23; right 36, 13, and 27); superior frontal gyrus (left −9, 31, and 45; right 17, 35, and 37).

17. Although the improvement in WM performance with cholinergic enhancement was a nonsignificant trend in the current study ($P = 0.07$), in a previous study (*9*) with a larger sample ($n = 13$) the effect was highly significant ($P < 0.001$). In the current study, we analyzed RT data for six of our seven subjects because the behavioral data for one subject were unavailable due to a computer failure. The difference in the significance of the two findings is simply a result of the difference in sample sizes. A power analysis shows that the size of the RT difference and variability in the current sample would yield a significant result ($P = 0.01$) with a sample size of 13. During the memory trials, mean RT was 1180 ms during placebo and 1119 ms during physostigmine. During the control trials, mean RT was 735 ms during placebo and 709 ms during physostigmine, a difference that did not approach significance ($P = 0.24$), suggesting that the effect of cholinergic enhancement on WM performance is not due to a nonspecific increase in arousal.
18. Matched-pair *t* tests (two-tailed) were used to test the significance of drug-related changes in the volume of regions of interest that showed significant response contrasts.
19. H. Sato, Y. Hata, H. Masui, T. Tsumoto, *J. Neurophysiol.* **55**, 765 (1987).
20. M. E. Hasselmo, *Behav. Brain Res.* **67**, 1 (1995).
21. M. G. Baxter, A. A. Chiba, *Curr. Opin. Neurobiol.* **9**, 178 (1999).
22. B. J. Everitt, T. W. Robbins, *Annu. Rev. Psychol.* **48**, 649 (1997).
23. R. Desimone, J. Duncan, *Annu. Rev. Neurosci.* **18**, 193 (1995).
24. P. C. Murphy, A. M. Sillito, *Neuroscience* **40**, 13 (1991).
25. M. Corbetta, F. M. Miezin, S. Dobmeyer, G. L. Shulman, S. E. Peterson, *J. Neurosci.* **11**, 2383 (1991).
26. J. V. Haxby *et al.*, *J. Neurosci.* **14**, 6336 (1994).
27. A. Rosier, L. Cornette, G. A. Orban, *Neuropsychobiology* **37**, 98 (1998).
28. M. E. Hasselmo, B. P. Wyble, G. V. Wallenstein, *Hippocampus* **6**, 693 (1996).
29. S. P. Mewaldt, M. M. Ghoneim, *Pharmacol. Biochem. Behav.* **10**, 1205 (1979).
30. M. Petrides, *Philos. Trans. R. Soc. London Ser. B* **351**, 1455 (1996).
31. M. E. Hasselmo, E. Fransen, C. Dickson, A. A. Alonso, *Ann. N.Y. Acad. Sci.* **911**, 418 (2000).
32. M. M. Mesulam, *Prog. Brain Res.* **109**, 285 (1996).
33. R. T. Bartus, R. L. Dean III, B. Beer, A. S. Lippa, *Science* **217**, 408 (1985).
34. N. Qizilbash *et al.*, *JAMA* **280**, 1777 (1998).
35. J. V. Haxby, J. Ma. Maisog, S. M. Courtney, in *Mapping and Modeling the Human Brain*, P. Fox, J. Lancaster, K. Friston, Eds. (Wiley, New York, in press).
36. We express our appreciation to S. Courtney, R. Desimone, Y. Jiang, S. Kastner, L. Latour, A. Martin, L. Pessoa, and L. Ungerleider for careful and critical review of the manuscript. We also thank M. B. Schapiro and S. I. Rapoport for input during early stages of this project. This research was supported by the National Institute on Mental Health and National Institute on Aging Intramural Research Programs.

7 August 2000; accepted 15 November 2000

A Global Geometric Framework for Nonlinear Dimensionality Reduction

Joshua B. Tenenbaum,[1]* Vin de Silva,[2] John C. Langford[3]

Scientists working with large volumes of high-dimensional data, such as global climate patterns, stellar spectra, or human gene distributions, regularly confront the problem of dimensionality reduction: finding meaningful low-dimensional structures hidden in their high-dimensional observations. The human brain confronts the same problem in everyday perception, extracting from its high-dimensional sensory inputs—30,000 auditory nerve fibers or 10^6 optic nerve fibers—a manageably small number of perceptually relevant features. Here we describe an approach to solving dimensionality reduction problems that uses easily measured local metric information to learn the underlying global geometry of a data set. Unlike classical techniques such as principal component analysis (PCA) and multidimensional scaling (MDS), our approach is capable of discovering the nonlinear degrees of freedom that underlie complex natural observations, such as human handwriting or images of a face under different viewing conditions. In contrast to previous algorithms for nonlinear dimensionality reduction, ours efficiently computes a globally optimal solution, and, for an important class of data manifolds, is guaranteed to converge asymptotically to the true structure.

