

Geometric diffusions for the analysis of data from sensor networks

Ronald R Coifman¹, Mauro Maggioni¹, Steven W Zucker¹ and Ioannis G Kevrekidis²

Harmonic analysis on manifolds and graphs has recently led to mathematical developments in the field of data analysis. The resulting new tools can be used to compress and analyze large and complex data sets, such as those derived from sensor networks or neuronal activity datasets, obtained in the laboratory or through computer modeling. The nature of the algorithms (based on diffusion maps and connectivity strengths on graphs) possesses a certain analogy with neural information processing, and has the potential to provide inspiration for modeling and understanding biological organization in perception and memory formation.

Addresses

¹Program of Applied Mathematics, Department of Mathematics, Yale University, 10 Hillhouse Avenue, New Haven, CT 06520, USA

²Department of Chemical Engineering, A-217 Engineering Quadrangle, Princeton University, Princeton, NJ 08544, USA

Corresponding author: Coifman, Ronald R (coifman@math.yale.edu)

Current Opinion in Neurobiology 2005, 15:XXX–XXX

This review comes from a themed issue on
New technologies
Edited by Gero Miesenböck and Richard GM Morris

0959-4388/\$ – see front matter
© 2005 Elsevier Ltd. All rights reserved.

Introduction

Data processing and analysis has always been a vital component of scientific research; increasingly so in our times [1–4,5], when highly resolved sensing in space and time gives rise to huge, high-dimensional datasets. The same holds when the data are the result of fine-grained computational modeling, rather than sensor output. In neuroscience, there are myriad sources of very high dimensional data. Perhaps the simplest example is a single spike train, or a sequence of 100 to 10,000 such trains [6]. The situation becomes much more interesting (and much more complicated) when one considers evaluating the information in electrode arrays in, for example, the retina [7], the hippocampus [8,9] or the motor cortex [10]. Apart from these foundational questions, ‘untangling the distributed code’ (e.g. [11,12]) is now a key question for developing man–machine interfaces [10,11], and is not unlike related questions for the analysis of EEG and MEG signals. The techniques described here should be relevant to many of these tasks, both for developing processing algorithms and for determining the level of structure and intrinsic information in the signals. The additional feature of extracting higher order concepts from data computationally resonates with the way such concepts are extracted from data physiologically. We comment on some such tentative ‘cognitive processing’ features of our data processing algorithms.

The mathematical theory underpinning these new data analysis algorithms is that of harmonic analysis on sets of data represented as points lying in n -dimensional Euclidean space, R^n , and on graphs constructed using this data. These graphs, connecting data points in a way to be described below, are in a way reminiscent of the interconnectivity graphs of sensor nodes (or neurons) in which the strength of the connections represents a high affinity between nodes. The main challenge involving the analysis of such complex structures lies in the ability to explain the transition from local ‘affinities’ of massive sensor outputs, or data, to some higher order concepts, regions of influence and connectivities on a macroscopic scale. The mathematical theory described here leads to various computational methodologies useful in data analysis and machine learning and, as such, provides a powerful tool for empirical modeling.

One goal of this review is to present these developments in data analysis; a second goal is to provide some insight into mathematical processing mechanisms. These might be useful to the scientist studying empirical data processing and biological information processing in the formulation of potential models of neuronal organization (or sensor fusion) at different levels of granularity. Our approach gives rise to Markov processes on graphs constructed using the data; and uses spectral theory and eigenfunctions of these Markov processes [1,2], leading to a natural geometric organization of complex data sets, providing a ‘nonlinear’ principal component analysis. We remark in passing that the top eigenfunction, corresponding to the highest eigenvalue, for the Web graph provides the ‘importance ranking’ used by ‘Google’® for webpage ranking, whereas the subsequent eigenfunctions provide a more detailed mapping. More importantly, we show how these eigenfunctions, viewed as a mathematical and computational tool, can be replaced by ‘aggregates of nodes’, equipped with a notion of multiscale affinity which can, in principle, be implemented biologically through various linking systems. This provides a potential theoretical mechanism for simple emergent organization and learning that might have biological relevance. Although related ideas appear in a variety of contexts of data analyses, such as spectral graph theory [13], manifold learning and nonlinear dimensionality reduction [14–17], we augment them by showing that the diffusion distances are key intrinsic geometric quantities linking spectral theory of Markov processes to the corresponding geometry of the data, relating localization in spectrum to localization in data space [2]. Existing dimensionality reduction techniques typically focus either on global or on local features of the data; our methodology integrates features at all scales in a coherent multiscale structure.

Geometric diffusions for global structure definition of data

In applied mathematics we often view ensembles of data as graphs with a large number of vertices, with each vertex being a data point (e.g. a visual stimulus), and edges connecting very similar data points (in an application-specific sense). For example, two visual stimuli could be considered similar if they excite a visual receptor in a very similar way.

Discovering large-scale structures and extracting information from such graphs is, in general, a very challenging task. Often the data are high-dimensional, that is, represented by long strings of numbers (vectors); however, physical or other constraints force the set of points or their probability densities to be intrinsically lower-dimensional, so they can, in principle, be described by a small number of degrees of freedom [1,2,14–17,18,19]. Our goal is to organize and process the data so as to reveal the low-dimensional structure. We use diffusion semigroups to generate various multiscale inference (or affinity) geometries (ontologies).

