Optimal Bases and Frames for Data-Driven Forecasting

Tyrus Berry

George Mason University

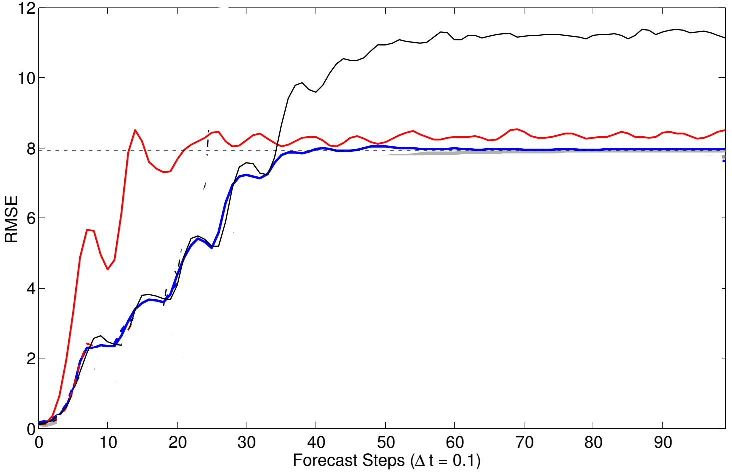
May 25, 2021

Joint work with John Harlim, PSU and Dimitris Giannakis, NYU
Supported by NSF-DMS 1854204 and 1723175

TYPES OF FORECASTING: DETERMINISTIC

- Deterministic Forecasting, $x_{k+1} = f(x_k)$
- Regression problem: Learn f from data
- Iterative Methods: $x_{k+n} = \tilde{f}^n(x_k)$ where $\tilde{f} \approx f$
- Direct Methods: $x_{k+n} = \tilde{f}_n(x_k)$ where $\tilde{f}_n \approx f^n$

DIRECT vs. **Iterative** vs PROBABILISTIC



Deterministic Forecasting

- Local Linear Regression (x_j near x):

$$f(x) \approx f(x_j) + Df(x_j)(x - x_j)$$

- Kernel Regression (h is bump function):

$$f(x) \approx \sum_j c_j k(x, x_j) = \sum_j c_j h(||x - x_j||)$$

- Neural Network (h is sigmoid):

$$f(x) \approx \sum_j c_j h(a_j^\top x + b_j) = \sum_j h(a_j^\top (x - \tilde{x}_j))$$

(where we write $b_j = a_j^\top \tilde{x}_j$)

- Deep Network: Composition of Neural Networks
- Reservoir Computer: Fix a_j, b_j, linear regression for c_j

Types of Forecasting: UQ

- Uncertainty Quantification, $p_{k+1} = f^* \circ p_k = p_k \circ f$
- Still a regression problem
- Option 1: Combine with ensemble forecast
- Option 2: Represent $\mathcal{L} = f^*$ in a basis

$$A_{ij} = \langle \phi_i, \mathcal{L}\phi_j \rangle = \langle \phi_i, \phi_j \circ f \rangle \approx \frac{1}{N} \sum_{k=1}^{N} \phi_i(x_k)\phi_j(x_{k+1})$$

TYPES OF FORECASTING: STOCHASTIC

- Stochastic Forecasting, $x_{k+1} = f(x_k, \omega_k)$
- *Not* a regression problem
- Don't just want $\bar{f} = \mathbb{E}_\omega[f(\cdot, \omega)]$
- We want the operator $p_{k+1} = \mathcal{L}p_k = \int p_k \circ f(\cdot, \omega)\, d\pi(\omega)$
- Note:

$$\int p_k \circ f(\cdot, \omega)\, d\pi(\omega) \neq p_k \circ \int f(\cdot, \omega)\, d\pi(\omega)$$

Stochastic forecasting = Operator Estimation

- Represent $\mathcal{L} = f^*$ in a basis

$$A_{ij} = \langle \phi_i, \mathcal{L}\phi_j \rangle = \langle \phi_i, \phi_j \circ f \rangle \approx \frac{1}{N}\sum_{k=1}^{N} \phi_i(x_k)\phi_j(x_{k+1})$$

- Error Sources: Bias, variance, and truncation
- Which basis?
  - Respect the measure $\Rightarrow$ Eliminate bias
  - Leverage smoothness $\Rightarrow$ Minimize variance
  - Capture global structure $\Rightarrow$ Minimize truncation

What is Manifold Learning?

