NONLINEAR PRINCIPAL COMPONENTS AND LONG-RUN IMPLICATIONS OF MULTIVARIATE DIFFUSIONS[1]

BY XIAOHONG CHEN[2], LARS PETER HANSEN[3] AND JOSÉ SCHEINKMAN[4]

Yale University, University of Chicago and Princeton University

We investigate a method for extracting nonlinear principal components (NPCs). These NPCs maximize variation subject to smoothness and orthogonality constraints; but we allow for a general class of constraints and multivariate probability densities, including densities without compact support and even densities with algebraic tails. We provide primitive sufficient conditions for the existence of these NPCs. By exploiting the theory of continuous-time, reversible Markov diffusion processes, we give a different interpretation of these NPCs and the smoothness constraints. When the diffusion matrix is used to enforce smoothness, the NPCs maximize long-run variation relative to the overall variation subject to orthogonality constraints. Moreover, the NPCs behave as scalar autoregressions with heteroskedastic innovations; this supports semiparametric identification and estimation of a multivariate reversible diffusion process and tests of the overidentifying restrictions implied by such a process from low-frequency data. We also explore implications for stationary, possibly nonreversible diffusion processes. Finally, we suggest a sieve method to estimate the NPCs from discretely-sampled data.

1. Introduction. Principal components are functions of the data that capture maximal variation in some sense. Often they are restricted to be linear functions of the underlying data as in original analyses of Pearson [27] and Hotelling [23]. In this paper we study the extraction of nonlinear principal components (NPCs) using information encoded in the probability density of the data. Formally, the NPCs maximize variation subject to orthogonality and smoothness constraints where smoothness constraints are enforced by a quadratic form f expressed in terms of the gradients of functions. Specifically, the quadratic form is

$$f(\phi, \psi) = \frac{1}{2} \int_{x \in \Omega} \nabla \phi(x)' \Sigma(x) \nabla \psi(x) q(x) \, dx,$$

where ∇ denotes the (weak) gradient operator, Ω is the state space, Σ is a state-dependent positive-definite matrix and q is the invariant density of a strictly stationary ergodic data, $\{x_i\}_{i=1}^{T}$.

Alternatively, NPCs are solutions to approximation problems. Suppose we wish to form the best finite-dimensional least squares approximation to an infinite-dimensional space of smooth functions where we use the form f to limit the class of functions to be approximated. In a sense that we make formal, a finite number of NPCs solves this problem. More stringent smoothness restrictions enforced by penalization limit the family of functions to be approximated while improving the overall quality of approximation. Thus our analysis of NPCs is in part an investigation of this approximation.

Previously Box and Tiao [7] proposed a canonical analysis of multivariate linear time series. This analysis produces linear principal components of the multivariate process that can be ordered from least to most predictable. Much later, in a seemingly unrelated paper, Salinelli [31] defined NPCs for multivariate absolutely continuous random variables and characterized these NPCs as eigenfunctions of a self-adjoint, differential operator. As we will show, these two methods are related. We share Salinelli [31] interest in NPCs, but our departure from his work is substantial. For Salinelli, the matrix Σ is the identity matrix, the state space Ω is compact, and the density q is bounded above and below for the bulk of his analysis. Our interest in probability densities q that do not have compact support, including densities with algebraic tails, leads us naturally to consider a more general class of smoothness penalties. By allowing for a more flexible specification for Σ and q, we entertain a larger class of smoothness constraints *vis a vis* Salinelli [31] with explicit links to the data generation. Establishing the existence of NPCs in our setup is no longer routine.

Salinelli [31] assumed that the data generation process is independent and identically distributed (i.i.d.). While our analysis is applicable to such an environment, we also explore the case in which data, $\{x_i\}_{i=1}^{T}$, is sampled in low frequency from a stationary Markov diffusion process. By considering such processes, we make a specific choice of the matrix Σ used to enforce smoothness. It is the local covariance or diffusion matrix. With this choice, the NPCs extracted with smoothness penalties are ordered by the ratio of their long-run variation to the overall variation as in Box and Tiao [7]. NPCs that capture variation subject to smoothness constraints also display low-frequency variation due to their high persistence. In effect, we provide an extension of the method of Box and Tiao [7] to nonlinear, multivariate Markov diffusions, and establish an explicit link to the method of Salinelli [31].

In this paper we do the following:

1. Formulate the NPCs extraction to include state dependence in the smoothness constraint and state spaces that have infinite Lebesgue measure.
2. Give sufficient conditions for the existence of these NPCs.

3. Provide a reversible Markov diffusion process for the data generation that supports the NPCs extraction method and generates testable implications.
4. Explore implications for a more general class of Markov diffusion processes.

The rest of the paper is organized as follows. In Section 2, we first define NPCs as functions that maximize variation subject to orthogonality conditions and smoothness bounds given by the quadratic form f. Section 3 presents existence results. In Section 4, we suppose the data are sampled from a multivariate nonlinear diffusion and establish the connection between our NPCs and the canonical analysis of Box and Tiao [7]. The results in Section 5 relate the NPCs to eigenfunctions of conditional expectations operators associated with a stationary Markov process $\{x_t\}$ defined using the diffusion matrix Σ and the stationary density q. Given an eigenfunction ψ, the process $\{\psi(x_t)\}$ behaves as a scalar autoregression. Thus the eigenfunctions we obtain satisfy testable implications when the data is generated by a Markov process. The Markov process constructed in Section 5 is time reversible. In Section 6, we characterize other Markov processes associated with the same q and Σ. Section 7 provides a sieve method to estimate these NPCs using discrete-time low-frequency observations $\{x_i\}_{i=1}^T$. Section 8 gives some concluding remarks and discusses applications of our results. The Appendix contain computations associated with an example and some proofs that are not stated in the main text.

2. Nonlinear principal components. To define a functional notion of principal components we require two quadratic forms. We start with an open connected $\Omega \subseteq \mathbb{R}^n$. Let q be a probability density on Ω with respect to Lebesgue measure. The implied probability distribution Q is the population counterpart to the empirical distribution of the data. The data could be i.i.d. as in Salinelli [31], but we are primarily interested in the case in which the data, $\{x_i\}_{i=1}^T$, are sampled in low frequency from a continuous-time, stationary Markov diffusion, $\{x_t : t \geq 0\}$. In this case q is the stationary density of x_t.

Let L^2 denote the space of Borel measurable square integrable functions with respect to the population probability distribution Q. The L^2 inner product (denoted $\langle \cdot, \cdot \rangle$) is one of the two forms of interest. We use the corresponding norm to define an approximation criterion.

The second form is used to measure smoothness. Consider a (quadratic) form f_o defined on C_K^2, the space of twice continuously differentiable functions with compact support in Ω, that can be parameterized in terms of the density q and a positive definite matrix Σ that can depend on the state:

(2.1)
$$f_o(\phi, \psi) = \frac{1}{2} \int_\Omega \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_j} \frac{\partial \psi}{\partial y_i} q,$$

where

$$\Sigma = [\sigma_{ij}].$$

ASSUMPTION 2.1. q is a positive, continuously differentiable probability density on Ω.

ASSUMPTION 2.2. Σ is a continuously differentiable, positive definite matrix function on Ω.

Assumptions 2.1 and 2.2 restrict the density q and the matrix Σ to be continuously differentiable. These assumptions are made for convenience. As argued by Davies [14] (see Theorem 1.2.5) these restrictions can be replaced by a less stringent requirement that entries of the matrix $q\Sigma$ are locally [in L^2(Lebesgue)], weakly differentiable.

While the f_o is constructed in terms of the product $q\Sigma$, the density q will play a distinct role when we consider extending the domain of the form to a larger set of functions.

To study the case in which Ω is not compact, we will consider a particular closed extension of the form f_o. We extend the form f_o to a larger domain $\bar{H} \subset L^2$ using the notion of a weak derivative.

$$\bar{H} \doteq \left\{ \phi \in L^2\!: \text{ there exists } g \text{ measurable, with } \int g'\Sigma g q < \infty, \right.$$

$$\left. \text{and } \int \phi \nabla \psi = - \int g \psi, \text{ for all } \psi \in C_K^1 \right\}.$$

The random vector g is unique (for each ϕ) and is referred to as the *weak derivative* of ϕ. From now on, for each ϕ in $\bar{H}$ we write $\nabla \phi = g$.

Notice that $\bar{H}$ is constructed exactly as a weighted Sobolev space except that instead of requiring that $g \in L^2$, we require that $\Lambda g \in L^2$ where Λ is the square root of Σ. Also we use C_K^1 test functions. One can show, using mollifiers, that allowing for this larger set of test functions is equivalent to using the more usual set of test functions, C_K^∞ (see Brezis [8], Remark 1, page 150). For any pair of functions ψ and ϕ in $\bar{H}$, we define,

$$f(\phi, \psi) = \frac{1}{2} \int_\Omega (\nabla \phi)' \Sigma (\nabla \psi) q,$$

which is an extension of f_o. In $\bar{H}$ we use the inner product $\langle \phi, \psi \rangle_{\bar{f}} = \langle \phi, \psi \rangle + f(\phi, \psi)$. With this inner product, $\bar{H}$ is complete and hence a Hilbert space (see Proposition A.1 in the Appendix). Thus $\bar{H}$ is taken to be the domain $\mathcal{D}(f)$ of the form f. Notice, in particular, that the unit function is in $\mathcal{D}(f) = \bar{H}$.

2.1. *Initial construction.*   NPCs maximize variation subject to smoothness constraints. In our generalization these NPCs are defined as follows.

DEFINITION 2.1.   The function ψ_j is the jth nonlinear principal component (NPC) for $j \geq 1$ if ψ_j solves,

$$\max_{\phi \in \bar{H}} \langle \phi, \phi \rangle,$$

subject to

$$f(\phi, \phi) = 1,$$
$$\langle \psi_s, \phi \rangle = 0, \qquad s = 0, \ldots, j-1,$$

where ψ_0 is initialized to be the constant function one.

There are two differences between our proposed extraction and that of Salinelli [31]. First, Salinelli [31] assumes that Σ is the identity matrix. To accommodate a richer class of densities, we allow Σ to be state dependent. Second, Salinelli [31] assumes that the data density q has finite Lebesgue measure and is bounded away from zero. We allow the Lebesgue measure of the state space to be infinite, and, accordingly, our density q is no longer assumed to be bounded from below.

NPCs are eigenfunctions of the quadratic forms f.

DEFINITION 2.2.   An eigenfunction ψ of the quadratic form f satisfies,

$$(2.2) \qquad f(\phi, \psi) = \delta \langle \phi, \psi \rangle$$

for all $\phi \in \mathcal{D}(f)$. The scalar δ is the corresponding eigenvalue.

Since f is positive semidefinite, δ must be nonnegative. The NPCs extracted in the manner given in (2.1) have eigenvalues δ_j that increase with j. If we renormalize the eigenfunctions to have a unit second moment, the NPCs will be ordered by their smoothness as measured by $\delta_j = f(\psi_j, \psi_j)$. Moreover, $f(\psi_j, \psi_k) = 0$ for $j \neq k$.

