Metric Embeddings Beyond Bi-Lipschitz Distortion via Sherali-Adams

Ainesh Bakshi AINESH@MIT.EDU
MIT

Vincent Cohen-Addad COHENADDAD@GOOGLE.COM
Google Research

Samuel B. Hopkins SAMHOP@MIT.EDU
MIT

Rajesh Jayaram RKJAYARAM@GOOGLE.COM
Google Research

Silvio Lattanzi SILVIOL@GOOGLE.COM
Google Research

Editors: Nika Haghtalab and Ankur Moitra

Abstract

Metric embeddings are a widely used method in algorithm design, where generally a "complex" metric is embedded into a simpler, lower-dimensional one. Historically, the theoretical computer science community has focused on *bi-Lipschitz* embeddings, which guarantee that every pairwise distance is approximately preserved. In contrast, alternative embedding objectives that avoid bi-Lipschitz distortion are commonly used in practice to map points to lower dimensions, yet these approaches have received relatively have received comparatively less study in theory.

In this paper, we focus on one such objective, Multi-dimensional Scaling (MDS), which embeds an n-point metric into low-dimensional Euclidean space. MDS is widely used as a data visualization tool in the social and biological sciences, statistics, and machine learning. Given a set of non-negative dissimilarities $\{d_{i,j}\}_{i,j\in[n]}$ over n points (which may or may not form a metric), the goal is to find an embedding $\{x_1,\dots,x_n\}\subset\mathbb{R}^k$ that minimizes

$$\mathsf{OPT}=\min_x \mathop{\mathbb{E}}_{i,j\in[n]}\left[\left(1-\frac{\|x_i-x_j\|}{d_{i,j}}\right)^2\right].$$

Despite its popularity, our theoretical understanding of MDS is extremely limited. Recently, Demaine, Hesterberg, Koehler, Lynch, and Urschel Demaine et al. (2021) gave the first approximation algorithm with provable guarantees for this objective, which achieves an embedding in constant dimensional Euclidean space with cost $\mathsf{OPT}+\varepsilon$ in $n^2\cdot 2^{\mathrm{poly}(\Delta/\varepsilon)}$ time, where Δ is the aspect ratio of the input dissimilarities. For metrics that admit low-cost embeddings, the aspect ratio Δ scales polynomially in n. In this work, we give the first approximation algorithm for MDS with quasi-polynomial dependency on Δ: for constant dimensional Euclidean space, we achieve a solution with cost $\tilde{\mathcal{O}}(\log\Delta)\cdot\mathsf{OPT}^{\Omega(1)}+\varepsilon$ in time $n^{\mathcal{O}(1)}\cdot 2^{\mathrm{poly}((\log(\Delta)/\varepsilon))}$. Our algoroithms are based on a novel *geometry-aware* analysis of a conditional rounding of the Sherali-Adams LP Hierarchy, allowing us to avoid exponential dependency on the aspect ratio of the input that would typically result from this rounding.

Keywords: Metric Embeddings, Dimensionality Reduction, Multi-Dimensional Scaling, Sherali-Adams

© 2025 A. Bakshi, V. Cohen-Addad, S.B. Hopkins, R. Jayaram & S. Lattanzi.

1. Introduction

Metric embeddings have long been a central topic in theoretical computer science, and are a foundational tool for understanding and simplifying complex structures in high-dimensional spaces. However, most of the literature on metric embeddings focuses on *bi-Lipschitz* embeddings, which preserve all distances up to a multiplicative distortion factor $L \geqslant 1$. That is, bi-Lipschitz embeddings satisfy $d_{i,j} \leqslant \|x_i - x_j\| \leqslant L \cdot d_{i,j}$ for all pairs i, j. While the theory of bi-Lipschitz embeddings is extremely rich (see Indyk et al. (2017) and references therein), strong computational hardness results are known for the problem of minimizing the distortion L over bi-Lipschitz embeddings of a given metric into $\mathbb{R}^k$ for fixed k Matoušek and Sidiropoulos (2010). Specifically, it is NP-hard approximate the best distortion L for a given n-point metric d to any factor better than $n^{\Theta(1/k)}$, and furthermore, this approximation factor can be obtained in polynomial time by first embedding $\{d_{i,j}\}_{i,j\in[n]}$ into high-dimensional Euclidean space using an approach of Bourgain Bourgain (1985) and then taking a random linear projection. If $\{d_{i,j}\}_{i,j\in[n]}$ is already a (high-dimensional) Euclidean metric, this approach is no better than projection to a random k-dimensional subspace.

Bi-Lipschitz-ness is, therefore, a poor yardstick for measuring the performance of efficient metric embedding algorithms, since a random embedding is essentially as bi-Lipschitz as the best embedding we can hope to obtain in polynomial time. Further, the worst-case distortion between any pair of points is not the only possible measure of the quality of an embedding. In fact, many popular dimensionality reduction objectives in practice focus on alternative embedding objectives, adopting a smoother notion of distortion. In this paper, we demonstrate that one such popular family of embedding objectives admits non-trivial approximation algorithms, making them a more fruitful testbed for algorithm design.

Specifically, we study the *multi-dimensional scaling* (MDS) objective. MDS dates at least to the 1930s Richardson (1938), and was popularized in the 1950s and 1960s psychometrics literature by influential works of Torgerson Torgerson (1952), Shephard Shepard (1962a,b), and Kruskal Kruskal (1964a,b). MDS has since become a bread-and-butter technique in applied statistics – it is the subject of several books Kruskal and Wish (1978); Cox and Cox (2000); Borg and Groenen (2005); Borg et al. (2012) and makes frequent appearances in statistics textbooks Davison and Sireci (2000); de Leeuw and Heiser (1982); Wichern et al.. Heuristic algorithms for MDS are built into several popular programming languages and packages, including Matlab, R, and scikit-learn Mat; De Leeuw and Mair (2011); Pedregosa et al. (2011). A Google Scholar search for "multidimensional scaling" turns up more than half a million results.

Formally, given an n-point metric $\{d_{ij}\}_{i,j\in[n]}$, the Kamada-Kawai cost-function for MDS, henceforth referred to as *KK*, is defined as follows:

$$\mathsf{OPT} = \min_{x_1,\ldots,x_n \in \mathbb{R}^k} \mathop{\mathbb{E}}_{i,j\sim[n]} \left(1 - \frac{\|x_i - x_j\|}{d_{i,j}}\right)^2 . \qquad \text{(Kamada-Kawai (KK))}$$

Here, $\mathbb{E}_{i,j\sim[n]}$ denotes a uniformly random pair i, j drawn from $[n] \times [n]$. KK measures the mean-squared distance between 1 and the multiplicative distortion $\|x_i - x_j\| / d_{i,j}$ experienced by the pair i, j. For normalization, note that placing all x_i at the same point gives objective value 1, so for any input $\{d_{ij}\}_{i,j\in[n]}$, the optimum value OPT is in $[0, 1]$.

