Graph approximations to geodesics on embedded manifolds

Mira Bernstein, Vin de Silva,
John C. Langford, Joshua B. Tenenbaum*

January 5, 2001

0 Introduction

In [1] Tenenbaum, de Silva and Langford consider the problem of non-linear dimensionality reduction: discovering intrinsically low-dimensional structures embedded in high-dimensional data sets. They describe an algorithm, called Isomap, and demonstrate its successful application to several real and synthetic data sets.

In this paper, we discuss some of the theoretical claims for Isomap made in [1]. In particular, we give a full proof of the asymptotic convergence theorem referred to in that paper.

Isomap deals with finite data sets of points in $\mathbb{R}^n$ which are assumed to lie on a smooth submanifold M^d of low dimension $d < n$. The algorithm attempts to recover M given only the data points. A crucial stage in the algorithm involves estimating the unknown *geodesic distance* in M between data points in terms of the *graph distance* with respect to some graph G constructed on the data points.

We show that the two distance metrics approximate each other arbitrarily closely, as the density of data points tends to infinity. Main Theorem A expresses this fact in terms of a sampling condition, and some conditions on

*The authors are listed alphabetically.

the graph. This can be checked explicitly in any given situation. The quality of the approximation depends, in part, on certain geometric parameters associated with the embedding of the submanifold.

Main Theorems B and C express this asymptotic convergence in probabilistic terms. Suppose that the data points are chosen randomly, with a certain density function α. We then seek theorems of the following form.

Asymptotic Convergence Theorem. *Given λ_1, λ_2, $\mu > 0$, then for α sufficiently large the inequalities*

$$1 - \lambda_1 \leq \frac{\text{graph distance}}{\text{geodesic distance}} \leq 1 + \lambda_2$$

hold with probability at least $1 - \mu$.

The graph in question is constructed in one of two ways, according to either an *ϵ-rule* or a *K-rule*, which we discuss later.

The paper is organised as follows. In Section 1 we give precise definitions of the set-up and of the various distance metrics considered. In Section 2, we describe the "sampling condition" which leads to the $1 + \lambda_2$ inequality in our convergence theorems. In Section 3, we explain how certain geometric parameters are involved in achieving the $1 - \lambda_1$ estimates. Main Theorem A appears at the end of Section 3.

In Sections 4 and 5 we give proofs of the probabilistic versions, Main Theorems B and C. Section 4 deals with the ϵ-rule, and Section 5 generalises that work to cover the K-rule, which requires more complicated arguments. In both cases, we consider submanifolds M without intrinsic curvature, because it is easy write down explicit formulas (of the kind cited in [1]). The case when M is curved is not essentially different, but it becomes harder to write down the required estimates explicitly.

In Section 6, we give an overview of how these results fit into the overall Isomap framework. Finally, in an Appendix, we give the complete proof of a geometric lemma used in Section 3.

1 Finite samples on submanifolds

Let $M = M^d$ be a compact d-dimensional smooth submanifold of the Euclidean space $\mathbb{R}^n$. Boundary and boundary corners are permitted, so for instance M may be a d-dimensional cube.

The natural Riemannian structure on M (induced from the Euclidean metric on $\mathbb{R}^n$) gives rise to a manifold metric d_M defined by:

$$d_M(x, y) = \inf_{\gamma}\{\text{length}(\gamma)\}$$

where γ varies over the set of (piecewise) smooth arcs connecting x to y in M. Note that $d_M(x, y)$ is generally different from the Euclidean distance $\|x - y\|$.

Let $\{x_i\} \subset M$ be a finite set, whose elements we will refer to as "data points" or "sample points". These points may be chosen randomly, or obtained in some other manner. The Isomap algorithm attempts to recover the manifold distances $d_M(x_i, x_j)$, given only the data points $\{x_i\} \subset \mathbb{R}^n$. Of course, this can only be done approximately.

The construction makes use of a graph G on the data points. Given such a graph we can define two further metrics, just on the set of data points. Let x, y belong to the set $\{x_i\}$. We define:

$$\begin{aligned} d_G(x, y) &= \min_P (\|x_0 - x_1\| + \ldots + \|x_{p-1} - x_p\|) \\ d_S(x, y) &= \min_P (d_M(x_0, x_1) + \ldots + d_M(x_{p-1}, x_p)) \end{aligned}$$

where $P = (x_0, \ldots, x_p)$ varies over all paths along the edges of G connecting x $(= x_0)$ to y $(= x_p)$.

Given the data points and graph G, one can compute d_G without knowledge of the manifold M. This is a key stage in the Isomap algorithm. Since the real goal is to estimate d_M, we must show that d_G is a good approximation to d_M, under favourable circumstances.

The metric d_S is an intermediate approximation. The proof that $d_G \approx d_M$ falls naturally into two parts: $d_M \approx d_S$ and $d_S \approx d_G$.

Proposition 1. We have the inequalities:

$$\begin{aligned} d_M(x, y) &\leq d_S(x, y) \\ d_G(x, y) &\leq d_S(x, y) \end{aligned}$$

Proof. The first expression is just the triangle inequality for the metric d_M. The second inequality holds because the Euclidean distances $\|x_i - x_{i+1}\|$ are smaller than the arc-length distances $d_M(x_i, x_{i+1})$. $\square$

Thus the main task will be to show that d_S is not too much bigger than d_M, and that d_G is not too much smaller than d_S.

2 The sampling condition

There are simple conditions on the data set $\{x_i\}$ and the graph G which guarantee that d_S is a good approximation to d_M.

Theorem 2. *Let ϵ and δ be positive, with $4\delta < \epsilon$. Suppose:*

1. *The graph G contains all edges xy for which $d_M(x, y) \leq \epsilon$.*
2. *For every point m in M there is a data point x_i for which $d_M(m, x_i) \leq \delta$.*

Then for all pairs of data points x, y we have

$$d_M(x, y) \leq d_S(x, y) \leq (1 + 4\delta/\epsilon) d_M(x, y)$$

We refer to the second condition in the theorem as the "δ-sampling condition".

Proof. Let γ be any piecewise-smooth arc connecting x to y, with length $\ell = \text{length}(\gamma)$. We will find a path from x to y along edges of G whose length $d_M(x_0, x_1) + \ldots + d_M(x_{p-1}, x_p)$ is less than $(1 + 4\delta/\epsilon)\ell$. The right-hand inequality will follow upon taking the infimum over γ.

