Principal Components and the Long Run

Xiaohong Chen
Department of Economics
London School of Economics
email: X.Chen1@lse.ac.uk

Lars Peter Hansen
Department of Economics
University of Chicago
email: l-hansen@uchicago.edu

José A. Scheinkman
Department of Economics
Princeton University
email: joses@princeton.edu *

October 4, 2000

Abstract

In this paper we suggest a method for extracting nonlinear principal components from the stationary distribution of a multivariate reversible diffusion process. These principal components a) maximize variation subject to smoothness and orthogonality constraints; and b) maximize long-run variation subject to overall variation and orthogonality constraints. Moreover, the principal components behave as scalar autoregressions with heteroskedastic innovations. This link between the stationary distribution, the long run dynamics and the transient dynamics supports parametric and semiparametric identification of a diffusion process and tests of the overidentifying restrictions implied by such a process. We provide sufficient conditions for the existence of principal components for diffusion processes with unbounded supports, and we study the limiting behavior of the corresponding eigenvalues.

1 Introduction

In this paper we develop a method for extracting nonlinear principal components from the stationary distribution of a multivariate, reversible diffusion process. These principal components are nonlinear functions of the Markov state and can be used to estimate and test models of Markov processes.

*We thank Anna Aslanyan, Henri Beresticky, E. Brian Davies, Nan Li, Gigliola Staffilani and Gauhar Turmuhambetova for useful conversations and Nan Li and Noah Williams for research assistance. Hansen and Scheinkman both acknowledge support from the National Science Foundation.

Our investigation of nonlinear principal components is facilitated by our use of continuous-time Markov models. Such models can be specified in a variety of ways. In economics and finance, researchers frequently describe a time series evolution by positing a stochastic differential equation with appropriate boundary conditions. This stochastic differential equation includes a drift (local mean or dt) contribution and a contribution from a weighting of a multivariate standard Brownian motion increment (dW_t). The matrix used in this weighting of the Brownian increment is a square root of the diffusion (local covariance) matrix.

Alternatively, a Markov process is depicted as a family (semigroup) of conditional expectation operators. The process may be modeled by the (infinitesimal) generator of this family of operators.[1] In the case of a multivariate diffusion, this generator is a second-order differential operator. First derivatives are weighted by the drift coefficient and second derivatives by the diffusion matrix.

We adopt a third strategy pioneered by Beurling and Deny (1958) and Fukushima (1971). As we will show, this approach is particularly valuable when constructing principal components of nonlinear time series. A symmetric Markov process is depicted by a positive semidefinite quadratic form defined on space of square integrable functions with respect to a prespecified measure with density q. The quadratic form is

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

where Σ is a state-dependent positive-definite matrix and ∇ denotes the (weak) gradient operator. Thus we parameterize a symmetric Markov diffusion in terms of a density q and a positive semidefinite weighting matrix Σ. The parameterized process is stationary when the density q is restricted to be integrable. Among other virtues, this approach allows us to extend the use of diffusion models of scalar stationary densities (suggested by Banon (1978) and Cobb, Kopstein, and Chen (1983)) to multivariate densities. The positive definite weighting matrix Σ in the Beurling and Deny (1958)/Fukushima (1971) parameterization is the diffusion (local covariance) matrix. A multivariate extension of a formula applied by Banon (1978) gives the drift (local mean) implied by the stationary density and the diffusion matrix.

Previously Salinelli (1998) defined nonlinear principal components for absolutely continuous random variables and characterized these principal components as eigenfunctions of a self-adjoint, differential operator. His principal component extraction uses information encoded in the density of the data. The principal components are chosen to maximize variation subject to orthogonality and smoothness constraints. Smoothness constraints are enforced by the quadratic form f expressed in terms of the gradients of functions.

Our departure from Salinelli (1998) is substantial in that he assumes the data generation is independent and identically distributed, while we study the case in which data are generated by stationary Markov process. To reinterpret the operators associated with Salinelli's principal component extraction, we are led naturally to a more general class of smoothness

[1] See Hansen and Scheinkman (1995) and Conley, Hansen, Luttmer, and Scheinkman (1998) for econometric applications of this approach.

penalties. For Salinelli, the matrix Σ is state independent, the state space is compact and the density q is bounded above and below. (For simplicity, q is set to constant for much of his analysis.) By allowing for a more general specification for Σ and q, we entertain a larger class of smoothness penalties *vis a vis* Salinelli (1998) with explicit links to the data generation. We use the principal components extracted from the stationary density to characterize martingale implications for time series data. The nonlinear principal components behave like scalar autoregressions with heteroskedastic innovations when viewed from the guise of a time series analyst. In addition, the principal components are ordered by the ratio of their long-run variation to the overall variation. Principal components that capture overall variation subject to smoothness restrictions also display low frequency variation due to their high persistence.

The diffusion model implicit in much of Salinelli (1998)'s analysis is a multivariate Brownian motion restricted to a compact state space with reflecting barriers at the boundary of the state space. This specification makes it easier to establish the existence of principal components. Many of the empirical models in economics and finance that motivate our paper imply processes that have non-attracting boundaries and stationary distributions with noncompact support. Although the state space can be transformed into a compact set, non-attracting boundaries will be preserved by the transformation. As a consequence either the density q or the diffusion matrix Σ will not be bounded away from zero. Establishing the existence of principal components in this setup is no longer routine.[2]

In this paper we do the following:

- Provide a Markov process for the data generation that supports the principal component extraction method and provides testable implications.
- Extend the principal component extraction of Salinelli (1998) to include state-dependence in the smoothness constraint and to allow for more general boundary behavior.
- Give sufficient conditions for the existence of these principal components and provide bounds on the persistence of the principal components.

Prior to developing formally these results, we give an overview and describe links to related literature.

2 Overview and Related Literature

We first review the functional principal component extraction method of Salinelli (1998) and describe extensions of this method required for Markov diffusion processes. We then discuss equivalent extraction procedures that aid our interpretation of the principal components. Finally, we delineate the testable time series implications for the principal components.

[2]In Salinelli (1998)'s words: "This problem seems very difficult to deal with; for this reason, we have restricted our attention to absolutely continuous r.v.s with positive bounded densities." Addressing this more difficult problem is necessary in our application to Markov processes with boundaries that are not attracting.

2.1 A Digression on Salinelli

Salinelli (1998) considers principal component extraction without reference to a diffusion. These principal maximize variation subject to smoothness constraints. In our generalization these principal components are defined as follows.

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

$$\max_{\phi} \int_{\Omega} (\phi)^2 q$$

subject to

$$\begin{aligned} \int_{\Omega} (\nabla\phi)'\Sigma(\nabla\phi) q &= 1 \\ \int_{\Omega} \phi\psi_s q &= 0, s = 0, ..., j-1 \end{aligned}$$

where ∇ is used to denote the (weak) gradient and ψ_0 is initialized to be the constant function one.

There are three differences between our proposed extraction and that of Salinelli (1998). First, Salinelli (1998) presumes the data are generated as a vector of independent and identically distributed observations whereas we consider stationary, reversible diffusion processes. The Σ matrix is then interpretable as a diffusion or local covariance matrix and q is the stationary density for the diffusion. In particular when q is uniform over a compact set and Σ is the identity matrix, the data generation under our reinterpretation would be that of a multivariate standard Brownian motion that reflects at the boundaries of the state space. Second, Salinelli (1998) assumes that Σ is state independent. To accommodate a rich class of stochastic processes, we allow Σ to be state dependent. Third, Salinelli (1998) assumes that the data density q has compact support and is bounded away from zero. We relax the assumption of a compact state space and accordingly the density q is no longer assumed to be bounded from below. This relaxation allows us to consider processes without attracting barriers, but it opens up the possibility that the principal component problem is no longer well posed. We address this by providing sufficient conditions for the existence of principal components.[3]

2.2 An Alternative Interpretation

Given a Markov process $\{x_t\}$ with stationary density q, we have at our disposal an alternative interpretation of the principal components based on long-run variation. For a test function

[3]There are other functional principal component constructions. For instance, Douxious and Pousse (1975) suggest a different way to construct nonlinear principal components for multivariate densities based on choosing pairs of functions that maximize cross correlations without penalizing derivatives. Darolles, Florens, and Renault (1998) and Darolles, Florens, and Gourieroux (2000) apply this construction to time series processes through the use of the joint density, say (x_{t+1}, x_t).

ϕ, we form the stochastic process $\{\phi(x_t)\}$. To measure the long-run variation of this process, we use the limit

$$\lim_{\alpha \downarrow 0} 2 \int_0^\infty \exp(-\alpha t) E[\phi(x_t)\phi(x_0)] dt$$

provided that this limit is well defined. As we will see this limit will turn out to be both the spectral density function for the process evaluated at frequency zero and the variance used in a martingale central-limit approximation for the process. As an alternative to the extraction suggested by Salinelli (1998) we may use:

Definition 2.2. *: The function ψ_j is said to be the j^{th} nonlinear principal component for $j \geq 1$ if ψ_j solves:*

$$\max_{\phi} \lim_{\alpha \to 0} 2 \int_0^\infty \exp(-\alpha t) E[\phi(x_t)\phi(x_0)] dt$$

subject to

$$\begin{aligned} \int_\Omega \phi^2 q &= 1 \\ \int_\Omega \phi \psi_s q &= 0, s = 0, ..., j-1 \end{aligned}$$

where we initialize ψ_0 to be one.

As we will see, this extraction will result in the same principal components as that given in Definition (2.1) provided that we use the diffusion matrix Σ to measure smoothness.[4]

In summary, we may think of principal component extraction as obtaining functions of the data with large long-run variances relative to their overall variance, or we may envision obtaining smooth functions that capture as much variation as possible. Smooth functions that capture variation in the data are also functions that have large long-run variances relative to their overall variance. It is the extended version of Salinelli (1998)'s extraction that gives rise to an operational method of computing principal components. These same components, however, are rank-ordered by the ratio of the long-run variation to the overall variation.

2.3 Testable Implications for Time Series

The principal components extracted by penalizing derivatives are also eigenfunctions of the expectation operators associated with the underlying Markov process. We may define a conditional expectation operator for each interval s of elapsed time. A principal component ψ satisfies:

$$E[\psi(x_{t+s})|x_t] = \exp(-\delta s)\psi(x_t), \tag{1}$$

[4] In fact any positive discount rate α can be used and the principal components will remain the same.

for some positive number δ and each transition interval s. This defines ψ as the eigenfunction of the conditional expectation operator for each time interval. Thus principal components extracted by one of the two procedures described previously will also satisfy the testable conditional moment implications (1). For any such principal component, 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 conditional heteroskedastic (have conditional variance that depends on the Markov state x_t).

2.4 Other Literature in Time Series

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, Scheinkman, and Touzi (1998), Darolles, Florens, and Renault (1998) and Florens, Renault, and Touzi (1998). In particular, Kessler and Sorensen (1999) use eigenfunctions to construct quasi-optimal estimators of parametric models of the drift and diffusion coefficients from discrete-time data. Demoura (1998), Hansen and Scheinkman (1995), Hansen, Scheinkman, and Touzi (1998), Darolles, Florens, and Renault (1998) and Florens, Renault, and Touzi (1998) study semiparametric and nonparametric identification and over-identification based on a principal component extraction that is closely related to the one analyzed here. The same principal components that we identified from the stationary density are solutions to constrained, maximal autocorrelation problems:

$$\exp(-\delta_j) = \sup_{\phi} E[\phi(x_{t+1})\phi(x_t)]$$

subject to:

$$\begin{aligned} \int \phi^2 q &= 1 \\ \int \phi \psi_s q &= 0, s = 0, 1, ..., j-1. \end{aligned}$$

Maximally autocorrelated functions also display maximal long-run variation relative to the overall variation.

Nonparametric estimation methods for time-series extraction based on autocorrelation have been proposed by Demoura (1998), Chen, Hansen, and Scheinkman (1998) and Darolles, Florens, and Gourieroux (2000).[5]

[5]Demoura (1998) assumes that the stationary density has exponentially thin tails, Chen, Hansen, and Scheinkman (1998) allow for algebraic tails, and Darolles, Florens, and Gourieroux (2000) formally treat the case in which the density has compact support and is bounded above and below. As mentioned previously, Darolles, Florens, and Renault (1998) and Darolles, Florens, and Gourieroux (2000) use the canonical principal component decompositions Douxious and Pousse (1975) applied to the joint density of (x_{t+1}, x_t). Their principal components coincide with the maximal autocorrelation construction when the process is reversible. Darolles, Florens, and Gourieroux (2000) use this insight to propose a test of time reversibility.

In addition to basing the extraction on the stationary density, we provide sufficient conditions for the existence of principal components for multivariate diffusions with unbounded supports. Our existence results also apply to principal component extraction based on maximal autocorrelations. On the other hand, formal justification of methods of inference is beyond the scope of this paper.

2.5 Threshold Model

We use a bivariate, continuous-time extension of the familiar threshold autoregression model to illustrate the construction of principal components.[6] Let $\Sigma = I$ and

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

The proportionality factor can be chosen so that the density q integrates to unity. The implied drift μ is given by:

$$\mu(x) = \begin{array}{cc} -(|x|-1)\frac{x}{|x|} & \text{if } \ |x| \geq 1 \\ 0 & \text{if } \ |x| < 1 \end{array}$$

Notice that along any ray from the origin, the drift is piece-wise linear: it is zero inside the unit circle and it pulls towards zero outside the unit circle. The density q that supports this construction has exponentially thin tails just as the stationary density implied by mean-reverting scalar threshold model. The radial symmetry is chosen for simplicity.

In our computations we assume the dimension of the process is two. The principal components are sometimes of the form:

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

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 ψ and ψ^* with the same smoothness and the same variance. One of these, say ψ 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 under the stationary distribution and are ordered by their smoothness (persistence). 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. By construction the principal component is smoother than the drift, which has a kink on the unit circle.

[6] See Tong (1990) for a discussion of discrete-time threshold autoregressive models. For analyses of other continuous-time threshold models see Stramer, Brockwell, and Tweedie (1996) and Stramer, Tweedie, and Brockwell (1996).