- **Manifold learning $\Leftrightarrow$ Estimating Laplace-Beltrami**
- Eigenfunctions $\Delta\varphi_i = \lambda_i\varphi_i$ orthonormal basis for $L^2(\mathcal{M})$
- Smoothest functions: φ_i minimizes the functional

$$\lambda_i = \min_{\substack{f\perp\varphi_k \\ k=1,\ldots,i-1}} \left\{ \frac{\int_{\mathcal{M}} ||\nabla f||^2 \, dV}{\int_{\mathcal{M}} |f|^2 \, dV} \right\}$$

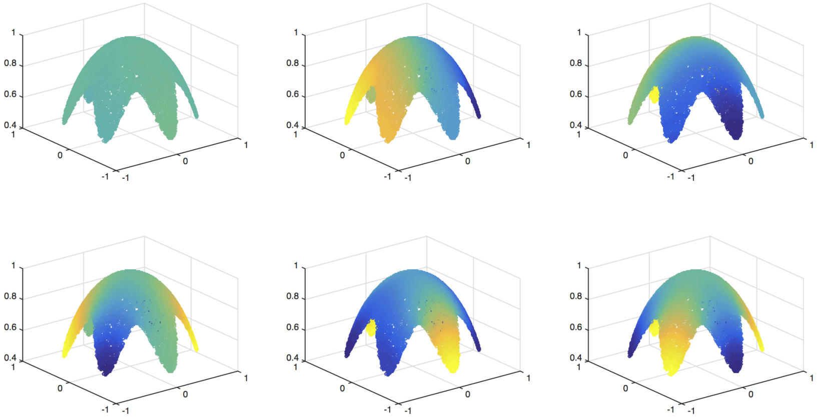
- Eigenfunctions of Δ are custom Fourier basis
  - Smoothest orthonormal basis for $L^2(\mathcal{M})$
  - Can be used to define wavelets
  - Define the Hilbert/Sobolev spaces on $\mathcal{M}$

DIFFUSION MAPS: GRAPH LAPLACIAN → MANIFOLD LAPLACIAN

- For data points $\{x_i\}_{i=1}^N \subset \mathcal{M} \subset \mathbb{R}^n$
- Define $J_{ij} = J(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{\delta^2}\right)$
- Define $D_{ii} = \sum_j J_{ij}$ (diagonal)
- Right normalization: $K = JD^{-1/2}$ and $\hat{D}_{ii} = \sum_j \hat{J}_{ij}$
- Left normalization: $\hat{K} = \hat{D}^{-1}K$
- Graph Laplacian: $L = \frac{1}{\delta^2}\left(I - \hat{K}\right)$
- **Theorem:** $L\vec{f} = \Delta_{p_{\mathrm{eq}}} + \mathcal{O}\left(\delta^2, N^{-1/2}\delta^{-1-d/2}\right)$

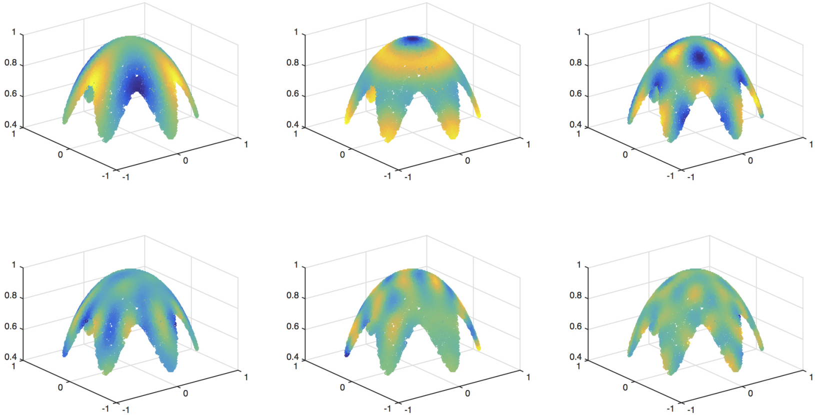
Harmonic Analysis on Manifolds/Data Sets

- Unit circle: $\Delta = \frac{d^2}{d\theta^2}$ eigenfunctions are Fourier basis
- General manifold or data set $\Rightarrow$ Custom Fourier basis



Harmonic Analysis on Manifolds/Data Sets

- Unit circle: $\Delta = \frac{d^2}{d\theta^2}$ eigenfunctions are Fourier basis
- General manifold or data set $\Rightarrow$ Custom Fourier basis



