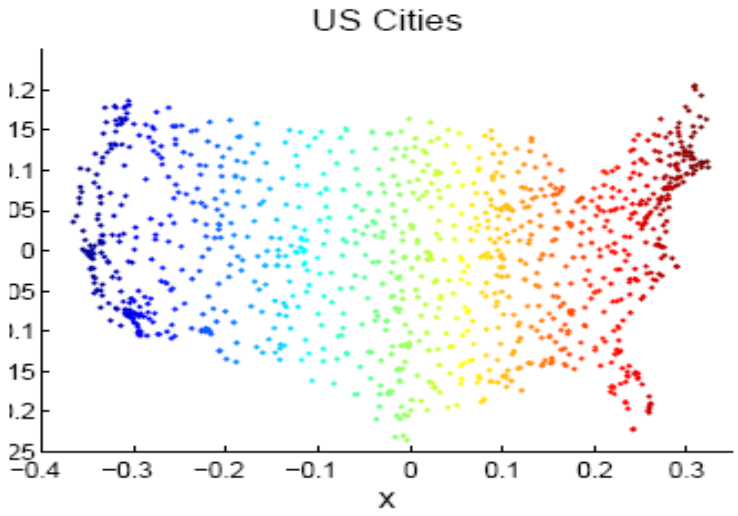
Eigenfunctions of Lapalace operators as Integrators of local models.

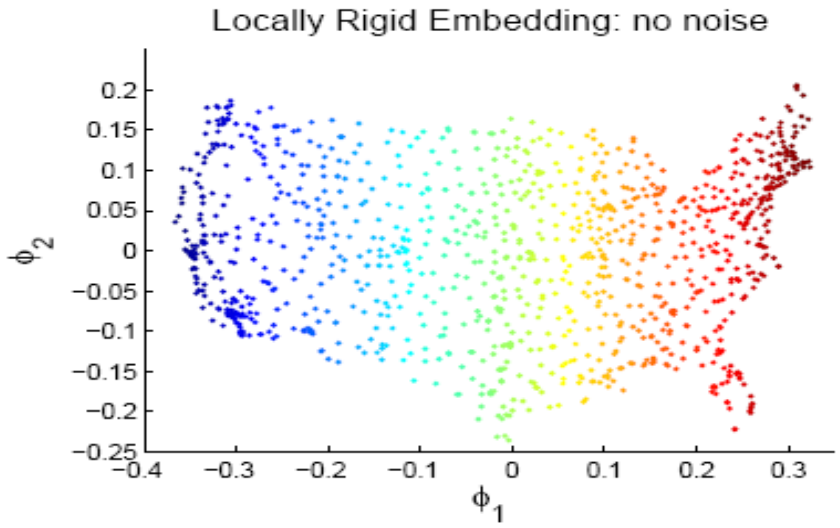
Ronald Coifman,Yale
Amit Singer, Princeton

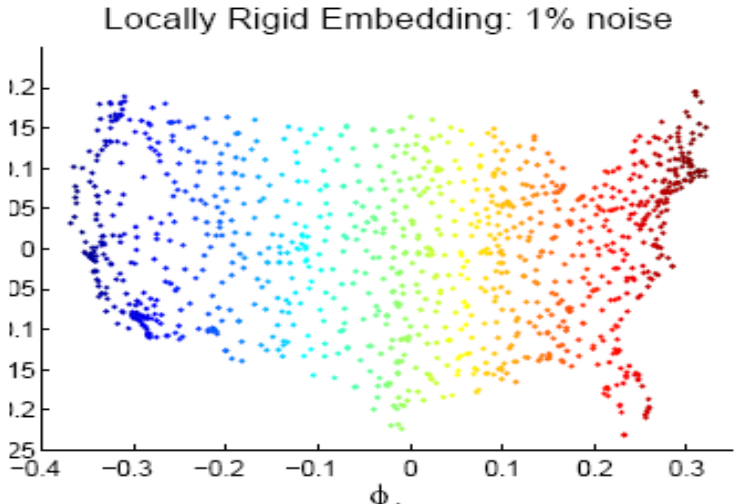
Our goal is to show that a variety of linear and nonlinear problems for which some local description exists can be integrated into global consistent models by finding the first few eigenvectors of an "***appropriate***" Laplace operator .

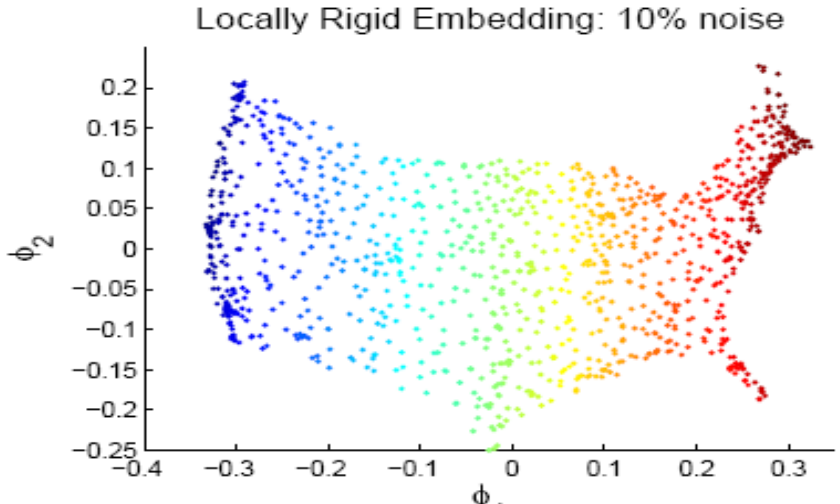
Examples include linear and nonlinear difference equations on graphs (sensor localization problem), nonlinear independent components analysis (inverse problems, intrinsic variables) , extraction of slow variables in dynamical systems , ***and of course dimensional reduction and knowledge organization.***

The sensor network local to global positioning problem. For each city in the US we know the distance to a few neighbors , how do we get the global position ?









Let P_i be the location of city (or sensor) i . From the knowledge of the distance to a few neighbors P_j we can easily calculate, from local connections, weights $A_{i,j}$ so that :

$$P_i = \sum_{j \sim i} A_{i,j} P_j \text{ where } \sum_{j \sim i} A_{i,j} = 1$$

Clearly both x and y coordinates (as well as 1) are eigenvetors of the matrix A, The coordinates satisfy the mean value theorem and are Harmonic. The matrix A is a local encapsulation of the relation between cities.

Generalization of the fundamental theorem of calculus .

Assume that at each site you know the difference of altitude between cities and some of their neighbours we get the global function as the z eigenfunction of the 3 dimensional version .

Basically find the altitude function from its local increments.

We observe that we can easily solve the Poisson equation on graphs $\Delta u=f$, where $\Delta = I - A$,

and A is any local averaging operator.

In fact let B= A+αAσ(I-A) with $\alpha = \frac{f}{A|f|}$, $\sigma = \mathrm{sgn}(f)$,

It is easy to check that

$$Bu=u$$

and therefore the solution to the Poisson equation is an eigenvector of B with eigenvalue 1. This is basically the obvious observation that we can reduce any nonhomegeneous linear problem to a homegeneous one by multiplying by a linear operator with f the only vector in its kernel.

Another example of a similar nonlinear equation that can be solved through the eigevectors of an appropriate Laplace operator is the following;

Invert the map f from R^n to R^m knowing only the transported metric at image points , i,e

$g_{i,j} = \sum_k f_{i,k} f_{j,k}$ where $f_{j,k} = \partial_{x_k} f_j$ j=1...m, k=1...n.

ie compute f^{-1} from $g_{i,j}$

Parametrize the domain of f by the eigenfunctions of a Laplace Beltrami operator ,(or a variant).

Around an image point use g^{-1} to define the metric

(so called Mahalanobis metric).and use the eigenvectors

of the corresponding Laplace operator as coordinates on the image .

Since the Laplace operator on the image was defined

to be identical to the Laplace operator on the domain of f ,

these coordinates give the inverse map f^{-1} modulo an isometry

on the domain of f.

