

**PRINCIPAL COMPONENTS AND LONG RUN IMPLICATIONS
OF MULTIVARIATE DIFFUSIONS**

By

Xiaohong Chen, Lars Peter Hansen, and José Scheinkman

April 2009

COWLES FOUNDATION DISCUSSION PAPER NO. 1694



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281**

<http://cowles.econ.yale.edu/>

Principal Components and Long Run Implications of Multivariate Diffusions

Xiaohong Chen
Department of Economics
New York University
email: xiaohong.chen@nyu.edu

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 ¹

First Draft: October 2000
Current Draft: December 2006

¹We thank Anna Aslanyan, Henri Berestycki, E. Brian Davies, Nan Li, Oliver Linton, Nour Meddahi, 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.

Abstract

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

Key words and phrases. Nonlinear principal components, discrete spectrum, eigenvalue decay rates, multivariate diffusion, quadratic form, conditional expectations operator.

1 Introduction

Principal components are functions of the data that capture maximal variation in some sense. Often they are restricted to be linear functions of the underlying data. In this paper we study the extraction of nonlinear principal components using information encoded in the density of the data. Formally, the principal components maximize variation subject to orthogonality and smoothness constraints where smoothness constraints are enforced by the quadratic form f expressed in terms of the gradients of functions. Specifically, the quadratic form is

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

where Σ is a state-dependent positive-definite matrix, q is the population density of the data, ∇ denotes the (weak) gradient operator and Ω is the state space.

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

Previously Salinelli (1998) defined nonlinear principal components for absolutely continuous random variables and characterized these principal components as eigenfunctions of a self-adjoint, differential operator. Our departure from Salinelli (1998) is substantial. For Salinelli, the matrix Σ is state independent, the state space Ω is compact and the density q is bounded above and below for the bulk of his analysis. Our interest in probability densities q that do not have compact support, including densities with algebraic tails leads us naturally to consider a more general class of smoothness penalties. By allowing for a more flexible specification for Σ and q , we entertain a larger class of smoothness constraints *vis a vis* Salinelli (1998) with explicit links to the data generation. Establishing the existence of principal components in this setup is no longer routine.

Salinelli (1998) assumes the data generation process is independent and identically distributed. While our analysis is applicable to such an environment, we also explore the case in which data are generated by a stationary diffusion process. By considering such processes, we give a specific interpretation of the matrix Σ used to enforce smoothness. It is the local covariance or diffusion matrix. With this interpretation, the principal components extracted with smoothness penalties are ordered by the ratio of their long run variation to the overall variation. That is, principal components that capture variation subject to smoothness constraints also display low frequency variation due to their high persistence.

In this paper we do the following:

1. Formulate the nonlinear principal component extraction to include state-dependence in the smoothness constraint and state spaces that have infinite Lebesgue measure.

2. Give sufficient conditions for the existence of these principal components and provide bounds on the incremental importance of these components when used in approximation.
3. Provide a reversible Markov diffusion process for the data generation that supports the principal component extraction method and generates testable implications.
4. Show that these sufficient conditions imply the existence of countably many eigenvalues and eigenfunctions for a more general class of Markov diffusion processes.

The rest of the paper is organized as follows. In Section 2, we first define principal components as functions that maximize variation subject to orthogonality conditions and smoothness bounds given by the quadratic form f . These principal components are eigenfunctions of the form f . We then characterize a family of optimization problems that principal components must solve and show that these functions also solve a problem of obtaining the “best” finite-dimensional approximations in a least-square sense. The quality of these approximations are related to the eigenvalues associated with the eigenfunctions. If the constraints of the optimization problem that yield the principal components satisfy a compactness criteria, principal components exist. Section 3 is dedicated to deriving primitive conditions that imply the compactness criteria. In Section 4 we specialize the conditions that imply compactness to the case when radial symmetry holds and discuss four parameterized examples. We then present bounds on the rate of divergence of the eigenvalues that are associated with the principal components. These bounds, in turn, are bounds on the rate of convergence of the approximation errors that obtain when we use principal components in approximation. The results in Section 5 relate the principal components to eigenfunctions of conditional expectations operators associated with a stationary Markov process $\{x_t\}$ defined using the diffusion matrix Σ and the stationary density q . Given an eigenfunction ψ , the process $\{\psi(x_t)\}$ behaves as a scalar autoregression. Thus the eigenfunction we obtain satisfy testable implications when the data is generated by a Markov process. Section 6 gives some concluding remarks and discusses applications our results. The appendices contain some proofs and computations associated with examples.

2 Principal Components

To define a functional notion of principal components we require two quadratic forms. We start with an open connected $\Omega \subseteq \mathbb{R}^n$. Let q be a probability density on Ω with respect to Lebesgue measure. The implied probability distribution is the population counterpart to the empirical distribution of the data. Let L^2 denote the space of Borel measurable square integrable functions with respect to population probability distribution. The L^2 inner product (denoted $\langle \cdot, \cdot \rangle$) is one of the two forms of interest. We use the corresponding norm to define an approximation criterion.

The second form is used to measure smoothness. Consider (quadratic) form f_o defined on C_K^2 , the space of twice continuously differentiable functions with compact support in Ω ,

that can be parameterized in terms of the density q and a positive definite matrix Σ that can depend on the state:

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

where

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

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

Assumption 2.2. : Σ is a continuously differentiable, positive definite matrix function on Ω .¹

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

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

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

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{H} we write $\nabla \phi = g$.² For any pair of functions ψ and ϕ in \bar{H} we define:

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

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

2.1 Initial Construction

Principal components maximize variation subject to smoothness constraints. In our generalization these principal components are defined as follows.³

¹Assumptions 2.1 and 2.2 restrict the density q and the 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 $L^2(\text{Lebesgue})$), weakly differentiable. (See Theorem 1.2.5.)

²Notice that \bar{H} is constructed exactly as a weighted Sobolev space except that instead of requiring that $g \in L^2$, we require that $\Lambda g \in L^2$ where Λ is the square root of Σ . Also we use C_K^1 test functions. One can show, using mollifiers, that allowing for this larger set of test functions is equivalent to using the more usual set of test functions, C_K^∞ (see Brezis (1983) Remark 1, page 150.)

³There are other functional principal component constructions. For instance, Dauxois and Pousse (1975) and Dauxois and Nkiet (1998) 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.

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

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

subject to

$$\begin{aligned} f(\phi, \phi) &= 1 \\ \langle \psi_s, \phi \rangle &= 0, s = 0, \dots, j-1 \end{aligned}$$

where ψ_0 is initialized to be the constant function one.

There are two differences between our proposed extraction and that of Salinelli (1998). First, Salinelli (1998) assumes that Σ is state independent. To accommodate a rich class of densities, we allow Σ to be state dependent. Second, Salinelli (1998) assumes that the data density q has finite Lebesgue measure and is bounded away from zero. We allow the Lebesgue measure of the state space to be infinite, and accordingly the density q is no longer assumed to be bounded from below.

As we will see, principal components are eigenfunctions of the quadratic forms f .

Definition 2.2. An eigenfunction ψ of the quadratic form f satisfies:

$$f(\phi, \psi) = \delta \langle \phi, \psi \rangle \tag{2}$$

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

Since f is positive semidefinite, δ must be nonnegative. The principal components extracted in the manner given in (2.1) have eigenvalues δ_j that increase with j . If we renormalize the eigenfunctions to have a unit second moment, the principal components will be ordered by their smoothness as measured by $\delta = f(\psi, \psi)$.

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

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

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

$$f(\phi, \psi) = \sum_{j=0}^{\infty} \delta_j \frac{\langle \phi, \psi_j \rangle \langle \psi, \psi_j \rangle}{\langle \psi_j, \psi_j \rangle^2}. \tag{3}$$

The following example is used to illustrate principal components.

Example 2.1. Let $\Sigma = I$ and

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

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

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

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

In addition, there are principal components that are not constant on circles. These principal components come in pairs. The principal component extraction will identify a two-dimensional space rather than the more familiar one-dimensional extraction. In particular there will be two orthogonal principal components 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 B. These principal components are scaled to have unit variance and are ordered by their smoothness. Principal components one and two come in a symmetric-antisymmetric pair and are shown in Figures 1 and 2. These functions are almost linear inside a circle of radius two. Beyond this circle the slope increases.

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

2.2 Benchmark Optimization Problem

Given a density q , principal components solve finite dimensional approximation problems. We will show that, in a specific sense, they can be used to build best approximating finite dimensional spaces. Let H be a closed in L^2 linear subspace of L^2 . Prior to establishing this result, consider the optimization problem:

Problem 2.1.

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

subject to

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

for some $\theta > 0$.

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

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

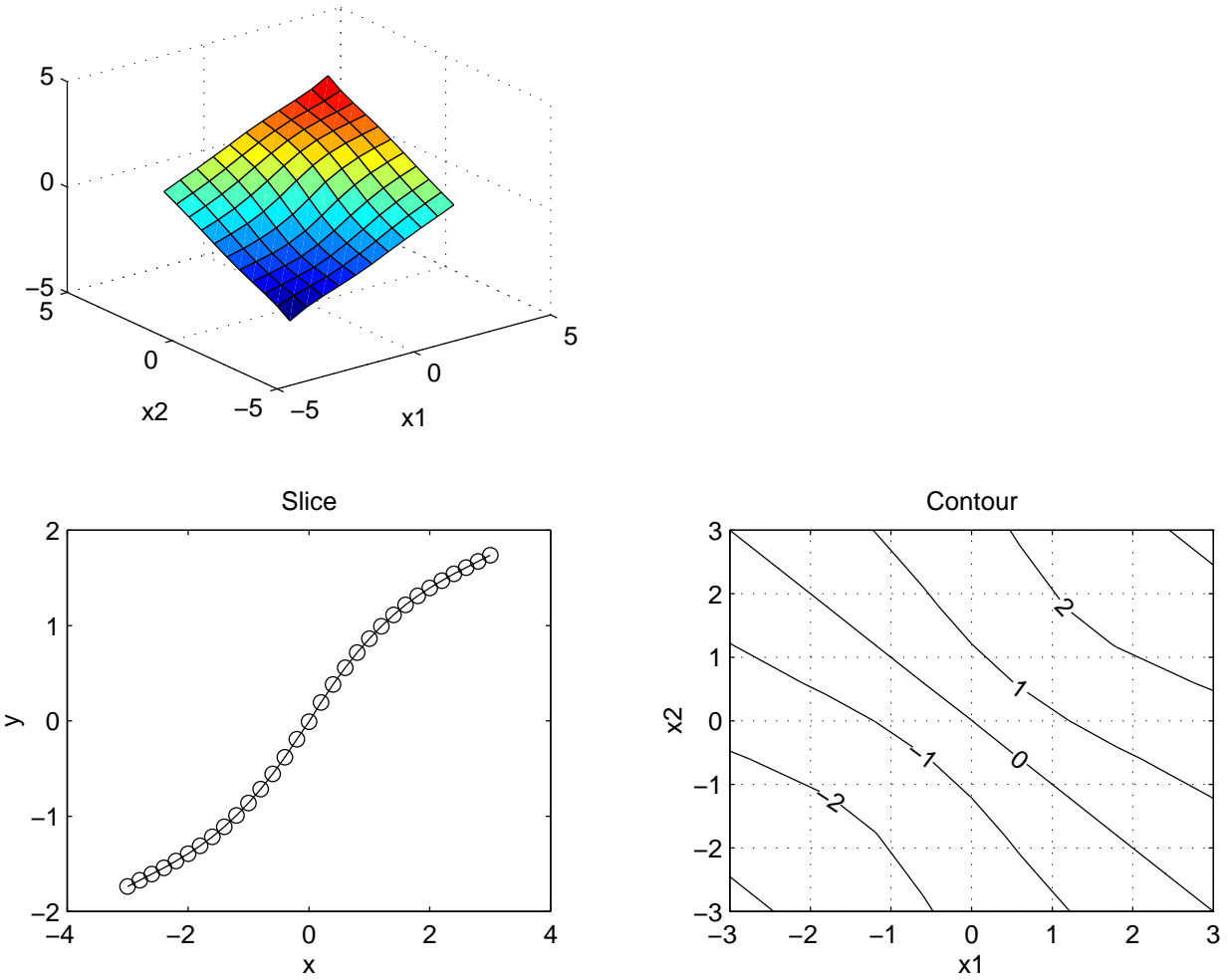


Figure 1: This figure displays the symmetric principal component in the first symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the other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

