Random Projections for Manifold Learning: Proofs and Analysis

Chinmay Hegde, Michael B. Wakin and Richard G. Baraniuk

September 14, 2007

Abstract

We derive theoretical bounds on the performance of manifold learning algorithms, given access to a small number of random projections of the input dataset. We prove that with the number of projections only logarithmic in the size of the original space, we may reliably learn the structure of the nonlinear manifold, as compared to performing conventional manifold learning on the full dataset.

1 Background

Suppose X is a set of points drawn from a uniform density supported on a compact K-dimensional submanifold $\mathcal{M}$. Often, the ambient dimension N is of intractable size, and we would like to represent our data in a space of reduced dimension. There are several nonlinear manifold learning algorithms which perform this task [1–3], and most of these algorithms attempt to preserve as best as possible the local metric structure of sample points. An alternative approach to dimensionality reduction is the method of *random projections* of manifold samples. Suppose we examine the effect of operating Φ, a random orthoprojector from $\mathbb{R}^N$ into $\mathbb{R}^M$, on a K-dimensional manifold $M \subset \mathbb{R}^N$. Then, if $M = O(K \log N)$, all pairwise geodesic and Euclidean distances between points on the manifold are well-preserved under this mapping with high probability [4].

In this technical report, we rigorously prove that with a sufficiently high number of "measurements" M per sample point, we may learn the structure of the manifold directly from such random projections. In Section 2, we establish a lower bound on the number of measurements M such that the intrinsic dimension of the manifold $\mathcal{M}$ can be reliably estimated from the projected dataset. In Section 3, we derive a bound on the performance of a popular manifold learning algorithm - Isomap - which operates directly on a specified small number of random projections of the dataset.

2 The GP algorithm

An important preprocessing step in manifold learning is the issue of estimating K, the intrinsic dimension (ID) of $\mathcal{M}$; this has been widely studied in the literature [5–7].

A common geometric approach for ID estimation is the Grassberger-Procaccia (GP) algorithm [5], which involves calculation and processing of pairwise distances between the data points. A sketch of the algorithm is as follows: given a set of points $X = \{x_1, x_2, \ldots\}$ sampled from $\mathcal{M}$, define $Q_j(r)$ as the number of points that lie within a ball of radius r centered around x_j. Let $Q(r)$

be the average of $Q_j(r)$ over all $j = 1, 2, \ldots, n$. Then, the estimated scale-dependent correlation dimension is the slope obtained by linearly regressing $\ln(Q(r))$ over $\ln(r)$ over the best possible linear part. It has been shown [8] that with increasing cardinality of the sample set X and decreasing scales, the estimated correlation dimension well approximates the intrinsic dimension of the underlying manifold.

In this section, we study the performance of the GP algorithm when given access to only a randomly projected version of the dataset. In particular, we establish a lower bound on the number of projections M required per sample point, such that the correlation dimension is estimated directly from the random projections up to any degree of accuracy.

2.1 Result

The notation used in this paper is identical to that in [4]. We have the following result to guarantee fidelity of the estimated correlation dimension:

Theorem 2.1 *Let $\mathcal{M}$ be a compact K-dimensional manifold in $\mathbb{R}^N$ having volume V and condition number $1/\tau$. Let $X = \{x_1, x_2, \ldots\}$ be a sequence of samples drawn from a* uniform *density supported on $\mathcal{M}$. Let $\widehat{K}$ be the dimension estimate of the GP algorithm on X over the range (r_{min}, r_{max}). Let $\beta = \ln(r_{max}/r_{min})$. Fix $0 < \delta < 1$ and $0 < \rho < 1$. Suppose the following condition holds:*

$$r_{max} < \tau/2. \tag{1}$$

Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with $M < N$ and

$$M \geq O\left(\frac{K \log(NV\tau^{-1}) \log(\rho^{-1})}{\beta^2 \delta^2}\right). \tag{2}$$

Let $\widehat{K}_\Phi$ be the estimated correlation dimension on ΦX in the projected space *over the range $(r_{min}\sqrt{M/N}, r_{max}\sqrt{M/N})$. Then, $\widehat{K}_\Phi$ is bounded by:*

$$(1-\delta)\widehat{K} \leq \widehat{K}_\Phi \leq (1+\delta)\widehat{K} \tag{3}$$

with probability exceeding $1 - \rho$.