A simple empirical diffusion matrix A can be constructed as follows

Let X_i ***represent normalized data ,we "soft truncate" the "affinity" matrix as***

$$A_{j,i} = \frac{\exp(-\left\| X_j - X_i \right\|^2 / \varepsilon)}{\sum_i \exp(-\left\| X_j - X_i \right\|^2 / \varepsilon)}$$

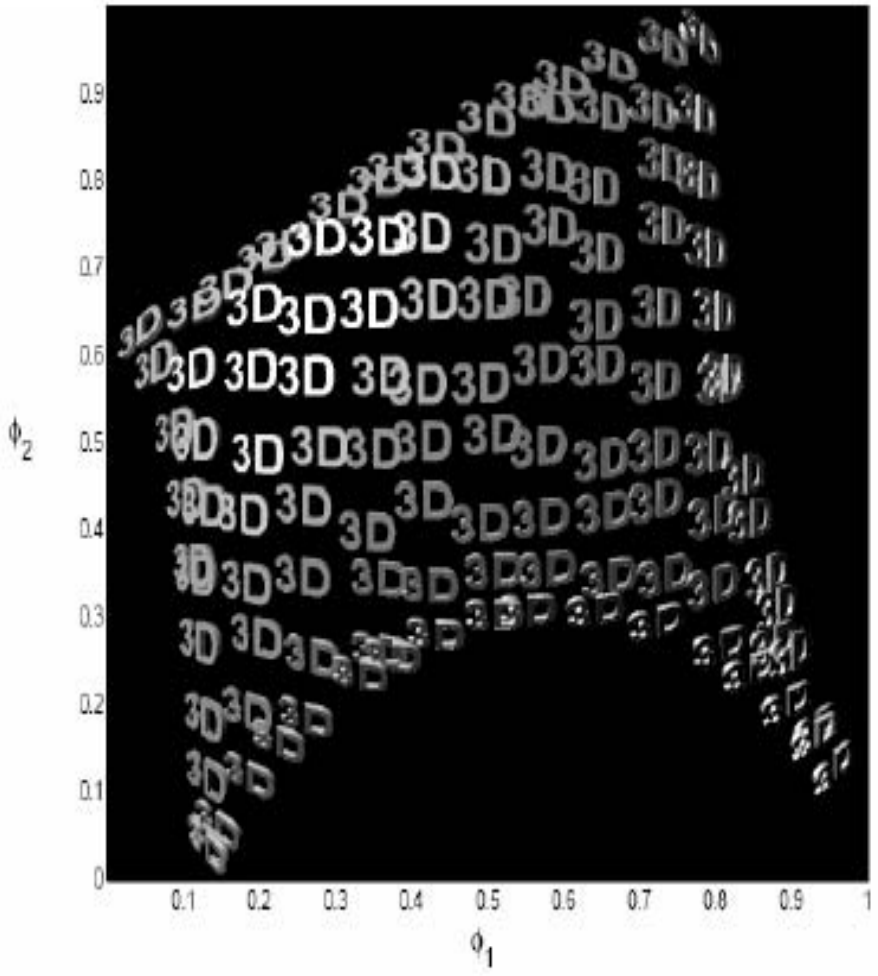
The eigenvectors of this matrix provide a local non linear local model of the data . Whose entries are the diffusion coordinates
These are also the eigenfunctions of the discrete Graph Laplace Operator.

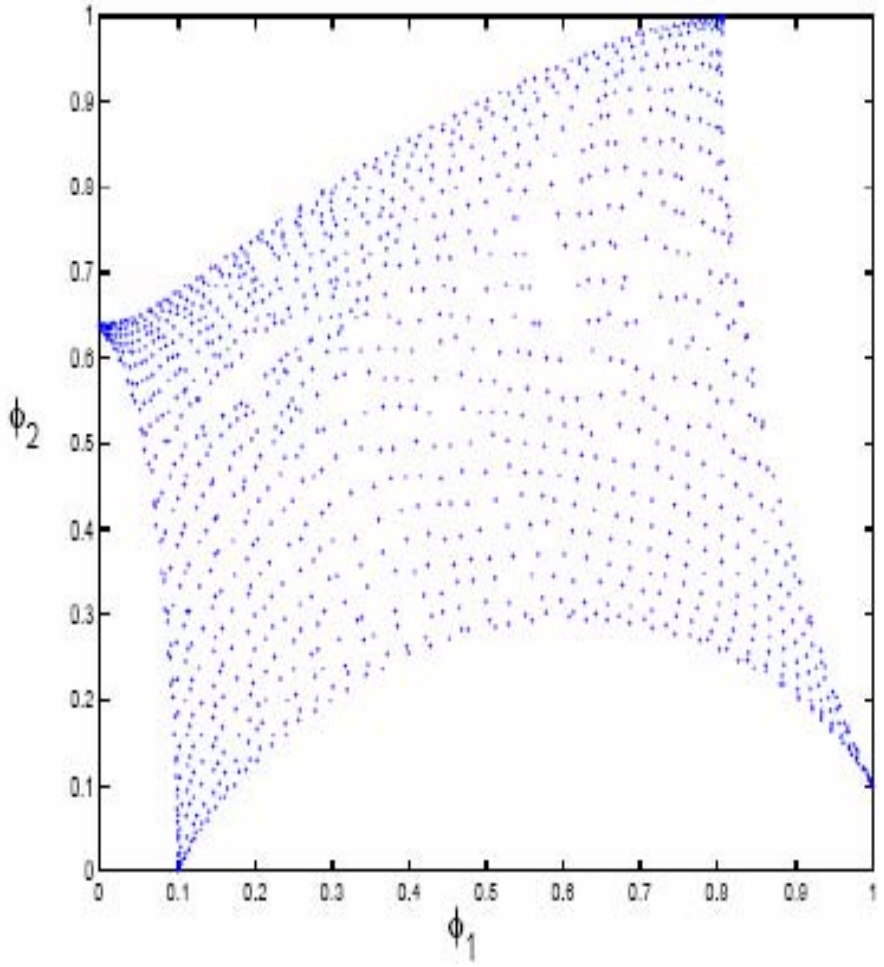
$$A^t = \sum \lambda_l^{2t} \phi_l(X_i)\phi_l(X_j) = a_t(X_i, X_j)$$

$$X_i^{(t)} \to (\lambda_1^t \phi_1(X_i), \lambda_2^t \phi_2(X_i), \lambda_3^t \phi_3(X_i), ..)$$

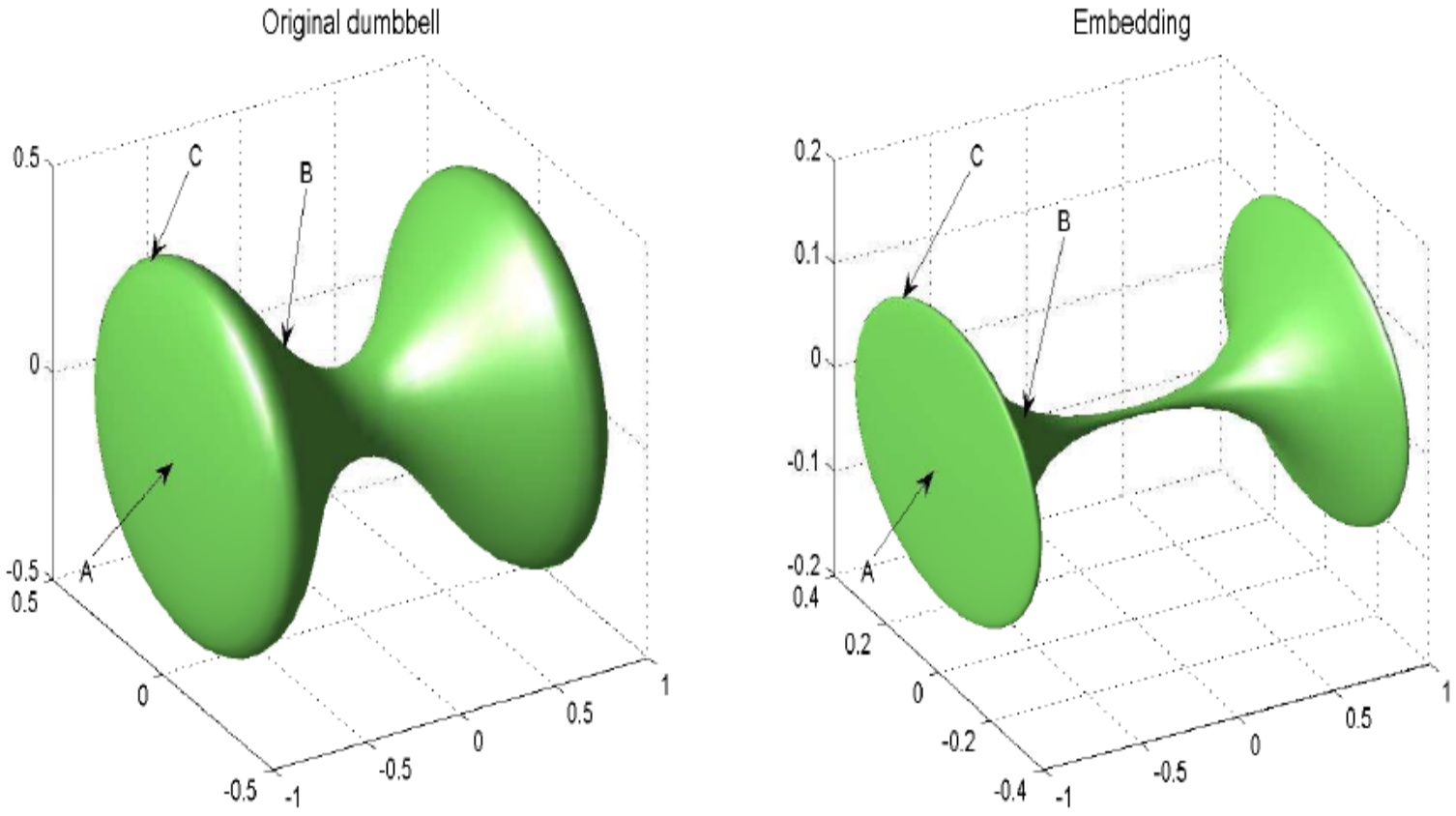
$$d_t^2(X_i, X_j) = a_t(X_i, X_i) + a_t(X_j, X_j) - 2a_t(X_i, X_j) = \left\| X_i^{(t)} - X_j^{(t)} \right\|^2$$

This map is a diffusion (at time t) embedding into Euclidean space





The First two eigenfunctions organize the small images which were provided in random order, in fact assembling the 3D puzzle.



