Principal Components and Long-run Implications of Multivariate Diffusions*

Xiaohong Chen†, Lars Peter Hansen‡ and José Scheinkman§

Department of Economics
Yale University
e-mail: xiaohong.chen@yale.edu

Departments of Economics and Statistics
University of Chicago
e-mail: lhansen@uchicago.edu

Department of Economics
Princeton University
e-mail: joses@princeton.edu

Abstract: We investigate a method for extracting nonlinear principal components. These principal components maximize variation subject to smoothness and orthogonality constraints; but we allow for a general class of constraints and multivariate densities, including densities without compact support and even densities with algebraic tails. We provide primitive sufficient conditions for the existence of these principal components. By exploiting the theory of continuous-time, reversible Markov processes, we give a different interpretation of the principal components and the smoothness constraints. When the diffusion matrix is used to enforce smoothness, the principal components maximize long-run variation relative to the overall variation subject to orthogonality constraints. Moreover, the principal components behave as scalar autoregressions with heteroskedastic innovations; this supports semiparametric identification 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 non-reversible diffusion processes.

AMS 2000 subject classifications: Primary 62H25, 47D07; secondary 35P05.

Keywords and phrases: nonlinear principal components, multivariate diffusion, quadratic form, conditional expectations operator.

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

*We thank Anna Aslanyan, Henri Berestycki, Jaroslav Borovička, E. Brian Davies, Nan Li, Oliver Linton, Peter McCullagh, Nour Meddahi, Gigliola Staffilani, Stephen Stigler and Gauhar Turmuhambetova for useful conversations and Nan Li and Noah Williams for research assistance.

†National Science Foundation support under Award Number SES0350770 and SES0718407.
In this paper we study the extraction of nonlinear principal components using information encoded in the density of the data. Formally, the principal components maximize variation subject to orthogonality and smoothness constraints where smoothness constraints are enforced by the quadratic form \( f \) expressed in terms of the gradients of functions. Specifically, the quadratic form is

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

where \( \Sigma \) is a state-dependent positive-definite matrix, \( q \) is the population density of the data, \( \nabla \) denotes the (weak) gradient operator and \( \Omega \) is the state space.

Alternatively, principal components 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 principal components 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 principal components is in part an investigation of this approximation.

Previously Box and Tiao (1977) 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 (1998) defined nonlinear principal components for multivariate absolutely continuous random variables and characterized these principal components as eigenfunctions of a self-adjoint, differential operator. As we will show these two methods are related. We share Salinelli (1998) interest in nonlinear principal components, but our departure from his work is substantial. For Salinelli, the matrix \( \Sigma \) is state independent, the state space \( \Omega \) 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 \( \Sigma \) and \( q \), we entertain a larger class of smoothness constraints \textit{vis a vis} Salinelli (1998) with explicit links to the data generation. Establishing the existence of principal components in this setup is no longer routine.

Salinelli (1998) assumed that the data generation process is independent and identically distributed. While our analysis is applicable to such an environment, we also explore the case in which data are generated by a stationary diffusion process. By considering such processes, we make a specific choice of the matrix \( \Sigma \) used to enforce smoothness. It is the local covariance or diffusion matrix. With this choice, the principal components extracted with smoothness penalties are ordered by the ratio of their long-run variation to the overall variation as in Box and Tiao (1977). Principal components 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 (1977) to nonlinear, multivariate Markov diffusions and establish an explicit link to the method of Salinelli (1998).

In this paper we do the following:

1. Formulate the nonlinear principal component 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 principal components.
3. Provide a reversible Markov diffusion process for the data generation that supports the principal component 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 principal components as functions that maximize variation subject to orthogonality conditions and smoothness bounds given by the quadratic form $f$. Section 3 contains existence results. In section 4 we suppose the data are generated by a multivariate nonlinear diffusion and establish the connection between our principal components and the canonical analysis of Box and Tiao (1977). The results in section 5 relate the principal components to eigenfunctions of conditional expectations operators associated with a stationary Markov process $\{x_t\}$ defined using the diffusion matrix $\Sigma$ and the stationary density $q$. Given an eigenfunction $\psi$, 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 $\Sigma$. Section 7 gives some concluding remarks and discusses applications of our results. The appendices contain some proofs and computations associated with examples.

2. 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 $\Omega$ with respect to Lebesgue measure. The implied probability distribution is the population counterpart to the empirical distribution of the data. The data could be iid and in Salinelli (1998), but we are primarily interested in the case in which the data are obtained by sampling a continuous-time, stationary Markov diffusion $\{x_t\}$. 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 population probability distribution. The $L^2$ inner product (denoted $<\cdot,\cdot>$) 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 $C^2_R$, the space of twice continuously differentiable functions with
compact support in \( \Omega \), that can be parameterized in terms of the density \( q \) and a positive definite matrix \( \Sigma \) that can depend on the state:

\[
fo(\phi, \psi) = \frac{1}{2} \int_\Omega \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_i} \frac{\partial \psi}{\partial y_j} q
\]

where

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

**Assumption 2.1.** \( q \) is a positive, continuously differentiable probability density on \( \Omega \).

**Assumption 2.2.** \( \Sigma \) is a continuously differentiable, positive definite matrix function on \( \Omega \).

Assumptions 2.1 and 2.2 restrict the density \( q \) and the matrix \( \Sigma \) to be continuously differentiable. These assumptions are made for convenience. As argued by Davies (1989) (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 \( \Omega \) 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.

\[
\bar{H} = \{ \phi \in L^2 : \text{there exists } g \text{ measurable, with } \int g' \Sigma g q < \infty, \text{ and } \int \phi \nabla \psi = -\int g \psi, \text{ for all } \psi \in C^1_K \}
\]

The random vector \( g \) is unique (for each \( \phi \)) and is referred to as the weak derivative of \( \phi \).