If $\ell \leq \epsilon - 2\delta$ then x and y are connected by an edge, which we can use as our path.

If $\ell > \epsilon - 2\delta$ we can write $\ell = \ell_0 + (\ell_1 + \ell_1 + \ldots + \ell_1) + \ell_0$, where $\ell_1 = \epsilon - 2\delta$ and $\epsilon - 2\delta \geq \ell_0 \geq (\epsilon - 2\delta)/2$. (There may be no ℓ_1 terms if ℓ is small.)

Now cut up the arc γ into pieces in accordance with this decomposition of its length ℓ. This gives us a sequence of points $\gamma_0 = x, \gamma_1, \ldots, \gamma_p = y$ along γ, which divide the length of γ as indicated.

Each point γ_i (for $i = 1, \ldots, p - 1$) lies within distance δ of a sample point x_i. We claim that the path $x x_1 x_2 \ldots x_{p-1} y$ satisfies our requirements. We estimate the length of each edge as follows:

$$\begin{aligned} d_M(x_i, x_{i+1}) &\leq d_M(x_i, \gamma_i) + d_M(\gamma_i, \gamma_{i+1}) + d_M(\gamma_{i+1}, x_{i+1}) \\ &\leq \delta + \ell_1 + \delta \\ &= \epsilon \\ &= \ell_1 \epsilon/(\epsilon - 2\delta) \end{aligned}$$

and similarly:

$$\begin{aligned} d_M(x, x_1) &\leq \ell_0\epsilon/(\epsilon - 2\delta) \\ d_M(x_{p-1}, y) &\leq \ell_0\epsilon/(\epsilon - 2\delta) \end{aligned}$$

Since $\ell_0\epsilon/(\epsilon - 2\delta) \leq \epsilon$, we find that each edge has length at most ϵ and so belongs to G. Adding the inequalities, we obtain:

$$\begin{aligned} d_M(x_0, x_1) + \ldots + d_M(x_{p-1}, x_p) &\leq \ell\epsilon/(\epsilon - 2\delta) \\ &< \ell(1 + 4\delta/\epsilon) \end{aligned}$$

The last step makes use of the general fact that $1/(1-t) < 1 + 2t$ when $0 < t < 1/2$. This completes the proof. □

Thus we see that d_S approximates d_M arbitrarily well, provided that the two conditions (in particular the δ-sampling condition) can be met.

3 Parameters of the manifold embedding

The assertion that $d_S \approx d_G$ requires us to consider various parameters of the embedded manifold.

The *minimum radius of curvature* $r_0 = r_0(M)$ is defined by:

$$\frac{1}{r_0} = \max_{\gamma,t} \{\|\ddot{\gamma}(t)\|\}$$

where γ varies over all unit-speed geodesics in M and t is in the domain D of γ. The second derivative is computed by regarding γ as a map $D \to \mathbb{R}^n$.

Any Euclidean sphere of radius r_0 has minimum radius of curvature equal to r_0; in particular this is true of circles of radius r_0 contained in some 2-dimensional plane. Intuitively, geodesics in M curl around “less tightly” than circles of radius less than $r_0(M)$.

The *minimum branch separation* $s_0 = s_0(M)$ is defined to be the largest positive number for which $\|x - y\| < s_0$ implies $d_M(x, y) \leq \pi r_0$, for $x, y \in M$.

The existence of r_0 and s_0 is guaranteed by the compactness of M. We now make use of these quantities.

Lemma 3. *If γ is a geodesic in M connecting points x and y, and if $\ell = \text{length}(\gamma) \leq \pi r_0$ then:*

$$2r_0 \sin(\ell/2r_0) \leq \|x - y\| \leq \ell$$

The right-hand inequality is just the fact that the line segment between x and y is the shortest arc connecting x to y. The proof of the left-hand inequality, which is somewhat technical, is deferred to the Appendix. Equality on the left is achieved in the case where γ is an arc of a circle of radius r_0. Thus Lemma 3 crystallises the notion that geodesics in M curl around less tightly than circles of radius r_0.

Using the fact that $\sin(t) \geq t - t^3/6$ for $t \geq 0$, we can write down a weakened form of Lemma 3:

$$(1 - \ell^2/24r_0^2)\ell \leq \|x - y\| \leq \ell$$

We make this change purely for cosmetic purposes; it makes it easy to see that when ℓ is small, all three terms are approximately equal.

There is also the "first-order" weakening

$$(2/\pi)\ell \leq \|x - y\| \leq \ell,$$

which is valid in the range $\ell \leq \pi r_0$.

We now have the main ingredients required to show $d_G \approx d_S$.

Corollary 4 *Let $\lambda > 0$ be given. Suppose the points x_i, x_{i+1} in M satisfy the conditions:*

$$\begin{aligned} \|x_i - x_{i+1}\| &< s_0 \\ \|x_i - x_{i+1}\| &\leq (2/\pi) r_0 \sqrt{24\lambda} \end{aligned}$$

Suppose also that there is a geodesic arc of length $d_M(x_i, x_{i+1})$ connecting x_i to x_{i+1}. Then:

$$(1 - \lambda) d_M(x_i, x_{i+1}) \leq \|x_i - x_{i+1}\| \leq d_M(x_i, x_{i+1})$$

Proof. The first assumption implies that $\ell = d_M(x_i, x_{i+1}) \leq \pi r_0$. Since we assume that this distance is represented by a geodesic arc, we can apply Lemma 3. First of all we have

$$\ell \leq (\pi/2)\|x_i - x_{i+1}\| \leq r_0\sqrt{24\lambda},$$

using the first-order approximation. It follows that $1 - \lambda \leq 1 - \ell^2/24r_0^2$ and so Lemma 3 implies the desired result. $\square$

It is clear how Corollary 4 relates to the crucial inequality $(1-\lambda)d_S \leq d_G \leq d_S$, since d_S and d_G are defined in terms of lengths of geodesics and lengths of straight line segments, respectively. However, one must be careful to ensure that all relevant graph edges have a corresponding geodesic arc.

This last assumption is not always valid. If M is a convex domain in R^n with a hole punched out of its interior, then points on opposite sides of the hole are not connected by a shortest geodesic.

