Time-Coupled Diffusion Maps

Nicholas F. Marshall

Mathematics Department, Yale University

December 2, 2014
Joint work with

Matthew J. Hirn
Département d’Informatique, École normale supérieure
Overview

1 Preliminaries

2 Time-coupled diffusion maps
Manifold learning

Data model
The data $X$ is an isometric embedding of a $k$-dimensional Riemannian manifold $(\mathcal{M}, g)$ in $d$-dimensional Euclidean space.

Cartoon
Diffusion maps

Setting

Data $X$ with distribution $\mu$ in $d$-dimensional Euclidean space

Affinity measure

Gaussian kernel $k_\varepsilon : X \times X \to \mathbb{R}$,

$$x \times y \mapsto \exp \left(-\frac{\|x - y\|^2}{4\varepsilon}\right)$$

Graph

Construct weighted graph $G = (X, k_\varepsilon)$

Random walk

Transition function $p_\varepsilon : X \times X \to \mathbb{R}$,

$$x \times y \mapsto \frac{k_\varepsilon(x,y)}{\int_X k_\varepsilon(x,y) d\mu(y)}$$
Illustration
Diffusion maps embedding

Diffusion distance

\[ D_t : X \times X \to \mathbb{R}, \quad x \times y \to \| p_\varepsilon^{(t)}(x, \cdot) - p_\varepsilon^{(t)}(y, \cdot) \|_{L^2} \]

Markov operator

\[ P_\varepsilon : L^2(X, \mu) \to L^2(X, \mu), \quad f \mapsto \int_X p_\varepsilon(\cdot, y)f(y)d\mu(y) \]

Markov kernel

The function \( p_\varepsilon^{(t)} \) is the kernel of the operator \( P_\varepsilon^t \).

Spectral representation

Under moderate conditions

\[ D_t(x, y) = \| \lambda^t \psi(x) - \lambda^t \psi(y) \|_{\ell^2} \]

where \( (\lambda_i) \) and \( (\psi_i) \) are the eigenvalues and eigenfunctions of \( P_\varepsilon \).
Illustration
Connection to diffusion

Theorem (Coifman and Lafon)

For any $t > 0$, the Neumann heat kernel $H^t$ can be approximated on $L^2(M)$ by $P_{\varepsilon}^{t\varepsilon}$:

$$\lim_{\varepsilon \to 0} P_{\varepsilon}^{t\varepsilon} = H^t$$

Cartoon
Problem Statement

Motivation

In practice, data often consists of multiple measurements \( \{X_1, \ldots, X_T\} \) resulting from different observations or processing.

Data model

The data \( \{X_\tau\} \) represents a sequence of isometric embeddings of a Riemannian manifold \( \mathcal{M} \) with a family of metrics \( g(\tau), \tau \in [0, T] \).

Cartoon

\[
\begin{align*}
\mathbb{R}^d &\quad \tau = 1, 2, \ldots, T \\
(\mathcal{M}, g(\tau)) & \\
\{X_1, X_2, X_3, \ldots, X_T\} &\quad \text{data set} \\
\text{manifold} &
\end{align*}
\]
Survey

Coifman, Lafon, Belkin, Niyogi

$P^t$ (homogeneous Markov chain)

Coifman and Hirn

$(P_1^t - P_2^t)^2 = P_1^{2t} + P_2^{2t} - P_1^t P_2^t - P_2^t P_1^t$

Wang, Jiangl, Wang, Zhou and Tu

$P_{(1)}^{t+1} = P^{(1)} P_{(2)}^t P^{(1)}'$ and $P_{(2)}^{t+1} = P^{(2)} P_{(2)}^t P^{(2)}'$

Ofir, Lederman, Talmon

$(P_1 P_2)^t$ and $(P_2 P_1)^t$

Time-coupled diffusion maps

$P_t P_{t-1} \cdots P_2 P_1$ (inhomogeneous Markov chain)
Construction

Evolving data

Sequence of embeddings \( \{X_i\}_{i=1}^T \) from a Riemannian manifold \( \mathcal{M} \) with a smooth 1-parameter family of metrics \( g(\tau), \tau \in [s, t] \).

