

Diffusion Maps for Changing Data

Matthew J. Hirn

Department of Mathematics
Yale University

July 26, 2012

Bell Labs

Collaborators

Joint work with Ronald Coifman and Roy Lederman.

Overview

- 1 High Dimensional Data, Low Dimensional Embeddings
 - Some Examples
 - Diffusion Maps
- 2 Diffusion Maps and Changing Graphs
 - Motivation
 - Generalizing the Diffusion Distance
- 3 Applications
 - Change Detection
 - Parameterized Difference Equations
- 4 Empirical Approximation

Overview

1 High Dimensional Data, Low Dimensional Embeddings

- Some Examples
- Diffusion Maps

2 Diffusion Maps and Changing Graphs

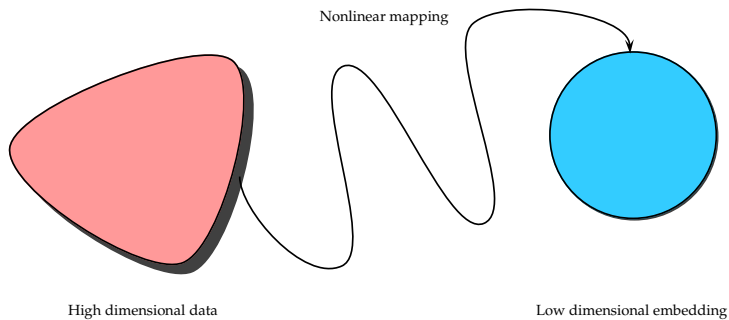
- Motivation
- Generalizing the Diffusion Distance

3 Applications

- Change Detection
- Parameterized Difference Equations

4 Empirical Approximation

Cartoon



Simple Example

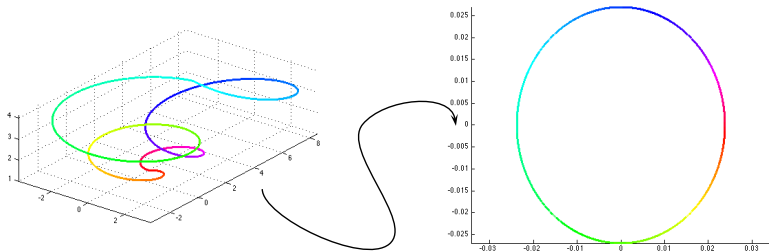
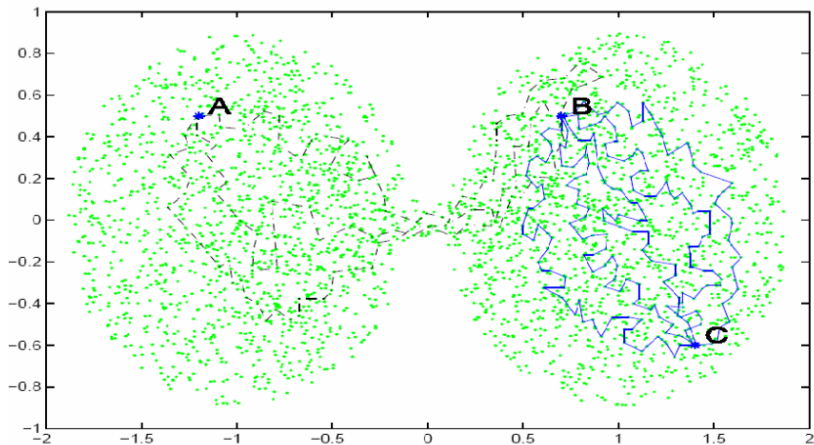


Figure: Left: A closed, non-self-intersecting curve in 3 dimensions. Right: Its embedding as a circle.

Euclidean vs Geodesic vs Diffusion



Created by Stéphane Lafon

Setup

- Let (X, μ) be a measure space.
 - The set X is our data set.
 - The measure μ represents the distribution of the points on X .
- Let $k : X \times X \rightarrow \mathbb{R}$ be a symmetric, positive kernel.
 - The value $k(x, y)$ measures the similarity between two points $x, y \in X$.
 - The kernel k should measure local similarities. For example,

$$k(x, y) = e^{-\|x-y\|^2/\varepsilon}.$$

- Let $G \triangleq (X, k)$ denote a weighted graph with:
 - Vertices given by our data set X .
 - Weighted edges with the weights given by k .