From now on, for each \( \phi \) 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 \( \Lambda \) is the square root of \( \Sigma \). Also we use \( C^1_K \) 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, \( C^\infty_K \) (see Brezis (1983) Remark 1, page 150.) For any pair of functions \( \psi \) and \( \phi \) 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 \( \langle \phi, \psi \rangle_f = \langle \phi, \psi \rangle + f(\phi, \psi) \). With this inner product, \( \bar{H} \) is complete (and hence a Hilbert space).

Thus \( \bar{H} \) is taken to be the domain \( D(f) \) of the form \( f \). Notice, in particular, that the unit function is in \( D(f) \).
2.1. Initial construction

Principal components maximize variation subject to smoothness constraints. In our generalization these principal components are defined as follows.

**Definition 2.1.** The function $\psi_j$ is the $j^{th}$ nonlinear principal component for $j \geq 1$ if $\psi_j$ solves:

$$\max_{\phi} <\phi, \phi>$$

subject to

$$f(\phi, \phi) = 1$$
$$<\psi_s, \phi> = 0, s = 0, ..., j - 1$$

where $\psi_0$ is initialized to be the constant function one.

There are two differences between our proposed extraction and that of Salinelli (1998). First, Salinelli (1998) assumes that $\Sigma$ is state independent. To accommodate a rich class of densities, we allow $\Sigma$ to be state dependent. Second, Salinelli (1998) 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 the density $q$ is no longer assumed to be bounded from below.

Principal components are eigenfunctions of the quadratic forms $f$.

**Definition 2.2.** An eigenfunction $\psi$ of the quadratic form $f$ satisfies:

$$f(\phi, \psi) = \delta <\phi, \psi>$$

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

Since $f$ is positive semidefinite, $\delta$ must be nonnegative. The principal components extracted in the manner given in (2.1) have eigenvalues $\delta_j$ that increase with $j$. If we renormalize the eigenfunctions to have a unit second moment, the principal components 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 principal components $\{\psi_j : j = 0, 1, \ldots\}$ exist with corresponding eigenvalues $\{\delta_j : j = 0, 1, \ldots\}$. Consider any $\phi$ in $L^2$. Then

$$\phi = \sum_{j=0}^{\infty} <\psi_j, \phi> \psi_j,$$

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

$$f(\phi, \psi) = \sum_{j=0}^{\infty} \delta_j \frac{<\phi, \psi_j> <\psi, \psi_j>}{<\psi_j, \psi_j>^2}.$$
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

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

for some $\theta > 0$.

A necessary condition for $\psi$ to be a principal component is that it satisfies an eigenvalue problem:

**Claim 2.1.** A solution $\psi$ 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 $\lambda$ 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.** $\{\phi \in \mathcal{D}(f) : f(\phi, \phi) + \theta \langle \phi, \phi \rangle \leq 1\}$ 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 principal components? 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 principal components 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 principal component used as an additional approximating function.

Initially solve Problem 2.1 for $H = L^2$, select a solution $\psi_0$ and denote the maximized objective as $\lambda_0$. 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 $H^\perp_{j-1}$ 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 $H = H^\perp_{j-1}$,
select a solution $\psi_j$, and form $\lambda_j$ 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 $\phi$ onto the closed (in $L^2$) linear space $\hat{H}$. The second moment of the approximation error is:

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

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

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

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

**Claim 2.3.**

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

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 $\psi_{N-1}$ may be replaced by the sum of $\psi_{N-1}$ plus some $\psi'$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 $N$.

The following example is used to illustrate principal components.

**Example 2.1.** Let $\Sigma = I$ and

$$q(x) \propto \begin{cases} \exp \left[ - (|x| - 1)^2 \right] & \text{if } |x| \geq 1 \\ 1 & \text{if } |x| < 1 \end{cases}$$

The proportionality factor can be chosen so that the density $q$ integrates to unity. The radial symmetry is chosen for simplicity.

For the sake of illustration, we assume that $n = 2$. Some principal components are of the form:

$$\psi(x) = \vartheta(|x|),$$

In addition, there are principal components that are not constant on circles. These principal components come in pairs. The principal component extraction will identify a two-dimensional space rather than the more familiar one-dimensional extraction. In particular there will be two orthogonal principal components $\psi$ and $\psi^*$ with the same smoothness and the same variance. One of
these, say $\psi$ will be symmetric: $\psi(x_1, x_2) = \psi(x_2, x_1)$ and another will be anti-symmetric: $\psi^*(x_1, x_2) = -\psi^*(x_2, x_1)$.

We display the first five principal components in the accompanying figures. The computational method is described in Appendix A. These principal components are scaled to have unit variance and are ordered by their smoothness. Principal components one and two come in a symmetric-antisymmetric pair and are shown in Figures 1 and 2. These functions are almost linear inside a circle of radius two. Beyond this circle the slope increases.

Principal components three and four also come in a symmetric-antisymmetric pair and are reported in Figures 3 and 4. Since the principal components are ordered by their smoothness, these functions oscillate more than the first pair. The fifth principal component is constant on circles and is depicted in Figure 5.