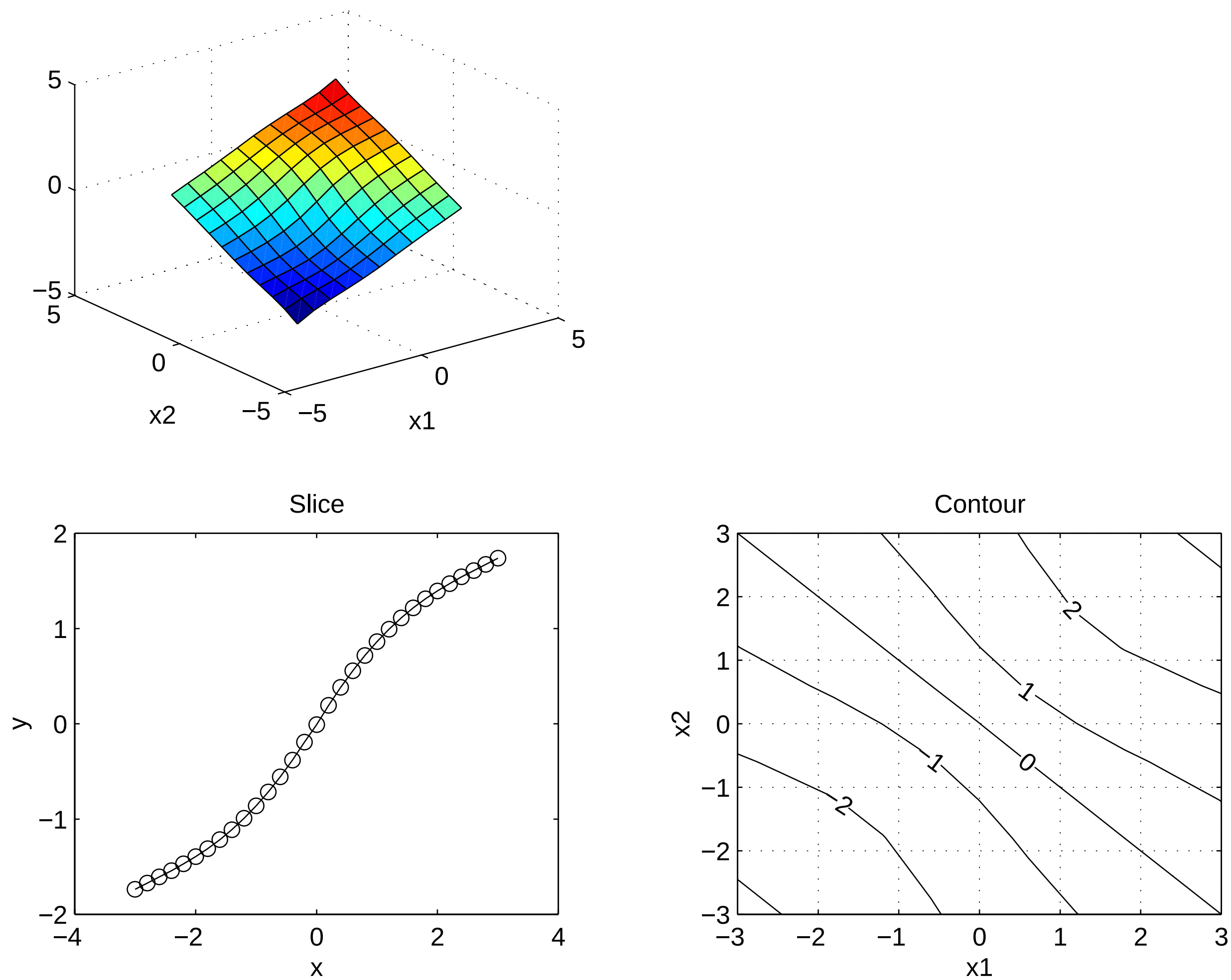

Figure 1: This figure displays the symmetric principal component in the first symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the other 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.

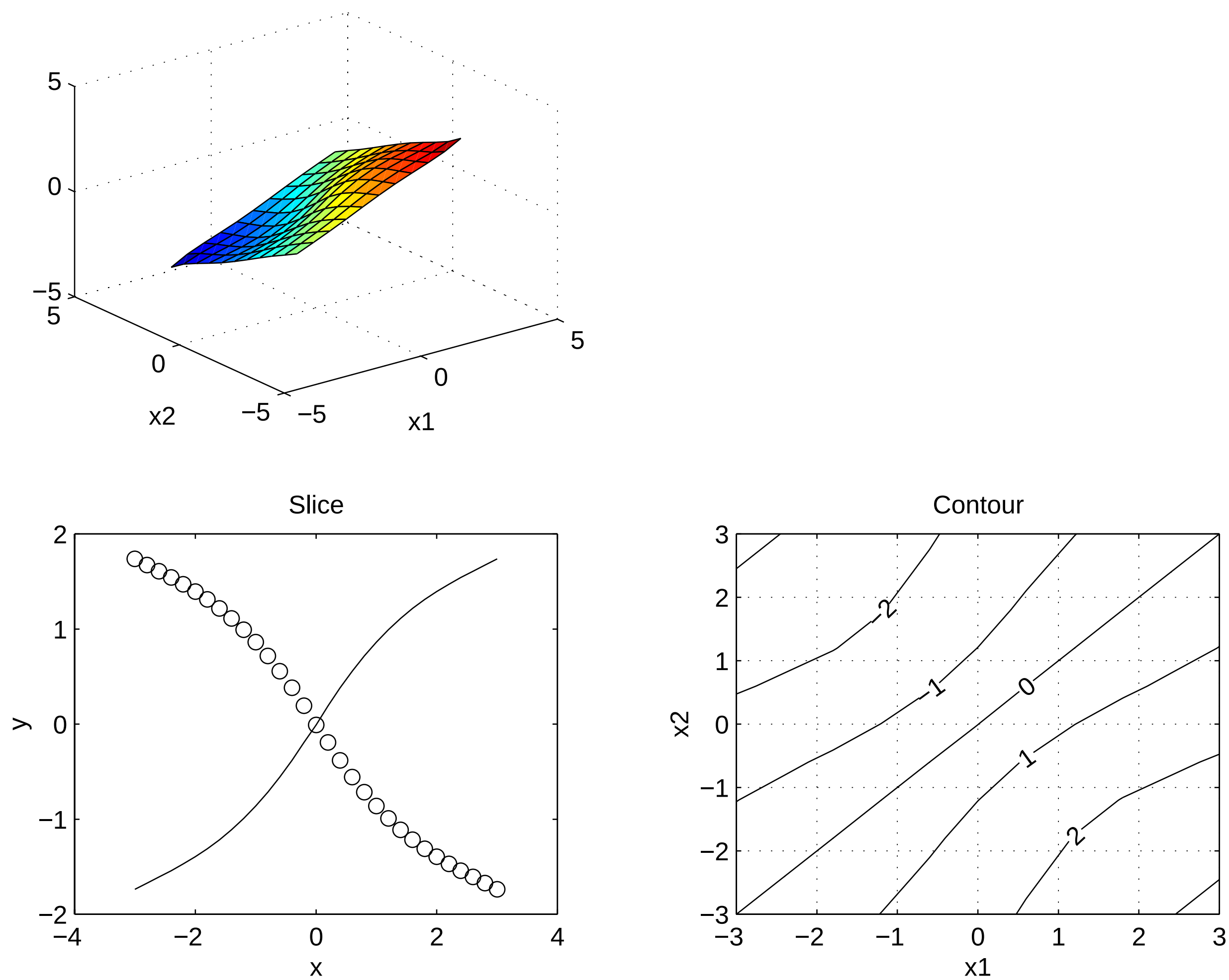

Figure 2: This figure displays the anti-symmetric principal component in the first symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the 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.

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.

Threshold models are often used in economics to study the convergence of relative prices across locations. These models have some heuristic appeal because in a world with transaction costs, one expects there to be more pull towards the center of the relative price distribution when the relative prices are further away from unity. Univariate threshold models are more typical in the empirical literature. The precise threshold specification used here and in empirical literature is hard to defend, however. In particular it is known that conditional volatility can play an important part in keeping distributions away from extreme values (see Conley, Hansen, Luttmer, and Scheinkman (1998)). The construction and use of principal components potentially can help to understand better the observed nonlinear dynamics in a time series. Given their implications for predictability, they offer convenient and revealing diagnostics.

The remainder of this paper is organized as follows. In sections 3 and 4 we show how to build a Markov process associated with functional principal component extraction using a state-dependent diffusion matrix and a density. In section 5 we provide a quadratic form and an operator characterization of the principal components, in section 6 we derive sufficient conditions for the existence of principal components when the state space is not compact, and in section 7 we study the limiting behavior of the eigenvalues. The asymptotic behavior of the eigenvalues helps us to understand how well a finite number of principal components approximate features of the stationary distribution and the temporal evolution of the process.

3 Parameterization

In this section we describe and justify a parameterization of a multivariate Markov diffusion. We limit our attention to processes that are stationary and reversible. We parameterize our candidate Markov process using two functions: a probability density q over the state space and a state-dependent positive semidefinite matrix Σ. We give sufficient conditions for there to exist a unique (reversible) Markov process associated with (q, Σ). The function q will turn out to be the stationary density for this process and Σ will be the diffusion or local covariance matrix.

We start with an open connected $\Omega \subseteq \mathbb{R}^n$, the state space for our processes, and a bilinear (quadratic) form f_o defined on $\mathcal{C}_K^2$, the space of twice continuously differentiable functions with compact support in Ω, that can be parameterized in terms of a density q and a positive definite matrix Σ that can depend on the state:

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

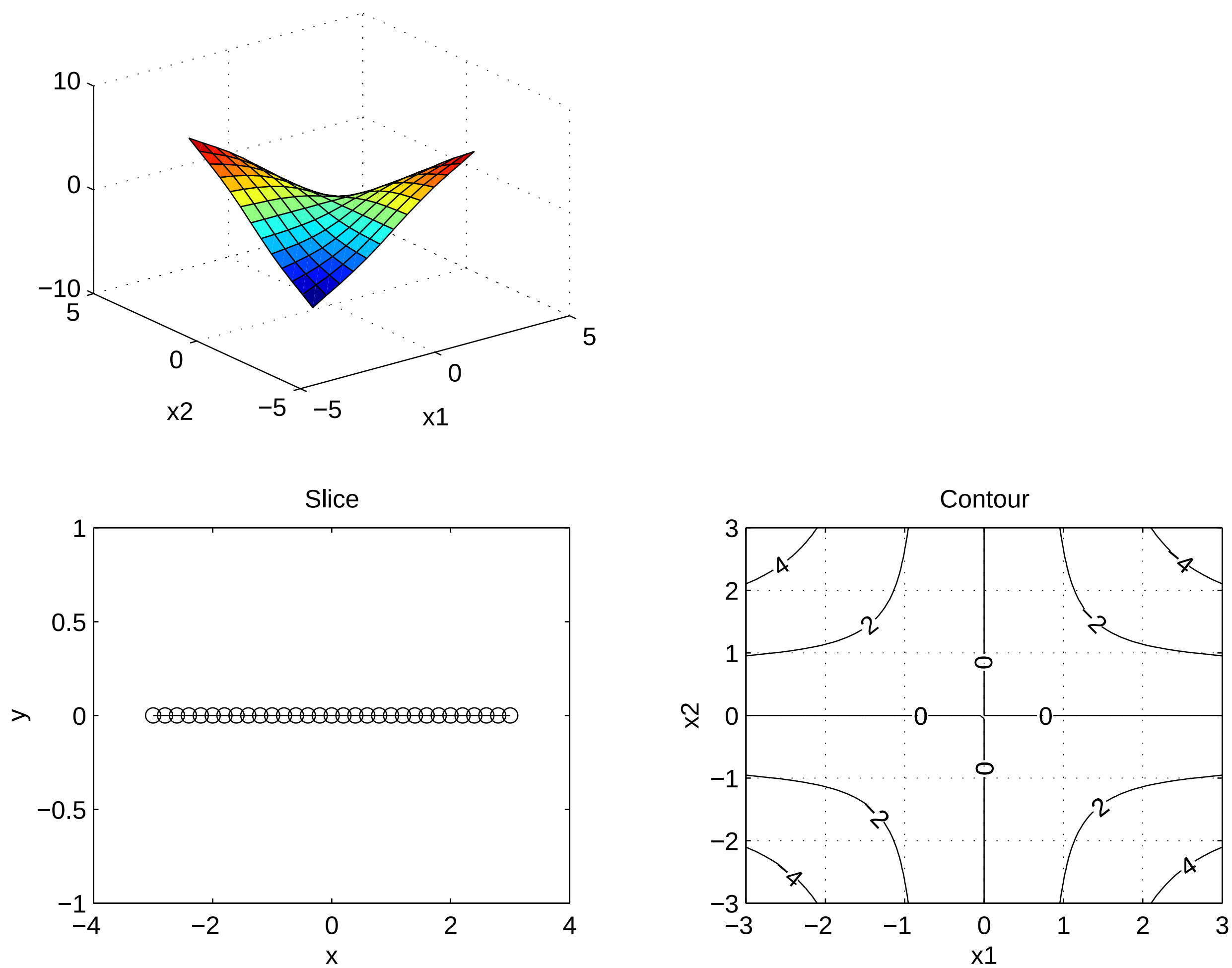

Figure 3: This figure displays the symmetric principal component in the second symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the other 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.

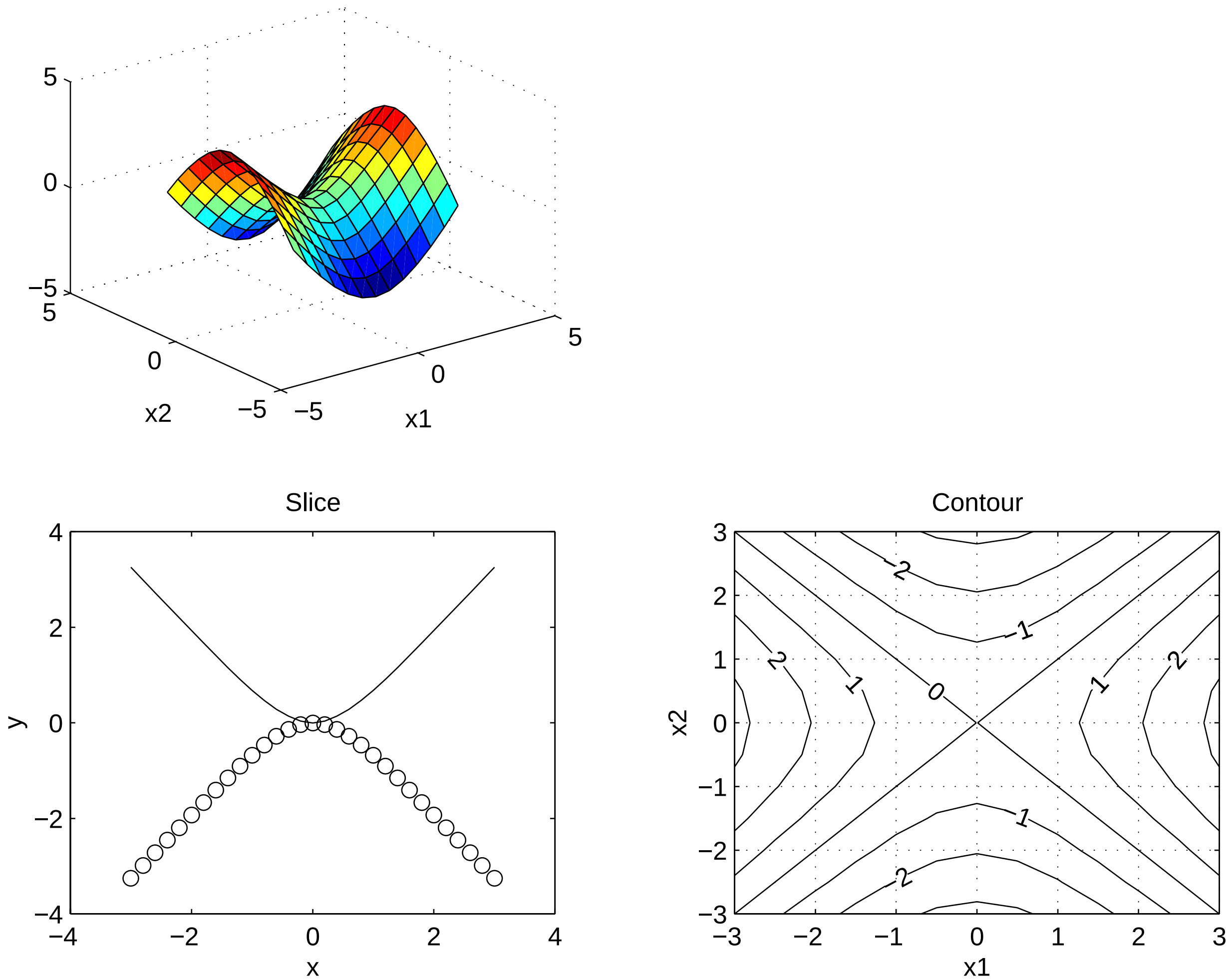

Figure 4: This figure displays the anti-symmetric principal component in the second symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero and the 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.

