Nonlinear Principal Components and Long Run Implications of Multivariate Diffusions

Xiaohong Chen, Lars Peter Hansen and José Scheinkman

Presented by Andrea Tamoni

February 9, 2011

Linear Principal Components Analysis

- Given an $n \times 1$ vector of mean-zero random variables x with $\mathrm{cov}(x, x') = \Sigma$
- PCA solve the question "What linear combinations of x has maximum variance, subject to the constraint that the sum of squared weights is one and each linear combination is orthogonal to the previous ones"

$$\max_{Q_i} \mathrm{var}(Q_i' x) = Q_i' \Sigma Q_i$$

$$\text{s.t.}$$

$$Q_i' Q_i = 1$$

$$Q_i' Q_j = 0 \quad i \neq j$$

- Solutions are given by eigenvectors of the covariance matrix $Q_i' \Sigma Q_i = \lambda_i$

Principal component in 2d

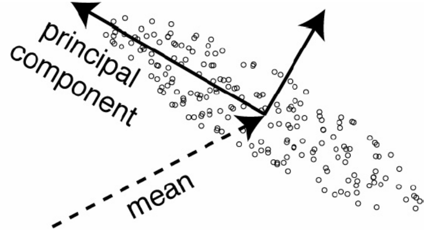

Figure: Geometrically, principal components represent a new coordinate system, with axes in the directions with maximum variability.

Objectives

- Define a functional notion of nonlinear principal components (NPC).
- Provide a time-reversible Markov diffusion process as data generating devices to interpret the NPCs extraction method.
- Relate NPC to eigenfunctions of conditional expectations operators associated with a time-reversible stationary Markov process

Nonlinear Principal Components

Definition

The function ψ_j is the j-th nonlinear principal component, for $j \geq 0$ if ψ_j solves for

$$\min_{\phi \in \bar{H}} \int_{x \in \Omega} \nabla \phi(x)^{\mathsf{T}} \Sigma(x) \nabla \phi(x) q(x) dx$$

s.t.

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

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

Markov Process and NPC

- The data $\{x_i\}_{i=1}^T$ are sampled from a continuous-time, stationary Markov diffusion $\{x_t : t \geq 0\}$ on the state space $\Omega \subseteq \mathbb{R}^n$. Specifically $\{x_t : t \geq 0\}$ solves the SDE

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

  with appropriate boundary restrictions and where $\{B_t : t \geq 0\}$ is an n-dimensional, standard Brownian motion
- Let q be the stationary density of x_t and $\Sigma(x_t) = \Lambda(x_t)\Lambda(x_t)^\mathsf{T}$
- Let $\phi(x_t) \in C_K^2$, the space of twice continuously differentiable functions with compact support in Ω. From Ito's lemma the (state dependent) local variance of the process $\{\phi(x_t)\}$ is

$$\nabla\phi(x)^\mathsf{T}\Sigma(x)\nabla\phi(x)$$

Markov Process and NPC

- The data $\{x_i\}_{i=1}^{T}$ are sampled from a continuous-time, stationary Markov diffusion $\{x_t : t \geq 0\}$ on the state space $\Omega \subseteq \mathbb{R}^n$. Specifically $\{x_t : t \geq 0\}$ solves the SDE

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

  with appropriate boundary restrictions and where $\{B_t : t \geq 0\}$ is an n-dimensional, standard Brownian motion
- Let q be the stationary density of x_t and $\Sigma(x_t) = \Lambda(x_t)\Lambda(x_t)^{\mathsf{T}}$
- Let $\phi(x_t) \in C_K^2$, the space of twice continuously differentiable functions with compact support in Ω. From Ito's lemma the (state dependent) local variance of the process $\{\phi(x_t)\}$ is

$$\nabla\phi(x)^{\mathsf{T}}\Sigma(x)\nabla\phi(x)$$

Quadratic form defined on C_K^2

- Compute the average of the local variance

$$f_0(\phi, \phi) = \frac{1}{2} \int_{x \in \Omega} \nabla \phi(x)^\intercal \Sigma(x) \nabla \phi(x) q(x) dx$$

  where ∇ is the gradient operator, $\Omega \subseteq \mathbb{R}^n$ is the state space and q is the stationary density of x_t

- The local variance is the measure of magnitude of the instantaneous forecast error in forecasting $\{\phi(x_t)\}$ over the next instant given the current Markov state.