2.4. Principal components revisited

In Problem 2.1, the constraint set gets larger as $\theta$ declines to zero. Reducing the smoothness penalty with a smaller $\theta$ enlarges the collection of functions that satisfy the constraint. Thus the maximized objective increases as $\theta$ is reduced. While this is true, it turns out the maximizing choice of $\phi$ does not depend on $\theta$ up to scale. This follows because the ranking over $\phi$’s implied by the ratio:

$$\frac{<\phi, \phi>}{\theta <\phi, \phi> + f(\phi, \phi)}$$

does not depend on the value of $\theta$. The same ranking is also implied by the ratio:

$$\frac{<\phi, \phi>}{f(\phi, \phi)}$$

provided that $H$ is orthogonal to all constant functions. Thus a scaled solution $\psi$ to Problem 2.1 also solves:

**Problem 2.2.**

$$\max_{\phi \in H} \quad <\phi, \phi>$$

subject to:

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

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

From Claim 2.1 we know that $\psi$ satisfies:

$$<\phi, \psi> = \lambda [\theta <\phi, \psi> + f(\phi, \psi)]$$

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

$$f(\phi, \psi) = \delta <\phi, \psi>$$
where
\[ \delta = 1 - \theta \lambda \frac{1}{\lambda}. \]
This is the eigenvalue associated with the principal component extraction. Solving for \( \lambda \),
\[ \lambda = \frac{1}{\theta + \delta}. \]
Since eigenvalues \( \delta \) of the form increase without bound, the corresponding sequence of \( \lambda \)'s converge to zero guaranteeing that approximation becomes arbitrarily accurate as the number of principal components 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 principal components. We allow for noncompact state spaces and provide alternative restrictions on the tail behavior of the the density \( q \) and the penalization matrix \( \Sigma \) that guarantee that the compactness criterion 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 \( \Sigma \) 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 \( \Sigma \) 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 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 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 (1998) established the existence of eigenfunctions by applying Rellich’s compact embedding theorem when the domain \( \Omega \) is compact with a continuous boundary. This approach requires a density that is bounded and bounded away from zero and a derivative penalty matrix \( \Sigma \) that is uniformly nonsingular.

### 3.2. Real Line

Perhaps surprisingly, the principal component extraction is nontrivial even for densities on the real line. This is because our are principal components can be nonlinear functions of the underlying random variable that is replicated over time. This we initially consider the case in which the state space is the reals, and write \( \Sigma = \varsigma^2 \). This gives us a particularly simple tradeoff between the tail density and the penalization function \( \varsigma \).
Proposition 3.1. Suppose

$$\int_0^\infty \frac{1}{\varsigma^2(x)q(x)} = +\infty, \int_0^\infty \frac{1}{\varsigma^2(-x)q(-x)} = +\infty$$  \hspace{1cm} (3.1)

$$\lim_{|x| \to \infty} -\frac{x}{|x|} \left[ \frac{q'(x)}{q(x)} + \varsigma'(x) \right] = +\infty.$$  \hspace{1cm} (3.2)

Then Condition 2.1 is satisfied.

When $\varsigma$ 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 $\varsigma$ 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 $\Sigma$. 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 $\mathcal{C}_0 \subset \mathcal{D}(f)$ is a core of $f$ if for any $\phi_0$ in the domain $\mathcal{D}(f)$, there exists a sequence $\{\phi_j\}$ in $\mathcal{C}_0$ such that

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

**Condition 3.1.** $C^2_K$ is a core of $f$.

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

Although their purpose was different, Fukushima et al. (1994) provide a convenient sufficient condition that implies Condition 3.1 in environments that interest us. Define:

$$\kappa(r) = \int_{|x| = 1} x'\Sigma(rx)xq(rx)dS(x)$$
where $dS$ is the measure (surface element) used for integration on the sphere $|x| = 1$. For functions $\psi$ and $\phi$ in $C^2_\mathbb{K}$ that are radially symmetric, i.e. $\phi(x) = \xi(|x|)$ and $\psi(x) = \zeta(|x|)$, we may depict the form $f_o$ as an integral over radii:

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

**Proposition 3.2.** Condition 3.1 is implied by:

$$\int_1^\infty \kappa(r)^{-1} r^{1-n} dr = \infty. \quad (3.3)$$

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

$$\min \left\{ \frac{1}{r_1}, \frac{1}{r_2} \right\} \geq \frac{1}{r_1 + r_2}.$$ 

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

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

Thus a sufficient condition for (3.3) is that

$$\int_1^\infty \frac{\kappa(r)}{r^2} r^{n-1} dr < \infty. \quad (3.4)$$

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

$$\varsigma^2(r) = \sup_{|x|=1} x^\top \Sigma(rx)x,$$

and

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

Notice that

$$\kappa(r) \leq \varsigma^2(r) \varrho(r).$$

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

$$\int_0^\infty \varrho(r) r^{n-1} dr = 1.$$
We may extend the previous argument by supposing instead that
\[ \varsigma^2(r) \leq c|r|^{2+2\delta} \]
for some positive \( \delta \). Then
\[
\frac{\kappa(r)}{r^2} \leq c r^{2\delta} \int_{|x|=1} q(rx) dS(x).
\]
Thus (3.4) is satisfied provided that
\[
\int |x|^{2\delta} q(x) dx < \infty.
\]
Hence we can allow for faster growth in \( \varsigma^2 \) if \( q \) has high enough moments.

So far we have produced a sufficient condition for approximation using functions in \( C^2_K \) (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 (1989)), 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,
\[
U_\theta = \{ \phi \in D(f) : f(\phi, \phi) + \theta < \phi, \phi > \leq 1 \}
\]
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 \( U(U_\theta) \) is pre-compact. We will actually construct a set that contains \( U(U_\theta) \) and is pre-compact in \( L^2(\text{leb}) \).