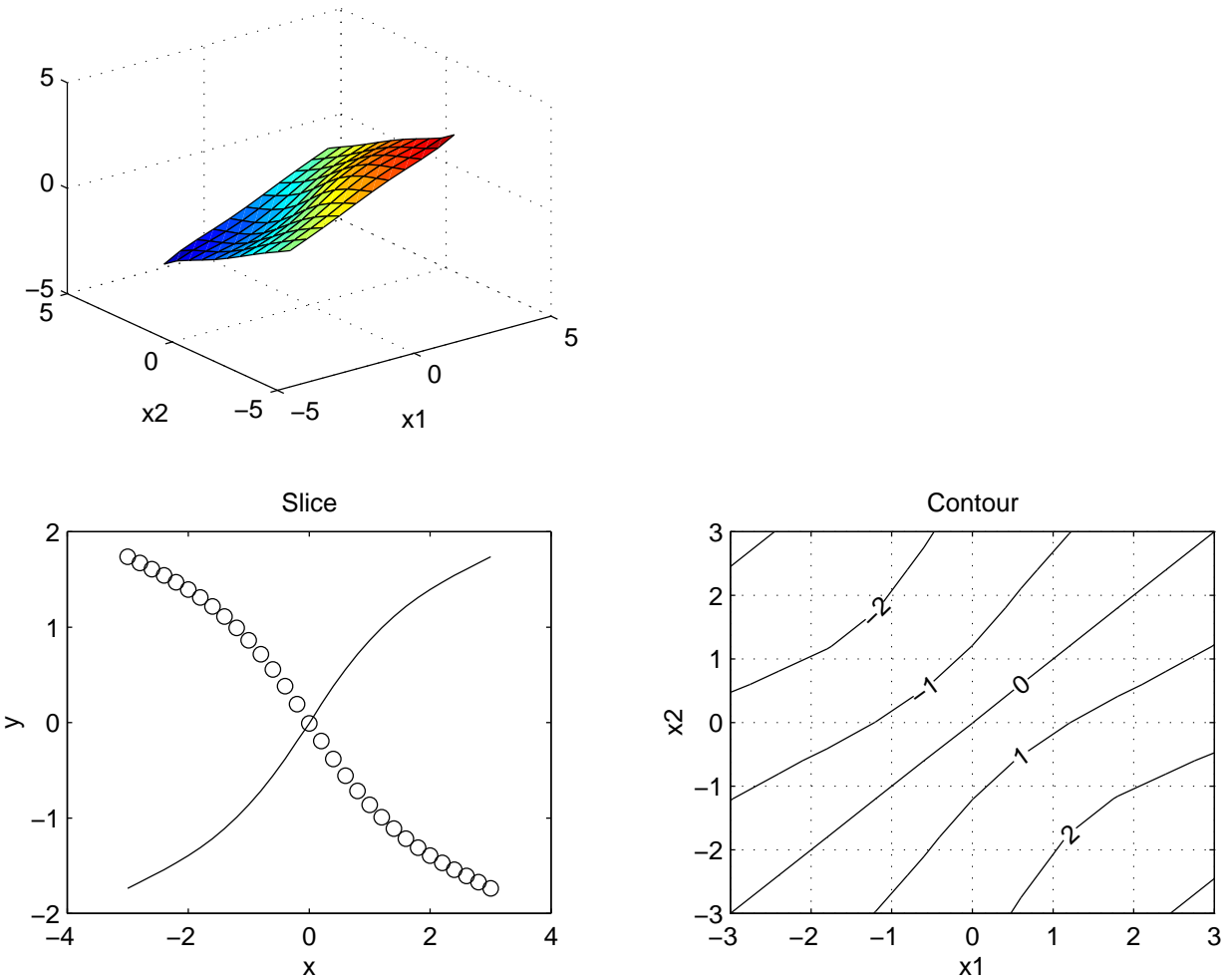


Figure 2: This figure displays the anti-symmetric principal component in the first symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the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

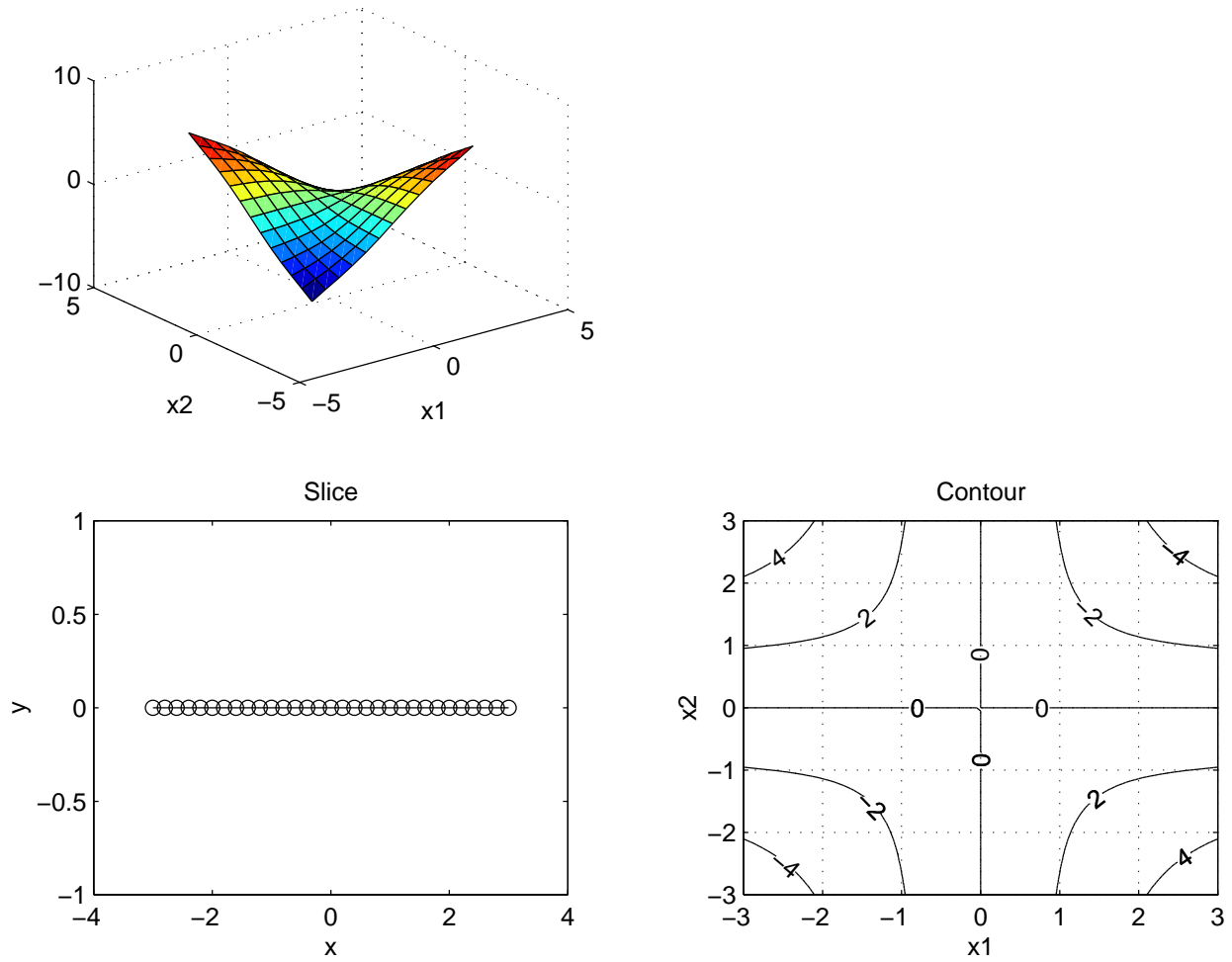


Figure 3: This figure displays the symmetric principal component in the second symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero, and the other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

