Ultra-Low-Dimensional Embeddings for Doubling Metrics*

T-H. Hubert Chan† Anupam Gupta† Kunal Talwar‡

We consider the problem of embedding a finite metric space (V, d) with small doubling dimension into low-dimensional Euclidean space, with provable bounds on the distortion. Roughly, a metric has *doubling dimension* $\dim_D = k$ if and only if it has 2^k points at roughly the same distance from each other, but no larger; a more formal definition can be found in [1]. A metric (V, d) is called *doubling* if the doubling dimension is a constant, independent of size $n = |V|$ of the metric space. Note that for a low-dimensional manifold $\mathcal{M}$, the geodesic distance metric $d_{\mathcal{M}}$ has doubling dimension depending only on the manifold dimension and not on the ambient dimension. Thus our results also give embeddings for a finite set of points lying on a low-dimensional manifold into a low-dimensional Euclidean space.

Dimension reduction in Euclidean spaces have been studied extensively. The celebrated and surprising "flattening" lemma of Johnson and Lindenstrauss [2] states that the dimension of any Euclidean metric on n points can be reduced to $O(\frac{\log n}{\epsilon^2})$ with $(1+\epsilon)$ distortion using a random linear map. Moreover, this result is existentially tight: a simple packing argument shows that any distortion-D embedding of a uniform metric on n points into Euclidean space requires at least $\Omega(\log_D n)$ dimensions—intuitively, there aren't enough distinct directions in a low dimensional Euclidean space to accommodate a large number of equidistant points. Hence we do need the $\Omega(\log n)$ dimensions, and even allowing $O(\log n)$ distortion cannot reduce the number of dimensions below $\Omega(\log n/\log\log n)$.

The Questions. The packing lower bound mentioned above shows that any metric requires $\Omega(\dim_D)$ dimensions for a constant-distortion embedding into Euclidean space: is this lower bound tight? Sadly, no: we now know the existence of n-point metrics with $\dim_D = O(1)$ that require $\Omega(\sqrt{\log n})$-distortion into Euclidean space with any number of dimensions [1]. But can we actually achieve this distortion with $o(\log n)$-dimensions? What if we are willing to incur a slightly higher distortion? Bourgain's classical result (along with the JL-lemma) shows that all metrics embed into Euclidean space of $O(\log n)$ dimensions and $O(\log n)$ distortion [4], but we do not even know if doubling metrics embed into $O(\log^{1-\epsilon} n)$ dimensions with $O(\log^{1-\epsilon} n)$ distortion for any constant $\epsilon > 0$.

If we restrict our attention to Euclidean doubling metrics, we know just as little: a tantalizing conjecture (first explicitly stated by Lang and Plaut [3]) asks if all Euclidean metrics with $\dim_D = O(1)$ embed into $O(1)$ dimensional Euclidean space with $O(1)$ distortion. However, the best result we know is still the JL-Lemma, which is completely oblivious to the doubling dimension; moreover, it is easy to show that linear maps cannot prove this conjecture. Again, we do not even know how to take a doubling *Euclidean* point set and flatten it into (say) $O(\log^{1-\epsilon} n)$ dimensions with $O(\log^{1-\epsilon} n)$ distortion!

The Answers. We make progress on the problem of embedding doubling metrics into Euclidean space with small dimension and distortion. (Our results hold for *all* doubling metrics, not just Euclidean ones.)

*An extended version of this abstract will appear in the *Proceedings of the 19th ACM-SIAM Symposium on Discrete Algorithms (SODA), 2008*.

†Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. This research was partly supported by the NSF grant CCR-0122581 (The ALADDIN project), the NSF CAREER award CCF-0448095, and by an Alfred P. Sloan Fellowship.

‡Microsoft Research, Silicon Valley Campus, 1065 La Avenida, Mountain View, CA 94043.

Details and proofs can be found in the extended abstract at `http://www.cs.cmu.edu/~hubert`.

Theorem 1 (Ultra-Low-Dimension Embedding) *Any metric with doubling dimension* $\dim_D < O(\frac{\log n}{\log\log n})$ *embeds into Euclidean space with* $O(\dim_D \log\log n)$ *dimensions and* $O(\frac{\log n}{\sqrt{\log\log n}})$ *distortion.*

Hence we can embed a doubling metric into *very few* Euclidean dimensions (i.e., $O(\dim_D)$ times a small $O(\log\log n)$ factor), and achieve a slightly smaller distortion than even Bourgain's embedding. Note that to achieve distortion $\tilde{O}(\log n)$, any metric with doubling dimension $\dim_D$ requires at least $\Omega(\frac{\dim_D}{\log\log n})$ Euclidean dimensions, and hence we within an $O(\log\log n)^2$ factor to the *optimal dimension* for this value of distortion.

The above result is a special case of our general theorem, which gives a smooth trade-off between the dimension of the target Euclidean space, and the distortion incurred by the map:

Theorem 2 (Main Trade-off Theorem) *Suppose* (V,d) *is a metric space with doubling dimension* $\dim_D$*. For any integer* T *such that* $\Omega(\dim_D \log\log n) \leq T \leq \ln n$*, there exists* $F : V \to \Re^T$ *into* T*-dimensional space such that for all* $x, y \in V$, $d(x,y) \leq \|F(x) - F(y)\|_2 \leq O\,(\sqrt{\frac{\dim_D}{T}}\ \log n) \cdot d(x,y)$.

Varying the target dimension T, we can get some interesting tradeoffs between the distortion and dimension. For instance, we can balance the two quantities and get $O(\log^{2/3} n)$ dimensions and $O(\log^{2/3} n)$ distortion for doubling metrics, as desired.

Future Directions.

Our work still leaves several questions open. Note that our results apply to *all doubling metrics*, and not just to Euclidean ones, and hence they are constrained by lower bound examples of metrics [1] where one cannot get a distortion smaller than $\Omega(\sqrt{\log n})$ while embedding into Euclidean space, even if one places no constraint on the dimension. An immediate question is whether one can start with a doubling metric already embedded in a Euclidean space, and obtain a better trade-off between dimension and distortion than the one obtained above. The ideal theorem along these lines would be to show that if a metric $\mathcal{M} = (V,d)$ embeds into Euclidean space with distortion D, then it can be embedded into $O(\dim_D(\mathcal{M}))$-dimensional Euclidean space with distortion $O(D)$. Moreover, while the embeddings promised by the above theorems are computable in polynomial time, the current running times are large, and another important question is to find efficient algorithms to find such embeddings.

References

[1] GUPTA, A., KRAUTHGAMER, R., AND LEE, J. R. Bounded geometries, fractals, and low–distortion embeddings. In *Proceedings of the 44th Symposium on the Foundations of Computer Science (FOCS)* (2003), pp. 534–543.

[2] JOHNSON, W. B., AND LINDENSTRAUSS, J. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics 26* (1984), 189–206.

[3] LANG, U., AND PLAUT, C. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata 87*, 1-3 (2001), 285–307.

[4] LINIAL, N., LONDON, E., AND RABINOVICH, Y. The geometry of graphs and some of its algorithmic applications. *Combinatorica 15*, 2 (1995), 215–245. (Preliminary version in *35th FOCS*, 1994).