How Much Distortion Can be Caused by One Bad Point?

Krzysztof P. Onak * Jonathan Lenchner * Liu Yang *

Introduction

Various authors have studied the problem of minimizing the distortion of embedding points from one metric space into another metric space. In this work we consider the problem of embedding $N+1$ points from $\mathbb{R}^{K+1}$ into $\mathbb{R}^K$ where all but one of the $N+1$ points are in $\mathbb{R}^K$. Given such a set of $N+1$ points, how much distortion must necessarily be incurred, by the best embedding, and how does that compare to the case where an arbitrary number of points can lie outside of $\mathbb{R}^K$ (i.e. outside of any hyperplane of $\mathbb{R}^{K+1}$)? Questions of this nature are important in many application areas, from data compression to machine learning.

Notation: Let Π be an embedding of one metric space, $\mathcal{M}_1$ into a second metric space, $\mathcal{M}_2$. Let $d_1(x,y)$ denote the distance between two points $x,y \in \mathcal{M}_1$ and let $d_2(x,y)$ denote the distance between two points $x,y \in \mathcal{M}_2$. If one of the metric spaces is a Euclidean space $\mathbb{R}^K$, then denote the distance between points by $d_K(x,y)$.

Definition: Let P be a finite point set in a metric space $\mathcal{M}_1$, and let $\Pi : P \to \mathcal{M}_2$ be a mapping (embedding) of P into $\mathcal{M}_2$. Then the **distortion** of the mapping Π, Dist(Π) is given by

$$\mathrm{Dist}(\Pi) = \max\left(\max_{x,y\in P} \frac{d_2(\Pi(x),\Pi(y))}{d_1(x,y)}, \max_{x,y\in P} \frac{d_1(x,y)}{d_2(\Pi(x),\Pi(y))}\right).$$

If Π is non-contracting then

$$\mathrm{Dist}(\Pi) = \max_{x,y\in P} \frac{d_2(\Pi(x),\Pi(y))}{d_1(x,y)},$$

while if Π is non-expanding then

$$\mathrm{Dist}(\Pi) = \max_{x,y\in P} \frac{d_1(x,y)}{d_2(\Pi(x),\Pi(y))}.$$

1 Background and Related Work

A fundamental reference that discusses the Lipschitz extension theorem of Kirszbraun (see next section), and the now classical Johnson-Lindenstrauss-Schechtman lemmas is [4]. [1] and [2] study embedding metric spaces into a line, and into the two-dimensional plane. Our work is most closely related to [3], which discusses online metric embeddings.

2 Embeddings Points on a Sphere into $\mathbb{R}^2$ and Points on a Circle into $\mathbb{R}$

Badiou [2] et al. showed that any embedding of a dense set of points on the unit sphere, under the Euclidean metric of $\mathbb{R}^3$ embeds into $\mathbb{R}^2$ with distortion $\Omega(\sqrt{N})$. The proof uses the Borsuk-Ulam Theorem together with Kirszbraun's Theorem [4], which says that a Liptschitz embedding of a subset of a Hilbert Space into another Hilbert Space can be extended to a Liptschitz embedding of the full space, with the same Liptschitz constant. The same argument can be used to show that any embedding of a dense set of points on the unit circle, under the Euclidean metric of $\mathbb{R}^2$ embeds into $\mathbb{R}$ with distortion $\Omega(N)$.

3 Embedding N Points on a Line and One Point off the Line onto a Line

Lemma 1 *Consider a collection of an odd number, N, of points on a line, each point one unit from the next, together with one additional point at height $\sqrt{N}$ above the center point of the points on the line. Then any embedding of these points into a line has distortion $\Omega(\sqrt{N})$.*

Proof. Label the points consecutively along the line by $P = \{p_1,\ldots,p_N\}$, and refer to the point above the line at distance $\sqrt{N}$ by q. Further, denote the central point among the points in P, $p_{\frac{N+1}{2}}$, by p_{cent}.