It is useful to have a criterion to identify when this assumption is valid. We say that M is *geodesically convex* if any two points x, y in M are connected by a geodesic of length $d_M(x, y)$.

Examples. Convex domains in $\mathbb{R}^n$ are geodesically convex, as are compact Riemannian manifolds without boundary. In general, if M is a compact Riemannian manifold, then M is geodesically convex if and only if its boundary is convex, in an appropriate sense.

This notion can be put to use in theorems of the following kind.

Main Theorem A. *Let M be a compact submanifold of $\mathbb{R}^n$ and let $\{x_i\}$ be a finite set of data points in M. We are given a graph G on $\{x_i\}$, and positive real numbers $\lambda_1, \lambda_2 < 1$. We also refer to positive real numbers ϵ_{min}, ϵ_{max} and δ. Suppose*

1. *The graph G contains all edges xy of length $\|x - y\| \leq \epsilon_{min}$;*
2. *All edges of G have length $\|x - y\| \leq \epsilon_{max}$;*
3. *The data set $\{x_i\}$ satisfies the δ-sampling condition in M;*
4. *The submanifold M is geodesically convex.*

Then, provided that

5. *$\epsilon_{max} < s_0$, where s_0 is the minimum branch separation of M,*
6. *$\epsilon_{max} \leq (2/\pi)r_0\sqrt{24\lambda_1}$, where r_0 is the minimum radius of curvature of M,*
7. *$\delta \leq \lambda_2\epsilon_{min}/4$,*

it follows that the inequalities

$$(1-\lambda_1)d_M(x,y) \leq d_G(x,y) \leq (1+\lambda_2)d_M(x,y)$$

are valid for all x, y in M.

The theorem gives us conditions for ensuring that $d_G \approx d_M$ to any given degree of accuracy, by setting λ_1 and λ_2 as small as required.

Proof. Since $\|x-y\| \leq d_M(x,y)$, condition 1 implies that G contains all edges xy for which $d_M(x,y) \leq \epsilon_{min}$. We can therefore apply Theorem 2, which with Proposition 1 gives

$$d_G(x,y) \leq d_S(x,y) \leq (1+4\delta/\epsilon_{min})d_M(x,y) \leq (1+\lambda_2)d_M(x,y),$$

establishing the right-hand inequality.

To prove the left-hand inequality, choose a path $x_0x_1\ldots x_p$ connecting x $(=x_0)$ to y $(=x_p)$ along graph edges which minimises the total graph length. The conditions on G and ϵ_{max} imply that each pair x_i, x_{i+1} satisfies the hypotheses of Corollary 4 with $\lambda = \lambda_1$. Then:

$$\begin{aligned} d_M(x,y) &\leq d_M(x_0,x_1) + \ldots + d_M(x_{p-1},x_p) \\ &\leq (1-\lambda_1)^{-1}\|x_0-x_1\| + \ldots + (1-\lambda_1)^{-1}\|x_{p-1}-x_p\| \\ &= (1-\lambda_1)^{-1}d_G(x,y) \end{aligned}$$

This establishes the left-hand inequality. □

Aside. It may seem that geodesic convexity is a much stronger condition than we really need for our proof. We could say that M is ϵ-geodesically convex if whenever $\|x-y\| \leq \epsilon$ there is a geodesic of length $d_M(x,y)$ connecting x to y. Taking $\epsilon = \epsilon_{max}$, this is all that is required in the proof. Curiously enough, it turns out that ϵ-geodesic convexity implies full geodesic convexity for a metrically complete manifold, such as a compact manifold, so nothing is really gained by doing this.

4 Probabilistic bounds

In this section, we consider what happens when the data points $\{x_i\}$ are chosen randomly and the graph G is determined explicitly according to an ϵ-rule or a K-rule (see below). In particular, we wish to show that the conditions of Main Theorem A can be satisfied with high probability, when there is a sufficiently high density of points.

First we discuss the random process. Let $\alpha : M \to \mathbb{R}_+$ be any positive real-valued function on M.

Convention. The sample set $\{x_i\}$ is chosen according to a Poisson process with density function α, meaning that for any measurable subset $A \subseteq M$,

$$Pr(A \text{ contains exactly } k \text{ points in } \{x_i\}) = e^{-a} a^k / k!,$$

where $a = \int_A \alpha$.

Under this convention, the expected number of points in A is just a. We write α_{min} and α_{max} for the minimum and maximum values of α on the compact manifold M.

Remark. The exact choice of random process is not especially significant, since the arguments we use are quite general. The Poisson process is constructed so that disjoint regions behave independently of each other. Another option would be to select N points independently from a fixed probability distribution $\hat{\alpha}$. It is just as easy to use this process when formulating our results. In any case, when N is large, this is well-approximated by the Poisson process with density $\alpha = N\hat{\alpha}$.

We now describe the two graph constructions mentioned above.

Definition. The ϵ-rule includes the edge xy in G whenever $\|x - y\| < \epsilon$, for some chosen $\epsilon > 0$.

Definition. The K-rule includes the edge xy in G whenever y is one of the K nearest neighbours of x (or vice versa); "nearest" in the Euclidean metric.

It turns out that the ϵ-rule is easier to work with; conditions 1 and 2 of Main Theorem A are automatically satisfied when $\epsilon_{min} \leq \epsilon \leq \epsilon_{max}$.

We will later establish that conditions 1 and 2 also hold (with high probability) for the K-rule, under suitable conditions. First we turn our attention to condition 3, the δ-sampling condition.

Sampling Lemma. *Let $\mu > 0$, $\delta > 0$ be given, and suppose $\{x_i\}$ is a data set in M chosen randomly with density function α. Then the δ-sampling condition is satisfied with probability at least $1 - \mu$, provided that*

$$\alpha_{min} > \log(V/\mu V_{min}(\delta/4))/V_{min}(\delta/2).$$

Here V is the volume of M, and $V_{min}(r)$ is defined to be the volume of the smallest metric ball in M of radius r (with respect to d_M):

$$V_{min}(r) = \min_{x \in M} Vol(B_x(r))$$

Note that the crucial lower bound $V_{min}(r)$ is positive-valued, by compactness of M.