KK is especially popular as a tool for force-directed graph drawing in the social and medical sciences. For instance, the KK objective has been used to visualize the spread of obesity and the dynamics of smoking in social networks Christakis and Fowler (2007, 2008), for mapping the structural core of the human cerebral cortex Hagmann et al. (2008) and identifying intellectual turning points in theoretical physics Chen (2004). In practice, the embedding $x_1, \ldots, x_n$ is typically obtained via gradient descent or other local search methods applied to the objective function. These objective functions are nonconvex, and we are aware of no provable guarantees for such local search procedures. We are aware of few works giving algorithms with provable guarantees for any MDS cost function, including KK. Before we turn to these and to our results, we contrast MDS with some familiar metric embedding problems studied in theoretical computer science.

MDS versus Principal Component Analysis. Principal component analysis (PCA) is a well-studied approach to dimension reduction of a high-dimensional Euclidean metric. It can be phrased via a similar-looking optimization problem as KK, applied to inner products rather than distances. If $y_1, \ldots, y_n \in \mathbb{R}^N$, then

$$\mathsf{PCA} = \min_{x_1,\ldots,x_n \in \mathbb{R}^k} \mathop{\mathbb{E}}_{i,j \in [n]} (\langle y_i, y_j \rangle - \langle x_i, x_j \rangle)^2 .$$

It is well known that the optimizing $x_1, \ldots, x_n$ are given by $\Pi_k y_1, \ldots, \Pi_k y_n$, where Π_k is the projector to the top k eigenvectors of the covariance matrix of $y_1, \ldots, y_n$. In particular, the optimal embedding is a linear projection of $y_1, \ldots, y_n$.

Since inner products are related to distances via $\langle y_i, y_j \rangle = \frac{1}{2}(\|y_i - y_j\|^2 - \|y_i\|^2 - \|y_j\|^2)$, is it tempting to imagine that the least-squares optimization problems defining PCA and MDS are equivalent or close to it. Indeed, a popular MDS heuristic, sometimes called *classical multidimensional scaling*, applies PCA to the matrix $(I - \frac{1}{n}11^\top)D^{(2)}(I - \frac{1}{n}11^\top)$, where $D^{(2)}$ is the entrywise square of the distance matrix D and 1 denotes the all-1s vector – if $\{d_{ij}\}_{i,j \in [n]}$ is a Euclidean metric $y_1, \ldots, y_n \in \mathbb{R}^N$, then this is equivalent to applying PCA to $y_1, \ldots, y_n$.

To the best of our knowledge, no provable approximation guarantees for KK (or any other MDS cost function) are known for this algorithm. In general, we do not expect the PCA-optimal embedding to be related to the KK-optimal embedding, and unlike PCA, the KK-optimal embedding can be a *nonlinear* mapping of $y_1, \ldots, y_n$ into $\mathbb{R}^k$. To demonstrate this power, we construct an explicit example where the KK-optimal embedding recovers interesting cluster structure which is lost by PCA in Section 11 of the supplementary. Moreover, in this example the solution produced by PCA has an arbitrarily worse MDS cost than the optimal MDS solution, despite being close to low-dimensional initially. Thus, even for pointsets that have a small optimal MDS cost to begin with, the quality resulting from using PCA as an algorithm for the MDS objective can be arbitrarily poor.

State of the Art: Large Embedding Dimension or Exponential Dependence on Aspect Ratio. Having argued the importance of designing algorithms with provable guaranteeds for MDS both because of its widespread use in practice and as a new approach to the algorithmic theory of metric embeddings, we turn to the algorithmic state of the art.

We are aware of only two works giving algorithms with provable approximation guarantees for MDS. Bartal, Fandina, and Neiman Bartal et al. (2019) observe that many MDS objectives become convex (semidefinite) programs if we ask to embed in n dimensions – this is true in particular for OPT. A simple Johnson-Lindenstrauss-style analysis shows that randomly projecting an n-dimensional KK solution, obtained via semidefinite programming, to k dimensions incurs an additive $O(1/k)$ loss in objective value. Thus, we can obtain a solution with cost $\mathsf{OPT} + O(1/k)$ in polynomial time. But, as we have already discussed, for small k (e.g. $k = 1, 2, 3$), random projection yields poor embeddings, and consequently this algorithm lacks nontrivial guarantees for such small k.

Demaine, Hesterberg, Koehler, Lynch, and Urschel give an algorithm that retains provable guarantees for small k, at the expense of much slower running time Demaine et al. (2021). They associate an important parameter to the input metric: Let

$$\Delta = \frac{\max_{i,j \in [n]} d_{ij}}{\min_{i,j \in [n]} d_{ij}}$$

measure the *aspect ratio* of the metric. Demaine et al. (2021) obtains an efficient algorithm when the aspect ratio (Δ) is small. Their result can be summarized as follows:

Theorem 1 (Approximation schemes scaling exponentially in aspect ratio Demaine et al. (2021)) *For every $k > 0$ and $p \geqslant 0$ there is an algorithm with running time $n^2 \cdot \exp((\Delta/\varepsilon)^{O(1)})$ which outputs an embedding with KK cost $\mathsf{OPT} + \varepsilon$.*

How should we think of the aspect ratio parameter Δ? If $\{d_{i,j}\}_{i,j \in [n]}$ represent, say, a 1-dimensional Euclidean metric, then $\Delta = \Omega(n)$. In general, for metrics with good low-dimensional embeddings, Demaine et al. (2021)'s algorithm has exponential running time, suggesting the question:

Is there an approximation algorithm for Kamada-Kawai which is efficient on inputs with large aspect ratio?

We note that inputs with large aspect ratio are of particular interest since they include instances that admit small-objective-value embeddings in low dimensions.

1.1. Our Results

We give an almost-exponential improvement on the running time compared to Demaine et al. (2021) with respect to Δ, at the cost of a somewhat worse approximation factor. As far as we know, we give the first nontrivial approximation algorithm for MDS which remains polynomial-time for super-poly-logarithmic values of Δ. Specifically, for every $k \geqslant 1$ we design an algorithm that runs in time$n^{O(1)} \cdot \exp((\log \Delta/\varepsilon)^{O(1)})$ and which outputs an embeddding with KK cost

$$\tilde{O}(\log \Delta) \cdot (\mathsf{OPT})^{\Omega(1)} + \varepsilon\,.$$

We now state our main theorem.