We prove the lemma by contradiction. Thus suppose we have a non-contracting embedding Π of the points $P \cup \{q\}$ into a line, which we suppose has distortion that is $o(\sqrt{N})$. We suppose first that some of the points in P are mapped under Π to one side of $\Pi(q)$ and some to the other. It must then be the case that some pair of adjacent points p_i and p_{i+1} are mapped by Π to opposite sides of $\Pi(q)$ (if not, then starting with p_1, p_2 we could inductively conclude that all points are to one side of $\Pi(q)$). But if p_i is mapped to one side of $\Pi(q)$ and p_{i+1} is mapped to the other, then, by non-contraction, the distance between $\Pi(q)$ and each of its closest neighbors is at least $\sqrt{N}$ and thus $d(\Pi(p_i),\Pi(p_{i+1}) \geq 2\sqrt{N}$ so the distortion in Π is at least $2\sqrt{N}$. We therefore conclude that for Π to have distortion $o(\sqrt{N})$ all points in the embedding $\Pi(p_i)$ must be to one side of $\Pi(q)$.

Since all points $\Pi(p_i)$ are to one side of $\Pi(q)$ there is a closest neighbor $\Pi(p^*)$ to $\Pi(q)$. Divide the points of P, as evenly as possible into four sequential quarters (being exact is unimportant since the argument will just involve orders of magnitude in N). p^* either comes from one of the outer quarters of $\{p_1,\ldots,p_N\}$ or from one of the two inner quarters. Label these quarters $P_{1/4}, P_{2/4}, P_{3/4}$ and $P_{4/4}$, respectively.

Suppose first that $p^* \in P_{1/4} \cup P_{4/4}$. Then consider the distortion in the mapping of q and p_{cent}:

*IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, {konak,lenchner,yangli}@us.ibm.com

$$\frac{d(\Pi(q),\Pi(p_{\text{cent}}))}{d(q,p_{\text{cent}})} = \frac{d(\Pi(q),\Pi(p^*))+d(\Pi(p^*),\Pi(p_{\text{cent}}))}{\sqrt{N}}$$

$$\geq \frac{\frac{N}{4}+\frac{N}{4}}{\sqrt{N}} = \frac{\sqrt{N}}{2}.$$

On the other hand, suppose that $p^* \in P_{2/4} \cup P_{3/4}$. Then without loss of generality we may assume we have a left to right ordering of points $\Pi(q),\Pi(p^*),\Pi(p_1),\Pi(p_N)$. Note that this ordering could equally well be $\Pi(q),\Pi(p^*),\Pi(p_N),\Pi(p_1)$, or the points $\Pi(p^*),\Pi(p_1),\Pi(p_N)$ could all be to the left of $\Pi(q)$ – these differences do not affect our argument. By a trivial argument analogous to the one we gave earlier, we can conclude that there must be consecutive points p_i, p_{i+1} with $\Pi(p_i)$ to one side of $\Pi(p_1)$ and $\Pi(p_{i+1})$ to the other side of $\Pi(p_1)$ and $p_i \in \{p_k, p_{k+1}, ..., p_{N-1}\}$ where $p^* = p_k$. But now consider the distortion in the mapping of p_i and p_{i+1}, keeping in mind that p_i and p_{i+1} are in the 2nd, 3rd or 4th quartile, and the mapping is non-contracting:

$$\frac{d(\Pi(p_i),\Pi(p_{i+1}))}{d(p_i,p_{i+1})} = \frac{d(\Pi(p_i),\Pi(p_1))+d(\Pi(p_1),\Pi(p_{i+1}))}{1}$$

$$\geq \frac{N}{4}+\frac{N}{4} = \frac{N}{2}.$$

In all cases then Π has distortion $\Omega(\sqrt{N})$ and the lemma is established. □