Random Walk on G

- Define the density $d : X \rightarrow \mathbb{R}$ as

$$d(x) \triangleq \int_X k(x, y) d\mu(y).$$

- Define a new kernel $p : X \times X \rightarrow \mathbb{R}$ as:

$$p(x, y) \triangleq \frac{k(x, y)}{d(x)}.$$

- p is the transition kernel of a random walk on G . In particular, $p(x, y)$ is the probability of transition from x to y in one time step.

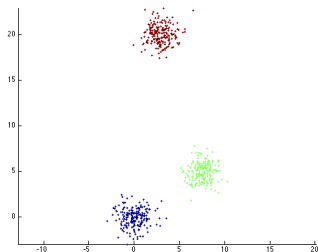
Diffusion Operator

- Define the diffusion operator $P : L^2(X, \mu) \rightarrow L^2(X, \mu)$ as:

$$(Pf)(x) \triangleq \int_X p(x, y) \cdot f(y) d\mu(y).$$

- Powers P^t , for t a positive integer, are equivalent to running the random walk forward in time t steps. Doing so reveals the relevant geometric structure of X at different scales.

Multiscale Analysis of Gaussian Clouds using P



(a) 3 Gaussian clouds

(b) Powers of P

Symmetrize the Random Walk

- One can also use a symmetric version of the random walk, which will simplify things later.
- Let $a : X \times X \rightarrow \mathbb{R}$ be the symmetric version of p , where

$$a(x, y) \triangleq d(x)^{\frac{1}{2}} \cdot p(x, y) \cdot d(y)^{-\frac{1}{2}}.$$

- Define the symmetric diffusion operator $A : L^2(X, \mu) \rightarrow L^2(X, \mu)$ as

$$(Af)(x) \triangleq \int_X a(x, y) \cdot f(y) d\mu(y).$$

- Let $a^{(t)} : X \times X \rightarrow \mathbb{R}$ be the kernel of the operator A^t , so that

$$(A^t f)(x) = \int_X a^{(t)}(x, y) \cdot f(y) d\mu(y).$$

Diffusion Distance

Definition

For each positive integer t , let $D^{(t)} : X \times X \rightarrow \mathbb{R}$ denote the **diffusion distance** at time t , where

$$D^{(t)}(x, y)^2 \triangleq \int_X \left(a^{(t)}(x, s) - a^{(t)}(y, s) \right)^2 d\mu(s).$$

Comparing Diffusions

Spectral Decomposition of A

- Under some mild assumptions on G and A , the operator A will have a discrete sequence of orthonormal eigenfunctions $\{\psi^{(\ell)}\}_{\ell \geq 1}$ with corresponding eigenvalues $\{\lambda^{(\ell)}\}_{\ell \geq 1}$, that satisfy:

- $(A\psi^{(\ell)})(x) = \lambda^{(\ell)} \cdot \psi^{(\ell)}(x).$

- $\langle \psi^{(\ell)}, \psi^{(m)} \rangle_{L^2(X, \mu)} = \delta(\ell - m).$

- $1 = \lambda^{(1)} > \lambda^{(2)} \geq \lambda^{(3)} \geq \dots \geq 0.$

- $\lim_{\ell \rightarrow \infty} \lambda^{(\ell)} = 0.$

- $\psi^{(1)}(x) = \sqrt{d(x)}.$

Diffusion Map

Definition

- We use the spectral decomposition of A to define the **diffusion map** $\Psi^{(t)} : X \rightarrow \ell^2$ as:

$$\Psi^{(t)}(x) \triangleq \left(\left(\lambda^{(\ell)} \right)^t \cdot \psi^{(\ell)}(x) \right)_{\ell \geq 1}.$$

- For any positive integer d , one can also define an **approximate diffusion map** $\Psi^{(t,d)} : X \rightarrow \mathbb{R}^d$ as:

$$\Psi^{(t,d)}(x) \triangleq \left(\left(\lambda^{(\ell)} \right)^t \cdot \psi^{(\ell)}(x) \right)_{\ell=1}^d.$$

Diffusion Map and the Diffusion Distance

Theorem (Coifman and Lafon)

- The diffusion distance on X is equivalent to the ℓ^2 distance on $\Psi^{(t)}(X)$, i.e.,

$$D^{(t)}(x, y) \triangleq \left\| a^{(t)}(x, \cdot) - a^{(t)}(y, \cdot) \right\|_{L^2(X, \mu)} = \left\| \Psi^{(t)}(x) - \Psi^{(t)}(y) \right\|_{\ell^2}.$$

- Furthermore, for any $\delta > 0$, there exists a positive integer d such that

$$D^{(t)}(x, y) \approx \left\| \Psi^{(t,d)}(x) - \Psi^{(t,d)}(y) \right\|_{\mathbb{R}^d},$$

to within accuracy δ .