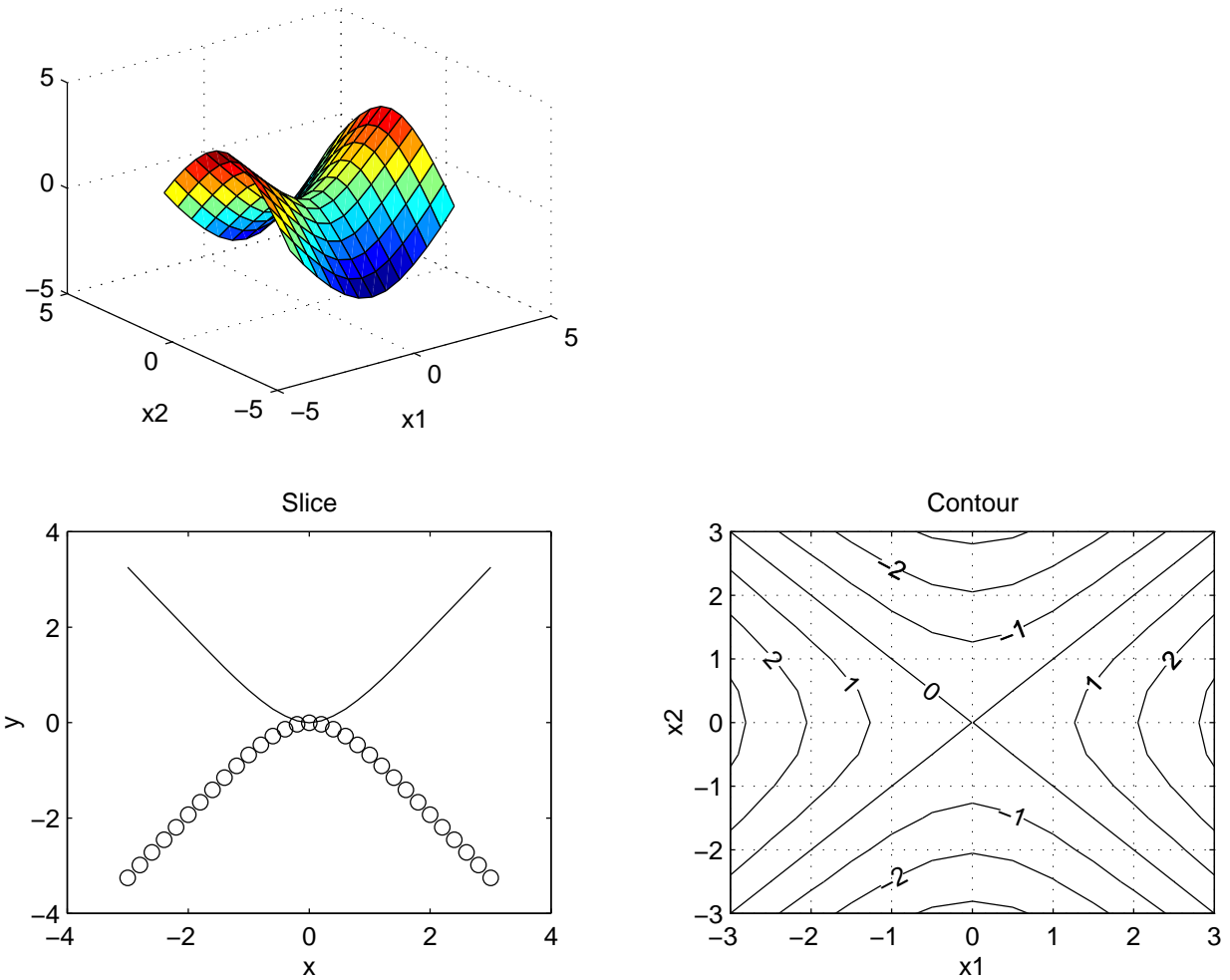


Figure 4: This figure displays the anti-symmetric principal component in the second symmetric-antisymmetric pair. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at zero and the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

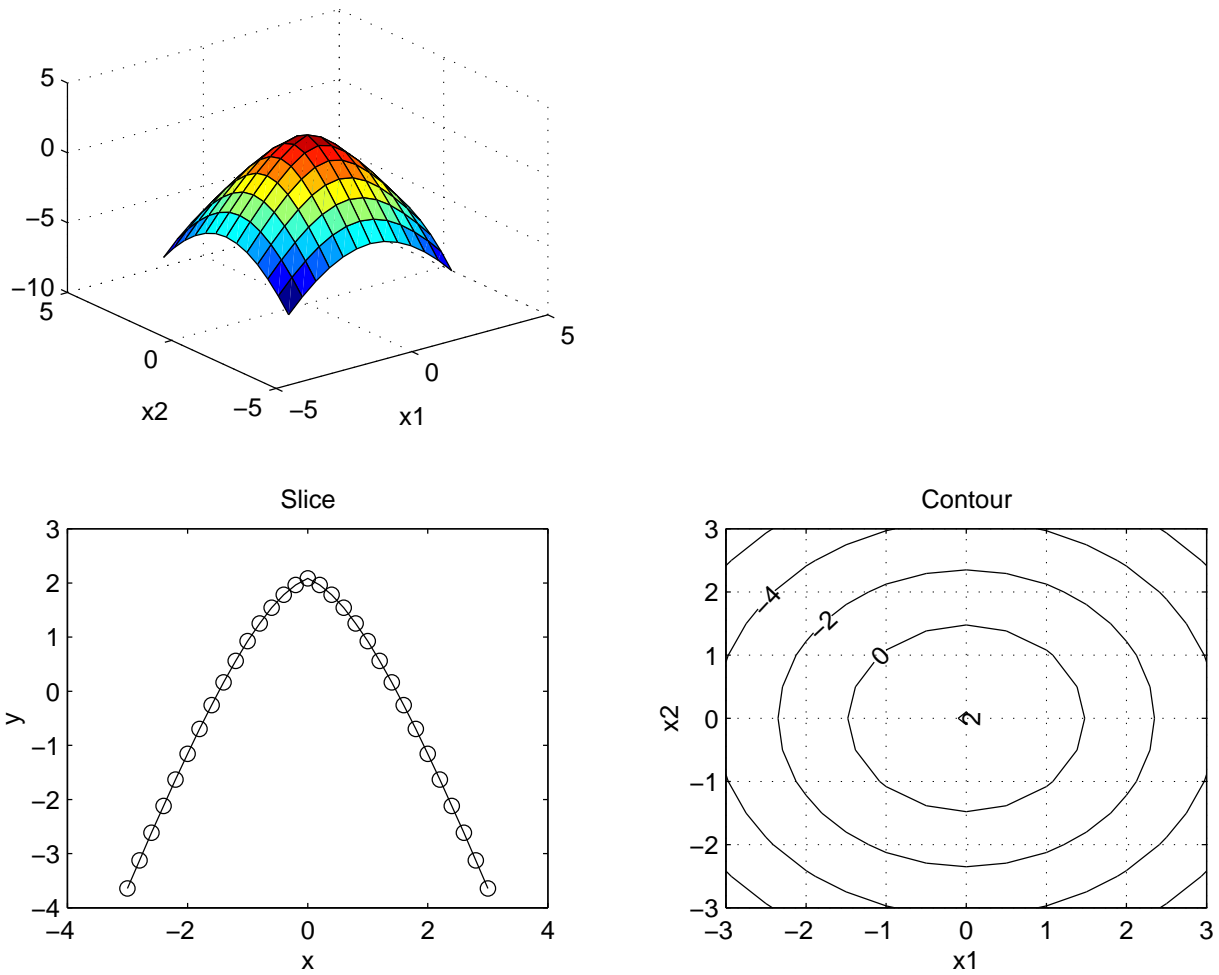


Figure 5: This figure displays the fifth principal component. This principal component is constant on circles centered at zero. The upper-left panel gives a three-dimensional plot of the principal component. The lower-left block gives two slices of the function. One slice fixes the first coordinate at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

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

We provide sufficient conditions for this compactness criterion in the next section. A necessary condition is that ϕ satisfies an eigenvalue problem:

Claim 2.1. *A solution ϕ^* to Problem 2.1 will also solve the eigenvalue problem:*

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

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

Proof. Let ϕ^* solve Problem 2.1. Then for each $\psi \in H$ and $r \in \mathbb{R}^+$ small

$$\frac{\langle \phi^* + r\psi, \phi^* + r\psi \rangle}{\langle \phi^* + r\psi, \phi^* + r\psi \rangle + \theta f(\phi^* + r\psi, \phi^* + r\psi)} \leq \lambda \doteq \frac{\langle \phi^*, \phi^* \rangle}{\langle \phi^*, \phi^* \rangle + \theta f(\phi^*, \phi^*)}.$$

That is:

$$2r \langle \phi^*, \psi \rangle + r^2 \langle \psi, \psi \rangle \leq \lambda [2r \langle \phi^*, \psi \rangle + r^2 \langle \psi, \psi \rangle + 2r\theta f(\phi^*, \psi) + r^2\theta f(\psi, \psi)].$$

Dividing by $2r$ and taking $\lim_{r \rightarrow 0}$

$$\langle \phi^*, \psi \rangle \leq \lambda [\langle \phi^*, \psi \rangle + \theta f(\phi^*, \psi)].$$

Since we could use also $-\psi$,

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

for each $\psi \in H$. □

2.3 Recursive approximation

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

Initially solve Problem 2.1 for $H = L^2$, select a solution ϕ_0 and denote the maximized objective as λ_0 . In fact $\phi_0 = 1$ and $\lambda_0 = 1$. Inductively, given $\phi_0, \phi_1, \dots, \phi_{j-1}$, form H_{j-1} as the j dimensional space generated by these j solutions constructed recursively. Let H_{j-1}^\perp denote the space of all elements of L^2 that are orthogonal to these j solutions and hence

orthogonal to H_{j-1} . Solve Problem 2.1 for $H = H_{j-1}^\perp$, select a solution ϕ_j , and form λ_j as the maximized value. The sequence $\{\lambda_j : j = 0, 1, \dots\}$ is decreasing because we are omitting components of the constraint set for the maximization problem as j increases.

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

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

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

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

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

$$\psi = \sum_{j=0}^N r_j \phi_j.$$

The objective can be expressed as:

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

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

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

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

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

Claim 2.3.

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

Proof. Write ψ as:

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

where φ is in H_{N-1}^\perp . Write:

$$Proj(\psi|H_{N-1}) = \sum_{j=0}^{N-1} r_j \phi_j$$

Using this decomposition, the objective can be written as:

$$\langle \varphi, \varphi \rangle,$$

and the constraint set can be written as:

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

because $\phi_1, \phi_2, \dots, \phi_{N-1}, \varphi$ are orthogonal, and $f(\phi_j, \varphi) = f(\phi_j, \phi_\ell) = 0$ for $j = 0, \dots, N-1$ and $\ell = j+1, j+2, \dots, N-1$. To maximize the objective, the coefficients r_j 's are set to zero and φ is chosen by solving Problem 2.1 for $H = H_{N-1}^\perp$. The conclusion follows. \square

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

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

Remark 2.2. *This analysis can be seen as a special case of the finite-dimensional approximation of linear operators. By using $\langle \phi, \psi \rangle + \theta f(\phi, \psi)$ as an inner product, we may construct a Hilbert space of functions. The embedding operator is the identity operator mapping this Hilbert space into L^2 . Approximation numbers from operator theory define formally the ability to approximate a linear operator such as this embedding operator by a best possible finite-dimensional operator (e.g. see Edmunds and Evans (1987), page 53). For this operator a solution is the least squares projection onto a finite number of the functions whose recursive construction we just characterized. The resulting approximation numbers $\{a_N : N \geq 0\}$ are given by $a_N = (\lambda_N)^{1/2}$.⁴*

2.4 Principal Components Revisited

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

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

⁴Our proof of Claim 2.2 follows closely that of Theorem 5.10 (page 91) in Edmunds and Evans (1987). It is included here for sake of completeness.

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

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

provided that H is orthogonal to all constant functions. Thus a scaled version of ϕ^* solves:

Problem 2.2.

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

subject to:

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

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

From Claim 2.1 we know that ϕ^* satisfies:

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

for $\psi \in H$. Rearranging terms,

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

where

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

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

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

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

3 Existence

In section 2.2 we noted that the compactness Criterion 2.1 implies the existence of principal components of the form f . We now consider more primitive sufficient conditions that imply this criterion.

This section is organized as follows. We first review the existence condition used by Salinelli (1998) and the one derived by Hansen et al. (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 the Lebesgue measure. This transformation allows us to apply known results for forms built using Lebesgue measure. Second, we study forms that are simpler but dominated by f . When the dominated forms satisfy Criterion 2.1 the same can be said of f .