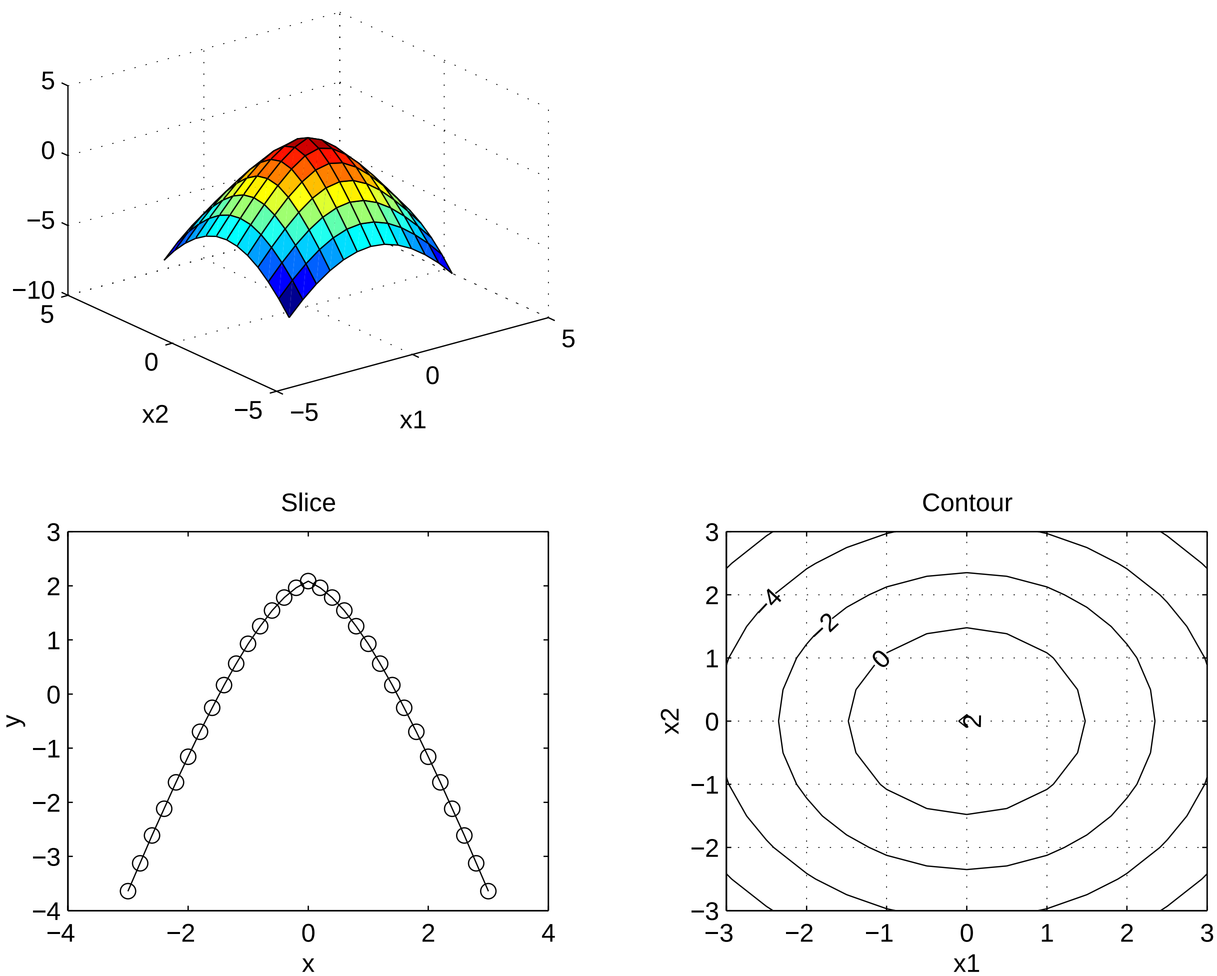

Figure 5: This figure displays the fifth principal component. This principal component is constant on circles centered at zero. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at 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.

where

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

Assumption 3.1. *: q is a strictly positive, continuously differentiable probability density on Ω.*

Assumption 3.2. *: Σ is a continuously differentiable, positive definite matrix function on Ω.*[7]

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. The domains of these extensions are in the space $\mathcal{L}^2$ of square integrable functions with respect to the measure induced by q. Let $< \cdot, \cdot >$ denote the usual inner product on $\mathcal{L}^2$. In this paper we consider exclusively the case in which q is integrable. As we will see, this restriction is used to produce a stationary Markov process. In fact, the Markov process construction we employ permits more general measures including infinite ones.

There is a differential operator L that is associated with the form f_o, which we construct using integration-by-parts. For any pair of functions ϕ and ψ in C_K^2:

$$\begin{aligned} f_o(\phi, \psi) &= \frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_j} \frac{\partial \psi}{\partial y_i} q \\ &= -\frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} \psi q - \frac{1}{2} \int \sum_{i,j} \frac{\partial (q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j} \psi \end{aligned} \tag{3}$$

where the second equality of (3) 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 (3) to motivate our interest in the differential operator L:

$$L\phi = \frac{1}{2} \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} + \frac{1}{2q} \sum_{i,j} \frac{\partial (q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j}. \tag{4}$$

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

$$\begin{aligned} f_o(\phi, \psi) &= - < L\phi, \psi > \\ &= - < \phi, L\psi > . \end{aligned}$$

where the second relation holds because we can interchange the role of ϕ and ψ in (3). Notice from (4) that operator L 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.

[7] Assumptions 3.1 and 3.2 restrict the density q and the diffusion matrix Σ to be continuously differentiable. As argued by Davies (1989) these restrictions can be replaced by a less stringent requirement that entries of the matrix $q\Sigma$ are locally (in $\mathcal{L}^2$(Lebesgue)), weakly differentiable. (See Theorem 1.2.5.)

Remark 3.3. *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:*

$$\mu_j = \frac{1}{2q} \sum_{i=1}^{n} \frac{\partial(\sigma_{ij} q)}{\partial y_i}. \tag{5}$$

Set

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

Then we may use Ito's Lemma to show that for each $\phi \in \mathcal{C}_K^2$

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

where this limit is taken with respect to the $\mathcal{L}^2$ *That is, L coincides with the infinitesimal generator of* $\{x_t\}$ *in* $\mathcal{C}_K^2$*. We use this link to the stochastic differential equation to motivate our labeling the matrix* Σ *as the diffusion matrix. In what follows, however, we will use a construction of the Markov process that does not make direct use of stochastic differential equations.*

As we will show, with this parameterization of L, the function q in Assumption 3.1 is the stationary density of the associated Markov process. Formula (5) restricts the drift and guarantees that the diffusion is reversible. Moreover, this formula provides a generalization of the formula used by Banon (1978) to recover the drift coefficient from the diffusion matrix and from the stationary density (see also Hansen and Scheinkman (1995), page 779). The implicit parameterization for the drift is restrictive when the dimension n of the diffusion exceeds one. By design it can accommodate any stationary distribution with a density that satisfies the smoothness conditions of Assumption 3.1.

4 Building the Markov Process

So far we have constructed a quadratic form f_o on $\mathcal{C}_K^2$. We now extend this form and build formally a generator for a Markov process from this extension.

4.1 Extension

Given a positive-semidefinite quadratic form f, we may construct an inner product

$$< \phi, \psi >_f = < \phi, \psi > + f(\phi, \psi)$$

on the domain of the form. If this domain is complete with this inner product (a Hilbert space), we say that the form is closed. The form f_o is not closed, but as we will now show it is *closable* (has a closed extension).

Let

$$\mathcal{L}^2 = \{\phi \in \Omega : \int (\phi)^2 q < \infty\}.$$

We extend the form f_o to a larger domain $\bar{\mathcal{H}} \subset \mathcal{L}^2$ using the notion of a weak derivative.

$$\begin{aligned}\bar{\mathcal{H}} &= \{\phi \in \mathcal{L}^2 : \text{there exists } g \text{ measurable, with } \int g'\Sigma g q < \infty, \\ &\quad \text{and} \int \phi \nabla \psi = -\int g\psi, \text{for all } \psi \in \mathcal{C}_K^1\}.\end{aligned}$$

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 $\bar{\mathcal{H}}$ we write $\nabla\phi = g$.[8] For any pair of functions ψ and ϕ in $\bar{\mathcal{H}}$ we define:

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

which is an extension of f_o. In $\bar{\mathcal{H}}$ we use the inner product $<\phi, \psi>_{\bar{f}} = <\phi, \psi> + \bar{f}(\phi, \psi)$.

Proposition 4.1. *$\bar{\mathcal{H}}$ is a Hilbert space.*

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

$$\begin{aligned}\phi &= \lim_{j\to\infty} \phi_j \\ v &= \lim_{j\to\infty} \Lambda\nabla\phi_j.\end{aligned}$$

For each $u \in \mathcal{C}_K^1$ we know that:

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

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

[8] Notice that $\bar{\mathcal{H}}$ is constructed exactly as a weighted Sobolev space except that instead of requiring that $g \in \mathcal{L}^2$, we require that $\Lambda g \in \mathcal{L}^2$ where Λ is the square root of Σ. Also we use $\mathcal{C}_K^1$ 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, $\mathcal{C}_K^\infty$. See Brezis (1983) Remark 1, page 150.

Since $\bar{\mathcal{H}}$ is a Hilbert space, it follows that the form $\bar{f}$ is a closed extension of f_o. There may, however, be other closed extensions. One such closed extension is the *minimal closed extension* of f_o. This extension is obtained by considering all the sequences $\phi_j \in \mathcal{C}_K^2$ that are Cauchy with respect to the norm induced by $< \cdot, \cdot >_{\bar{f}}$. Since the form $\bar{f}$ is closed, any such Cauchy sequence has a limit point in $\mathcal{D}(\bar{f}) = \bar{\mathcal{H}}$. We let $\mathcal{D}(\hat{f})$ denote the space of all such limit points, and we define $\hat{f}$ to be the restriction of $\bar{f}$ to $\mathcal{D}(\hat{f})$. The form $\hat{f}$ is closed, and any closed extension f of f_o must have a domain that includes $\mathcal{D}(\hat{f})$. This justifies calling $\hat{f}$ the minimal closed extension.

4.2 Generator

Associated with any closed extension f (including $\bar{f}$ and $\hat{f}$) defined on a subspace $\mathcal{D}(f)$ 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 $\phi \in \mathcal{L}^2$, define $G_\alpha \phi \in \mathcal{D}(f)$ to be the solution to

$$f(G_\alpha \phi, \psi) + \alpha < G_\alpha \phi, \psi >=< \phi, \psi > \tag{6}$$

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 (1994) pages 15 and 19). In particular, G_α is a one-to-one mapping from $\mathcal{L}^2$ into $G_\alpha(\mathcal{L}^2)$.

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

$$F\phi = (G_\alpha)^{-1}\phi - \alpha\phi \tag{7}$$

defined on the domain $G_\alpha(\mathcal{L}^2)$. It can be shown that F is independent of α.[9] This construction of the generator F guarantees the form representation:

$$f(\phi, \psi) =< F\phi, \psi > \tag{8}$$

for ψ in the domain of the form.

We also use the family of resolvent operators to build a semigroup of conditional expectation operators. A natural candidate for this semigroup is $\{\exp(-tF)\}$ where we call $-F$ the *generator* of this semigroup. 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:

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

[9] 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: $f(\phi, \psi) =< \sqrt{F}\phi, \sqrt{F}\psi >$ (*e.g.* see Fukushima, Oshima, and Takeda (1994) Theorem 1.3.1).

as a Neumann series expansion. Notice that (7) implies

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

Instead of the direct use of a series expansion, we use the limit

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

often referred to as Yosida approximation to construct formally a strongly continuous, semigroup of operators indexed by $t \geq 0$.[10]

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:

$$G_\alpha \phi = \int_0^\infty \exp(-\alpha t)\exp(-tF)\phi dt \tag{9}$$

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 $\{G_\alpha : \alpha > 0\}$.

Consider now the forms $\hat{f}$ and $\bar{f}$ constructed as extensions of f_o. Let $\hat{F}$ and $\bar{F}$ be given by formula (7). For this operator construction to be of interest we want $\hat{F}$ and $\bar{F}$ to be extensions of $-L$. The operator $\hat{F}$ is known to be an extension of $-L$ (*e.g.* see Lemma 3.3.1 of Fukushima, Oshima, and Takeda (1994)), and it is commonly referred to as the Friedrichs extension. The operator $\bar{F}$ is also known to be an extension of $-L$. We include the following argument for sake of completeness. (See Lemma 3.3.4 of Fukushima, Oshima, and Takeda (1994).[11])

Claim 4.2. *$\bar{F}$ is an extension of $-L$.*

Proof. See the Appendix B. ∎

The unit function is in the space $\bar{\mathcal{H}}$, and

$$\bar{f}(1, \phi) = 0$$

for any $\phi \in \bar{\mathcal{H}}$. Notice that $\bar{G}_\alpha 1 = \frac{1}{\alpha}$ solves equation (6). Thus the unit function is also in the domain of the operator $\bar{F}$ and $\bar{F}1 = 0$. This property of $\bar{F}$ will turn out to be important when we use $\bar{F}$ to construct a Markov semigroup. The unit function will not generally be in the domain of the Friedrichs extension $\hat{F}$, however.

[10] Strong continuity requires that $\exp(-tF)\phi$ converges in $\mathcal{L}^2$ to ϕ as t declines to zero.

[11] Fukushima, Oshima, and Takeda (1994) assume extra smoothness in establishing this and other results. The extra smoothness conditions are not used in their proof of Lemma 3.3.4.

4.3 Cores

While the domain $\mathcal{D}(\bar{f})$ has a simple characterization, this may not be true of other extensions f. In addition the associated self-adjoint operators F may fail to have domains with simple descriptions. Since domains of forms and operators are sometimes hard to specify explicitly, in what follows we will make reference to the *core* of a closed extension f and the associated self-adjoint operator F. These cores are collections of functions that are rich a particular sense and are often easier to characterize than the corresponding domains.