We show that appropriately selected eigenfunctions of Markov matrices describing local transitions, or affinities in the system, lead to coarse-grained, macroscopic structures at different scales.

In particular, the leading eigenfunctions enable a low dimensional geometric embedding of the dataset into a lower-dimensional Euclidean space, so that the ordinary Euclidean distance in the embedding space measures intrinsic diffusion (inference, affinity or relevance) metrics of the data.

The Euclidean correlation in R^n , for large n is, in general, not a good measure of affinity, except possibly for very close-by data points. This is the reason for the introduction of the ‘closeness’ parameter ϵ in the formula below. The premise is that the Euclidean distance provides a meaningful measure of ‘affinity’ for data lying closer than a cutoff distance quantified by this ϵ ; and is meaningless for data beyond this cutoff. One of the main contributions is to find an embedding space such that the Euclidean distance in this space is truly representative of the closeness (‘affinity’) among the data.

Mathematical background

Think of a point X_i in Euclidean space as representing a string of outputs from a neuron labeled by i (data vector, sensor output stream, and so on). A matrix of local affinities can be constructed as:

$$A_\varepsilon = [X_i \cdot X_j]_\varepsilon := \exp\left\{-\frac{(1 - X_i \cdot X_j)}{\varepsilon}\right\}$$

$$\|X_i\| = 1$$

The strength of such a data-correlation based affinity decays rapidly with the distance of outputs (other data affinities are possible, including chemical). We renormalize this matrix to a Markov matrix A (or more precisely A_ε), with sums of the entries of each row equaling one. A measures local similarities, and corresponds to one step of a random walk on the data [1**,2**,20]; its powers A^t correspond to propagation of the local similarities by the Markov process after t steps (time) of the random walk. This random walk on the data gives rise to a geometric diffusion (analogous to the derivation of the diffusion equation from Brownian motion). For large t , all similarities are integrated along all paths, yielding information about global structures in the data. Remarkably, these can be efficiently computed: let $\varphi_l(i) = \varphi_l(X_i)$ be the l^{th} eigenvector of A evaluated at data point i , satisfying $A\varphi_l(i) = \lambda_l^2\varphi_l(i)$ (λ_l^2 are arranged in decreasing order). Then

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

$$= a_t(i, j) \equiv a_t(X_i, X_j).$$

We consider the map

$$X_i^{(t)} \rightarrow (\lambda_1^t \varphi_1(X_i), \lambda_2^t \varphi_2(X_i), \dots, \lambda_m^t \varphi_m(X_i)) = \hat{X}_i^{(t)}$$

called the ‘diffusion map’, embedding into R^m at time t . The square of the ‘diffusion distance’ at time t , measuring ‘divergence’ between nodes i and j , is:

$$d_t^2(i, j) = a_t(i, i) + a_t(j, j) - 2a_t(i, j) = \sum_1^m \lambda_l^{2t} (\varphi_l(i) - \varphi_l(j))^2 = \|\hat{X}_i^{(t)} - \hat{X}_j^{(t)}\|^2.$$

For large t this can be computed very accurately using only the corresponding first few eigenfunctions, because only a few of the terms λ_l^{2t} are above the level of precision of interest (Figure 1). This provides a diffusion map embedding of output data into a new low-dimensional Euclidean space, converting diffusion distance on the data points into Euclidean distance in the embedding space.

As a first simple example of data reorganization provided by the diffusion embedding, we consider a sampled geometric hourglass surface, idealizing a set of data points with two weakly connected clusters, see Figure 2. We embed the point cloud into three-dimensional Euclidean space so that the diffusion distance in the original space can be computed as the ordinary Euclidean length of the chord connecting them in embedding space. Because the diffusion is slower through the bottleneck, the two components are farther apart in the diffusion metric.

In Figure 3, we illustrate the organizational ability of the diffusion maps on a collection of images given in random order. The inputs are 2-D gray scale pictures of the object in ‘3D’ in various positions, each viewed as a $32 \times 32 = 1024$ dimensional vector. To calculate the embedding, one constructs the Markov matrix as above, and computes the first few eigenfunctions. The top two eigenfunctions reveal the orientation of ‘3D’, and organize the data accordingly, see Figure 3.

Next, we organize a heterogeneous material, consisting of two component materials (nodes, represented by circles and crosses), possessing different conductivities (Figure 4). Although the gross statistics of circles and crosses are identical on both lobes, the left lobe happens to have more highly conductive links, which reduces the diffusion distance between its constituent nodes. The left-to-right bottleneck increases the

diffusion distance between the two lobes, because there are fewer paths connecting the left and right lobe. The actual long-time affinity structure is described in terms of the eigenfunctions (Figure 4): on the left all points are tightly linked, whereas on the right they maintain some distance. The map has accounted for the preponderance of connections through all paths of all lengths between the nodes.

The next example (Figure 5) represents an organization of the configuration space of lip images that arise from a single speaker. No structure is assumed. The local similarity between images, viewed as high-dimensional vectors, organizes them as above in the first three diffusion coordinates. Different locations in the diffusion plot correspond to different clusters of strongly related lip images.

