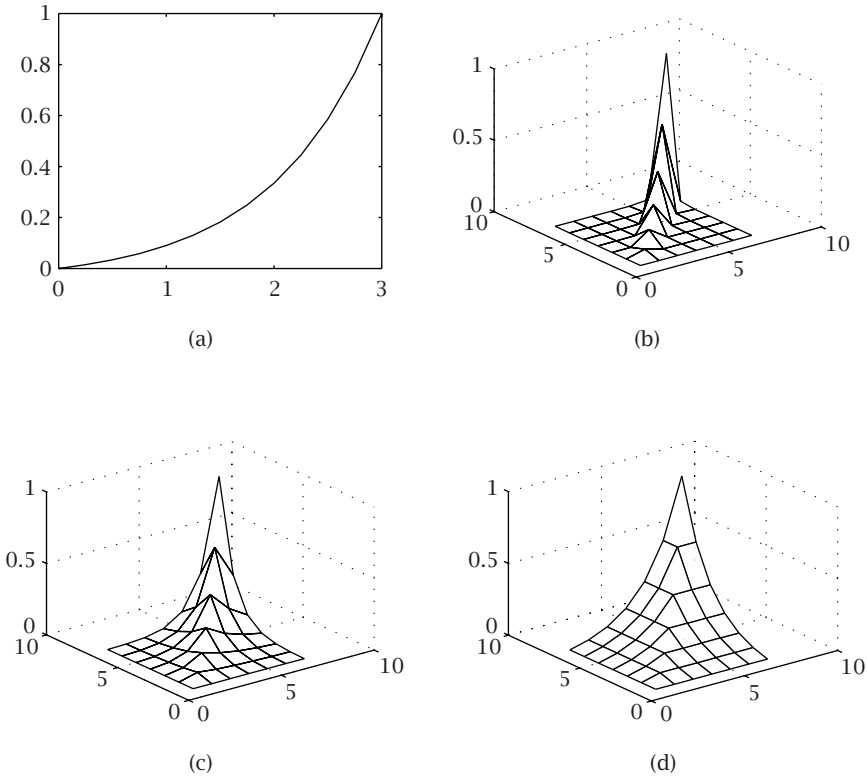
FIGURE 3.1. Covariance kernels, with $R_i(x,x)$, $0 \leq x \leq 3$, plotted in (a) and (b), (c), and (d) represent various versions of possible kernels. For clarity, the x and y axes are labeled $0, 1, \ldots, 7$ for the partition $x_j = y_j = 3j/7$.

4. Simulations. Starting with such a nice model for the covariance kernel in Theorem 3.7, one might ask if stochastic linearization contributes anything. That is, if the only information available for the discrete process Y_t is the mean $\mathbf{m}_t$ and the covariance kernel R_t, then everything we can know of the approximating normal process determined by the first two moments is captured by the discrete distribution function obtained as follows. The finite-dimensional density function is given by

$$f_t(\mathbf{x}) = (2\pi)^{-n/2} \det(R_t)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_t)^T R_t^{-1}(\mathbf{x} - \mathbf{m}_t)\right). \tag{4.1}$$

The finite-dimensional distribution function is then given by

$$\mathcal{P}(\mathbf{a} \leq Y_t \leq \mathbf{b}) = \int_{\mathbf{a}}^{\mathbf{b}} f_t(\mathbf{x}) d\mathbf{x} = \int_{a(0)}^{b(0)} \cdots \int_{a(n)}^{b(n)} f_t(x_0, \ldots, x_n) dx_0 \cdots dx_n. \tag{4.2}$$

However, a problem remains with the slow convergence of numerical evaluation of the iterated integral, especially when n is large, bigger than eight. Many methods rely on some kind of simulation to speed convergence [20].
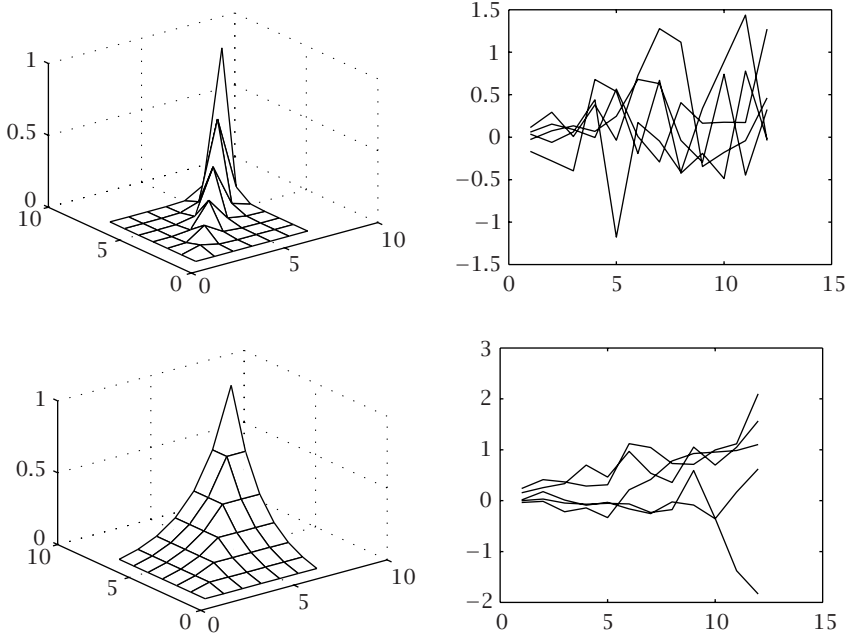
FIGURE 4.1. Simulations.