The natural diffusion on the surface of the dumbell is mapped out in the embedding . Observe that A is closer to B than to C ,and that the two lobes are well separated by the bottleneck.

Construction of the family of diffusions

(1) Fix $\alpha \in \mathbb{R}$, and a rotation-invariant kernel $k_\varepsilon(x,y) = h\left(\frac{\|x-y\|^2}{\varepsilon}\right)$.

(2) Let

$$q_\varepsilon(x) = \int_X k_\varepsilon(x,y) q(y) dy \,,$$

and form the new kernel

$$k_\varepsilon^{(\alpha)}(x,y) = \frac{k_\varepsilon(x,y)}{q_\varepsilon^\alpha(x) q_\varepsilon^\alpha(y)} \,.$$

(3) Apply the weighted graph Laplacian normalization to this kernel by setting

$$d_\varepsilon^{(\alpha)}(x) = \int_X k_\varepsilon^{(\alpha)}(x,y) q(y) dy \,,$$

and by defining the anisotropic transition kernel

$$p_{\varepsilon,\alpha}(x,y) = \frac{k_\varepsilon^{(\alpha)}(x,y)}{d_\varepsilon^{(\alpha)}(x)} \,.$$

$$L_{\varepsilon,\alpha} = \frac{I - P_{\varepsilon,\alpha}}{\varepsilon}$$

be the infinitesimal generator of the Markov chain. Then for a fixed $K > 0$,

we have on E_K

$$\lim_{\varepsilon \to 0} L_{\varepsilon,\alpha} f = \frac{\Delta(f q^{1-\alpha})}{q^{1-\alpha}} - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}} f \,.$$

In other words, the eigenfunctions of $P_{\varepsilon,\alpha}$ *can be used to approximate those of the following symmetric Schrödinger operator:*

$$\Delta\phi - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}\phi,$$

where $\phi = f q^{1-\alpha}$.

Let's write $q = e^{-U}$, then the generator becomes

$$\Delta\phi - \left(\frac{\|\nabla U\|^2}{4} - \frac{\Delta U}{2}\right)\phi\,.$$

It is shown in (29) that a simple conjugation of this specific Schrödinger operator leads to the forward Fokker-Plank equation

$$\frac{\partial q}{\partial t} = \nabla \cdot (\nabla q + q\nabla U)\,,$$

where $q(x,t)$ represents the density of points at position x and time t of a dynamical system satisfying the Langevin equation

$$\dot{x} = -\nabla U(x) + \sqrt{2}\dot{w}\,, \tag{2}$$

We now describe the inverse map problem mentioned before , as a tool for data analysis and intrinsic data parameterization.

The basic model provides for an intrinsic organization of data driven by a stochastic process or by a black box f which is a nonlinear mapping from an unknown parameter space (independent parameters) into high dimensions , we assume that f is smooth and invertible on its range .

The goal is to extract a parameterization which is independent of f or the mode of observations .

We achieve this by locally undistorting the observations using the inverse of the square Jacobian of f , as computed through the local covariance matrix of the data

Asymptotically the local covariance of the data is computable through the Jacobian of f , this information suffices to construct the inverse map specifically:

The basic idea is that any smooth map between smooth manifolds $f : \mathcal{M}_X \mapsto \mathcal{M}_Y$ can be linearly approximated in a local neighborhood of any given point by its differential. The first-order Taylor expansion near x_0 reads

$$y = f(x) = y_0 + J_f(x_0)(x - x_0) + O\left(\|x - x_0\|^2\right), \tag{38}$$

where $J_f(x_0)$ is the Jacobian of f at x_0 and $y_0 = f(x_0)$.[3] This gives a first-order approximation for the distances

$$\|y - y_0\|^2 = \left\|J_f(x_0)(x - x_0)\right\|^2 + O\left(\|x - x_0\|^3\right). \tag{39}$$

Similarly, for the inverse map $f^{-1} : \mathcal{M}_Y \mapsto \mathcal{M}_X$ we have

$$\|x - x_0\|^2 = \left\|J_{f^{-1}}(y_0)(y - y_0)\right\|^2 + O\left(\|x - x_0\|^3\right), \tag{40}$$

Nonlinear independent components Analysis and the discrete graph laplacian L

$$L = D^{-1}W - I,$$

$$W_{ij} = \exp\left\{-\frac{\|J^{-1}(\boldsymbol{y}^{(i)})(\boldsymbol{y}^{(j)} - \boldsymbol{y}^{(i)})\|^2 + \|J^{-1}(\boldsymbol{y}^{(j)})(\boldsymbol{y}^{(j)} - \boldsymbol{y}^{(i)})\|^2}{4\varepsilon}\right\}$$

$$\frac{1}{N}\sum_{j=1}^{N} W_{ij} q_Y(\boldsymbol{y}^{(j)}) \to \tag{4.2}$$

$$\int_Y \exp\left\{-\frac{\|J^{-1}(\boldsymbol{y}^{(i)})(\boldsymbol{y} - \boldsymbol{y}^{(i)})\|^2 + \|J^{-1}(\boldsymbol{y})(\boldsymbol{y} - \boldsymbol{y}^{(i)})\|^2}{4\varepsilon}\right\} q_Y(\boldsymbol{y}) p_Y(\boldsymbol{y})\, d\boldsymbol{y},$$

Leading to $$\mathcal{L}q = \Delta q - \nabla U \cdot \nabla q,$$

This Fokker Plank operator ın the x coordınates ,is a sum of independent operators in each variable whose first eigenvectors are monotone in the x variables, these are the intrinsic variables .

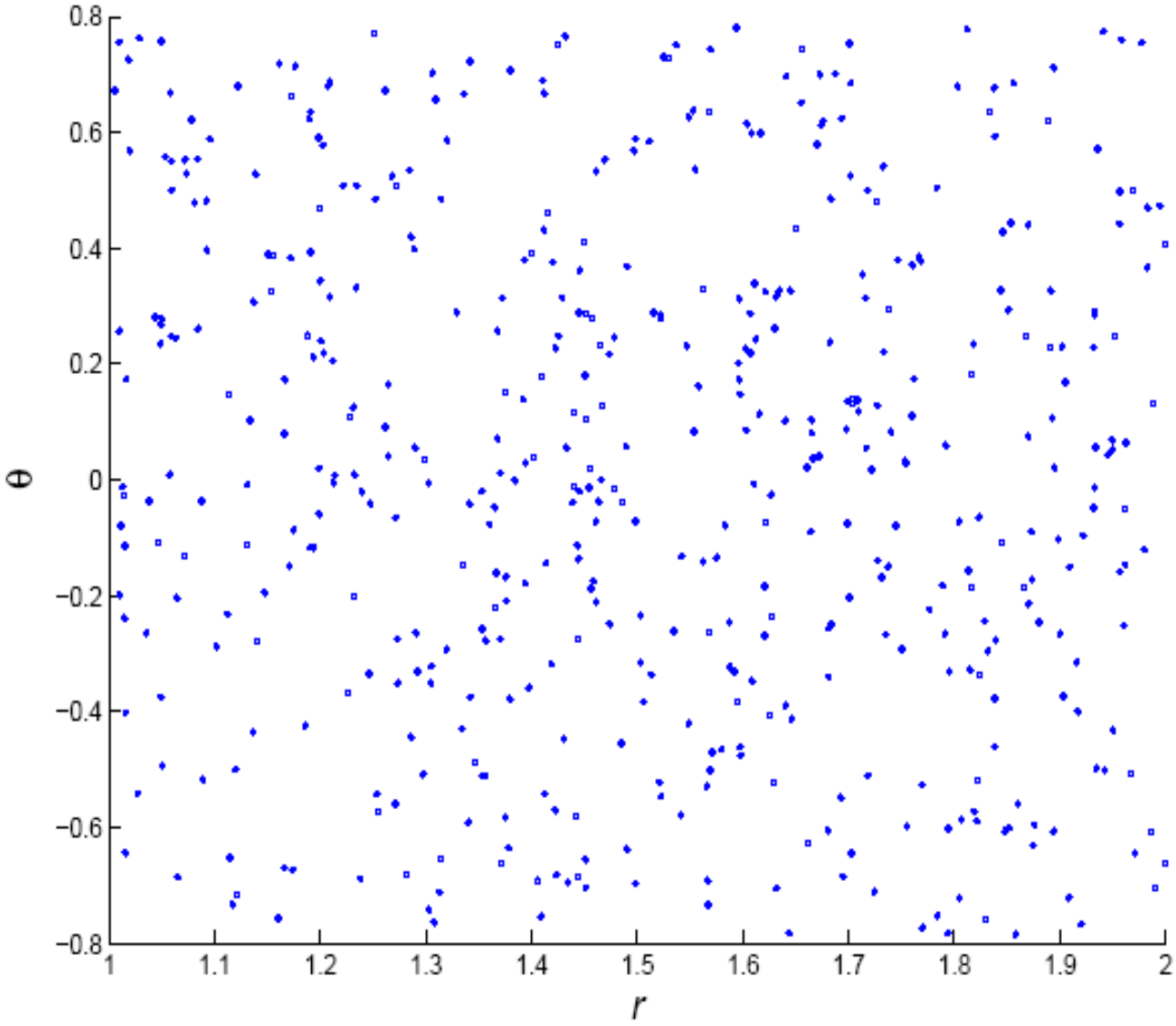


Figure 1: Data sample of $N = 500$ points drawn randomly from the stationary distributions of r (5.1) and θ (5.2).