Dynamic learning through diffusion geometry

We now use these ideas to describe various learning methodologies in which the diffusion mechanism is iteratively adjusted to improve accuracy.

First, we generalize the basic affinity matrix to enable purely empirical and dynamical modeling and learning.

Assume that a data point set (sensor output, individual neuron output strings, and so on) has been generated by a process, the local statistical characteristics of which vary from location to location. For each point x , we view the neighboring data points as generated by a local unknown diffusion process, the probability density of which is estimated by $p_x(y) = c_x \exp(-q_x(x-y))$, where q_x is a quadratic form obtained empirically (for example by local principal component analysis [21]) from the data in a small ‘neighborhood’ of x .

We use the matrix $\sum_y p_x(y)p_z(y) = a(x,z)$ to model the corresponding data-driven diffusion. The

distance defined by this kernel is $d(x,z) = (\sum_y |p_x(y) - p_z(y)|^2)^{1/2}$, which can be viewed as the natural distance on the ‘statistical tangent space’ to the point cloud.

In a dynamical learning situation we can start with a data point x , use its Euclidean neighborhood to define $p_x(y)$ at x , then find the z ’s that can be reached from x to compute locally $a(x,z)$. We then propagate a density in a neighborhood of x via powers of A , stopping when the propagation by diffusion slows down.

When labels are available, separating (a subset of) the data in different classes, the information they provide can be incorporated in p_x , by locally warping the metric so that the diffusion starting in one class stays in that class without leaking to others. This could be obtained, for example, by using any kind of local discriminant analysis [21] to build a local metric, the ‘fast’ directions of which are parallel to the boundary between classes and the ‘slow’ directions of which are transverse to the class boundaries. We also suggest that an iterative, partially supervised procedure can lead to good results in many practical situations.

In Figure 6 we represent a diffusion from labeled samples, from three different types of tissue, seeking to identify all related samples in the image. Here, each pixel has an absorption spectrum, with 128 spectral dimensions. The middle image shows the failure of conventional ‘nearest neighbor’ classification, whereas the diffusion distance yields a better classification.

Multiscale analysis of diffusion and spectral analysis

Our goal is to replace the analytic construction of the eigenfunctions by direct combinatorial link organizations. We show that the emergent organization discovered above with the help of the eigenfunctions can be translated into a multiscale hierarchical geometry of data points. This point of view can be used as a guide for theoretical processing models in biological systems.

The first few eigenfunctions of the matrix A (or equivalently, of the Laplacian on a graph [13]) detect and organize global structures on the data-based graph [1,16]. It is often the case, in biological and other complex systems, that several organizational structures exist at different ‘scales’. Sensor outputs can be grouped (compressed) into ensembles at different scales of complexity, to perform tasks at different levels of complexity or abstraction, and integrating the tasks performed at lower levels of complexity.

We sketch a technique for constructing these sets of structures at different scales on a set of outputs or data, starting from the finest granularity, and building up to more complex structures, all inter-related at each scale and across scales, culminating in the global structures detected and described by the analysis with eigenfunctions described above. In the case of clouds of data points, this translates into a multiscale analysis of the cloud of points; at each scale we have a set of aggregates of points, and relationships among these groups are determined by a power of the diffusion operator at that scale. We claim (see [2]) that the embedding provided by the eigenfunctions can also be achieved by a hierarchical regrouping of data, using affinity at different diffusion time scales as a grouping mechanism.

The construction alluded to above is most easily explained in terms of conventional semantic analysis of text documents, each document being a data point. Each document has coordinates that represent the frequency of occurrence of words in it. We correlate only documents with strong similarity of vocabulary. Given a document x , we can build a folder around it of documents with strong immediate affinity (i.e. nearest neighbors). This becomes a folder at ‘scale 1’. To obtain a folder at ‘scale 2’ we consider all documents, y , that are nearest neighbors to a nearest neighbor of x (i.e. they are linked by a chain of length 2 to x), and measure affinity as the sum of strength of all these chains of length 2 linking y to x ; we keep only those, y , with strong affinity to form a folder at scale 2. We repeat this process for all chains of length 4 and less. One can easily build a directory structure of folders at all dyadic scales, with folders at a fixed scale being disjoint. From our point of view, every sensor (every neuron) can be viewed as a document for which a string of sensor outputs are the coordinates (elementary semantic content), whereas the folders are groups of outputs combining similar or highly related outputs at different resolution (or abstraction) levels. In Figure 7 the elementary documents are various 6x6 patches of the image in the first panel. The folders at different levels of resolution correspond to higher level features of the image.

To relate this description to a mathematical formulation we start by observing, as above (Figure 1), that the numerical rank of $(A_\varepsilon)^{t/\varepsilon}$ decreases rapidly as t increases. In particular, if we consider the expansions $a_t(x, y) = \sum \lambda_i^{2t/\varepsilon} \varphi_i(x) \varphi_i(y)$, for $t = \varepsilon 2^j$, obtained by successive squaring, then for any fixed precision the summation can be restricted to smaller and smaller sets of indices.

Secondly, the columns $a_t(x, y)$ of the matrix $(A_\varepsilon)^{t/\varepsilon}$ represent the probability of transition in t steps from x to y .