Theorem 2 (Main Theorem) *Given an instance of MDS under the Kamada-Kawai objective, with aspect ratio $1 \leqslant \Delta$, target dimension $k \in \mathbb{N}$, and target accuracy $0 < \varepsilon < 1$, there exists an algorithm that outputs an embedding $\{\hat{x}_i\}_{i \in [n]}$ such that with probability at least $99/100$,*

$$\mathop{\mathbb{E}}_{i,j}\left[\left(1 - \frac{\|\hat{x}_i - \hat{x}_j\|_2}{d_{i,j}}\right)^2\right] \leqslant \begin{cases} \mathcal{O}\Big(\sqrt{\mathsf{OPT}} \cdot \log(\Delta/\varepsilon)\Big) + \varepsilon & \text{if } k = 1 \\ \mathcal{O}\Big(\sqrt{\mathsf{OPT}} \cdot \log(1/\varepsilon) \cdot \log(\Delta/\varepsilon)\Big) + \varepsilon & \text{if } k = 2 \\ \mathcal{O}\Big(\mathsf{OPT}^{1/k} \cdot \log(\Delta/\varepsilon)\Big) + \varepsilon & \text{otherwise.} \end{cases}$$

Further, the algorithm runs in $(n\Delta/\varepsilon)^{\mathcal{O}(k\cdot\tau)}$ time, where

$$\tau = \begin{cases} \mathcal{O}((\log(\log(\Delta)/\varepsilon)\log(\Delta/\varepsilon))/\varepsilon) & \text{if } k = 1 \\ \mathcal{O}\Big((\log^2(\log(\Delta)/\varepsilon)\log(\Delta/\varepsilon))/\varepsilon\Big) & \text{if } k = 2 \\ \mathcal{O}\big((k\log(\log(\Delta)/\varepsilon)\log(\Delta/\varepsilon)^{k/2})/\varepsilon^{k/2}\big) & \text{otherwise.} \end{cases}$$

We refer the reader to Remark 3 below for discussion of the dependence of the hidden constants on k. Our result does not actually require that $\{d_{i,j}\}_{i,j\in[n]}$ form a metric – it is enough for each to be a non-negative number in $[1,\Delta]$ (this is also true for Theorem 1).

Remark 3 *See the supplementary material for a slight strengthening of Theorem 2 where we sparsify the linear program to obtain a* $\mathrm{poly}(n)$ *running time for constant* Δ, k, ε.

Also note: we have optimized the choices of parameters in Theorem 2 to maximize the exponent of OPT *in our approximation guarantee. Different tradeoffs among the parameters are possible. As one example, if we follow an argument in line with technical overview, we can obtain a running time with* Δ*-dependence of the form* $\exp((\log\Delta)^{O(1)}/\varepsilon^{O(k)})$ *rather than* $\exp(((\log\Delta)/\varepsilon))^{O(k)})$*, at the cost that instead of* $\mathsf{OPT}^{1/k}$ *we get* $\mathsf{OPT}^{\Omega(1/k)}$.

Remark 4 (Comparison with Demaine et al. (2021)) *Consider the illustrative toy example where the input distances correspond to a small perturbation of the line metric on* n *points. Here, the* $\{d_{i,j}\}_{i,j\in[n]}$ *correspond to scalars* $\{x_i^*\}_{i\in[n]}$ *such that* $d_{i,j} = (1\pm\zeta)|x_i^* - x_j^*|$*, where the perturbation is bounded, i.e.* $|\zeta| \leqslant 1/\mathrm{poly}(n)$*. We note that such instances are of particular interest, since they are precisely the ones that admit good low-dimensional embeddings.*

It is easy to see that $\Delta = \Omega(n)$ *and when the target dimension* $k = 1$*, the KK objective has value* $\mathsf{OPT} = 1/\mathrm{poly}(n)$*. Then, given an* $1/\mathrm{poly}(n) < \varepsilon < 1$*, Demaine et al. (2021) runs in time* $2^{\mathrm{poly}(n/\varepsilon)}$ *and outputs an embedding with cost* ε*. In contrast, our algorithm runs in time* $n^{\mathrm{poly}(\log(n),1/\varepsilon)}$ *and also ouputs an embedding with cost* ε.

Our approach is to round a Sherali-Adams linear programming relaxation of the KK objective function.[1] We use $((\log\Delta)/\varepsilon)^{O(k)}$ levels of the hierarchy, leading to an LP with $\approx n^{((\log\Delta)/\varepsilon)^{O(k)}}$ variables and constraints. (We can sparsify this LP to obtain our final running time.)

The conditioning-based rounding scheme we consider is a natural one – it goes back at least to the *global correlation rounding* algorithm of Barak et al. (2011) for dense constraint satisfaction problems. In fact, a naïve analysis of the same rounding scheme, applied instead to $(\Delta/\varepsilon)^{O(1)}$ levels of Sherali-Adams, is one way to obtain the result in Demaine et al. (2021). In their proof of Theorem 1, the authors of Demaine et al. (2021) take a different perspective, directly reducing MDS to a dense constraint satisfaction problem.

Our main contribution is a new *geometry-aware* analysis of conditioning-based Sherali-Adams rounding, which we describe in Section 2. More broadly, we believe that new approaches to analyzing convex relaxations are a promising avenue for new metric embedding algorithms. We outline a number of open problems in this direction in Section 3.

1.2. Related Work

Multi-dimensional scaling has been studied since at least the 1950s, with too vast a literature to survey properly here. We refer the reader to the survey Young (2013) and book Cox and Cox (2000) for review of its history and applications. See also the recent book Agrawal et al. (2021) for a perspective from applied optimization. We discuss here some related works from the theoretical computer science literature.

1. We could also use a more powerful Sum-of-Squares semidefinite programming relaxation, with no change to asymptotic running times or analysis, but do not require it.

t-SNE and Friends. Another popular non-linear dimension reduction technique/objective function, closely related to MDS, is t-SNE Van der Maaten and Hinton (2008); Hinton and Roweis (2002). Notably Arora, Hu, and Kothari Arora et al. (2018) showed that the gradient-descent algorithm used in practice for t-SNE performs well if the input exhibits some clusterability properties. Other non-linear dimension reduction techniques used in practice include the Isomap embedding Balasubramanian and Schwartz (2002), or spectral embedding methods Von Luxburg (2007).

Approximation Algorithms via LP Hierarchies. There is a vast literature on approximation algorithms using LP hierarchies. Notable examples include works on scheduling with communication delay Davies et al. (2020), the matching polytope Mathieu and Sinclair (2009), numerous works on CSPs and Dense-CSPs Barak et al. (2011); Guruswami and Sinop (2012), Bin Packing Karlin et al. (2011), and Correlation Clustering Cohen-Addad et al. (2022).

Approximation Algorithms for Metric Embeddings with Additive or Multiplicative Distortion Metric embeddings have been studied extensively for 30+ years in theoretical computer science, with extensive applications. See e.g. references in the recent paper Sidiropoulos et al. (2017). Strong inapproximability results Matousek and Sidiropoulos (2008) for minimizing bi-Lipschitz distortion led to some efforts to design approximation algorithms for relaxed measures of distortion, among them MDS. We are not aware of works in this vein whose techniques apply directly to the Kamada-Kawai problem we consider here, so we defer further discussion to Section 10 of the supplementary.

2. Technical Overview

We now overview our techniques. We adopt the convention $1 \leqslant d_{i,j} \leqslant \Delta$ for all i, j. This overview is meant to aid the reader in understanding the main ideas in our arguments, so the quantitative claims often ignore constant or logarithmic factors. See the supplementary material for rigorous arguments.