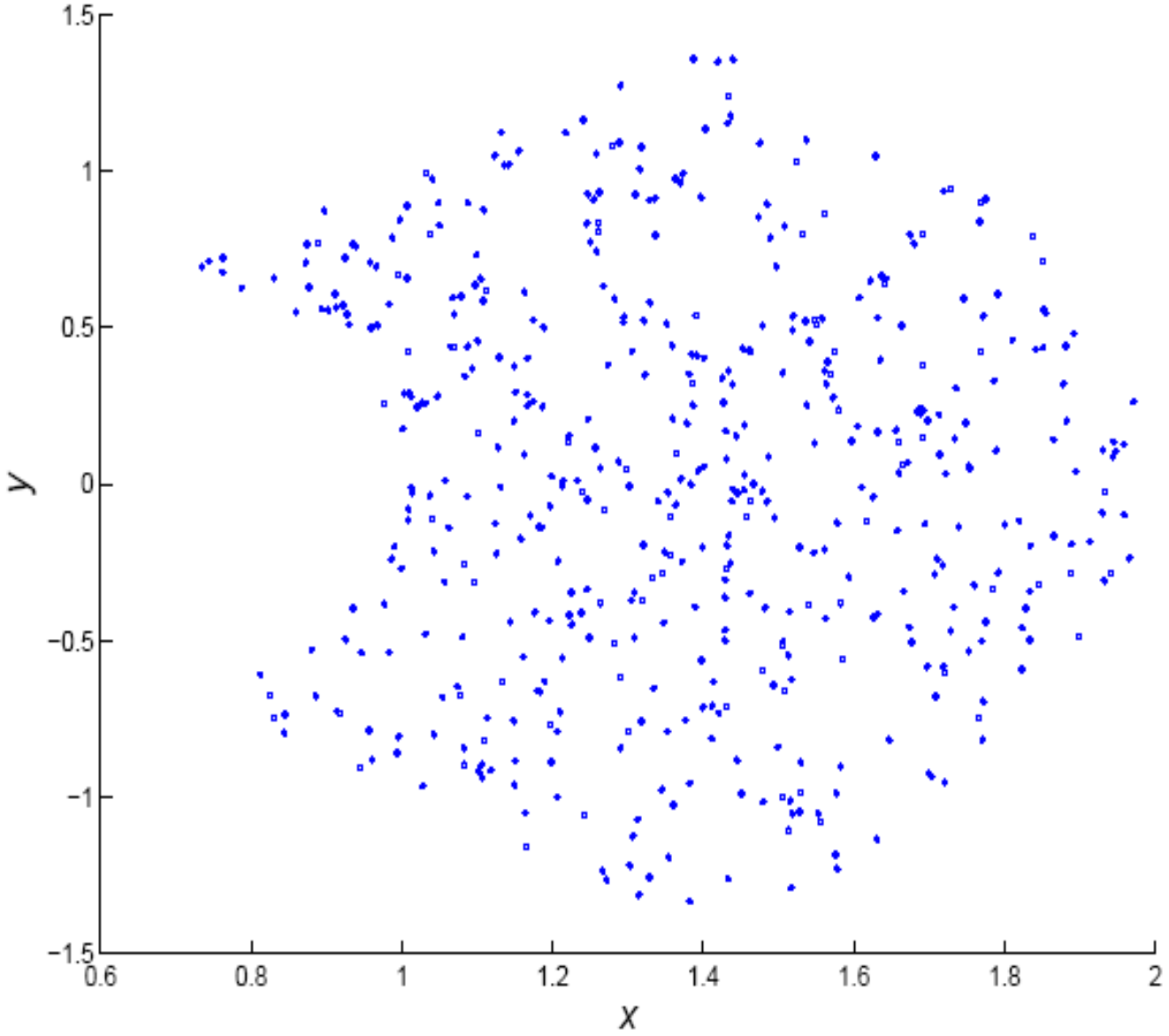


Figure 2: Cartesian coordinates of data points $x^{(i)} = r^{(i)} \cos\theta^{(i)}$, $y^{(i)} = r^{(i)} \sin\theta^{(i)}$, $i = 1, \ldots, N$.

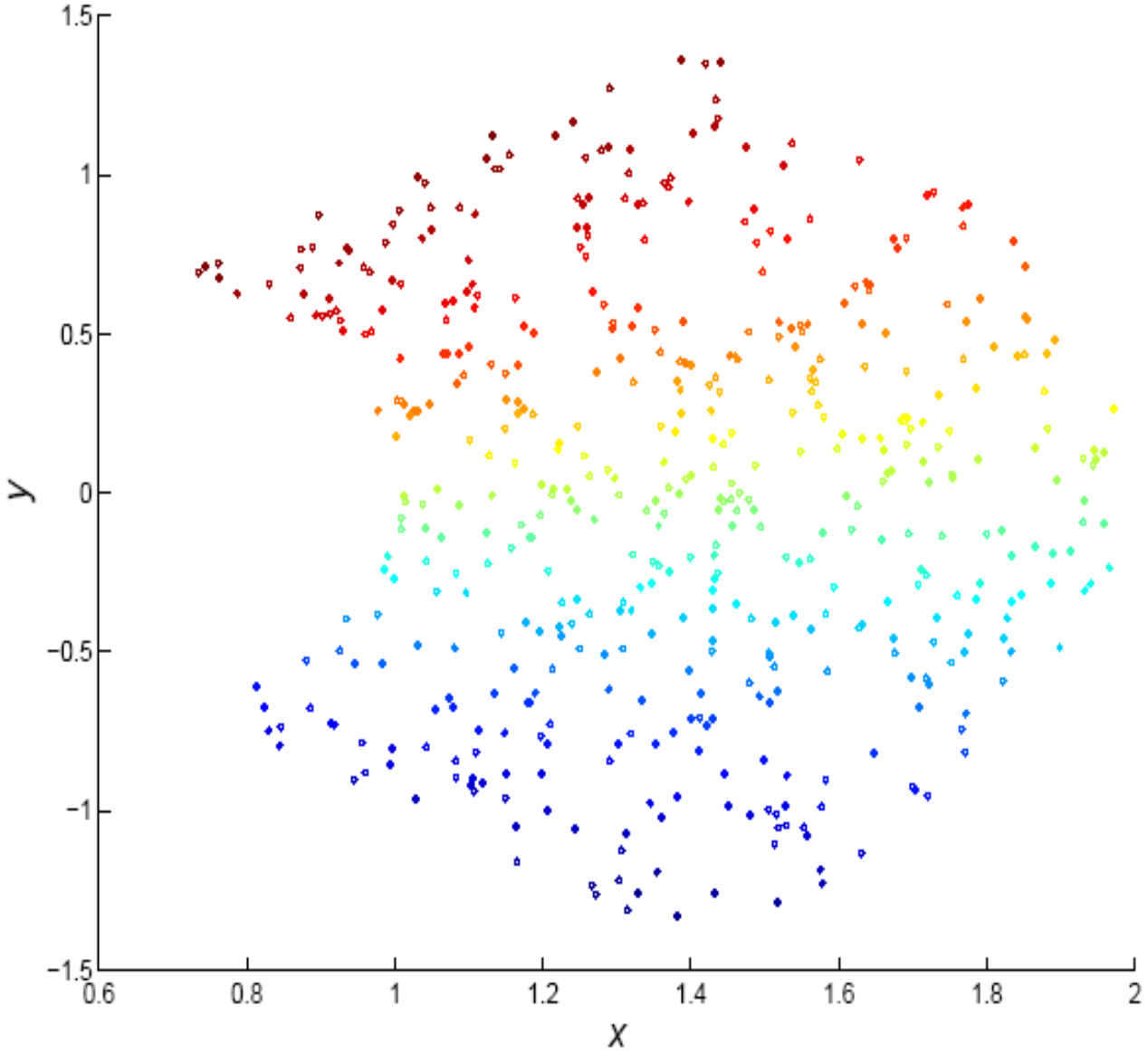


Figure 3: Color map plot of ϕ^1 shows its level lines to be rays aminating from the origin, thus ϕ^1 reveals the θ coordinate.