We can also interpret the x column of the matrix A^{2^j} , $a_{2^j}(x, y)$, as a rank of affinity between sensor (neuron) output x and sensor (neuron) output y at scale j , and the collection of points y , such that $a_{2^j}(x, y) > \delta$ could represent all sensor (neuron) outputs y similar to x .

We present a very simple method for obtaining a hierarchical ‘sensor folder’ (or ‘neuron group’) organization, as described above for the text documents. A minimal collection of clusters organizing the whole set of points at different levels of granularity is obtained as follows: let $\{x_k^{j+1}\}$ be a maximal subcollection of points in $\{x_k^j\}$ (key-points at scale j), such that $1/2 \leq d_{2^j}(x_k^{j+1}, x_i^{j+1})$, where $\{x_k^0\}$ are the original points. Then any point is at distance at most $1/2$ at scale j from one of the

selected ‘key-points’ at that scale, enabling us to create a document folder labeled by the key-point. It is easy to modify this construction to obtain a tree of non-overlapping folders.

This construction, when applied to text documents (equipped with semantic coordinates), builds an automatic folder structure with corresponding key documents characterizing the folders.

A detailed, refined construction of scaling functions (columns of A^t) and wavelets representing this multiscale organization of the graph is provided in Coifman and Maggioni [2], and connections with related algorithms in numerical analysis in Brandt [22]. This analysis of aggregation at different times (and corresponding scales), enables us to perform multiscale wavelet analysis on manifolds and graphs in a natural way. Applications include compression of functions on the dataset, denoising of such functions, and learning (in the sense of classification and regression) of functions on the dataset. Although the description of the analysis given above refers only to organization of existing data, we point out that the tools developed also enable the incorporation of new data points into the structure in a consistent way, and the extension of functions modeled on the data to new sensor outputs [1'',2'',4''].

The multiscale construction enables structure to emerge at different scales as a function of connectivity. In Figure 7 we show several small patches from a simple image. If all patches are considered, edge filters (at the finer scales) and blob filters (at the coarser scales) naturally arise. Note the clear curvature in their structure [23]. Restricting the number of patches would result in more V1-like ‘receptive-fields’ [24–27].

Stochasticity and coherence

Global geometric diffusions can be applied to data driven by a Langevin equation [19''] that is used to model many biological systems [28–30], for example, stochastic unsynchronized neuronal pulse trains. The macroscopic probability density behavior of such systems is governed by the Fokker–Planck operator [19''], the eigenfunctions of which can be empirically approximated as described above, leading to efficient descriptions of likely, long-time probability configurations and geometries [2'',9,19]. The connections between Bayesian learning and Fokker-Planck equations date back to Verrelst [31] and references therein.

Diffusion wavelets and global diffusion have both been applied successfully to learning processes in a variety of (stochastic) environments, where an agent (e.g. robot) learns optimal behavior for achieving certain tasks from past experiences [18''].

Conclusions

Diffusion geometries can reveal structure in data at different levels of organization. Because many sources of data in neuroscience are high-dimensional, understanding their primary, low-dimensional intrinsic structure can be insightful. It has been indicated that image patch structure can suggest receptive field properties, and that different properties emerge at different levels. The intrinsic dimensionality can also be useful for efficient data analysis. Many applications of these techniques in neuroscience remain to be tried, from spike train analysis to olfaction and the electroencephalogram (EEG). But perhaps more exciting is the possibility that emergent structure across levels will open a theoretical door into cognitive neuroscience and memory organization.

Matlab scripts for the computations involved in diffusion maps and multiscale analysis of diffusion are available online [32] or upon request from M Maggioni.

Acknowledgements

The authors would like to thank Y Keller, S Lafon, AB Lee and B Nadler for providing some of the figures, and P Jones for useful comments during the preparation of the manuscript. Research partially supported by Defence Advanced Research Projects Agency. M Maggioni is partially supported by National Science Foundation grant DMS-0512050.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

••1. Coifman RR, Lafon S: **Diffusion maps**. *Appl Comp Harm Anal* 2005, in press.

The authors present an introduction to diffusion maps and their applications.

- 2. Coifman RR, Maggioni M: **Diffusion wavelets**. Tech Rep YALE/DCS/TR-1303. *Appl Comp Harm Anal* 2005, in press.

The authors provide an in-depth presentation of the multiscale construction of diffusion geometries for multiscale analysis on graphs and manifolds.

- 3. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner FJ, Zucker SW: **Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part I: Diffusion maps**. *Proc Nat Acad Sci* 2005, **102**: 7426-7431.

The authors provide an introduction to geometric diffusions with applications to analysis of data sets and simulated physical systems.

- 4. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner FJ, Zucker SW: **Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part II: Multiscale methods**. *Proc Natl Acad Sci USA* 2005, **102**:7432-7437.

The authors provide a short introduction to the construction of multiscale diffusion geometries, and techniques for out-of-sample extension of Laplacian eigenfunctions and diffusion wavelets.