Proof. We begin by covering M with a finite family of metric balls (with respect to d_M) of radius $\delta/2$. Choose the sequence of centers $p_1, p_2, \ldots$ in such a way that

$$p_{i+1} \notin \bigcup_{j=1}^{i} B_{p_j}(\delta/2).$$

When this is no longer possible, the job is done. Meanwhile, note that the smaller balls $B_{p_i}(\delta/4)$ are all disjoint (since no two of the p_i are within distance $\delta/2$ of each other). It follows that at most $V/V_{min}(\delta/4)$ points can be chosen before the process necessarily terminates.

Now every x in M belongs to some ball $B_i = B_{p_i}(\delta/2)$. If we can show that each ball B_i contains a data point, then the δ-sampling condition must be satisfied, since the diameter of B_i is at most δ.

Now we can compute:

$$\begin{aligned}
Pr(\delta\text{-sampling condition holds}) &= Pr(\text{no ball } B_i \text{ is empty}) \\
&= 1 - Pr(\text{some ball } B_i \text{ is empty}) \\
&\geq 1 - \sum_i Pr(B_i \text{ is empty})
\end{aligned}$$

Now

$$\begin{aligned}
Pr(B_i \text{ is empty}) &= \exp(-\int_{B_i} \alpha) \\
&\leq \exp(-V_{min}(\delta/2)\alpha_{min})
\end{aligned}$$

Since there are at most $V/V_{min}(\delta/4)$ balls, we can finally estimate:

$$Pr(\delta\text{-sampling condition holds}) \geq 1 - \frac{V}{V_{min}(\delta/4)} \exp(-V_{min}(\delta/2)\alpha_{min})$$

Simple algebraic manipulation shows that if α_{min} satisfies the condition in the statement of the lemma, then the right hand side must be greater than $1-\mu$. □

Remark. A similar result can be derived in the scenario where we sample N points independently from the fixed probability distribution $\hat{\alpha}$. This time the condition

$$N\hat{\alpha}_{min} > \log(V/\mu V_{min}(\delta/4))/V_{min}(\delta/2)$$

ensures that the δ-sampling condition holds with probability at least $1-\mu$. The proof is virtually identical.

Before we can make effective use of the Sampling Lemma, we must say something about the bounds $V_{min}(r)$.

First consider the case where M has no boundary.

1. If M is intrinsically flat, then balls of radius r look exactly like Euclidean balls (at least, so long as r is small enough to ensure that no ball overlaps itself). Thus we have the exact formula $V_{min}(r) = \eta_d r^d$, where η_d is the volume of the unit ball in $\mathbb{R}^d$.

2. Even if M is not intrinsically Euclidean, it is true that very small balls in M are almost Euclidean. In particular $V_{min}(r)/r^d \to \eta_d$ as $r \to 0$. If r is expected to be small, this indicates that it is reasonable to approximate $V_{min}(r) \approx \eta_d r^d$.

3. On the other hand, strict lower bounds for $V_{min}(r)$ may be derived using so-called "volume comparison theorems", in terms of the minimum radius of curvature r_0.

When M has boundary (but not corners), the situation is more complicated.

4. In the Euclidean case, metric balls $B_x(r)$ close to the boundary may be sliced by the boundary. If the boundary is itself flat, then the worst case is that half the ball will be missing.

5. If the boundary is not flat, it is still a reasonable first-order approximation to assume that $Vol(B_x(r)) \geq \eta_d r^d/2$. This is true also when M is not flat. Indeed, $Vol(B_x(r))/r^d \to \eta_d r^d/2$ as $r \to 0$ whenever x is a boundary point of M.

6. One may derive more complicated comparison theorems to give correct lower bounds for $V_{min}(r)$ in general. The curvature of the boundary will necessarily appear as a parameter in these theorems.

On the other hand, it is not hard to argue that these complicated boundary terms play comparatively small part in the workings of the Sampling Lemma. The number of boundary balls in a covering by balls of radius r is of order $1/r^{d-1}$, whereas the number of interior balls is of order $1/r^d$, as $r \to 0$.

In the interests of simplicity, we will henceforth disregard boundary effects. Correct modifications to the proofs can be found, if desired, by using $V_{min}(r)$ in place of the Euclidean formula $\eta_d r^d$.

Main Theorem B. *Let M be a compact submanifold of $\mathbb{R}^n$ isometrically equivalent to a convex domain in $\mathbb{R}^d$. Let λ_1, λ_2 and μ be given, and let $\epsilon > 0$ be chosen so that $\epsilon < s_0$ and $\epsilon \leq (2/\pi)r_0\sqrt{24\lambda_1}$. A sample data set $\{x_i\}$ is chosen randomly from a Poisson distribution with density function α, and the ϵ-rule is used to construct a graph G on $\{x_i\}$.*

Suppose also that:

$$\alpha_{min} > [\log(V/\mu\eta_d(\lambda_2\epsilon/16)^d)]/\eta_d(\lambda_2\epsilon/8)^d,$$

Then, with probability at least $1-\mu$, and neglecting boundary effects, the inequalities

$$(1-\lambda_1)d_M(x,y) \leq d_G(x,y) \leq (1+\lambda_2)d_M(x,y)$$

hold for all x, y in M.

Proof. This is a straightforward combination of Main Theorem 1 and the Sampling Lemma. □

5 Extending Main Theorem B to the K-rule

In this section we will extend our results to cover the K-rule. What is new is that conditions 1 and 2 in Main Theorem A are no longer automatically satisfied. We will find high-probability guarantees for both conditions. Here they are again:

1. The graph G contains xy as an edge whenever $\|x-y\| \leq \epsilon_{min}$.

2. All edges of G have length $\|x - y\| \le \epsilon_{max}$

The intuitive idea may be described very simply. Consider a Poisson distribution in $\mathbb{R}^d$ with constant density function α. The expected number of points in a ball of radius ℓ is $\alpha \eta_d \ell^d$. Putting it another way, if we set $K + 1 = \alpha \eta_d \ell^d$ and construct a graph according to the K-rule, then we can expect the maximum edge length at any given vertex to be around ℓ (in the manifold metric).

If we choose $\ell_{min} < \ell < \ell_{max}$, then the neighbourhood $B_x(\ell_{min})$ of the data point x is likely to contain fewer than K other data points; and the neighbourhood $B_x(\ell_{max})$ is likely to contain more than K other data points. If both of these things are true at x, then the K nearest-neighbour edges at x will:

1. include all edges xy where $\|x - y\| \le \epsilon_{min} = (2/\pi)\ell_{min}$;
2. not include any edge xy where $\|x - y\| > \epsilon_{max} = \ell_{max}$.