First notice that \( U \) and \( U^{-1} \) leave \( C^2_K \) invariant, and for any \( \psi \in C^2_K \) 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 \( \psi \in C^2_K \), it follows that
\[ \int (\nabla h)' \Sigma [\nabla (\psi \psi^*)] = -\sum_{i,j} \sigma_{i,j} \frac{\partial^2 h}{\partial y_i \partial y_j} \psi \psi^* - \sum_{i,j} \frac{\partial \sigma_{i,j}}{\partial y_i} \frac{\partial h}{\partial y_j} \psi \psi^*. \]

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

where the potential function \( V \) is given by:
\[ V = -\sum_{i,j} \sigma_{i,j} \frac{\partial^2 h}{\partial y_i \partial y_j} - \sum_{i,j} \frac{\partial \sigma_{i,j}}{\partial y_i} \frac{\partial h}{\partial y_j} + (\nabla h)' \Sigma (\nabla h). \]  
(3.6)

**Proposition 3.3.** Suppose that \( C^2_K \) is a core for \( f \), \( \psi = U \phi \) for some \( \phi \in \overline{H} \) and \( V \) is bounded from below. Then \( \psi \) is weakly differentiable,
\[ \nabla \psi = \exp(-h)(-\phi \nabla h + \nabla \phi) \]
and
\[ \frac{1}{2} \int (\nabla \phi)' \Sigma \nabla \phi \psi^2 = \frac{1}{2} \int (\nabla \psi)' \Sigma (\nabla \psi) + \frac{1}{2} \int V \psi^2. \]  
(3.7)

A consequence of this proposition is that
\[ V_\theta = \{ \psi \in L^2(\text{leb}) : \int \left( \theta + \frac{1}{2} V \right) \psi^2 + \frac{1}{2} \int (\nabla \psi)' \Sigma (\nabla \psi) \leq 1 \} \supset U(U_\theta), \]
and it thus suffices to show that \( V_\theta \) 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 \( \Sigma \) 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 \( V_\theta \) we impose:

**Assumption 3.2.** The derivative penalty matrix \( \Sigma \geq \varsigma I \) for some \( \varsigma > 0 \).
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 \( \Sigma \) by a lower bound. Given Assumption 3.2 we can always construct a twice continuously differentiable function \( \varsigma(x) \) with

\[
\Sigma(x) \geq \varsigma(x)^2 I. \tag{3.8}
\]

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

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

on the space \( C^2_K \). Then

\[
\hat{f}_o(\phi, \phi) \leq f_o(\phi, \phi).
\]

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

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

in place of \( V \) in demonstrating compactness.

**Criterion 3.2.** Equation (3.8) is satisfied and

\[
\lim_{|x| \to \infty} \hat{V}(x) = +\infty.
\]

To derive some sufficient conditions for this criterion we parameterize \( \varsigma(x) \) as:

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

Then an alternative formula for \( \hat{V} \) is:

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

An alternative to Criterion 3.2 is:

**Criterion 3.3.** Equation (3.8) is satisfied with \( \varsigma(x) = \exp[v(x)] \) and,

a) \[
\lim_{|x| \to \infty} \frac{|\nabla v(x)|}{|\nabla h(x)|} = 0;
\]
Proposition 3.5. Suppose Assumptions 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 $\varsigma$ 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

$$\lim_{|x| \to \infty} |\nabla v(x)| = 0.$$ 

Write

$$\int \varsigma^2 \nabla \phi \cdot \nabla \phi = \varsigma \int \nabla \phi \cdot \nabla \phi + \int (\varsigma^2 - \varsigma) \nabla \phi \cdot \nabla \phi.$$ 

We now deduce a convenient lower bound on:

$$\int (\varsigma^2 - \varsigma) \nabla \phi \cdot \nabla \phi,$$

following an approach of Davies (1989) (see Theorem 1.5.12). Construct an additional potential function:

$$\tilde{W}(x) = (\varsigma^2 + \varsigma)(\nabla v \cdot \nabla v) + (\varsigma^2 - \varsigma) \text{trace} \left( \frac{\partial^2 v}{\partial x_i \partial x_j} \right).$$

Lemma 3.1. If equation (3.8) holds,

$$\int \tilde{W} \phi^2 \leq \int (\varsigma^2 - \varsigma) \nabla \phi \cdot \nabla \phi$$

for all $\phi \in C^2_K$.

Note that

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

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

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

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

**Proposition 3.6.** Suppose Assumptions 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 \( \Sigma \) limits the functions that we hope to approximate using principal components. Also for \( C^2_k \) to be a core for the form \( f \) we require limits on growth in \( \Sigma \) (see subsection 3.3.1.)

Our use of \( \tilde{W} \) in addition to \( \tilde{V} \) in effect replaces \( -\zeta^2 |\nabla v|^2 \) with a second derivative term:

\[ \zeta(x)^2 \text{trace} \left[ \frac{\partial^2 v(x)}{\partial x_i \partial x_j} \right]. \]

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 \( \tilde{c} = \log \xi. \) Thus \( \zeta \) 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

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

Notice that both terms converge to zero as \( |x| \) gets large, but that the squared gradient scaled by \( \zeta^2 \) 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 \( \beta > 1. \)

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 construction of principal components. The first one, Criterion 3.3 gives the most flexibility in terms of the penalization matrix \( \Sigma \); 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

\[ 1 \text{ We have previously established an alternative compactness criterion for } n = 1 \text{ that does not involve second derivatives that may be preferred to Criterion 3.4.} \]
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 $C^2_K$ ceases to be a core for the form used in the eigenfunction extraction. Thus our results feature a tradeoff between how thick the tail is in the density of the data and how much penalization is required in order for the resolvent to be compact.