5. Donoho D: **Data! Data! Data! Challenges and opportunities of the coming data deluge**. *Michelson Memorial Lecture Series* 2001, available online at: <http://www-stat.stanford.edu/~donoho/Lectures/AMS2000/AMS2000.html>
6. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W: *Spikes, exploring the neural code*. MIT Press, 1997.
7. Warland D, Reinagel P, Meister M: **Decoding visual information from a population of retinal ganglion cells**. *J Neurophys* 1997, **78**: 2336-2350.
8. Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ: **Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells**. *J Neurophysiol* 1998, **79**:1017-1044.
9. Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA: **A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells**. *J Neurosci* 1998, **18**:7411-7425.
10. Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK, Kim J, Biggs SJ, Srinivasan MA, Nicolelis MA: **Real-time prediction of hand trajectory by ensembles of cortical neurons in primates**. *Nature* 2000, **408**:361–365.
11. Donoghue J, Nurmikko A, Friehs G, Black M: **Development of neural motor prostheses for humans**. *Advances in Clinical Neurophysiology (Supplements to Clinical Neurophysiology, Vol. 57)* Editors: Hallett M, Phillips LH, Schomer DL, Massey JM. 2004.
12. Georgopoulos AP, Kettner RE, Schwartz AB: **Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of**

- the direction of movement by a neuronal population.** *J Neurosci* 1988, **8**:2928-2937.
13. Chung F: *Spectral Graph Theory*, (92). American Mathematical Society: CMBS-AMS series in Mathematics; 1997.
 14. Ham J, Lee DD, Mika S: **Schölkopf: a kernel view of the dimensionality reduction of manifolds.** In *Proceedings of the XXI Conference on Machine Learning, Banff, Canada, 2004*.
- The authors provide an overview of several nonlinear dimensionality reduction and manifold learning techniques, under the umbrella of kernel methods.
15. Roweis ST, Saul LK: **Nonlinear dimensionality reduction by locally linear embedding.** *Science* 2000, **290**:2323-2326.
 16. Belkin M, Niyogi P: **Laplacian eigenmaps for dimensionality reduction and data representation.** *Neural Comp* 2003, **15**: 1373-1396.
 17. Tenenbaum JB, de Silva V, Langford JC: **A global geometric framework for nonlinear dimensionality reduction.** *Science* 2000, **290**:2319-232.
 - 18 Mahadevan S, Maggioni M: **Value function approximation with diffusion wavelets and Laplacian eigenfunctions.** *Proc NIPS* 2005, in press.
- Introduction to the application of diffusion wavelets and Laplacian eigenfunctions to Markov decision processes and learning.
- 19. Coifman RR, Lafon S, Kevrekidis Y, Nadler B: **Diffusion maps, spectral clustering and reaction coordinates of dynamical systems.** *Appl Comp Harm Anal* 2005, in press.
- The authors supply an interpretation of diffusion geometries for data generated by simulations of certain classes of physical systems with several examples.
20. Szummer M, Jaakkola T: **Partially labeled classification with Markov random walks.** *Advances in Neuronal Information Processing Systems* 2001, **14**: 945-952.
 21. Hastie T, Tibshirani R, Friedman JH: *The Elements of Statistical Learning*. Springer-Verlag; 2001.
 22. Brandt A: **Algebraic multigrid theory: the symmetric case.** *Appl Math Comp* 1986, **19**: 23-56.
 23. Dobbins A, Zucker SW, Cynader M: **Endstopped neurons in the visual cortex as a substrate for calculating curvature.** *Nature* 1987, **329**:438-441.
 24. Bell AJ, Sejnowski TJ: **The independent components of natural images are edge filters.** *Vision Research* 1997, **57**: 3327-3338.
 25. van Hateren J, van der Schaaf A: **Independent component filters of natural images compared with simple cells in primary visual cortex.** *Proc R Soc London B* 1997, **265**: 259-366.
 26. Olhausen BA, Field DJ: **Sparse coding with an overcomplete basis set: a strategy employed by V1?** *Vision Research* 1997, **37**:3311-3325.

27. Caywood MS, Willmore, B, Tolhurst DJ: **Independent components of color natural scenes resemble V1 neurons in their spatial and color tuning.** *J Neurophysiol* 2004, **91**:2859-2873.
28. Rao CV, Wolf DM, Arkin AP: **Control, exploitation and tolerance of intracellular noise.** *Nature* 2002, **420**:231-237.
29. Vogels TP, Rajan K, Abbot LF: **Neural network dynamics.** *Annu Rev Neurosci* 2005, **28**:357-376.
30. Miesenböck G, Kevrekidis IG: **Optical imaging and control of genetically designated neurons in functioning circuits.** *Annu Rev Neurosci* 2005, **28**:533-563.
31. Verrelst H, Suykens J, Vandewalle J, De Moor B, **Bayesian learning and the Fokker-Planck machine.** In *Proceedings of the International Workshop on Advanced Black-box Techniques for Nonlinear Modeling, Leuven, Belgium, 1998.* 55-61.
32. Maggioni M: Homepage, URL: www.math.yale.edu/~mmm82