These are precisely the conditions we wish to prove, with high probability, for all points x. Therefore the task is to bound the number of points in $B_x(\ell_{min})$, $B_x(\ell_{max})$ below $K+1$ and above $K+1$ (respectively).

For a small fixed value of α, these bounds are not likely to be very effective. However, we can let α tend to infinity. In order to keep the length scale ϵ fixed, we preserve the ratio $(K+1)/\alpha$; so K also tends to infinity. Under these circumstances the strong law of large numbers prevails, and the probability that $B_x(\ell_{min})$ contains too many points, or that $B_x(\ell_{max})$ contains too few points, becomes exponentially small.

We now give the details of this argument, in greater generality. We return to the setting of a smooth compact submanifold $M \subset \mathbb{R}^n$, with data points $\{x_i\}$ chosen at random according to a Poisson process with density function α, which is bounded $\alpha_{min} \le \alpha \le \alpha_{max}$.

The ℓ_{min} Theorem. *Let ℓ_{min} be chosen to satisfy:*

$$\alpha_{max} V_{max}(2\ell_{min}) < (K+1)/2$$

where $V_{max}(r)$ is the volume of the largest metric ball in M of radius r. Then, with probability at least $1 - \mu$, no ball $B_x(\ell_{min})$ of radius ℓ_{min} contains more than $K+1$ data points. Here

$$\mu = (e/4)^{(K+1)/2} \frac{V}{V_{min}(\ell_{min}/2)}.$$

The ℓ_{max} Theorem. *Let ℓ_{max} be chosen to satisfy:*

$$\alpha_{min} V_{min}(\ell_{max}/2) > 2(K+1).$$

Then, with probability at least $1-\mu$, no ball $B_x(\ell_{max})$ of radius ℓ_{max} contains fewer than $K+1$ data points. Here

$$\mu = e^{-(K+1)/4} \frac{V}{V_{min}(\ell_{max}/4)}.$$

The proof of these results hinges on the following lemma, which quantifies the strong law of large numbers (in this case, for the Poisson distribution).

Lemma (Chernoff bounds). *If X is a random variable with Poisson distribution with mean m, then for all $t > 0$*

$$Pr(X > (1+t)m) < \left[\frac{e^t}{(1+t)^{1+t}} \right]^m$$

and

$$Pr(X < (1-t)m) < e^{-mt^2/2}.$$

Proof. See [2]. □

In particular, we can deduce:

$$\begin{array}{rcl} Pr(X > 2m) & < & (e/4)^m \\ Pr(X < m/2) & < & e^{-m/8} \end{array}$$

Proof of ℓ_{min} and ℓ_{max} theorems.

To prove the ℓ_{min} theorem, we first cover M with metric balls of radius ℓ_{min}. As discussed previously, this requires at most $V/V_{min}(\ell_{min}/2)$ balls. Let p be one of the centers of these balls. Consider the ball $B_p(2\ell_{min})$ of twice the radius. The expected number of points in this ball is at most $\alpha_{max} V_{max}(2\ell_{min})$ which by hypothesis is less than $(K+1)/2$. Applying the Chernoff bound we find that:

$$Pr(B_p(2\ell_{min}) \text{ contains} > K+1 \text{ data points}) < (e/4)^{(K+1)/2}$$

Since each point x in M lies within ℓ_{min} of one of the centers p, it follows that each ball $B_x(\ell_{min})$ is contained in some $B_p(2\ell_{min})$. Since there are at most $V/V_{min}(\ell_{min}/2)$ centers, we can estimate the error probability

$$Pr(\text{Some } B_x(\ell_{min}) \text{ contains} > K+1 \text{ points}) \leq (e/4)^{(K+1)/2} \frac{V}{V_{min}(\ell_{min}/2)}$$

as required.

Similarly, to prove the ℓ_{max} theorem, we cover M with metric balls of radius $\ell_{max}/2$, which requires at most $V/V_{min}(\ell_{max}/4)$ balls. Let p be one of the centers of these balls. The expected number of points in $B_p(\ell_{max}/2)$ is at least $\alpha_{min} V_{min}(\ell_{max}/2)$ which by hypothesis is greater than $2(K+1)$. Applying the Chernoff bound we find that:

$$Pr(B_p(\ell_{max}/2) \text{ contains} < K+1 \text{ data points}) < e^{-(K+1)/4}$$

Now each point x in M lies in $B_p(\ell_{max}/2)$ for some p, and $B_x(\ell_{max}) \supset B_p(\ell_{max}/2)$. Since there are at most $V/V_{min}(\ell_{max}/4)$ centers, we can estimate the error probability:

$$Pr(\text{Some } B_x(\ell_{max}) \text{ contains} < K+1 \text{ points}) \leq e^{-(K+1)/4} \frac{V}{V_{min}(\ell_{max}/4)}$$

This completes the proof of both theorems. □

We can put these results together to obtain our main result for the K-rule.

Main Theorem C. *Let M be a compact submanifold of $\mathbb{R}^n$ isometrically equivalent to a convex domain in $\mathbb{R}^d$. Let λ_1, λ_2 and μ be given, and let $\epsilon > 0$ be chosen so that $\epsilon < s_0$ and $\epsilon \leq (2/\pi) r_0 \sqrt{24\lambda_1}$. A sample data set $\{x_i\}$ is chosen randomly from a Poisson distribution with density function α, which has bounded variation $A = \alpha_{max}/\alpha_{min}$. Fix the ratio*

$$\frac{K+1}{\alpha_{min}} = \frac{\eta_d(\epsilon/2)^d}{2}$$

and use the K-rule to construct a graph G on $\{x_i\}$.

Suppose also that:

$$\alpha_{min} > [4 \log(8V/\mu\eta_d(\lambda_2\epsilon/32\pi)^d)]/\eta_d(\lambda_2\epsilon/16\pi)^d,$$

and that:

$$\begin{aligned} e^{-(K+1)/4} &\leq \mu\eta_d(\epsilon/4)^d/4V \\ (e/4)^{(K+1)/2} &\leq \mu\eta_d(\epsilon/8)^d/16AV \end{aligned}$$

Then, with probability at least $1 - \mu$, and neglecting boundary effects, the inequalities

$$(1-\lambda_1)d_M(x,y) \le d_G(x,y) \le (1+\lambda_2)d_M(x,y)$$

hold for all x, y in M.