2.1. Our Algorithm

Discretization. We begin by discretizing the KK problem. This makes it easier to formulate the Sherali-Adams LP hierarchy for KK. Specifically, we show that for any k and input $\{d_{i,j}\}_{i,j\in[n]}$, the optimal embedding among those which place $x_1, \ldots, x_n$ at distinct points in a $(\Delta/\varepsilon)^{O(k)}$-size net of $\mathbb{R}^k$ has cost at most ε more than the optimal embedding into $\mathbb{R}^k$. In particular, this requires us to demonstrate that there is an ε-approximate optimal embedding which has aspect ratio at most $O(\Delta/\varepsilon)$; we do this by proving that any optimal embedding can be projected onto a ball of radius $O(\Delta/\varepsilon)$ without significantly increasing the cost. Since our approximation guarantees anyway lose an additive ε, we can restrict attention to embeddings $x_1, \ldots, x_n$ into a discrete subset of $\mathbb{R}^k$ with $(\Delta/\varepsilon)^{O(k)}$ points. In this overview we will not distinguish carefully between discretized and non-discretized versions of OPT.

Sherali-Adams Hierarchy. For an integer $t \geqslant 1$, the level-t Sherali-Adams linear programming (LP) relaxation for (discretized) KK is an LP with $(n\Delta/\varepsilon)^{O(kt)}$ variables and constraints, whose solutions are *Sherali-Adams pseudoexpectations* (or "pseudoexpectations" for short), denoted $\tilde{\mathbb{E}}$.

A level-t pseudoexpectation is a collection of $\binom{n}{t}$ "local distributions", one for each $S \subseteq [n]$ of size $|S| = t$. The local distribution μ_S is a joint probability distribution over assignments of $\{x_i\}_{i\in S}$ to elements of $\mathbb{R}^k$. The constraints of the Sherali-Adams LP ensure that local distributions

μ_S, μ_T have the same marginal distribution on $S \cap T$. For purposes of designing a rounding scheme for the LP, we often pretend that $\tilde{\mathbb{E}}$ defines an honest-to-goodness joint probability distribution over assignments of $x_1, \ldots, x_n$ to $\mathbb{R}^k$, although we can only look at this distribution "locally". This mentality dates to Barak et al. (2011); Raghavendra and Tan (2012) and has since been used extensively to design algorithms by rounding Sherali-Adams LPs and Sum-of-Squares semidefinite programs.

Pseudoexpectations and variance: For any real-valued function f depending on a subset S of at most t out of $x_1, \ldots, x_n$, f has a well-defined "pseudo-expected" value, denoted $\tilde{\mathbb{E}} f = \mathbb{E}_{x_S \sim \mu_S} f(x_S)$. We can extend $\tilde{\mathbb{E}}$ to be a linear operator which assigns a real value to any linear combination of such t-local functions. We can similarly measure the variance of any t-local function via $\tilde{\mathbb{E}}(f - \tilde{\mathbb{E}} f)^2$, since $(f - \tilde{\mathbb{E}} f)^2$ is itself a t-local function.

Conditioning: For any level-t pseudoexpectation $\tilde{\mathbb{E}}$, index $i \in [n]$, and $z \in \mathbb{R}^k$ such that $\mathbf{Pr}(x_i = z) > 0$,[2] we can define a level-$(t-1)$ conditional pseudoexpectation $\tilde{\mathbb{E}}[\cdot \mid x_i = z]$ by defining μ_S to be $\mu_{S \cup \{x_i\}}$ conditioned on $x_i = z$.

Rounding by Conditioning. Our algorithm first solves the Sherali-Adams LP to find a level t pseudoexpectation minimizing

$$\min_{\tilde{\mathbb{E}}} \tilde{\mathbb{E}} \mathop{\mathbb{E}}_{i,j} \left(1 - \frac{\|x_i - x_j\|}{d_{i,j}}\right)^2 \leqslant \mathsf{OPT}\,, \tag{1}$$

where the inequality holds because the LP is a relaxation of the minimization problem defining OPT. We ignore some technicalities in this overview; in particular, our purpose here we will set $t \approx (\log \Delta)/\varepsilon^{O(k)}$. Later, in the full analysis in the supplementary, with the goal of optimizing the hidden constants in the approximation guarantee, we will actually set $t = ((\log \Delta)/\varepsilon)^{O(k)}$, although noting that the approach given in this overview is valid but would give a different runtime-approximation trade-off — see the discussion after Theorem 2.

To round $\tilde{\mathbb{E}}$ we choose a random set $S = \{i_1, \ldots, i_{t-2}\} \subset [n]$ of size $|S| = t - 2$, where each of $i_1, \ldots, i_{t-2}$ is drawn uniformly and independently at random. Then, we draw a sample $z_S \sim \mu_S$ from the local distribution for S and output the expected locations of each x_i conditioned on $x_S = z_S$; i.e., we embed the i-th point at $\tilde{\mathbb{E}}[x_i \mid x_S = z_S] \in \mathbb{R}^k$.

Efficient Kamada-Kawai (informal)

Input: Nonnegative numbers $\mathcal{D} = \{d_{i,j}\}_{i,j \in [n]}$, target dimension $k \in \mathbb{N}$, target accuracy $0 < \varepsilon < 1$.

Operation:

1. Let $t = (\log \Delta)/\varepsilon^{O(k)}$. Let $\tilde{\mathbb{E}}$ be a degree t Sherali-Adams pseudoexpectation over a $(\Delta/\varepsilon)^{O(k)}$ size net of $\mathbb{R}^k$, optimizing

$$\min_{\tilde{\mathbb{E}}} \ \tilde{\mathbb{E}} \mathop{\mathbb{E}}_{i,j \sim [n]} \left(1 - \frac{\|x_i - x_j\|}{d_{ij}}\right)^2 .$$

2. We can define this probability via $\tilde{\mathbb{E}}\, 1_{x_i = z}$.

2. Let $S \subset [n]$ be a random subset of size $t-2$ and let z_S be a draw from the local distribution μ_S induced by $\tilde{\mathbb{E}}$.

Output: The embedding $\left\{\tilde{\mathbb{E}}\left[x_i \mid x_S = z_S\right]\right\}_{i\in[n]}$.

2.2. Our Analysis

Our goal is now to bound the expected objective value of the rounded solution

$$\mathop{\mathbb{E}}_{S} \mathop{\mathbb{E}}_{z_S\sim\mu_S} \mathop{\mathbb{E}}_{i,j} \left(1 - \frac{\|\tilde{\mathbb{E}}[x_i \mid x_S = z_S] - \tilde{\mathbb{E}}[x_j \mid x_S = z_S]\|}{d_{i,j}}\right)^2$$

in terms of the LP value (1) and $\varepsilon > 0$.