Theorem2.1 prescribes a sufficient number of random projections required to estimate the correlation dimension of the dataset X, within an error fraction δ of the dimension estimate obtained by the ordinary GP algorithm on the original dataset. Note that the bound on M in the above theorem is in the worst case, and depends on several parameters, some intrinsic to the manifold (K, τ, V, β), as well as some that we can explicitly choose to obtain a desired performance guarantee (δ, ϵ).

2.2 Proof

The proof of Theorem 2.1 is based on the results describing the effect of the operator Φ on the metric structure of nearby points on a given K-dimensional manifold, as described in Section 3.2.4 in [4]. A quick sketch of the proof is as follows. Given a "tolerance" parameter, we can calculate the worst case metric distortion suffered by the manifold under the action of Φ, such that the estimated correlation dimension of the projected set ΦX is within $(1 + \delta)$ times the correlation dimension of

X. We impose a suitable bound on the largest scale used to estimate the correlation dimension (as specified in Equation 2). Finally, we use the main result in [4] to obtain a lower bound in the number of projections required to guarantee the desired accuracy of the ID estimate.

We say that a set X has isometry constant ϵ, if for every $x \in X$, the following relation holds:

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x\|_2}{\|x\|_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}}. \tag{4}$$

Next, we make use of the following lemmata to lead up to the proof.

Lemma 1 Suppose $r < r_{\max}$, the maximum permissible ball radius used to estimate box counting dimension around any point $x \in \mathcal{M}$. Let $Q_x(r)$ be the number of points within the ball. If (a) the manifold $\mathcal{M}$ is sampled densely and uniformly, and (b) $2r_{\max} < \tau$ hold true, then the maximum possible value for $Q_{\Phi x}(r\sqrt{M/N}) = Q_x(r)(1-\epsilon)^{-\widehat{K}}$, where $\widehat{K}$ is the estimated intrinsic dimension of $\mathcal{M}$.

Proof: This follows from the assumption that around r, the slope of the graph of $\ln Q$ versus $\ln r$ is linear with slope $\widehat{K}$. Under projection, owing to Lemma 1, the set of points within the ball experience an isometry constant ϵ. Hence, the worst case increase in Q is when every point in a ball of radius $r/(1-\epsilon)$ gets mapped into a ball of radius $r\sqrt{(M/N)}$ in the projected space. Hence, by linearity of the graph, the new number of points within the projected ball equals $Q_x(r) \times (1-\epsilon)^{-\widehat{K}}$. Hence, by a simple union bound argument, the worst case increase in $Q(r)$, obtained by averaging $Q_x(r)$ over all x, is also by a factor of $(1-\epsilon)^{-\widehat{K}}$. This seems to be a pathological case, but we retain it for our derivation of the greatest possible lower bound on the number of projections. By an identical argument, the worst case decrease in Q(r) is by a factor of $(1+\epsilon)^{\widehat{K}}$.
Care has to be taken to ensure that $r_{\max}$ is not too large, so as to avoid counting points from "faraway regions" in the manifold. This is captured in assumption (b), which relates $r_{\max}$ to the condition number τ of the manifold. (For properties relating distances between points and condition number, refer to, for instance, Section 2 in [4].)

Lemma 2 Suppose $\beta = \ln(r_{\max}/r_{\min})$. Then, if $\epsilon < \beta\delta/2$, the ID estimate of the projected data set is *guaranteed* to be within δ times the ID estimate of the original data set.

Proof: Assuming that the regression is done over $(r_{\min}, r_{\max})$, the slope of the linear region is given simply by $(\ln Q(r_{\max}) - \ln Q(r_{\min}))/\beta$. The worst case, which is again pathological, occurs when the two extremes have been multiplied by $(1+\epsilon)^{\widehat{K}}$ and $(1-\epsilon)^{\widehat{K}}$. Therefore, the worst possible error in calculation of the slope $= \widehat{K} \times \ln(((1+\epsilon))/(1-\epsilon))$. Converting to the natural logarithm and applying a Taylor series, we obtain the worst case slope estimation error as $2\widehat{K}\epsilon\beta$. However, we need this error to be less than $\widehat{K}\delta$. Rearranging, we get the required upper bound on ϵ.

Using Lemmas 1 and 2, we are now ready to prove our result. We have shown that in the worst case, ϵ has to be smaller than the bound given in Lemma 2. Hence eliminating ϵ from the two equations, we get Equation 2 in Theorem 2.1 as the constraint. The bound on the number of projections is obtained by simply plugging in the maximum allowable value of ϵ in Theorem 3.1 in [4]. This completes the proof.