A canonical problem in dimensionality reduction from the domain of visual perception is illustrated in Fig. 1A. The input consists of many images of a person's face observed under different pose and lighting conditions, in no particular order. These images can be thought of as points in a high-dimensional vector space, with each input dimension corresponding to the brightness of one pixel in the image or the firing rate of one retinal ganglion cell. Although the input dimensionality may be quite high (e.g., 4096 for these 64 pixel by 64 pixel images), the perceptually meaningful structure of these images has many fewer independent degrees of freedom. Within the 4096-dimensional input space, all of the images lie on an intrinsically three-dimensional manifold, or constraint surface, that can be parameterized by two pose variables plus an azimuthal lighting angle. Our goal is to discover, given only the unordered high-dimensional inputs, low-dimensional representations such as Fig. 1A with coordinates that capture the intrinsic degrees of freedom of a data set. This problem is of central importance not only in studies of vision (*1*–*5*), but also in speech (*6*, *7*), motor control (*8*, *9*), and a range of other physical and biological sciences (*10*–*12*).

[1]Department of Psychology and [2]Department of Mathematics, Stanford University, Stanford, CA 94305, USA. [3]Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15217, USA.

*To whom correspondence should be addressed. E-mail: jbt@psych.stanford.edu

The classical techniques for dimensionality reduction, PCA and MDS, are simple to implement, efficiently computable, and guaranteed to discover the true structure of data lying on or near a linear subspace of the high-dimensional input space (*13*). PCA finds a low-dimensional embedding of the data points that best preserves their variance as measured in the high-dimensional input space. Classical MDS finds an embedding that preserves the interpoint distances, equivalent to PCA when those distances are Euclidean. However, many data sets contain essential nonlinear structures that are invisible to PCA and MDS (*4*, *5*, *11*, *14*). For example, both methods fail to detect the true degrees of freedom of the face data set (Fig. 1A), or even its intrinsic three-dimensionality (Fig. 2A).

Here we describe an approach that combines the major algorithmic features of PCA and MDS—computational efficiency, global optimality, and asymptotic convergence guarantees—with the flexibility to learn a broad class of nonlinear manifolds. Figure 3A illustrates the challenge of nonlinearity with data lying on a two-dimensional "Swiss roll": points far apart on the underlying manifold, as measured by their geodesic, or shortest path, distances, may appear deceptively close in the high-dimensional input space, as measured by their straight-line Euclidean distance. Only the geodesic distances reflect the true low-dimensional geometry of the manifold, but PCA and MDS effectively see just the Euclidean structure; thus, they fail to detect the intrinsic two-dimensionality (Fig. 2B).

Our approach builds on classical MDS but seeks to preserve the intrinsic geometry of the data, as captured in the geodesic manifold distances between all pairs of data points. The crux is estimating the geodesic distance between faraway points, given only input-space distances. For neighboring points, input-space distance provides a good approxima-

tion to geodesic distance. For faraway points, geodesic distance can be approximated by adding up a sequence of "short hops" between neighboring points. These approximations are computed efficiently by finding shortest paths in a graph with edges connecting neighboring data points.

The complete isometric feature mapping, or Isomap, algorithm has three steps, which are detailed in Table 1. The first step determines which points are neighbors on the manifold M, based on the distances $d_X(i,j)$ between pairs of points i,j in the input space X. Two simple methods are to connect each point to all points within some fixed radius ϵ, or to all of its K nearest neighbors (*15*). These neighborhood relations are represented as a weighted graph G over the data points, with edges of weight $d_X(i,j)$ between neighboring points (Fig. 3B).

In its second step, Isomap estimates the geodesic distances $d_M(i,j)$ between all pairs of points on the manifold M by computing their shortest path distances $d_G(i,j)$ in the graph G. One simple algorithm (*16*) for finding shortest paths is given in Table 1.

The final step applies classical MDS to the matrix of graph distances $D_G = \{d_G(i,j)\}$, constructing an embedding of the data in a d-dimensional Euclidean space Y that best preserves the manifold's estimated intrinsic geometry (Fig. 3C). The coordinate vectors $\mathbf{y}_i$ for points in Y are chosen to minimize the cost function

$$E = \|\tau(D_G) - \tau(D_Y)\|_{L^2} \qquad (1)$$

where D_Y denotes the matrix of Euclidean distances $\{d_Y(i,j) = \|\mathbf{y}_i - \mathbf{y}_j\|\}$ and $\|A\|_{L^2}$ the L^2 matrix norm $\sqrt{\sum_{i,j} A_{ij}^2}$. The τ operator