Definition 4.3. *A family of functions $\mathcal{C}o \subset \mathcal{D}(f)$ is a* core *of f if for any ϕ_o in the domain $\mathcal{D}(f)$, there exists a sequence $\{\phi_j\}$ in $\mathcal{C}o$ such that*

$$\lim_{j\to\infty} < \phi_j - \phi_o, \phi_j - \phi_o >_f = 0.$$

The construction of the minimal extension, $\hat{f}$, in section 4.1 guarantees that $\mathcal{C}_K^2$ is a form core.

Definition 4.4. *A family of functions $\mathcal{C}o \subset \mathcal{D}(F)$ is a* core *of F if for any ϕ_o in the domain $\mathcal{D}(F)$, there exists a sequence $\{\phi_j\}$ in $\mathcal{C}o$ such that*

$$\begin{aligned} \lim_{j\to\infty} \phi_j &= \phi_o \\ \lim_{j\to\infty} F\phi_j &= F\phi_o \end{aligned}$$

in $\mathcal{L}^2$.

4.4 Markov Semigroup

As we have just seen, associated with a closed form f, there is an operator F and a (strongly continuous) semigroup $\exp(-tF)$ on $\mathcal{L}^2$. 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 $\mathcal{L}^2$, $0 \leq \exp(-Ft)\phi \leq 1$. A semigroup satisfying this property is called *submarkov* in the language of Beurling and Deny (1958). Second we require, for each $t \geq 0$, $\exp(-Ft)1 = 1$. A semigroup satisfying this property is said to *conserve probabilities*. We refer to a submarkov semigroup that conserves probabilities as a Markov semigroup.[12] Finally we must make sure that the Markov semigroup is actually the family of conditional expectation operators of a Markov process, a question we will address in section 4.5.

The following condition is sufficient for a closed form to generate a submarkov semigroup (*e.g* see Davies (1989) section 1.3).

[12]Fukushima (1971) and others do not use the term *submarkov* when the semigroup fails to conserve probabilities. Fukushima (1971) shows that when the operators fail to conserve probabilities, a Markov process construction is still possible, but on an extended state space. We will describe this construction subsequently.

Condition 4.5. *(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 $\exp(-Ft)$ is submarkov, and for each $t \geq 0$, $\exp(-Ft)$ is an $\mathcal{L}^2$ contraction ($\|\exp(-Ft)\phi\|_2 \leq \|\phi\|_2$). This contraction property is also satisfied for the $\mathcal{L}^p$ norm for $1 \leq p \leq \infty$ (Davies (1989) Theorem 1.3.3). In particular, we may extend the semigroup from $\mathcal{L}^2$ to $\mathcal{L}^1$ while preserving the contraction property.

The semigroup conserves probability when the unit function is in the domain of F and

$$F1 = 0. \tag{10}$$

Notice that $\bar{f}$ satisfies the Beurling-Deny conditions and $\bar{F}$ satisfies (10). Hence the semigroup $\exp(-\bar{F}t)$ is a Markov semigroup and from Claim 4.2 its generator $\bar{F}$ is an extension of $-L$. This gives a formal way to construct a Markov semigroup from the differential operator (4) or equivalently from the form $\bar{f}$.

Similarly, the minimal extension, $\hat{f}$, defined in Section 4.1 satisfies the Beurling-Deny criteria (Davies (1989) Theorem 1.3.5). Thus there exists a self-adjoint operator $\hat{F}$ associated with $\hat{f}$ which is an extension of $-L$ and generates a submarkov semigroup $\exp(-\hat{F}t)$. Without further restrictions, this semigroup will not necessarily conserve probabilities.

4.5 Markov Processes and the Minimal Closed Extension

We have already argued that there is a self-adjoint operator, $\hat{F}$, associated with the minimal extension, $\hat{f}$, that is an extension of the differential operator $-L$. The operator $\hat{F}$ generates a submarkov semigroup $\exp(-\hat{F}t)$ on $\mathcal{L}^2$. Since $\mathcal{C}_K^2$ is a core for the minimal extension, Theorem 7.2.1 of Fukushima, Oshima, and Takeda (1994) guarantees that there exists a Markov process $\{x_t\}$ constructed on an extended state space that has $\exp(-\hat{F}t)$ as its semigroup of conditional expectations.[13] Since the semigroup may be submarkov (fails to conserve probabilities) the construction allows for the existence of an absorbing state. The semigroup is defined for functions that are zero at this additional state. The probabilistic interpretation is as follows. When the semigroup fails to conserve probabilities, a Markov process $\{x_t\}$ is constructed with boundaries that are attainable. The process dies when the boundary is contacted and is formally absorbed into the additional state.[14] The process $\{x_t\}$ can be constructed to have continuous sample paths with probability one prior to hitting the boundary. (See Theorem 7.2.2 of Fukushima, Oshima, and Takeda (1994).)

[13]Actually it is a *Hunt process*, a special strong Markov process that possesses certain sample path continuity properties. (See Fukushima, Oshima, and Takeda (1994), Appendix A.2 for a definition.)

[14]We may measure the probability of not hitting the boundary over an interval of time by $\exp(-\hat{F})1$. In contrast, a diffusion process associated with the form $\bar{F}$, when it exists, reflects off the boundary when the boundary is hit and continues, thus conserving probabilities.

4.6 Uniqueness

In this section we show that if the minimal extension conserves probability, it coincides with the form $\bar{f}$, and in fact with any closed extension of f_o that satisfies the Beurling-Deny Condition. Since any closed extension that leads to a submarkov semigroup must satisfy the Buerling-Deny condition (see Theorem 1.4.1, page 23 of Fukushima, Oshima, and Takeda (1994)), if the minimal extension conserves probabilities, the constructed Markov process must be "unique". For this reason most of the rest of this paper deals with the case where the minimal extension conserves probabilities.

We also discuss sufficient conditions on the parameterization (Σ, q) that guarantees that the minimal extension conserves probability. The following result is due essentially to Davies (1985) and Azencott (1974)

Proposition 4.6. *Suppose that $\hat{F}1 = 0$. Let f be any other closed extension of f_o that satisfies Condition 4.5. Then $\hat{f} = f$ and $\hat{F} = F$ where F is the operator associated with the form f.*

Proof. Since $\hat{F}1 = 0$, it follows that

$$\hat{G}_\alpha 1 = \frac{1}{\alpha}.$$

As we remarked earlier, the Buerling-Deny criterion implies that the semigroup $\exp(-\hat{F}t)$ can be extended to a semigroup in $\mathcal{L}^1$. Let $\hat{F}^1$ denote the generator of the $\mathcal{L}^1$ semigroup. Since $\hat{G}_\alpha 1$ is constant and bounded away from zero, then the argument of Theorem 2.2 of Davies (1985) and Corollary 2.5 of Azencott (1974) can be easily modified to show that $\mathcal{C}^2_K$ is a core of $\hat{F}^1$. While Davies (1985) assumes extra smoothness conditions for Σ and q than we have here, the proof of his Lemma 2.3 is applicable more generally, and in particular to our environment. As a consequence, the second part of his proof of Theorem 2.2 remains valid for our problem.

Since $-F$ also satisfies the Buerling-Deny criteria, the semigroup $\exp(-Ft)$ may also be extended to $\mathcal{L}^1$. Let F^1 be the generator of this extended semigroup. Note that for any $\phi \in \mathcal{C}^2_K$

$$\begin{aligned} -L\phi &= F\phi = \hat{F}\phi \\ &= F^1\phi = \hat{F}^1\phi. \end{aligned} \tag{11}$$

Since $\mathcal{C}^2_K$ is a core for $\hat{F}^1$, $(\hat{F}^1 + \alpha I)(\mathcal{C}^2_K)$ is dense in $\mathcal{L}^1$. It follows that $\mathcal{C}^2_K$ must also be core for F^1 and that $F^1 = \hat{F}^1$. However, since q is integrable, $\mathcal{L}^2$ convergence implies $\mathcal{L}^1$ convergence and consequently F and $\hat{F}$ are restrictions of F^1 and $\hat{F}^1$, respectively. The conclusion follows. □

For the semigroup $\exp(-\hat{F}t)$ to conserve probability, it suffices to show that the unit function is in the domain of the minimal extension $\hat{f}$. Fukushima, Oshima, and Takeda

(1994) provide a convenient sufficient condition.[15] 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 ψ and ϕ in $\mathcal{C}_K^\infty$ that are radially symmetric, *i.e.* $\phi(x) = \phi^*(|x|)$ and $\psi(x) = \psi^*(|x|)$, we may depict the form f_o as an integral over radii:

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

A sufficient condition for the conservation of probabilities is:

$$\int_0^\infty \kappa(r)^{-1}r^{1-n}dr = \infty. \tag{12}$$

(See Fukushima, Oshima, and Takeda (1994) Theorem 1.6.6 and Theorem 1.6.7.) When this condition is met, a sequence of functions ϕ_j in $\mathcal{C}_K^2$ can be constructed that converge to the unit function in $\mathcal{L}^2$ and $f_o(\phi_j, \phi_j)$ converges to zero.[16] This places the unit function in the domains of all of the form closures of f_o including $\hat{f}$. Moreover, $\hat{f}(1, \phi) = 0$ for all ϕ in the domain of the minimal extension.

Notice that (12) 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:

$$\begin{aligned} +\infty &= \left(\int_0^\infty \frac{1}{r}dr\right)^2 \\ &\leq \int_0^\infty \kappa(r)^{-1}r^{1-n}dr \int_0^\infty \kappa(r)r^{n-3}dr. \end{aligned}$$

Thus a sufficient condition for (12) is that

$$\int_0^\infty \frac{\kappa(r)}{r^2}r^{n-1}dr < \infty. \tag{13}$$

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

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

[15] Another set of sufficient conditions can be obtained by observing that a recurrent semigroup conserves probabilities (Fukushima, Oshima, and Takeda (1994) Lemma 1.6.5). Hasminskii (1960) and Stroock and Varadhan (1979) have suggested the following approach for demonstrating recurrence. Find a smooth, strictly positive function V that tends to ∞ as $|x|$ gets arbitrarily large and satisfies: $LV \leq V$. This approach has given rise to sufficient conditions for recurrence based on alternative choices of V.

[16] An approximating sequence of functions with compact support is supplied by Fukushima, Oshima, and Takeda (1994) in the proof of Theorem 1.6.7. This sequence can be smoothed using a suitable regularization to produce a corresponding approximating sequence in $\mathcal{C}_K^2$.

and

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

Notice that

$$\kappa(r) \leq \sigma^2(r)\lambda(r).$$

Suppose for instance, $\sigma(r)$ is dominated by a quadratic function (in r). Then (13) and hence (12) are satisfied because the density q is integrable:

$$\int_0^\infty \lambda(r) r^{n-1} dr = 1.$$

Quadratic bounds rule out many interesting examples including ones studied by Conley, Hansen, Luttmer, and Scheinkman (1998) in which stationarity is *volatility-induced*. These latter examples allow for growth in the diffusion coefficient that may be faster than quadratic. We may extend the previous argument by supposing instead that

$$\sigma^2(r) \leq c|r|^{2+2\delta}$$

for some positive δ. Then

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

Thus (13) is satisfied provided that

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

Hence we can allow for faster growth in σ if q has high enough moments.

Remark 4.7. *For a scalar diffusion, let s denote the familiar scale density used to classify whether a boundary is attracting or not. The scale density is proportional to the reciprocal of the stationary density times the diffusion coefficient. Thus*

$$\kappa(r) = \frac{1}{s(r)} + \frac{1}{s(-r)},$$

and hence

$$\kappa(r)^{-1} \leq \min\{s(r), s(-r)\}.$$

When restriction (12) is satisfied, both infinite boundaries are not attracting.

4.7 Long-run

In this section we show that if the minimal extension conserves probability then q is a stationary density of this Markov process, justifying our intuition about the parameterization. In fact we will show that given any operator F satisfying $F1 = 0$, a Markov process with semigroup of conditional expectations given by $exp(-tF)$ is recurrent and has q as its stationary density.

Stationarity (with density q) is equivalent to

$$\int \exp(-tF)\phi q = \int \phi q$$

for $\phi \in \mathcal{L}^2$. Recurrence requires that

$$\lim_{\alpha \downarrow 0} G_\alpha \phi = 0 \quad \text{or} \quad \infty$$

for any nonnegative function ϕ in $\mathcal{L}^1$.[17]

Claim 4.8. *When F satisfies (10), then* $\exp(-Ft)$ *is recurrent and has q as its stationary density.*

Proof. To verify that the semigroup is conservative, note that

$$\exp(-Ft)1 - 1 = -\int_0^t \exp(-Fs)F1 ds = 0.$$

Since $\exp(-Ft)$ is self-adjoint, for any $\phi \in \mathcal{L}^2$,

$$\begin{aligned} \int \exp(-Ft)\phi q &= < \exp(-Ft)\phi, 1 > \\ &= < \phi, \exp(-Ft)1 > \\ &= \int \phi q, \end{aligned}$$

which shows that q is a stationary density for the semigroup $\exp(-Ft)$. Since the density q is integrable, the unit function is in $\mathcal{L}^p$ for $1 \leq p \leq \infty$ and $\alpha G_\alpha 1 = 1$. Thus $G_\alpha 1$ tends to $+\infty$ as α declines to zero. Since the unit function is q-integrable, it follows from part (b) Lemma 1.6.4 of Fukushima, Oshima, and Takeda (1994)) that semigroup is recurrent. □

[17] A recurrent semigroup is always conservative (see Lemma 1.6.5 of Fukushima, Oshima, and Takeda (1994)), even if q is not integrable. As we will see in the next proof, when q is integrable a conservative semigroup is always recurrent. Also Assumption 3.2 guarantees that the form $\hat{f}$ and hence the semigroup $\exp(-t\hat{F})$ is irreducible (see Example 4.6.1 on page 173 of Fukushima, Oshima, and Takeda (1994)).

4.8 Long-Run Variance

In this subsection we construct a second quadratic form used to depict the long-run variance of stochastic processes constructed from the Markov process $\{x_t\}$. This construction is only of interest when the Markov process is stationary (when $\exp(-tF)$ conserves probability).

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 $\mathcal{S}(F)$ of functions in $\mathcal{L}^2$ 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 (6) or (7).

In light of equation (9)

$$< G_\alpha \phi, \psi >= \int_0^\infty \exp(-\alpha t) E[\phi(x_t)\psi(x_0)]dt. \tag{14}$$

Hence, using (7), we obtain:

$$\begin{aligned} g(\phi, \psi) &= \lim_{\alpha \downarrow 0} 2 < G_\alpha \phi, \psi > \\ &= \lim_{\alpha \downarrow 0} 2 < (\alpha I + F)^{-1} \phi, \psi > . \end{aligned}$$

Notice that this form is symmetric because the resolvent operator is self adjoint for any positive α. Using (14) we may write this form as

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

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 (1982) and Hansen and Scheinkman (1995), we may use this martingale construction to justify:

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

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.

5 Eigenfunctions

Principal components are eigenfunctions of the quadratic forms f and g. In this section we define eigenfunctions of forms and their associated operators. In the next section we study the existence of eigenfunctions.

