The Isomap Algorithm and Topological Stability

Tenenbaum *et al*. (*1*) presented an algorithm, Isomap, for computing a quasi-isometric, low-dimensional embedding of a set of high-dimensional data points. Two issues need to be raised concerning this work. First, the basic approach presented by Tenenbaum *et al*. is not new, having been described in the context of flattening cortical surfaces using geodesic distances and multidimensional scaling (*2*, *3*) and in later work that used Dijkstra's algorithm to approximate geodesic distances (*4*, *5*). These ideas generalize to arbitrary dimensionality if the connectivity and metric information of the manifold are correctly supplied.

Second, and more important, this approach is topologically unstable and can only be used after careful preprocessing of the data (*6*). In the application domain of cortical flattening, it is necessary to check manually for connectivity errors, so that points nearby in 3-space (for example, on opposite banks of a cortical sulcus) are not taken to be nearby in the cortical surface. If such care is taken, this method represents the preferred method for quasi-isometric cortical flattening.

What is new about the Isomap algorithm is how it defines the connectivity of each data point via its nearest Euclidean neighbors in the high-dimensional space. This step is vulnerable to short-circuit errors if the neighborhood is too large with respect to folds in the manifold on which the data points lie or if noise in the data moves the points slightly off the manifold. Even a single short-circuit error can alter many entries in the geodesic distance matrix, which in turn can lead to a drastically different (and incorrect) low-dimensional embedding.

We illustrate this failure in Fig. 1, using the MATLAB code published by Tenenbaum *et al*. along with their "Swiss roll" data, to which we have added a small amount of noise (*7*). Clearly, the algorithm is topologically unstable: Small errors in the data connectivity (topology) can lead to large errors in the solution. Choosing a very small neighborhood is not a satisfactory solution, as this can fragment the manifold into a large number of disconnected regions. Choosing the neighborhood "just right" requires a priori information about the global geometry of the high-dimensional data manifold (*8*), but, presumably, it is exactly in the absence of such information that one would need to use an algorithm to find "meaningful low-dimensional structures hidden in high-dimensional observations" (*1*). In summary, the basic idea

of Isomap has long been known, and the new component introduced by Tenenbaum *et al*. provides an unreliable estimate of surface connectivity, which can lead to failure of the algorithm to perform as claimed.

Mukund Balasubramanian
Eric L. Schwartz
Department of Cognitive and Neural Systems and
Department of Electrical and Computer Systems
Boston University
Boston, MA 02215, USA

References and Notes
1. J. B. Tenenbaum, V. de Silva, J. C. Langford, *Science* **290**, 2319 (2000).
2. E. L. Schwartz, A. Shaw, E. Wolfson, *IEEE Trans. Pattern Anal. Machine Intell.* **11**, 1005 (1989).
3. E. Wolfson, E. L. Schwartz, *IEEE Trans. Pattern Anal. Machine Intell.* **11**, 1001 (1989).
4. B. Fischl, M. I. Sereno, A. M. Dale, *NeuroImage* **9**, 195 (1999).
5. For a surface represented as a polyhedral mesh, exact geodesics can be computed in exponential time (*2*). However, geodesics between all pairs of vertices on the mesh can be approximated in $O(N \log N)$ time using Dijkstra's algorithm to compute the shortest path along the edges of the polygons in the mesh (*4*). For meshes with $O(10^3)$ vertices, exact geodesics can be computed in about 10 min on a 500-MHz Pentium III PC; for much larger meshes, faster approximation schemes are desirable.
6. E. L. Schwartz, B. Merker, E. Wolfson, A. Shaw, *IEEE Comput. Graphics Appl.* **8**, 13 (1988).
7. We added zero-mean, normally distributed noise to the coordinates of each point, with the standard deviation of the noise chosen to be 2% of smallest dimension of the bounding box enclosing the data set (Fig. 1). Using smaller errors (for example, 1%) also led to failures, but we have not systematically studied the precise nature of algorithm failure, algorithm parameters, noise definition, and degree of failure of the solution.
8. Tenenbaum *et al*. provide proofs to show that if quantities such as the minimal radius of curvature of the manifold, the minimal branch separation, the volume of the manifold, and the density of the data points are known, then the neighborhood size can be bounded so that the algorithm performs correctly [note 18 of (*1*)]. In practice, however, these quantities will not be known in advance for a given set of high-dimensional data points. Furthermore, there is no prescription for how to compute these bounds in the presence of noise.