Nonlinear Principal Components

Definition

The function ψ_j is the j-th nonlinear principal component, for $j \geq 0$ if ψ_j solves for

$$\min_{\phi \in C_K^2} f_0(\phi, \phi)$$

s.t.

$$\langle \phi, \phi \rangle = \int \phi^2(x) q(x) dx = 1$$

$$\langle \psi_s, \phi \rangle = \int \psi(x) \phi(x) q(x) dx = 0 \quad s = 0, \dots, j-1$$

- The NPCs are extracted by making the average of local forecast error small for functions with unit second moments plus orthogonality.

Extension of the form f_0

- We want to study the case where Ω is not compact
- Quadratic form f expressed in terms of the gradients of functions

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

  where ∇ is the weak gradient operator, $\Omega \subseteq \mathbb{R}^n$ is a state space (with possibly infinite Lebesgue measure), Σ is a *state dependent* positive definite matrix, q is the invariant density of a strictly stationary ergodic data
- The smoothness constraint include state dependence

Domain of the form f

- Let L^2 denote the space of Borel measurable square integrable functions with respect to the population probability distribution Q with density q
- The form f is defined for any functions $\phi(\cdot)$ and $\psi(\cdot)$ in $\bar{H}$ where

$$\bar{H} = \{\phi \in L^2 : \exists\, g \text{ measurable, with } \int g(x)^{\intercal}\Sigma(x)g(x)q(x)dx < \infty$$
$$\text{and } \int \phi(x)\nabla\psi(x)dx = -\int g(x)\psi(x)dx \ \ \forall \psi \in C_K^1\}$$

- g is referred to as the weak derivative of ϕ. If $\phi \in \bar{H}$ then we write $\nabla\phi = g$. Notice that
  - g is unique
  - $\Lambda(x_t)g \in L^2$

Forms and Markov Processes

- Use the form f to build a Markov process. How?
- Associated with the form f, there is a second-order differential operator F that generates the semigroup of a Markov diffusion
- The diffusion process has $\Sigma(x_t)$ as its local covariance matrix and q as it stationary density.
- The construction of F is unique provided that we restrict the process to be time reversible.

A differential operator on C_K^2

- Associated with the form f_0, there is a second-order differential operator F_0. For any ϕ and ψ in C_K^2

$$
\begin{aligned}
f_0(\phi,\psi) &= \frac{1}{2}\int \sum_{i,j} \sigma_{i,j} \frac{\partial \phi}{\partial x_j}\frac{\partial \psi}{\partial x_i} q \\
&= \underbrace{-\frac{1}{2}\int \sum_{i,j} \sigma_{i,j} \frac{\partial^2 \phi}{\partial x_i \partial x_j}\psi q - \frac{1}{2}\int \sum_{i,j} \frac{\partial (q\sigma_{ij})}{\partial x_i}\frac{\partial \phi}{\partial x_j}\psi}_{\langle F_0\phi,\psi\rangle}
\end{aligned}
$$

where we implicitly define

$$
F_0\phi = -\frac{1}{2}\sum_{i,j} \sigma_{i,j} \frac{\partial^2 \phi}{\partial x_i \partial x_j} - \frac{1}{2q}\sum_{i,j} \frac{\partial (q\sigma_{ij})}{\partial x_i}\frac{\partial \phi}{\partial x_j}
$$

Implied Generator

- The generator $\mathcal{A}$ of a multivariate Markov diffusion process is

$$\mu(x) \cdot \frac{\partial \phi}{\partial x} + \frac{1}{2} Tr\left(\Sigma \frac{\partial^2 \phi}{\partial x \partial x^{\mathsf{T}}}\right)$$

- Recall the differential operator F_0 associated with the form f

$$-F_0\phi = \frac{1}{2q}\sum_{i,j}\frac{\partial(q\sigma_{ij})}{\partial x_i}\frac{\partial \phi}{\partial x_j} + \frac{1}{2}\sum_{i,j}\sigma_{i,j}\frac{\partial^2\phi}{\partial x_i \partial x_j}$$