Suppose that the NPCs $\{\psi_j : j = 0, 1, \ldots\}$ exist with corresponding eigenvalues $\{\delta_j : j = 0, 1, \ldots\}$. Consider any ϕ in L^2. Then

$$\phi = \sum_{j=0}^{\infty} \frac{\langle \psi_j, \phi \rangle}{\langle \psi_j, \psi_j \rangle} \psi_j,$$

$$\langle \phi, \phi \rangle = \sum_{j=0}^{\infty} \frac{\langle \psi_j, \phi \rangle^2}{\langle \psi_j, \psi_j \rangle},$$

and for any $\phi, \psi \in \mathcal{D}(f)$,

$$(2.3) \qquad f(\phi, \psi) = \sum_{j=0}^{\infty} \delta_j \frac{\langle \phi, \psi_j \rangle \langle \psi, \psi_j \rangle}{\langle \psi_j, \psi_j \rangle}.$$

2.2. *Benchmark optimization problem.* Let H be a closed linear subspace of L^2, and consider the optimization problem.

PROBLEM 2.1.

$$\max_{\phi \in H} \langle \phi, \phi \rangle,$$

subject to

$$\theta \langle \phi, \phi \rangle + f(\phi, \phi) \le 1$$

for some $\theta > 0$.

A necessary condition for ψ to be a NPC is that it satisfies an eigenvalue problem:

CLAIM 2.1. *A solution ψ to Problem* 2.1 *will also solve the eigenvalue problem,*

$$\langle \phi, \psi \rangle = \lambda [\theta \langle \phi, \psi \rangle + f(\phi, \psi)]$$

for some positive λ and all $\phi \in H$.

To establish the existence of a solution to Problem 2.1, it suffices to suppose the following:

CONDITION 2.1 (Existence). $\{\phi \in \mathcal{D}(f) : f(\phi, \phi) + \theta \langle \phi, \phi \rangle \le 1\}$ is precompact (has compact closure) in L^2.

The precompactness restriction guarantees that we may extract an L^2 convergent sequence in the constraint set, with objectives that approximate the supremum. The limit point of convergent sequence used to approximate the supremum, however, will necessarily be in the constraint set because the constraint set is convex and the form is closed.

2.3. *Approximation.* Why do we care about NPCs? One way to address this is to explore the construction of the best, finite dimensional, least squares approximations. Specifically, suppose we wish to construct the best finite-dimensional set of approximating functions for the space of functions that are square integrable with respect to a probability measure Q with density q. We now motivate NPCs as the recursive solution to such a problem. The N-dimensional problem is solved by solving N one-dimensional problems using a sequence of H's that remove one dimension in each step. The outcome at each step is a NPC used as an additional approximating function.

Initially solve Problem 2.1 for $H = L^2$, select a solution ψ_0 and denote the maximized objective as λ_0. Inductively, given $\psi_0, \psi_1, \ldots, \psi_{j-1}$, form H_{j-1} as

the j-dimensional space generated by these j solutions constructed recursively. Let $H_{j-1}^{\perp}$ denote the space of all elements of L^2 that are orthogonal to these j solutions and hence orthogonal to H_{j-1}. Solve Problem 2.1 for $H = H_{j-1}^{\perp}$, select a solution ψ_j, and form λ_j as the maximized value. The sequence $\{\lambda_j : j = 0, 1, \ldots\}$ is decreasing because we are omitting components of the constraint set for the maximization problem as j increases.

In what sense is such a recursive procedure optimal? In answering this question, let $\text{Proj}(\phi|\hat{H})$ denote the least squares projection of ϕ onto the closed (in L^2) linear space $\hat{H}$. The second moment of the approximation error is,

$$\langle \phi - \text{Proj}(\phi|\hat{H}), \phi - \text{Proj}(\phi|\hat{H}) \rangle = \langle \phi, \phi \rangle - [\text{Proj}(\phi|\hat{H})]^2.$$

CLAIM 2.2. *Let $\hat{H}$ denote any N-dimensional subspace of L^2. Then*

$$\max_{\{\phi : \theta \langle \phi, \phi \rangle + f(\phi, \phi) \leq 1\}} \{\langle \phi, \phi \rangle - [\text{Proj}(\phi|\hat{H})]^2\} \geq \lambda_N.$$

Our next result shows that the bound deduced in Claim 2.2 is attained by H_{N-1}.

CLAIM 2.3.

$$\max_{\{\phi : \theta \langle \phi, \phi \rangle + f(\phi, \phi) \leq 1\}} \{\langle \phi, \phi \rangle - [\text{Proj}(\phi|H_{N-1})]^2\} = \lambda_N.$$

Taken together, these two claims justify H_{N-1} as a good N-dimensional space of approximating functions.

REMARK 2.1. There exist N-dimensional spaces other than H_{N-1} that attain the bound given in Claim 2.2. One reason is that there may be multiple solutions to Problem 2.1. Even when the solution to Problem 2.1 is unique, at each stage of the construction, ψ_{N-1} may be replaced by the sum of ψ_{N-1} plus some ψ's that is orthogonal to all of the solutions to Problem 2.1 with $H = H_{N-1}^{\perp}$. Such a choice cannot necessarily be used in a recursive construction of optimal approximating spaces with dimension greater than N.

2.4. *Nonlinear principal components revisited.* In Problem 2.1, the constraint set gets larger as θ declines to zero. Reducing the smoothness penalty with a smaller θ enlarges the collection of functions that satisfy the constraint. Thus the maximized objective increases as θ is reduced. While this is true, it turns out the maximizing choice of ϕ does not depend on θ up to scale. This follows because the ranking over ϕ's implied by the ratio,

$$\frac{\langle \phi, \phi \rangle}{\theta \langle \phi, \phi \rangle + f(\phi, \phi)},$$

does not depend on the value of θ. The same ranking is also implied by the ratio,

$$\frac{\langle \phi, \phi \rangle}{f(\phi, \phi)},$$

provided that H is orthogonal to all constant functions. Thus a scaled solution ψ to Problem 2.1 also solves.

PROBLEM 2.2.

$$\max_{\phi \in H} \langle \phi, \phi \rangle,$$

subject to

$$f(\phi, \phi) = 1.$$

Restricting H to be orthogonal to constant functions is equivalent to limiting attention to functions ϕ that have mean zero under the population data distribution Q. Recall that our construction of NPCs is based on the recursive application of this problem.

From Claim 2.1 we know that ψ satisfies

$$\langle \phi, \psi \rangle = \lambda[\theta \langle \phi, \psi \rangle + f(\phi, \psi)]$$

for all $\phi \in H$. Rearranging terms,

$$f(\phi, \psi) = \delta \langle \phi, \psi \rangle,$$

where

$$\delta = \frac{1 - \theta \lambda}{\lambda}.$$

This is the eigenvalue associated with the NPC extraction. Solving for λ,

$$\lambda = \frac{1}{\theta + \delta}.$$

Since eigenvalues δ of the form increase without bound, the corresponding sequence of λ's converge to zero guaranteeing that approximation becomes arbitrarily accurate as the number of NPCs increases.

3. Existence. In this section, we consider more primitive sufficient conditions that imply Condition 2.1, which as we noted in Section 2, guarantees the existence of NPCs. We allow for noncompact state spaces and provide alternative restrictions on the tail behavior of the the density q and the penalization matrix Σ that guarantee that the compactness criterion (Condition 2.1) is satisfied. Roughly speaking, when the tails of the density q are exponentially thin, the compactness criterion can be established without requiring that the matrix Σ becomes large (in the sense of positive definite matrices) in the tails. On the other hand, when the tails of q are algebraic and hence thicker, divergence of Σ in the tails can play an important role in establishing Condition 2.1.

We start by reviewing some known existence conditions, which we extend using two devices. First, we transform the function space and hence the (quadratic) form so that distribution induced by q is replaced by the Lebesgue measure. This transformation allows us to apply known results for forms built using the Lebesgue measure. Second, we study forms that are simpler but dominated by f. When the dominated forms satisfy Condition 2.1, the same can be said of f.

3.1. *Compact domain.*   Salinelli [31] established the existence of eigenfunctions by applying Rellich's compact embedding theorem when the domain Ω is compact with a continuous boundary. This approach requires a density q that is bounded and bounded away from zero and a derivative penalty matrix Σ that is uniformly nonsingular.

3.2. *Real line.*   Perhaps surprisingly, the NPC extraction is nontrivial even for densities on the real line. This is because our NPCs can be nonlinear functions of the underlying Markov state. We initially consider the case in which the state space is the real line.

PROPOSITION 3.1.   *Suppose $\Sigma = \varsigma^2$, and*

$$\int_0^\infty \frac{1}{\varsigma^2(x)q(x)} = +\infty,$$

(3.1)

$$\int_0^\infty \frac{1}{\varsigma^2(-x)q(-x)} = +\infty,$$

(3.2) $$\lim_{|x|\to\infty} -\frac{x}{|x|}\left[\varsigma(x)\frac{q'(x)}{q(x)} + \varsigma'(x)\right] = +\infty.$$

Then Condition 2.1 is satisfied.

When ς is constant, the compactness condition (3.2) reduces to

$$\lim_{|x|\to\infty} -\frac{x}{|x|}\left[\frac{q'(x)}{q(x)}\right] = +\infty,$$

which rules out densities with algebraic tails (tails that decay slower than $|x|$ raised to a negative power). By allowing for ς to increase, we can accommodate densities with algebraic tails. We now extend this analysis to higher dimensions.

3.3. $\mathbb{R}^n$.   In the subsections that follow, we will provide multivariate extensions for both sources of compactness: growth in the logarithmic derivative of the density q and growth in the derivative penalty Σ. For simplicity, we will concentrate in the case where the state space is all of $\mathbb{R}^n$.

3.3.1. *Cores.* The compactness Condition 2.1 involves the domain of the form f which is often rather complicated to describe. For this reason, we will focus on cases where this domain can be well approximated by smooth functions. The adequate notion of approximation is that of a core:

DEFINITION 3.1. A family of functions $\mathcal{C}o \subset \mathcal{D}(f)$ is a *core* of f if for any ϕ_0 in the domain $\mathcal{D}(f)$, there exists a sequence $\{\phi_j\}$ in $\mathcal{C}o$ such that

$$\lim_{j\to\infty} \langle \phi_j - \phi_0, \phi_j - \phi_0 \rangle + f(\phi_j - \phi_0, \phi_j - \phi_0) = 0.$$

CONDITION 3.1. C_K^2 is a core of f.

Let $\hat{f}$ denote the *minimal extension*, the smallest closed extension of the form f_o defined in (2.1). Condition 3.1 is equivalent to $f = \hat{f}$.