4. Forms and Markov processes

So far we considered the role of the penalization matrix $\Sigma$ in the construction and approximation properties of principal components. We now use stationary Markov diffusions to give an explicit interpretation of this penalization matrix.

We proceed as follows. Suppose the data are generated by a Markov diffusion by sampling say at integer points in time. Specifically, $\{x_t\}$ solves

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

for some n-dimensional vector function $\mu$ and some $n \times n$ matrix function $\Lambda$ 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 $\mu$ that are implicit in such a construction. Let $\phi$ be in $C^2_K$. Then it follows from Itô’s lemma that 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 measure of magnitude of the instantaneous forecast error in forecasting $\{\phi(x_t)\}$ over the next instant given the current Markov state.

The principal component extraction given by Definition 2.1 performed equivalently as:

**Definition 4.1.** The function $\psi_j$ is the $j^{th}$ nonlinear principal component for $j \geq 0$ if $\psi_j$ solves:

$$\min_{\phi} f(\phi, \phi)$$

subject to

$$< \phi, \phi > = 1$$

$$< \psi_s, \phi > = 0, s = 0, ..., j - 1.$$
(1977) 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 \( f(\psi_j, \psi_k) \) the (on 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 non-compact state space of the Markov process. Thus imposing uniform bounds on both \( q \) and the matrix \( \Sigma \) over compact state spaces is too restrictive. The existence results in Section 3 avoid such restrictions.


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 \( \Sigma \) as its local covariance matrix and \( q \) as its 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 \( f_o \), which we construct using integration-by-parts. For any pair of functions \( \phi \) and \( \psi \) in \( C^2_K \):

\[
f_o(\phi, \psi) = \frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_i} \frac{\partial \psi}{\partial y_j} q = -\frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} \psi q = -\frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_i} \frac{\partial \psi}{\partial y_j} q \quad (5.1)
\]

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

\[
\int \sum_{i,j} \frac{\partial (q \sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j} \psi = -\int \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} \psi q - \int \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_i} \frac{\partial \psi}{\partial y_j} q.
\]
We use (5.1) to motivate our interest in the differential operator \( F_o \):

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

(5.2)

This operator is constructed so that the form \( f_o \) can be represented as:

\[
f_o(\phi, \psi) = <F_o \phi, \psi> = <\phi, F_o \psi>.
\]

where the second relation holds because we can interchange the role of \( \phi \) and \( \psi \) 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 \( \{x_t\} \) 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 Ito’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 \textit{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 \( \Sigma \) 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 (1964) 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 (1978) and Cobb et al. (1983) extended this analysis in part by taking as given the diffusion coefficient and constructing the implied drift coefficient for
the stochastic differential equation the generates a prescribed stationary density. Banon (1978) and Cobb et al. (1983) 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 (1978)’s aim for generality, but at the same time we retain Wong (1964)’s interest in spectral decompositions.

As in Wong (1964); Banon (1978); Cobb et al. (1983), we parameterize diffusion processes using the stationary density \( q \) and a (possibly state dependent) diffusion coefficient \( \Sigma \) in contrast to the more typical approach of starting with a drift and the diffusion coefficients. In contrast to Wong (1964); Banon (1978); Cobb et al. (1983), we allow the diffusion process to be multivariate on a state space \( \Omega \). 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 \( (q, \Sigma) \). Associated with the closed extension \( f \) is a family of resolvent operators \( G_\alpha \) indexed by a positive parameter \( \alpha \). 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 \( G_\alpha \) is constructed as follows. Given a function \( \phi \in L^2 \), define \( G_\alpha \phi \in D(f) \) to be the solution to

\[
  f(G_\alpha \phi, \psi) + \alpha < G_\alpha \phi, \psi >= < \phi, \psi >
\]

for all \( \psi \in D(f) \). 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 et al. (1994) pages 15 and 19). In particular, \( G_\alpha \) 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:

\[
  F\phi = (G_\alpha)^{-1}\phi - \alpha \phi
\]

defined on the domain \( G_\alpha(L^2) \). It can be shown that \( F \) is independent of \( \alpha \). 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 \( D(f) \) and: \( f(\phi, \psi) = < \sqrt{F}\phi, \sqrt{F} \psi > \) (e.g. see Fukushima et al. (1994) Theorem 1.3.1). Moreover, it is an extension of the operator \( F_\alpha \) because \( f \) is an extension of \( f_\alpha \) (e.g. see Lemma 3.3.1 of Fukushima et al. (1994)).

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 \( \alpha \) and any \( t \), we may form the exponential:

\[
  \exp(t\alpha^2G_\alpha - \alpha t) I
\]
as a Neumann series expansion. Notice that (5.4) implies
\[ t\alpha^2 G_\alpha - t\alpha I = t\alpha[(I + \frac{1}{\alpha}F)^{-1} - I] \]
\[ = -tF \left( I + \frac{1}{\alpha}F \right)^{-1}. \]

Instead directly using a series expansion, we use the limit
\[ \lim_{\alpha \to \infty} \exp[(t\alpha^2 G_\alpha) - \alpha tI] = \exp(-tF) \]
only 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:
\[ G_\alpha \phi = \int_0^\infty \exp(-\alpha t) \exp(-tF)\phi dt \quad (5.5) \]
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 submarkov in the language of Beurling and Deny (1958). 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 submarkov 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 submarkov semigroup (e.g., see Davies (1989) section 1.3).