Another Circle



(c) BTR-60

(d) BTR driving in a circle

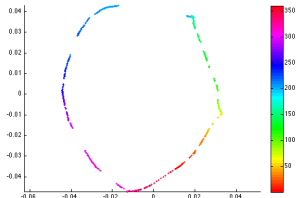


Figure: Embedding of the BTR video. The color corresponds to the angle on the circle at which the BTR is located.

Overview

1 High Dimensional Data, Low Dimensional Embeddings

- Some Examples
- Diffusion Maps

2 Diffusion Maps and Changing Graphs

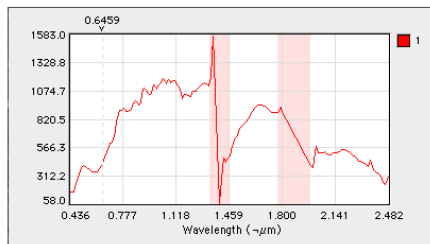
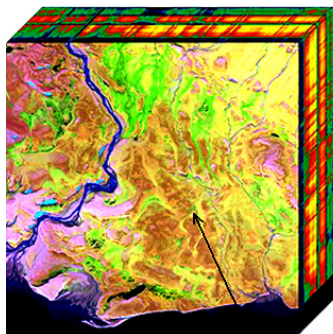
- Motivation
- Generalizing the Diffusion Distance

3 Applications

- Change Detection
- Parameterized Difference Equations

4 Empirical Approximation

Hyperspectral Imagery



Change Detection

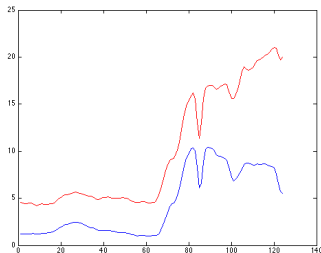


(a) September

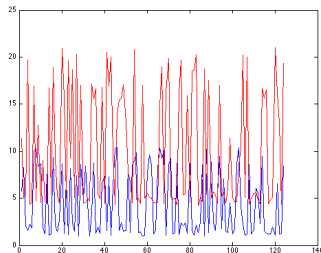


(b) October

Sample Spectra



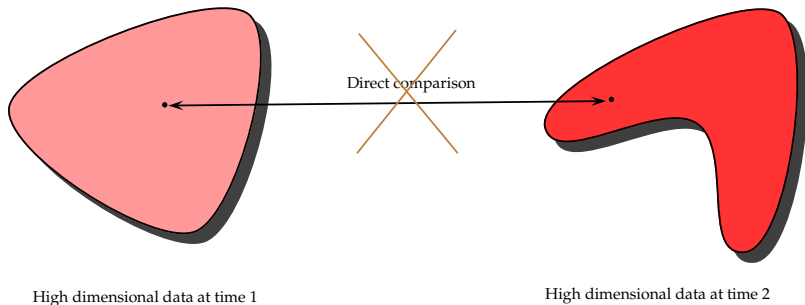
(a) Spectral signature of same pixel



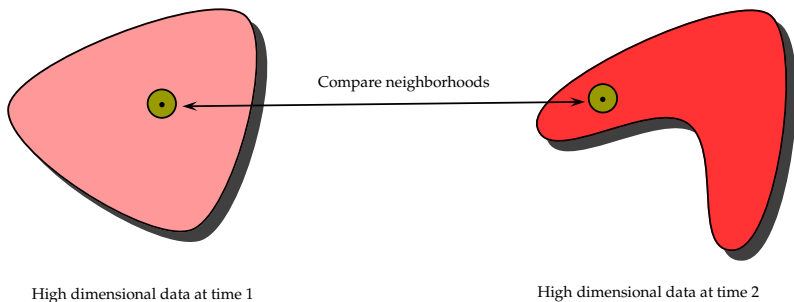
(b) Independently randomized spectral signatures