Although their purpose was different, Fukushima, Oshima and Takeda [18] provide a convenient sufficient condition that implies Condition 3.1 in environments that interest us. Define

$$\kappa(\mathsf{r}) = \int_{|x|=1} x' \Sigma(\mathsf{r}x) x q(\mathsf{r}x) \, dS(x),$$

where dS is the measure (surface element) used for integration on the sphere $|x| = 1$. For functions ψ and ϕ in C_K^2 that are radially symmetric, that is, $\phi(x) = \xi(|x|)$ and $\psi(x) = \zeta(|x|)$, we may depict the form f_o as an integral over radii,

$$f_o(\psi, \phi) = \int_0^\infty \frac{d\xi(\mathsf{r})}{d\mathsf{r}} \frac{d\zeta(\mathsf{r})}{d\mathsf{r}} \kappa(\mathsf{r}) \mathsf{r}^{n-1} \, d\mathsf{r}.$$

PROPOSITION 3.2. *Condition 3.1 is implied by*

$$(3.3) \qquad \int_1^\infty \kappa(\mathsf{r})^{-1} \mathsf{r}^{1-n} \, d\mathsf{r} = \infty.$$

Restriction (3.3) implies the scalar restriction (3.1) of Proposition 3.1. This follows since for any nonnegative reals r_1 and r_2,

$$\min\left\{\frac{1}{\mathsf{r}_1}, \frac{1}{\mathsf{r}_2}\right\} \geq \frac{1}{\mathsf{r}_1 + \mathsf{r}_2}.$$

Notice that (3.3) is a joint restriction on Σ and q. We may relate this condition to the moments of q and the growth of Σ using the inequality,

$$\infty = \left(\int_1^\infty \frac{1}{\mathsf{r}} \, d\mathsf{r}\right)^2 \leq \int_1^\infty \kappa(\mathsf{r})^{-1} \mathsf{r}^{1-n} \, d\mathsf{r} \int_1^\infty \kappa(\mathsf{r}) \mathsf{r}^{n-3} \, d\mathsf{r}.$$

Thus a sufficient condition for (3.3) is that

$$(3.4) \qquad \int_1^\infty \frac{\kappa(\mathsf{r})}{\mathsf{r}^2} \mathsf{r}^{n-1} \, d\mathsf{r} < \infty.$$

This latter inequality displays a tradeoff between growth in the penalization matrix and moments of the distribution. Define

$$\varsigma^2(\mathsf{r}) = \sup_{|x|=1} x'\Sigma(\mathsf{r}x)x$$

and

$$\varrho(\mathsf{r}) = \int_{|x|=1} q(\mathsf{r}x)\,dS(x).$$

Notice that

$$\kappa(\mathsf{r}) \le \varsigma^2(\mathsf{r})\varrho(\mathsf{r}).$$

Suppose, for instance, $\varsigma^2(\mathsf{r})$ is dominated by a quadratic function (in r). Then (3.4) and hence (3.3) are satisfied because the density q is integrable,

$$\int_0^\infty \varrho(\mathsf{r})\mathsf{r}^{n-1}\,d\mathsf{r} = 1.$$

We may extend the previous argument by supposing instead that

$$\varsigma^2(\mathsf{r}) \le c|\mathsf{r}|^{2+2\delta}$$

for some positive δ. Then

$$\frac{\kappa(\mathsf{r})}{\mathsf{r}^2} \le c\mathsf{r}^{2\delta} \int_{|x|=1} q(\mathsf{r}x)\,dS(x).$$

Thus (3.4) is satisfied, provided that

$$\int |x|^{2\delta} q(x)\,dx < \infty.$$

Hence we can allow for faster growth in ς^2 if q has high enough moments.

So far we have produced a sufficient condition for approximation using functions in C_K^2 (Condition 3.1). We provide sufficient conditions for the original compactness condition (Condition 2.1) by transforming the probability measure.

3.3.2. *Transforming the measure.* In this subsection, we map the original probability space L^2 into a Lebesgue counterpart $L^2(\text{leb})$. The transformation is standard (see Davies [14]), but it is often applied in the reverse direction. By using this transformation we may appeal to some existing mathematical results on compactness to establish Criterion 2.1,

$$\mathcal{U}_\theta = \{\phi \in \mathcal{D}(f) : f(\phi, \phi) + \theta\langle\phi, \phi\rangle \le 1\}$$

is precompact in L^2 for some $\theta > 0$.

Given q, write

$$q^{1/2} = \exp(-h).$$

ASSUMPTION 3.1. The function h is twice continuously differentiable.

This assumption imposes some extra smoothness on the density that was not required in our previous analysis.

Map the space L^2 into $L^2(\text{leb})$ by the (invertible) unitary transformation,

$$\psi = U\phi \equiv \exp(-h)\phi.$$

Since U is unitary, it suffices to show that $U(\mathcal{U}_\theta)$ is pre-compact. We will actually construct a set that contains $U(\mathcal{U}_\theta)$ and is pre-compact in $L^2(\text{leb})$.

First, notice that U and U^{-1} leave C_K^2 invariant, and for any $\psi \in C_K^2$, the corresponding $\phi = U^{-1}\psi$ satisfies

$$\nabla\phi = \exp(h)(\psi\nabla h + \nabla\psi).$$

Thus

$$f(U^{-1}\psi, U^{-1}\psi^*) = \frac{1}{2}\int (\nabla\psi)'\Sigma(\nabla\psi^*) + \frac{1}{2}\int (\nabla h)'\Sigma[\nabla(\psi\psi^*)]$$

$$+ \frac{1}{2}\int (\nabla h)'\Sigma(\nabla h)\psi\psi^*.$$

Applying integration-by-parts to $\psi \in C_K^2$, it follows that

$$\int (\nabla h)'\Sigma[\nabla(\psi\psi^*)] = -\int \sum_{i,j} \sigma_{i,j}\frac{\partial^2 h}{\partial y_i\,\partial y_j}\psi\psi^* - \int \sum_{i,j}\frac{\partial\sigma_{i,j}}{\partial y_i}\frac{\partial h}{\partial y_j}\psi\psi^*.$$

Therefore,

$$(3.5) \qquad f(U^{-1}\psi, U^{-1}\psi^*) = \frac{1}{2}\int (\nabla\psi)'\Sigma(\nabla\psi^*) + \frac{1}{2}\int V\psi\psi^*,$$

where the *potential* function V is given by

$$(3.6) \qquad V = -\sum_{i,j}\sigma_{i,j}\frac{\partial^2 h}{\partial y_i\,\partial y_j} - \sum_{i,j}\frac{\partial\sigma_{i,j}}{\partial y_i}\frac{\partial h}{\partial y_j} + (\nabla h)'\Sigma(\nabla h).$$

PROPOSITION 3.3. *Suppose that C_K^2 is a core for f, $\psi = U\phi$ for some $\phi \in \bar{H}$ and V is bounded from below. Then ψ is weakly differentiable,*

$$\nabla\psi = \exp(-h)(-\phi\nabla h + \nabla\phi)$$

and

$$(3.7) \qquad \frac{1}{2}\int (\nabla\phi)'\Sigma\nabla\phi q = \frac{1}{2}\int (\nabla\psi)'\Sigma(\nabla\psi) + \frac{1}{2}\int V\psi^2.$$

A consequence of this proposition is that

$$\mathcal{V}_\theta = \left\{ \psi \in L^2(\text{leb}) : \int \left(\theta + \frac{1}{2} V \right) \psi^2 + \frac{1}{2} \int (\nabla \psi)' \Sigma (\nabla \psi) \le 1 \right\} \supset U(\mathcal{U}_\theta)$$

and it thus suffices to show that $\mathcal{V}_\theta$ is precompact in $L^2(\text{leb})$ for some $\theta > 0$.

We consider two methods for establishing that this last property is satisfied. We first focus on the behavior of the potential V used in the quadratic form, $\int (\theta + \frac{1}{2} V) \psi^2$, and then we study extensions that exploit growth in the derivative penalty matrix Σ used in the quadratic form, $\int (\nabla \psi)' \Sigma (\nabla \psi)$.

3.3.3. *Divergent potential.* In this section, we use the tail behavior of the potential V. To simplify the treatment of the term $\int (\nabla \psi)' \Sigma (\nabla \psi)$ in the definition of $\mathcal{V}_\theta$, we impose:

ASSUMPTION 3.2. The derivative penalty matrix $\Sigma \ge \underline{c} I$ for some $\underline{c} > 0$.

This assumption rules out cases in which the derivative penalty matrix diminishes to zero for arbitrarily large states.

We also suppose that the potential function diverges at the boundary,

CRITERION 3.1. $\lim_{|x| \to \infty} V(x) = +\infty$.

PROPOSITION 3.4. *Under Assumptions* 3.1 *and* 3.2, *if Criterion* 3.1 *is satisfied, then Condition* 2.1 *is satisfied.*

Direct verification of Criterion 3.1 may be difficult because formula (3.6) is a bit complicated. However, we may replace the Σ by a lower bound. Given Assumption 3.2, we can always construct a twice continuously differentiable function $\varsigma(x)$ with

$$(3.8) \qquad \Sigma(x) \ge \varsigma(x)^2 I \ge \underline{c} I \qquad \text{for some } \underline{c} > 0.$$

We now show how growth conditions on $\varsigma(x)$ can help in delivering compactness.
Let

$$L \check{f}_o(\phi, \phi^*) = \frac{1}{2} \int \nabla \phi(x) \cdot \nabla \phi^*(x) \varsigma(x)^2 q(x)$$

on the space $\mathcal{C}_K^2$. Then

$$\check{f}_o(\phi, \phi) \le f_o(\phi, \phi).$$

Let $\check{f}$ be the minimal extension of $\check{f}_o$. If f is the minimal extension of f_o, when (3.8) holds, the domain of $\check{f}$ contains the domain of f. Applying Proposition 3.3 to $\check{f}$, it suffices to use

$$\check{V}(x) = \varsigma(x)^2 \left(-\text{trace} \left[\frac{\partial^2 h(x)}{\partial x_i \, \partial x_j} \right] - \frac{2 \nabla \varsigma(x) \cdot \nabla h(x)}{\varsigma(x)} + |\nabla h(x)|^2 \right)$$

in place of V in demonstrating compactness.

CRITERION 3.2. *Equation* (3.8) *is satisfied, and*

$$\lim_{|x| \to \infty} \check{V}(x) = +\infty.$$

To derive some sufficient conditions for this criterion, we parameterize ς as

$$\varsigma(x) = \exp[v(x)].$$

Then an alternative formula for $\check{V}$ is

$$\check{V}(x) = -\varsigma(x)^2 \operatorname{trace}\left[\frac{\partial^2 h(x)}{\partial x_i \, \partial x_j}\right] + \varsigma(x)^2 |\nabla h(x) - \nabla v(x)|^2$$

$$- \varsigma(x)^2 \nabla v(x) \cdot \nabla v(x).$$

An alternative to Criterion 3.2 is:

CRITERION 3.3. *Equation* (3.8) *is satisfied with* $\varsigma(x) = \exp[v(x)]$ *and*:

(a)
$$\lim_{|x| \to \infty} \frac{|\nabla v(x)|}{|\nabla h(x)|} = 0;$$

(b)
$$\lim_{|x| \to \infty} \varsigma(x)^2 \left(-\operatorname{trace}\left[\frac{\partial^2 h(x)}{\partial x_i \, \partial x_j}\right] + \nabla h(x) \cdot \nabla h(x)\right) = +\infty.$$

PROPOSITION 3.5. *Suppose Assumption* 3.1 *is satisfied. Then Criterion* 3.3 *implies Condition* 2.1.