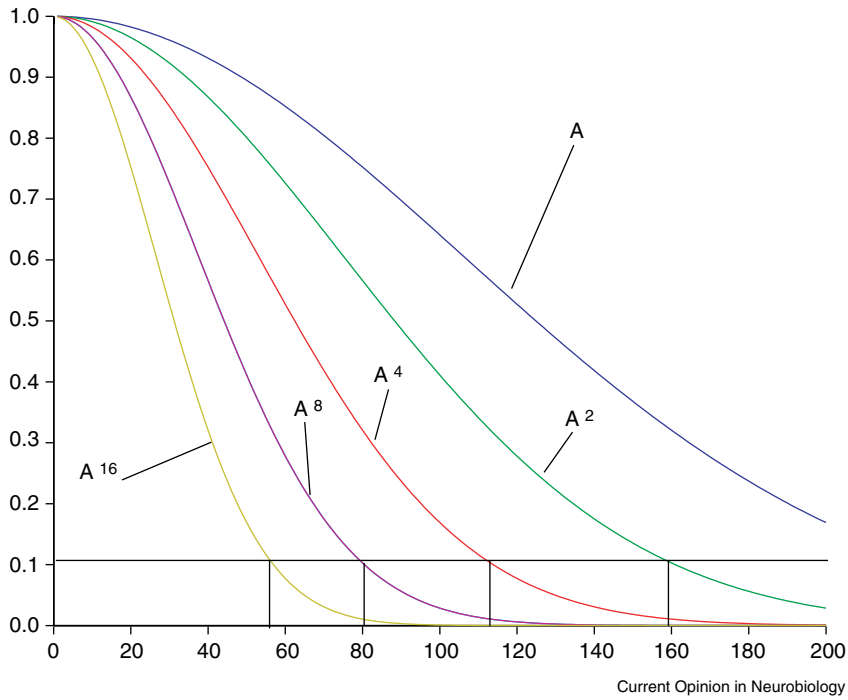


Figure 1
 The spectra of powers of A . Some examples of the spectra of the dyadic powers of A . The x axis is the index of the eigenvalue, and the y axis the eigenvalue itself. Eigenvalues are positive and are arranged in nonincreasing order.

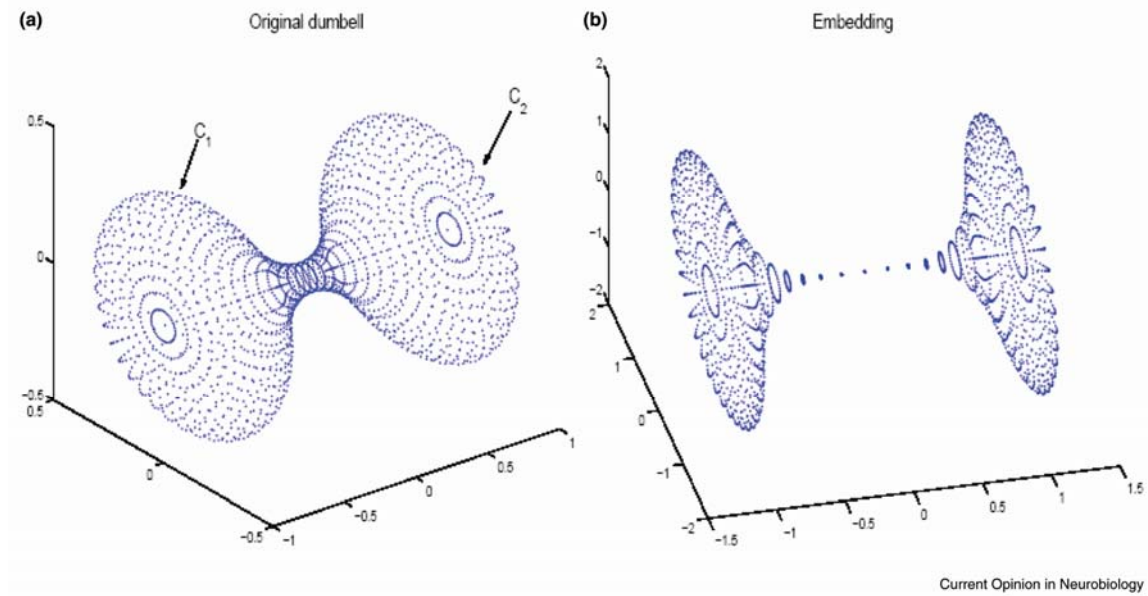


Figure 2
 Diffusion embedding of a sampled hourglass manifold. **(a)** An original set of points sampled on a hourglass manifold, as a model for two weakly-connected clusters C_1 and C_2 , and **(b)** their embedding using the eigenfunctions of the diffusion matrix A . The Euclidean distance in image in **(b)** is equivalent to large-time t diffusion distance on the original set of points in **(a)**. The two 'clusters' get flattened and move further apart in the new space. The axes just provide a reference frame.