5.1 Eigenfunctions of a Form

An eigenfunction ϕ of the quadratic form f satisfies:

$$f(\phi, \psi) = \delta < \phi, \psi > \tag{15}$$

for all $\psi \in \mathcal{D}(f)$. The scalar δ is the corresponding eigenvalue. Since f is positive semidefinite, δ must be nonnegative. As we will see, the eigenfunctions of the form f, when they exist, are also the principal components.

5.2 Eigenfunctions of Operators

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 equation (6):

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

From this formula, we may verify that

$$G_\alpha \phi = \frac{1}{\delta + \alpha} \phi.$$

In fact f and G_α must share eigenfunctions for any $\alpha > 0$. The eigenvalues are related via the formula:

$$\lambda = \frac{1}{\delta + \alpha} \tag{16}$$

where λ is the eigenvalue of G_α and δ is the corresponding eigenvalue of f. In establishing the existence of principal components, it is convenient to study eigenvalues of the corresponding resolvent operators. The resolvent operators have $\mathcal{L}^2$ as their domain and are bounded.

Given relation (8) between the form f and the operator F, we expect them to share eigenvalues and eigenfunctions. This is verified formally using the formula

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

and relation (16). 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 ϕ is an eigenfunction of the form f with eigenvalue δ, ϕ is an eigenfunction of $\exp(-tF)$ with eigenvalue $\exp(-t\delta)$ for any positive t. Thus eigenfunctions of the form f satisfy the time series conditional moment restriction (1) described in section 2.

Since the form g is given by:

$$\lim_{\alpha \downarrow 0} 2 < G_\alpha \phi, \psi >,$$

g shares eigenfunctions with f with eigenvalues related via the formula: $\lambda_g = 2/\delta$.[18] As we will see, t is the fact that f and g share eigenvalues that delivers the *long run* interpretation of the principal component extraction suggested by Salinelli (1998).

5.3 Extracting Eigenfunctions from Resolvent Operators

Eigenfunctions, when they exist, may be extracted from the resolvent operator by solving a sequence of maximization problems. We know that: $f(\phi, \psi) + \alpha < \phi, \psi >$ defines an inner product on domain $\mathcal{D}(f)$ of the form f. We will use this inner product in the following sequence of maximization problems, defined recursively. Let $\mathcal{H}_j$ denote the space of functions in $\mathcal{D}(f)$ that are orthogonal to the first ψ_s for $s = 0, 1, ..., j-1$ where $\psi_0 = 1$. Construct the j^{th} eigenfunction ψ_j by solving

$$\begin{aligned} \lambda_j &= \sup_{\mathcal{H}_j} f(G_\alpha \phi, \phi) + \alpha < G_\alpha, \phi > \\ &\quad \text{subject to} \quad f(\phi, \phi) + \alpha < \phi, \phi >= 1. \end{aligned}$$

Since $f(G_\alpha \phi, \phi) + \alpha < G_\alpha \phi, \phi >=< \phi, \phi >$, this maximization problem may be written equivalently as:

$$\begin{aligned} \lambda_j &= \sup_{\mathcal{H}_j} < \phi, \phi > \\ &\quad \text{subject to} \quad f(\phi, \phi) + \alpha < \phi, \phi >= 1. \end{aligned} \tag{17}$$

The corresponding eigenfunction ψ_j is the solution to this problem. Eigenfunctions associated with distinct eigenvalues are orthogonal with respect to f and with respect to $< \cdot, \cdot >$. Thus in computing eigenvalues, we may use the inner product $< \cdot, \cdot >$ to construct the spaces $\mathcal{H}_j$. While this change in inner products will alter the spaces, it will not alter the

[18] Since the order zero eigenfunction of the form f is constant, it is not in the domain of the form g.

solutions to the maximization problems. The $\alpha = 0$ limit of this sequence of maximization problems gives our generalization of the Salinelli's principal component extraction described in Section 2 (see Definition 2.1).

As an alternative extraction algorithm we could use the $< \cdot, \cdot >$ inner product at the outset. This would give rise to:

$$\begin{aligned} \lambda_j &= \sup_{\mathcal{H}_j} < G_\alpha \phi, \phi > \\ & \quad \text{subject to} \quad < \phi, \phi >= 1. \end{aligned}$$

The $\alpha = 0$ limit of this extraction gives the *long run* principle component extraction described in Section 2 (see Definition 2.2).

In the next section we present sufficient conditions for solutions to exist to the eigenfunction extraction problems. As a precursor to this discussion, notice that the constraint:

$$f(\phi, \phi) + \alpha < \phi, \phi >= 1$$

in extraction problem (17) can be relaxed by replacing the equality with the weak inequality $\leq$. To establish the existence of a solution, it suffices to suppose the following:

Criterion 5.1. $\{\phi \in \mathcal{D}(f) : f(\phi, \phi) + \alpha < \phi, \phi > \leq 1\}$ *is precompact (has compact closure) in* $\mathcal{L}^2$ *for some* α.

The precompactness restriction guarantees that we may extract an $\mathcal{L}^2$ 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.[19] In checking *compactness Criterion* 5.1 we are free to choose α in a convenient manner.

We provide sufficient conditions for this compactness criterion in the next section.

Remark 5.2. *It is known that Criterion 5.1 is equivalent to either of the following:*

1. *F has a complete orthonormal basis of eigenfunctions.*

2. *The resolvent operators G_α are compact (map the unit ball in $\mathcal{L}^2$ into compact sets).*

(See Corollary 4.2.3 of Davies (1995) and Theorem IV.2.9 Edmunds and Evans (1987).) In particular, when we provide sufficient conditions for Criterion 5.1, they will imply that the resolvent operators are compact. These conditions will also imply that the conditional expectation operators $\exp(-Ft)$ *for* $t > 0$ *are compact.*

[19]This result follows because of what is sometimes referred to as the parallelogram law. Consider the Cauchy sequence $\{\phi_j\}$ used to approximate the supremum. Then

$$f(\phi_j - \phi_k, \phi_j - \phi_k) = 2f(\phi_j, \phi_j) + 2f(\phi_k, \phi_k) - 4f\left(\frac{\phi_j + \phi_k}{2}, \frac{\phi_j + \phi_k}{2}\right).$$

The existence of solution follows because $\frac{1}{2}(\phi_j + \phi_k)$ satisfies the constraint with a weak inequality and the form f is closed. In fact as Reed and Simon (1978) argue, once we know that principal component problems are well posed, it may be shown that the constraint set is closed and hence compact (see page 245).

6 Existence

In section 5 we noted that the compactness Criterion 5.1 implies the existence of principal components for the forms f and g. We now consider more primitive sufficient conditions that imply this criterion. Specifically, we consider the $\hat{f}$ form, which is the minimal extension of the form f_o. This is done for mathematical convenience. We are primarily interested in the case in which the resulting generator and semigroup conserve probabilities, implying that the $\hat{f}$ and $\bar{f}$ forms coincide.

This section is organized as follows. We first review the existence condition used by Salinelli (1998) and the one derived by Hansen, Scheinkman, and Touzi (1998) for the real line. We then extend their results to multivariate processes defined on $\mathbb{R}^n$ using two devices. First, we transform the function space and hence the form so that distribution induced by q is replaced by Lebesgue measure. This transfomation allows us to apply known results for forms built using Lebesgue measure. Second, we study forms that are simpler but dominated by $\hat{f}$. When the dominated forms satisfy Criterion 5.1 the same can be said of $\hat{f}$.

6.1 Compact Domain

Salinelli (1998) used Rellich's compact embedding theorem applied to a compact state space to establish existence of principal components when the domain Ω is bounded with a continuous boundary. His approach can be employed here provided that the density is bounded and bounded away from zero and the diffusion matrix is uniformly nonsingular. Under these conditions, the boundaries of the diffusion are attracting. The boundary behavior implicit in our analysis is reflection. In what follows we will study models in which boundaries are not attracting.

6.2 Real Line

Consider first results on the real line. Hansen, Scheinkman, and Touzi (1998) consider scalar diffusions, whose boundaries are not attracting. An equivalent statement of their compactness condition, written in terms of the stationary density and the scalar diffusion coefficient ρ^2 is:

$$\lim_{|y|\to\infty} -\frac{y}{|y|}\left(\rho\frac{q'}{q} + \rho'\right) = +\infty. \tag{18}$$

The second term in parenthesis in (18) does not help when ρ has at most linear growth , *i.e.*

$$\limsup_{|y|\to\infty} -\frac{y}{|y|}\rho' < +\infty.$$

Then compactness reduces to the requirement that

$$\lim_{|y|\to\infty} -\rho\frac{y}{|y|}\frac{q'}{q} = +\infty. \tag{19}$$

In particular, the stationary density must not have algebraic tails, tails that decay faster than $|y|$ raised to a negative power. On the other hand, the resolvent operators may be compact when volatility growth is greater than linear, that is when

$$\lim_{|y|\to\infty} \frac{y}{|y|}\rho' = +\infty.$$

Fat tails in the stationary density are sometimes permitted when the volatility growth is faster than linear. The second term in compactness criterion (18) comes into play, however.

We cannot simply increase the growth of volatility, while holding fixed the stationary density, to satisfy the compactness condition, since we will eventually make $\pm\infty$ attracting barriers. To see this, recall that the scale density function for a one-dimensional diffusion is proportional to $\frac{1}{\rho^2 q}$. Boundaries are attracting when the scale density is integrable. Thus when the diffusion coefficient growth more than offsets the density decay, the infinite boundaries can be attained in finite time. The form $\hat{f}$ is associated with a submarkov semigroup: a semigroup that fails to be conservative.

To illustrate these phenomena, we study a family of diffusion models with fat-tailed distributions.

Example 6.1. *Consider a scalar diffusion with $q(y) = (1+y^2)^{-\gamma}$ and $\rho^2 = (1+y^2)^{\beta}$. For q to be integrable we require that $\gamma > 1/2$. Compactness condition (18) requires that $\beta > 1$ and*

$$2\gamma > \beta. \tag{20}$$

This compactness condition, however, was derived for processes with boundaries that are not attainable. Both boundaries are nonattracting when

$$2\gamma \geq 2\beta - 1$$

which, since $\beta > 1$, implies (20).

In the subsections that follow, we will provide multivariate extensions for both sources of compactness: growth in the logarithmic derivative of the density and growth in the diffusion coefficient. For simplicity, we take the state space to be $\mathbb{R}^n$.

6.3 Transforming the Measure

In this subsection we transform the space $\mathcal{L}^2$ into its Lebesgue counterpart $\mathcal{L}^2(leb)$ and study the implied quadratic form $\tilde{f}$ that is induced by $\hat{f}$. 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 the existence of eigenfunctions for quadratic forms.

Given q write:

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

Assumption 6.2. *The function h is twice continuously differentiable.*

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

Map the space $\mathcal{L}^2$ into $\mathcal{L}^2(leb)$ by the unitary transformation:

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

Notice that U leaves $\mathcal{C}_K^2$ invariant. Associated with the minimal extension $\hat{f}$ is a closed form $\tilde{f}$ defined by:

$$\begin{aligned} \tilde{f}(\psi, \psi^*) &= \tilde{f}(U\phi, U\phi^*) \\ &= \hat{f}(\phi, \phi^*) \end{aligned}$$

on the domain $\mathcal{D}(\tilde{f}) = U\mathcal{D}(\hat{f})$. Since $\hat{f}$ is the minimal closed extension of f_o, $\tilde{f}$ is the minimal closed extension of the form $\tilde{f}$ restricted to $\mathcal{C}_K^2$.

To depict the induced form, $\tilde{f}$, on $\mathcal{C}_K^2$, we use the product formula:

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

Then

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

Applying integration-by-parts for $\phi \in \mathcal{C}_K^2$, it follows that

$$\frac{1}{2}\int(\nabla h)'\Sigma[\nabla(\psi\psi^*)] = -\frac{1}{2}\int\sum_{i,j}\sigma_{i,j}\frac{\partial^2 h}{\partial y_i \partial y_j}\psi\psi^* - \frac{1}{2}\int\sum_{i,j}\frac{\partial\sigma_{i,j}}{\partial y_i}\frac{\partial h}{\partial y_j}\psi\psi^*.$$

Therefore,

$$\tilde{f}(\psi, \psi^*) = \frac{1}{2}\int(\nabla\psi)'\Sigma(\nabla\psi^*) + \frac{1}{2}\int V\psi\psi^* \tag{21}$$

where the *potential* function V is given by:

$$V = -\sum_{i,j}\sigma_{i,j}\frac{\partial^2 h}{\partial y_i \partial y_j} - \sum_{i,j}\frac{\partial\sigma_{i,j}}{\partial y_i}\frac{\partial h}{\partial y_j} + (\nabla h)'\Sigma(\nabla h). \tag{22}$$

The form $\tilde{f}$ will be positive semidefinite because $\hat{f}$ is. Moreover, these forms share the same eigenvalues. Nevertheless, the potential function V can be negative. We will eventually restrict V to be bounded from below implying the existence of a positive number α such that $2\alpha + V$ is strictly positive. Thus a possibly different closed extension of $\tilde{f}$ may be built by using the formula:

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

defined on the domain:

$$\mathcal{D}(f^*) = \{\psi \in \mathcal{L}^2(leb) : \phi \text{ has a weak derivative and } \int (2\alpha + V)(\psi)^2 + \int (\nabla\psi)'\Sigma(\nabla\psi) < \infty\}.$$

We may show the form f^* is closed by imitating the proof of Proposition 4.1. Thus f^* must be an extension of the minimal extension $\tilde{f}$. In particular, the representation of $\tilde{f}$ by (21) is valid on its entire domain.

As we argued in Section 5, for the existence of a principal component decomposition of $\hat{f}$, we seek sufficient conditions for Criterion 5.1: the collection of functions:

$$\mathcal{U}_\alpha = \{\phi \in \mathcal{D}(\hat{f}) : \hat{f}(\phi, \phi) + \alpha < \phi, \phi > \leq 1\}$$

is precompact in $\mathcal{L}^2$ for some $\alpha > 0$. In terms of the transformed space, it suffices to show that

$$\mathcal{V}_\alpha = \{\psi \in \mathcal{D}(f^*) : \frac{1}{2} \int (2\alpha + V)\, \psi^2 + \frac{1}{2} \int (\nabla\psi)'\Sigma(\nabla\psi) \leq 1\}$$

is precompact in $\mathcal{L}^2(leb)$ for some $\alpha > 0$.

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

6.4 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 $\mathcal{V}_\alpha$ we impose:

Assumption 6.3. *The diffusion matrix $\Sigma \geq \underline{c}I$ for some positive $\underline{c} > 0$.*

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

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

Criterion 6.4. $\lim_{|x| \to \infty} V(x) = +\infty$.