3 Isomap

Manifold learning consists of finding a low-dimensional representation of data in order to facilitate visualization, classification, interpolation and other types of signal analysis. Eigenvalue-based learning algorithms such as PCA (Principal Components Analysis) assume that the data resides in a *subspace* of low dimension, and thus try to discover a linear transformation (such as a rotation or a scaling) that results in an approximate low-dimensional representation of the sample points. On the other hand, data modeled by nonlinear manifolds cannot be well approximated by linear subspaces. This has led to the development of *geometry-based* learning algorithms [1–3]. These algorithms produce a nonlinear mapping that embeds the data in a low-dimensional space, simultaneously trying to minimize a suitable objective function that represents the fidelity of the mapping in a local or global sense.

Isomap [1] is an nonlinear algorithm used to generate a mapping from a set of sample points belonging to a K-dimensional manifold $\mathcal{M}$, into a *Euclidean* space of dimension K. The implicit assumption is that the data manifold is itself Euclidean, i.e., the geodesic distances between points are equal to the ℓ_2 distances between their corresponding preimages in some unknown parameter space. Isomap, in essence, tries to discover the coordinate structure of that K-dimensional space.

Isomap works in two stages:

- A suitably defined graph G is constructed with the input data points acting as the vertices. Typically, nearby points are connected with edges and faraway points are not. By computing piecewise linear path lengths in the graph G, estimates of the geodesic distances between every pair of sample points are obtained.

- This set of geodesic distances is provided as input to the multidimensional scaling (MDS) algorithm. MDS has been used in classical statistics to discover the coordinate structure of objects, given a measure of pairwise "dissimilarities" between the input objects. Essentially, MDS performs an eigenanalysis of a suitable linear transformation of the matrix of squared geodesic distances, and the rank-K approximation of this new matrix yields the best possible K-dimensional coordinate structure of the input sample points in the mean-squared sense.

The minimum squared error in MDS is called *stress*; in the manifold learning literature, it is usually referred to as *residual variance.* The residual variance R serves as a global metric for how well Isomap manages to embed the input data. If the original data manifold is Euclidean, residual variance measures, in the ℓ_2 sense, the distance between a linearly transformed matrix of original manifold coordinates, and the matrix of functional space coordinates estimated by Isomap. Now, suppose that merely a few random projections of the data samples are available. The results in [4] place bounds on the distortion suffered by pairwise geodesic and ℓ_2 distances, given a certain number of measurements. The question now becomes: how well does Isomap perform with the only projected version of the data acting as the input? In other words, can we bound the error in the residual variance as a function of increasing number of measurements?

3.1 Result

The second of our main theoretical results prescribes a minimum number of measurements per sample point such that the residual variance produced by Isomap in the measurement domain is within an arbitrary additive constant of the Isomap produced with the full data.

Theorem 3.1 *Let $\mathcal{M}$ be a compact K-dimensional manifold in $\mathbb{R}^N$ having volume V and condition number $1/\tau$. Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of samples drawn from a sufficiently fine density (specified in the proof) supported on $\mathcal{M}$. Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with $M < N$. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Suppose*

$$M \geq O\left(\frac{K \log(NV\tau^{-1}) \log(\rho^{-1})}{\epsilon^2}\right).$$

Define the diameter Γ *of the dataset as follows:*

$$\Gamma = \max_{1 \leq i,j \leq n} d_{iso}(x_i, x_j)$$

where $d_{iso}(x, y)$ stands for the Isomap estimate of the geodesic distance between points x and y. Define R and R_Φ to be the residual variances obtained when Isomap generates a K-dimensional embedding of the original dataset X and projected dataset ΦX respectively. Under suitable constructions of the Isomap connectivity graphs (as outlined in the proof), R_Φ is bounded by:

$$R_\Phi < R + C\Gamma^2\epsilon$$

with probability exceeding $1 - \rho$. C is a function only on the number of sample points n.

The parameter ϵ is arbitrary, and hence we may choose a large enough M (which is still only logarithmic in N) such that the residual variance yielded by Isomap on the randomly projected version of the dataset is arbitrarily close to the variance produced with the native data. Again, the bound on M in the above theorem is in the worst case, and depends on several intrinsic (manifold) constants (such as K, τ , V , Γ), as well as extrinsic parameters (such as ρ and ϵ).