Forecasting with the Shift Map

- Stochastic evolution operator: $\mathcal{M}_\tau p(x,t) = p(x,t+\tau)$

$$\mathbb{E}_{p(\cdot,t+\tau)}[f] = \langle f, p(x,t+\tau)\rangle = \langle f, \mathcal{M}_\tau p(x,t)\rangle$$

- Dual is the shift map: $S_\tau f(x(t)) = f(x(t+\tau))$

$$\mathbb{E}_{p(\cdot,t+\tau)}[f] = \mathbb{E}\left[\langle f(x(t+\tau)), p(x,t)\rangle\right] = \mathbb{E}\left[\langle S_\tau f(x(t)), p(x,t))\rangle\right]$$

FORECASTING WITH THE SHIFT MAP

$$p(x,t) \xrightarrow{\text{Diffusion Forecast}} p(x,t+\tau)$$

$$\Big\downarrow \langle p, \varphi_j \rangle \qquad\qquad \Big\uparrow \textstyle\sum_j c_j \varphi_j p_{\mathrm{eq}}$$

$$\vec{c}(t) \xrightarrow{A_{lj} \equiv \mathbb{E}[\langle \varphi_j, S\varphi_l \rangle_{p_{\mathrm{eq}}}]} \vec{c}(t+\tau) = A\vec{c}(t).$$

Assuming ergodicity and mixing:

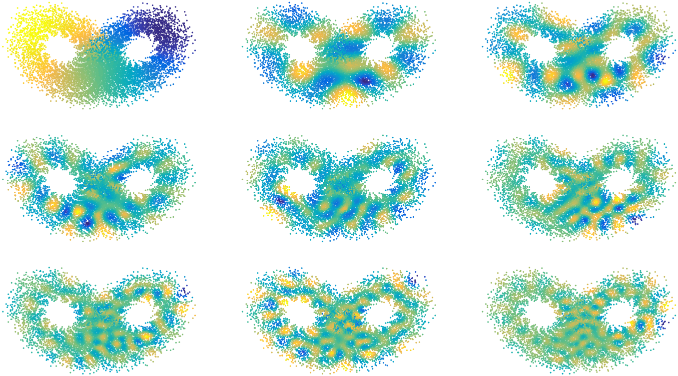
$$\mathbb{E}[\langle \varphi_j, S\varphi_l \rangle_{p_{\mathrm{eq}}}] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varphi_j(x_i)\varphi_l(x_{i+1})$$

Choosing a basis

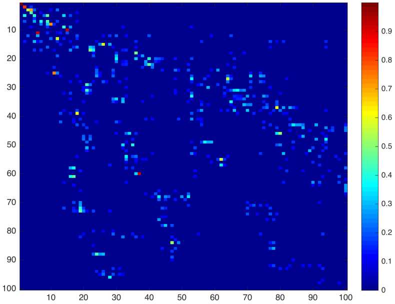
- Variance of $\frac{1}{N}\sum_{i=1}^{N}\varphi_j(x_i)\varphi_l(x_{i+1})$ is $\propto ||\nabla\varphi_l||_{p_{\mathrm{eq}}}$
- Minimizers of $||\nabla\varphi_l||_{p_{\mathrm{eq}}}$ are a generalized Fourier basis
- Let $\Delta_{p_{\mathrm{eq}}} = \Delta + \frac{\nabla p_{\mathrm{eq}}}{p_{\mathrm{eq}}}\cdot\nabla$ be the Laplacian weighted by p_{eq}
- The eigenfunctions $\Delta_{p_{\mathrm{eq}}}\varphi_j = \lambda_j\varphi_j$ minimize $||\nabla\varphi_j||_{p_{\mathrm{eq}}} = \lambda_j$
- How do we find φ_j? Manifold Learning: Diffusion Maps

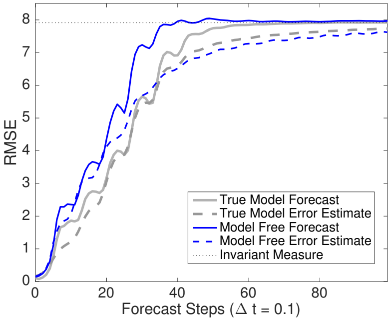
MANIFOLD LEARNING $\Rightarrow$ CUSTOM 'FOURIER' BASIS

- **Optimal basis:** Minimum variance $A_{lj} \equiv \mathbb{E}[\langle \varphi_j, S\varphi_l \rangle_q]$



SHIFT MAP ⇒ MARKOV MATRIX





DIFFUSION FORECAST EXAMPLE

(Loading Video...)