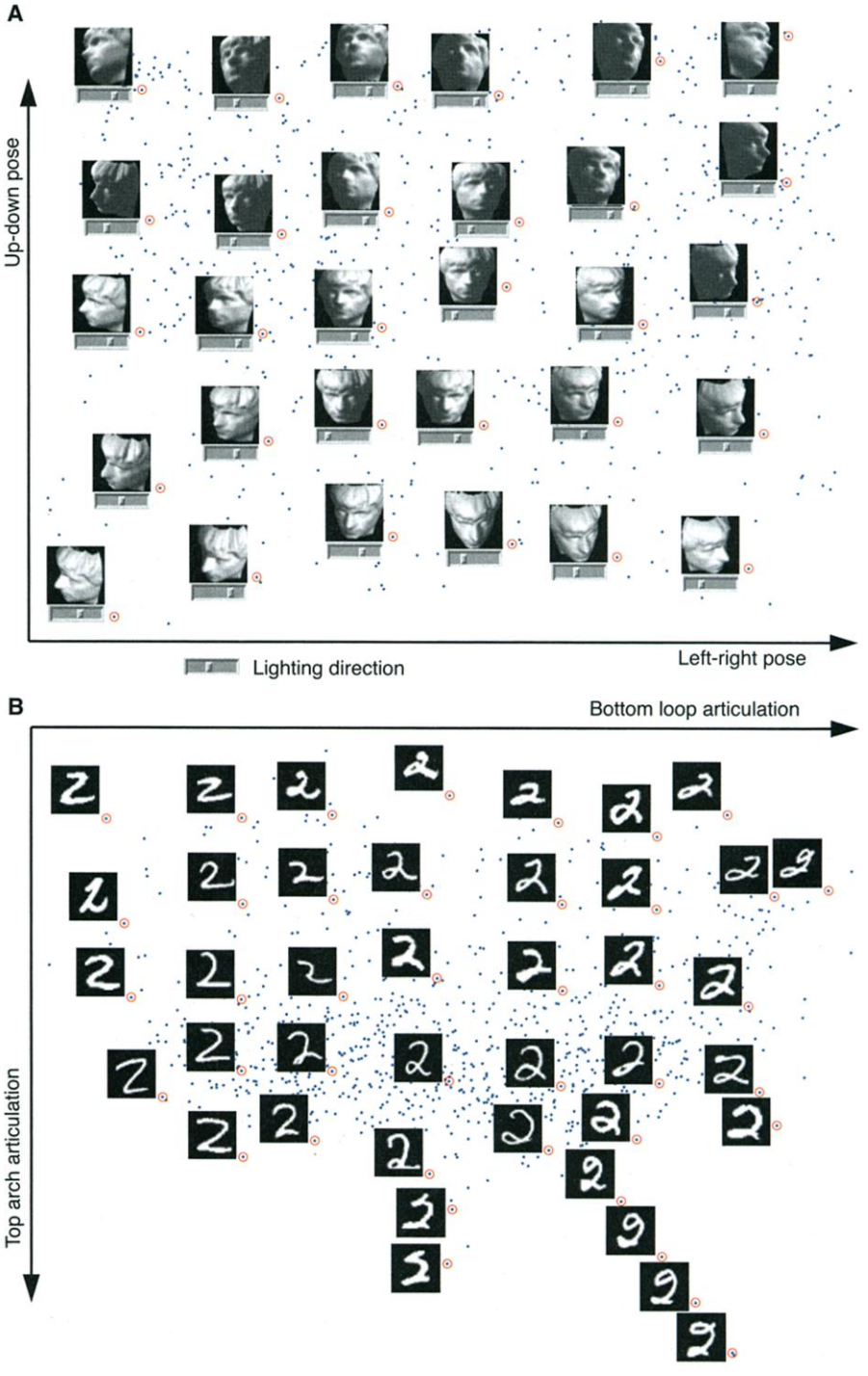

Fig. 1. (**A**) A canonical dimensionality reduction problem from visual perception. The input consists of a sequence of 4096-dimensional vectors, representing the brightness values of 64 pixel by 64 pixel images of a face rendered with different poses and lighting directions. Applied to $N = 698$ raw images, Isomap ($K = 6$) learns a three-dimensional embedding of the data's intrinsic geometric structure. A two-dimensional projection is shown, with a sample of the original input images (red circles) superimposed on all the data points (blue) and horizontal sliders (under the images) representing the third dimension. Each coordinate axis of the embedding correlates highly with one degree of freedom underlying the original data: left-right pose (x axis, $R = 0.99$), up-down pose (y axis, $R = 0.90$), and lighting direction (slider position, $R = 0.92$). The input-space distances $d_X(i,j)$ given to Isomap were Euclidean distances between the 4096-dimensional image vectors. (**B**) Isomap applied to $N = 1000$ handwritten "2"s from the MNIST database (*40*). The two most significant dimensions in the Isomap embedding, shown here, articulate the major features of the "2": bottom loop (x axis) and top arch (y axis). Input-space distances $d_X(i,j)$ were measured by tangent distance, a metric designed to capture the invariances relevant in handwriting recognition (*41*). Here we used ϵ-Isomap (with $\epsilon = 4.2$) because we did not expect a constant dimensionality to hold over the whole data set; consistent with this, Isomap finds several tendrils projecting from the higher dimensional mass of data and representing successive exaggerations of an extra stroke or ornament in the digit.

converts distances to inner products (*17*), which uniquely characterize the geometry of the data in a form that supports efficient optimization. The global minimum of Eq. 1 is achieved by setting the coordinates $\mathbf{y}_i$ to the top d eigenvectors of the matrix $\tau(D_G)$ (*13*).