Figure: Blue: September; Red: October

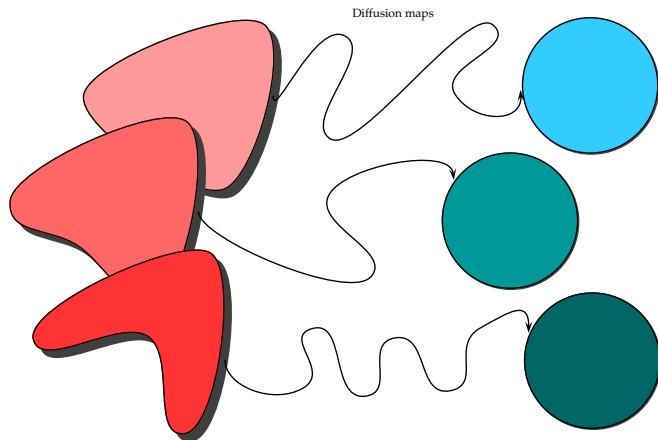
Compare Neighborhoods



Compare Neighborhoods



Multiple Low Dimensional Embeddings



Changing high dimensional data

Low dimensional embeddings

Changing Data Model

- Once again let (X, μ) be the data set.
- Let $\mathcal{I}$ denote the set of parameters (or times) for which we view/measure/observe (X, μ) .
- Encode the changes in (X, μ) across $\mathcal{I}$ in a family of kernels $\{k_\alpha : X \times X \rightarrow \mathbb{R} \mid \alpha \in \mathcal{I}\}$.
- Let $G_\alpha \triangleq (X, k_\alpha)$ denote the corresponding weighted graph for each $\alpha \in \mathcal{I}$.

Symmetric Diffusion Operators, Revisited

- For each $\alpha \in \mathcal{I}$, we have the symmetric transition kernels $a_\alpha^{(t)} : X \times X \rightarrow \mathbb{R}$ for the random walk on G_α , along with their corresponding symmetric diffusion operators:

$$(A_\alpha^t f)(x) = \int_X a_\alpha^{(t)}(x, y) \cdot f(y) d\mu(y).$$

- Let $\{\psi_\alpha^{(\ell)}\}_{\ell \geq 1}$ be the eigenfunctions of A_α with corresponding eigenvalues $\{\lambda_\alpha^{(\ell)}\}_{\ell \geq 1}$.

Generalized Diffusion Distance

■ Set $x_\alpha \triangleq (x, \alpha) \in X \times \mathcal{I}$.

Definition

For each positive integer t , let $D^{(t)} : (X \times \mathcal{I}) \times (X \times \mathcal{I}) \rightarrow \mathbb{R}$ denote the **diffusion distance** at time t , where

$$D^{(t)}(x_\alpha, y_\beta)^2 \triangleq \int_X \left(a_\alpha^{(t)}(x, s) - a_\beta^{(t)}(y, s) \right)^2 d\mu(s).$$

Diffusion Maps, Revisited

- For each $\alpha \in \mathcal{I}$ we have the diffusion map $\Psi_\alpha^{(t)} : X \rightarrow \ell^2$, which gives

$$D^{(t)}(x_\alpha, y_\alpha) = \left\| \Psi_\alpha^{(t)}(x) - \Psi_\alpha^{(t)}(y) \right\|_{\ell^2}.$$

- On the other hand, if $\alpha, \beta \in \mathcal{I}$ and $\alpha \neq \beta$, then

$$D^{(t)}(x_\alpha, y_\beta) \neq \left\| \Psi_\alpha^{(t)}(x) - \Psi_\beta^{(t)}(y) \right\|_{\ell^2}.$$

- In particular, the ℓ^2 space $\Psi_\alpha^{(t)}(X)$ is different than the ℓ^2 space $\Psi_\beta^{(t)}(X)$.