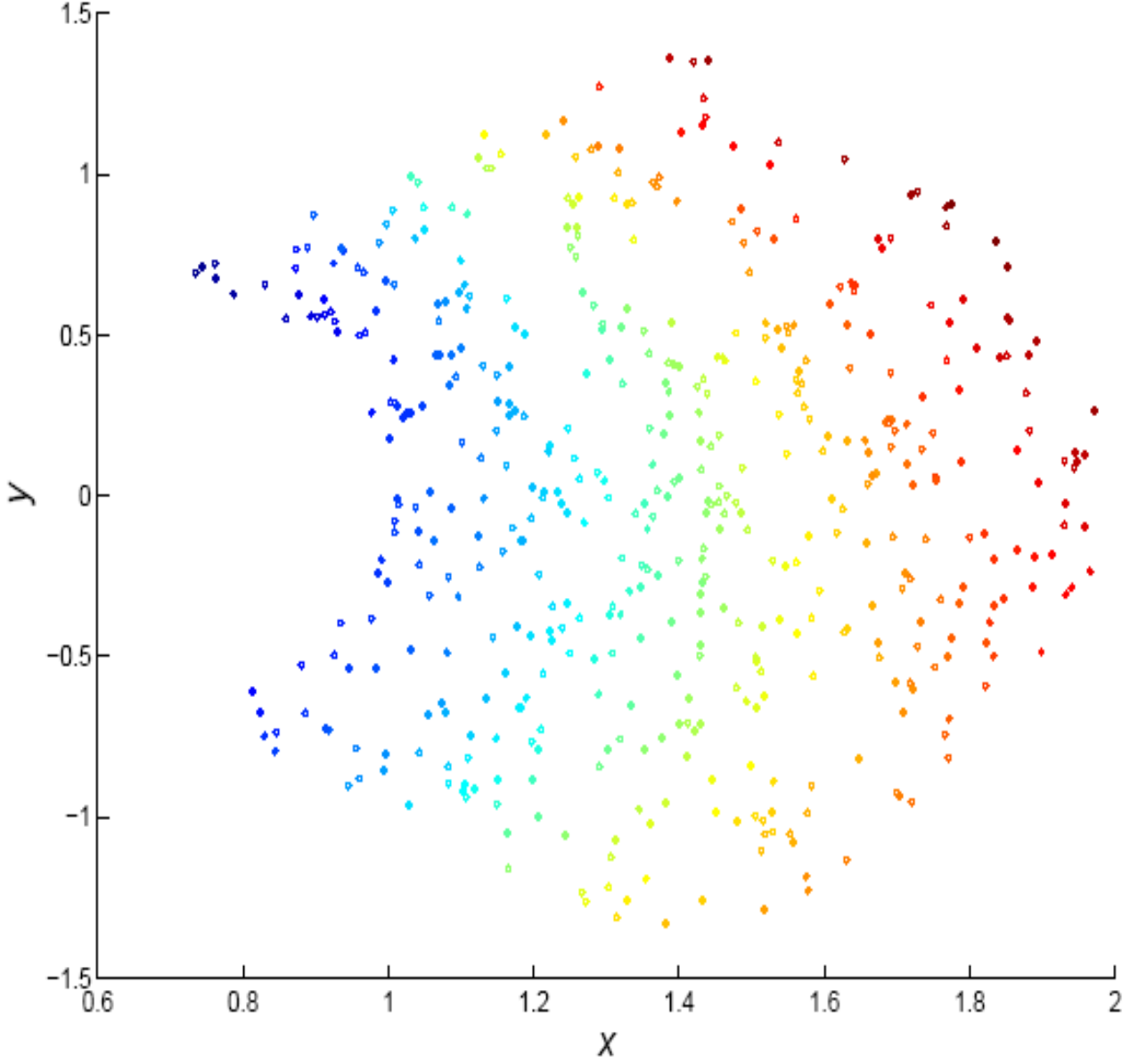


Figure 4: Color map plot of ϕ^2 shows its level lines to be concentric circles, thus ϕ^2 reveals the r coordinate.

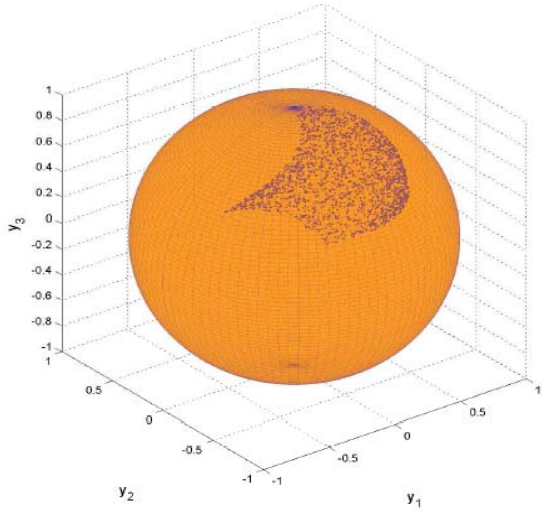
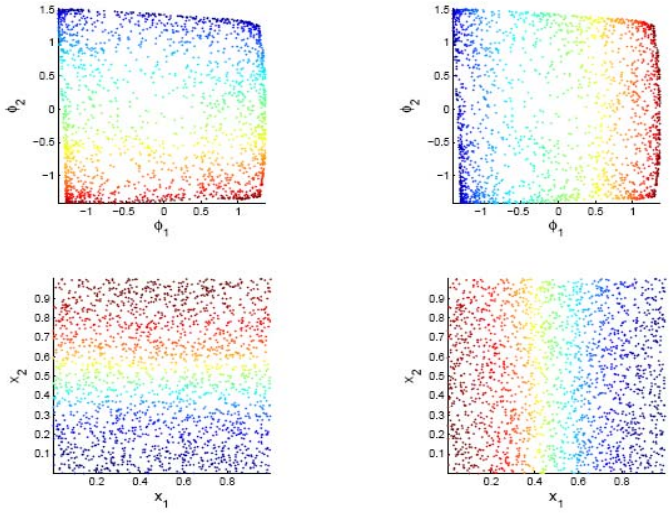


Fig. 6. Points (y_1, y_2, y_3) on the unit sphere are obtained by mapping uniformly sampled points in the unit square by (52).



Oscillating "half-moons"

Next, consider the system of stochastic differential equations

$$\mathrm{d}u = a_1\,\mathrm{d}t + a_2\,\mathrm{d}w_1, \qquad [20]$$

$$\mathrm{d}v = a_3(1-v)\,\mathrm{d}t + a_4\,\mathrm{d}w_2, \qquad [21]$$

where a_i, $i = 1, 2, 3, 4$, are constants and w_i, $i = 1, 2$ are independent δ-correlated white noises (Wiener processes). We consider (20)–(21) together with the following nonlinear transformation of variables

$$x = v\cos(u+v-1), \qquad y = v\sin(u+v-1). \qquad [22]$$

We will assume that the observables x and y are the actual observables, while u and v are unknown. We choose the values of parameters as: $a_1 = a_2 = 10^{-3}$, $a_3 = a_4 = 10^{-1}$.

The illustrative trajectory which starts at $[x(0), y(0)] = [1, 0]$ is plotted in the left panel of Figure 3. The trajectory is colored according to time. We run simulations for a longer time 8×10^4, which accounts for about 12-13 rotations, and record 2000 data points at equidistant time intervals of length $8 \times 10^4 / 2000 = 40$. This data set is plotted in the middle panel of Figure 3. Again, points are colored according to time. We clearly see that there is no correlation between time and the slow variable (which is u MOD 2π) because of oscillations.

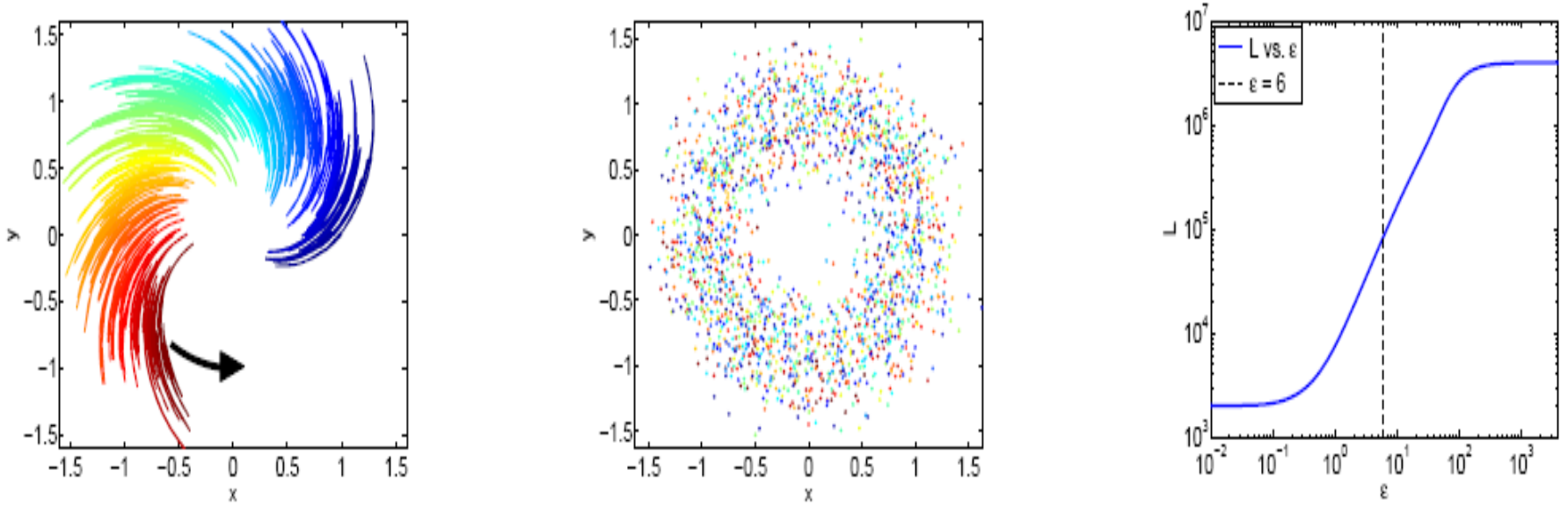


Fig. 3: *Oscillating half moons: the short illustrative trajectory of* (20)–(22) *which starts at* $[x(0), y(0)] = [1, 0]$. *The trajectory is colored according to time (left panel). The representative data set sampled at equal time steps from a longer stochastic simulation. The points are colored according to time (middle panel). Plot of* $L(\varepsilon)$ *given by* (23) *(right panel).*