As with PCA or MDS, the true dimensionality of the data can be estimated from the decrease in error as the dimensionality of Y is increased. For the Swiss roll, where classical methods fail, the residual variance of Isomap correctly bottoms out at $d = 2$ (Fig. 2B).

Just as PCA and MDS are guaranteed, given sufficient data, to recover the true structure of linear manifolds, Isomap is guaranteed asymptotically to recover the true dimensionality and geometric structure of a strictly larger class of nonlinear manifolds. Like the Swiss roll, these are manifolds whose intrinsic geometry is that of a convex region of Euclidean space, but whose ambient geometry in the high-dimensional input space may be highly folded, twisted, or curved. For non-Euclidean manifolds, such as a hemisphere or the surface of a doughnut, Isomap still produces a globally optimal low-dimensional Euclidean representation, as measured by Eq. 1.

These guarantees of asymptotic convergence rest on a proof that as the number of data points increases, the graph distances $d_G(i,j)$ provide increasingly better approximations to the intrinsic geodesic distances $d_M(i,j)$, becoming arbitrarily accurate in the limit of infinite data (*18*, *19*). How quickly $d_G(i,j)$ converges to $d_M(i,j)$ depends on certain parameters of the manifold as it lies within the high-dimensional space (radius of curvature and branch separation) and on the density of points. To the extent that a data set presents extreme values of these parameters or deviates from a uniform density, asymptotic convergence still holds in general, but the sample size required to estimate geodesic distance accurately may be impractically large.

Isomap's global coordinates provide a simple way to analyze and manipulate high-dimensional observations in terms of their intrinsic nonlinear degrees of freedom. For a set of synthetic face images, known to have three degrees of freedom, Isomap correctly detects the dimensionality (Fig. 2A) and separates out the true underlying factors (Fig. 1A). The algorithm also recovers the known low-dimensional structure of a set of noisy real images, generated by a human hand varying in finger extension and wrist rotation (Fig. 2C) (*20*). Given a more complex data set of handwritten digits, which does not have a clear manifold geometry, Isomap still finds globally meaningful coordinates (Fig. 1B) and nonlinear structure that PCA or MDS do not detect (Fig. 2D). For all three data sets, the natural appearance of linear interpolations between distant points in the low-dimensional coordinate space confirms that Isomap has captured the data's perceptually relevant structure (Fig. 4).

Previous attempts to extend PCA and MDS to nonlinear data sets fall into two broad classes, each of which suffers from limitations overcome by our approach. Local linear techniques (*21*–*23*) are not designed to represent the global structure of a data set within a single coordinate system, as we do in Fig. 1. Nonlinear techniques based on greedy optimization procedures (*24*–*30*) attempt to discover global structure, but lack the crucial algorithmic features that Isomap inherits from PCA and MDS: a noniterative, polynomial time procedure with a guarantee of global optimality; for intrinsically Euclidean man-

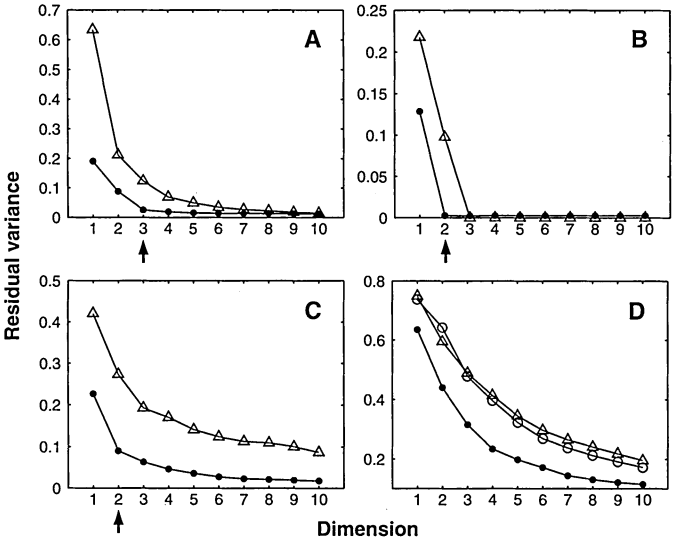

Fig. 2. The residual variance of PCA (open triangles), MDS [open triangles in (A) through (C); open circles in (D)], and Isomap (filled circles) on four data sets (*42*). (**A**) Face images varying in pose and illumination (Fig. 1A). (**B**) Swiss roll data (Fig. 3). (**C**) Hand images varying in finger extension and wrist rotation (*20*). (**D**) Handwritten "2"s (Fig. 1B). In all cases, residual variance decreases as the dimensionality d is increased. The intrinsic dimensionality of the data can be estimated by looking for the "elbow" at which this curve ceases to decrease significantly with added dimensions. Arrows mark the true or approximate dimensionality, when known. Note the tendency of PCA and MDS to overestimate the dimensionality, in contrast to Isomap.