To get a sense of the challenge we face, let us see what could go wrong if we output $\tilde{\mathbb{E}}\, x_i$ as the embedded location of point i, without first doing the conditioning step. Even if OPT is 0, meaning that $\{d_{ij}\}_{i,j\in[n]}$ has a perfect embedding into $\mathbb{R}^k$, the pseudoexpectation $\tilde{\mathbb{E}}$ could have $\tilde{\mathbb{E}}\, x_i = 0$ for all i, yielding the worst-possible rounded cost of 1. For instance, this could arise by first taking a perfect embedding $y_1, \ldots, y_n \in \mathbb{R}^k \setminus \{0\}$ of the input (i.e. $\|y_i - y_j\|_2 = d_{ij}$ for all i, j), and obtaining a probability distribution over embeddings which takes the values $(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$ with probability $1/2$ and $(x_1, \ldots, x_n) = (-y_1, \ldots, -y_n)$ with probability $1/2$.

Let us see that conditioning addresses the issue in the context of this example. If we sample, say, $z_1 = y_1$ and condition on $x_1 = z_1$, the conditional distribution of x_i places all its mass y_i, giving the perfect embedding we are looking for.

In general of course we cannot assume that $\mathsf{OPT} = 0$ or that the pseudoexpectation $\tilde{\mathbb{E}}$ obtained via (1) is an actual distribution over embeddings, much less one obtained by taking a single fixed embedding and randomly negating it. Instead, we develop a more robust version of the above argument. Although after conditioning on $x_S = z_S$ the distributions of the remaining x_i will not be supported on a single point as they were in the above example, we will be able to show that for most i the conditional distribution has small variance. Next we show that such a variance bound is enough to bound the rounded cost.

Small Variance Implies Small Rounded Cost. Consider the contribution of a pair (i, j) to the rounded objective. Suppose we are lucky, in the following sense: after conditioning on $x_S = z_S$, the variance of both x_i and x_j is small compared to $d_{i,j}^2$. That is, suppose:

$$\mathop{\mathbb{E}}_{z_S\sim\mu_S} \tilde{\mathbb{E}}[\|x_i - \tilde{\mathbb{E}}[x_i \mid x_S = z_S]\|^2 \mid x_S = z_S], \mathop{\mathbb{E}}_{z_S\sim\mu_S} \tilde{\mathbb{E}}[\|x_j - \tilde{\mathbb{E}}[x_j \mid x_S = z_S]\|^2 \mid x_S = z_S] \leqslant \varepsilon \cdot d_{i,j}^2 . \tag{2}$$

Now, for any jointly-distributed $\mathbb{R}^k$-valued random variables X, Y, we have $\|\mathbb{E}\, X - \mathbb{E}\, Y\| = \mathbb{E}\, \|X - Y\| \pm (\mathbb{E}\, \|X - \mathbb{E}\, X\| + \mathbb{E}\, \|Y - \mathbb{E}\, Y\|)$. Applying this fact with (X, Y) being the joint distribution of (x_i, x_j) conditioned on $z_S = x_S$, we can bound the contribution of the pair (i, j) to the rounded cost:

$$\mathop{\mathbb{E}}_{z_S\sim\mu_S} \left(1 - \frac{\|\tilde{\mathbb{E}}[x_i \mid x_S = z_S] - \tilde{\mathbb{E}}[x_j \mid x_S = z_S]\|}{d_{i,j}}\right)^2$$

$$\leqslant 2 \mathop{\mathbb{E}}_{z_S \sim \mu_S} \left(1 - \frac{\tilde{\mathbb{E}}[\|x_i - x_j\| \mid x_S = z_S]}{d_{i,j}}\right)^2 + 4\varepsilon$$
$$\leqslant 2 \mathop{\mathbb{E}}_{z_S \sim \mu_S} \tilde{\mathbb{E}} \left[\left(1 - \frac{\|x_i - x_j\|}{d_{i,j}}\right)^2 \mid x_S = z_S\right] + 4\varepsilon \quad \text{(Jensen's)}\,.$$

By the law of total expectation, this is (twice) the contribution of the pair (i, j) to the objective value of the LP, plus a negligible 4ε. Since the objective value of the LP is in total at most OPT, if we were lucky like this for every pair (i, j) we would obtain the bound $O(\mathsf{OPT} + \varepsilon)$ on the rounded cost.

As an aside, the "global correlation rounding" analysis of Barak et al. (2011); Raghavendra and Tan (2012) can easily be adapted to show that the above variance bound holds for most (i, j) pairs if we take $|S| = (\Delta/\varepsilon)^{O(1)}$. This would recover the result of Demaine et al. (2021).

We aim to take $|S| \approx (\log \Delta)/\varepsilon^{O(k)}$, in which case we will show that this variance bound holds for all but a roughly $(\varepsilon + |S|\mathsf{OPT})$-fraction of pairs. To this end, we introduce a "geometry-aware" strategy to obtain a (weakened version of) (2). (We will eventually be able to argue that the small fraction of pairs where this variance bound fails do not contribute much to the rounded cost.)

Variances Are as Small as the Distance to the Closest Conditioned Point. This brings us to the question: for a given (i, j) pair, how big are the variances $\tilde{\mathbb{E}}\, \|x_i - \tilde{\mathbb{E}}\, x_i\|^2, \tilde{\mathbb{E}}\, \|x_j - \tilde{\mathbb{E}}\, x_j\|^2$ after we condition on $x_S = z_S$? Consider any jointly-distributed $\mathbb{R}^k$-valued random variables (X, Y) – now think of X as one of the x_i's, and Y as some x_ℓ for $\ell \in S$. Simple probability arguments show that when we fix a value y for Y and condition on $Y = y$, the typical conditional variance of X is bounded by the typical square distance from X to Y. That is, conditioning on Y "localizes" X to a ball around Y of squared radius $\mathbb{E}\, \|X - Y\|^2$. Formally,

$$\mathop{\mathbb{E}}_{y \sim Y} (\mathbb{E}[\|X - \mathbb{E}[X \mid Y = y]\| \mid Y = y]) \leqslant \mathbb{E}\, \|X - Y\|^2\,.$$

For a typical pair (i, j) we want to localize x_i and x_j each to a ball of squared radius $\varepsilon \cdot d_{i,j}^2$, as in (2). This localization happens if S contains i', j' such that $\tilde{\mathbb{E}}\, \|x_i - x_{i'}\|^2 \leqslant \varepsilon \cdot d_{i,j}^2$ and similarly $\tilde{\mathbb{E}}\, \|x_j - x_{j'}\|^2 \leqslant \varepsilon \cdot d_{i,j}^2$. (Conditioning on the remaining $S \setminus \{i', j'\}$ can only reduce the expected variances further.) To see why we should hope that S will contain such i', j', we need to take a detour to obtain some control over $\tilde{\mathbb{E}}\, \|x_i - x_{i'}\|^2$ for $i' \in S$.

Most Pairwise Distances are Close to Euclidean. We will argue that for most (a, b) pairs, $\tilde{\mathbb{E}}\, \|x_a - x_b\|^2$ is close to the distance $\|x_a^* - x_b^*\|^2$, where x_a^* and x_b^* are members of some actual n-point k-dimensional Euclidean metric.