4 Embedding N Points on a Plane and One Point off the Plane onto a Plane

Lemma 2 *Consider a dense set of N points, each approximately unit distance from its closest neighbors, inside a disk of radius $\sqrt{N}$ and one additional point at height $N^{1/4}$ above the center point of the points in the disk. Then any embedding of these points into the plane has distortion $\Omega(N^{1/4})$.*

Proof (sketch). Suppose we have a non-expanding embedding Π of the N points, P, in the disk, together with the point above the center of the disk, which we again call q, into the plane. Extend Π to be a non-expanding embedding of all of $\mathbb{R}^3$ into $\mathbb{R}^2$ by Kirszbraun's Theorem. Since Π is Lipschitz (with Lipschitz constant at most 1), Π is continuous. Let p_{cent} be the centerpoint in P directly below q. Consider the image under Π of the vertical diameter of the disk $\Pi(\text{diam})$. This image is a continuous curve through $\Pi(p_{\text{cent}})$. Color the top half of $\Pi(\text{diam})$ red and the bottom half green. Now consider the image $\Pi(\text{diam})$ as the diameter turns through 180 degrees. Continue to color $\Pi(\text{top-half})$ red and $\Pi(\text{bottom-half})$ green. A straight forward argument shows that either the endpoints of the red and green halves of these curves collectively form a closed curve with $\Pi(p_{\text{cent}})$ in its interior *or* at some point in the turning of the diameter either the end point of the green curve intersects the red curve or the end point of the red curve intersects the green curve. Suppose one of these latter two cases holds, say it is that the end point of the red curve intersects the green curve. If p_{r_e} is the pre-image of the end point of the red curve at this juncture, then there is a point p_g which is the pre-image of a point along the green curve such that $d(\Pi(p_{r_e}),\Pi(p_g)) \approx 1$ while the points p_{r_e}, p_g lie along a diameter and are at least $\sqrt{N}$ apart in the pre-image. Thus, in this case, $Dist(\Pi) = \Omega(\sqrt{N})$.

On the other hand, if $\Pi(p_{\text{cent}})$ is in the interior of the image of the disk then consider $\Pi(q)$, the image of the point above p_{cent}. If $\Pi(q)$ lies inside the image of the boundary of the disk, then since Π is non-expanding there is a point of the disk that is approximately distance 1 or less from $\Pi(q)$. Since the point started at least at distance $N^{1/4}$, the incurred distortion is $\Omega(N^{1/4})$. On the other hand, if the boundary of the disk lies between $\Pi(q)$ and $\Pi(p_{\text{dent}})$ then we again find a distortion of $\Omega(N^{1/4})$. □

Future Work

These results are just the first of a hoped for more detailed characterization of how one incurs distortion on a point-by-point basis embedding from one Euclidean space into another of smaller dimension. In general if N points in $\mathbb{R}^k$ can incur some maximum distortion when the points are embedded in $\mathbb{R}^{k'}$, for $k' < k$, how much distortion can be incurred from a point set of the same size N but where all but M of the N points lie in some k'-flat, where $M = o(N)$?

References

[1] M. Bădoiu, J. Chuzhoy, P. Indyk, and A. Sidiropoulos. Low-distortion embeddings of general metrics into the line. In *In STOC05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 225–233. ACM, 2005.

[2] M. Bădoiu, K. Dhamdhere, A. Gupta, Y. Rabinovich, H. Räcke, R. Ravi, and A. Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 119–128, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[3] P. Indyk, A. Magen, A. Sidiropoulos, and A. Zouzias. Online embeddings. In M. Serna, R. Shaltiel, K. Jansen, and J. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 6302 of *Lecture Notes in Computer Science*, pages 246–259. Springer Berlin Heidelberg, 2010.

[4] J. Lee and A. Naor. Absolute Lipschitz extendability. In *C. R. Acad. Sci. Paris*, volume 338, pages 859–862, 2003.