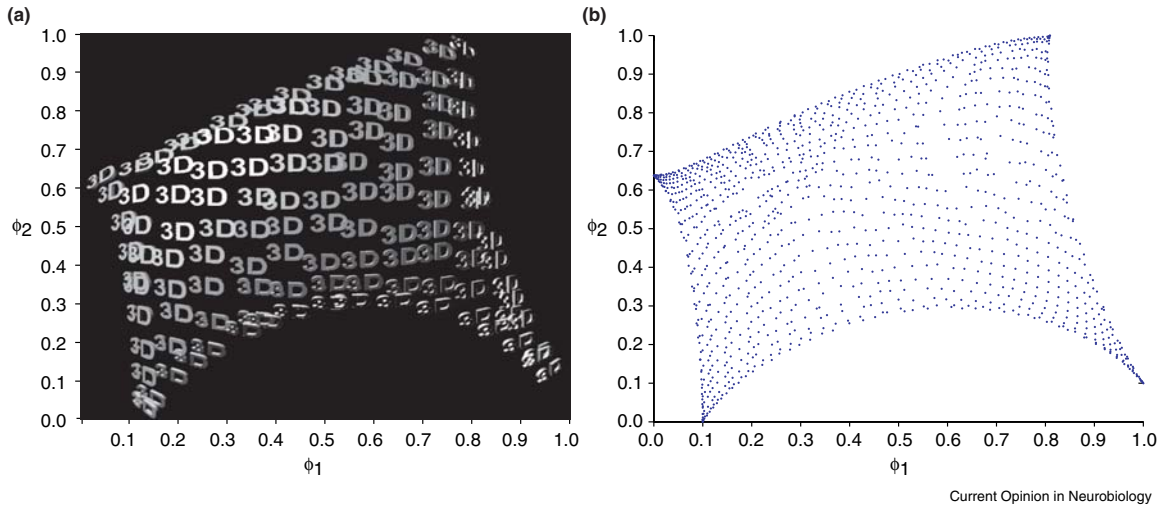


Figure 3

Diffusion embedding of a set of pictures of “3D”. Organization emerging from a collection of images given in random order (data = $\{x_i\}$). (a) The images are displayed according to their location in the two-dimensional diffusion embedding

$(\phi_1(x_i), \phi_2(x_i))$, displayed in (b). The coordinates capture (perceive) the orientation of the picture in 3D.

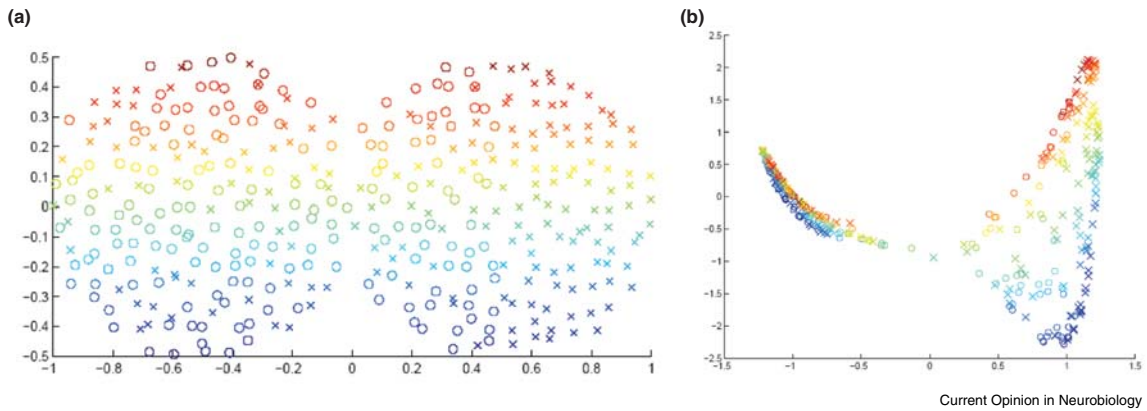
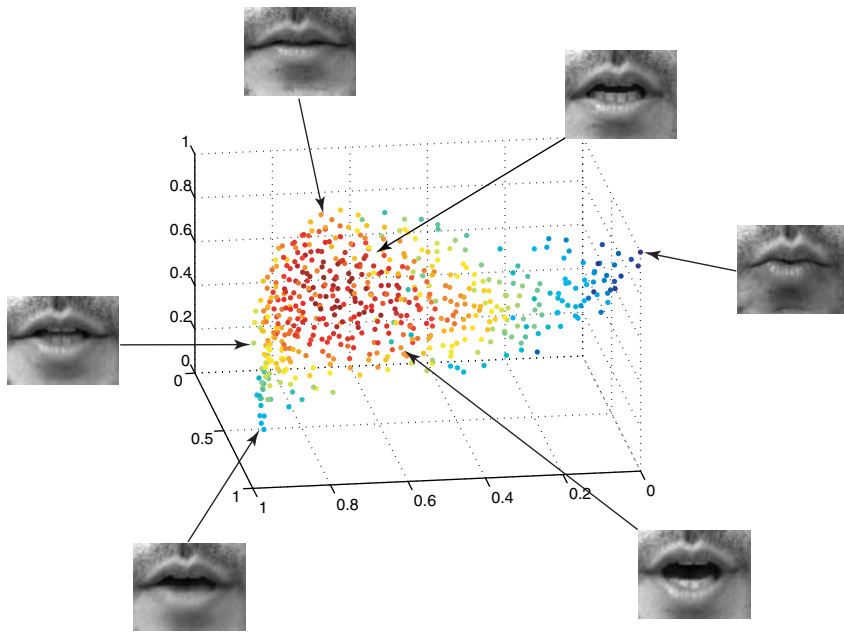