Proof of Main Theorem C. Let $\epsilon_{max} = \ell_{max}$ and let $\epsilon_{min} = \ell_{min}$. Note that, by the first-order version of Lemma 3, for any $p \in M$, the Euclidean ball around p of radius ϵ_{min} in the ambient space $\mathbb{R}^n$ contains at most as many points $\{x_i\}$ as $B_p(\ell_{min})$, provided that $\epsilon_{min} < s_0$. Analogously, the Euclidean ball around p or radius ϵ_{max} in the ambient space $\mathbb{R}^n$ contains at least as many points $\{x_i\}$ as $B_p(\ell_{max})$, provided that $\epsilon_{max} < s_0$.

Thus we can replace the ℓ_{min} and ℓ_{max} theorems by analogous ϵ_{min} and ϵ_{max} theorems. This is useful because the ϵ- and K-rules are defined in terms of Euclidean distances, and not manifold distances. From the first parts of the ℓ_{min} and the ℓ_{max} theorems we obtain the conditions:

$$\begin{aligned} 2\eta_d(\pi\epsilon_{min})^d &\le (K+1)/\alpha_{max} \\ \frac{\eta_d(\epsilon_{max}/2)^d}{2} &\le (K+1)/\alpha_{min} \end{aligned}$$

We might as well take equality in both conditions, which makes

$$\epsilon_{min} = \frac{\epsilon_{max}}{2\pi(4A)^{1/d}}$$

(remembering that $\alpha_{max} = A\alpha_{min}$).

The theorem is now a straightforward combination of the Sampling Lemma, the ℓ_{min} theorem and the ℓ_{max} theorem. We obtain a lower bound on α_{min} to guarantee the Sampling Condition holds with probability at least $1-\mu/2$; and lower bounds on K to obtain error probabilities of at most $\mu/4$ in the ℓ_{min} and ℓ_{max} theorems. We can write these bounds in terms of $\epsilon = \epsilon_{max}$. The overall error probibilty is then at most $\mu/2 + \mu/4 + \mu/4 = \mu$. □

6 Concluding Remarks

We close with a brief overview of how these results yield the guarantees of asymptotic convergence described in the basic Isomap paper [1]. There are two main claims in [1]: that d_G converges asymptotically to d_M, and therefore that Isomap asymptotically recovers the true structure for data manifolds that are isometric to a convex domain of Euclidean space. We consider each of these pieces in turn.

The claim that d_G converges asymptotically to d_M is just a straightforward application of our Theorems B and C. The specific bounds on the deviation of d_G from d_M given in note 18 of [1] come from assuming zero intrinsic curvature and a constant Poisson density $\alpha = \alpha_{min} = \alpha_{max}$, and neglecting boundary effects. These bounds can be extended to deal with intrinsic curvature and boundary effects by deriving precise estimates on the quantities $V_{min}(r)$ for $r > 0$, as discussed in Section 4 following the proof of the Sampling Lemma. Asymptotically, introducing intrinsic curvature does not change the fact that when r is small, $V_{min}(r) \approx \eta_d r^d$ (except on the boundary where $V_{min}(r) \approx \eta_d r^d/2$).

The assertion that Isomap asymptotically recovers the true structure of intrinsically Euclidean data sets is really a statement about how the d-dimensional Euclidean embedding Y produced by Isomap relates to a hypothetical Euclidean configuration that can be thought of as the first stage of a generative model of the data. We imagine that the data were generated by first sampling points from V, a convex region of d-dimensional Euclidean space, and then mapping those points isometrically (and in general, nonlinearly) into a high-dimensional observation space X. The observed points in X (or more properly, their distances d_X) are then the inputs to Isomap, and the goal of the algorithm is to invert the nonlinear isometric mapping $V \to X$. Let $\mathbf{v}_i$ denote the coordinates of point i in the original Euclidean space V, which we would like to recover, and let $\mathbf{y}_i$ denote the coordinates of i in the embedding produced by Isomap. Our claim is that in the limit of infinitely many data points, the embedding coordinates $\{\mathbf{y}_i\}$ produced by applying MDS to any fixed finite set of points converge asymptotically to their original coordinates $\{\mathbf{v}_i\}$, up to Euclidean isomorphism (translation, rotation, scaling).

This claim rests on several properties of the algorithm for classical multidimensional scaling (MDS), which serves as the final step of Isomap. MDS takes as input the pairwise distances for a set of points and constructs a d-dimensional Euclidean embedding that captures those distances as closely as possible. When MDS is given as input actual Euclidean distances for a configuration of points in a d-dimensional space, it will construct a d-dimensional embedding that captures those distances perfectly and thereby recovers exactly the original configuration of the points, up to isomorphism. In general, any lower dimensional embedding will not capture the distances perfectly, so it is easy to detect that d is the true dimensionality of the data. Moreover, MDS is continuous: small bounded variations in the input distances lead to small bounded variations in the output embedding. These properties are described in detail by Cox & Cox [3].

As established in this note, the second step of Isomap produces distance estimates d_G that converge to the true geodesic distances d_M. If the original coordinates $\{\mathbf{v}_i\}$ were sampled from a convex Euclidean domain V, then the geodesic distances d_M will be equal to the Euclidean distances in V. These two facts guarantee that MDS (and hence Isomap) will asymptotically recover the original Euclidean configuration of the data. We can obtain quantitative bounds on how quickly the embedding coordinates $\{\mathbf{y}_i\}$ converge to the original coordinates $\{\mathbf{v}_i\}$ by combining the bounds in our Main Theorems B or C with the perturbation analysis of MDS on page 38 of [3].

Finally, we note that we have not said anything here about the ability of Isomap to recover low-dimensional nonlinear geometry in the presence of observation noise. Suppose that after being sampled from a convex d-dimensional Euclidean domain V and then isometrically embedded in the n-dimensional observation space X, the data points are perturbed by some low-amplitude isotropic noise process. We expect that the asymptotic result of applying Isomap to this noisy data will be a "thickened" embedding, with the top d embedding dimensions accounting for most of the variance in graph distances and the bottom $n-d$ dimensions accounting for some small residual variance whose amplitude can be bounded in terms of the magnitude of the noise. We leave a formal analysis of Isomap with noisy data to future work.