Restriction (b) of Criterion 3.3 limits the second derivative contribution from offsetting that of the squared gradient of h. This criterion is convenient to check when h displays polynomial growth, or equivalently when q has exponentially thin tails. Even if $|\nabla h|$ becomes arbitrarily small for large $|x|$, the compactness criterion can still be satisfied by having the penalization ς increase to more than offset this decline.

Next we consider a way to exploit further growth in $\nabla \varsigma$. This approach gives us a way to enhance the potential function, and may be used when $\liminf_{|x| \to \infty} \frac{|\nabla v(x)|}{|\nabla h(x)|} > 0$. Write

$$\int \varsigma^2 \nabla \phi \cdot \nabla \phi = \underline{c} \int \nabla \phi \cdot \nabla \phi + \int (\varsigma^2 - \underline{c}) \nabla \phi \cdot \nabla \phi.$$

We now deduce a convenient lower bound on

$$\int (\varsigma^2 - \underline{c}) \nabla \phi \cdot \nabla \phi,$$

following an approach of Davies [14] (see Theorem 1.5.12). Construct an additional potential function,

$$\check{W}(x) = (\varsigma^2 + \underline{c})(\nabla v \cdot \nabla v) + (\varsigma^2 - \underline{c}) \operatorname{trace}\left(\frac{\partial^2 v}{\partial x_i \, \partial x_j}\right).$$

LEMMA 3.1. *If* (3.8) *holds*, *then*

$$\int \check{W}\phi^2 \leq \int (\varsigma^2 - \underline{c})\nabla\phi \cdot \nabla\phi \qquad \text{for all } \phi \in C_K^2.$$

Note that

$$\check{V}(x) + \check{W}(x) = \varsigma(x)^2 \operatorname{trace}\left[\frac{\partial^2 v(x)}{\partial x_i \, \partial x_j} - \frac{\partial^2 h(x)}{\partial x_i \, \partial x_j}\right] + \varsigma(x)^2 |\nabla h(x) - \nabla v(x)|^2$$

$$+ \underline{c}\left[\nabla v(x) \cdot \nabla v(x) - \operatorname{trace}\left(\frac{\partial^2 v(x)}{\partial x_i \, \partial x_j}\right)\right].$$

CRITERION 3.4. *Equation* (3.8) *is satisfied for* $\varsigma(x) = \exp[v(x)]$, *and*:

(a) $$\lim_{|x|\to\infty}\left[\nabla v(x) \cdot \nabla v(x) - \operatorname{trace}\left(\frac{\partial^2 v(x)}{\partial x_i \, \partial x_j}\right)\right] = 0;$$

(b) $$\lim_{|x|\to\infty} \varsigma(x)^2 \operatorname{trace}\left[\frac{\partial^2 v(x)}{\partial x_i \, \partial x_j} - \frac{\partial^2 h(x)}{\partial x_i \, \partial x_j}\right] + \varsigma(x)^2 |\nabla h(x) - \nabla v(x)|^2 = +\infty.$$

PROPOSITION 3.6. *Suppose Assumption* 3.1 *and Condition* 3.1 *are satisfied. Then Criterion* 3.4 *implies Condition* 2.1.

Restriction (a) of Criterion 3.4 limits the tail growth of the penalization. There are two reasons that such growth should be limited. The fast growth in Σ limits the functions that we hope to approximate using NPCs. Also, for C_K^2 to be a core for the form f, we require limits on growth in Σ (see Section 3.3.1).

Our use of $\check{W}$ in addition to $\check{V}$, in effect, replaces $-\varsigma^2 |\nabla v|^2$ with a second derivative term,

$$\varsigma(x)^2 \operatorname{trace}\left[\frac{\partial^2 v(x)}{\partial x_i \, \partial x_j}\right].$$

The following example illustrates the advantage of this replacement.

EXAMPLE 3.1. Let

$$v(x) = \frac{\beta}{2}\log(1 + |x|^2) + \frac{\tilde{c}}{2},$$

where $\tilde{c} = \log\underline{c}$. Thus ς grows like $|x|^\beta$ in the tails. Simple calculations result in

$$-\nabla v(x) \cdot \nabla v(x) = -\beta^2 \frac{|x|^2}{(1 + |x|^2)^2}$$

and

$$\operatorname{trace}\left[\frac{\partial^2 v(x)}{\partial x_i \, \partial x_j}\right] = \beta\left[\frac{n + (n-2)|x|^2}{(1 + |x|^2)^2}\right].$$

Notice that both terms converge to zero as $|x|$ gets large, but that the squared gradient scaled by ς^2 becomes arbitrarily large when $\beta > 1$. The first term is always negative, but the second one is nonnegative, provided that $n \geq 2$. Even when $n = 1$, the second term is larger than the first, provided that $\beta > 1$.[1] This example illustrates when Criterion 3.4 is preferred to Criterion 3.3. The distinction can be important when densities have algebraic tails.

This section contains our main existence results which we now summarize. We provided two criteria for constructing penalization functions that support the existence of countable many NPCs. The first one, Criterion 3.3, gives the most flexibility in terms of the penalization matrix Σ; but it is applicable for densities that have relatively thin tails. Densities with algebraic tails are precluded. The second one, Criterion 3.4, allows for densities with algebraic tails but requires that the penalization be more severe in the extremes to compensate for the tail thickness. Making the penalization more potent limits the class of functions that are approximated. Moreover, when the penalization is too extreme, we encounter an additional approximation problem: the family of functions C_K^2 ceases to be a core for the form used in the NPCs extraction.

4. Forms and Markov processes. So far, we considered the role of the penalization matrix Σ in the construction and approximation properties of NPCs. We now use stationary Markov diffusions to give an explicit interpretation of this penalization matrix.

We proceed as follows. Suppose the data, $\{x_i\}_{i=1}^T$, are generated by a Markov diffusion by sampling, say, at integer points in time. Specifically, $\{x_t : t \geq 0\}$ solves

$$dx_t = \mu(x_t)\,dt + \Lambda(x_t)\,dB_t$$

for some n-dimensional vector function μ and some n by n matrix function Λ of the Markov state with appropriate boundary restrictions where $\{B_t : t \geq 0\}$ is an n-dimensional, standard Brownian motion. Suppose further that this process has q as its stationary density and that $\Sigma = \Lambda \Lambda'$. We will have more to say in Section 6 about the restrictions on μ that are implicit in such a construction. Let ϕ be in C_K^2. Then it follows from Itô's lemma that the local variance of the process $\{\phi(x_t)\}$ is

$$(\nabla \phi)' \Sigma (\nabla \phi),$$

which is state dependent. Note that $f(\phi, \phi)$ is the average of this local variance. The local variance is the measure of magnitude of the instantaneous forecast error in forecasting $\{\phi(x_t)\}$ over the next instant given the current Markov state.

The NPC extraction given by Definition 2.1 can be performed equivalently as:

---

[1] We have previously established an alternative compactness criterion for $n = 1$ that does not involve second derivatives that may be preferred to Criterion 3.4.

DEFINITION 4.1.   The function ψ_j is the jth nonlinear principal component (NPC) for $j \geq 0$ if ψ_j, solves

$$\min_{\phi \in \bar{H}} f(\phi, \phi),$$

subject to

$$\langle \phi, \phi \rangle = 1,$$
$$\langle \psi_s, \phi \rangle = 0, \qquad s = 0, \ldots, j - 1.$$

Thus the NPCs are extracted by making the local forecast error (appropriately scaled) small for functions with unit second moments plus orthogonality. It is a continuous-time counterpart to the $(1 - R^2)$ in a forecasting regression. Recall that the NPCs satisfy $\langle \psi_j, \psi_k \rangle = 0$ and $f(\psi_j, \psi_k) = 0$ for $j \neq k$. These properties are nonlinear counterparts to the canonical components in the extraction of Box and Tiao [7]. Box and Tiao [7] show that their canonical analysis produces k component series that: (i) are ordered from least predictable to most predictable; (ii) are contemporaneously uncorrelated; and (iii) have contemporaneously uncorrelated forecast errors. In verifying our counterpart to the third property, notice that in continuous time the unpredictable component is $\nabla \psi_j(x_t) \Lambda(x_t) dB_t$, and thus $f(\psi_j, \psi_k)$ is the (average) local covariance of $\psi_j(x_t)$ and $\psi_k(x_t)$.

For financial and economics applications it is important to allow for barriers that are not attracting, and it is desirable to allow for a noncompact state space of the Markov process. Thus imposing uniform bounds on both q and the matrix Σ over compact state spaces is too restrictive. Our existence results in Section 3 avoid such restrictions.

Our construction of NPCs supports the estimation and testing of multivariate Markov diffusion models. There are other functional principal components constructions. For instance, Dauxois and Nkiet [12] construct nonlinear principal components for multivariate densities by choosing pairs of functions that maximize cross correlations without penalizing derivatives. Zhou and He [34] propose L^1-norm constrained principal components for the purpose of dimension reduction and variable filtering. Ramsay and Silverman [28] provide detailed discussions on functional principal component analysis for i.i.d. realization of curves.

5. Reversible diffusions.
We next consider how to use the form f to build a Markov process. Specifically associated with the form f, there is a second-order differential operator F that generates the semigroup of a Markov diffusion. The diffusion process has Σ as its local covariance matrix and q as it stationary density. The construction of F is *unique* provided that we restrict the process to be time reversible.

5.1. *A differential operator.*   There is a differential operator F_o that is associated with the form f_o [given in (2.1)], which we construct using integration-by-

parts. For any functions ϕ and ψ in C_K^2,

$$f_o(\phi, \psi) = \frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_j} \frac{\partial \psi}{\partial y_i} q$$

(5.1)

$$= -\frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} \psi q - \frac{1}{2} \int \sum_{i,j} \frac{\partial (q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j} \psi,$$

where the second equality of (5.1) follows from the integration-by-parts formula,

$$\int \sum_{i,j} \frac{\partial (q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j} \psi = -\int \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} \psi q - \int \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_j} \frac{\partial \psi}{\partial y_i} q.$$

We use (5.1) to motivate our interest in the differential operator F_o,

(5.2) $$F_o \phi = -\frac{1}{2} \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} - \frac{1}{2q} \sum_{i,j} \frac{\partial (q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j}.$$

This operator is constructed so that the form f_o can be represented as

$$f_o(\phi, \psi) = \langle F_o \phi, \psi \rangle = \langle \phi, F_o \psi \rangle,$$

where the second relation holds because we can interchange the role of ϕ and ψ in (5.1). Notice from (5.2) that operator F_o has both a first derivative term and a second derivative term. Symmetry (with respect to q) is built into the construction of this operator because of its link to the symmetric form f_o.

We are interested in the operator F_o because of its use in modeling Markov diffusions. Suppose that $\{x_t : t \geq 0\}$ solves the stochastic differential equation,

$$dx_t = \mu(x_t)\, dt + \Lambda(x_t)\, dB_t$$

with appropriate boundary restrictions where $\{B_t : t \geq 0\}$ is an n-dimensional, standard Brownian motion, and

$$\mu_j = \frac{1}{2q} \sum_{i=1}^{n} \frac{\partial (\sigma_{ij} q)}{\partial y_i}.$$

Set

$$\Sigma = \Lambda \Lambda'.$$

Then we may use Itô's lemma to show that for each $\phi \in C_K^2$,

$$-F_o \phi = \lim_{t \downarrow 0} \frac{E[\phi(x_t) | x_0 = x] - \phi(x)}{t},$$

where this limit is taken with respect to the L^2. That is, $-F_o$ coincides with the *infinitesimal generator* of $\{x_t\}$ in C_K^2. We use this link to the stochastic differential equation to motivate our use of the matrix Σ for penalizing derivatives. This matrix will also be the diffusion matrix for a continuous-time Markov process with stationary density q.