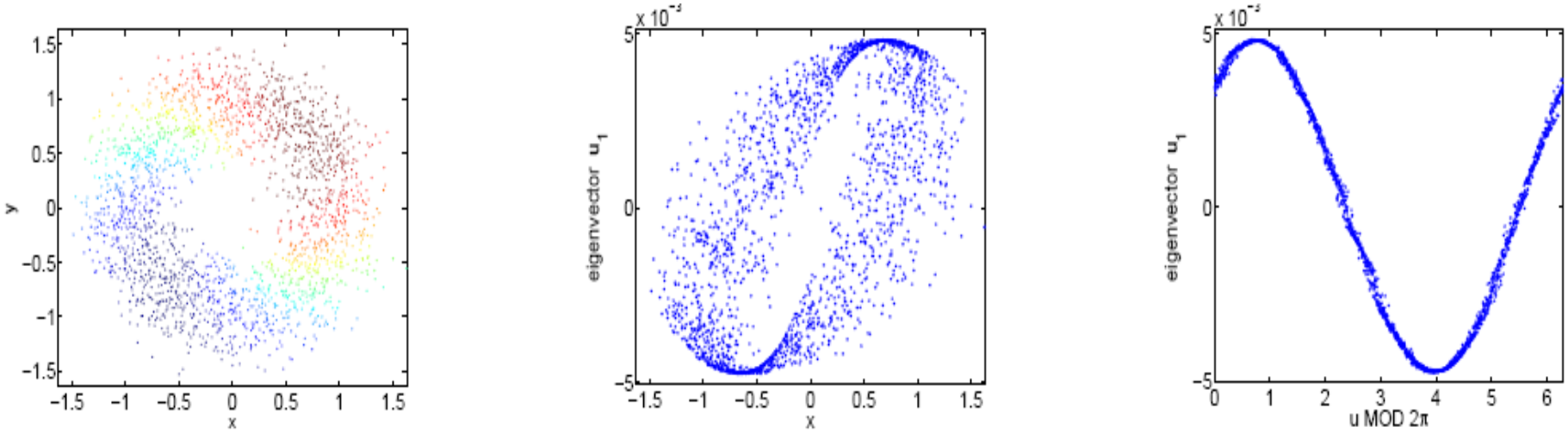


Fig. 4: *Oscillating half moons: the data set with each point colored according to* $\mathbf{u}_1$ *(left panel). Vector* $\mathbf{u}_1$ *as a function of* x *(middle panel). Vector* $\mathbf{u}_1$ *as a function of* u *MOD* 2π *(right panel).*

Inherently non-linear chemical reactions

We consider the following set of chemical reactions

$$X \xrightarrow{k_1} X + Z, \qquad Y + Z \xrightarrow{k_2} Y, \qquad [24]$$

$$\emptyset \xrightarrow{k_3} Y, \qquad Y \xrightarrow{k_4} \emptyset, \qquad [25]$$

$$\emptyset \xrightarrow{k_5} X. \qquad [26]$$

The first two reactions (24) are production and degradation of Z (catalyzed by X and Y, respectively). The production and degradation of Z is assumed to be happening on a fast time scale. Reactions (25) are production and degradation of Y. They are assumed to occur on an intermediate time scale (i.e. slower than the fast time scale, but faster than the slow time scale). The reaction (26) is production of X which is assumed to be slow. We choose the values of the rate constants as

$$k_1 = 1000, \quad k_2 = 1, \quad k_3 = 40, \quad k_4 = 1, \quad k_5 = 1. \qquad [27]$$

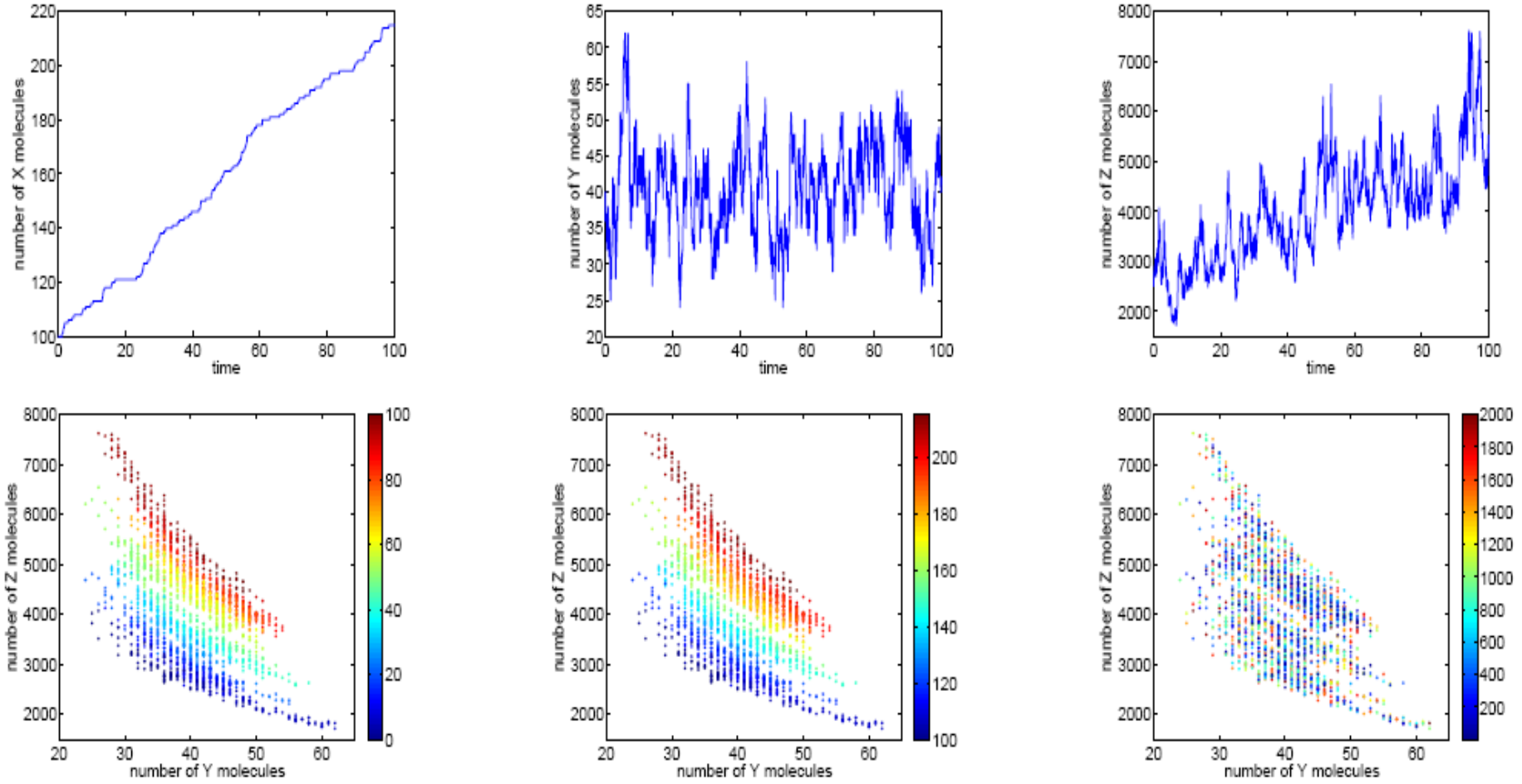


Fig. 5: *Inherently non-linear chemical reactions: the time evolution of X (top left panel), Y (top middle panel) and Z (top right panel) given by the stochastic simulation of the chemical system (24) – (26). The same trajectory (2000 data points, saved at equal time intervals $\Delta t = 0.05$ apart) plotted in the Y-Z plane is shown in the bottom panels. We color the points according to time (bottom left panel) and according to the number of X molecules (bottom middle panel). To emphasize the strength of our approach, we randomize the order of the data points – we color the resulting data set according to the order in the new list in the right panel (bottom right panel).*