3.1 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 derivative penalty matrix Σ is uniformly nonsingular.⁵

3.2 Real Line

Consider next results on the real line. Thus q is a positive density on the real line. The positive definite matrix Σ is now simply a scalar, which we parameterize as

$$\Sigma = \zeta^2.$$

Proposition 3.1. *Suppose that*

$$\int_0^\infty \frac{1}{\zeta^2(x)q(x)} = +\infty, \int_0^\infty \frac{1}{\zeta^2(-x)q(-x)} = +\infty \quad (4)$$

$$\lim_{|x| \rightarrow \infty} -\frac{x}{|x|} \left[\zeta(x) \frac{q'(x)}{q(x)} + \zeta'(x) \right] = +\infty. \quad (5)$$

Then Criterion 2.1 is satisfied.

Proof. See Appendix C. □

When ζ is constant, the compactness condition (5) reduces to:

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

which rules out densities with algebraic tails (tails that decay slower than $|x|$ raised to a negative power.) By allowing for ζ to increase, we can accommodate densities with algebraic tails.

3.3 \mathbb{R}^n

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

⁵Salinelli (1998) also gives results for some very special cases of unbounded domains (see section five of his paper). Among other restrictions, he considers multivariate domains with finite Lebesgue measure.

3.3.1 Core

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

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

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

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

Let \hat{f} denote the *minimal extension*, the smallest closed extension of f_o that is defined in (1). Condition 3.1 is equivalent to $f = \hat{f}$.

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

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

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

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

Proposition 3.2. *Condition 3.1 is implied by:*

$$\int_1^\infty \kappa(r)^{-1} r^{1-n} dr = \infty. \quad (6)$$

Proof. See Appendix C. □

Restriction (6) implies the scalar restriction (4) of Proposition 3.1. This follows since for any non-negative reals r_1 and r_2 ,

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

Notice that (6) 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:

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

Thus a sufficient condition for (6) is that

$$\int_1^\infty \frac{\kappa(r)}{r^2} r^{n-1} dr < \infty. \quad (7)$$

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

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

and

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

Notice that

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

Suppose for instance, $\varsigma^2(r)$ is dominated by a quadratic function (in r). Then (7) and hence (6) are satisfied because the density q is integrable:

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

Remark 3.1. Consider the specification given in Example 2.1. Since Σ is an identity matrix, it suffices that q be integrable. Thus C_K^2 is a core for the form f in this example.

We may extend the previous argument by supposing instead that

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

for some positive δ . Then

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

Thus (7) is satisfied provided that

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

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

3.3.2 Transforming the Measure

In this subsection we transform the space L^2 into a Lebesgue counterpart $L^2(leb)$. The transformation is standard (see Davies (1989)), but it is often applied in the reverse direction. By using this transformation we may appeal to some existing mathematical results on compactness to establish Criterion 2.1,

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

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

Given q write:

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

Assumption 3.1. *The function h is twice continuously differentiable.*

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

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

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

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

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

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

Thus

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

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

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

Therefore,

$$f(U^{-1}\psi, U^{-1}\psi^*) = \frac{1}{2} \int (\nabla\psi)' \Sigma (\nabla\psi^*) + \frac{1}{2} \int V \psi\psi^* \quad (8)$$

where the *potential* function V is given by:

$$V = - \sum_{i,j} \sigma_{i,j} \frac{\partial^2 h}{\partial y_i \partial y_j} - \sum_{i,j} \frac{\partial \sigma_{i,j}}{\partial y_i} \frac{\partial h}{\partial y_j} + (\nabla h)' \Sigma (\nabla h). \quad (9)$$

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

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

and

$$\int (\nabla\phi)' \Sigma \nabla\phi q = \int V\psi^2 + \int (\nabla\psi)' \Sigma (\nabla\psi). \quad (10)$$

Proof. See Appendix C. □

A consequence of this proposition is that

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

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

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

3.3.3 Divergent Potential

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

Assumption 3.2. *The derivative penalty matrix $\Sigma \geq \underline{c}I$ for some $\underline{c} > 0$.*

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

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

Criterion 3.1. $\lim_{|x| \rightarrow \infty} V(x) = +\infty$.

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

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

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

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

Direct verification of Criterion 3.1 may be difficult because formula (9) is a bit complicated. However, we may replace the Σ by a lower bound. In what follows we suppose that

Assumption 3.3. *The derivative penalty matrix Σ satisfies*

$$\Sigma(x) \geq \varsigma(x)^2 I.$$

where ς is twice continuously differentiable and $\varsigma(x)^2 \geq \underline{c} > 0$.

Let:

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

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

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

Let \check{f} be the minimal extension of \check{f}_o . If f is the minimal extension of f_o , Assumption 3.3 insures that the domain of \check{f} contains the domain of f . Applying Proposition 3.3 to \check{f} , it suffices to use

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

in place of V in demonstrating compactness.

Criterion 3.2. *Assumption 3.3 is satisfied and*

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

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

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

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

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

An alternative to criterion 3.2 is:

Criterion 3.3. *Assumption 3.3 is satisfied,*

a)

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

b)

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

Proposition 3.5. *Suppose Assumptions 3.1 is satisfied. Then Criterion 3.3 implies Criterion 2.1.*

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

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

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

We now deduce a convenient lower bound on:

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

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

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

Lemma 3.1. *Suppose Assumption 3.3 is satisfied. Then*

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

for all $\phi \in C_K^2$.

Proof. See Appendix C. □

Note that

$$\begin{aligned} \check{V}(x) + \check{W}(x) &= \varsigma(x)^2 \text{trace} \left[\frac{\partial^2 v(x)}{\partial x_i \partial x_j} - \frac{\partial^2 h(x)}{\partial x_i \partial x_j} \right] + \varsigma(x)^2 |\nabla h(x) - \nabla v(x)|^2 \\ &\quad + \underline{c} \left[\nabla v(x) \cdot \nabla v(x) - \text{trace} \left(\frac{\partial^2 v(x)}{\partial x_i \partial x_j} \right) \right]. \end{aligned}$$

Criterion 3.4. *Assumption 3.3 is satisfied,*

a)

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

b)

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

Proposition 3.6. *Suppose Assumptions 3.1 and Condition 3.1 are satisfied. Then Criterion 3.4 implies Criterion 2.1.*

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

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

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

The following example illustrates the advantage of this replacement.

Example 3.1.

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

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

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

and

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

Notice that both terms converge to zero as $|x|$ gets large, but that the squared gradient scaled by ζ^2 becomes arbitrarily large when $\beta > 1$. The first term is always negative, but the second one is nonnegative provided that $n \geq 2$. Even when $n = 1$ the second term is larger than the first provided that $\beta > 1$.⁶ This example illustrates when Criterion 3.4 is preferred to Criterion 3.3. The distinction can be important when densities have algebraic tails.

We conclude this section by summarizing our findings so far. We provided two criteria for constructing penalization functions that support principal component approximation. The first one, Criterion 3.3 gives the most flexibility; but it is applicable for data densities that have relatively thin tails. The second one, Criterion 3.4, allows for densities with algebraic tails but requires that the penalization be more severe in the extremes to compensate for the tail thickness. Making the penalization more potent limits the class of functions that are approximated. Moreover, when the penalization is too extreme, we encounter an additional approximation problem: the family of functions C_K^2 ceases to be a core for the form used in the principal component extraction.

The remainder of our paper can be read in two different ways. The next section considers refinements of results in this section, but for a limited class of densities. These results in effect illustrate formally and in a more precise way how the quality of approximation is altered by the tail behavior of the density and the penalization. Section 5 uses the theory of Markov processes to provide an explicit rationale for the choice of penalization. Neither this section nor the next builds on one another. Each can be read separately without knowledge of the other.

4 Illustration and Refinement

The optimization problem 2.1 uses a form built with a penalization matrix Σ scaled by a density q . The density also defines the underlying sense of approximation. For a given density, increasing the penalization effectively limits the class of functions that can be approximated. This should result in a corresponding accuracy gain. That is, for a fixed finite number of principal components we should be able to approximate better a smaller class of functions. Changing the density q changes the norm of the space L^2 . In particular, when we make the tail of the density thicker, the sense of approximation becomes more stringent. While it is too ambitious to establish a full array of exact results, we will be able to illustrate these effects for some examples.

In this section we first specialize the general compactness criteria of section 3 to the case of a radially symmetric density and penalization function. We study four parameterized classes of examples, two with densities that have exponential tails and two with densities

⁶We have previously established an alternative compactness criterion for $n = 1$ that does not involve second derivatives that may be preferred to Criterion 3.4.

that have thicker algebraic tails. We first derive bounds on the parameters that guarantee that we can apply radial counterparts to our compactness criteria 3.3 and 3.4 above. We then present bounds on the rate of divergence of the eigenvalues δ_j associated with the principal component extraction. Recall that these eigenvalues measure the smoothness of the principal components and are related to the bounds, λ_j , on the least squares approximation error second moments via the formula:

$$\lambda_j = \frac{1}{1 + \theta\delta_j}. \quad (11)$$

Thus by bounding the rate of divergence of the eigenvalues, we bound the rate of convergence of the approximation errors.

4.1 Compactness under Radial Symmetry

Prior to our analysis of some examples, we consider the specialization of our compactness criteria when radial symmetry is imposed. Recall that $q(x) = \exp[-2h(x)]$ and $\Sigma(x) \geq \exp[2v(x)]$. Suppose that

$$\begin{aligned} h(x) &= \eta(|x|) \\ v(x) &= \nu(|x|), \end{aligned}$$

then

$$\nabla h(x) = \frac{x}{|x|}\eta'(|x|), \quad \text{trace} \left[\frac{\partial^2 h(x)}{\partial x \partial x'} \right] = \frac{(n-1)}{|x|}\eta'(|x|) + \eta''(|x|),$$

where when necessary we use the weak notion of a derivative. The radial counterpart to compactness Criterion 3.3 is

Criterion 4.1. *Assumption 3.3 is satisfied,*

a)

$$\lim_{r \rightarrow \infty} \frac{\nu'}{\eta'} = 0;$$

b)

$$\lim_{r \rightarrow \infty} \exp(2\nu) \left[-\frac{(n-1)}{r}\eta' - \eta'' + \eta'^2 \right] = +\infty.$$

Remark 4.1. *Consider again Example 2.1. This example is radially symmetric and can be parameterized as in this section. Exploiting the radial symmetry, let*

$$\eta(r) \doteq \begin{cases} \frac{1}{2}(r-1)^2 + \text{constant} & \text{if } 1 \leq r \\ \text{constant} & \text{if } 0 \leq r < 1 \end{cases}$$

where the constant is chosen so that $q = \exp(-2h)$ is a density and $h(x) = \eta(|x|)$. The function ν is identically one. Since η grows as a quadratic in r , Criterion 4.1 is satisfied.