5.2. *Generating reversible diffusions.* Wong [33] constructed scalar diffusion models with stationary densities in the Pearson class and produced a spectral or eigenfunction decomposition of the associated one-parameter semigroup of conditional expectation operators. Banon [3] and Cobb, Kopstein and Chen [10] extended this analysis, in part, by taking the diffusion coefficient as given and constructing the implied drift coefficient for the stochastic differential equation that generates a prescribed stationary density. Banon [3] and Cobb, Kopstein and Chen [10] did not analyze the implied spectral decomposition of the associated conditional expectation operators. In all these analyses, the stationary density of the diffusion process is taken as one of the starting points of a model builder. In this section, we share Banon's [3] aim for generality, but at the same time, we retain Wong's [33] interest in spectral decompositions.

As in Banon [3], Cobb, Kopstein and Chen [10], Wong [33], we parameterize diffusion processes using the stationary density q and a (possibly state dependent) diffusion coefficient Σ in contrast to the more typical approach of starting with a drift and the diffusion coefficients. In contrast to Banon [3], Cobb, Kopstein and Chen [10], Wong [33], we allow the diffusion process to be multivariate on a state space Ω. For this to result in a unique diffusion, we require that the diffusion be time reversible.

A stochastic process is time reversible if its forward and backward transition probabilities are the same. Multivariate reversible diffusions can be parameterized directly by the pair (q, Σ). Associated with the closed extension f is a family of resolvent operators G_α indexed by a positive parameter α. We use the resolvent operators to build a semigroup of conditional expectation operators for a Markov process, and in particular, the generator of that semigroup.

For any $\alpha > 0$, the resolvent operator G_α is constructed as follows. Given a function $\phi \in L^2$, define $G_\alpha \phi \in \mathcal{D}(f)$ to be the solution to

$$(5.3) \qquad f(G_\alpha \phi, \psi) + \alpha \langle G_\alpha \phi, \psi \rangle = \langle \phi, \psi \rangle$$

for all $\psi \in \mathcal{D}(f)$. The Riesz representation theorem guarantees the existence of the $G_\alpha \phi$. This family of resolvent operators is known to satisfy several convenient restrictions (e.g., see Fukushima, Oshima and Takeda [18], pages 15 and 19). In particular, G_α is a one-to-one mapping from L^2 into $G_\alpha(L^2)$.

We associate with the form f the self-adjoint, positive semidefinite operator,

$$(5.4) \qquad F\phi = (G_\alpha)^{-1}\phi - \alpha\phi,$$

defined on the domain $G_\alpha(L^2)$. It can be shown that F is independent of α. Since the operator F is self-adjoint and positive semidefinite, we may define a unique positive semidefinite square root $\sqrt{F}$. While F may only be defined on a reduced domain, the domain of its square root may be extended uniquely to the entire space $\mathcal{D}(f)$ and $f(\phi, \psi) = \langle \sqrt{F}\phi, \sqrt{F}\psi \rangle$ (e.g., see Fukushima, Oshima and Takeda [18], Theorem 1.3.1). Moreover, it is an extension of the operator F_o

because f is an extension of f_o (e.g., see Lemma 3.3.1 of Fukushima, Oshima and Takeda [18]).

We also use the family of resolvent operators to build a semigroup of conditional expectation operators. A natural candidate for this semigroup is $\{\exp(-tF) : t \geq 0\}$. Formally, the expression $\exp(-tF)$ is not well defined as a series expansion. However, for any α and any t, we may form the exponential,

$$\exp(t\alpha^2 G_\alpha - \alpha t I)$$

as a Neumann series expansion. Notice that (5.4) implies

$$t\alpha^2 G_\alpha - t\alpha I = t\alpha\left[\left(I + \frac{1}{\alpha}F\right)^{-1} - I\right] = -tF\left(I + \frac{1}{\alpha}F\right)^{-1}.$$

Instead of directly using a series expansion, we use the limit,

$$\lim_{\alpha \to \infty} \exp[(t\alpha^2 G_\alpha) - \alpha t I] = \exp(-tF),$$

often referred to as "Yosida approximation" to construct formally a strongly continuous semigroup of operators indexed by $t \geq 0$.

We have just seen how to construct resolvent operators and the semigroup of conditional expectation operators from the form. We may *invert* this latter relation and obtain

$$(5.5) \qquad\qquad G_\alpha \phi = \int_0^\infty \exp(-\alpha t)\exp(-tF)\phi\,dt,$$

which is the usual formula for the resolvents of a semigroup of operators. The operator $-F$ is referred to as the generator of both the semigroup $\{\exp(-tF) : t \geq 0\}$ and of the family of resolvent operators $\{G_\alpha : \alpha > 0\}$.

As we have just seen, associated with a closed form f, there is an operator F and a (strongly continuous) semigroup $\{\exp(-tF) : t \geq 0\}$ on L^2. To establish that there is a Markov process associated with this semigroup, we need first to verify that the semigroup satisfies two properties. First we require, for each $t \geq 0$ and each $0 \leq \phi \leq 1$ in L^2, $0 \leq \exp(-tF)\phi \leq 1$. A semigroup satisfying this property is called *sub-Markov* in the language of Beurling and Deny [5]. Second we require, for each $t \geq 0$, $\exp(-tF)1 = 1$. A semigroup satisfying this property is said to *conserve probabilities*. We refer to a sub-Markov semigroup that conserves probabilities as a Markov semigroup. Finally we must make sure that the Markov semigroup is actually the family of conditional expectation operators of a Markov process.

The following condition is sufficient for a closed form to generate a sub-Markov semigroup (e.g., see Davies [14], Section 1.3).

CONDITION 5.1 (Beurling–Deny).   For any $\phi \in \mathcal{D}(f)$, ψ given by the truncation,

$$\psi = (0 \vee \phi) \wedge 1$$

is in $\mathcal{D}(f)$ and

$$f(\psi, \psi) \leq f(\phi, \phi).$$

When this condition is satisfied, the semigroup $\exp(-tF)$ is sub-Markov, and for each $t \geq 0$, $\exp(-tF)$ is an L^2 contraction ($\|\exp(-tF)\phi\|_2 \leq \|\phi\|_2$). This contraction property is also satisfied for the L^p norm for $1 \leq p \leq \infty$ (Davies [14], Theorem 1.3.3). In particular, we may extend the semigroup from L^2 to L^1 while preserving the contraction property.

PROPOSITION 5.1. *There exists a self adjoint operator F associated with f, which is an extension of F_o and generates a semigroup $\{\exp(-tF) : t \geq 0\}$. The density q is the stationary density for this diffusion, the matrix Σ is the diffusion matrix and $\exp(-tF)$ is the conditional expectation operator over an interval of time t.*

5.3. *Nonlinear principal components and eigenfunctions.* Continuous time Markov process models are typically specified in terms of their local dynamics. Given the nonlinearity in the state variables, it is a nontrivial task to infer the global dynamics, and, in particular, the long-run behavior from this local specification. Characterizing eigenfunctions of conditional expectation operators offer a way of approximating intermediate and long-term dynamics in ways that are typically disguised from the local dynamics in nonlinear settings.

Eigenfunctions of the closed form f will also be eigenfunctions of the resolvent operators G_α and of the generator F. For convenience, we rewrite (5.3),

$$f(G_\alpha \phi, \psi) + \alpha \langle G_\alpha \phi, \psi \rangle = \langle \phi, \psi \rangle.$$

From this formula, we may verify that f and G_α must share eigenfunctions for any $\alpha > 0$. The eigenvalues are related via the formula,

$$\lambda = \frac{1}{\delta + \alpha},$$

where λ is the eigenvalue of G_α and δ is the corresponding eigenvalue of f.

Given the relation between the generator F and the resolvent operator G_α,

$$F\phi = (G_\alpha)^{-1}\phi - \alpha\phi,$$

these two operators must share eigenfunctions. Moreover, eigenfunctions of the operators F, G_α and the form f must belong to the domain of F or equivalently to the image of G_α. This domain is contained in the domain of the form f. Similarly, we may show that if ϕ is an eigenfunction of the form f with eigenvalue δ, then ϕ is an eigenfunction of $\exp(-tF)$ with eigenvalue $\exp(-t\delta)$ for any positive t.

An eigenfunction ψ of the generator F satisfies

(5.6) $$E[\psi(x_{t+s})|x_t] = \exp(-\delta s)\psi(x_t)$$

for some positive number δ and each transition interval s. Thus the NPCs described previously will also satisfy the testable conditional moment restrictions (5.6). The

scalar process $\{\psi(x_t)\}$ should behave as a scalar autoregression with autoregressive coefficient $\exp(-\delta s)$ for sample interval s. The forecast error: $\psi(x_{t+s}) - \exp(-\delta s)\psi(x_t)$ will typically be conditionally heteroskedastic (have conditional variance that depends on the Markov state x_t).

Since the form f can be depicted using a principal component decomposition as in (2.3), analogous decompositions are applicable to F and $\exp(-tF)$,

$$F\phi = \sum_j \delta_j \frac{\langle \phi, \psi_j \rangle}{\langle \psi_j, \psi_j \rangle} \psi_j,$$

$$\exp(-tF)\phi = \sum_j \exp(-t\delta_j) \frac{\langle \phi, \psi_j \rangle}{\langle \psi_j, \psi_j \rangle} \psi_j,$$

where the first expansion is only a valid L^2 series when ϕ is in the domain of the operator F. When the eigenvalues δ_j of the form increase rapidly (in j), the term $\exp(-t\delta_j)$ will decline to zero rapidly (in j), more so when the time horizon t becomes large. As a consequence, it becomes easier to approximate the conditional expectation operator over a finite transition interval t with a smaller number of NPCs. On the other hand, slow eigenvalue divergence of the form will make it challenging to approximate the transition operators with a small number of NPCs. Our results in Section 3 give primitive conditions based on the tail behaviors of stationary density and diffusion matrix for the existence of the above eigenfunction decompositions. In an earlier, longer version of our paper, we provided primitive sufficient conditions for the speed of eigenvalue decays.

5.4. *An alternative form.* In this subsection we construct a second quadratic form used to depict the long-run variance of a stochastic processes constructed from the Markov process $\{x_t\}$.

This quadratic form is defined to be the limit

$$g(\phi, \psi) = 2 \lim_{\alpha \downarrow 0} \langle G_\alpha \phi, \psi \rangle$$

and is well defined on a subspace $\mathcal{S}(F)$ of functions in L^2 for which

$$\lim_{\alpha \downarrow 0} \langle G_\alpha \phi, \phi \rangle < \infty.$$

While the form f is used to define the operator F, the form g may be used to define F^{-1} as is evident from formulas (5.3) or (5.4). The forms f and g share eigenfunctions. The g eigenvalues are the reciprocals of the f eigenvalues.