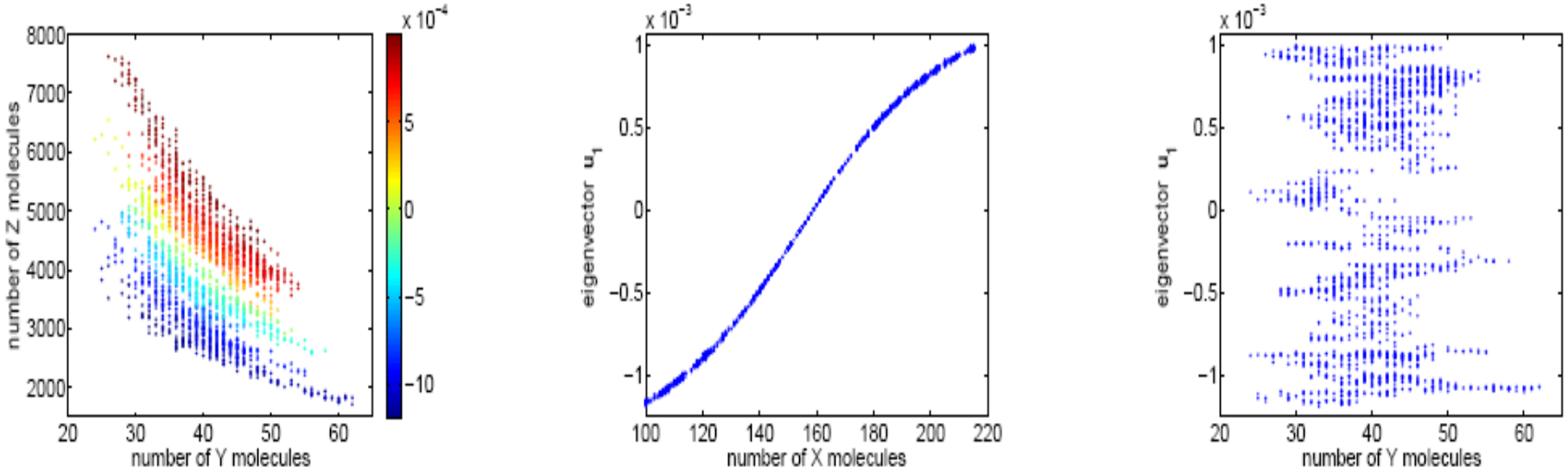
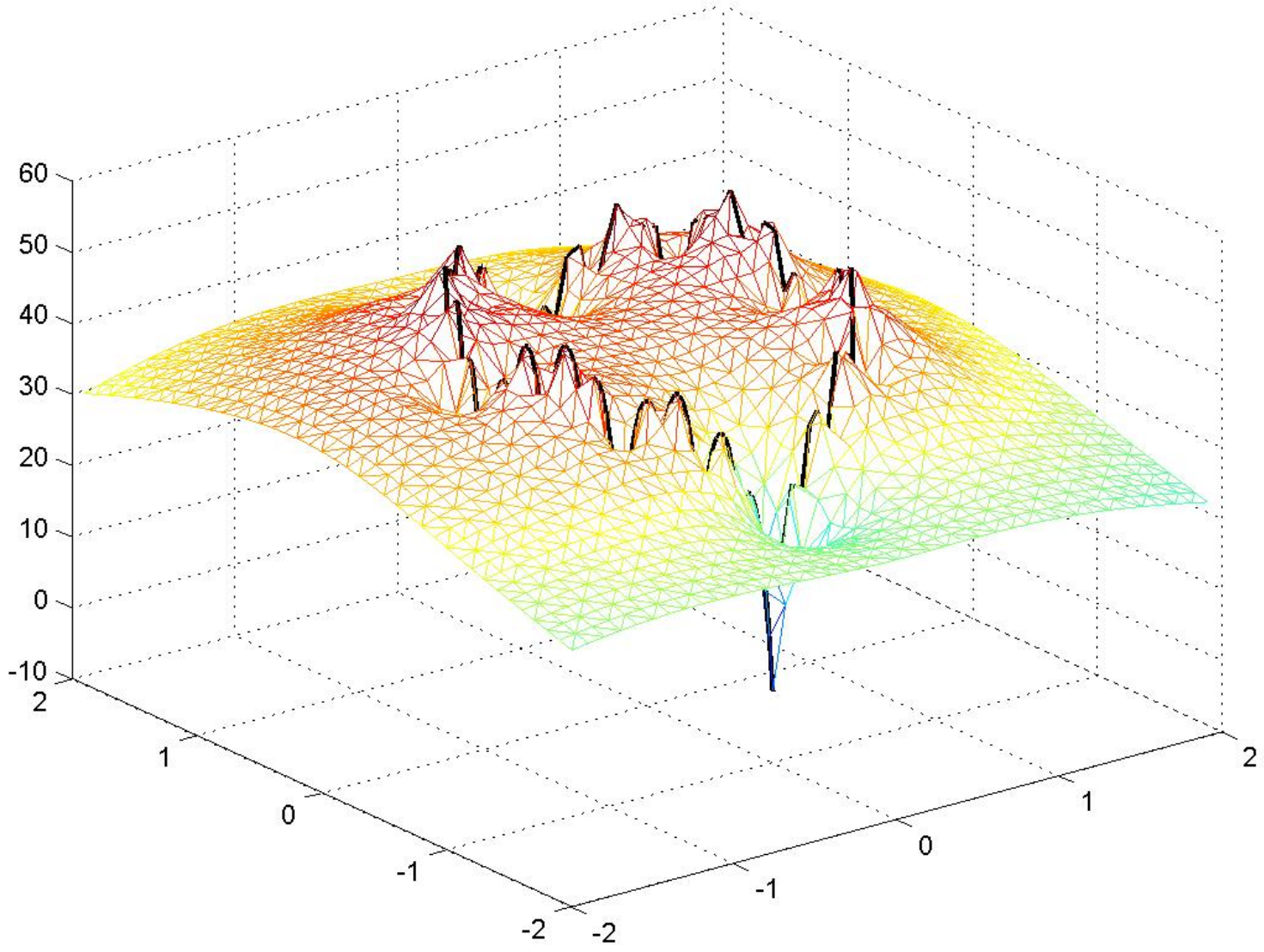


Fig. 6: *Inherently non-linear chemical reactions: the data set in the Y-Z plane with each point colored according to* $\mathbf{u}_1$ *(left panel). Vector* $\mathbf{u}_1$ *as a function of X (middle panel). Vector* $\mathbf{u}_1$ *as a function of Y (right panel).*

For these intrinsic parametrizations to be useful for modeling and prediction it is essential to extend their definition beyond the known empirical data .

This can be achieved by extending inverse of the local data correlation matrix ,viewed as an empirical matrix valued function on the data , our basic assumption is that this function is smooth and therefore can be extended by the multiscale algorithm already discussed previously .

After such an extension has been achieved , we can use the Gaussian defined by the extended Jacobian to extend the corresponding coordinates , this should be compared to a direct extension of the coordinates as described below .


60
50
40
30
20
10
0
-10
2
1
0
-1
-2
-2
-1
0
1
2

Multiscale Littlewood Paley expansions with Gaussian kernels at different scales, described below

We consider an empirical function f defined on a data set X_i
We diagonalize various scaled Gaussian kernels restricted to the data as follows;

$e^{-4^m|x_i-x_j|^2} = \sum_l \lambda^m{}_l \Phi^m{}_l(x_i)\Phi^m{}_l(x_j)$ each $\Phi^m{}_l$ has an extension " to distance $\approx 2^{-m}$ " given by