Similarly, the radial counterpart to compactness Criterion 3.4 is:

Criterion 4.2. *Assumption 3.3 is satisfied,*

a)

$$\lim_{r \rightarrow \infty} -\frac{(n-1)}{r} \nu' - \nu'' + (\nu')^2 = 0;$$

b)

$$\lim_{r \rightarrow \infty} \exp(2\nu) \left[\nu'' - \eta'' + \frac{(n-1)}{r} (\nu' - \eta') + (\eta' - \nu')^2 \right] = +\infty$$

We will apply these criteria in the examples that follow.

4.2 Four Examples

Example 4.1. (Exponential Decay I) *Suppose that*

$$\begin{aligned} \eta(r) &= c_1(1+r^2)^{\frac{\xi}{2}} + c_2, \\ \exp[2\nu(r)] &= \underline{c}(1+r^2)^\beta. \end{aligned}$$

For this example we use compactness Criterion 4.1. A simple calculation shows that part a) is satisfied and that part b) is dominated by $\exp(2\nu)(\eta')^2$. The resulting potential function behaves like

$$\underline{c}(c_1)^2 \xi^2 (1+r^2)^\omega$$

for large r where

$$\omega = \beta + \xi - 1.$$

Compactness requires that ω exceeds zero and thus depends on the sum of β and ξ . This dependence depicts a simple tradeoff between the tail behavior of the density and the potency of the penalization. In particular, as long as $\xi > 1$, we still have compactness even when the penalization matrix Σ is the identity matrix.

Our second example includes a limiting case of Example 4.1 in which $\xi = 1$. If β were also set to zero, principal components would fail to exist because we would be attempting to approximate a class of functions that is too large. We shrink the set by including a logarithmic specification of ς .

Example 4.2. (Exponential Decay II) *Suppose that η is the same as in Example 4.1 but that we consider a less potent penalty function:*

$$\begin{aligned} \eta(r) &= c_1(1+r^2)^{\frac{\xi}{2}} + c_2, \\ \exp[2\nu(r)] &= \bar{c}[\log(1+r^2)]^\tau + \underline{c} \end{aligned}$$

where $\xi \geq 1$ and $\tau \geq 0$.

As in Example 4.1, we apply Criterion 4.1. In this case, however, the potential function behaves like:

$$\bar{c}(c_1 \xi)^2 [\log(1+r^2)]^\tau (1+r^2)^{\xi-1}$$

for large r . Notice that the potential function diverges even when $\xi = 1$ provided that $\tau > 0$.

Our next example considers densities with algebraic tails.

Example 4.3. (Algebraic Decay I)⁷ *Suppose*

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

where $\gamma > n/2$ for the resulting q to be integrable. Consistent with our previous example and Example 3.1, let

$$\nu(\mathbf{r}) = \frac{\beta}{2} \log(1 + r^2) + \frac{\tilde{\mathbf{c}}}{2}$$

where $\tilde{\mathbf{c}} = \log \underline{\mathbf{c}}$. We now restrict $\beta > 1$.

In this case we use compactness Criterion 4.2. Simple calculations result in

$$\begin{aligned} (\eta' - \nu')^2 &= (\beta - \gamma)^2 \frac{r^2}{(1 + r^2)^2} \\ \frac{(n-1)}{r}(\nu' - \eta') &= (n-1)(\beta - \gamma) \frac{1}{1 + r^2} \\ \nu'' - \eta'' &= (\beta - \gamma) \frac{1 - r^2}{(1 + r^2)^2} \end{aligned}$$

Provided that $\gamma - \beta > n - 2$ the potential function Criterion 4.2 is guaranteed to be positive and to diverge for large r . Moreover, Proposition 3.2 guarantees that C_K^2 is a form core when $\Sigma = \zeta^2 I$.⁸

For this example, γ has to be large relative to β to ensure that the potential function has the correct sign in the tails and that C_K^2 is a form core. On the other hand, it does not influence the rate at which the potential function diverges. Instead the potential function behaves like

$$[(\beta - \gamma)^2 + (\beta - \gamma)(n - 2)](1 + r^2)^\omega$$

for $\omega = \beta - 1$. The value of γ only alters the scaling factor.

The following example studies a limiting case of the algebraic tail, but with a weak penalty. Thus we set $\beta = 1$, a value not allowed in our previous investigation, but we include a more modest penalty term.

Example 4.4. (Algebraic Decay II) *As in our previous example, suppose*

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

where $\gamma > n/2$, and let

$$\exp[2\nu(\mathbf{r})] = \underline{\mathbf{c}}(1 + r^2)(1 + [\log(1 + r^2)])^\tau$$

for $\tau > 0$.

⁷Pang (1996) gives an extended analysis of an example similar to this.

⁸By optimizing over the choice of the potential function, this inequality can be improved to be: $\gamma - \beta > \frac{n}{2} - 1$ and Proposition 3.2 may still be used to establish that C_K^2 is a form core.

In this example it can be shown that the potential function of Criterion 4.2 behaves like scale multiple of $(1 + [\log(1 + r^2)])^\tau$ for large r . Consistent with our previous example, we restrict $\gamma > n - 1$ to ensure that the potential function is positive and that C_K^2 is a core for the form.

The four examples illustrate the application of the compactness criteria. When the potential functions diverge more rapidly, approximation becomes easier in the sense that we obtain sharper bounds on the decay rate of the eigenvalues. We show this formally in the next subsections.

4.3 Some benchmark results on eigenvalue decay

Consider two forms f_1 and f_2 with common cores such that $f_2 \geq f_1$ on a core for the form f_2 . Then the j^{th} eigenvalue of the form f_2 is greater than or equal to the j^{th} eigenvalue of the form f_1 (e.g. see Edmunds and Evans (1987), Lemma 2.3). This gives us an operational way to use eigenvalue divergence for special forms to bound the eigenvalue divergence of the forms of interest. For this reason we start by deriving sharp bounds on parameterized versions of two forms studied previously in the literature:

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

and

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

defined on the appropriately defined subspaces of $L^2(\text{leb})$. We will apply the following three results in our subsequent analysis. Our first result gives a characterization of the eigenvalues of a parameterization of f_w .

Claim 4.1. *Let $w(x) = (1 + |x|^2)^\omega$ for some $\omega > 1/2$. Then the j^{th} eigenvalue of f_w satisfies:⁹*

$$\delta_j \sim (j + 1)^{\frac{2\omega}{n(1+\omega)}}.$$

Remark 4.2. *We consider again Example 2.1. The potential function for this example is quadratic in $r = |x|$ in the tails and behaves like $w(x) = (1 + |x|^2)^\omega$ in the tails for $\omega = 1$. A modification of the above claim that is easy to establish shows that $\delta_j \sim (j + 1)^{\frac{1}{n}}$.*

The second two results consider two alternative parameterizations of the form f_w^d .

Claim 4.2. *Let $w(x) = (1 + |x|^2)^\omega$ for $\omega > 0$. The eigenvalues δ_j of f_w^d satisfy:*

- *If $\omega > 1$, then $\delta_j \sim (j + 1)^{\frac{2}{n}}$.*
- *If $0 < \omega < 1$, then $\delta_j \sim (j + 1)^{\frac{2\omega}{n}}$.*

⁹The notation $\delta_j \sim a_j$ means that there are two positive finite constants c_ℓ and c_u such that $c_\ell a_j \leq \delta_j \leq c_u a_j$ for all j .

- If $\omega = 1$, then $\delta_j \sim \left(\frac{j+1}{\log(j+1)}\right)^{\frac{2}{n}}$.

Claim 4.3. Let $w(x) = [1 + \log(1 + |x|^2)]^\tau$ for $\tau > 0$. Then the eigenvalues δ_j of f_w^d satisfy

$$\delta_j \sim [\log(j + 1)]^\tau.$$

Notice that in comparing the results in Claims 4.2 and 4.3, we see that weaker penalties slow the decay rate in the eigenvalues.

In the remainder of this section we shall apply these claims to obtain bounds on the eigenvalues of the form f .

4.4 Lower bounds

We now reconsider the four examples described previously using bounding arguments. Consider first Example 4.1. In this example we use either $\omega = \beta + \xi - 1$ in conjunction with Claim 4.1 or $\omega = \min\{\beta + \xi - 1, \beta\}$ in conjunction with Claim 4.2 to obtain a lower bound on the divergence rate of the eigenvalues depending upon which claim gives the best results. The outcome of this comparison is depicted in Figure 6.

This figure depicts a tradeoff between the thinness of tail density as captured by the parameter ξ and the strictness of the penalty β . In the far right region the divergence rate matches that when the state space is compact with minimum smooth boundary and the density is bounded and bounded away from zero (see *e.g.* Edmunds and Evans (1987), Theorem V.6.5). When $\beta = 0$, we fail to obtain a bound on the eigenvalue divergence when $1 < \xi < 3/2$ even though we know that the potential function diverges, albeit at a slow rate. This is evidently a case in which approximation is particularly difficult and eigenvalue decay is slow.

Consider Example 4.2. For $\xi \geq 3/2$ we can bound the eigenvalue decay just as in Example 4.1 with $\beta = 0$ by applying Claim 4.1. For $1 \leq \xi < 3/2$, we use Claim 4.3 to bound the eigenvalues. Combining these results gives:

- (Claim 4.3) If $1 \leq \xi < 3/2$, then $\delta_j \geq c[\log(j + 1)]^\tau$
- (Claim 4.1) If $\xi \geq 3/2$, then $\delta_j \geq c(j + 1)^{\frac{2(\xi-1)}{n\xi}}$

The first region for ξ is of particular interest to us. In Example 4.1, when $\beta = 0$ we ruled out $\xi = 1$ to achieve existence, and we eliminated entirely this first region for ξ to obtain bounds on the eigenvalues. We include this region by imposing a weak penalty, but obtaining a correspondingly weak bound on the eigenvalue decay.

Consider next Example 4.3. In this case the eigenvalue decay bounds are dictated by the magnitude of β . Applying Claim 4.2 with $\omega = \beta - 1$, we obtain:¹⁰

¹⁰Notice that when $\beta > 3/2$, Claim 4.1 also can be applied with $\omega = \beta - 1$. However, Claim 4.2 provides sharper eigenvalue bounds.