Inhomogeneous Markov chain

For each \( X_i \), the diffusion framework defines a Markov operator \( P_{i,\varepsilon} \). We consider \( \{P_{i,\varepsilon}\}_{i=1}^T \) as a inhomogeneous Markov chain.

Adjusting for nonuniform distribution

The \( \{P_{i,\varepsilon}\} \) can be defined approximately independent of the possibly changing distribution of points induced by point movement.
Time-coupled diffusion maps embedding

Time-coupled diffusion distance

\[ D_{\varepsilon, s}^{(t)} : X \times X \to \mathbb{R}, \quad x \times y \mapsto \| p_{\varepsilon, s}^{(t)}(x, \cdot) - p_{\varepsilon, s}^{(t)}(y, \cdot) \|_{L^2}^2 \]

Markov kernel

Let \( p_{\varepsilon, s}^{(t)} \) denote the kernel of \( P_{\varepsilon, s}^{(t)} \triangleq P_{\varepsilon, \tau_{T-1}} P_{\varepsilon, \tau_{T-2}} \cdots P_{\varepsilon, \tau_1} P_{\varepsilon, \tau_0} \), where \( (\tau_i) \) is an \( \varepsilon \) uniform partition of \([s, t]\)

Theorem

The time-coupled diffusion distance

\[ D_{\varepsilon, s}^{(t)}(x, y) = \left\| \lambda_{s}^{(t)} \psi_{s}^{(t)}(x) - \lambda_{s}^{(t)} \psi_{s}^{(t)}(y) \right\|_{\ell^2} \]

where \( \left( \lambda_{s, i}^{(t)} \right), \left( \psi_{s, i}^{(t)} \right) \) are the eigenvalues, eigenfunctions of \( P_{s}^{(t)} \)
Time-dependent heat equation

Fundamental solution

Let $\mathcal{M}$ be a Riemannian manifold with a smooth 1-parameter family of metrics $g(\tau), \tau \in [0, T]$. A sufficiently smooth function

$$Z : \mathcal{M} \times [0, T] \times \mathcal{M} \times [0, T] \to \mathbb{R}$$

is a fundamental solution to the heat equation if:

1. $$\left( \frac{\partial}{\partial t} - \Delta_{g(t)} \right) Z(x, t; y, s) = 0$$

2. $$\lim_{t \searrow s} Z(x, t; y, s) = \delta_y(x)$$
Intuition

Static metric \((M, g)\)

We can represent the heat kernel \(H^t = e^{t\Delta}\) which has expansion

\[
e^{t\Delta} = e^{\epsilon\Delta} \cdots e^{\epsilon\Delta} \approx (1 + \epsilon\Delta)^{t/\epsilon}
\]

Evolving metric \((M, g(\tau))\)

We would expect the time-dependent heat kernel to resemble

\[
e^{\int_s^t \Delta g(\tau) d\tau} = e^{\int_s^{s+\epsilon} \Delta g(\tau) d\tau + \int_{s+\epsilon}^{s+2\epsilon} \Delta g(\tau) d\tau + \cdots + \int_{t-\epsilon}^t \Delta g(\tau) d\tau}
\]

However, the family \(\{\Delta g(\tau)\}\) is not in general commutative, and

\[
e^{\int_s^t A_\tau d\tau} = e^{\int_s^b A_\tau d\tau} e^{\int_b^t A_\tau d\tau} \quad \text{and} \quad \frac{\partial}{\partial t} e^{\int_s^t A_\tau d\tau} = A_t e^{\int_s^t A_\tau d\tau}
\]

only if \(\{A_\tau\}\) is commutative.
Series representation

Ordered Exponential

For a family of operators \( \{A_\tau\}_{\tau \in [s,t]} \) the ordered exponential

\[
\text{OE}_s^t A_\tau \triangleq I + \int_s^t A_\tau d\tau + \int_s^t A_\tau \int_s^\tau A_{\tau_1} d\tau_1 d\tau + \int_s^t A_\tau \int_s^\tau A_{\tau_1} \int_s^{\tau_1} A_{\tau_2} + \cdots
\]

where \( \int_s^t A_\tau d\tau : X \to X \) maps functions \( f \mapsto \int_s^t A_\tau f d\tau \).