3.2 Proof

The proof follows in two stages. First, we derive a bound on the errors incurred in estimating pairwise geodesic distances from a suitably constructed graph involving only the randomly projected data points as vertices. Next, given that the errors in estimating geodesic distances are bounded, we derive a bound on the overall residual variance produced by the MDS step in Isomap.

A crucial component in the Isomap algorithm is the calculation of geodesic distances using the graph G. It has been rigorously proved [9] that if the underlying manifold is sampled with sufficiently high density, the estimates of geodesic distances using G well-approximate the lengths of the true underlying geodesics.

3.2.1 Graph distances using random measurements

Let $\mathcal{M}$ and $\{x_i\}$ be as specified in Theorem 3.1 and suppose G is a graph satisfying the conditions in Main Theorem A of [9]. Note that G is a graph with vertices $\{x_i\}$ and edges that connect some (but perhaps not all) pairs of points. The graph distance between two points u, v in the set $\{x_i\}$ is defined as

$$d_G(u, v) = \min_P(\|x_0 - x_1\|_2 + \cdots + \|x_{p-1} - x_p\|_2),$$

which considers piecewise Euclidean distances over all paths P in the graph G that join $u = x_0$ to $v = x_p$. Supposing that assumptions 1-7 are met in Main Theorem A (part of the assumption is that only nearby points are connected in G), the conclusion is that

$$(1 - \lambda_1)d_M(u, v) \leq d_G(u, v) \leq (1 + \lambda_1)d_M(u, v)$$

for all points u, v in the set $\{x_i\}$. (It claims this holds for all $u, v \in \mathcal{M}$, but d_G is defined only for points in G.)

Suppose the data has native dimension N, and let $\Phi : \mathbb{R}^N \to \mathbb{R}^M$ be a projection operator such that

$$(1 - \beta) \|u - v\|_2 \leq \|\Phi u - \Phi v\|_2 \leq (1 + \beta) \|u - v\|_2$$

for all $u, v \in \mathcal{M}$. (Note that this might be an orthoprojector renormalized by $\sqrt{N/M}$. Also note that this property actually need hold only for all $u, v \in \{x_i\}$.) Suppose also that

$$\frac{1+\beta}{1-\beta} < \frac{\epsilon_{\max}}{\epsilon_{\min}}.$$

Now, suppose projections $\{y_i = \Phi x_i\}$ of the samples are collected. With these nodes $\{y_i\}$ as vertices we would like to construct a new graph ΦG, and the key question is which nodes should be joined by edges. This must be concluded purely from the projections $\{y_i\}$ themselves, and not using the original connectivity of G as side information. Once the connectivity for ΦG is defined, we can define a new distance metric on the points $w, z \in \{y_i\}$:

$$d_{\Phi G}(w, z) = \min_P(\|y_0 - y_1\|_2 + \cdots + \|y_{p-1} - y_p\|_2)$$

which considers piecewise Euclidean distances over all paths P in the graph ΦG that join $w = y_0$ to $z = y_p$. Ultimately we hope to conclude that

$$d_M(u, v) \approx d_{\Phi G}(\Phi u, \Phi v)$$

for all $u, v \in \{x_i\}$.

Now, which nodes nodes in ΦG should be joined by edges? To define connectivity in ΦG, our rule is that two nodes $w, z \in \{y_i\}$ should be joined by an edge iff

$$\|w - z\|_2 \leq (1 + \beta)\epsilon_{\min}.$$

(Actually it is also acceptable to optionally permit any edge of length up to $(1 - \beta)\epsilon_{\max}$, but no greater.) Let us furthermore define a second graph G' on the native data points $\{x_i\}$ as follows: two nodes u, v in G' are connected by an edge iff their projections $\Phi u, \Phi v \in \{y_i\}$ are joined by an edge in ΦG. It is easy to check that G' (like G) meets all assumptions in Main Theorem A. To check assumption 1, suppose $\|u - v\|_2 \leq \epsilon_{\min}$. Then $\|\Phi u - \Phi v\|_2 \leq (1 + \beta) \|u - v\|_2 \leq (1 + \beta)\epsilon_{\min}$, and so Φu connects to Φv in ΦG and hence u connects to v in G'. To check assumption 2, suppose $\|u - v\|_2 > \epsilon_{\max}$. Then $\|\Phi u - \Phi v\|_2 \geq (1 - \beta) \|u - v\|_2 > (1 - \beta)\epsilon_{\max} > \frac{1+\beta}{1-\beta}\epsilon_{\min} > (1 + \beta)\epsilon_{\min}$, and so Φu does not connect to Φv in ΦG and hence u does not connect to v in G'.