Response: Balasubramanian and Schwartz claim that "the basic idea" of our Isomap technique for nonlinear dimensionality reduction (*1*) has "long been known" in the context of flattening cortical surfaces. However, the problem of cortical flattening differs in crucial ways from our problem of finding low-dimensional structure in a cloud of high-dimensional data points. State-of-the-art procedures for cortical flattening (*2*, *3*) take as input a triangulated mesh of fixed topology and dimensionality, which represents additional structure not available in the general problem we attempted to solve. We take as input only a collection of unorganized data points; the topology and dimensionality of the underlying manifold are unknown and
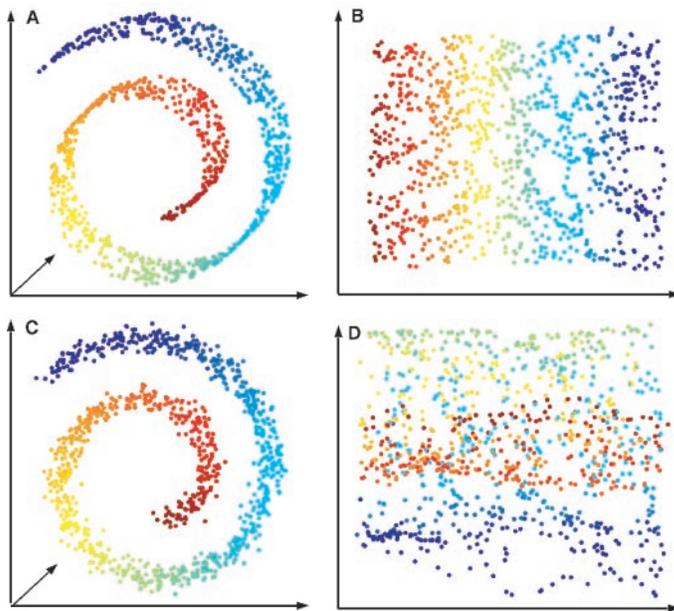
Fig. 1. (**A**) The "Swiss roll" data used by Tenenbaum *et al*. (*1*) to illustrate their algorithm (*n* = 1000). (**B**) The two-dimensional (2D) representation computed by the ε−Isomap variant of the Isomap algorithm, with ε = 5. Nearby points in the 2D embedding are also nearby points in the 3D manifold, as desired. (**C**) Data shown in A, with zero-mean normally distributed noise added to the coordinates of each point, where the standard deviation of the noise was chosen to be 2% of smallest dimension of the bounding box enclosing the data. (**D**) The Isomap (ε = 5) solution for the noisy data. There are gross "folds" in the embedding, and neither the metric nor the topological structure of the solution in (B) is preserved.

need to be estimated in the process of constructing a faithful low-dimensional embedding. Our algorithm provides a simple method for estimating the intrinsic geometry of a data manifold based on a rough estimate of each data point's neighbors on the manifold. The simplicity of our algorithm, in contrast to the specialized methods used in cortical flattening (2–4), makes it highly efficient and generally applicable to a broad range of data sources and dimensionalities, and makes possible our theoretical analyses of asymptotic convergence (1).

Balasubramanian and Schwartz also assert that Isomap is "topologically unstable," based on their failure to construct a topology-preserving 2D embedding of a Swiss roll data set that has been corrupted by additive Gaussian noise. As we dis-

cussed in our original report (1), the success of Isomap depends on being able to choose a neighborhood size (ε or K) that is neither so large that it introduces "short-circuit" edges into the neighborhood graph, nor so small that the graph becomes too sparse to approximate geodesic paths accurately. Short-circuit edges can lead to low-dimensional embeddings that do not preserve a manifold's true topology, as Balasubramanian and Schwartz illustrate with a choice of neighborhood size (ε = 5) that works well for our original noiseless data but that is too large given the level of noise (5) in their data. However, it is misleading to characterize our algorithm as "topologically unstable" without considering its stability over a range of neighborhood sizes. Indeed, this same level of noise

poses no difficulty for constructing a topology-preserving embedding if the neighborhood size is chosen just slightly smaller, with ε between 4.3 and 4.6 (Fig. 1D).