- $-F_0$ coincides with the infinitesimal generator of $\{x_t\}$ in C_K^2 once we set $\sigma_{i,j}$ to be the (i,j) element of the matrix Σ and we identify the implicit drift

$$\mu_j \equiv \frac{1}{2q}\sum_i \frac{\partial(q\sigma_{ij})}{\partial x_i}$$

Implied Generator

- The generator $\mathcal{A}$ of a multivariate Markov diffusion process is

$$\mu(x) \cdot \frac{\partial \phi}{\partial x} + \frac{1}{2} Tr \left(\Sigma \frac{\partial^2 \phi}{\partial x \partial x^{\mathsf{T}}} \right)$$

- Recall the differential operator F_0 associated with the form f

$$-F_0 \phi = \frac{1}{2q} \sum_{i,j} \frac{\partial (q \sigma_{ij})}{\partial x_i} \frac{\partial \phi}{\partial x_j} + \frac{1}{2} \sum_{i,j} \sigma_{i,j} \frac{\partial^2 \phi}{\partial x_i \partial x_j}$$

- $-F_0$ coincides with the infinitesimal generator of $\{x_t\}$ in C_K^2 once we set $\sigma_{i,j}$ to be the (i,j) element of the matrix Σ and we identify the implicit drift

$$\mu_j \equiv \frac{1}{2q} \sum_i \frac{\partial (q \sigma_{ij})}{\partial x_i}$$

Implied Generator

- The generator $\mathcal{A}$ of a multivariate Markov diffusion process is

$$\mu(x) \cdot \frac{\partial \phi}{\partial x} + \frac{1}{2} Tr \left(\Sigma \frac{\partial^2 \phi}{\partial x \partial x^{\mathsf{T}}} \right)$$

- Recall the differential operator F_0 associated with the form f

$$-F_0 \phi = \frac{1}{2q} \sum_{i,j} \frac{\partial (q \sigma_{ij})}{\partial x_i} \frac{\partial \phi}{\partial x_j} + \frac{1}{2} \sum_{i,j} \sigma_{i,j} \frac{\partial^2 \phi}{\partial x_i \partial x_j}$$

- $-F_0$ coincides with the infinitesimal generator of $\{x_t\}$ in C_K^2 once we set $\sigma_{i,j}$ to be the (i,j) element of the matrix Σ and we identify the implicit drift

$$\mu_j \equiv \frac{1}{2q} \sum_i \frac{\partial (q \sigma_{ij})}{\partial x_i}$$

Implied Generator - Cont'd

- We parameterize diffusion processes using
  - the stationary density q
  - a (possibly state dependent) diffusion matrix Σ
  - the drift μ is implicit in such a construction
- We have shown how to go from the forms to the generator of Markov processes. Just find the symmetric solution to

$$f_0(\phi, \psi) = \langle F_0\phi, \psi\rangle = \langle \phi, F_0\psi\rangle \tag{1}$$

  where the second relation holds because we can interchange the role of ϕ and ψ in the form
- There are typically also nonsymmetric solutions to

$$f_0(\phi, \psi) = \langle F_0\phi, \psi\rangle$$

- These nonsymmetric solutions are also generators of diffusions.
- While the diffusion matrix is the same for the operator and its adjoint, the drift vectors differ.

Implied Generator - Cont'd

- We parameterize diffusion processes using
  - the stationary density q
  - a (possibly state dependent) diffusion matrix Σ
  - the drift μ is implicit in such a construction
- We have shown how to go from the forms to the generator of Markov processes. Just find the symmetric solution to

$$f_0(\phi, \psi) = \langle F_0\phi, \psi\rangle = \langle \phi, F_0\psi\rangle \tag{1}$$

  where the second relation holds because we can interchange the role of ϕ and ψ in the form
- There are typically also nonsymmetric solutions to

$$f_0(\phi, \psi) = \langle F_0\phi, \psi\rangle$$

- These nonsymmetric solutions are also generators of diffusions.
- While the diffusion matrix is the same for the operator and its adjoint, the drift vectors differ.