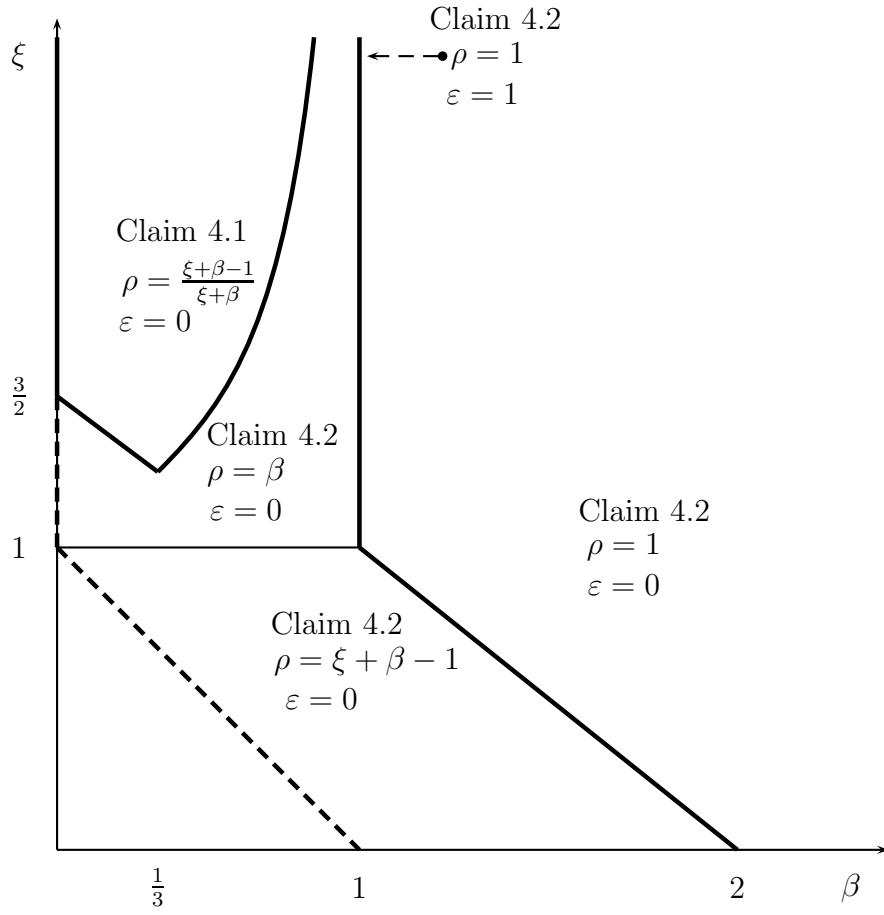


Figure 6: For exponential tail density, the j^{th} eigenvalue $\delta_j \geq c(j+1)^{\frac{2\rho}{n}} [\log(j+1)]^{-\frac{2\varepsilon}{n}}$

- If $\beta > 2$ then $\delta_j \geq c(j+1)^{\frac{2}{n}}$.
- If $\beta = 2$ then $\delta_j \geq c(j+1)^{\frac{2}{n}}[\log(j+1)]^{-\frac{2}{n}}$.
- If $1 < \beta < 2$, then $\delta_j \geq c(j+1)^{\frac{2(\beta-1)}{n}}$.

The value of γ limits how large β can be because of the form core restriction.

Finally, the eigenvalue decay rate for Example 4.4 is given by Claim 4.3. The rate becomes arbitrarily slow as τ declines to zero.

4.5 Upper bounds

So far we have derived lower bounds on the decay of the eigenvalues of the form. As we saw from formula (11), these lower bounds give upper bounds on the ability to approximate with a finite number of principal components. Thus these bounds are informative as to when approximation problems are relatively easy. Upper bounds on the decay of the eigenvalues of the form can also be valuable. Not only do they allow us to infer when the lower bounds are relatively sharp, they can be used to inform as to when approximation problems are difficult. Upper bounds on eigenvalues can be obtained by using upper bounds on forms as we now illustrate.

Consider Example 4.1 with $\beta = 0$. We have already verified that the potential function diverges for $1 < \xi < 3/2$, but we failed to give a lower bound on the eigenvalue decay. Applying Claim 4.1 in conjunction with divergence rate of the potential function, we obtain the upper bound:

$$\delta_j \leq c(j+1)^{\frac{2(\xi-1)}{n}},$$

but in this case the bound is necessarily weak. When $\xi = 1$ and we include the limiting penalization given in Example 4.2, we find that

$$\delta_j \leq c[\log(j+1)]^\tau$$

and our bound is now sharp. The slow eigenvalue decay in this example implies that the resulting approximation problem is a particularly challenging one.

5 Forms and Markov Processes

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

We proceed as follows. We suppose that the data are generated by a Markov diffusion. Associated with the form f , there is a second-order differential operator F that generates the semigroup of a Markov diffusion. The diffusion process has Σ as its local covariance matrix and q as its stationary density. The construction of F is unique provided that we restrict the

process to be time reversible. The construction will give us an alternative interpretation for the matrix Σ when the data are samples of a continuous time diffusion process.

As we will see when we relax reversibility there are multiple diffusion processes that are consistent with the form f . All of these diffusion processes have the same diffusion matrix Σ and the same stationary density q , but they have different drift functions. Our existence results for principal components will turn out to also imply the existence of eigenfunctions of the conditional expectations operators of these alternative Markov processes. The eigenfunctions behave like scalar autoregressions with heteroskedastic disturbances that satisfy conditional moment restrictions in discrete time. When the process is reversible these two objects coincide, but more generally they diverge.

5.1 A Differential Operator

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

$$\begin{aligned} f_o(\phi, \psi) &= \frac{1}{2} \int_{\Omega} \sum_{i,j} \sigma_{ij} \frac{\partial \phi}{\partial y_j} \frac{\partial \psi}{\partial y_i} q \\ &= -\frac{1}{2} \int \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} \psi q - \frac{1}{2} \int \sum_{i,j} \frac{\partial(q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j} \psi \end{aligned} \quad (12)$$

where the second equality of (12) 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 (12) to motivate our interest in the differential operator F_o :

$$F_o \phi = -\frac{1}{2} \sum_{i,j} \sigma_{ij} \frac{\partial^2 \phi}{\partial y_i \partial y_j} - \frac{1}{2q} \sum_{i,j} \frac{\partial(q\sigma_{ij})}{\partial y_i} \frac{\partial \phi}{\partial y_j}. \quad (13)$$

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

$$\begin{aligned} f_o(\phi, \psi) &= \langle F_o \phi, \psi \rangle \\ &= \langle \phi, F_o \psi \rangle. \end{aligned}$$

where the second relation holds because we can interchange the role of ϕ and ψ in (12). Notice from (13) that operator F_o has both a first derivative term and a second derivative term. Symmetry (with respect to q) is built into the construction of this operator because of its link to the symmetric form f_o .

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

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

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

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

Set

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

Then we may use Ito's Lemma to show that for each $\phi \in C_K^2$

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

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

5.2 Generating Reversible Diffusions

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

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

$$f(G_\alpha\phi, \psi) + \alpha \langle G_\alpha\phi, \psi \rangle = \langle \phi, \psi \rangle \tag{14}$$

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

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

$$F\phi = (G_\alpha)^{-1}\phi - \alpha\phi \tag{15}$$

defined on the domain $G_\alpha(L^2)$. It can be shown that F is independent of α .¹¹ Moreover, it is an extension of the operator F_o because f is an extension of f_o (*e.g.* see Lemma 3.3.1 of Fukushima et al. (1994)).

¹¹Since the operator F is self-adjoint and positive semidefinite, we may define a unique positive semidefinite square root \sqrt{F} . While F may only be defined on a reduced domain, the domain of its square root may be extended uniquely to the entire space $\mathcal{D}(f)$ and: $f(\phi, \psi) = \langle \sqrt{F}\phi, \sqrt{F}\psi \rangle$ (*e.g.* see Fukushima et al. (1994) Theorem 1.3.1).

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

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

as a Neumann series expansion. Notice that (15) implies

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

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

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

often referred to as Yosida approximation to construct formally a strongly continuous, semigroup of operators indexed by $t \geq 0$.¹²

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

$$G_\alpha \phi = \int_0^\infty \exp(-\alpha t) \exp(-tF) \phi dt \quad (16)$$

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

As we have just seen, associated with a closed form f , there is an operator F and a (strongly continuous) semigroup $\{\exp(-tF) : t \geq 0\}$ on L^2 . To establish that there is a Markov process associated with this semigroup, we need first to verify that the semigroup satisfies two properties. First we require, for each $t \geq 0$ and each $0 \leq \phi \leq 1$ in L^2 , $0 \leq \exp(-Ft)\phi \leq 1$. 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.¹³ Finally we must make sure that the Markov semigroup is actually the family of conditional expectation operators of a Markov process.

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

¹²Strong continuity requires that $\exp(-tF)\phi$ converges in L^2 to ϕ as t declines to zero.

¹³Fukushima (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 5.1. (*Beurling-Deny*) For any $\phi \in \mathcal{D}(f)$, ψ given by the truncation:

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

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

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

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

The form f satisfies the Beurling-Deny criteria (Davies (1989) Theorem 1.3.5). Thus there exists a self-adjoint operator F associated with f , which is an extension of F_o and generates a submarkov semigroup $\exp(-Ft)$. Theorem 7.2.1 of Fukushima et al. (1994) guarantees that there exists a Markov process $\{x_t\}$ that has $\exp(-Ft)$ as its semigroup of conditional expectations.¹⁴ The semigroup $\exp(-Ft)$ conserves probability because the unit function is in the domain of the form f and $f(1, \phi) = 0$ for any $\phi \in \mathcal{D}(f)$. As a consequence, the unit function is also in the domain of the operator F ,

$$F1 = 0.$$

We illustrate the Markov process construction by reconsidering Example 2.1. The Markov process associated with this form has the identity as the diffusion matrix Σ . The drift can be inferred from formula (13):

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

This construction results in a continuous-time extension of the familiar threshold autoregression model.¹⁵

5.3 Principal Components and Eigenfunctions

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 (14):

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

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

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

¹⁴Actually it is a *Hunt process*, a special strong Markov process that possesses certain sample path continuity properties. (See Fukushima et al. (1994), Appendix A.2 for a definition.)

¹⁵See Tong (1990) for a discussion of discrete-time threshold autoregressive models. For analyses of other continuous-time threshold models see Stramer et al. (1996a) and Stramer et al. (1996b).

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

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

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

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

An eigenfunction ψ of the generator F satisfies:

$$E[\psi(x_{t+s})|x_t] = \exp(-\delta s)\psi(x_t), \quad (17)$$

for some positive number δ and each transition interval s . Thus the principal components described previously will also satisfy the testable conditional moment implications (17). The scalar process $\{\psi(x_t)\}$ should behave as a scalar autoregression with autoregressive coefficient $\exp(-\delta s)$ for sample interval s . The forecast error: $\psi(x_{t+s}) - \exp(-\delta s)\psi(x_t)$ will typically be conditionally heteroskedastic (have conditional variance that depends on the Markov state x_t).

Since the form can be depicted using a principal component decomposition as in (3), analogous decompositions are applicable to F and $\exp(-Ft)$:

$$\begin{aligned} F\phi &= \sum_j \delta_j \frac{\langle \phi, \psi_j \rangle}{\langle \psi_j, \psi_j \rangle} \psi_j \\ \exp(-tF)\phi &= \sum_j \exp(-t\delta_j) \frac{\langle \phi, \psi_j \rangle}{\langle \psi_j, \psi_j \rangle} \psi_j \end{aligned}$$

where the first expansion is only a valid L^2 series when ϕ is in the domain of the operator F . When the eigenvalues of the form increase rapidly, the term $\exp(-t\delta_j)$ will decline to zero, more so when the time horizon t becomes large. As a consequence, it becomes easier to approximate the conditional expectation operator over a finite transition interval t with a smaller number of principal components. On the other hand, slow eigenvalue divergence of the form will make it challenging to approximate the transition operators with a small number of principal components. Our results in previous sections give primitive conditions based on the behavior of stationary density and diffusion matrices for the existence of eigenfunction decompositions and eigenvalue decay rates.