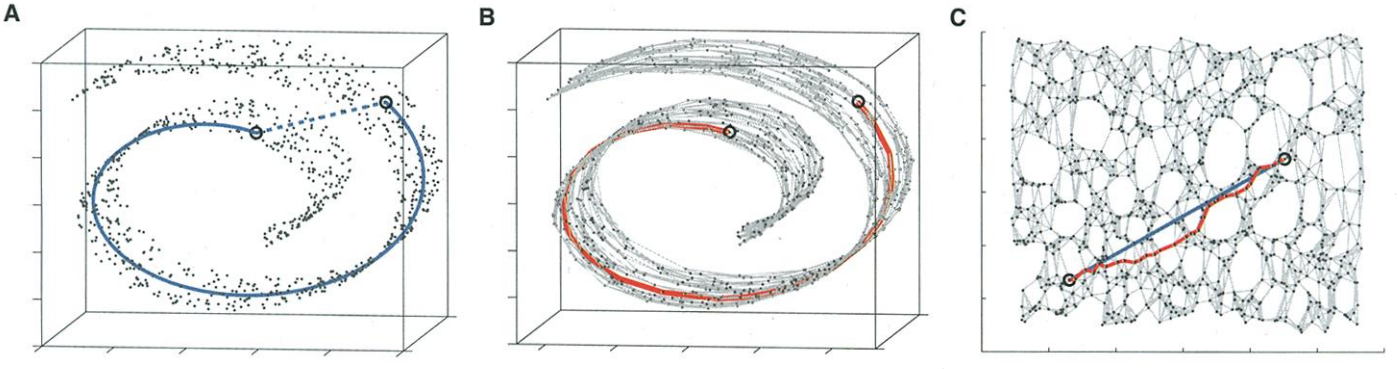

Fig. 3. The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (**A**) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect their intrinsic similarity, as measured by geodesic distance along the low-dimensional manifold (length of solid curve). (**B**) The neighborhood graph G constructed in step one of Isomap (with $K = 7$ and $N = 1000$ data points) allows an approximation (red segments) to the true geodesic path to be computed efficiently in step two, as the shortest path in G. (**C**) The two-dimensional embedding recovered by Isomap in step three, which best preserves the shortest path distances in the neighborhood graph (overlaid). Straight lines in the embedding (blue) now represent simpler and cleaner approximations to the true geodesic paths than do the corresponding graph paths (red).

Table 1. The Isomap algorithm takes as input the distances $d_X(i,j)$ between all pairs i,j from N data points in the high-dimensional input space X, measured either in the standard Euclidean metric (as in Fig. 1A) or in some domain-specific metric (as in Fig. 1B). The algorithm outputs coordinate vectors $\mathbf{y}_i$ in a d-dimensional Euclidean space Y that (according to Eq. 1) best represent the intrinsic geometry of the data. The only free parameter (ϵ or K) appears in Step 1.

Step		
1	Construct neighborhood graph	Define the graph G over all data points by connecting points i and j if [as measured by $d_X(i,j)$] they are closer than ϵ (ϵ-Isomap), or if i is one of the K nearest neighbors of j (K-Isomap). Set edge lengths equal to $d_X(i,j)$.
2	Compute shortest paths	Initialize $d_G(i,j) = d_X(i,j)$ if i,j are linked by an edge; $d_G(i,j) = \infty$ otherwise. Then for each value of $k = 1, 2, \ldots, N$ in turn, replace all entries $d_G(i,j)$ by $\min\{d_G(i,j), d_G(i,k) + d_G(k,j)\}$. The matrix of final values $D_G = \{d_G(i,j)\}$ will contain the shortest path distances between all pairs of points in G (*16*, *19*).
3	Construct d-dimensional embedding	Let λ_p be the p-th eigenvalue (in decreasing order) of the matrix $\tau(D_G)$ (*17*), and v_p^i be the i-th component of the p-th eigenvector. Then set the p-th component of the d-dimensional coordinate vector $\mathbf{y}_i$ equal to $\sqrt{\lambda_p} v_p^i$.