BASIC DISCUSSION. We propose another simulation methodology based on stochastic linearization. Let X_t be an n-dimensional normally distributed random row vector. Note that $EX_t = 0$ and $EX_t^T X_t = I_n$. If K_t^u is the upper Cholesky factor of K_t, then $W_t(p) = [X_t K_t^u](p)$.

To see this, note that $EX_t K_t^u = 0$ and

$$E(X_t K_t^u)^T X_t K_t^u = (K_t^u)^T E X_t^T X_t K_t^u = K_t. \tag{4.3}$$

For a given finite-dimensional stochastic linearization A_t with matrix representation $(K_t^u)^T R_t^u$, we have

$$[A_t W](t_p) = [W_t^T (K_t)^{-1} (K_t^u)^T R_t^u](p) = [W_t^T (K_t^u)^{-1} R_t^u](p) = [X_t R_t^u](p). \tag{4.4}$$

In Figure 4.1 we can compare the behavior of the system examples with $c = 8$ and 0. The probability of a sample path generated by the first system on the interval $[0,3]$ exceeding $a = 1.0$ is 0.3750. For the second system, the probability is 0.2262.

OPERATOR NORM. In the system monitoring problem, there is a need to measure the distance between two nonlinear systems. One possiblity is to measure the distance between the systems' discrete linearizations rather than between the systems directly. If the systems have covariance kernels R_1 and R_2, respectively, with upper Cholesky factorizations R_1^u and R_2^u, then an approximate operator norm for the difference of

TABLE 4.1. Table of distances.

	Ru1	Ru2	Ru3
Ru1	0.0	—	—
Ru2	0.7961	0.0	—
Ru3	1.3039	0.5471	0.0

the discrete linearizations could be computed with the following scrap of MatLab code:

```
% input: InvKt, Ru1, Ru2
x = randn(1000, n);
y = x*(Ru1 - Ru2);
z = diag(y*InvKt*y');
w = diag(x*x');
max((z./w).^.5)
```

The distances between the systems represented by the covariance kernels given in Figure 3.1 are given in Table 4.1.

In the absence of an absolute scale, we can only conclude that the third system is further from the first than it is from the second. This certainly fits our intuition.

5. Significance of work. We need a more robust condition on the finite-dimensional covariances $\{R_t\}$ implying convergence of the finite-dimensional operators $\{A_t\}$. Theorems 3.5 and 3.7 are too delicate for application to estimates of R_t. That is, the theorems assume we know R_t exactly or, in other words, we have an infinite amount of data at our disposal.

One can easily move from statistics of observations of inputs and outputs to confidence intervals and other measures of the accuracy of the estimates of R_t. We need to extend these possibilities to results on the quality of the estimates of A_t.

Much of this material can be extended immediately to vector processes. Examples of vector processes have been explored; for instance, a Lorenz system [6, 17, 26] with a one-dimensional noise input and a two-dimensional observation. The notion of convergence introduced for scalar inputs and outputs extends to the vector case. However, until the scalar case is settled, the condition which implies convergence for the vector case is hard to visualize.

Again, the dimension of the state space for the underlying system does not enter. The first example has an infinite-dimensional state space. So the method of linearization under investigation, if we can carry out our program, will apply to some systems governed by nonlinear partial differential equations as well.

6. Comments on the literature. Most work reported in the literature [4, 7, 8, 9, 10, 12, 13, 14, 15, 16, 22, 23, 27] assumes a model. As noted earlier [19] common practice when confronted by a system known only from input/output data is to fit a linear model.

The statistical linearization as presented in [25] is based on an assumption of the form of the underlying nonlinear system. Data enters the problem from simulations of

a known nonlinear system. Applications are made to marine structures such as drilling platforms.

Application of Hilbert space ideas to system problems requires an additional time structure which can be used to guarantee the operators are realizable, that is, causal. This requirement as discussed in [8] can be satisfied in several different settings [2, 10, 29]. The framework of Hellinger integrable functions, associated with the covariance function of the Wiener process, has a built-in time structure. The elements of $\mathscr{A} \cup \mathscr{B}$ are immediately causal.

The starting point for the work in this investigation differs from that of [8, 10, 27] in that the covariance R, known only partially as R_t from data, is the matrix representation of several positive definite Hermitian operators depending on choices made for the Hilbert space. Further, no assumption is made concerning the factorization of this operator. We are searching for conditions on R which yield the existence of a factorization.