The suggestion by Balasubramanian and Schwartz that there is no way to choose an appropriate neighborhood size without "a priori information about the global geometry" of the data is incorrect. Such a priori information is useful in bounding the worst-case performance of our algorithm as a function of the number of data points (1), but it is not necessary to use the algorithm in practice. Fig. 1 illustrates one practical approach to selecting an appropriate neighborhood size, based on a trade-off between maximizing the number of points captured in the Euclidean embedding and minimizing the distortion of the geodesic distance estimates (6). This method successfully picks out appropriate neighborhood sizes for both the noiseless (Fig. 1E) and noisy (Fig. 1F) Swiss roll data sets. In both cases, the topology-preserving solution is stable, with a range of neighborhood sizes that achieve close to zero distortion of the geodesic distances while preserving the integrity of the underlying manifold.

For the Swiss roll with 1000 data points and Gaussian additive noise, the stability analysis illustrated in Fig. 1 can find neighborhood sizes that yield topology-preserving embeddings as long as the noise standard deviation is less than approximately 12% of the separation between branches of the manifold (7). Given a larger number of data points or a less curved manifold, the degree of noise tolerance could be substantially better. For instance, with 2000 data points on the Swiss roll, the noise standard deviation may be as high as 20% of the branch separation. For the face images from our original paper (1), topology-preserving embeddings are reliably found in the presence of Gaussian pixel noise with standard deviation as high 70% of each pixel's standard deviation in the noiseless data. More generally, short-circuit edges pose a threat to any attempt to infer the global geometry of a nonlinear data set from a bottom-up estimate of the data's topology; the locally linear embedding (LLE) algorithm (8) embodies a different approach to this same general goal and suffers from the same problem of short-circuit edges in the presence of noisy or sparse data (9). Improving the robustness of these dimensionality reduction algorithms in the presence of high noise levels—where short-circuit edges cannot be eliminated simply by shrinking the neighborhood size—is an important area for future research.

Fig. 1. (**A**) Our Swiss roll data set (1), shaded according to one intrinsic dimension of the underlying manifold. (**B**) The Swiss roll data plus Gaussian noise, with zero mean and standard deviation equal to approximately 7.5% of the separation between branches of the manifold (comparable to the noise level used by Balasubramanian and Schwartz). (**C**) An Isomap embedding of the noiseless data in (A) with neighborhood size ε = 5 preserves both the intrinsic geometry and the topological order of the underlying manifold. (**D**) An Isomap embedding of the noisy data in (B) with an appropriately chosen neighborhood size (ε = 4.6) also preserves both the intrinsic geometry and the topological order of the underlying manifold, just as in (C). (**E**) An appropriate neighborhood size for embedding the noiseless
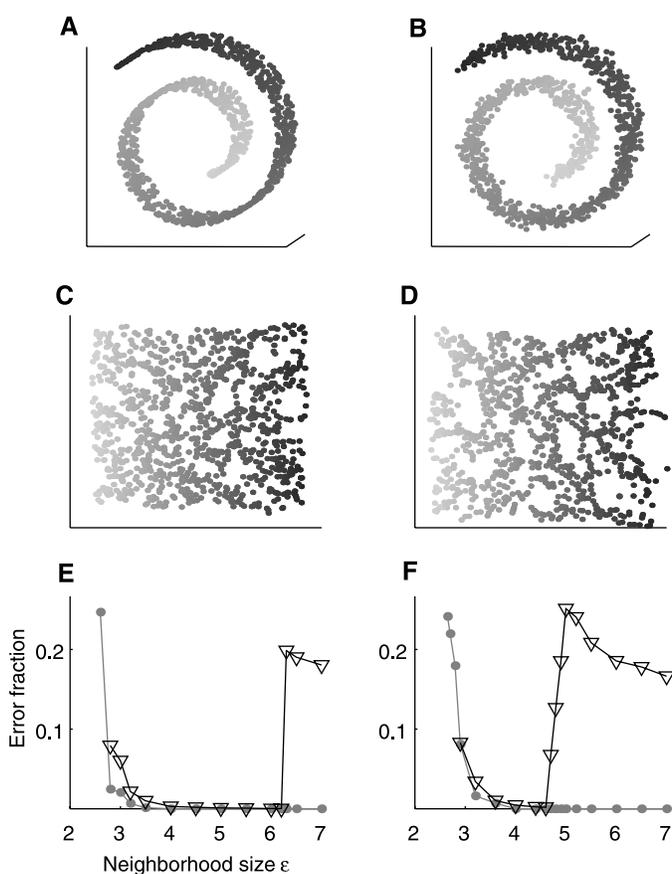
data into 2D Euclidean space can be determined based on a trade-off between two cost functions: the fraction of the variance in geodesic distance estimates not accounted for in the Euclidean embedding [(6); black triangles], and the fraction of points not included in the largest connected component of the neighborhood graph, and thus not included in the Euclidean embedding (gray filled circles, with vertical scale three times that shown). Setting the neighborhood size (x axis) too large (above ε = 6.2) masks the intrinsic low-dimensional structure of the data set, leading to geodesic distance estimates that are poorly represented by a 2D Euclidean embedding and thus a much higher residual variance. Setting the neighborhood size too small (below ε = 3.5) leaves out many points from the largest connected component of the neighborhood graph, leading to the deletion of those points from the Isomap solution and increased residual variance for the remaining points. A stable range of neighborhood sizes between these two extremes leads to embeddings of the entire data set with near-zero residual variance. (**F**) The same analysis can be used to select a reasonable neighborhood size for the noisy data. A neighborhood size that works well for the noiseless data (ε = 5) is too large here, yielding a residual variance of 0.25. However, a stable range of neighborhood sizes just slightly smaller (ε between 3.5 and 4.6) yields reasonable results, with residual variances all less than or equal to 0.01.