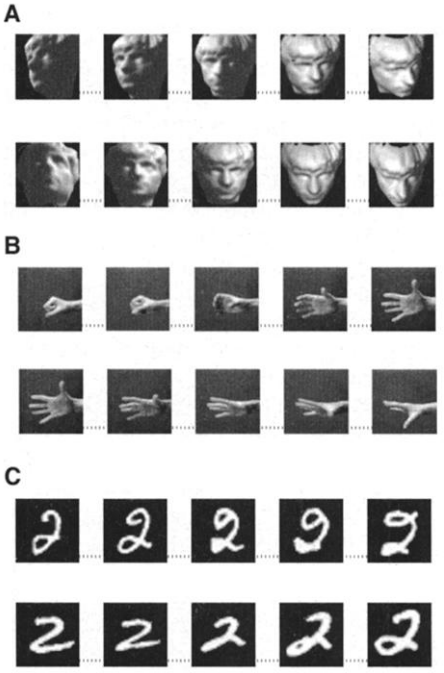

Fig. 4. Interpolations along straight lines in the Isomap coordinate space (analogous to the blue line in Fig. 3C) implement perceptually natural but highly nonlinear "morphs" of the corresponding high-dimensional observations (*43*) by transforming them approximately along geodesic paths (analogous to the solid curve in Fig. 3A). (**A**) Interpolations in a three-dimensional embedding of face images (Fig. 1A). (**B**) Interpolations in a four-dimensional embedding of hand images (*20*) appear as natural hand movements when viewed in quick succession, even though no such motions occurred in the observed data. (**C**) Interpolations in a six-dimensional embedding of handwritten "2"s (Fig. 1B) preserve continuity not only in the visual features of loop and arch articulation, but also in the implied pen trajectories, which are the true degrees of freedom underlying those appearances.

ifolds, a guarantee of asymptotic convergence to the true structure; and the ability to discover manifolds of arbitrary dimensionality, rather than requiring a fixed d initialized from the beginning or computational resources that increase exponentially in d.

Here we have demonstrated Isomap's performance on data sets chosen for their visually compelling structures, but the technique may be applied wherever nonlinear geometry complicates the use of PCA or MDS. Isomap complements, and may be combined with, linear extensions of PCA based on higher order statistics, such as independent component analysis (*31*, *32*). It may also lead to a better understanding of how the brain comes to represent the dynamic appearance of objects, where psychophysical studies of apparent motion (*33*, *34*) suggest a central role for geodesic transformations on nonlinear manifolds (*35*) much like those studied here.

References and Notes

1. M. P. Young, S. Yamane, *Science* **256**, 1327 (1992).
2. R. N. Shepard, *Science* **210**, 390 (1980).
3. M. Turk, A. Pentland, *J. Cogn. Neurosci.* **3**, 71 (1991).
4. H. Murase, S. K. Nayar, *Int. J. Comp. Vision* **14**, 5 (1995).
5. J. W. McClurkin, L. M. Optican, B. J. Richmond, T. J. Gawne, *Science* **253**, 675 (1991).
6. J. L. Elman, D. Zipser, *J. Acoust. Soc. Am.* **83**, 1615 (1988).
7. W. Klein, R. Plomp, L. C. W. Pols, *J. Acoust. Soc. Am.* **48**, 999 (1970).
8. E. Bizzi, F. A. Mussa-Ivaldi, S. Giszter, *Science* **253**, 287 (1991).
9. T. D. Sanger, *Adv. Neural Info. Proc. Syst.* **7**, 1023 (1995).
10. J. W. Hurrell, *Science* **269**, 676 (1995).
11. C. A. L. Bailer-Jones, M. Irwin, T. von Hippel, *Mon. Not. R. Astron. Soc.* **298**, 361 (1997).
12. P. Menozzi, A. Piazza, L. Cavalli-Sforza, *Science* **201**, 786 (1978).
13. K. V. Mardia, J. T. Kent, J. M. Bibby, *Multivariate Analysis*, (Academic Press, London, 1979).
14. A. H. Monahan, *J. Clim.*, in press.
15. The scale-invariant K parameter is typically easier to set than ϵ, but may yield misleading results when the local dimensionality varies across the data set. When available, additional constraints such as the temporal ordering of observations may also help to determine neighbors. In earlier work (*36*) we explored a more complex method (*37*), which required an order of magnitude more data and did not support the theoretical performance guarantees we provide here for ϵ- and K-Isomap.
16. This procedure, known as Floyd's algorithm, requires $O(N^3)$ operations. More efficient algorithms exploiting the sparse structure of the neighborhood graph can be found in (*38*).
17. The operator τ is defined by $\tau(D) = -HSH/2$, where S is the matrix of squared distances $\{S_{ij} = D_{ij}^2\}$, and H is the "centering matrix" $\{H_{ij} = \delta_{ij} - 1/N\}$ (*13*).
18. Our proof works by showing that for a sufficiently high density (α) of data points, we can always choose a neighborhood size (ϵ or K) large enough that the graph will (with high probability) have a path not much longer than the true geodesic, but small enough to prevent edges that "short circuit" the true geometry of the manifold. More precisely, given arbitrarily small values of λ_1, λ_2, and μ, we can guarantee that with probability at least $1 - \mu$, estimates of the form

$$(1 - \lambda_1)d_M(i,j) \le d_G(i,j) \le (1 + \lambda_2)d_M(i,j)$$

will hold uniformly over all pairs of data points i,j. For ϵ-Isomap, we require

$$\epsilon \le (2/\pi)r_0\sqrt{24\lambda_1}, \qquad \epsilon < s_0,$$

$$\alpha > [\log(V/\mu\eta_d(\lambda_2\epsilon/16)^d)]/\eta_d(\lambda_2\epsilon/8)^d$$

where r_0 is the minimal radius of curvature of the manifold M as embedded in the input space X, s_0 is the minimal branch separation of M in X, V is the (d-dimensional) volume of M, and (ignoring boundary effects) η_d is the volume of the unit ball in Euclidean d-space. For K-Isomap, we let ϵ be as above and fix the ratio $(K + 1)/\alpha = \eta_d(\epsilon/2)^d/2$. We then require

$$e^{-(K+1)/4} \le \mu\eta_d(\epsilon/4)^d/4V,$$

$$(e/4)^{(K+1)/2} \le \mu\eta_d(\epsilon/8)^d/16V,$$

$$\alpha > [4\log(8V/\mu\eta_d(\lambda_2\epsilon/32\pi)^d)]/\eta_d(\lambda_2\epsilon/16\pi)^d$$