**Condition 5.1.** (Beurling-Deny) For any \( \phi \in D(f) \), \( \psi \) given by the truncation:
\[ \psi = (0 \lor \phi) \land 1 \]
is in \( D(f) \) and
\[ f(\psi, \psi) \leq f(\phi, \phi). \]
When this condition is satisfied, the semigroup \( \exp(-tF) \) is submarkov, 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 (1989) 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 $F_0$ and generates a semigroup $\{\exp(-tF) : t \geq 0\}$. The density $q$ is the stationary density for this diffusion, the matrix $\Sigma$ is the diffusion matrix and $\exp(-tF)$ is the conditional expectation operator over an interval of time $t$.

Proof. The form $f$ satisfies the Beurling-Deny criteria (Davies (1989) Theorem 1.3.5). Thus there exists a self-adjoint operator $F$ which is an extension of $F_0$ and generates a submarkov semigroup $\exp(-tF)$. Theorem 7.2.1 of Fukushima et al. (1994) 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 $\phi \in D(f)$. As a consequence, the unit function is also in the domain of the operator $F$.

$F1 = 0.$

We illustrate the Markov process construction by reconsidering Example 2.1. The Markov process associated with this form has the identity as the diffusion matrix $\Sigma$. The drift can be inferred from formula (5.2):

$$
\mu = -\frac{1}{2} \nabla \log q = \begin{cases} 
-(|y|-1) \frac{\bar{w}}{|y|} & \text{if } |y| \geq 1 \\
0 & \text{if } |y| < 1.
\end{cases}
$$

This construction results in a continuous-time extension of the familiar threshold autoregression model. See Tong (1990) and Fan and Yao (2005) for discussions of discrete-time threshold autoregressive models. For analyses of other continuous-time threshold models see Stramer et al. (1996a) and Stramer et al. (1996b).

5.3. 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 $G_\alpha$ and of the generator $F$. For convenience, we rewrite equation (5.3):

$$
f(G_\alpha \phi, \psi) + \alpha < G_\alpha \phi, \psi >= < \phi, \psi >.
$$

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

$$
\lambda = \frac{1}{\delta + \alpha}
$$
where $\lambda$ is the eigenvalue of $G_\alpha$ and $\delta$ is the corresponding eigenvalue of $f$.

Given the relation between the generator $F$ and the resolvent operator $G_\alpha$,

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

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

An eigenfunction $\psi$ of the generator $F$ satisfies:

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

for some positive number $\delta$ and each transition interval $s$. Thus the principal components described previously will also satisfy the testable conditional moment implications (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 can be depicted using a principal component decomposition as in (2.3), analogous decompositions are applicable to $F$ and $\exp(-tF)$:

$$F \phi = \sum_j \delta_j \frac{<\phi, \psi_j>}{<\psi_j, \psi_j>} \psi_j,$$

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

where the first expansion is only a valid $L^2$ series when $\phi$ is in the domain of the operator $F$. When the eigenvalues of the form increase rapidly, the term $\exp(-t\delta_j)$ will decline to zero, 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 principal components. On the other hand, slow eigenvalue divergence of the form will make it challenging to approximate the transition operators with a small number of principal components. Our results in Section 3 give primitive conditions based on the behavior of stationary density and diffusion matrices for the existence of eigenfunction decompositions and eigenvalue decay rates.

### 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

\[ g(\phi, \psi) = 2 \lim_{\alpha \downarrow 0} < G_\alpha \phi, \psi > \]

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

\[ \lim_{\alpha \downarrow 0} < G_\alpha \phi, \phi > < \infty. \]

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 equation (5.5)

\[ < G_\alpha \phi, \psi > = \int_0^\infty \exp(-\alpha t) E[\phi(x_t)\psi(x_0)] dt. \quad (5.7) \]

Hence, using (5.4), we obtain:

\[ g(\phi, \psi) = \lim_{\alpha \downarrow 0} 2 < G_\alpha \phi, \psi > = \lim_{\alpha \downarrow 0} 2 < (\alpha I + F)^{-1} \phi, \psi >. \]

Notice that this form is symmetric because the resolvent operator is self-adjoint for any positive \( \alpha \). 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 \( j^{th} \) principal component \( \psi_j \) for \( j \geq 1 \) solves:

\[ \max_{\phi} g(\phi, \phi) \]

subject to

\[ < \phi, \phi > = 1 \]

\[ < \psi_s, \phi > = 0, s = 0, ..., j - 1 \]

where \( \psi_0 \) is initialized to be the constant function one.

Recall that the spectral density function at frequency \( \theta \) 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 \( \psi \) 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 (1982) and Hansen and Scheinkman (1995), 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 \( T \) becomes large.

This gives us an alternative interpretation of our extended version of the principal component extraction of Salinelli (1998). We may base the extraction on maximizing \( g(\phi, \phi) \) subject to \( <\phi, \phi> = 1 \) over recursively constructed spaces \( H \). 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 \( \Sigma \), we have seen how to construct a reversible diffusion, but typically there are other diffusions that share the same density and diffusion matrix. We now characterize the drifts of such processes.

Instead of constructing a Markov process implied by a form, suppose instead 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 (1958) and Hansen and Scheinkman (1995) 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 \( C^2 \), 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
\[- \int \psi(A\phi)q = f_o(\phi, \psi)\]
on $C^2_K$.
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 $F$ that we 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 (1958) that the adjoint operator has the same diffusion matrix, but a different drift vector $\mu$. The drift for the adjoint operator $A^*$ is given by:
\[\mu^* = -\mu + \frac{1}{q} \sum_{i,j} \partial(q\sigma_{ij}) \frac{\partial \phi}{\partial y_i} \frac{\partial \phi}{\partial y_j}\]
The adjoint operator generates the semigroup of expectation operators for the reverse time diffusion. From the formula for reverse time drift, $\mu^*$, it follows that
\[\frac{\mu + \mu^*}{2} = \frac{1}{2q} \sum_{i,j} \partial(q\sigma_{ij}) \frac{\partial \phi}{\partial y_i} \frac{\partial \phi}{\partial y_j},\]
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, $\mu$ and $\mu^*$, and the common diffusion matrix, $\Sigma$.