Figure 4

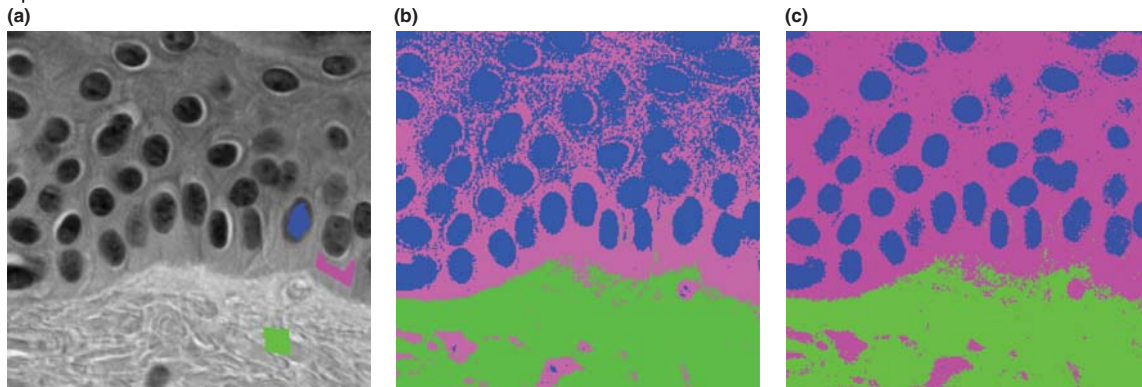
Diffusion embedding of a heterogeneous material. (a) A heterogeneous material and (b) its long-term diffusion embedding $(\phi_2(x_i), \phi_3(x_i))$. This structure could be interpreted as a map of trees (circles) and shrubs (crosses), with the links representing the probability of fire propagating among them. From (b) it is clear that the risk of fire propagating from top to bottom is higher on the left side of the forest. Color is included so that points can be matched across the two pictures.



Current Opinion in Neurobiology

Figure 5

Diffusion embedding of images of lips. The lip alphabet is learnt from a set of pictures of the lips of a speaker. The manifold structure and its parameters are parametrized by the three top eigenfunctions (axes in the figure) of the diffusion, and this parametrization can be used to lip-read. An interpretation of the low order eigenfunctions is openness of the mouth and exposure of teeth.



Current Opinion in Neurobiology

Figure 6

Classification of tissue types in a hyperspectral image through diffusion. **(a)** A slice of a hyperspectral image with three selected regions that correspond to three different biologically significant types of tissue: nuclei (blue), cytoplasm of epidermal cells (pink) and collagen in the underlying dermis (green). **(b)** Predictions of tissue type by a standard nearest neighbor classifier, trained on the set in (a). **(c)** Predictions made by the diffusion classifier described above, with the training set represented in (a).

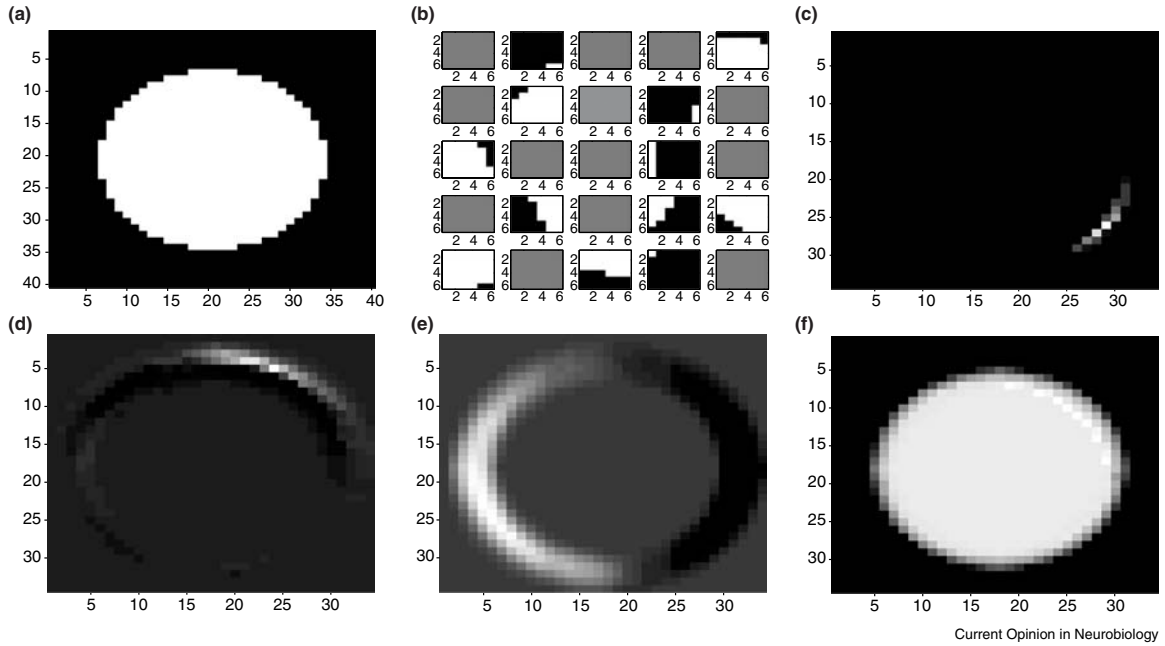


Figure 7
 Multiscale folders. **(a)** Original picture. **(b)** A subset of 6×6 pixel patches extracted from the image. **(c)** A folder at scale 2 is a weighted aggregate of patches, representing a higher level feature. **(d)** Another folder at scale 2 is an edge detector. **(e and f)** Two folders at scale 3 that represent weighted aggregates of patches ('attributes' or 'features') at an even coarser scale.