The exact content of these conditions—but not their general form—depends on the particular technical assumptions we adopt. For details and extensions to nonuniform densities, intrinsic curvature, and boundary effects, see http://isomap.stanford.edu.
19. In practice, for finite data sets, $d_G(i,j)$ may fail to approximate $d_M(i,j)$ for a small fraction of points that are disconnected from the giant component of the neighborhood graph G. These outliers are easily detected as having infinite graph distances from the majority of other points and can be deleted from further analysis.
20. The Isomap embedding of the hand images is available at *Science* Online at www.sciencemag.org/cgi/content/full/290/5500/2319/DC1. For additional material and computer code, see http://isomap.stanford.edu.
21. R. Basri, D. Roth, D. Jacobs, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1998), pp. 414–420.
22. C. Bregler, S. M. Omohundro, *Adv. Neural Info. Proc. Syst.* **7**, 973 (1995).
23. G. E. Hinton, M. Revow, P. Dayan, *Adv. Neural Info. Proc. Syst.* **7**, 1015 (1995).
24. R. Durbin, D. Willshaw, *Nature* **326**, 689 (1987).
25. T. Kohonen, *Self-Organisation and Associative Memory* (Springer-Verlag, Berlin, ed. 2, 1988), pp. 119–157.
26. T. Hastie, W. Stuetzle, *J. Am. Stat. Assoc.* **84**, 502 (1989).
27. M. A. Kramer, *AIChE J.* **37**, 233 (1991).
28. D. DeMers, G. Cottrell, *Adv. Neural Info. Proc. Syst.* **5**, 580 (1993).
29. R. Hecht-Nielsen, *Science* **269**, 1860 (1995).
30. C. M. Bishop, M. Svensén, C. K. I. Williams, *Neural Comp.* **10**, 215 (1998).
31. P. Comon, *Signal Proc.* **36**, 287 (1994).
32. A. J. Bell, T. J. Sejnowski, *Neural Comp.* **7**, 1129 (1995).
33. R. N. Shepard, S. A. Judd, *Science* **191**, 952 (1976).
34. M. Shiffrar, J. J. Freyd, *Psychol. Science* **1**, 257 (1990).

35. R. N. Shepard, *Psychon. Bull. Rev.* **1**, 2 (1994).
36. J. B. Tenenbaum, *Adv. Neural Info. Proc. Syst.* **10**, 682 (1998).
37. T. Martinetz, K. Schulten, *Neural Netw.* **7**, 507 (1994).
38. V. Kumar, A. Grama, A. Gupta, G. Karypis, *Introduction to Parallel Computing: Design and Analysis of Algorithms* (Benjamin/Cummings, Redwood City, CA, 1994), pp. 257–297.
39. D. Beymer, T. Poggio, *Science* **272**, 1905 (1996).
40. Available at www.research.att.com/~yann/ocr/mnist.
41. P. Y. Simard, Y. LeCun, J. Denker, *Adv. Neural Info. Proc. Syst.* **5**, 50 (1993).
42. In order to evaluate the fits of PCA, MDS, and Isomap on comparable grounds, we use the residual variance $1 - R^2(\hat{D}_M, D_Y)$. D_Y is the matrix of Euclidean distances in the low-dimensional embedding recovered by each algorithm. $\hat{D}_M$ is each algorithm's best estimate of the intrinsic manifold distances: for Isomap, this is the graph distance matrix D_G; for PCA and MDS, it is the Euclidean input-space distance matrix D_X (except with the handwritten "2"s, where MDS uses the tangent distance). R is the standard linear correlation coefficient, taken over all entries of $\hat{D}_M$ and D_Y.
43. In each sequence shown, the three intermediate images are those closest to the points 1/4, 1/2, and 3/4 of the way between the given endpoints. We can also synthesize an explicit mapping from input space X to the low-dimensional embedding Y, or vice versa, using the coordinates of corresponding points $\{\mathbf{x}_i, \mathbf{y}_i\}$ in both spaces provided by Isomap together with standard supervised learning techniques (*39*).
44. Supported by the Mitsubishi Electric Research Laboratories, the Schlumberger Foundation, the NSF (DBS-9021648), and the DARPA Human ID program. We thank Y. LeCun for making available the MNIST database and S. Roweis and L. Saul for sharing related unpublished work. For many helpful discussions, we thank G. Carlsson, H. Farid, W. Freeman, T. Griffiths, R. Lehrer, S. Mahajan, D. Reich, W. Richards, J. M. Tenenbaum, Y. Weiss, and especially M. Bernstein.