We can see that distances in G' and ΦG must be similar. Let let P' be a path in G' joining $u, v \in \{x_i\}$ and let ΦP be the corresponding path in ΦG joining $\Phi u, \Phi v \in \{y_i\}$. Then stepping along ΦP,

$$\begin{aligned} \|y_0 - y_1\|_2 + \cdots + \|y_{p-1} - y_p\|_2 &= \|\Phi x_0 - \Phi x_1\|_2 + \cdots + \|\Phi x_{p-1} - \Phi x_p\|_2 \\ &\leq (1 + \beta)(\|x_0 - x_1\|_2 + \cdots + \|x_{p-1} - x_p\|_2). \end{aligned}$$

This holds for every path, and a similar lower bound holds for every path. It follows that

$$(1-\beta)d_{G'}(u,v) \leq d_{\Phi G}(\Phi u, \Phi v) \leq (1+\beta)d_{G'}(u,v)$$

for all $u, v \in \{x_i\}$ and hence that

$$(1-\beta)(1-\lambda_1)d_M(u,v) \leq d_{\Phi G}(\Phi u, \Phi v) \leq (1+\beta)(1-\lambda_2)d_M(u,v)$$

for all $u, v \in \{x_i\}$. Thus the projected graph distances (which can be computed purely from the projected data $\{y_i\}$) provide a faithful approximation of the geodesic distance on the manifold.

3.2.2 Isomap residual variance using perturbed input distances

The final question to be addressed is as follows: how does the performance of Isomap on the data set change, given the maximum perturbation ϵ that each of our input distances can possibly suffer? We use the residual variance (or stress) as the metric to quantitatively describe the performance of Isomap on a given dataset.

To define stress a little more clearly, let n be the number of sample points and $\mathbf{D} = (d_{rs})^2$ be the $n \times n$ matrix of squared geodesic distances between sample points r and s. Isomap computes the matrix $\mathbf{B} = (b_{rs})$:

$$b_{rs} = -\frac{d_{rs}^2 - \frac{1}{n}\sum_r d_{rs}^2 - \frac{1}{n}\sum_s d_{rs}^2 + \frac{1}{n^2}\sum_{r,s} d_{rs}^2}{2}.$$

$\mathbf{B}$ is shown to satisfy the relation $\mathbf{B} = \mathbf{X^T X}$, where $\mathbf{X}$ (size $K \times n$) is the (centered) set of K-dimensional Euclidean coordinates that represents the presumed embedding of the n points. The final step finds the K-dimensional representation of every point by performing an eigenvalue decomposition of $\mathbf{B}$ and obtaining the coordinates by projecting X onto the subspace spanned by the eigenvectors vi corresponding to the K largest eigenvalues $\lambda_i, i = 1, 2, \ldots, K$. The stress, or residual variance R, is defined as the sum of the $n - K$ smallest (positive) eigenvalues of $\mathbf{B}$. In an ideal scenario, $\mathbf{B}$ would be of rank K and the smallest $n - K$ eigenvalues (consequently, R) would be equal to zero. R represents the deviation from Euclidean-ness, i.e. the inability of Isomap to embed all distances in Euclidean K-dimensional space.

Now, suppose that d_{rs} is perturbed by a fraction (smaller than ϵ). We know that an isometry constant of ϵ implies a squared isometry constant of 3ϵ. Since we have both upper and lower bounds on the perturbation of d_{rs}, we can immediately write down the following bound on the distortion suffered by b_{rs}:

$$|\Delta b_{rs}| < 6\epsilon\Gamma^2,$$

where Γ is the square root of the largest entry of $\mathbf{D}$, or the estimated diameter of our compact manifold. Therefore, under perturbation ϵ, the matrix B varies as:

$$\mathbf{B}(\epsilon) = \mathbf{B} + 6\epsilon\Gamma^2\mathbf{C},$$

where $\mathbf{C}$ is a matrix whose entries are from the interval $(-1, 1)$.