First, we argue that for most (a, b) pairs, $\tilde{\mathbb{E}}\, \|x_a - x_b\|^2 \approx d_{a,b}^2$. Indeed, a simple Markov inequality applied to the objective function shows this. We have $\mathbb{E}_{a,b}\, \tilde{\mathbb{E}}(1 - \|x_a - x_b\|/d_{a,b})^2 \leqslant \mathsf{OPT}$. Hence, with probability at least $1 - O(\mathsf{OPT})$ over a random pair $a, b \in [n]$, we have $\tilde{\mathbb{E}}(1 - \|x_a - x_b\|/d_{a,b})^2 \leqslant 0.01$. For any such a, b pair, $\tilde{\mathbb{E}}\, \|x_a - x_b\|^2 = (1 \pm 0.1) \cdot d_{a,b}^2$.

We have related $\tilde{\mathbb{E}}\, \|x_a - x_b\|^2$ to the input metric distances $d_{a,b}$. Now we want to relate the input metric distances $d_{a,b}$ to distances in an actual low-dimensional Euclidean metric. Let $x_1^*, \ldots, x_n^* \in \mathbb{R}^k$ be an optimal KK solution, achieving objective cost OPT. By the same Markov argument as above, with probability $1 - O(\mathsf{OPT})$ over a, b, we have $\|x_a^* - x_b^*\|^2 = (1 \pm 0.1) \cdot d_{a,b}^2$. Putting these together, again with probability $1 - O(\mathsf{OPT})$ over a, b, we have $\tilde{\mathbb{E}}\, \|x_a - x_b\|^2 = (1 \pm 0.2)\|x_a^* -$

$x_b^*\|^2$. In conclusion, if OPT is small, then for most pairs a, b, the pseudoexpected (square) distances agree with those of some actual n-point k-dimensional Euclidean metric, up to a multiplicative constant.

x^*-Distance to Closest Conditioned Point Is Small. Let us return to the mission of showing that for a typical pair i, j, the random set S contains i', j' such that $\tilde{\mathbb{E}}\, \|x_i - x_{i'}\|^2 \leqslant \varepsilon \cdot d_{i,j}^2$ and similarly for j, j'. Since we have showed that most pairwise distances satisfy $\tilde{\mathbb{E}}\, \|x_a - x_b\|^2 \approx d_{a,b}^2 \approx \|x_a^* - x_b^*\|^2$ for some k-dimsional metric $x_1^*, \dots, x_n^*$, it will help to first show an analogous statement where $\tilde{\mathbb{E}}\, \|x_i - x_{i'}\|^2$ and $d_{i,j}^2$ are replaced with $\|x_i^* - x_{i'}^*\|^2$ and $\|x_i^* - x_j^*\|^2$, respectively.

Consider a simple example. Suppose $k = 1$ and $x_1^*, \dots, x_n^*$ are the uniformly-spaced metric where $x_i^* = i$. In this example, we can see that a random pair i, j will have distance $\|x_i^* - x_j^*\| = \Omega(n)$.

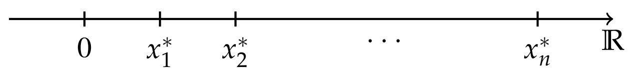

Figure 1: Example – uniformly-spaced metric $x_1*, \dots, x_n^*$

Fixing i and taking S to contain t random indices from $[n]$, we also have

$$\mathop{\mathbb{E}}_{S} \min_{i' \in S} \|x_i^* - x_{i'}^*\| \approx \frac{n}{t} \approx \frac{\|x_i^* - x_j^*\|}{t}\,,$$

since around a $1/t$-fraction of the points $x_{i'}^*$ are at distance $\approx n/t$ to x_i. If $k > 1$ but we maintain the “uniform spacing” assumption, we would similarly obtain $\mathbb{E}_S \min_{i' \in S} \|x_i^* - x_{i'}^*\| \approx \frac{n^{1/k}}{t^{1/k}} \approx \frac{\|x_i^* - x_j^*\|}{t^{1/k}}$. So, if we take $t = 1/\varepsilon^{O(k)}$, then at least for a typical pair (i, j), we expect to have $\min_{i' \in S} \|x_i^* - x_{i'}^*\|^2 \leqslant \varepsilon \cdot d_{i,j}^2$.

While the reasoning above seems specific to the uniformly-spaced metric, we can actually prove the following version of it which applies to a general k-dimensional discrete metric, and also quantifies the fraction of i, j pairs for which the inequality holds:

Lemma 5 (Key lemma on k-dimensional discrete metrics) *Let $x_1^*, \dots, x_n^* \in \mathbb{R}^k$ be a discrete metric with aspect ratio Δ. For every $\delta > 0$, with probability at least $1 - O(\delta \log(\Delta))$ over $i, j \sim [n]$ chosen uniformly at random, with probability at least $1 - \delta$ over $S \subseteq [n], |S| = t$ is uniformly random,*

$$\min_{a \in S} \|x_i^* - x_a^*\|^2 + \min_{a \in S} \|x_j^* - x_a^*\|^2 \leqslant \tilde{O}\left(\frac{1}{\delta t}\right)^{1/k} \cdot \|x_i^* - x_j^*\|^2\,.$$

We prove Lemma 5 by splitting all the pairs (i, j) according to $\log \Delta$ geometrically-increasing scales, and arguing separately at each scale. The resulting dependence on $\log \Delta$ in Lemma 5 is the source of the dependence of our approximation guarantees on $\log \Delta$. This dependence is tight in the sense that it cannot be removed from the statement of Lemma 5, as we show in Example 1 below. In the rest of this overview we will ignore the failure probability over S in Lemma 5.

Example 1 *Consider the case $k = 1$ where we let $n = \log \Delta$ and $x_i^* = 2^i$. Then for every pair $i < j$, we have $2^{j-1} \leqslant \|x_i^* - x_j^*\| \leqslant 2^j$, thus there does not exist even a single point x_k^* such that $\|x_j^* - x_k^*\| \leqslant (1/3) \cdot \|x_i^* - x_j^*\|$.*

From x^* Distances Back to Variances. Our goal, as before, is to show that for most i, j, $\mathbb{E}_S \min_{a\in S} \tilde{\mathbb{E}} \|x_i - x_a\|^2 \leqslant \varepsilon \cdot d_{i,j}^2$ and similarly for j. (We still have to do something about the rounded cost of the small set of (i, j) pairs for which this fails.) Our progress so far is: there is k-dimensional Euclidean metric $x_1^*, \ldots, x_n^*$ such that for random i, j,

1. with probability $1 - O(\mathsf{OPT})$, $\tilde{\mathbb{E}} \|x_i - x_j\|^2 = (1 \pm 0.1) \cdot \|x_i^* - x_j^*\|^2 = (1 \pm 0.1) \cdot d_{i,j}^2$, and

2. with probability $1 - \varepsilon$, $\min_{a\in S} \|x_i^* - x_a^*\|^2 + \min_{a\in S} \|x_j^* - x_a^*\|^2 \leqslant \varepsilon \cdot \|x_i^* - x_j^*\|^2$, if we take $t \approx \frac{\log \Delta}{\varepsilon^{O(k)}}$.