10 August 2000; accepted 21 November 2000

Nonlinear Dimensionality Reduction by Locally Linear Embedding

Sam T. Roweis[1] and Lawrence K. Saul[2]

Many areas of science depend on exploratory data analysis and visualization. The need to analyze large amounts of multivariate data raises the fundamental problem of dimensionality reduction: how to discover compact representations of high-dimensional data. Here, we introduce locally linear embedding (LLE), an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs. Unlike clustering methods for local dimensionality reduction, LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations do not involve local minima. By exploiting the local symmetries of linear reconstructions, LLE is able to learn the global structure of nonlinear manifolds, such as those generated by images of faces or documents of text.

How do we judge similarity? Our mental representations of the world are formed by processing large numbers of sensory inputs—including, for example, the pixel intensities of images, the power spectra of sounds, and the joint angles of articulated bodies. While complex stimuli of this form can be represented by points in a high-dimensional vector space, they typically have a much more compact description. Coherent structure in the world leads to strong correlations between inputs (such as between neighboring pixels in images), generating observations that lie on or close to a smooth low-dimensional manifold. To compare and classify such observations—in effect, to reason about the world—depends crucially on modeling the nonlinear geometry of these low-dimensional manifolds.

Scientists interested in exploratory analysis or visualization of multivariate data (*1*) face a similar problem in dimensionality reduction. The problem, as illustrated in Fig. 1, involves mapping high-dimensional inputs into a low-dimensional "description" space with as many coordinates as observed modes of variability. Previous approaches to this problem, based on multidimensional scaling (MDS) (*2*), have computed embeddings that attempt to preserve pairwise distances [or generalized disparities (*3*)] between data points; these distances are measured along straight lines or, in more sophisticated usages of MDS such as Isomap (*4*), along shortest paths confined to the manifold of observed inputs. Here, we take a different approach, called locally linear embedding (LLE), that eliminates the need to estimate pairwise distances between widely separated data points. Unlike previous methods, LLE recovers global nonlinear structure from locally linear fits.

The LLE algorithm, summarized in Fig. 2, is based on simple geometric intuitions. Suppose the data consist of N real-valued vectors $\vec{X}_i$, each of dimensionality D, sampled from some underlying manifold. Provided there is sufficient data (such that the manifold is well-sampled), we expect each data point and its neighbors to lie on or close to a locally linear patch of the manifold. We characterize the local geometry of these patches by linear coefficients that reconstruct each data point from its neighbors. Reconstruction errors are measured by the cost function

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2 \qquad (1)$$

which adds up the squared distances between all the data points and their reconstructions. The weights W_{ij} summarize the contribution of the jth data point to the ith reconstruction. To compute the weights W_{ij}, we minimize the cost

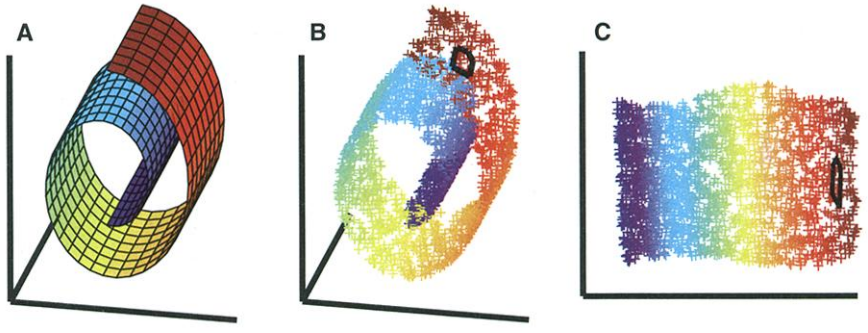

Fig. 1. The problem of nonlinear dimensionality reduction, as illustrated (*10*) for three-dimensional data (B) sampled from a two-dimensional manifold (A). An unsupervised learning algorithm must discover the global internal coordinates of the manifold without signals that explicitly indicate how the data should be embedded in two dimensions. The color coding illustrates the neighborhood-preserving mapping discovered by LLE; black outlines in (**B**) and (**C**) show the neighborhood of a single point. Unlike LLE, projections of the data by principal component analysis (PCA) (*28*) or classical MDS (*2*) map faraway data points to nearby points in the plane, failing to identify the underlying structure of the manifold. Note that mixture models for local dimensionality reduction (*29*), which cluster the data and perform PCA within each cluster, do not address the problem considered here: namely, how to map high-dimensional data into a single global coordinate system of lower dimensionality.

[1]Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR, UK. [2]AT&T Lab—Research, 180 Park Avenue, Florham Park, NJ 07932, USA.

E-mail: roweis@gatsby.ucl.ac.uk (S.T.R.); lsaul@research.att.com (L.K.S.)