It can be shown [10] that if $\mathbf{B}$ is perturbed by $6\epsilon\Gamma^2\mathbf{C}$, the eigenvalues λ_i of the new matrix are approximated by the following relation (again, neglecting quadratic terms):

$$\lambda_i(\epsilon) = \lambda_i + 6\epsilon\Gamma^2\mathbf{v_i^T C v_i}.$$

Assume that there is a cutoff to distinguish between the K^{th} and the $(K+1)^{\text{th}}$ largest eigenvalues, so that there is no significant reordering of eigenvalues. (A strong way to enforce this would be to impose the condition that ϵ should be small enough that the first $K+1$ eigenvalues $\lambda_i, i = 1, 2, \ldots, K, K+1$ maintain their respective positions after re-sorting according to absolute value.) Hence, the residual variance as a function of ϵ can be written as:

$$\begin{aligned} R(\epsilon) &= \sum_{i=K+1}^{n} \lambda_i(\epsilon) \\ &= \left(\sum_{i=K+1}^{n} \lambda_i\right) + 6\epsilon\Gamma^2 \left(\sum_{i=K+1}^{n} \mathbf{v_i^T C v_i}\right) \\ &= R + 6\epsilon\Gamma^2 \left(\sum_{i=K+1}^{n} \mathbf{v_i^T C v_i}\right). \end{aligned}$$

Since all eigenvectors are orthonormal, the quantity $\mathbf{v_i^T C v_i}$ can be bounded by the maximum eigenvalue Λ of the matrix $\mathbf{C}$. Rearranging, we get the following upper bound on the error on the change in the residual variance ΔR:

$$\begin{aligned} \Delta R(\epsilon) &< 6\epsilon\Gamma^2\Lambda(n-K) \\ &\approx 6\epsilon\Gamma^2\Lambda n. \end{aligned}$$

for small K. Therefore, the change in the "average" embedding distortion R_{av} *per sample point*, under the effect of random projections and sampling the manifold, varies with ϵ as:

$$\Delta R_{av}(\epsilon) < 6\epsilon\Gamma^2\Lambda.$$

3.2.3 Bound on the number of projections M

In Section 3.2.1, we proved that given a number of random projections of the data, the distances between points calculated by the Isomap algorithm using a suitable connectivity graph well-approximate the actual geodesic distances. By combining β and $\max(|\lambda_i|)$, we can derive an overall "isometry constant" ϵ [1] which is guaranteed under the action of the operator Φ. (If both β and $\max(|\lambda_i|)$ are small, a good approximation to ϵ is their sum $(\beta + \max(|\lambda_i|))$.)

The final equation in Section 3.2.2 gives us a condition on the number of random projections M required to obtain arbitrarily small ϵ. This is obtained by plugging the desired value of ϵ into the main result in [4]. Note that Λ and Γ could potentially be large, thus yielding a large prescribed value for the number of projections. Λ is bounded (since $\mathbf{C}$ is a bounded linear operator) and depends only on the size of $\mathbf{C}$. However, this is just a sufficiency condition and in practice, we can make do with far fewer measurements. Also, there is no dependence on N, the ambient dimension, in the bound derived above. Thus, the advantages of analyzing random measurements become highly evident as N grows intractably large.

[1] Note that this distortion parameter ϵ is separate from $\epsilon_{\max}$ and $\epsilon_{\min}$, which are parameters used in the graph approximation step in Isomap. Also, λ_i in this context is unrelated to the eigenvalues of the inner product matrix $\mathbf{B}$.

4 Conclusion

The bounds derived in this technical report prove the utility of using random projections as a tool for obtaining a reduced representation of high-dimensional data, which can be used to perform both ID estimation and subsequent manifold learning.

References

[1] J. B. Tenenbaum, V.de Silva, and J. C. Landford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[2] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[3] D. Donoho and C. Grimes. Hessian eigenmaps: locally linear embedding techniques for high dimensional data. *Proc. of National Academy of Sciences*, 100(10):5591–5596, 2003.

[4] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. 2006. Preprint. Available at http://www.acm.caltech.edu/ wakin/publications.html .

[5] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D Nonlinear Phenomena*, 9:189–208, 1983.

[6] F. Camastra. Data dimensionality estimation methods: a survey. *Pattern Recognition*, 36:2945–2954, 2003.

[7] J. A. Costa and A. O. Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. Signal Processing*, 52(8):2210–2221, August 2004.

[8] B. Kégl. Intrinsic dimension estimation using packing numbers. In *Advances in NIPS*, volume 14. MIT Press, 2002.

[9] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds, 2000. Technical report, Stanford University.

[10] T. Cox and M. Cox. Multidimensional scaling, 1994. Chapman & Hall.