**Remark 6.1.** The density $q$ and the diffusion matrix $\Sigma$ do place other restrictions on the drift vector $\mu$. Since $q$ is the stationary density, $\mu$ and $\mu^*$ must also satisfy:
\[
\frac{\partial (\mu q)}{\partial y} = \frac{\partial}{\partial y} \sum_{i,j} \partial(q\sigma_{ij}) \frac{\partial \phi}{\partial y_i} \frac{\partial \phi}{\partial y_j},
\]
\[
\frac{\partial (\mu^* q)}{\partial y} = \frac{\partial}{\partial y} \sum_{i,j} \partial(q\sigma_{ij}) \frac{\partial \phi}{\partial y_i} \frac{\partial \phi}{\partial y_j}.
\]
While there is typically one solution $\mu$ (or $\mu^*$) 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 matrices; but the average of the forward and backward drift can be inferred.
Remark 6.2. The principal component 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” equation 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 (1973), Theorem 4.25, page 108.)

Given $\alpha$ the resolvent operator will be compact provided that the image of $R_\alpha$ of the $L^2$ unit ball has compact closure. Consider a function $\varphi$ given by

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

Then $\varphi \in D(A)$ and

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

Thus it suffices to show that

$$\{ \varphi \in D(A) : \alpha^2 < \varphi, \varphi > + 2\alpha f(\varphi, \varphi) \leq 1 \}$$

has compact closure. This set will have compact closure, if, and only if, compactness Condition 2.1 is satisfied for

$$\theta = \frac{2}{\alpha}.$$ 

7. Conclusions and related literature

We have studied principal components 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. Meddahi (2001), Andersen et al. (2004) and Corradi and Distaso (2006) exploit the approximation properties of principal components of the type we characterized here in volatility modeling.\(^2\) Relatedly, given their ability to capture variation in multivariate data, Bontemps and Meddahi (2005) use the implied principal components from parametric densities to devise tests of multivariate distributional assumptions.

We also used multivariate Markov diffusions as data generating devices to interpret principal components. These principal components are eigenfunctions

\(^2\)These papers make reference to an earlier version of our paper.
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 (1995) that reversible diffusions can be identified nonparametrically from discrete-time low frequency stationary observations.

For more general diffusions, these principal components are orthogonal and have orthogonal innovations analogous to those from the canonical analysis of Box and Tiao (1977) for linear multiple time series models. (See also Pan and Yao (2007).) Thus our principal component construction provides a convenient way to summarize implications of multivariate nonlinear diffusion models. 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 based on low frequency data. Our characterization of principal components 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 has been suggested previously by Demoura (1998), Hansen and Scheinkman (1995), Kessler and Sorensen (1999), Hansen et al. (1998), Florens et al. (1998), Chen et al. (1998) and Gobet et al. (2004). In particular, Kessler and Sorensen (1999) use eigenfunctions to construct quasi-optimal estimators of parametric scalar models of the drift and diffusion coefficients from discrete-time data in the special case in which the functional forms of eigenfunctions are known \textit{a priori}. Hansen and Scheinkman (1995), Hansen et al. (1998), Chen et al. (1998), Gobet et al. (2004) and Darolles et al. (2004) study semiparametric and nonparametric identification and over-identification based on a principal component extraction that is closely related to the one analyzed here; see Fan (2005) for a recent review. This previous literature focuses primarily on scalar diffusion models and in some cases to 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 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., Ramsay and Silverman (2005), Silverman (1996), Hall et al. (2006), Kneip and Utikal (2001) and Benko et al. (2007). These existing results can in principle be modified to establish asymptotic properties of estimated principal components from discrete-time low frequency realizations of an underlying multivariate Markov diffusion model.
Appendix A: Computation and figures

For convenience in our numerical calculations, we transform the state space using polar coordinates: $x' = [r \cos(\omega), r \sin(\omega)]$ for $\omega$ in $(-\pi, \pi]$ and $r \geq 0$.

Consider the quadratic form in the level. For a given $\phi$, define $\xi$ as

$$\xi(r, \omega) = \phi[r \cos(\omega), r \sin(\omega)],$$

and define $\xi^*$ analogously from $\phi^*$. Then

$$\int_{\mathbb{R}^2} \phi \phi^* q = \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^{\pi} \xi(r, \omega)\xi^*(r, \omega) d\omega q^*(r) dr \quad (A.1)$$

where

$$q^*(r) \propto \begin{cases} r \exp[-(r-1)^2] & \text{if } r \geq 1 \\ r & \text{if } r < 1 \end{cases}.$$