5.4 An Alternative Form

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

This quadratic form is defined to be the limit

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

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

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

While the form f is used to define the operator F , the form g may be used to define F^{-1} as is evident from formulas (14) or (15). The forms f and g share eigenfunctions. The g eigenvalues are the reciprocals of the f eigenvalues.

In light of equation (16)

$$\langle G_\alpha \phi, \psi \rangle = \int_0^\infty \exp(-\alpha t) E[\phi(x_t) \psi(x_0)] dt. \quad (18)$$

Hence, using (15), we obtain:

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

Notice that this form is symmetric because the resolvent operator is self-adjoint for any positive α . Using (18) 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:

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

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

This gives us an alternative interpretation of our extended version of the principal component extraction of Salinelli (1998). We may base the extraction on maximizing $g(\phi, \phi)$ subject to $\langle \phi, \phi \rangle = 1$ over recursively constructed spaces H . In words we are maximizing long run variation while constraining the overall variation. Smooth functions of a Markov state are also highly persistent and as a consequence maximize long run variation.

5.5 Irreversible Processes

The stationary Markov construction we used in the previous section resulted in a generator that was self adjoint and hence a process that was time reversible. Even among the class of stationary Markov diffusions, reversibility is special when the process has multiple dimensions. Given a stationary density q and a diffusion matrix Σ , we have seen how to construct a reversible diffusion, but typically there are other diffusions that share the same density and diffusion matrix. We now characterize the drifts of such processes and show how our compactness results can be extended.

5.5.1 Misspecified Reversibility

Instead of constructing a Markov process implied by a form, suppose instead we have specified the process as a semigroup of conditional expectation operators indexed by the transition interval. We suppose this process has stationary density q . Following Nelson (1958) and Hansen and Scheinkman (1995) we study the semigroup of conditional expectation operators on the space L^2 . This semigroup has a generator A defined on a dense subspace of L^2 . Consistent with our construction of F , on the subspace of C_K^2 , we suppose that A can be represented as a second-order differential operator:

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

and that

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

It may be shown that

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

on C_K^2 .

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

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

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

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

It follows from Nelson (1958) that the adjoint operator has the same diffusion matrix, but a different drift vector μ . The drift for the adjoint operator A^* is given by:

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

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

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

which is the negative of the second term in representation (13) for F_o . Thus if the generator A of the semigroup is not self adjoint, then the operator F implied by the form is a second order differential operator built using a simple average of the forward and reverse time drift coefficients, μ and μ^* , and the common diffusion matrix, Σ .

Remark 5.1. *The density q and the diffusion matrix Σ do place other restrictions on the drift vector μ . Since q is the stationary density, μ and μ^* must also satisfy:*

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

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

5.5.2 Resolvent Compactness

For a Markov process with generator A we may construct a family of resolvent operators:

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

While the generator is an unbounded operator on L^2 , the resolvent operators are bounded. When the resolvents are compact operators, they have well defined eigenfunctions and eigenvalues, but they may be complex valued. (See Rudin (1973), Theorem 4.25, page 108.)

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

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

Then

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

Thus it suffices to show that

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

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

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

Given these relations, our sufficient conditions for the existence of principal components can also be applied to establish the existence of eigenfunctions even when the generator is not self adjoint. These eigenfunctions continue to behave as conditionally heteroskedastic autoregressions and therefore satisfy an extensive array of testable conditional moment restrictions.¹⁶ Also they can be used to approximate dynamics over finite time intervals and can be ordered by their importance in contributing to long run dynamics by the magnitudes of the eigenvalues.

5.6 Related Literature

The idea of using eigenfunctions of conditional expectation operators for estimation and testing of Markov processes has been suggested previously by Demoura (1998), Hansen and Scheinkman (1995), Kessler and Sorensen (1999), Hansen et al. (1998), Florens et al. (1998) and Gobet et al. (2004). In particular, Kessler and Sorensen (1999) use eigenfunctions to construct quasi-optimal estimators of parametric scalar models of the drift and diffusion coefficients from discrete-time data in the special case in which the functional forms of eigenfunctions are known *a priori*. Demoura (1998), Hansen and Scheinkman (1995), Hansen et al. (1998), Gobet et al. (2004) and Darolles et al. (2004) study semiparametric and non-parametric identification and over-identification based on a principal component extraction that is closely related to the one analyzed here; see Fan (2005) for a recent review. This previous literature focuses on applications primarily to scalar diffusions and in some cases to scalar diffusions on compact state spaces with reflective boundaries.

In this section we have shown how the principal component extraction for Markov diffusions extends to multivariate settings applicable to processes without attracting barriers. While the principal components we construct will typically not be eigenfunctions of conditional expectation operators when the underlying Markov process fails to be reversible, our existence results for principal components still apply to the existence of eigenfunctions.

¹⁶The eigenfunctions and eigenvalues may be complex in the irreversible case. When they are complex, the real and imaginary parts of the eigenfunction can be used to construct a bi-variate, first-order autoregression with an error term that has conditional mean zero.

6 Concluding Remarks

We have studied principal components from multiple vantage points. We have explored their role in capturing variation subject to smoothness constraints and their role in capturing long run variation in time series modeling. We have also considered their use in approximation where the smoothness constraints limit the family of functions to be approximated. Along this latter vein, Meddahi (2001), Andersen et al. (2004) and Corradi and Distaso (2005) exploit their approximation properties of principal components in volatility modeling. Relatedly, given their ability to capture variation in multivariate data, Bontemps and Meddahi (2005) use the implied principal components to devise tests of multivariate distributional assumptions.

There are a variety of other applications that can be explored in both time series and cross sectional analyses. For instance, the forms used to define principal components give a way to model multivariate Markov processes. While we focused on diffusion processes, more general processes including processes with jumps can be accommodated by expanding the types of forms that are considered. As argued by Hansen and Scheinkman (1995) reversible diffusions can be identified nonparametrically from discrete time data. Moreover, for any given nonlinear Markov process model fit to data, the implied principal components or more generally the eigenfunctions give natural devices for characterizing the dynamics and for supplying diagnostics and testable implications. For instance, under reversibility we have seen that principal components should behave as eigenfunctions of conditional expectation operators and hence satisfy testable conditional moment restrictions.

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

In addition to developing some of the alternative interpretations of principal components in a functional setting, this paper has provided sufficient conditions for their existence and in some circumstances characterized the behavior of the corresponding eigenvalues that measure the incremental importance of the principal components. Formally, we link the existence of the principal components to the behavior of the density and the behavior of the penalization function used to measure smoothness. By supplying this analysis we demonstrate how high level existence assumptions can be supplanted by more careful inspection of the inputs into the extraction or approximation problems.

In this paper we have focused on the natural first step by establishing the existence of and characterizing functional principal components. Inferential issues, while crucial, are beyond the scope of this paper. Since these methods have already found use in applications, some statistical results have been obtained in special cases. Our main interest in these methods is in the characterization of distributions and the dynamics that underly those distributions. Formalizing statistical comparisons of models and data in a multivariate setting is an obvious next step, supported by either parametric, semiparametric or nonparametric estimation of

local dynamics.

A Form Closure

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

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

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

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

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

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

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

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

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

B Computation

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

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

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

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

$$\int_{\mathbb{R}^2} \phi\phi^*q = \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi \xi(r, \omega)\xi^*(r, \omega)d\omega q^*(r)dr \quad (19)$$

where

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

Consider next the quadratic form for the derivatives. Note that

$$\nabla\phi = \frac{1}{r} \begin{bmatrix} r \cos(\omega) & -\sin(\omega) \\ r \sin(\omega) & \cos(\omega) \end{bmatrix} \begin{bmatrix} \xi_r \\ \xi_\omega \end{bmatrix}.$$

Thus

$$(\nabla\phi) \cdot (\nabla\phi^*) = (\xi_r \xi_r^*) + \frac{1}{r^2} (\xi_\omega \xi_\omega^*).$$

We may evaluate the form:

$$\begin{aligned} \int_{\mathbb{R}^2} (\nabla\phi) \cdot (\nabla\phi^*) q &= \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi \xi_r(r, \omega) \xi_r^*(r, \omega) d\omega q^*(r) dr \\ &+ \frac{1}{2\pi} \int_0^\infty \frac{1}{r^2} \int_{-\pi}^\pi \xi_\omega(r, \omega) \xi_\omega^*(r, \omega) d\omega q^*(r) dr. \end{aligned} \quad (20)$$

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

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

Thus the form in (19) is:

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

and the form in (20) is:

$$\begin{aligned} \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi \xi_r(r, \omega) \xi_r^*(r, \omega) d\omega q^*(r) dr &+ \frac{1}{2\pi} \int_0^\infty \frac{1}{r^2} \int_{-\pi}^\pi \xi_\omega(r, \omega) \xi_\omega^*(r, \omega) d\omega q^*(r) dr \\ &= \frac{1}{2} \int_0^\infty \left[\frac{k^2}{r^2} p(r) p^*(r) + p'(r) p^{*'}(r) \right] q^*(r) dr. \end{aligned}$$

For computational purposes we may use these two forms in p and solve scalar problems. Notice that the second form depends on k . The $k = 0$ problem gives rise to the principal components that are constant on circles. For $k \geq 1$ we may compute principal components of the form $p(r) \cos(k\omega)$ and $p(r) \sin(k\omega)$. The sum of the two will be symmetric and the difference will be anti-symmetric when converted to the original coordinates.

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

1. Basis functions. We used as basis functions Hermite polynomials constructed to be orthogonal relative to the density: $\exp(-y^2)$.

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

C Additional Proofs

Consider first the proof of Proposition 3.1.

Proof. Hansen et al. (1998) consider densities from stationary scalar diffusions, whose boundaries are not attracting. This proposition gives an equivalent statement of their compactness condition, written in terms of the stationary density. The scalar diffusion coefficient in their analysis is ζ^2 .

□

We now present a criteria for Condition 3.1 to hold. This result is due essentially to Azencott (1974) and Davies (1985).

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

Proof. As explained in Section 5.2, associated with the forms f and \hat{f} we may construct operators F and \hat{F} and resolvents G and \hat{G} . Integration by parts can be used to show that the operators F and \hat{F} are extensions of the differential operator

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

defined on C_K^2 . The form \hat{f} and hence the form f satisfies the Beurling-Deny Criterion 5.1 (Davies (1989) Theorem 1.3.5). Hence as stated in Davies (1985) the operators F and \hat{F} can be extended to subspaces of L^1 . Similarly the resolvents G and \hat{G} can be extended to L^1 . We will denote the extended operators as F^1 , \hat{F}^1 , G^1 , and \hat{G}^1 . Since q is integrable, L^2 convergence implies L^1 convergence and consequently F and \hat{F} are restrictions of F^1 and \hat{F}^1 , respectively. Similarly for the resolvent operators.