In light of (5.5),

$$(5.7) \qquad \langle G_\alpha \phi, \psi \rangle = \int_0^\infty \exp(-\alpha t) E[\phi(x_t)\psi(x_0)]\, dt.$$

Hence using (5.4), we obtain

$$g(\phi, \psi) = \lim_{\alpha \downarrow 0} 2 \langle G_\alpha \phi, \psi \rangle = \lim_{\alpha \downarrow 0} 2 \langle (\alpha I + F)^{-1}\phi, \psi \rangle.$$

Notice that this form is symmetric because the resolvent operator is self-adjoint for any positive α. Using (5.7) we may write this form as

$$g(\phi, \psi) = \int_{-\infty}^{+\infty} E[\phi(x_t)\psi(x_0)]\, dt = \int_{-\infty}^{+\infty} E[\psi(x_t)\phi(x_0)]\, dt.$$

PROPOSITION 5.2.   *The jth nonlinear principal component ψ_j for $j \geq 1$ solves*

$$\max_{\phi \in \mathcal{S}(F)} g(\phi, \phi),$$

subject to

$$\langle \phi, \phi \rangle = 1,$$

$$\langle \psi_s, \phi \rangle = 0, \qquad s = 0, \ldots, j - 1,$$

where ψ_0 is initialized to be the constant function one.

Recall that the spectral density function at frequency θ for a stochastic process $\{\phi(x_t)\}$ is defined to be

$$\int_{-\infty}^{+\infty} \exp(-i\theta t) E[\phi(x_t)\phi(x_0)]\, dt,$$

whenever this integral is well defined. In particular $g(\phi, \phi)$ is the spectral density of the process $\{\phi(x_t)\}$ at frequency zero, a well-known measure of the long-run variance.

For an alternative but closely related defense of the term *long-run* variance, suppose that $\phi = F\psi$ for some ψ in the domain of F. Then

$$M_T = \psi(x_T) - \psi(x_0) + \int_0^T \phi(x_s)\, ds$$

is a martingale adapted to the Markov filtration. Following Bhattacharya [6] and Hansen and Scheinkman [21], we may use this martingale construction to justify

$$\frac{1}{\sqrt{T}} \int_0^T \phi(x_s)\, ds \Rightarrow \text{Normal}(0, g(\phi, \phi)).$$

Thus $g(\phi, \phi)$ is the limiting variance for the process $\{\frac{1}{\sqrt{T}} \int_0^T \phi(x_s)\, ds\}$ as the sample length T goes to infinity.

This gives us an alternative interpretation of our NPCs. We may base the extraction on maximizing $g(\phi, \phi)$ subject to $\langle \phi, \phi \rangle = 1$ and orthogonality constraint. In words, we are maximizing long-run variation while constraining the overall variation. Smooth functions of a Markov state are also highly persistent and as a consequence maximize long-run variation.

6. Irreversible diffusions. The stationary Markov construction we used in the previous section resulted in a generator that was self adjoint and hence a

process that was time reversible. Even among the class of stationary Markov diffusions, reversibility is special when the process has multiple dimensions. Given a stationary density q and a diffusion matrix Σ, we have seen how to construct a reversible diffusion, but typically there are other diffusions that share the same density and diffusion matrix, but are not reversible. We now characterize the drifts of such processes.

Instead of constructing a Markov process implied by a form, suppose we have specified the process as a semigroup of conditional expectation operators indexed by the transition interval. We suppose this process has stationary density q. Following Nelson [25] and Hansen and Scheinkman [21] we study the semigroup of conditional expectation operators on the space L^2. This semigroup has a generator A defined on a dense subspace of L^2. Consistent with our construction of F, on the subspace of C_K^2, we suppose that A can be represented as a second-order differential operator,

$$A\phi = \frac{1}{2} \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \, \partial y_j} + \sum_j \mu_j \frac{\partial \phi}{\partial y_j}$$

and that

$$\int A\phi q = 0.$$

It may be shown that

$$-\int \psi (A\phi) q = f_o(\phi, \psi) \qquad \text{on } C_K^2.$$

This construction does not require that $A = -F$ or that A be self-adjoint. How can the adjoint be represented? The adjoint must satisfy

$$-\int \phi (A^* \psi) q = f_o(\phi, \psi),$$

implying that the F constructed previously must satisfy $F = -(A + A^*)/2$. Moreover, since q is also the stationary density of the reverse time process,

$$\int A^* \phi q = 0.$$

It follows from Nelson [25] that the adjoint operator has the same diffusion matrix, but a different drift vector. The drift for the adjoint operator A^* is given by

$$\mu^* = -\mu + \frac{1}{q} \sum_{i,j} \frac{\partial(q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j}.$$

The adjoint operator generates the semigroup of expectation operators for the reverse time diffusion. From the formula for reverse time drift, μ^*, it follows that

$$\frac{\mu + \mu^*}{2} = \frac{1}{2q} \sum_{i,j} \frac{\partial(q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j},$$

which is the negative of the second term in representation (5.2) for F_o. Thus if the generator A of the semigroup is not self adjoint, then the operator F implied by the form is a second order differential operator built using a simple average of the forward and reverse time drift coefficients, μ and μ^*, and the common diffusion matrix, Σ.

REMARK 6.1. The density q and the diffusion matrix Σ do place other restrictions on the drift vector μ. Since q is the stationary density, μ and μ^* must also satisfy

$$\frac{\partial(\mu q)}{\partial y} = \frac{\partial}{\partial y} \sum_{i,j} \frac{\partial(q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j},$$

$$\frac{\partial(\mu^* q)}{\partial y} = \frac{\partial}{\partial y} \sum_{i,j} \frac{\partial(q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j}.$$

While there is typically one solution μ (or μ^*) to this equation for the scalar case, multiple solutions will exist for the multivariate case. That is, unless reversibility is imposed a priori, the drift cannot be identified from the density and diffusion matrix; but the average of the forward and backward drift can be inferred.

REMARK 6.2. The NPCs existence results of Section 3 have an immediate extension to the existence of eigenfunctions of the semigroup of conditional expectation operators when the Markov diffusion is not reversible. For a semigroup with generator A, we may "invert" (5.4) to construct a family of resolvent operators,

$$R_\alpha \phi = \int_0^\infty \exp(-\alpha t) \exp(At)\phi \, dt = (\alpha I - A)^{-1}\phi,$$

and a form $f(\phi, \psi) = \langle \phi, A\psi \rangle$, which is not necessarily symmetric. While the generator is an unbounded operator on L^2, the resolvent operators are bounded. When the resolvents are compact operators, they have well defined eigenfunctions and eigenvalues, but they may be complex valued (see Rudin [30], Theorem 4.25, page 108).

Given α, the resolvent operator will be compact provided that the image of R_α of the L^2 unit ball has compact closure. Consider a function φ given by

$$\varphi = (\alpha I - A)^{-1}\phi.$$

Then $\phi \in \mathcal{D}(A)$ and

$$\langle \phi, \phi \rangle = \alpha^2 \langle \varphi, \varphi \rangle - 2\alpha \langle \varphi, A\varphi \rangle + \langle A\varphi, A\varphi \rangle \geq \alpha^2 \langle \varphi, \varphi \rangle + 2\alpha f(\varphi, \varphi).$$

Thus it suffices to show that

$$\{\varphi \in \mathcal{D}(A) : \alpha^2 \langle \varphi, \varphi \rangle + 2\alpha f(\varphi, \varphi) \leq 1\}$$

has compact closure. This set will have compact closure, if, and only if, compactness Condition 2.1 is satisfied for $\theta = \alpha/2$.

7. Extraction of NPCs from data.   In this section, we suggest a sieve method to estimate NPCs based on discretely-sampled data from a Markov process. Here we only sketch the construction and leave the formal justification and detailed analysis of rates of convergence for subsequent research.

Let $\{x_i\}_{i=1}^T$ be a discrete-time sample of the underlying continuous-time, ergodic Markov diffusion $\{x_t : t \geq 0\}$ on the state space $\Omega \subseteq \mathbb{R}^n$.[2] Suppose that the penalization matrix Σ is either known or could be consistently estimated. The invariant probability measure Q is unknown but is consistently estimated by the empirical distribution of the data $\{x_i\}_{i=1}^T$.

Let $\{H_m : m = 1, 2, \ldots\}$ be a sequence of increasing finite-dimensional linear (sieve) spaces that approximate the Hilbert space $\bar{H}$ (the domain of the form f) as m goes to infinity. For notational convenience, let m be the dimension of H_m, and suppose that m goes to infinity slowly as the sample size T goes to infinity. One strategy is to extract the finite sample approximations sequentially as in optimization problem given in Definition 2.1. For a finite-dimensional sieve approximation, it suffices to solve a generalized eigenvector problem.

Since the space H_m is finite dimensional,

$$H_m = \left\{ \phi(x) = \sum_{k=1}^m b_k B_k(x) \right\},$$

where the basis functions $\{B_k(x) : k \geq 1\}$ are used to construct the sieve. For example, $\{B_k(x) : k \geq 1\}$ could be one of the following: (i) thin-plate splines, radial-basis wavelets or tensor-product wavelets if q has algebraic tails on its support $\Omega = \mathbb{R}^n$; (ii) tensor-product Hermite polynomial basis or Gaussian radial basis if q has exponential thin tails on its support $\Omega = \mathbb{R}^n$.

Form a vector Ψ_m of functions of x by stacking terms $B_k(x)$, $k = 1, \ldots, m$, and form two matrices:

$$\mathbf{V}_T = \frac{1}{2T} \sum_{i=1}^T [\nabla \Psi_T(x_i)] \Sigma(x_t) [\nabla \Psi_T(x_t)]^*,$$

$$\mathbf{W}_T = \frac{1}{T} \sum_{i=1}^T \Psi_T(x_i) \Psi_T(x_i)^*,$$

where $*$ denotes transpose. Both matrices are symmetric and positive semidefinite by construction. The matrix $\mathbf{W}_T$ is typically nonsingular while $\mathbf{V}_T$ is singular

---

[2]If available a continuous-time record could be used, but statistical approximation remains an issue because the length of the record is finite.

when a constant function is in the sieve space H_m. Stack the coefficients on the sieve basis into a vector a_T. The sample counterpart to the NPC (or eigenfunction) problem is the following generalized eigenvector problem:

$$(7.1) \qquad \mathbf{V}_T \widehat{a}_T = \widehat{\delta}_T \mathbf{W}_T \widehat{a}_T,$$

where $\widehat{a}_T$ is a generalized eigenvector and $\widehat{\delta}_T$ a generalized eigenvalue. Since $\mathbf{W}_T$ is positive definite, we may apply the Cholesky decomposition to transform this generalized eigenvector problem into a standard eigenvector problem.

Associated with each eigenvector solution to (7.1), is an eigenfunction formed by multiplying the coefficient entries of the eigenvector by the sieve basis functions. We have constructed this sample problem so that one of the approximating eigenfunctions will be constant whenever there is a nonzero constant function in H_m, and the associated eigenvalue is zero.[3]

8. Conclusions and related literature.

We have studied NPCs from multiple vantage points. We have explored their role in capturing variation subject to smoothness constraints and their role in capturing long-run variation in time-series modeling. We have also considered their use in approximation where the smoothness constraints limit the family of functions to be approximated.