Consider next the quadratic form for the derivatives. Note that

$$\nabla \phi = \frac{1}{r} \begin{bmatrix} r \cos(\omega) & -\sin(\omega) \\ r \sin(\omega) & \cos(\omega) \end{bmatrix} \begin{bmatrix} \xi_r \\ \xi_\omega \end{bmatrix}.$$
Fig 2. This figure displays the anti-symmetric principal component in the first symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the other slice fixes the second coordinate at zero. The value of the principal component is given on the vertical axis. The lower-right panel reports level curves of the principal component holding fixed the value of the principal component at different levels.
Fig 3. This figure displays the symmetric principal component in the second symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the other slice fixes the second coordinate at zero. The value of the principal component is given on the vertical axis. The lower-right panel reports level curves of the principal component holding fixed the value of the principal component at different levels.
Fig 4. This figure displays the anti-symmetric principal component in the second symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero and the other slice fixes the second coordinate at zero. The value of the principal component is given on the vertical axis. The lower-right panel reports level curves of the principal component holding fixed the value of the principal component at different levels.
Fig 5. This figure displays the fifth principal component. This principal component is constant on circles centered at zero. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the other slice fixes the second coordinate at zero. The value of the principal component is given on the vertical axis. The lower-right panel reports level curves of the principal component holding fixed the value of the principal component at different levels.
Thus
\[(\nabla \phi) \cdot (\nabla \phi^*) = (\xi, \xi^*) + \frac{1}{r^2} (\xi, \xi^*_0).\]

We may evaluate the form:
\[
\int_{\mathbb{R}^2} (\nabla \phi) \cdot (\nabla \phi^*) q = \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi \xi_r(r,\omega)\xi^*_r(r,\omega)dwq^*(r)dr + \frac{1}{2\pi} \int_0^\infty \frac{1}{r^2} \int_{-\pi}^\pi \xi_\omega(r,\omega)\xi^*_\omega(r,\omega)dwq^*(r)dr. \quad (A.2)
\]

In our calculations we use basis functions of the form: \(p(r)\cos(k\omega)\) and \(p(r)\sin(k\omega)\) where \(p\) is a scalar Hermite polynomial in \(r\) and \(k\) is a nonnegative integer. We exploit the orthogonality of \(\cos(k\omega)\) and \(\sin(k\omega)\) for a given \(k\) and the orthogonality of \(\cos(k\omega)\) with \(\sin(\ell\omega)\) and \(\cos(\ell\omega)\) for \(\ell\) different from \(\ell\) all with respect to the uniform distribution on \([-\pi, \pi]\). This orthogonality allows us to separate the problem in two ways, by choice of \(k\) and by choice of cosine or sine for a given \(k\).

With this separation in mind, consider two functions: \(\xi(r,\omega) = p(r)\cos(k\omega)\) and \(\xi^*(r,\omega) = p^*(r)\cos(k\omega)\) for some positive integer \(k\). Recall that
\[
\frac{1}{2\pi} \int_{-\pi}^\pi \cos(k\omega)^2d\omega = \frac{1}{2\pi} \int_{-\pi}^\pi \sin(k\omega)^2d\omega = \frac{1}{2}.
\]

Thus the form in (A.1) is:
\[
\frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi \xi(r,\omega)\xi^*(r,\omega)dwq^*(r)dr = \frac{1}{2} \int_0^\infty p(r)p^*(r)q^*(r)dr,
\]
and the form in (A.2) is:
\[
\frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi \xi_r(r,\omega)\xi^*_r(r,\omega)dwq^*(r)dr + \frac{1}{2\pi} \int_0^\infty \frac{1}{r^2} \int_{-\pi}^\pi \xi_\omega(r,\omega)\xi^*_\omega(r,\omega)dwq^*(r)dr = \frac{1}{2} \int_0^\infty \left[\frac{k^2}{r^2}p(r)p^*(r) + p'(r)p^*'(r)\right]q^*(r)dr.
\]

For computational purposes we may use these two forms in \(p\) and solve scalar problems. Notice that the second form depends on \(k\). The \(k = 0\) problem gives rise to the principal components that are constant on circles. For \(k \geq 1\) we may compute principal components of the form \(p(r)\cos(k\omega)\) and \(p(r)\sin(k\omega)\).

The sum of the two will be symmetric and the difference will be anti-symmetric when converted to the original coordinates.

To solve the principal component problem numerically, we selected a finite-dimensional family of basis functions, evaluated two quadratic forms using numerical integration, and solved a generalized eigenvector problem.

1. Basis functions. We used as basis functions Hermite polynomials constructed to be orthogonal relative to the density: \(\exp(-y^2)\).
2. Numerical integration. We performed numerical integration over $r$ using Monte Carlo sampling the implied density $q^*$ for $r = |x|$.

3. Generalized eigenvectors. The previous two steps resulted in the construction of two positive semidefinite matrices. One for the form $\int \phi \psi q$ and the other for the form $\frac{1}{2} \int (\nabla \phi) \cdot (\nabla \psi) q$. Call the first matrix $V$ and the second matrix $W$. We factored $V = A'A$ using a Cholesky decomposition, and computed the spectral (eigenvalue-eigenvector) decomposition of $A^{-1}WA^{-1}$ using the Schur decomposition to construct the principal components.

Appendix B: Proofs

Proof of Claim 2.2. In solving the maximization component of the problem, first limit the $\phi$’s to be in $H_N$ but orthogonal to $\hat{H}$. This can only reduce maximized value. The space of such $\phi$’s contains more than just the zero element because $H_N$ has $N+1$ dimensions. Write $\phi$ as:

$$\phi = \sum_{j=0}^{N} r_j \psi_j.$$

The objective can be expressed as:

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

since $\text{Proj}(\phi|\hat{H}) = 0$. The constraint set implies that

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

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

Proof of Claim 2.3. Write $\phi$ as:

$$\phi = \text{Proj}(\phi|H_{N-1}) + \varphi$$

where $\varphi$ is in $H^\perp_{N-1}$. Write:

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

Using this decomposition, the objective can be written as:

$$\langle \varphi, \varphi \rangle,$$
and the constraint set can be written as:

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

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

References