Proposition 6.5. *Under Assumptions 6.2 and 6.3, if Criterion 6.4 is satisfied, then Criterion 5.1 is satisfied for $\hat{f}$.*

Proof. Since V is continuous and diverges at the boundaries, it must be bounded from below. Also, it follows from Assumption 6.3 that

$$\mathcal{V}_\alpha \subset \{\psi \in \mathcal{L}^2(leb) : \phi \text{ has a weak derivative and } \frac{1}{2}\int (2\alpha + V)(\psi)^2 + \frac{1}{2\underline{c}} \int |\nabla\psi)|^2 \leq 1\}.$$

We may apply the argument in the proof of Theorem XIII.67 of Reed and Simon (1978) to establish that $\mathcal{V}_\alpha$ is precompact in $\mathcal{L}^2(leb)$. □

Example 6.6. *Suppose that $\Sigma = I$. Then the potential function V is given by*

$$V = -\text{trace}\left(\frac{\partial^2 h}{\partial x \partial x'}\right) + (\nabla h) \cdot (\nabla h).$$

For compactness Criterion 6.4, we require that:

$$\lim_{|x| \to \infty} -\text{trace}\left(\frac{\partial^2 h}{\partial x \partial x'}\right) + (\nabla h) \cdot (\nabla h) = +\infty.$$

In this example, Criterion 6.4 gives rise to a rather simple characterization for compactness based on the tail behavior of stationary density. This simplicity is achieved by making the diffusion matrix independent of the state vector as is implicit in principal component formulation of Salinelli (1998). More generally, direct verification of Criterion 6.4 may be difficult because formula (22) is a bit complicated.

An alternative potential function can be constructed by replacing q and Σ by conveniently chosen lower bounds. As we will see in the next subsections this bounding approach can lead to sufficient conditions that are easier to check. It also allows us to impose the extra smoothness condition required by Assumption 6.2 on a lower bound to the density function rather than directly on the density function q. Finally, by using a lower bound on the diffusion matrix Σ, we will be able to produce an additional nonnegative term that can be added to the potential function.

6.5 Radial Bounds

To derive more primitive or easily interpretable restrictions, we now place radial bounds on the density q and the diffusion matrix Σ.

Assumption 6.7. *There exists a twice continuously differentiable function $\nu : \mathbb{R}^+ \to \mathbb{R}$ such that $q \geq \exp[-2\nu(|x|)]$.*

Notice that the density q does not have to be twice differentiable. It is the function ν used to depict the radial bounds that is assumed to display sufficient smoothness.

Assumption 6.8. *The diffusion matrix Σ satisfies*

$$\Sigma(x) \geq \rho(|x|)^2 I$$

where $\rho(r) = \underline{c}(1 + r^2)^{\frac{\beta}{2}}$ and $\beta \geq 0$.

Construct the form:

$$\check{f}(\phi, \phi) = \frac{1}{2} \int |\nabla \phi|^2 \exp[-2\nu(|x|)] \rho(|x|)^2$$

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

$$\check{f}(\phi, \phi) \leq \hat{f}(\phi, \phi).$$

We may now use $\check{f}$ in place of $\hat{f}$ in the previous construction of the potential V. For instance, we may use:

$$h(x) = \nu(|x|).$$

Then

$$\nabla h(x) = \nu'(|x|)\frac{x}{|x|}.$$

As a consequence,

$$\sum_i \frac{\partial^2 h}{\partial x_i^2} = \nu''(|x|) + (n-1)\frac{\nu'(|x|)}{|x|}$$

The radial counterpart to the potential function is:

$$\check{V} = \rho^2[-\nu'' - (n-1)\frac{\nu'}{r} + (\nu')^2 - 2\frac{\rho'}{\rho}\nu']. \tag{23}$$

The principal components exist under the following restrictions:

Criterion 6.9. *The diffusion* $\Sigma(x) \geq \rho(|x|)^2 I$*, the density* $q \geq \exp[-2\nu(|x|)]$ *and*

$$\lim_{r\to\infty} \check{V}(r) = \infty.$$

This compactness criterion can sometimes be simplified.

Proposition 6.10. *Suppose that a) Assumption 6.7 is satisfied; b) Assumption 6.8 is satisfied for* $\beta \leq 1$*; c) the limits* $\lim_{r\to\infty} \rho\nu'$ *and* $\lim_{r\to\infty} \check{V}(r)$ *are well defined. Then* $\check{V}$ *converges to* $+\infty$ *at* ∞ *if, and only if* $\rho\nu'$ *converges to* $+\infty$ *at the same point.*

Proof. See Appendix B. □

This claim justifies the compactness criterion

$$\lim_{r\to\infty} \rho\nu' = +\infty$$

when the growth in ρ is no more than linear ($\beta \leq 1$). It gives a partial extension of the compactness criterion of Hansen, Scheinkman, and Touzi (1998) and is a multivariate counterpart to restriction (19).

Remark 6.11. *Proposition 6.10 assumes that the potential function* $\check{V}$ *has a well defined limit at* ∞*. It is evident from the proof of this proposition that this requirement may be replaced by instead requiring that* $\rho^2[-\nu'' + (\nu')^2]$ *has a well defined limit at the boundary. Given the volatility bounding function* ρ *behaves like* r^β *for large* r*, this latter requirement is a restriction on the density bound that should be easy to verify.*

Proposition 6.10 only works for stationary densities with exponential tails. In the next subsection we study how volatility growth can be exploited to allow a weaker bound on the stationary density.

6.6 Volatility-Aided Compactness

Growth in the diffusion matrix Σ can sometimes be used to extend our compactness criterion. To study this phenomenon we strengthen our lower bound on the diffusion matrix. The outcome of our analysis in this subsection is an additional term:

$$\check{W}(r) = (\frac{n}{2} - 1 + \beta)^2 \frac{\rho(r)^2}{(1+r^2)}$$

that we can add to our previously derived potential functions V or $\check{V}$. The additional term will only be of assistance when the growth in the diffusion matrix is more than quadratic ($\beta > 1$).

Criterion 6.12. *The diffusion matrix* $\Sigma(x) \geq \rho(|x|)^2 I$ *for* $\beta > 1$ *and*

$$\lim_{|x| \to \infty} V(x) + (1-\epsilon)\check{W}(|x|) = +\infty.$$

for some $\epsilon > 0$.

Criterion 6.13. *The diffusion matrix* $\Sigma(x) \geq \rho(|x|)^2 I$ *for* $\beta > 1$*, the density* $q(x) \geq \exp[-2\nu(|x|)]$ *and*

$$\lim_{r \to \infty} \check{V}(r) + (1-\epsilon)\check{W}(r) = +\infty$$

for some $\epsilon > 0$.

Before justifying these criteria, we need the following inequality:

Proposition 6.14. *Suppose that Assumption 6.8 is satisfied for* $\beta > 1$*. Then*

$$\frac{1}{2}\int |\nabla\psi(x)|^2 \rho(|x|)^2 \geq \frac{(\frac{n}{2} - 1 + \beta)^2}{2} \int \frac{\rho(|x|)^2}{(1+|x|^2)}\psi(x)^2 \qquad (24)$$

for all $\psi \in \mathcal{C}_K^2$.

Proof. Let J be the second-order differential operator associated with this form. Apply J to the function

$$\chi(x) = (1+|x|^2)^{-c/2}$$

for some $c > 0$ and compute:

$$\begin{aligned} J\chi &= \frac{c}{2}\chi\frac{\rho^2}{(1+|x|^2)}[n + 2\beta\frac{|x|^2}{1+|x|^2} - (c+2)\frac{|x|^2}{(1+|x|^2)}] \\ &\geq \frac{c}{2}\frac{\rho^2}{(1+|x|^2)}(n - c - 2 + 2\beta)\chi \end{aligned}$$

It follows from Theorem 1.5.12 of Davies (1989) that

$$\frac{1}{2}\int |\nabla\psi|^2\rho(|x|)^2 \geq \frac{1}{2}\int W_c\psi^2 \tag{25}$$

on $\mathcal{C}_K^2$ where W_c is the function:

$$W_c(x) = c\frac{\rho^2}{(1+|x|^2)}[n-c-2+2\beta].$$

We choose the parameter c to make the lower bound (25) as sharp as possible. Choosing c to maximize the quadratic function $c(n-c-2+2\beta)$ leads to inequality (24). □

While the inequality in Proposition 6.14 is established on $\mathcal{C}_K^2$, it can be extended to the domain of a closed form with a $\mathcal{C}_K^2$ core and in particular to the domain of the minimal extension of any form defined in $\mathcal{C}_K^2$.

Two corollaries now follow.

Corollary 6.15. *If Criterion 6.12 is satisfied, then Criterion 5.1 is satisfied for $\hat{f}$.*

Corollary 6.16. *If Criterion 6.13 is satisfied, then Criterion 5.1 is satisfied for $\hat{f}$.*

The construction of the original potential function V can be simplified at the outset by replacing Σ with its lower bound $\rho(|x|)^2$ but continuing to use the density $q = \exp(-2h)$. This results in the function:

$$V^*(x) = \rho^2[-\sum_i \frac{\partial^2 h}{\partial x_i^2} + |\nabla h|^2 - 2\frac{\rho'}{\rho}(\frac{x}{|x|})\cdot\nabla h]$$

The function V^* may be used in place V in the compactness criterion 6.12.

Example 6.17. *We consider a simplification of the compactness criterion 6.13 when the density bound is algebraic:*

$$\nu = \frac{\gamma}{2}\log(1+r^2) + c^*.$$

For the density lower bound to be integrable, we need $\gamma > \frac{n}{2}$. Using formula (23) we may show that $\check{V}$ behaves like:

$$\left(\frac{\rho^2}{1+r^2}\right)\left[2\gamma\left(1-\frac{n}{2}-\beta\right)+\gamma^2\right]$$

for large r. Thus Criterion 6.13 will be satisfied provided that

$$\left(\gamma-\frac{n}{2}+1-\beta\right)^2 > 0.$$

For instance, we may exploit volatility growth to demonstrate compactness of the resolvent operators provided that

$$\beta < \gamma - \frac{n}{2} + 1. \tag{26}$$

The resolvent operators associated with $\hat{f}$ will also be compact in Example 6.17 when $\beta > \gamma - \frac{n}{2} + 1$. Recall, however, that by making the volatility growth large, while holding fixed the density, the semigroup $\exp(-t\hat{F})$ may eventually fail to conserve probabilities (be submarkov). That is, eventually the forms $\hat{f}$ and $\bar{f}$ will differ because the infinite boundaries may be attainable. As we saw in our previous discussion, we may use upper bounds on q and Σ to obtain sufficient conditions for the conservation of probabilities for the Markov process associated with the minimal extension $\hat{f}$. To establish compactness of resolvent operators we are led to use lower bounds on q and Σ.

Example 6.18. *Pang (1996) studies differential operators in which the density $q \propto (\kappa)^{-2\gamma}$ and $\Sigma \propto (\kappa)^{2\beta} I$ where κ is a positive scalar function that is bounded above and below by positive scalar multiples of $1 + |x|$.*[20] *This nicely worked out example suggests that there is little scope for refining our results using radial bounds. Specifically, Pang (1996) establishes the following:*

- *When $\beta > 1$ and $\beta < \gamma - \frac{n}{2} + 1$, the resolvent operators associated with $\hat{f}$ are compact and the semigroup conserves probability.*
- *When $\beta > 1$ and $\beta > \gamma - \frac{n}{2} + 1$, the resolvent operators associated with $\hat{f}$ are compact, but the corresponding semigroup fails to conserve probability.*
- *When $\beta < 1$ the resolvent operators associated with $\hat{f}$ fail to be compact.*

It is the first of these cases that is primarily interest to us and is consistent with our use of radial bounds and inequality (26). The second case illustrates that increasing volatility elasticity β holding fixed the stationary density can cause infinite boundaries to be attracting.

7 Eigenvalue Divergence

In this section we briefly survey and apply some existing results on how fast the eigenvalues of the form diverge. If δ is an eigenvalue of the form f, $\frac{1}{\delta}$ is the corresponding eigenvalue a principal component extracted by maximizing the variance subject to smoothness constraints. Similarly, $\frac{2}{\delta}$ is the corresponding eigenvalue of the form g used to define the extraction based on maximizing long-run variation. Finally, $\exp(-t\delta)$ is the eigenvalue of the conditional expectation operator over an interval of time t. Thus large eigenvalues of the form f correspond to small eigenvalues of the associated principal component extractions and of the conditional expectation operators. Finite-dimensional approximation of conditional expectation operators based on principal components will work better when the eigenvalues of the form diverge more quickly.

[20] Pang (1996) uses σ instead of κ to denote the scalar building-block function for the density and diffusion matrix. Our parameters β and γ are linked to his parameters a and b via the formulas: $2\beta = a - b + 2$ and $2\gamma = a + n$.

To motivate the analysis in this section, consider a spectral decomposition of a positive semidefinite operator F used to generate a Markov process:

$$F\phi = \sum_j \delta_j < \phi, \psi_j > \psi_j$$

for ϕ in $\mathcal{D}(F)$. The eigenvalues δ_j are listed in increasing order and repeated according to their multiplicity, and the eigenfunctions ψ_j are scaled to have unit norms. The corresponding spectral decomposition of the conditional expectation operator over an interval t is

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

for ϕ in $\mathcal{L}^2$. When the eigenvalues of F increase rapidly, ignoring tail terms in this sum is less problematic because the $< \phi, \psi_j >$ is discounted by $\exp(-t\delta_j)$ for large j. Notice that the implied transition density is given by the expansion:

$$\sum_j \exp(-t\delta_j)\psi_j(y)\psi_j(x)q(y)dy$$

over an interval t where x is the value of the current Markov state. Again eigenvalues of F are used as discount rates for terms in this expansion. Large tail eigenvalues of F limit the approximation errors induced by ignoring higher-order terms.

Our characterizations of eigenvalue divergence use scale multiples of forms with known eigenvalue divergence as lower bounds on forms that interest us. From what is referred to as the $\max - \min$ principle for closed symmetric forms, we know that if we can order the forms, we can order the eigenvalues of the forms. Consider a form f and a form f_ℓ such that $\mathcal{D}(f) \subset \mathcal{D}(f_\ell)$ and $f(\phi, \phi) \geq f_\ell(\phi, \phi)$ for all ϕ in their common domain $\mathcal{D}(f)$. Then the j^{th} eigenvalue of the form f_ℓ is less than or equal to the j^{th} eigenvalue of the form f (*e.g.* see Edmunds and Evans (1987), Lemma 2.3 page 499). This gives us an operational way to use known eigenvalue decay characterizations for three special forms to bound the eigenvalue divergence for the forms that interest us.

All forms studied in this section are constructed on $\mathcal{L}^2(leb)$. We saw in Section 6 how to transform the $\mathcal{L}^2$ forms into corresponding forms on $\mathcal{L}^2(leb)$ with the use of potential functions.

7.1 Compact Support

To provide a benchmark, we first consider the Ω is an open subset of $\mathbb{R}^n$ with compact closure and a boundary that is *minimally smooth* (see Edmunds and Evans (1987) page 255 for a definition). Consider the form:

$$f_K(\phi, \psi) = \int_\Omega \phi\psi + \int_\Omega \nabla\phi\nabla\psi$$

defined on the corresponding Sobolev space. The eigenvalues of the form f_K satisfy:

$$\delta_j \geq c^*(j+1)^{2/n}$$

for some positive number c^* (*e.g.* see Edmunds and Evans (1987) Theorem 6.5 page 292).[21]

If we scale a form by a positive number, we scale the eigenvalues of the form. Thus only the constant in the approximation formula changes. To relate back to Salinelli (1998)'s functional principal components, consider the case of multivariate Brownian motion on Ω with reflection at the boundary. This gives rise to a form proportional to:

$$\int_\Omega \nabla\phi\nabla\psi. \tag{27}$$

The eigenvalues of this form are equal to the eigenvalues of f_K up to a translation by minus one. Thus the eigenvalue bounds for f_K also apply to the gradient quadratic form (27).