We also used multivariate Markov diffusions as data generating devices to interpret our NPCs. These NPCs are eigenfunctions of conditional expectation operators when the Markov process is reversible and hence imply conditional moment restrictions. Our analysis expands on the result of Hansen and Scheinkman [21] that reversible diffusions can be identified nonparametrically from discrete-time, low-frequency, stationary observations.

For more general diffusions, these NPCs are orthogonal and have orthogonal innovations analogous to those from the canonical analysis of Box and Tiao [7] for linear multiple time series models (see also Pan and Yao [26]). Thus our NPC construction provides a convenient way to summarize implications of multivariate nonlinear diffusion models. Given the nonlinearity in the state variables, it is a non-trivial task to infer the global dynamics, and, in particular, the long-run behavior from this local specification based on low-frequency data. Our characterization of NPCs offers a way to characterize features of the implied time series that are typically disguised from the local dynamics. While we featured diffusion processes, more general processes, including processes with jumps, can be accommodated by expanding the types of forms that are considered.

The idea of using eigenfunctions of conditional expectation operators for estimation and testing of Markov processes based on low-frequency data has been suggested previously by Chen, Hansen and Scheinkman [9], Demoura [15], Florens,

---

[3]For reversible diffusions we could instead approximate the NPCs nonparametrically by maximizing autocorrelations. For scalar diffusion models, this method has been already considered in Chen, Hansen and Scheinkman [9] and Gobet, Hoffmann and Reib [19]. Both papers used a wavelet sieve as the approximating space.

Renault and Touzi [17], Gobet, Hoffmann and Reib [19], Hansen and Scheinkman [21], Hansen, Scheinkman and Touzi [22] and Kessler and Sorensen [24]. In particular, Kessler and Sorensen [24] use eigenfunctions to construct quasi-optimal estimators of parametric scalar diffusion models of the drift and diffusion coefficients from discrete-time data in the special case in which the functional forms of eigenfunctions are known a priori. Chen, Hansen and Scheinkman [9], Darolles, Florens and Gourieroux [11], Gobet, Hoffmann and Reib [19], Hansen and Scheinkman [21] and Hansen, Scheinkman and Touzi [22] study semiparametric and nonparametric identification and over-identification based on an eigenfunction extraction that is closely related to the one analyzed here (see Fan [16] for a recent review). This previous literature focuses primarily on scalar diffusion models, and, in some cases, on scalar diffusions on compact state spaces with reflective boundaries. Our analysis of Markov diffusions extends to multivariate settings applicable to processes without attracting barriers.

In this paper, we have characterized a particular type of functional principal components motivated in part by long-run implications of multivariate Markov diffusions. This is a natural first step. Inferential issues, while crucial, are beyond the scope of this paper. Formalizing statistical comparisons of models and data in a multivariate setting is an obvious next step, supported by either parametric, semiparametric or nonparametric estimation. There are a number of recent statistical results on estimation and inference of functional principal components of covariance operators associated with i.i.d. or longitudinal sample of curves (see, e.g., Benko, Hardle and Kneip [4], Hall, Muller and Wang [20], Ramsay and Silverman [28], Silverman [32] and Zhou, Huang and Carroll [35]). These existing results can, in principle, be modified to establish asymptotic properties of our estimated NPCs from discrete-time, low-frequency realizations of an underlying multivariate Markov diffusion model.

APPENDIX: PROOFS

PROOF OF CLAIM 2.2.   In solving the maximization component of the problem, first limit the ϕ's to be in H_N but orthogonal to $\hat{H}$. This can only reduce maximized value. The space of such ϕ's contains more than just the zero element because H_N has $N + 1$ dimensions. Write ϕ as $\phi = \sum_{j=0}^{N} r_j \psi_j$. Since $\text{Proj}(\phi|\hat{H}) = 0$, the objective can be expressed as $\sum_{j=0}^{N} (r_j)^2 \lambda_j$. The constraint set implies that

$$\sum_{j=0}^{N} (r_j)^2 \leq 1,$$

because $f(\psi_j, \psi_\ell) = \langle \psi_j, \psi_\ell \rangle = 0$ for $j \neq \ell$. While the coefficients r_j cannot be freely chosen (ϕ must be orthogonal to $\hat{H}$), they can be scaled so that the constraint is satisfied with equality. Since the sequence of λ_j's is decreasing, the maximized objective must be no less than λ_N.   □

PROOF OF CLAIM 2.3.   Write ϕ as: $\phi = \text{Proj}(\phi|H_{N-1}) + \varphi$ where φ is in $H_{N-1}^{\perp}$. Write

$$\text{Proj}(\phi|H_{N-1}) = \sum_{j=0}^{N-1} \mathsf{r}_j \psi_j.$$

Using this decomposition, the objective can be written as $\langle \varphi, \varphi \rangle$, and the constraint set can be written as

$$\sum_{j=0}^{N-1} (\mathsf{r}_j)^2 + \langle \varphi, \varphi \rangle + \theta f(\varphi, \varphi) \leq 1,$$

because $\psi_1, \psi_2, \ldots, \psi_{N-1}, \varphi$ are orthogonal, and $f(\psi_j, \varphi) = f(\psi_j, \psi_\ell) = 0$ for $j = 0, \ldots, N-1$ and $\ell = j+1, j+2, \ldots, N-1$. To maximize the objective, the coefficients r_j's are set to zero, and φ is chosen by solving Problem 2.1 for $H = H_{N-1}^{\perp}$. The conclusion follows. $\square$

PROOF OF PROPOSITION 3.1.   Hansen, Scheinkman and Touzi [22] consider densities from stationary scalar diffusions, whose boundaries are not attracting. This proposition gives an equivalent statement of their compactness condition, written in terms of the stationary density. The scalar diffusion coefficient in their analysis is ς^2. $\square$

To show that the form f is closed extension of f_o, we verify that $\bar{H}$ is a Hilbert space.

PROPOSITION A.1. $\bar{H}$ is a Hilbert space.

PROOF.   Let Λ be the symmetric square root of the penalty matrix Σ. If $\{\phi_j\}$ is a Cauchy sequence in $\bar{H}$, then $\{\phi_j\}$ and the entries of $\{\Lambda \nabla \phi_j\}$ form Cauchy sequences in L^2. Denote the limits in L^2 as

$$\phi = \lim_{j \to \infty} \phi_j,$$

$$v = \lim_{j \to \infty} \Lambda \nabla \phi_j.$$

For each $u \in C_K^1$ we know that

$$\int \phi_j \frac{\partial u}{\partial x} = -\int (\nabla \phi_j) u,$$

where $\frac{\partial u}{\partial x}$ is the partial derivative of u with respect to x. Since Σ is positive definite and continuous on any compact subset of Ω, and u vanishes outside any such set, it follows that

$$\int \phi \frac{\partial u}{\partial x} = -\int (\Lambda^{-1} v) u.$$

Hence $\phi \in \bar{H}$ with $\nabla \phi = \Lambda^{-1} v$. Moreover, $\phi_n \to \phi$ in $\bar{H}$. $\square$

X. CHEN, L. P. HANSEN AND J. SCHEINKMAN

We now present a criteria for Condition 3.1 to hold. This result is due, essentially, to Azencott [2] and Davies [13].

PROPOSITION A.2. *Consider a form f_o that satisfies the Beurling–Deny Criterion 5.1. Let $\hat{f}$ denote the minimal extension of f_o with domain $\mathcal{D}(\hat{f})$. Suppose that $1 \in \mathcal{D}(\hat{f})$ and $\hat{f}(1, \phi) = 0$ for all $\phi \in \mathcal{D}(\hat{f})$. Then $\hat{f} = f$.*

PROOF. As explained in Section 5, associated with the forms f and $\hat{f}$, we may construct operators F and $\hat{F}$ and resolvents G and $\hat{G}$. Integration by parts can be used to show that the operators F and $\hat{F}$ are extensions of the differential operator

$$\hat{L}\phi = -\frac{1}{q} \sum_{i,j} \frac{\partial}{\partial x_i} \left(q\sigma_{ij} \frac{\partial \phi}{\partial x_j} \right),$$

defined on C_K^2. The form $\hat{f}$ and hence the form f satisfies the Beurling–Deny Criterion 5.1 (Davies [14], Theorem 1.3.5). Hence as stated in Davies [13], the operators F and $\hat{F}$ can be extended to subspaces of L^1. Similarly, the resolvents G and $\hat{G}$ can be extended to L^1. We will denote the extended operators as F^1, $\hat{F}^1$, G^1 and $\hat{G}^1$. Since q is integrable, L^2 convergence implies L^1 convergence and consequently F and $\hat{F}$ are restrictions of F^1 and $\hat{F}^1$, respectively. The same argument applies to the resolvent operators.

If $\hat{f}(1, \phi) = 0$ for all $\phi \in \mathcal{D}(\hat{f})$, then $\hat{F}1 = 0$ and $\hat{G}1 = 1$. Consequently, $G^1 1 = 1$. It follows from Theorem 2.2 in Davies [13] that C_K^2 is a core for F^1, in the sense that F^1 is the closure in L^1 of $\hat{L}$.[4] Hence C_K^2 is a core for F^2, and thus a core for f, or, equivalently, f and $\hat{f}$ coincide. □

PROOF OF PROPOSITION 3.2. Since $\hat{f}$ is the minimal closed extension, it has C_K^2 as its core. When this condition is met, a sequence of functions ϕ_j in C_K^2 can be constructed that converge to 1 in L^2, and $f(\phi_j, \phi_j)$ converges to zero. See Fukushima, Oshima and Takeda [18], Theorems 1.6.6 and 1.6.7. An approximating sequence of functions with compact support is supplied by Fukushima, Oshima and Takeda [18] in the proof of Theorem 1.6.7. This sequence can be smoothed using a suitable regularization to produce a corresponding approximating sequence in C_K^2. Thus the unit function is in the domain of $\hat{f}$ and $\hat{f}(1, \phi) = 0$ for $\phi \in C_K^2$ and hence for $\phi \in \mathcal{D}(\hat{f})$. As we established above, this is sufficient for Condition 3.1. □

---

[4]Davies [13] assumes that the coefficients of $\hat{L}$ are C^∞. However, the proof holds for C^2 coefficients since elliptic regularity holds even when the coefficients are only Lipschitz (see Theorem 6.3 of Agmon [1]).

PROOF OF PROPOSITION 3.3. Since V is bounded from below, we may choose a $\theta > 0$ such that $V + \theta$ is nonnegative. Construct the space,

$$\check{H} \doteq \left\{ \psi \in L^2(\text{leb}) \colon \int (V + \theta)(\psi)^2 < \infty, \text{ there exists } g \text{ measurable, with} \right.$$

$$\left. \int g' \Sigma g < \infty, \text{ and } \int \psi \nabla \varphi = - \int g \varphi, \text{ for all } \psi \in C_K^1 \right\}.$$

As in the proof of Proposition A.1, it follows that $\check{H}$ is a Hilbert space with inner product,

$$\int (V + \theta + 1)\psi \tilde{\psi} + \int (\nabla \psi)' \Sigma (\nabla \tilde{\psi}).$$

We show that $U \bar{H} \subset \check{H}$.