References

[1]   N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.

[2]   W. Arveson, *Interpolation problems in nest algebras*, J. Funct. Anal. **20** (1975), no. 3, 208–233.

[3]   A. K. Bose, A. S. Cover, and J. A. Reneke, *On point dissipative n-dimensional systems of differential equations with quadratic nonlinearity*, Int. J. Math. Math. Sci. **16** (1993), no. 1, 139–148.

[4]   S. N. Chow and J. K. Hale, *Methods of Bifurcation Theory*, Grundlehren der Mathematischen Wissenschaften, vol. 251, Springer-Verlag, New York, 1982.

[5]   A. S. Cover, J. A. Reneke, M. Fryer, S. Lenhart, and V. Protopopescu, *RKH space methods for low level monitoring and control of nonlinear systems*, Math. Models Methods Appl. Sci. **6** (1996), no. 1, 77–96.

[6]   A. S. Cover, J. A. Reneke, S. Lenhart, and V. Protopopescu, *RKH space methods for low level monitoring and control of nonlinear systems. II. A vector-case example: The Lorenz system*, Math. Models Methods Appl. Sci. **7** (1997), no. 6, 823–845.

[7]   J. B. Cruz and P. V. Kokotović (eds.), *Feedback Systems*, McGraw-Hill, New York, 1972.

[8]   A. Feintuch and R. Saeks, *System Theory. A Hilbert Space Approach*, Pure and Applied Mathematics, vol. 102, Academic Press, New York, 1982.

[9]   P. A. Fuhrmann, *Linear Systems and Operators in Hilbert Space*, McGraw-Hill, New York, 1981.

[10] I. Gohberg and M. A. Kaashoek (eds.), *Constructive Methods of Wiener-Hopf Factorization*, Operator Theory: Advances and Applications, vol. 21, Birkhäuser Verlag, Basel, 1986.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Series in the Mathematical Sciences, vol. 3, Johns Hopkins University Press, Maryland, 1983.

[12] A. N. Gündeş and C. A. Desoer, *Algebraic Theory of Linear Feedback Systems with Full and Decentralized Compensators*, Lecture Notes in Control and Information Sciences, vol. 142, Springer-Verlag, Berlin, 1990.

[13] D. A. Herrero, *Approximation of Hibert Space Operators*, Wiley, New York, 1981.

[14] P. A. Ioannou and P. V. Kokotović, *Adaptive Systems with Reduced Models*, Lecture Notes in Control and Information Sciences, vol. 47, Springer-Verlag, Berlin, 1983.

[15] P. V. Kokotović (ed.), *Foundations of Adaptive Control*, Lecture Notes in Control and Information Sciences, vol. 160, Springer-Verlag, Berlin, 1991.

[16] P. V. Kokotović, A. Benoussan, and G. L. Blankenship (eds.), *Singular Perturbations and Asymptotic Analysis in Control Systems*, Springer-Verlag, Berlin, 1987.

[17] E. N. Lorenz, *Deterministic non-periodic flow*, J. Atmospheric Sci. **20** (1963), 130–141.

[18] J. S. Mac Nerney, *Hellinger integrals in inner product spaces*, J. Elisha Mitchell Sci. Soc. **76** (1960), 251–273.

[19] M. Morari, *Theory and practice of process control*, SIAM Conference on Control and Its Applications, Minnesota, 1992, invited talk.

[20] W. J. Morokoff, *Generating quasi-random paths for stochastic processes*, SIAM Rev. **40** (1998), no. 4, 765–788.

[21] E. Parzen, *An approach to time series analysis*, Ann. Math. Statist. **32** (1961), 951–989.

[22] W. R. Perkins and J. B. Cruz Jr., *Engineering of Dynamic Systems*, Wiley, New York, 1969.

[23] G. Pisier, *Factorization of Linear Operators and Geometry of Banach Spaces*, CBMS Regional Conference Series in Mathematics, vol. 60, American Mathematical Society, Rhode Island, 1986.

[24] J. A. Reneke, R. E. Fennell, and R. B. Minton, *Structured Hereditary Systems*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 107, Marcel Dekker, New York, 1987.

[25] J. B. Roberts and P.-T. D. Spanos, *Random Vibration and Statistical Linearization*, John Wiley & Sons, Chichester, 1990.

[26] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*, Applied Mathematical Sciences, vol. 41, Springer-Verlag, New York, 1982.

[27] M. Vidyasagar, *Control System Synthesis. A Factorization Approach*, MIT Press Series in Signal Processing, Optimization, and Control, vol. 7, MIT Press, Massachusetts, 1985.

[28] H. L. Weinert (ed.), *Reproducing Kernel Hilbert Spaces; Applications in Statistical Signal Processing*, Hutchinson Ross Publishing, Pennsylvania, 1982.

[29] J. C. Willems, *Stability, instability, invertibility and causality*, SIAM J. Control **7** (1969), 645–671.

James A. Reneke: Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975, USA

E-mail address: reneke@ces.clemson.edu