If $\hat{f}(1, \phi) = 0$ for all $\phi \in \mathcal{D}(\hat{f})$ then $\hat{F}1 = 0$ and $\hat{G}1 = 1$. Consequently $G^11 = 1$. It follows from Theorem 2.2 in Davies (1985) that C_K^2 is a core for F^1 , in the sense that F^1 is

the closure in L^1 of \hat{L} .¹⁷ Hence C_K^2 is a core for F^2 , and thus a core for f or, equivalently, f and \hat{f} coincide. □

We now prove Proposition 3.2.

Proof. Since \hat{f} is the minimal closed extension, it has C_K^2 as its core. When this condition is met, a sequence of functions ϕ_j in C_K^2 can be constructed that converge to 1 in L^2 and $f(\phi_j, \phi_j)$ converges to zero. See Fukushima et al. (1994) Theorem 1.6.6 and Theorem 1.6.7. An approximating sequence of functions with compact support is supplied by Fukushima et al. (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 C_K^2 . Thus the unit function is in the domain of \hat{f} and $\hat{f}(1, \phi) = 0$ for $\phi \in C_K^2$ and hence for $\phi \in \mathcal{D}(\hat{f})$. As we established above, this is sufficient for Condition 3.1. □

Next, we establish Proposition 3.3

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

$$\begin{aligned} \check{H} \doteq & \{ \psi \in L^2(\text{leb}) : \int (V + \theta)(\psi)^2 < \infty, \text{ there exists } g \text{ measurable, with } \int g' \Sigma g < \infty, \\ & \text{and } \int \psi \nabla \varphi = - \int g \varphi, \text{ for all } \varphi \in C_K^1 \}. \end{aligned}$$

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

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

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

Since C_K^2 is a core for f , there exists a sequence $\{\phi_j : j = 1, 2, \dots\}$ in C_K^2 that converges to ϕ in the Hilbert space norm of \bar{H} . Hence this sequence is Cauchy in that norm. Writing $\psi_j = U\phi_j$ and applying equation (8) we obtain:

$$\begin{aligned} \int (\phi_j - \phi_\ell)^2 (1 + \theta) q + \int (\nabla \phi_j - \nabla \phi_\ell)' \Sigma (\nabla \phi_j - \nabla \phi_\ell) q \\ = \int (V + \theta + 1) (\psi_j - \psi_\ell)^2 + \int (\nabla \psi_j - \nabla \psi_\ell)' \Sigma (\nabla \psi_j - \nabla \psi_\ell). \end{aligned}$$

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

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

¹⁷Davies (1985) assumes that the coefficients of \hat{L} are C^∞ . However the proof holds for C^2 coefficients since elliptic regularity holds even when the coefficients are only Lipschitz (see Theorem 6.3 of Agmon (1965))

$$\begin{aligned}
&= \lim_{j \rightarrow \infty} (\nabla \phi_j)' \Sigma (\nabla \phi_j) q \\
&= \int (\nabla \phi)' \Sigma (\nabla \phi) q.
\end{aligned}$$

This proves (10).

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

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

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

$$\int \psi \nabla \varphi = - \int g \varphi, \quad (21)$$

and

$$\int g' \Sigma g < \infty. \quad (22)$$

We check relation (21) by applying integration by parts,

$$\begin{aligned}
- \int \nabla \psi \varphi &= - \int [\exp(-h)(-\phi \nabla h + \nabla \phi)] \varphi = - \int \nabla \phi \exp(-h) \varphi + \int \exp(-h) \varphi \phi \nabla h \\
&= \int \phi [\exp(-h) \nabla \varphi - \nabla h \exp(-h) \varphi] + \int \exp(-h) \varphi \phi \nabla h = \int \psi \nabla \varphi.
\end{aligned}$$

Inequality (22) follows from (10). □

We now establish Lemma 3.1

Proof. Consider a positive function

$$\chi(x) = \frac{1}{\zeta},$$

and note that

$$[\zeta(x)^2 - \underline{\zeta}] \nabla \chi(x) = -\zeta(x) \nabla v + \underline{\zeta} \frac{\nabla v(x)}{\zeta(x)}.$$

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

$$\begin{aligned}
\int (\zeta^2 - \underline{\zeta}) \nabla \chi \cdot \nabla \phi &= \int \left[\left(\frac{\zeta^2 + \underline{\zeta}}{\zeta} \right) (\nabla v \cdot \nabla v) + \left(\frac{\zeta^2 - \underline{\zeta}}{\zeta} \right) \text{trace} \left(\frac{\partial^2 v}{\partial x_i \partial x_j} \right) \right] \phi \\
&= \int \check{W} \chi \phi
\end{aligned}$$

The conclusion follows from Theorem 1.5.12 of Davies (1989). While Davies (1989) uses test functions ϕ in C_K^∞ , the same proof applies to C_K^2 test functions. □

Finally, we consider the three results related to eigenvalue decay. The proof of Claim 4.1 is based on a result of Reed and Simon (1978).

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 of¹⁸

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

where

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

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 $r^{\frac{n}{2} + \frac{n}{2\omega}}$. An analogous upper bound can be formed with an appropriate adjustment in the constant term. \square

Claim 4.2 follows directly from Proposition 4.1 of Mynbaev and Otel'baev (1988).

Proof. Use the form f_w^d to construct an inner product and a corresponding Hilbert space, and consider the embedding operator (identity operator) mapping this space into $L^2(\text{leb})$. Mynbaev and Otel'baev (1988) established the exact order of the approximation numbers $\{a_j : j \geq 0\}$ for this embedding operator. The result follows by using the known relation $(\delta_j)^{-1/2} = a_j$ between the approximation numbers of an embedding operator and eigenvalues. (See Remark 2.2 and Theorem 5.10 of Edmunds and Evans (1987).) \square

In turn, Claim 4.3 follows from the analysis of Haroske (1997).

Proof. Again we use the form f_w^d to construct an embedding operator as in the previous proof. For the choice of w given in this claim, Haroske (1997) (Proposition 4.4) established the exact order of its entropy numbers: $e_j \sim [\log(j + 1)]^{-\frac{\tau}{2}}$. Entropy number e_j for an operator mapping into $L^2(\text{leb})$ is the infimum of all $\epsilon > 0$ for which there exist 2^j balls in $L^2(\text{leb})$ with radius ϵ which cover the image of the operator over the unit ball implied by the form. For this problem, entropy numbers behave like approximation numbers. For instance, inequality (3.0.9) in Carl and Stephani (1990), page 120 states that $a_j \leq 2e_j$, and the Theorem in Triebel (1994) implies that $e_j \leq ca_j$ for some positive number c .¹⁹ Given the essential equivalence of the behavior of approximation numbers and entropy numbers, the conclusion follows from the relationship $(\delta_j)^{-1/2} = a_j$ mentioned above. \square

¹⁸While Reed and Simon (1978) only prove this result for $n \geq 2$, they argue that the result extends to the $n = 1$ case.

¹⁹The second inequality follows from part (ii) of the Theorem in Triebel (1994) by setting Triebel's $f(j)$ equal to the reciprocal of the entropy numbers. With this construction the inequality is an implication of Triebel's equation (5).

References

- Agmon, S. 1965. *Lectures on Elliptic Boundary Problems*. Princeton: Vos Nostrand.
- Andersen, T. G., T. Bollerslev, and N. Meddahi. 2004. Analytical Evaluation of Volatility Forecasts. *International Economic Review* 45:1079–1110.
- Azencott, R. 1974. Behavior of Diffusion Semi-groups at Infinity. *Bull. Soc. Math. France* 102:193–240.
- 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 Wahrscheinlichkeits-theorie und verwandte Gebiete* 60:185–201.
- Bontemps, C. and N. Meddahi. 2005. Testing Distributional Assumptions: A GMM Approach. Universite de Montreal.
- Brezis, H. 1983. *Analyse Fonctionnelle*. Paris: Masson.
- Carl, B. and I. Stephani. 1990. *Entropy, Compactness and the Approximation of Operators*. Cambridge University Press.
- Corradi, V. and W. Distaso. 2005. Semiparametric Comparison of Stochastic Volatility Models using Realized Measures. *Review of Economic Studies* forthcoming.
- Darolles, S., J. P. Florens, and C. Gourieroux. 2004. Kernel Based Nonlinear Canonical Analysis and Time Reversibility. *Journal of Econometrics* 119:323–353.
- Dauxois, J. and G. M. Nkiet. 1998. Nonlinear Canonical Analysis and Independence Tests. *Annals of Statistics* 26 (4):1254–1278.
- Dauxois, J. and A. Pousse. 1975. Une Extension de l'Analyse Canonique: Quelques Applications. *Ann. Inst. H. Poincare* 11:355–379.
- Davies, E. B. 1985. L^1 Properties of Second Order Operators. *Bulletin of the London Mathematical Society* 17:417–436.
- . 1989. *Heat Kernels and Spectral Theory*. Cambridge University Press.
- Demoura, S. 1998. The Nonparametric Estimation of the Expected Value Operator. Mimeo.
- Edmunds, D. E. and W.D. Evans. 1987. *Spectral Theory and Differential Operators*. Oxford University Press.
- Fan, J. 2005. A Selective Overview of Nonparametric Methods in Financial Econometrics (with discussion). *Statistical Science* 20:317–357.

- 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.
- Gobet, E., M. Hoffmann, and M. Reib. 2004. Nonparametric Estimation of Scalar Diffusions based on Low Frequency Data. *Annals of Statistics* 26:2223–2253.
- 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. 1997. Some Limiting Embeddings in Weighted Function Spaces and Related Entropy Numbers. Forschungsergebnisse Math/Inf 97/04, Universitat Jena, Germany.
- Kessler, M. and M. Sorensen. 1999. Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process. *Bernoulli* 5:299–314.
- Meddahi, N. 2001. An Eigenfunction Approach for Volatility Modeling. Universite de Montreal.
- Mynbaev, K. and M. Otel’baev. 1988. *Weighted Functional Spaces and Differential Operator Spectrum*. Nauka.
- Nelson, E. 1958. The Adjoint Markov Process. *Duke Mathematical Journal* 25:671–690.
- 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.
- Rudin, W. 1973. *Functional Analysis*. New York: McGraw-Hill.
- Salinelli, E. 1998. Nonlinear Principal Components I: Absolutely Continuous Variables. *Annals of Statistics* 26:596–616.
- Stramer, O., P. J. Brockwell, and R. Tweedie. 1996a. Continuous-time Threshold AR(1) Models. *Advances in Applied Probability* 28:728–746.
- Stramer, O., R. Tweedie, and P. J. Brockwell. 1996b. Existence and Stability of Continuous Time Threshold ARMA Models. *Statistica Sinica* 6:715–732.

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

Triebel, H. 1994. Relations between Approximation Numbers and Entropy Numbers. *Journal of Approximation Theory* 78.