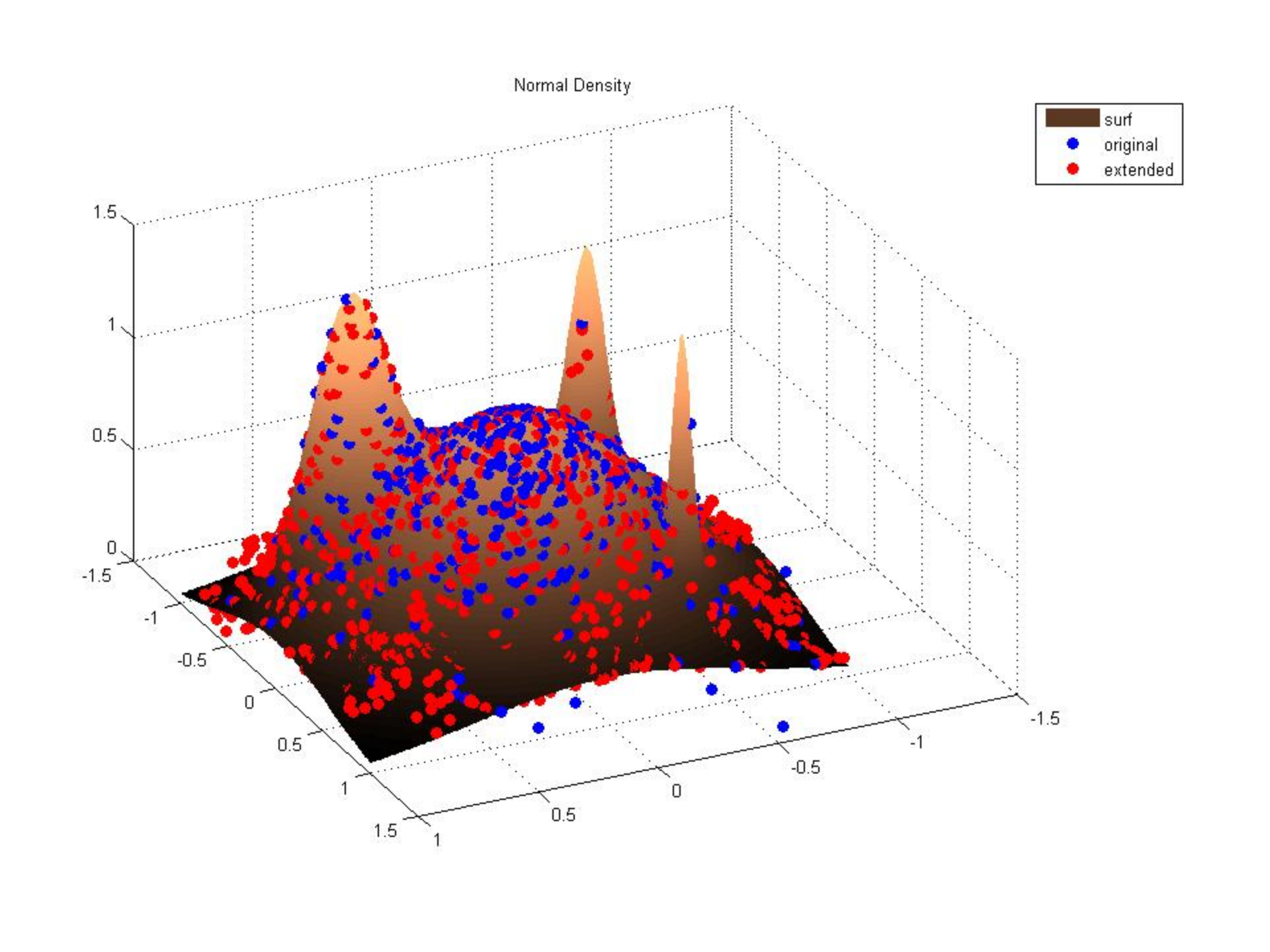
$$\Phi^m{}_l(x) = \frac{1}{\lambda^m{}_l}\sum_i e^{-4^m|x-x_i|^2}\Phi^m{}_l(x_i) \qquad \text{if } \lambda^m{}_l / \lambda^m{}_0 \geq \delta = 0.1$$

We define the projection $\sum_{l;} \left\langle g, \Phi^m{}_l \right\rangle \Phi^m{}_l(x_i) = P_m(g)$ where l are as above.

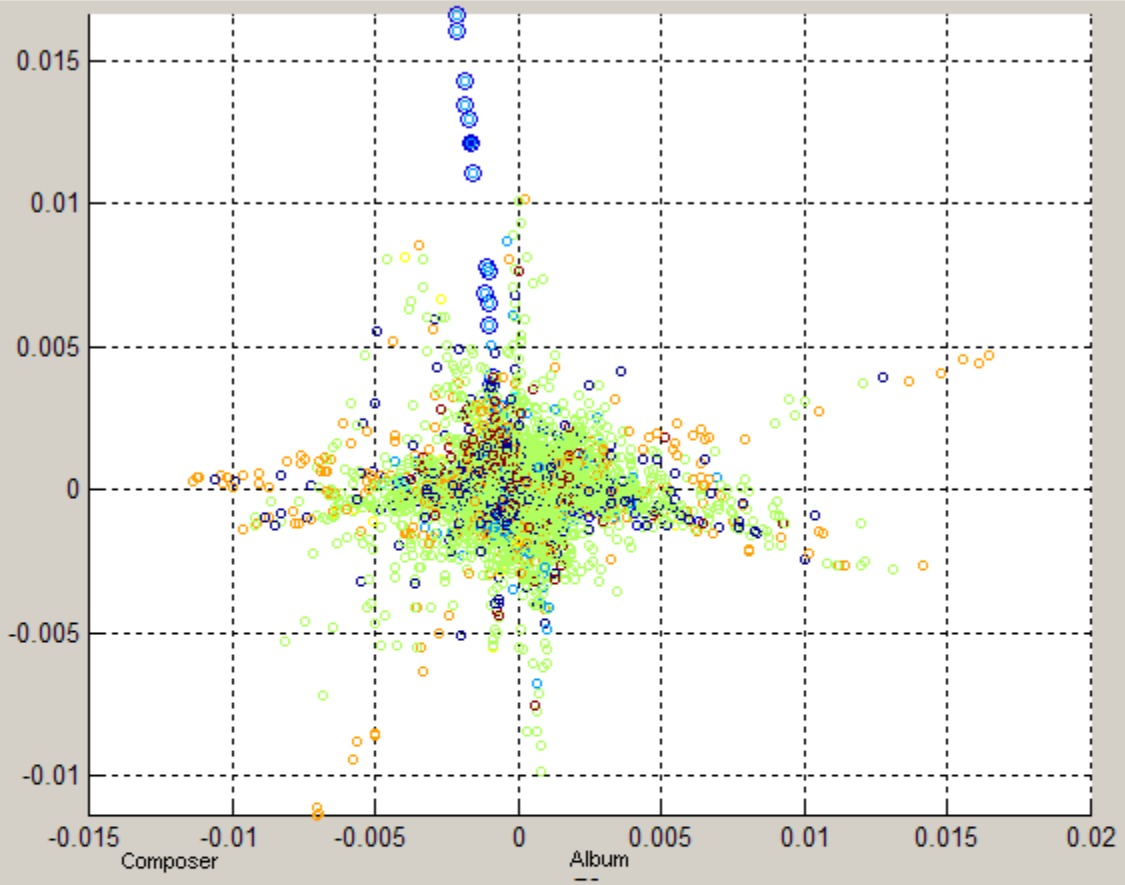
Given f defined on the data x_i let $P_1(f)$ be the coarse scale approximation.

We then expand the residual f - $P_1(f)$ using $e^{-4|x_i-x_j|^2}$, which is half as wide, to build $P_2(f - P_1(f))$ which extends "to distance $\approx 1/2$". If $P_1(f) + P_2(f - P_1(f))$ is a good approximation we stop, otherwise we repeat at the next scale.

YALE


Normal Density
surf
original
extended
1.5
1
0.5
0
-1.5
-1
-0.5
0
0.5
1
1.5
1
0.5
0
-0.5
-1
-1.5

GUI_music_02



Display options

Scale by eigenvalues

Axis equal

3D

Rotate

Time for diffusion map

Choose axes

25

42

35

Show spectrum

Search Options

Dimensions for distance

Top 100

Number of neighbors for search

trich Fischer-Dieskau	Schubert Winterreise	13 - Dietrich Fischer-Dieskau - Die Post
trich Fischer-Dieskau	Schubert Schwanengesang und 7 Lieder	14 - Fischer-Dieskau, Dietrich - Schwanengesang D. 957 - 1
trich Fischer-Dieskau	Schubert Winterreise	20 - Dietrich Fischer-Dieskau - Der Wegweiser
trich Fischer-Dieskau	Schubert Schwanengesang und 7 Lieder	15 - Fischer-Dieskau, Dietrich - An die Musik D. 547
trich Fischer-Dieskau	Schubert Winterreise	10 - Dietrich Fischer-Dieskau - Rast
trich Fischer-Dieskau	Schubert Winterreise	01 - Dietrich Fischer-Dieskau - Gute Nacht
trich Fischer-Dieskau	Schubert Schwanengesang und 7 Lieder	01 - Fischer-Dieskau, Dietrich - Schwanengesang D. 957 - 1
trich Fischer-Dieskau	Schubert Schwanengesang und 7 Lieder	15 - Fischer-Dieskau, Dietrich - An die Musik D. 547
trich Fischer-Dieskau	Schubert Winterreise	24 - Dietrich Fischer-Dieskau - Der Leiermann
trich Fischer-Dieskau	Schubert Schwanengesang und 7 Lieder	05 - Fischer-Dieskau, Dietrich - Schwanengesang D. 957 - 5
trich Fischer-Dieskau	Schubert Schwanengesang und 7 Lieder	07 - Fischer-Dieskau, Dietrich - Schwanengesang D. 957 - 7
ank Sinatra		

> PLAY

[] STOP

Make Track New Center

Save Song List

Save MP3s

Search in the whole database

->

->

No track selected

No track selected

Empty

Play K mean center

Playback options

Use WM activeX

Many of the applications of Laplacean eigenvectors to geometrize data and dimensional reduction are pervasive through the machine learning community, as well as the spectral graph theory community. Our point here is, that tuning the Laplace operator to the task enables global integration of local models (differential and integral calculus).

A simple synthetic description of many of these ideas and more on diffusion geometries is given in the July 2006 issue of Applied and Computational Harmonic Analysis.

and

R Coifman , A. Singer, Non-linear independent component analysis with diffusion maps, *Appl. Comput. Harmon, Anal.* 25 (2008) 226-239.

A. Singer, “A Remark on Global Positioning from Local Distances”,Proceedings of the National Academy of Sciences, 105 (28):9507-9511 (2008).