Remark

If \( A_\tau = A \) is constant in \( \tau \). Then \( \text{OE}_s^t A_\tau = e^{(t-s)A} \)

Definition

Let the operator \( H_s^t \triangleq \text{OE}_s^t \Delta g(\tau) \).
Proposition

The operator $H_t^s$ is the time-dependent heat kernel, that is,

1. $\| H_t^s \| < \infty$
2. $\frac{\partial}{\partial t} H_t^s = \Delta g(t) H_t^s$
3. $\lim_{t \downarrow s} H_t^s = I$

Proposition

The operator $H_t^s$ admits the follows asymptotic expansion

$$H_t^s = I + \varepsilon \Delta g(s) + o(\varepsilon).$$

where $\varepsilon \triangleq t - s$. Furthermore, $H_t^s$ satisfies the semi-group property.
Main result

Theorem

*On* $L^2(\mathcal{M})$ the heat kernel $H^t_s$ can be approximated by the operator $P^{(t)}_{\varepsilon,s}$

$$\lim_{\varepsilon \to 0} P^{(t)}_{\varepsilon,s} = H^t_s$$
Main result

Theorem

On $L^2(M)$ the heat kernel $H^t_s$ can be approximated by the operator $P^{(t)}_{\varepsilon,s}$

$$\lim_{\varepsilon \to 0} P^{(t)}_{\varepsilon,s} = H^t_s$$
Flares description

Application

Suppose changing data represents nonlinear fluctuations from an invariant underlying geometric structure

Simulation construction

Begin by adding a small smooth bump to a unit sphere in $\mathbb{R}^3$: 

To create random fluctuations, we simultaneously compound bumps with random locations, radii, durations, and magnitudes.
Flares simulation videos
Flares embedding video

We plot the first 3 coordinates from the Flares data embeddings:

Left

Diffusion Maps
Input: $X_t$

Right

Time-Coupled Diffusion Maps
Input: $\{X_\tau\}_{\tau=1}^t$
Flares illustrations
Windmills description

Application

Suppose the changing data contains intermingled substructures which move cohesively with respect to the evolution parameter.

Simulation construction

We define two three prong (windmill) point clouds:

Both windmill spin at the same radial velocity in the positive direction such that opposing blades are always interacting.
Windmills videos

Data

Graph
Windmills illustrations
Algorithmic complexity

Diffusion maps (DM) \( n \) points, \( k \) neighbors

Nearest neighbor DM implementation in \( O(k^2 n \log n) \). How? Compute \( k \)-nearest neighbors and use sparse matrices.

Time-coupled Diffusion Maps (TCDM)

Using sparse matrices alone will be ineffective for TCDM as the product \( P_T \cdots P_1 \) spreads matrix values destroying sparsity.

![Sparsity vs. Number of Matrix Products](image)
Implementation idea (preliminary)

Fixed numerical rank

Approximate the product of the sequence $\{P_i\}_{i=1}^T$ by assuming each $P_i$ has a target rank of $r$ and an oversampling factor $e$.

Create random basis ($y_j$)

Generate $\lceil T(r + e)/c \rceil$ random vectors $\omega_j$ and define $y_j = P_{ij}\omega_j$ where $i_j$ is chosen uniformly at random from $1, \ldots, T$. The continuity factor $1 \leq c \leq T$ encodes how smoothly $\{P_i\}$ changes.

Matrix product

After expressing the family $\{P_i\}_{i=1}^T$ with respect to $(y_j)$ we can compute the product in $O(T^3n/c^2)$ operations.
References

Hiba Abdallah.

*Processus de Diffusion sur un Flot de Variétés Riemanniennes.*

M. Belkin and P. Niyogi.

Towards a theoretical foundation for laplacian-based manifold methods.
2005.

Bennett Chow, Sun-Chin Chu, David Glickenstein, Christine Guenther, James Isenber, Tom Ivey, Dan Knopf, Peng Lu, Feng Luo, and Lei Ni.

AMS, 2008.

Ronald R. Coifman and Matthew J. Hirn.

Diffusion maps for changing data.

Ronald R. Coifman and Stphane Lafon.

Diffusion maps.

Christine M. Guenther.

The fundamental solution on manifolds with time-dependent metrics.
Thank you

math.yale.edu/~nfm2