A differential operator associated with f

- Given a function $\phi \in L^2$ define $G_\alpha \phi \in \bar{H}$ to be the solution to

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

  $\forall \psi \in \bar{H}$ and for any $\alpha > 0$. (The Riesz representation theorem guarantees the existence of $G_\alpha \phi$)
- Associate with the form f the operator

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

  defined on $G_\alpha(L^2)$
- Since the operator F is self-adjoint and positive semidefinite, we may define a unique positive semidefinite square root $\sqrt{F}$ which can be extended uniquely to the entire space $\bar{H}$.

A differential operator associated with f - Cont'd

- Associated with a closed form f, there is an operator F and a (strongly continuous) semigroup of operators indexed by $t \geq 0$.
- A natural candidate for this semigroup is $\{\exp(-tF) : t \geq 0\}$ on L^2 (to be interpreted as "Yosida approximation")
- To establish that there is a Markov process associated with this semigroup, we must show that it satisfies contraction and that it conserves probabilities. Under certain conditions (see Beurling and Deny, 1958) these two properties are satisfied.
- In all, the semigroup is actually the family of conditional expectation operators of a Markov process

$$\exp(-tF)\phi(x_0) = E[\phi(x_t) \mid x_0]$$

A differential operator associated with f - Cont'd

- Overall we show that

Theorem

There exists a self adjoint operator F associated with f, which is an extension of F_0 and generates a semigroup $\{\exp(-tF) : t \geq 0\}$. The density q is the stationary density for this diffusion, the matrix Σ is the diffusion matrix and $\exp(-tF)$ is the conditional expectation operator over an interval of time t.

NPC and Eigenfunctions

Definition

An eigenfunction ψ of the quadratic form f satisfies

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

$\forall\ \phi \in \bar{H}$. The scalar $\delta > 0$ is the corresponding eigenvalue.

- A necessary condition for ψ to be a NPC is that it must be an eigenfunction of the quadratic forms f
- Eigenfunctions of the closed form f will also be eigenfunctions of the resolvent operators G_α and of the generator F.

NPC and Eigenfunctions

Definition

An eigenfunction ψ of the quadratic form f satisfies

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

$\forall \, \phi \in \bar{H}$. The scalar $\delta > 0$ is the corresponding eigenvalue.

- A necessary condition for ψ to be a NPC is that it must be an eigenfunction of the quadratic forms f
- Eigenfunctions of the closed form f will also be eigenfunctions of the resolvent operators G_α and of the generator F.

Principal Component Decompositions

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

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

- It follows that for any $\phi, \psi \in \bar{H} \subset L^2$

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

- We also show that $f_0(\phi, \psi) = \langle F_0 \phi, \psi \rangle$
- The generator therefore has spectral decomposition

$$-F\phi = \sum_{j=0}^{\infty} -\delta_j \frac{\langle \psi_j, \phi \rangle}{\langle \psi_j, \psi_j \rangle} \psi_j \qquad (2)$$

Principal Component Decompositions - Cont'd

- We can use the generator F to build a (strongly continuous) semigroup of conditional expectation operators.
- We then have the decomposition for the semigroup

$$\underbrace{\exp(-tF)\phi}_{\int \phi(y) P_t(x,dy)} = \sum_{j=0}^{\infty} \exp\left(-t\delta_j\right) \frac{\langle \psi_j, \phi \rangle}{\langle \psi_j, \psi_j \rangle} \psi_j \quad (3)$$

- Take $\phi = \psi_i$, the i-th eigenfunction of the quadratic form f (and of the generator F). We then have

$$E\left[\psi_i(x_{s+t}) \mid x_s\right] = \exp(-\delta_i t)\psi_i(x_s)$$

Statistical behavior of NPC

- Recall that if ψ is an eigenfunction of the form f then

$$E\left[\psi(x_{s+t}) \mid x_s\right] = \exp(-\delta t)\psi(x_t)$$

- 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 have conditional variance that depends on the Markov state x_t

An alternative form

Definition

The function ψ_j is the j-th nonlinear principal component, for $j \geq 1$ if ψ_j solves for

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

s.t.

$$\langle \phi, \phi \rangle = \int \phi^2(x) q(x) dx = 1$$

$$\langle \psi_s, \phi \rangle = \int \psi(x) \phi(x) q(x) dx = 0 \quad s = 0, \ldots, j-1$$

where ψ_0 is initialized to be the constant function one and $\mathcal{S}(F)$ is a well defined subspace of functions in L^2 for which $g(\phi, \phi) < \infty$.