Acknowledgments

We thank S. Dasgupta, R. Lehrer, S. Mahajan, and H. Moore for helpful comments on this work.

References

[1] J. B. Tenenbaum, V. de Silva, J. C. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** (5500), 22 Dec 2000.

[2] H. Chernoff (1952). A measure of asymptotic efficiency for test of a hypothesis based upon the sum of observations. *Annals of Mathematical Statistics* **23**, 493-507.

[3] T. F. Cox and M. A. A. Cox (1994). *Multidimensional Scaling.* London: Chapman and Hall.

Appendix: Proof of Lemma 3

Lemma 3 comes from the following fundamental estimate:

Minimum Length Lemma. *Let $\gamma : [0, T] \to \mathbb{R}^n$ be a smooth arc which is parametrised by arc-length*

$$\|\dot{\gamma}(t)\| = 1$$

and which satisfies

$$\|\ddot{\gamma}(t)\| \leq 1/r$$

for all t in $[0, T]$. Then for all $0 \leq t \leq \pi r$ we have:

$$\|\gamma(t) - \gamma(0)\| \geq 2r \sin(t/2r)$$

In fact, the Minimum Length Lemma is valid in the range $0 \leq t \leq 2\pi r$, but we only require the restricted version of the lemma. For simplicity, that is what we will prove.

We adopt the following proof strategy. The arc γ is approximated by a polygonal arc. The condition $\ddot{\gamma}(t) \leq 1/r$ leads to upper bounds on the change in angle from one edge to the next. We prove a polygonal version of the Minimum Length Lemma which is stated in terms of these bounds. Finally, we can make the approximation arbitrarily accurate, which gives us the Minimum Length Lemma in the limit.

We begin by putting our approximations on a sound footing. Consider a piece of the arc γ of length ℓ. More precisely, let $x = \gamma(t)$ and $y = \gamma(t + \ell)$, where $0 \leq t \leq t + \ell \leq T$.

Lemma 5. *Let $\epsilon > 0$ be given. There exists $\delta > 0$ such that whenever we have x and y as above with $\ell < \delta$ the following inequalities hold.*

1. *The Euclidean edge length $\|y - x\|$ satisfies:*

$$\ell(1 - \epsilon) < \|y - x\| \leq \ell$$

2. The angle θ between the vector $y-x$ and the vector $\dot\gamma(t)$ satisfies:

$$\theta < \frac{\ell}{2(r-\epsilon)}$$

The inequality $\|y-x\| \le \ell$ is the familiar fact that straight line segments are shortest.

Remark. We could spare ourselves the effort of proving Lemma 3 by using instead the first inequality in Lemma 5, which is much easier to prove and has the same general form. The main results of the paper can be derived from Lemma 5, with slightly different explicit formulas. However we feel it is worth carrying out the extra work needed to obtain the 'correct' estimates.

Proof of Lemma 5. The fundamental theorem of calculus gives us the formula:

$$\begin{aligned} y-x &= \gamma(t+\ell)-\gamma(t) \\ &= \int_t^{t+\ell} \dot\gamma(a)\,da \\ &= \int_t^{t+\ell} [\dot\gamma(t) + \int_t^a \ddot\gamma(b)\,db]\,da \\ &= \ell\dot\gamma(t) + \int_t^{t+\ell}\int_t^a \ddot\gamma(b)\,db\,da \end{aligned}$$

which leads to the estimate:

$$\begin{aligned} \|(y-x)-\ell\dot\gamma(t)\| &\le \int_t^{t+\ell}\int_t^a \|\ddot\gamma(b)\|\,db\,da \\ &\le \int_t^{t+\ell}\int_t^a (1/r)\,db\,da \\ &= \ell^2/2r \end{aligned}$$

When ℓ is small this implies that the vectors $y-x$ and $\ell\dot\gamma(t)$ are very close. More precisely we have:

$$\begin{aligned} \|y-x\| &\ge \|\ell\dot\gamma(t)\| - \ell^2/2r \\ &= \ell(1-\ell/2r) \end{aligned}$$

which immediately gives the first inequality, by taking $\delta = 2r\epsilon$.

Moreover, the angle θ must be small when ℓ is small, since $\ell^2/2r$ is small when compared to ℓ. In more detail we can estimate:

$$\begin{aligned}\sin\theta &\leq \|(y-x)-\ell\dot{\gamma}(t)\|/\|\ell\dot{\gamma}(t)\| \\ &\leq (\ell^2/2r)/\ell \\ &= \ell/2r\end{aligned}$$

Since $(1-\epsilon)\theta < \sin\theta$ for θ close to zero, the second inequality in Lemma 5 also follows, for suitable δ. $\square$

We will now state and prove the required estimate for polygonal arcs. Afterwards we will show how to use Lemma 5 to deduce the Minimal Length Lemma.

Polygonal Minimal Length Lemma. *Let $V = V_0V_1\ldots V_p$ be a polygonal arc in $\mathbb{R}^n$, and let $\theta_1,\ldots,\theta_p$ be positive real numbers satisfying*

$$\theta_1 + \ldots + \theta_p < \pi.$$

Suppose the angle at vertex V_i satisfies the constraint

$$\angle V_{i-1}V_iV_{i+1} \geq \pi - \theta_i$$

for $i = 1,\ldots,p$. Let $W = W_0W_1\ldots W_p$ be a polygonal arc in $\mathbb{R}^2$ with the same edge lengths $\|W_{i+1}-W_i\| = \|V_{i+1}-V_i\|$ and with angles $\angle W_{i-i}W_iW_{i+1} = \pi - \theta_i$, consistently oriented clockwise or anticlockwise. Then

$$\|V_p - V_0\| \geq \|W_p - W_0\|$$

wth equality if and only if V lies in a plane and is isometrically equivalent to W.

Note that the angle constraints on V can be combined in a simple way to deduce the following inequalities:

$$\begin{aligned}\angle V_0V_iV_{i-1} &\leq \theta_1 + \ldots + \theta_{i-1} \\ \angle V_pV_iV_{i+1} &\leq \theta_{i+1} + \ldots + \theta_p\end{aligned}$$

These can be proved by induction; indeed

$$\angle V_0V_{i-1}V_i \geq (\pi - \angle V_0V_{i-1}V_{i-2}) - \theta_i$$

so

$$\begin{aligned}\angle V_0V_iV_{i-1} &\leq \pi - \angle V_0V_{i-1}V_i \\ &\leq \angle V_0V_{i-1}V_{i-2} + \theta_{i-1} \\ &\leq \theta_1 + \ldots + \theta_{i-2} + \theta_{i-1}\end{aligned}$$

as desired. The second inequality is similar.