7.2 $\Omega = \mathbb{R}^n$

In this section we report the eigenvalue behavior of two forms studied previously. We depict these forms using the weighting function:

$$w(x) = (1+|x|^2)^\omega.$$

The forms of considered in this section are:

$$f_w(\phi,\psi) = \int \phi\psi w + \int \nabla\phi\nabla\psi$$

and

$$f_w^d(\phi,\psi) = \int \phi\psi w + \int \nabla\phi\nabla\psi w$$

defined on the appropriately defined subspaces of $\mathcal{L}^2(leb)$.

Notice that the difference in the two forms is that f_w^d has a common weighting for both function levels and (weak) derivatives whereas the weighting only applies to the function levels for the form f_w. In what follows we will take as the domain $\mathcal{D}(f_w)$ the space of all weakly differentiable functions ϕ in $\mathcal{L}^2(leb)$ such that

$$\int \phi\phi w + \int \nabla\phi\nabla\phi < \infty,$$

[21]Edmunds and Evans (1987) provide upper and lower bounds on the approximation numbers for the embedding of a Sobolev space into $\mathcal{L}^2(leb)$. The $(j+1)^{st}$ approximation number is the square roots of the reciprocal of the j^{th} eigenvalue of the form f_K. The reciprocal of the squared upper bound of the $(j+1)^{st}$ approximation number gives the lower bound of interest to us.

and similarly the domain $\mathcal{D}(f_w^d)$ includes the weakly differentiable functions in $\mathcal{L}^2(leb)$ for which:

$$\int \phi\phi w + \int \nabla\phi\nabla\phi w < \infty.$$

Both forms are clearly closed on these domains. While these domains may be larger than those of the minimal extensions using $\mathcal{C}_K^2$ functions, we may still use these forms as lower bounds because of the max-min principal.

Claim 7.1. *(Reed and Simon (1978)) Suppose that $\omega > 1/2$. Then the j^{th} eigenvalue of f_w satisfies:*

$$\delta_j \geq c^*(j+1)^{\frac{2\omega}{n(1+\omega)}}$$

for some positive constant c^*.

Proof. Reed and Simon (1978) (Theorem XIII.81) show that the number of eigenvalues of f_w that are less than or equal to $r > 0$ can be bounded above and below by scale multiples of[22]

$$\int_0^{\hat{u}} [r - (1+u^2)^\omega]^{\frac{n}{2}} u^{n-1} du.$$

where

$$\hat{u} = (r^\omega - 1)^{\frac{1}{2}}.$$

Since $r \geq r - (1+u^2)^\omega$, the integral is dominated by

$$r^{\frac{n}{2}} \int_0^{\hat{u}} u^{n-1} du = \frac{1}{n} r^{\frac{n}{2}} \hat{u}^n$$

Thus the integral can be dominated by a scale multiple of $r^{\frac{n}{2}+\frac{n}{2\omega}}$. (See also the discussion on pages 511 and 512 of Edmunds and Evans (1987).) The eigenvalue inequality follows.[23] □

Notice that as ω becomes large, we approach the eigenvalue divergence rate for the problem with compact support.

Mynbaev and Otel'baev (1988) produce approximation number bounds for the embedding of a weighted Sobolev space (constructed with weight function $w^{\frac{1}{2}}$ into $\mathcal{L}^2(leb)$. (See also Haroske (1995).) By inverting the square of the upper bound for the $(j+1)^{st}$ approximation number, we obtain the following lower bounds for the j^{th} eigenvalue.

[22]While Reed and Simon (1978) only prove this result for $n \geq 2$, they argue that the result extends to the $n = 1$ case.

[23]An analogous upper bound can be formed with an appropriate adjustment in the constant term. This shall not concern us since it is only the lower bound that we use in our subsequent analysis.

Claim 7.2. *(Mynbaev and Otel'baev (1988)) The eigenvalues of f_w^d satisfy the following.*

- *If $\omega > 1$, then $\delta_j \geq c^*(j+1)^{\frac{2}{n}}$*
- *If $0 < \omega < 1$, then $\delta_j \geq c^*(j+1)^{\frac{2\omega}{n}}$*
- *If $\omega = 1$, then $\delta_j \geq c^* \left(\frac{j+1}{\log(j+1)}\right)^{\frac{2}{n}}$*

for appropriate choices of c^*.

With the volatility weighting included in the form f_w^d, we obtain a faster rate of eigenvalue divergence. In contrast to Claim 7.1, we now have bounds that apply for $0 < \omega \leq 1$ and for $\omega > 1$ the divergence rate is the same as that for an unweighted form on a compact state space.[24]

7.3 Applications

In section 6 we constructed forms that were conveniently chosen lower bounds on the form $\hat{f}$. These forms exploited radial bounds on the density q and the diffusion matrix Σ:

$$\begin{aligned} q(x) &\geq \exp[-2\nu(|x|)] \\ \Sigma(x) &\geq \rho(|x|)^2 = \underline{c}(1+|x|^2)^{\beta}. \end{aligned}$$

7.3.1 Exponentially Thin Tails

Initially we constructed a potential function $\check{V}$ and a form:

$$\check{f}(\phi, \psi) = \int \phi(x)\psi(x)\check{V}(|x|) + \int \nabla\phi(x)\nabla\psi(x)\rho(|x|)^2$$

that is a lower bound on $\hat{f}$. Suppose that the density q has exponentially thin tails and that ν used in the density bound is given by

$$\nu = c_1(1+r^2)^{\frac{\xi}{2}} + c_0$$

for some $\xi > 0$ and some positive constants c_1 and c_0. In this case the potential function satisfies

$$\check{V} + \check{c} \geq \tilde{c}(1+r^2)^{\omega}$$

[24]The eigenvalue bounds delivered by Claim 7.1 and Claim 7.2 imply that the associated conditional expectation operators are Hilbert-Schmidt operators. An operator is Hilbert-Schmidt when the sum of squares of its eigenvalues are finite. Darolles, Florens, and Renault (1998) impose the Hilbert-Schmidt restriction in their analyses of principal components extracted from conditional expectation operators.

for some positive constants $\check{c}$ and $\tilde{c}$ and

$$\omega = \xi - 1 + \beta \tag{28}$$

Adding a constant to the potential function $\check{V}$ will cause the eigenvalues of the form to be increased by that same constant. Since the eigenvalues are increasing and diverging, we will not be concerned by translations. To apply Claim 7.1 we must have $\omega > \frac{1}{2}$. Notice from formula (28) that a thin tail (as reflected by a large ξ) and a high volatility elasticity (as reflected by a large β) imply faster rates of divergence of the eigenvalues.

To apply Claim 7.2 we may use this same inequalities, but we may have to employ a different weighting function with say parameter ω^*. In order that for scale multiples of the weighting function to be dominated by both a translation of the potential function $\check{V}$ and ρ^2, ω^* should satisfy:

$$\begin{aligned} \omega^* &= \min\{\omega, \beta\} \\ &= \min\{\xi - 1, 0\} + \beta. \end{aligned}$$

While ω^* is less than or equal to ω, when $\omega = \omega^*$, the Claim 7.2 eigenvalue bounds (eventually) exceed those given by Claim 7.1. It is of interest to know more generally which bounds are superior. There are four cases:

- $\beta = 0$. If $0 < \xi \leq 3/2$, Claims 7.1 and 7.2 are not applicable. If $\xi > 3/2$, Claim 7.1 is applicable but not Claim 7.2.

- $0 < \beta \leq 1/3$. If $0 < \xi \leq 3/2 - \beta$, Claim 7.1 is not applicable, but Claim 7.2 is applicable. Conversely, if $\xi > 3/2 - \beta$, then Claim 7.1 is applicable and provides superior eigenvalue bounds.

- $1/3 < \beta < 1$. If $0 < \xi < \frac{1+\beta^2-\beta}{1-\beta}$, the Claim 7.2 eigenvalue bounds are superior to those of Claim 7.1. (For some points in this region Claim 7.1 is not applicable.) Conversely, if $\xi > \frac{1+\beta^2-\beta}{1-\beta}$, the Claim 7.1 eigenvalues bounds are superior to those of Claim 7.2. Finally, if $\xi = \frac{1+\beta^2-\beta}{1-\beta}$ the bounds from the two claims are of the same order.

- $\beta \geq 1$ and $\xi > 0$. Claim 7.2 eigenvalue bounds always apply and dominate those of Claim 7.1 whenever the latter bounds are applicable.

The eigenvalues bounds implied by Claims 7.1 and 7.2 for the form $\hat{f}$ are depicted in Figure 6. We depict the bounds in four regions of the (β, ξ) parameter space. In two of the four regions, there is an explicit tradeoff between how thin the tail is of the density lower bound and how large the volatility elasticity is of the lower bound of the diffusion matrix. The eigenvalue bounds are the largest given in Claim 7.2 ($\delta_j \geq c^*(j+1)^{\frac{2}{n}}$) provided that either $0 < \xi \leq 1$ and $\beta > 2 - \xi \geq 1$, or $\xi \geq 1$ and $\beta > 1$. These bounds are of the same order as those for the case in which Σ and q are constant and the support of the state space is compact.

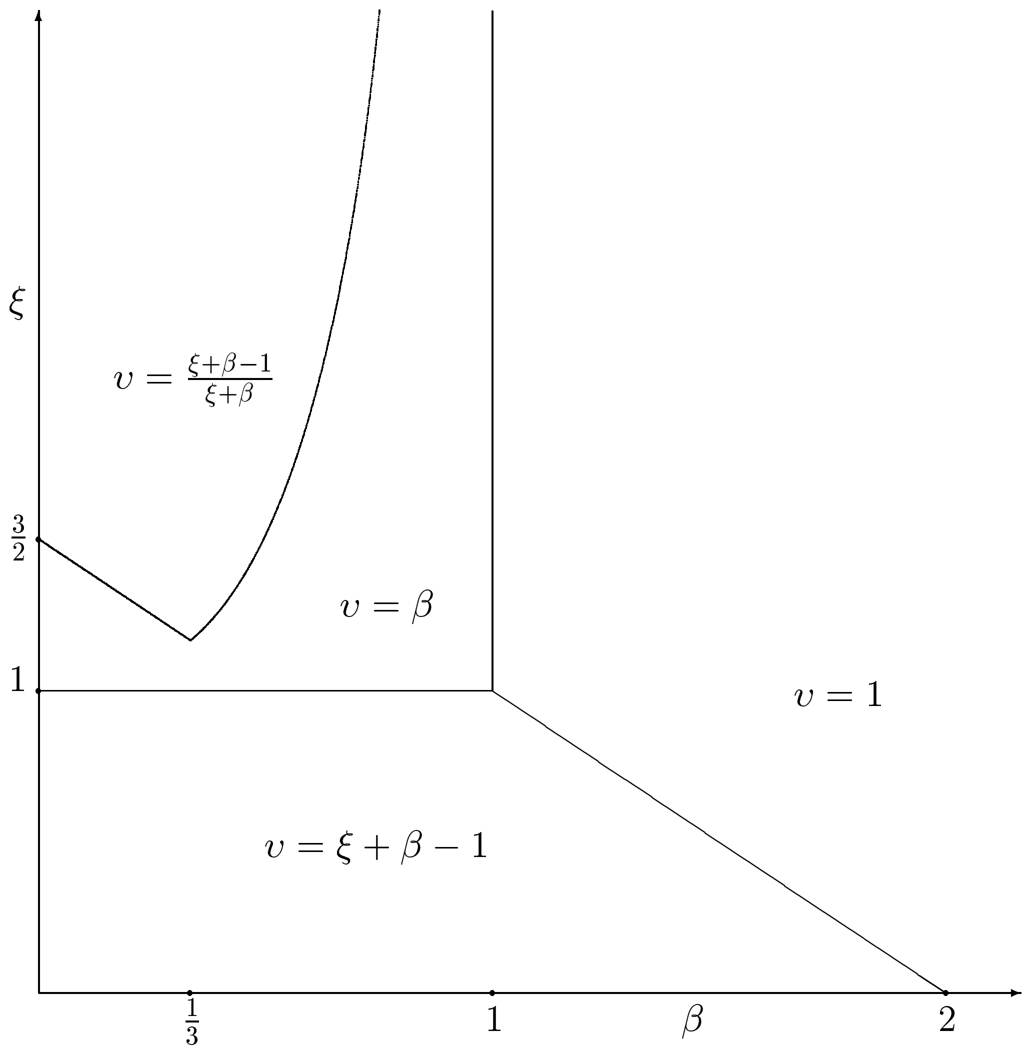

Figure 6: This figure displays the lower bounds on the divergence of the eigenvalues when the density has an exponentially thin tail. The eigenvalue bounds are of the form $c^*(j+1)^{\frac{2\upsilon}{n}}$ where n is the dimension of the underlying Markov process. The values for υ are depicted in each of four regions. These values depend explicitly on the density lower bound, parameterized by ξ, and on the volatility lower bound, parameterized by β. Claim 7.1 is used for the upper-left region and Claim 7.2 for the other three regions. Since υ is always less than or equal to one, the best bound is that given in the upper-right region.

Example 7.3. *Davies (1989) considers an example in which* $q(x) \propto \exp(-2|x|^\xi)$ *and* $\Sigma = I$. *The density tail in this example is essentially equivalent to our choice of* $\nu = c_1(1+r^2)^{\frac{\xi}{2}} + c_0$ *for the same value of* ξ. *Thus Claim 7.1 provides eigenvalue bounds as long as* $\xi > 3/2$. *When* $\xi > 2$, *Davies shows that the implied conditional expectation operators are bounded operators mapping* $\mathcal{L}^2$ *into* $\mathcal{L}^\infty$ *(conditional expectations are ultracontractive). As a consequence, the eigenfunctions are bounded and the transition density (defined relative to q) can be uniformly approximated by the eigenfunction expansion. (See Davies (1989) Example 4.7.2 on page 133 and Theorem 2.1.4 on pages 60 and 61.) Uniform approximation is often more tractable than* $\mathcal{L}^2$ *approximation once statistical sampling error is formally introduced into the analysis.*

From Theorem 4.7.3 of Davies (1989), the conditional expectation operators will fail to be ultracontractions in the region $0 < \xi < 2$ *for this example. Nevertheless, there will continue to exist* $\mathcal{L}^2$ *expansions of the conditional expectation operators in terms of eigenfunctions.*

7.3.2 Algebraic Tails

Recall that compactness Criterion 6.13 can be used when $\beta > 1$. This criterion is based on:

$$\check{V} + (1-\epsilon)\check{W}$$

for some $\epsilon > 0$ instead of just $\check{V}$. We may imitate the analysis in the case of low volatility growth by finding a positive number ω such that

$$\check{V} + (1-\epsilon)\check{W} + \check{c} \geq \tilde{c}(1+r^2)^\omega$$

for some positive constants $\check{c}$ and $\tilde{c}$.