We would like to use 1 to replace $\|x_i^* - x_a^*\|^2, \|x_j^* - x_a^*\|^2, \|x_i^* - x_j^*\|^2$ in 2 with $\tilde{\mathbb{E}} \|x_i - x_a\|^2, \tilde{\mathbb{E}} \|x_j - x_a\|^2, \tilde{\mathbb{E}} \|x_i - x_j\|^2$, respectively, perhaps at the cost of weakening the inequality by a multiplicative constant. We can do this via a simple counting argument. Because S is random, every i for which $\min_{a\in S} \|x_i^* - x_a^*\|^2 \gg \min_{a\in S} \tilde{\mathbb{E}} \|x_i - x_a\|^2$ witnesses around n/t pairs i, a which fail the inequality in (1). So, with probability roughly $1 - O(t \cdot \mathsf{OPT})$ over a randomly chosen i, we will have

$$\min_{a\in S} \tilde{\mathbb{E}} \|x_i - x_a\|^2 \leqslant O(1) \cdot \min_{a\in S} \|x_i^* - x_a^*\|^2 .$$

Putting things together, if we draw a random pair i, j, with probability at least $1 - O(t \cdot \mathsf{OPT} + \varepsilon)$, we will have

$$\mathop{\mathbb{E}}_{z_S\sim\mu_S} \tilde{\mathbb{E}}[\|x_i - \tilde{\mathbb{E}}[x_i \mid x_S = z_S]\|^2 \mid x_S = z_S] + \mathop{\mathbb{E}}_{z_S\sim\mu_S} \tilde{\mathbb{E}}[\|x_j - \tilde{\mathbb{E}}[x_j \mid x_S = z_S]\|^2 \mid x_S = z_S] \leqslant \varepsilon \cdot d_{i,j}^2 , \tag{3}$$

giving us the inequality (2) that we have been looking for. The final step is to address the rounded cost of the "bad" i, j pairs, occurring with probability at most $O(t \cdot \mathsf{OPT} + \varepsilon)$.

Small Sets of Pairs Don't Contribute Much to Rounded Cost. Suppose that $B \subseteq [n] \times [n]$. We show that the contribution of B to the rounded objective cost is (in expectation) $O(\mathsf{OPT} + \mathbf{Pr}_{(i,j)\sim\phi}((i, j) \in B))$, giving us a bound on the contribution to the rounded cost from bad pairs.

Fix a pair (i, j), a subset $S \subseteq [n]$, $|S| = t - 2$, and an assignment z_S for x_S. The contribution of (i, j) to the rounded cost obeys the bound

$$\begin{aligned}\left(1 - \frac{\|\tilde{\mathbb{E}}[x_i \mid x_S = z_S] - \tilde{\mathbb{E}}[x_j \mid x_S = z_S]\|}{d_{i,j}}\right)^2 &\leqslant 1 + \frac{\|\tilde{\mathbb{E}}[x_i \mid x_S = z_S] - \tilde{\mathbb{E}}[x_j \mid x_S = z_S]\|^2}{d_{i,j}^2} \\ &\leqslant 1 + \left(\tilde{\mathbb{E}}\left[\frac{\|x_i - x_j\|}{d_{i,j}} \mid z_S = x_S\right]^2\right) \quad \text{(Jensen's)} \\ &\leqslant 3 + 2 \cdot \tilde{\mathbb{E}}\left[\left(1 - \frac{\|x_i - x_j\|}{d_{i,j}}\right)^2 \mid z_S = x_S\right] ,\end{aligned} \tag{4}$$

where for the last step we added and subtracted 1 from the expression inside the square. If we take the expectation of (4) over $z_S \sim \mu_S$, we get $O(1) +$ (contribution of i, j to LP objective). So,

$$\mathop{\mathbb{E}}_{i,j} \mathop{\mathbb{E}}_{z_S\sim\mu_S} \left(1 - \frac{\|\tilde{\mathbb{E}}[x_i \mid x_S = z_S] - \tilde{\mathbb{E}}[x_j \mid x_S = z_S]\|}{d_{i,j}}\right)^2 \cdot \mathbf{1}_{(i,j)\in B} \leqslant O(|B|/n^2 + \mathsf{OPT}) . \tag{5}$$

Putting It All Together. Putting together our conclusions (3) about variance (after conditioning) of all but an $O(t \cdot \mathsf{OPT} + \varepsilon)$-fraction of pairs (5) about the contribution of those remaining pairs, and recalling that we chose $t \approx (\log \Delta)/\varepsilon^{O(k)}$, we obtain the bound

$$(\text{rounded cost}) \leqslant \frac{\log \Delta}{\varepsilon^{O(k)}} \cdot \mathsf{OPT} + \varepsilon\,. \tag{6}$$

The minimum over all ε which we can be achieved above is $\log \Delta \cdot \mathsf{OPT}^{\Omega(1/k)}$, leading to our final bound. As a reminder, in this overview, we have ignored some technical details – see the supplementary material for the rigerous proofs.

3. Open Problems

Our work gives, as far as we know, the first polynomial-time approximation algorithm for any MDS-style objective function with provable approximation guarantees when Δ is super-logarithmic in n. We hope that the idea of using convex programming hierarchies to Kamada-Kawai opens the door to further progress on related optimization-based formulations of dimension reduction. To this end, we offer a non-exhaustive list of some open problems, focusing on MDS-style objectives.

Improved algorithms for Kamada-Kawai. The only hardness of approximation we are aware of for the Kamada-Kawai objective is the result of Demaine et al. (2021) showing that an FPTAS is not possible unless $P = NP$. But this does not rule out, say, constant-factor approximation:

Problem 6 *Design a constant-factor approximation algorithm for the Kamada-Kawai function which is polynomial time for superlogarithmic aspect ratio.*

Weighted least squares. A wider range of MDS-style objectives (many available in R, scikit-learn, Matlab, etc.) is captured via the following generalization of Kamada-Kawai. A weighted least-squares instance is specified by dissimilarities $\{d_{ij}\}_{i,j\in[n]}$ and nonnegative weights $\{w_{ij}\}_{i,j\in[n]}$:

$$\min_{x_1,\dots,x_n\in\mathbb{R}^k} \mathop{\mathbb{E}}_{i\neq j} w_{ij}\left(1 - \frac{\|x_i - x_j\|}{d_{ij}}\right)^2 = \min_{x_1,\dots,x_n\in\mathbb{R}^k} \mathop{\mathbb{E}}_{i\neq j} \frac{w_{ij}}{d_{ij}^2}(d_{ij} - \|x_i - x_j\|)^2\,.$$

Setting all $w_{ij} = 1$ recovers the Kamada-Kawai objective; other choices lead to other popular MDS objectives. For instance, $w_{ij} = d_{ij}$ is "Sammon mapping" and $w_{ij} = d_{ij}^2$ is "raw stress". By analogy to CSPs, it's natural to associate a parameter to the weights measuring their density: let $W = \max w_{ij} / \mathbb{E}\, w_{ij}$. The algorithm of Demaine et al. (2021) extends to an additive PTAS for weighted MDS, with running time exponential in W. The case of "raw stress" is particuarly interesting, because it corresponds to least-squares projection of the given metric into the set of k-dimensional Euclidean distance matrices.