*Joshua B. Tenenbaum**
Department of Psychology
Stanford University
Stanford, CA 94305—2130, USA
E-mail: jbt@psych.stanford.edu

TECHNICAL COMMENTS

Vin de Silva
Department of Mathematics
Stanford University

John C. Langford
Department of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15217, USA

---

*Current address: Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

References and Notes

1. J. B. Tenenbaum, V. de Silva, J. C. Langford, *Science* **290**, 2319 (2000).
2. B. Fischl, M. I. Sereno, A. M. Dale, *NeuroImage* **9**, 195 (1999).
3. B. Wandell, S. Chial, B. Backus, *J. Cognit. Neurosci.* **12**, 739 (2000).
4. E. Wolfson, E. L. Schwartz, *IEEE Trans. Pattern Anal. Machine Intell.* **11**, 1001 (1989).
5. Balasubramanian and Schwartz add Gaussian noise with standard deviation equal to approximately 7.5% of the minimal separation between branches of the manifold as embedded in the 3D observation space. We characterize the magnitude of the noise with respect to the manifold's minimal branch separation, rather than the minimal extent of the data's bounding box, as do Balasubramanian and Schwartz, because for nonlinear manifolds, the former parameter influences the susceptibility to short-circuit errors (*1, 10*) while the latter does not.
6. We use a measure of residual variance [note 42 of (*1*)] to characterize how well the low-dimensional Euclidean embedding captures the geodesic distances estimated from the neighborhood graph. Lower residuals indicate better-fitting solutions, with less metric distortion.
7. It is possible to analyze some of the effects of noise through an extension of our asymptotic convergence arguments for noiseless data (*1, 10*). The principal danger with noisy data is that short-circuit edges may appear in the neighborhood graph, significantly changing its topology. In the noiseless situation, there is an upper bound ε_{max} for ε, below which we avoid these gross topological errors. If the noise is bounded by a constant c, then we must reduce this upper bound to $\varepsilon_{max} - 2c$. For Gaussian noise, one looks for a value of c where the probability that the noise is not bounded by c is as small as required. The calculation depends on the number of data points N and the variance of the noise; if the variance is not too large, this works reasonably well for realistic values of N.
8. S. T. Roweis, L. K. Saul, *Science* **290**, 2323 (2000).
9. We have not studied in detail the behavior of LLE with noisy data, but in the cases we have explored, its behavior seems to be qualitatively similar to that of Isomap. For instance, a neighborhood size of $\varepsilon = 5.5$ allows LLE to construct a topology-preserving embedding for the noiseless Swiss roll data in Fig. 1A but yields substantial topological errors for the noisy data in Fig. 1B.
10. M. Bernstein, V. de Silva, J. C. Langford, J. B. Tenenbaum, "Graph approximations to geodesics on embedded manifolds," preprint available at http://isomap.stanford.edu.
11. Supported by NSF, the DARPA Human ID program, and ONR. We thank M. Bernstein and an anonymous referee for helpful comments.