Mapping $\Psi_\alpha^{(t)}(X)$ into $\Psi_\beta^{(t)}(X)$

- Let $u = (u_\ell)_{\ell \geq 1} \in \ell^2$.
- For each $\alpha, \beta \in \mathcal{I}$, define the map $O_{\alpha \rightarrow \beta} : \ell^2 \rightarrow \ell^2$ as:

$$O_{\alpha \rightarrow \beta}(u) \triangleq \left(\sum_{m \geq 1} u_m \cdot \langle \psi_\alpha^{(m)}, \psi_\beta^{(\ell)} \rangle_{L^2(X, \mu)} \right)_{\ell \geq 1}.$$

- The map $O_{\alpha \rightarrow \beta}$ has the effect of mapping $\Psi_\alpha^{(t)}(X)$ into $\Psi_\beta^{(t)}(X)$.

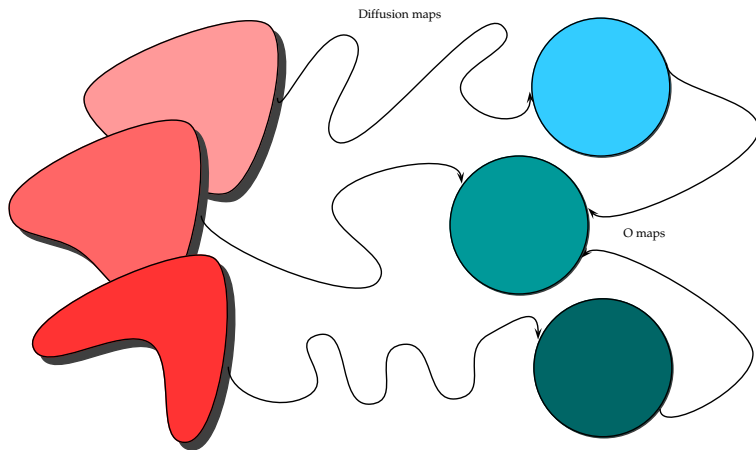
Composing the Two Maps

Theorem (Coifman and H.)

Fix $\gamma \in \mathcal{I}$. Then, for any positive integer t and $x_\alpha, y_\beta \in X \times \mathcal{I}$,

$$\begin{aligned} D^{(t)}(x_\alpha, y_\beta) &\triangleq \left\| a_\alpha^{(t)}(x, \cdot) - a_\beta^{(t)}(y, \cdot) \right\|_{L^2(X, \mu)} \\ &= \left\| O_{\alpha \rightarrow \gamma} \Psi_\alpha^{(t)}(x) - O_{\beta \rightarrow \gamma} \Psi_\beta^{(t)}(y) \right\|_{\ell^2}. \end{aligned}$$

Final Cartoon



Changing high dimensional data

Low dimensional embeddings

Overview

1 High Dimensional Data, Low Dimensional Embeddings

- Some Examples
- Diffusion Maps

2 Diffusion Maps and Changing Graphs

- Motivation
- Generalizing the Diffusion Distance

3 Applications

- Change Detection
- Parameterized Difference Equations

4 Empirical Approximation

Change Detection via Diffusion Maps



(a) September



(b) October

(c) $D^{(t)}(x_{\text{Sep}}, x_{\text{Oct}})$ as $t \rightarrow \infty$

The Standard Map

- The standard map is an area preserving chaotic map from a square with side length 2π onto itself.
- It is defined as:

$$p_{n+1} = p_n + K \cdot \sin(\theta_n)$$

$$\theta_{n+1} = \theta_n + p_{n+1}$$

- $K = 0$: the map is linear, and only periodic and quasi-periodic orbits are possible.
- $K > 0$: the map is nonlinear, and certain initial conditions give chaotic dynamics. The larger K , the more chaotic.

Diffusion Embedding of the Standard Map

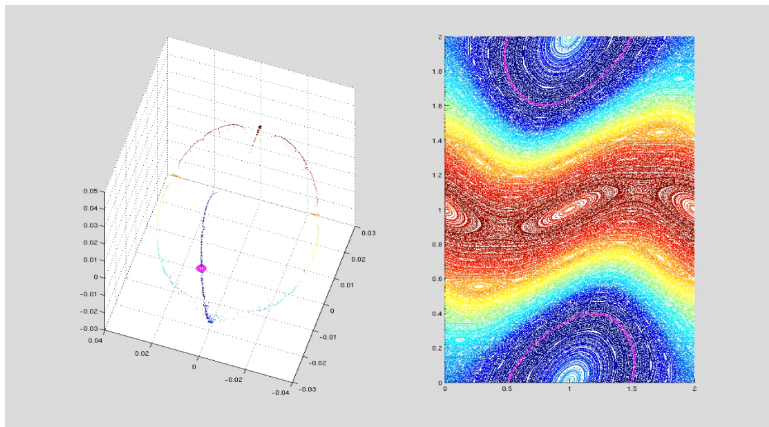


Figure: Standard map on right; diffusion embedding on left. The orbit of a particular color on the right corresponds to the embedded point of the same color on the left.