RELATIONSHIP TO CLASSICAL METHODS

- For partial observations, use Takens' reconstruction
- Local linear representations
  - Based on nearest neighbor interpolation
  - Kernel regression also interpolates from neighbors ($\approx$ linear for large data set near manifold)
  - Diffusion forecast extends the map to distributions
- Partition state space $\Rightarrow$ Markov matrix
  - Also uses the shift map, just a different basis
  - Diffusion forecast is optimal basis for estimation

Relationship to reservoir computers

- Create a random (recurrent) network $v_k \in \mathbb{R}^N$

$$v_{k+1} = f(Av_k + Bx_k)$$

- Continuously feed in the time series x_k

$$\begin{aligned} v_{k+1} &= f(Af(A \cdots f(Av_{k-\tau} + Bx_{k-\tau}) + \cdots) + Bx_k) \\ &= g(x_k, x_{k-1}, ..., x_{k-\tau}) \end{aligned}$$

- Predict: $x_{k+1} = Wv_k = Wg(x_k, ..., x_{k-\tau})$
- Since $\lambda_{\max}(A) < 1$ network forgets distant past
- Chooses a random diffeomorphism of a delay embedding
- Uses a linear combination W of a *random* basis

NEXT STEPS: FRAMES

- Frames for function space:
  - Instead of using a basis for L^2, use a wavelet frame $\Psi_{\ell,j}$
  - Can we reduce variance with an optimal frame?
- Frames for differential forms:
  - Stochastic evolution operator also acts on forms
  - Spectral Exterior Calculus: $\{\phi_i d\phi_j\}$ is a frame for 1-forms
  - Plan: Represent the SEO on forms in this frame

Projections of High Dimensional Dynamics

- Consider the 40-dimensional Lorenz-96 system:

$$\dot{x}_i = x_{i-1}x_{i+1} - x_{i-1}x_{i-2} - x_i + 8$$

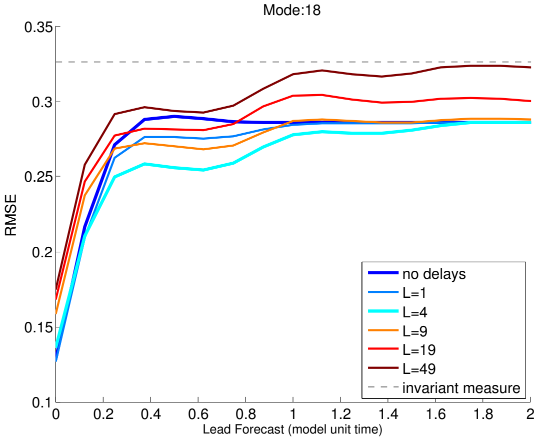
- Assume we only observe a projection of this system

$$y = h(x_1, ..., x_{40})$$

- Evolution of y is not closed, sometimes modeled by SDEs

Attractor Reconstruction

- Evolution of $y = h(x)$ is not closed
- Adding some delays helps, but adding too many hurts



NEXT STEPS: MORI-ZWANZIG FORMALISM

- Evolution of $y = h(x)$ is not closed
- Delay-embedding, $\tilde{y}_t$ only yeilds partial reconstruction
- Projections of dynamical systems can be closed as

Mori-Zwanzig formalism: $$\frac{d}{dt}\tilde{y} = V + K + R$$

- Diffusion Forecast includes: V (Markovian), R (stochastic)
- Missing the memory term: $K = \int_{-\infty}^{t} K(s, \tilde{y}_t, \tilde{y}_s)\tilde{y}_s \, ds$

Code and papers available at:

http://math.gmu.edu/~berry/

Building the basis

- B. and Sauer, *Consistent Manifold Representation for Topological Data Analysis.*
- Coifman and Lafon, *Diffusion maps.*
- B. and Harlim, *Variable Bandwidth Diffusion Kernels.*
- B. and Sauer, *Local Kernels and Geometric Structure of Data.*

Diffusion forecast

- B., Giannakis, and Harlim, *Nonparametric forecasting of low-dimensional dynamical systems.*
- B. and Harlim, *Forecasting Turbulent Modes with Nonparametric Diffusion Models.*