As a special case, we reconsider Example 6.17 in which the density q has thick tails. Specifically the function ν used to bound the density is given by:

$$\nu = \frac{\gamma}{2}\log(1+r^2) + c^*$$

for some $\gamma > \frac{n}{2}$. Previously we argued that if $\beta < \gamma - \frac{n}{2} + 1$, then $\check{V} + (1-\epsilon)\check{W}$ will behave like a constant multiple of $(1+r^2)^{\beta-1}$ for large r. Thus Claim 7.2 can be applied with $\omega = \beta - 1$. Notice that when $\beta > 3/2$, Claim 7.1 also can be applied with $\omega = \beta - 1$. The Claim 7.2 eigenvalue bounds are sharper, however.

8 Extensions

In this paper we have used a modeling device of Beurling and Deny (1958) and Fukushima (1971) to establish the existence of and interpret functional principal components extracted by maximizing variation subject to smoothness constraints. The smoothness restriction is linked explicitly to the diffusion matrix and the resulting principal components have nice time series properties. They maximize the long run variation holding constant the overall variation, and behave like scalar autoregressions with heteroskedastic innovations. This latter property gives us a way to extract testable conditional moment implications for multivariate diffusion models that are stationary and time reversible. These conditional moment implications can form the basis of parametric estimation as suggested by Kessler and Sorensen (1999) or they can provide the basis of parametric and semiparametric testing. In this paper, however, we have not confronted formally issues of statistical inference.

Our assumption that the density q is integrable is not required in the depiction of symmetric diffusions using quadratic forms. A rich class of diffusion processes, including many nonstationary diffusions, can be constructed by specifying the pair (q, Σ). In fact multivariate Brownian is a symmetric Markov process with a constant Σ matrix and a density $q \equiv 1$. Fukushima, Oshima, and Takeda (1994) provide conditions on Σ and q that insure the conservation of probabilities (even when q is not integrable) and that the implied Markov process is recurrent. Thus this same modeling approach is likely to be valuable beyond the analysis of functional principal component analysis.

Our application of principal component analysis is tied to the Markov diffusion model. The diffusion matrix penalizes the derivatives. Symmetric markov jump models can also be constructed using quadratic forms. The counterparts to principal components can still be defined, but the extraction method would be different. The gradient quadratic form used to enforce smoothness would be replaced by an alternative quadratic form with a different weighting scheme. The treatment of existence would, of course, have to be changed.

A Computations

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

Consider the quadratic form in the level. For a given ϕ, define ψ as

$$\psi(r,\theta) = \phi[r\cos(\theta), r\sin(\theta)],$$

and define ψ^* analogously from ϕ^*. Then

$$\int_{\mathbb{R}^2} \phi\phi^* q = \frac{1}{2\pi}\int_0^\infty \int_{-\pi}^{\pi} \psi\psi^* d\theta q^*(r) dr \tag{29}$$

where

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

Consider next the quadratic form for the derivatives. Note that

$$\nabla\phi = \frac{1}{r}\begin{bmatrix} r\cos(\theta) & -\sin(\theta) \\ r\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \psi_r \\ \psi_\theta \end{bmatrix}.$$

Thus

$$(\nabla\phi)\cdot(\nabla\phi^*) = (\psi_r\psi_r^*) + \frac{1}{r^2}(\psi_\theta\psi_\theta^*).$$

We may evaluate the form:

$$\int_{\mathbb{R}^2} (\nabla\phi)\cdot(\nabla\phi^*) q = \frac{1}{2\pi}\int_0^\infty \int_{-\pi}^{\pi} \psi_r\psi_r^* d\theta q^*(r) dr + \frac{1}{2\pi}\int_0^\infty \frac{1}{r^2}\int_{-\pi}^{\pi} \psi_\theta\psi_\theta^* d\theta q^*(r) dr. \tag{30}$$

In our calculations we use basis functions of the form: $p(r)\cos(k\theta)$ and $p(r)sin(k\theta)$ where p is a scalar Hermite polynomial in r and k is a nonnegative integer. We exploit the orthogonality of $\cos(k\theta)$ and $\sin(k\theta)$ for a given k and the orthogonality of $\cos(k\theta)$ with $\sin(\ell\theta)$ and $\cos(\ell\theta)$ for k different from ℓ 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: $\psi(r,\theta) = p(r)\cos(k\theta)$ and $\psi^*(r,\theta) = p^*(r)\cos(k\theta)$ for some strictly positive integer k. Recall that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(k\theta)^2 d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sin(k\theta)^2 d\theta = \frac{1}{2}.$$

Thus the form in (29) is:

$$\frac{1}{2\pi}\int_0^\infty \int_{-\pi,\pi} \psi\psi^* d\theta q^*(r) dr = \frac{1}{2}\int_0^\infty p(r)p^*(r)q^*(r) dr,$$

and the form in (30) is:

$$\frac{1}{2\pi}\int_0^\infty \int_{-\pi}^{\pi} \psi_r \psi_r^* d\theta q^*(r) dr \quad + \quad \frac{1}{2\pi}\int_0^\infty \frac{1}{r^2}\int_{-\pi}^{\pi} \psi_\theta \psi_\theta^* d\theta q^*(r) dr$$
$$= \quad \frac{1}{2}\int_0^\infty \left[\frac{k^2}{r^2} p(r) p^*(r) + p'(r) p^{*\prime}(r)\right] q^*(r) dr.$$

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 $p(r)\cos(k\theta)$ and $p(r)\sin(k\theta)$. 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 a Cholesky decomposition, and computed the spectral (eigenvalue-eigenvector) decomposition of $A'^{-1}WA^{-1}$ using the Schur decomposition to construct the principal components.

B Additional Proofs

In this appendix we supply the proofs for some of our results.

Consider first Claim 4.2.

Claim B.1. *$\bar{F}$ is an extension of $-L$.*

Proof. Suppose that $\psi \in \mathcal{L}^2$ and $\phi \in \mathcal{C}_K^2$. By imitating our earlier demonstration of the symmetry of L, and exploiting the fact that $\nabla\phi$ is a $\mathcal{C}^1$ function with compact support:

$$\bar{f}(\bar{G}_\alpha\psi, \phi) =< \bar{G}_\alpha\psi, -L\phi > .$$

From the definition of $\bar{G}_\alpha$, we also have that

$$\bar{f}(\bar{G}_\alpha\psi, \phi) =< \psi, \phi > -\alpha < \bar{G}_\alpha\psi, \phi > .$$

Thus

$$
\begin{aligned}
< \bar{G}_\alpha \psi, (\alpha I - L)\phi > &= < (\alpha I - L^*)\bar{G}_\alpha \psi, \phi > \\
&=< \phi, \psi >
\end{aligned}
$$

where L^* is the adjoint of L. Consequently,

$$(\alpha I - L^*)\bar{G}_\alpha \psi = \psi,$$

which guarantees that $\bar{G}_\alpha \psi$ is in the domain of $-L^*$. Therefore, $-L^*$ is an extension of $\bar{F}$, and hence $\bar{F}$ is an extension of $-L$. □

Consider next Proposition 6.10.

Proposition B.2. *Suppose that a) Assumption 6.7 is satisfied; b) Assumption 6.8 is satisfied for $\beta \leq 1$; c) the limits $\lim_{r\to\infty} \rho\nu'$ and $\lim_{r\to\infty} \check{V}(r)$ are well defined. Then $\check{V}$ converges to $+\infty$ at ∞ if, and only if $\rho\nu'$ converges to $+\infty$ at the same point.*

Proof. First, suppose that $\rho\nu'$ converges to $+\infty$ at ∞. Since $\beta \leq 1$,

$$\lim_{r\to\infty} \frac{(n-1)\nu'/r}{(\nu')^2} = \lim_{r\to\infty} \frac{(n-1)}{\nu' r} = 0,$$

and

$$\lim_{r\to\infty} \frac{(\rho'/\rho)\nu'}{(\nu')^2} = \lim_{r\to\infty} \frac{\beta}{\nu' r} = 0.$$

Thus in studying the tail behavior of $\check{V}$ we may ignore the terms $(n-1)\frac{\nu'}{r}$ and $2\frac{\rho'}{\rho}\nu'$. Moreover,

$$\lim_{r\to\infty} \check{V} = \lim_{r\to\infty} \rho^2[-\nu'' + (\nu')^2]. \tag{31}$$

Suppose that limits in (31) are not $+\infty$. Then dividing the right-hand side of (31) by $\rho^2(\nu')^2$, it follows

$$\lim_{r\to\infty} \frac{-\nu''}{(\nu')^2} \geq -1.$$

Since

$$\frac{-\nu''}{(\nu')^2} = \left(\frac{1}{\nu'}\right)'.$$

We have the inequality,

$$\frac{1}{\nu'(r)} \geq c_o - \frac{1}{2} r$$

for some constant c_o and r sufficiently large. Thus

$$r\nu'(r) \leq \frac{2r}{2c_o - r}.$$

The right-hand side has a finite limit as r goes to ∞, which contradicts the requirement that

$$\lim_{r\to\infty} \rho\nu' = +\infty.$$

Therefore, $\rho^2[-\nu'' + (\nu')^2]$ and hence the potential $\check{V}$ must converge to $+\infty$ as r tends to ∞.

Next suppose that $\rho\nu'$ has a finite limit at ∞ that is strictly positive. We will show that $\check{V}$ has a finite limit at ∞. Note that

$$\lim_{r\to+\infty} r\nu'$$

has a well defined strictly positive limit, possibly $+\infty$. As a consequence,

$$\lim_{r\to\infty} \frac{(n-1)\nu'/r}{(\nu')^2} = \lim_{r\to\infty} \frac{(n-1)}{\nu' r}$$

is well defined and finite as is

$$\lim_{r\to\infty} \frac{(\rho'/\rho)\nu'}{(\nu')^2} = \lim_{r\to\infty} \frac{\beta}{\nu' r}.$$

The remaining term in $\check{V}$ that could diverge is $-\rho^2\nu''$. Suppose to the contrary that

$$\lim_{r\to\infty} -\rho^2\nu'' = +\infty.$$

Then

$$\lim_{r\to\infty} -\frac{\nu''}{(\nu')^2} = +\infty$$

which implies that for any $c_1 > 0$, there exists an c_2 such that

$$\frac{1}{\nu'} \geq c_1 r + c_2$$

for large r. This contradicts the integrability of $\exp[-2\nu(|x|)]$. Therefore,

$$\lim_{r\to\infty} -\frac{\nu''}{(\nu')^2}$$

cannot be infinite.

Finally, suppose that $\rho\nu'$ converges to a nonpositive number. Then for any $c_1 > 0$ for sufficiently large r

$$\nu' \leq c_1 r^{-\beta}.$$

Since $\beta \leq 1$ and c_1 can be made arbitrarily small when $\beta = 1$, $\exp[-2\nu(|x|)]$ is not integrable. Thus we again have a contradiction. □

References

Azencott, R. (1974). Behavior of diffusion semi-groups at infinity. *Bull. Soc. Math. France 102*, 193–240.

Banon, G. (1978). Nonparametric identification of diffusions. *SIAM Journal on Control 16*, 380–395.

Beurling, A. and J. Deny (1958). Espaces de dirichlet i, le cas elementaire. *Acta Math. 99*, 203–224.

Bhattacharya, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for markov processes. *Zeitschrift fur Wahrsheinlichkeits-theorie und verwandte Gebiete 60*, 185–201.

Brezis, H. (1983). *Analyse Fonctionelle.* Paris: Masson.

Chen, X., L. Hansen, and J. Scheinkman (1998). Shape-preserving estimation of diffusions. Mimeo.

Cobb, L. P., P. Kopstein, and N. Chen (1983). Estimation and moment recursion relations for multimodal distributions in the exponential family. *Journal of the American Statistical Association 78*, 124–130.

Conley, T., L. P. Hansen, E. Luttmer, and J. Scheinkman (1998). Short-term interest rates as subordinated diffusions. *Review of Financial Studies 10*, 525–577.

Darolles, S., J. P. Florens, and C. Gourieroux (2000). Kernel based nonlinear canonical analysis and time reversibility. Mimeo.

Darolles, S., J. P. Florens, and E. Renault (1998). Nonlinear principal components and inference on a conditional expectation operator. Conference Presentation in the 1998 Paris-Berlin Seminar, Garchy.

Davies, E. B. (1985). l^1 properties of second order operators. *Bulleton of the London Mathematical Society 17*, 417–436.

Davies, E. B. (1989). *Heat Kernels and Spectral Theory.* Cambridge University Press.

Davies, E. B. (1995). *Spectral Theory and Differential Operators.* Cambridge University Press.

Demoura, S. (1998). The nonparametric estimation of the expected value operator. Mimeo.

Douxious, J. and A. Pousse (1975). Une extension de l'analyse canonique. queslques applications. *Ann. Inst. H. Poincare 11*, 355–379.

Edmunds, D. E. and W. Evans (1987). *Spectral Theory and Differential Operators.* Oxford University Press.

Florens, J. P., E. Renault, and N. Touzi (1998). Testing for embeddability by stationary reversible continuous-time markov processes. *Econometric Theory 69*, 744–69.

Fukushima, M. (1971). Dirichlet spaces and strong markov processes. *Transactions of the American Mathematics Society 162*, 185–224.

Fukushima, M., Y. Oshima, and M. Takeda (1994). *Dirichlet Forms and Symmetric Markov Processes.* Walter de Gruyter.

Hansen, L. P. and J. Scheinkman (1995). Back to the future. *Econometrica 63*, 767–804.

Hansen, L. P., J. Scheinkman, and N. Touzi (1998). Spectral methods for identifying scalar diffusions. *Journal of Econometrics 86*, 1–32.

Haroske, D. (1995). Approximation numbers in some weighted function spaces. *Journal of Approximation Theory 83*, 104–136.

Hasminskii, R. Z. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the solution of the cauchy problem for parabolic equations. *Theory, Probability and Applications 5*, 179–196.

Kessler, M. and M. Sorensen (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli 5*, 299–314.

Mynbaev, K. and M. Otel'baev (1988). *Weighted Functional Spaces and Differential Operator Spectrum.* Nauka.

Pang, M. M. H. (1996). l^1 and l^2 properties of a class of singular second order elliptic operators on $\mathbb{R}^n$ with measurable coefficients. *Journal of Differential Equations 129*, 1–17.

Reed, M. and B. Simon (1978). *Methods of Modern Mathematical Physics IV: Analysis of Operators*. Academic Press.

Salinelli, E. (1998). Nonlinear principal components i: Absolutely continuous variables. *Annals of Statistics 26*, 596–616.

Stramer, O., P. J. Brockwell, and R. Tweedie (1996). Continuous-time threshold ar(1) models. *Advances in Applied Probability 28*, 728–746.

Stramer, O., R. Tweedie, and P. J. Brockwell (1996). Existence and stability of continuous time threshold arma models. *Statistica Sinica 6*, 715–732.

Stroock, D. W. and S. R. S. Varadhan (1979). *Multidimensional Diffusion Processes*. Springer.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical Systems Approach*. Oxford University Press.