An alternative form - Cont'd

- The quadratic form $g(\phi, \psi)$ is defined as

$$g(\phi, \psi) = 2 \lim_{\alpha \downarrow 0} \langle G_\alpha \phi, \psi \rangle = \int_{-\infty}^{\infty} E[\phi(x_t)\psi(x_0)]dt$$

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

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

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

$$\frac{1}{\sqrt{T}} \int_0^T \phi(x_s)ds \xrightarrow[T \to \infty]{} N(0, g(\phi, \phi))$$

An alternative form - Cont'd

- The quadratic form $g(\phi, \psi)$ is defined as

$$g(\phi, \psi) = 2 \lim_{\alpha \downarrow 0} \langle G_\alpha \phi, \psi \rangle = \int_{-\infty}^{\infty} E[\phi(x_t)\psi(x_0)]dt$$

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

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

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

$$\frac{1}{\sqrt{T}} \int_0^T \phi(x_s)ds \xrightarrow[T \to \infty]{} N(0, g(\phi, \phi))$$

Additional Results

- Nonlinear principal components (NPC) are functions of the data
  - that capture maximal variation in some sense
  - subject to orthogonality conditions and smoothness constraints enforced by a quadratic form f
- Alternatively NPC are 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

Conclusions

Why are we fussing about this?

- Take a multivariate nonlinear diffusion models. Given the nonlinearity in the state variables, it is a nontrivial task to infer the global dynamics
- ... in particular, it is difficult to infer the long-run behavior from this local specification based on low-frequency data
- These NPCs are eigenfunctions of conditional expectation operators *when the Markov process is reversible*
- ...hence imply conditional moment restrictions.
- NPCs offers a way to characterize intermediate and long-run features of the implied time series that are typically disguised from the local dynamics in non-linear settings.
- Using eigenfunctions of conditional expectation operators for estimation and testing of Markov processes

Conclusions

Why are we fussing about this?

- Take a multivariate nonlinear diffusion models. Given the nonlinearity in the state variables, it is a nontrivial task to infer the global dynamics
- ... in particular, it is difficult to infer the long-run behavior from this local specification based on low-frequency data
- These NPCs are eigenfunctions of conditional expectation operators *when the Markov process is reversible*
- ...hence imply conditional moment restrictions.
- NPCs offers a way to characterize intermediate and long-run features of the implied time series that are typically disguised from the local dynamics in non-linear settings.
- Using eigenfunctions of conditional expectation operators for estimation and testing of Markov processes

Conclusions

Why are we fussing about this?

- Take a multivariate nonlinear diffusion models. Given the nonlinearity in the state variables, it is a nontrivial task to infer the global dynamics
- ... in particular, it is difficult to infer the long-run behavior from this local specification based on low-frequency data
- These NPCs are eigenfunctions of conditional expectation operators *when the Markov process is reversible*
- ...hence imply conditional moment restrictions.
- NPCs offers a way to characterize intermediate and long-run features of the implied time series that are typically disguised from the local dynamics in non-linear settings.
- Using eigenfunctions of conditional expectation operators for estimation and testing of Markov processes

Conclusions

Why are we fussing about this?

- Take a multivariate nonlinear diffusion models. Given the nonlinearity in the state variables, it is a nontrivial task to infer the global dynamics
- ... in particular, it is difficult to infer the long-run behavior from this local specification based on low-frequency data
- These NPCs are eigenfunctions of conditional expectation operators *when the Markov process is reversible*
- ...hence imply conditional moment restrictions.
- NPCs offers a way to characterize intermediate and long-run features of the implied time series that are typically disguised from the local dynamics in non-linear settings.
- Using eigenfunctions of conditional expectation operators for estimation and testing of Markov processes