Since C_K^2 is a core for f, there exists a sequence $\{\phi_j : j = 1, 2, \ldots\}$ in C_K^2 that converges to ϕ in the Hilbert space norm of $\bar{H}$. Hence this sequence is Cauchy in that norm. Writing $\psi_j = U\phi_j$ and applying (3.5) we obtain

$$\int (\phi_j - \phi_\ell)^2 (1 + \theta) q + \int (\nabla \phi_j - \nabla \phi_\ell)' \Sigma (\nabla \phi_j - \nabla \phi_\ell) q$$

$$= \int (V + \theta + 1)(\psi_j - \psi_\ell)^2 + \int (\nabla \psi_j - \nabla \psi_\ell)' \Sigma (\nabla \psi_j - \nabla \psi_\ell).$$

Thus $\{\psi_j : j = 1, 2, \ldots\}$ is Cauchy in the Hilbert space norm of $\check{H}$ and the limit point ψ must satisfy $\psi = U\phi$. Notice that $\int V(\psi)^2 + \int (\nabla \psi)' \Sigma (\nabla \psi)$ equals $\check{H}$ squared norm minus $\theta + 1$ times the $L^2(Q)$ squared norm. Thus

$$\int V(\psi)^2 + \int (\nabla \psi)' \Sigma (\nabla \psi) = \lim_{j \to \infty} \int (\nabla \psi_j)' \Sigma (\nabla \psi_j) q$$

$$= \lim_{j \to \infty} (\nabla \phi_j)' \Sigma (\nabla \phi_j) q$$

$$= \int (\nabla \phi)' \Sigma (\nabla \phi) q.$$

This proves (3.7).

For a given $\psi = U\phi$, our candidate for the weak derivative is

$$g \doteq \exp(-h)(-\phi \nabla h + \nabla \phi).$$

To verify that g is indeed the weak derivative, we must show that for any $\varphi \in C_K^1$,

(A.1) $$\int \psi \nabla \varphi = - \int g \varphi$$

and

(A.2) $$\int g' \Sigma g < \infty.$$

We check relation (A.1) by applying integration by parts,

$$-\int \nabla\psi\varphi = -\int [\exp(-h)(-\phi\nabla h + \nabla\phi)]\varphi$$

$$= -\int \nabla\phi \exp(-h)\varphi + \int \exp(-h)\varphi\phi\nabla h$$

$$= \int \phi[\exp(-h)\nabla\varphi - \nabla h \exp(-h)\varphi] + \int \exp(-h)\varphi\phi\nabla h = \int \psi\nabla\varphi.$$

Inequality (A.2) follows from (3.7). $\square$

PROOF OF PROPOSITION 3.4. Since V is continuous and diverges at the boundaries, it must be bounded from below. Also, it follows from Assumption 3.2 that

$$\mathcal{V}_\theta \subset \left\{\psi \in L^2(\text{leb}) : \psi \text{ has a weak derivative and}\right.$$

$$\left.\int \left(\theta + \frac{1}{2}V\right)(\psi)^2 + \frac{\mathsf{c}}{2}\int |\nabla\psi|^2 \leq 1\right\}.$$

We may then apply the argument in the proof of Theorem XIII.67 of Reed and Simon [29] to establish that $\mathcal{V}_\theta$ is precompact in $L^2(\text{leb})$. $\square$

PROOF OF LEMMA 3.1. Consider a positive function

$$\chi(x) = \frac{1}{\varsigma}$$

and note that

$$[\varsigma(x)^2 - \mathsf{c}]\nabla\chi(x) = -\varsigma(x)\nabla v + \mathsf{c}\frac{\nabla v(x)}{\varsigma(x)}.$$

For ϕ in C_K^2, we may apply integration by parts to show that

$$\int (\varsigma^2 - \mathsf{c})\nabla\chi \cdot \nabla\phi = \int \left[\left(\frac{\varsigma^2 + \mathsf{c}}{\varsigma}\right)(\nabla v \cdot \nabla v) + \left(\frac{\varsigma^2 - \mathsf{c}}{\varsigma}\right)\text{trace}\left(\frac{\partial^2 v}{\partial x_i\,\partial x_j}\right)\right]\phi$$

$$= \int \check{W}\chi\phi.$$

The conclusion follows from Theorem 1.5.12 of Davies [14]. While Davies [14] uses test functions ϕ in C_K^∞, the same proof applies to C_K^2 test functions. $\square$

PROOF OF PROPOSITION 5.1. The form f satisfies the Beurling–Deny criteria (Davies [14], Theorem 1.3.5). Thus there exists a self-adjoint operator F which is an extension of F_o and generates a sub-Markov semigroup $\exp(-tF)$. Theorem 7.2.1 of Fukushima, Oshima and Takeda [18] guarantees that there exists a

Markov process $\{x_t\}$ that has $\exp(-tF)$ as its semigroup of conditional expectations. The semigroup $\exp(-tF)$ conserves probability because the unit function is in the domain of the form f and $f(1, \phi) = 0$ for any $\phi \in \mathcal{D}(f)$. As a consequence, the unit function is also in the domain of the operator F, $F1 = 0$. $\quad \square$

REFERENCES

[1] AGMON, S. (1965). *Lectures on Elliptic Boundary Problems*. Van Nostrand, Princeton, NJ. MR0178246

[2] AZENCOTT, R. (1974). Behavior of diffusion semi-groups at infinity. *Bull. Soc. Math. France* **102** 193–240. MR0356254

[3] BANON, G. (1978). Nonparametric identification of diffusions. *SIAM J. Control Optim.* **16** 380–395. MR0492159

[4] BENKO, M., HARDLE, W. and KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343

[5] BEURLING, A. and DENY, J. (1958). Espaces de Dirichlet *i*, le cas elementaire. *Acta Math.* **99** 203–224. MR0098924

[6] BHATTACHARYA, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete* **60** 185–201. MR0663900

[7] BOX, G. E. P. and TIAO, G. C. (1977). Canonical analysis of multiple time series. *Biometrika* **64** 355–365. MR0519089

[8] BREZIS, H. (1983). *Analyse Fonctionelle*. Masson, Paris. MR0697382

[9] CHEN, X., HANSEN, L. P. and SCHEINKMAN, J. (1998). Shape-preserving estimation of diffusions. Working paper, Univ. Chicago.

[10] COBB, L. P., KOPSTEIN, P. and CHEN, N. (1983). Estimation and moment recursion relations for multimodal distributions in the exponential family. *J. Amer. Statist. Assoc.* **78** 124–130. MR0696856

[11] DAROLLES, S., FLORENS, J. P. and GOURIEROUX, C. (2004). Kernel based nonlinear canonical analysis and time reversibility. *J. Econometrics* **119** 323–353. MR2057103

[12] DAUXOIS, J. and NKIET, G. M. (1998). Nonlinear canonical analysis and independence tests. *Ann. Statist.* **26** 1254–1278. MR1647653

[13] DAVIES, E. B. (1985). l^1 properties of second order operators. *Bull. London Math. Soc.* **17** 417–436. MR0806008

[14] DAVIES, E. B. (1989). *Heat Kernels and Spectral Theory*. Cambridge Univ. Press, Cambridge. MR0990239

[15] DEMOURA, S. (1998). The nonparametric estimation of the expected value operator. Workshop Presentation, Univ. Chicago.

[16] FAN, J. (2005). A selective overview of nonparametric methods in financial econometrics (with discussion). *Statist. Science* **20** 317–357. MR2210224

[17] FLORENS, J. P., RENAULT, E. and TOUZI, N. (1998). Testing for embeddability by stationary reversible continuous-time Markov processes. *Econometric Theory* **69** 744–69. MR1666704

[18] FUKUSHIMA, M., OSHIMA, Y. and TAKEDA, M. (1994). *Dirichlet Forms and Symmetric Markov Processes*. Walter de Gruyter, Berlin. MR1303354

4312   X. CHEN, L. P. HANSEN AND J. SCHEINKMAN

[19] GOBET, E., HOFFMANN, M. and REIß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **26** 2223–2253. MR2102509
[20] HALL, P., MULLER, H. G. and WANG, J. L. (2006). Properties of principal components methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365
[21] HANSEN, L. P. and SCHEINKMAN, J. (1995). Back to the future. *Econometrica* **63** 767–804. MR1343081
[22] HANSEN, L. P., SCHEINKMAN, J. and TOUZI, N. (1998). Spectral methods for identifying scalar diffusions. *J. Econometrics* **86** 1–32. MR1650012
[23] HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **17** 417–441.
[24] KESSLER, M. and SORENSEN, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5** 299–314. MR1681700
[25] NELSON, E. (1958). The adjoint Markov process. *Duke Math. J.* **25** 671–690. MR0101555
[26] PAN, J. and YAO, Q. (2008). Modelling multiple time series via common factors. *Biometrika* **95** 365–379. Available at http://biomet.oxfordjournals.org/cgi/reprint/95/2/365.pdf, http://biomet.oxfordjournals.org/cgi/content/abstract/95/2/365.
[27] PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2** 559–572.
[28] RAMSAY, J. and SILVERMAN, B. W. (2005). *Functional Data Analysis*. Springer, New York. MR2168993
[29] REED, M. and SIMON, B. (1978). *Methods of Modern Mathematical Physics IV: Analysis of Operators*. Academic Press, San Diego. MR0493421
[30] RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill, New York. MR0365062
[31] SALINELLI, E. (1998). Nonlinear principal components i: Absolutely continuous variables. *Ann. Statist.* **26** 596–616. MR1626079
[32] SILVERMAN, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24** 1–24. MR1389877
[33] WONG, E. (1964). The construction of a class of stationary Markov processes. In *Stochastic Processes in Mathematical Physics and Engineering* (R. E. Bellman, ed.). *Proceedings of Symposia in Applied Mathematics* **17** 264–276. Amer. Math. Soc., Providence, RI. MR0161375
[34] ZHOU, J. and HE, X. (2008). Dimension reduction based on constrained cononical correlation and variable filtering. *Ann. Statist.* **36** 1649–1668. MR2435451
[35] ZHOU, L., HUANG, J. Z. and CARROLL, R. J. (2008). Joint modelling of paried sparse functional data using principal components. *Biometrika* **95** 601–619.

X. CHEN
COWLES FOUNDATION
 FOR RESEARCH IN ECONOMICS
DEPARTMENT OF ECONOMICS
YALE UNIVERSITY
BOX 208281
NEW HAVEN, CONNECTICUT 06520-8281
USA
E-MAIL: xiaohong.chen@yale.edu

L. P. HANSEN
DEPARTMENTS OF ECONOMICS
 AND STATISTICS
THE UNIVERSITY OF CHICAGO
1126 EAST 59TH STREET
CHICAGO, ILLINOIS 60637
USA
E-MAIL: lhansen@uchicago.edu

J. SCHEINKMAN
DEPARTMENT OF ECONOMICS
PRINCETON UNIVERSITY
26 PROSPECT AVENUE
PRINCETON, NEW JERSEY 08540-5296
USA
E-MAIL: joses@princeton.edu