Diffusion Embedding as K Changes

Figure: Standard map on right; diffusion embedding on left.

Common Embedding

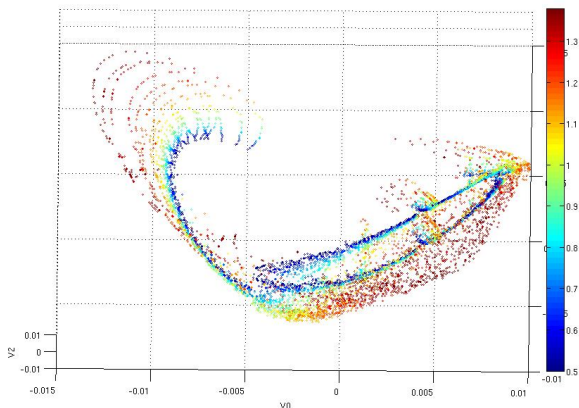


Figure: Diffusion embeddings of the standard map for various K all mapped into the same ℓ^2 space. The color of the points corresponds to the value of K (see color bar on right).

Overview

1 High Dimensional Data, Low Dimensional Embeddings

- Some Examples
- Diffusion Maps

2 Diffusion Maps and Changing Graphs

- Motivation
- Generalizing the Diffusion Distance

3 Applications

- Change Detection
- Parameterized Difference Equations

4 Empirical Approximation

Motivation

- In practice, one may only have samples $X_n = \{x^{(1)}, \dots, x^{(n)}\}$ from the continuous data set X .
- Alternatively, one may wish to sample data on hand in order to reduce computational complexity.
- In either case, can we be sure that the results we get on the samples are accurate when compared to the ideal case?

Random Sampling and Empirical Operators

- Suppose μ is a probability measure.
- Let $X_n = \{x^{(1)}, \dots, x^{(n)}\} \subset X$ be sampled i.i.d. according to μ .
- Set the following $n \times n$ matrices:

$$\mathbf{K}_\alpha[i, j] \triangleq \frac{1}{n} \cdot k_\alpha(x^{(i)}, x^{(j)}),$$

$$\mathbf{D}_\alpha[i, i] \triangleq \sum_{j=1}^n \mathbf{K}_\alpha[i, j],$$

$$\mathbf{A}_\alpha \triangleq \mathbf{D}_\alpha^{-\frac{1}{2}} \mathbf{K}_\alpha \mathbf{D}_\alpha^{-\frac{1}{2}}.$$

Empirical Diffusion Distance

Definition

For each positive integer t , let $D_n^{(t)} : (X_n \times \mathcal{I}) \times (X_n \times \mathcal{I}) \rightarrow \mathbb{R}$ denote the **empirical diffusion distance** at time t , where

$$D_n^{(t)}(x_\alpha^{(i)}, x_\beta^{(j)})^2 \triangleq n^2 \cdot \sum_{\ell=1}^n (\mathbf{A}_\alpha^t[i, \ell] - \mathbf{A}_\beta^t[j, \ell])^2.$$

- Recall that, even when evaluated just on X_n , the original diffusion distance was defined as:

$$D^{(t)}(x_\alpha^{(i)}, x_\beta^{(j)})^2 \triangleq \int_X \left(a_\alpha^{(t)}(x^{(i)}, s) - a_\beta^{(t)}(x^{(j)}, s) \right)^2 d\mu(s).$$

Approximation Theorem

Theorem Coifman and H.

Let $\tau > 0$. Then, with probability $1 - 2e^{-\tau}$,

$$D_n^{(t)}(x_\alpha^{(i)}, x_\beta^{(j)}) \approx D^{(t)}(x_\alpha^{(i)}, x_\beta^{(j)}),$$

to within accuracy $C \cdot \frac{\sqrt{\tau}}{\sqrt{n}}$.

Thank you

www.math.yale.edu/~mh644