We now begin the proof proper.

Proof of Polygonal Minimal Length Lemma. Fixing the starting edge V_0V_1 we note that we have a compact configuration space of paths V satisfying the given constraint. We can therefore assume that V is minimising for the function $\|V_p - V_0\|$ and endeavour to prove that V is isometrically equivalent to W.

Butterfly Lemma. *Let XQP and YQR be two triangles in $\mathbb{R}^n$ of fixed size and shape, and common vertex Q. Let ξ, η denote the angles $\xi = \angle XQP$, $\eta = \angle YQR$. Suppose also that we have the constraint*

$$\angle PQR \geq \pi - \theta$$

for some θ, where

$$\xi + \theta + \eta \leq \pi$$

Consider all configurations of triangles as above, with the same size and shape. Then the distance $\|Y - X\|$ is minimised if and only if the two triangles lie in the same plane, the angle $\angle PQR$ equals $\pi - \theta$, and the angles $\angle XQP$, $\angle YQR$ are contained in $\angle PQR$ (so in particular $\angle XQY = \pi - \xi - \theta - \eta$).

Assuming the Butterfly Lemma, we can now complete the proof. Suppose V is a minimising configuration. For any $0 < i < p$, consider the triangles $V_0V_iV_{i-1}$ and $V_pV_iV_{i+1}$. Since V is minimising, this configuration of triangles is minimising in the sense of the Butterfly Lemma, with $\theta = \theta_i$. We calculate:

$$\begin{aligned} \xi + \theta + \eta &= \angle V_0V_iV_{i-1} + \theta_i + \angle V_pV_iV_{i+1} \\ &\leq \theta_1 + \ldots + \theta_{i-1} + \theta_i + \theta_{i+1} + \ldots + \theta_p \\ &< \pi \end{aligned}$$

so the Butterfly Lemma applies. We deduce that all five points V_0, V_{i-1}, V_i, V_{i+1}, V_p lie in the same plane, that the angle $\angle V_{i-1}V_iV_{i+1}$ equals $\pi - \theta_i$, and that $\angle V_0V_iV_{i-1}$, $\angle V_pV_iV_{i+1}$ are interior to $\angle V_{i-1}V_iV_{i+1}$.

Since this is true for every $0 < i < p$, it follows that V is isometrically equivalent to W. This completes the proof of the Polygonal Minimal Length Lemma, modulo the Butterfly Lemma. □

Proof of Butterfly Lemma. It is equivalent to consider the problem of minimising the angle $\angle XQY$, since the lengths XQ, QY are fixed. Let x, p, r, y be points in the unit sphere S^{n-1} corresponding to the vector directions

$(X-Q)$, $(P-Q)$, $(R-Q)$, $(Y-Q)$ respectively, and let $\bar{p}$ denote the antipodal point to p. Angles at Q can now be represented as geodesic distances in the sphere:

$$\begin{aligned}
\angle XQY &= d(x,y) \\
\angle XQP &= d(x,p) \\
\angle YQR &= d(y,r) \\
\angle PQR &= d(p,r)
\end{aligned}$$

The constraint on $\angle PQR$ may be interpreted as:

$$d(r,\bar{p}) \leq \theta$$

We can now put this all together. By the triangle inequality for S^{n-1} we know that

$$d(p,\bar{p}) \leq d(p,x) + d(x,y) + d(y,r) + d(r,\bar{p})$$

with equality precisely when x, y, r lie (in that order) along a minimising geodesic arc from p to $\bar{p}$. Thus:

$$\pi \leq \xi + d(x,y) + \eta + d(r,\bar{p})$$

It follows that $d(x,y) \geq \pi - \xi - \theta - \eta$ with equality precisely when $pxyr\bar{p}$ is a minimising geodesic and $d(r,\bar{p}) = \theta$. Such a geodesic certainly exists when $\xi + \theta + \eta \leq \pi$. All the conclusions stated in the lemma now follow. $\square$

We are now in a position to prove the main result.

Proof of Minimum Length Lemma Let $\gamma : [0,T] \to \mathbb{R}^n$ be an arc of length T, with $\|\dot{\gamma}(t)\| = 1$ and $\|\ddot{\gamma}(t)\| \leq 1/r$ for all t.

We may assume that $T < \pi r$. (The case $T = \pi r$ follows by continuity). In fact, assume that $T \leq \pi(r - \epsilon)$ for some $\epsilon > 0$. Let $\delta = \delta(\epsilon)$ be as in Lemma 5.

Choose a large integer p and let $\ell = T/p$. Assume p is large enough to ensure that $\ell < \delta$. Let $U = U_0U_1 \ldots U_p$ be the polygonal arc in $\mathbb{R}^n$ with $U_i = \gamma(i\ell)$.

By Lemma 5 we have a uniform lower bound $\pi - \ell/2(r - \epsilon)$ on the angles $\angle U_{i-1}U_iU_{i+1}$. By assumption, $p\ell/2(r - \epsilon) \leq \pi$, so the Polygonal Minimal Length Lemma will apply.

Before we use the lemma, let us first modify U. Consider the polygonal arc $V = V_0V_1 \ldots V_p$ constructed so that $V_{i+1} - V_i$ is parallel to $U_{i+1} - U_i$ and rescaled to have length ℓ.

Applying the Polygonal Minimal Length Lemma, we find that $\|V_p - V_0\| \geq \|W_p - W_0\|$, where $W = W_0 W_1 \ldots W_p$ is a polygonal arc in the plane with equal edges of length ℓ and exterior angles of size $\ell/2(r-\epsilon)$.

Now we can allow ϵ to tend to zero. Using Lemma 5, we find that $V_p - V_0$ converges to $\gamma(T) - \gamma(0)$. By inspection, or by using Lemma 5, we find that $W_p - W_0$ converges to a chord in a circle of radius r spanning an arc of length T. This completes the proof. $\square$

Remark. We expect that there is a shorter proof of the Minimum Length Lemma using calculus.