Problem 7 *Provide nontrivial polynomial-time approximation guarantees for weighted least-squares when Δ is superlogarithmic.*

Krustal Stress-1 and Stress-2. Another common objective function is Kruskal's "Stress-1":

$$\text{Stress-1}(x_1, \ldots, x_n) = \frac{\mathbb{E}_{i \neq j}(d_{ij} - \|x_i - x_j\|)^2}{\mathbb{E}_{i \neq j} \|x_i - x_j\|^2}\,.$$

Stress-1 is different from other objectives discussed here due to the normalization by $\|x_i - x_j\|$, meaning that the objective value penalizes contracting pairs somewhat more than the weighted least-squares objective. A local-search routine for minimizing Stress-1 is the default implementation in Matlab's built-in `mdscale`, and is also available in scikit-learn. Stress-2 is yet another variant:

$$\text{Stress-2}(x_1, \ldots, x_n) = \frac{\left(\mathbb{E}_{i \neq j}(d_{ij} - \|x_i - x_j\|)^2\right)^2}{\mathbb{E}_{i \neq j} \|x_i - x_j\|^4}\,.$$

Problem 8 *Provide a nontrivial poly-time approximation algorithm for Kruskal Stress-*1 *or Stress-*2*.*

Acknowledgments

We thank Moses Charikar, Piotr Indyk, and Robert Kleinberg for helpful conversations. SBH was funded by NSF CAREER award no. 2238080 and MLA@CSAIL. AB was funded by the Foundations of Data Science Institute, NSF award no. 2023505.

References

URL https://www.mathworks.com/help/stats/multidimensional-scaling.html.

Akshay Agrawal, Alnur Ali, Stephen Boyd, et al. Minimum-distortion embedding. *Foundations and Trends® in Machine Learning*, 14(3):211–378, 2021.

Sanjeev Arora, Wei Hu, and Pravesh K. Kothari. An analysis of the t-sne algorithm for data visualization. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 1455–1462. PMLR, 2018. URL http://proceedings.mlr.press/v75/arora18a.html.

Mukund Balasubramanian and Eric L Schwartz. The isomap algorithm and topological stability. *Science*, 295(5552):7–7, 2002.

Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *2011 ieee 52nd annual symposium on foundations of computer science*, pages 472–481. IEEE, 2011.

Yair Bartal, Nova Fandina, and Ofer Neiman. Dimensionality reduction: theoretical perspective on practical measures. *Advances in Neural Information Processing Systems*, 32, 2019.

Ingwer Borg and Patrick JF Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.

Ingwer Borg, Patrick JF Groenen, and Patrick Mair. *Applied multidimensional scaling*. Springer Science & Business Media, 2012.

Jean Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics*, 52:46–52, 1985.

Chaomei Chen. Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences*, 101(suppl_1):5303–5310, 2004.

Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

Nicholas A Christakis and James H Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

Vincent Cohen-Addad, Euiwoong Lee, and Alantha Newman. Correlation clustering with sherali-adams. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 651–661. IEEE, 2022. doi: 10.1109/FOCS54457.2022.00068. URL https://doi.org/10.1109/FOCS54457.2022.00068.

Trevor F Cox and Michael AA Cox. *Multidimensional scaling*. CRC press, 2000.

Sami Davies, Janardhan Kulkarni, Thomas Rothvoss, Jakub Tarnawski, and Yihao Zhang. Scheduling with communication delays via lp hierarchies and clustering. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 822–833. IEEE, 2020.

Mark L Davison and Stephen G Sireci. Multidimensional scaling. In *Handbook of applied multivariate statistics and mathematical modeling*, pages 323–352. Elsevier, 2000.

Jan de Leeuw and Willem Heiser. 13 theory of multidimensional scaling. *Handbook of statistics*, 2: 285–316, 1982.

Jan De Leeuw and Patrick Mair. Multidimensional scaling using majorization: Smacof in r. 2011.

Erik Demaine, Adam Hesterberg, Frederic Koehler, Jayson Lynch, and John Urschel. Multidimensional scaling: Approximation and complexity. In *International Conference on Machine Learning*, pages 2568–2578. PMLR, 2021.

Venkatesan Guruswami and Ali Kemal Sinop. Faster SDP hierarchy solvers for local rounding algorithms. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 197–206. IEEE Computer Society, 2012. doi: 10.1109/FOCS.2012.58. URL https://doi.org/10.1109/FOCS.2012.58.

Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159, 2008.

Geoffrey E Hinton and Sam Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15, 2002.

Piotr Indyk, Jiří Matoušek, and Anastasios Sidiropoulos. 8: low-distortion embeddings of finite metric spaces. In *Handbook of discrete and computational geometry*, pages 211–231. Chapman and Hall/CRC, 2017.

Anna R Karlin, Claire Mathieu, and C Thach Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. In *Integer Programming and Combinatoral Optimization: 15th International Conference, IPCO 2011, New York, NY, USA, June 15-17, 2011. Proceedings 15*, pages 301–314. Springer, 2011.

Joseph B Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964a.

Joseph B Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2): 115–129, 1964b.

Joseph B Kruskal and Myron Wish. *Multidimensional scaling*. Sage, 1978.

Claire Mathieu and Alistair Sinclair. Sherali-adams relaxations of the matching polytope. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 293–302. ACM, 2009. doi: 10.1145/1536414.1536456. URL https://doi.org/10.1145/1536414.1536456.

Jirí Matousek and Anastasios Sidiropoulos. Inapproximability for metric embeddings into r^d. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 405–413. IEEE Computer Society, 2008. doi: 10.1109/FOCS.2008.21. URL https://doi.org/10.1109/FOCS.2008.21.

Jiří Matoušek and Anastasios Sidiropoulos. Inapproximability for metric embeddings into $\mathbb{R}^d$. *Transactions of the American Mathematical Society*, 362(12):6341–6365, 2010.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Prasad Raghavendra and Ning Tan. Approximating csps with global cardinality constraints using sdp hierarchies. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 373–387. SIAM, 2012.

Marion Webster Richardson. Multidimensional psychophysics. *Psychological bulletin*, 35:659–660, 1938.

Roger N Shepard. The analysis of proximities: multidimensional scaling with an unknown distance function. i. *Psychometrika*, 27(2):125–140, 1962a.

Roger N Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. ii. *Psychometrika*, 27(3):219–246, 1962b.

Anastasios Sidiropoulos, Dingkang Wang, and Yusu Wang. Metric embeddings with outliers. In *SODA*, pages 670–689, 2017.

Warren S Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17:395–416, 2007.

Dean W Wichern et al. Applied multivariate statistical analysis.

Forrest W Young. *Multidimensional scaling: History, theory, and applications*. Psychology Press, 2013.