Embedding Approximately Low-Dimensional ℓ_2^2 Metrics into ℓ_1

Amit Deshpande[1], Prahladh Harsha[2], and Rakesh Venkat[3]

1 **Microsoft Research, Bangalore, India**
amitdesh@microsoft.com
2 **Tata Institute of Fundamental Research, Mumbai, India**
prahladh@tifr.res.in
3 **Tata Institute of Fundamental Research, Mumbai, India**
rakesh09@gmail.com

Abstract

Goemans showed that any n points $x_1, \ldots x_n$ in d-dimensions satisfying ℓ_2^2 triangle inequalities can be embedded into ℓ_1, with worst-case distortion at most $\sqrt{d}$. We consider an extension of this theorem to the case when the points are *approximately* low-dimensional as opposed to exactly low-dimensional, and prove the following analogous theorem, albeit with *average distortion* guarantees: There exists an ℓ_2^2-to-ℓ_1 embedding with average distortion at most the *stable rank*, $\mathrm{sr}(M)$, of the matrix M consisting of columns $\{x_i - x_j\}_{i<j}$. Average distortion embedding suffices for applications such as the Sparsest Cut problem. Our embedding gives an approximation algorithm for the Sparsest Cut problem on low threshold-rank graphs, where earlier work was inspired by Lasserre SDP hierarchy, and improves on a previous result of the first and third author [Deshpande and Venkat, In *Proc. 17th APPROX*, 2014]. Our ideas give a new perspective on ℓ_2^2 metric, an alternate proof of Goemans' theorem, and a simpler proof for *average* distortion $\sqrt{d}$.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Metric Embeddings, Sparsest Cut, Negative type metrics, Approximation Algorithms

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2016.10

1 Introduction

A finite metric space consists of a pair $(\mathcal{X}, d)$, where $\mathcal{X}$ is a finite set of points, and $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a distance function on pairs of points in $\mathcal{X}$. Finite metric spaces arise naturally in combinatorial optimization (e.g., the ℓ_1 space in cut problems), and in practice (e.g., edit-distance between strings over some alphabet Σ). Since the input space may not be amenable to efficient optimization, or may not admit efficient algorithms, one looks for *embeddings* from these input spaces to easier spaces, while minimizing the *distortion* incurred. Given its importance, various aspects of such embeddings have been investigated such as dimension, distortion, efficient algorithms, and hardness results (refer to surveys [10, 16, 14] and references therein). In this paper, we provide better distortion guarantees for embedding *approximately* low-dimensional points in the ℓ_2^2-metric into ℓ_1, and give applications to the Sparsest Cut problem.

In the Sparsest Cut problem, we are given graphs C, D on the same vertex set V, with $|V| = n$, called the *cost* and *demand* graphs, respectively. They are specified by non-negative

36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2016).
Editors: Akash Lal, S. Akshay, Saket Saurabh, and Sandeep Sen; Article No. 10; pp. 10:1–10:13
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

edge weights $c_{ij}, d_{ij} \geq 0$, for $i < j \in [n]$ and the *(non-uniform) sparsest cut* problem, henceforth referred to as Sparsest Cut, asks for a subset $S \subseteq V$ that minimizes

$$\Phi(S) := \frac{\sum_{i<j} c_{ij} \left|\mathbb{I}_S(i) - \mathbb{I}_S(j)\right|}{\sum_{i<j} d_{ij} \left|\mathbb{I}_S(i) - \mathbb{I}_S(j)\right|},$$

where $\mathbb{I}_S(i)$ is the indicator function giving 1, if $i \in S$, and 0, otherwise. We denote the optimum by $\Phi^* := \min_{S \subseteq V} \Phi(S)$. When the demand graph is a complete graph on n vertices with uniform edge weights, the problem is then commonly referred to as the Uniform Sparsest Cut problem.

The best known (unconditional) approximation guarantee for the Uniform Sparsest Cut problem is $O(\sqrt{\log n})$, due to Arora, Rao and Vazirani [3] (henceforth referred to as the ARV algorithm). Building on techniques in this work, Arora, Lee and Naor [2] give a $O(\sqrt{\log n}\log\log n)$ algorithm for non-uniform Sparsest Cut. These results come from a semi-definite programming (SDP) relaxation to produce solutions in the ℓ_2-squared metric space, i.e., a set of vectors $\{x_i\}_{i \in V}$ in some high dimensional space that satisfy triangle inequality constraints on the squared distances in the following sense.

$$\|x_i - x_j\|_2^2 + \|x_j - x_k\|_2^2 \geq \|x_i - x_k\|_2^2 \qquad \forall\ i, j, k \in [n].$$

Since the ℓ_1 metric lies in the non-negative cone of cut (semi-)metrics, ARV [3] and Arora-Lee-Naor[2] round their solutions via low-distortion embeddings of the above ℓ_2^2 solution into ℓ_1 metric. Embeddings with low *average-distortion* suffice for applications to the Sparsest Cut problem.

Any n points satisfying ℓ_2^2 triangle inequalities make only acute angles among themselves, and therefore must lie in $\Omega(\log n)$ dimensions (Chapter 15, [1]). However, for low *threshold-rank* graphs, or more generally, when the r-th smallest generalized eigenvalue of the cost and demand graphs satisfies $\lambda_r(C, D) \gg \Phi_{SDP}$, the above SDP solution is known to be *approximately* low-dimensional, that is, the span of its top r eigenvectors contains nearly all of its total eigenmass (implicit in [9]). Moreover, it can be embedded into ℓ_1 using solutions of higher-levels of the Lasserre SDP hierarchy to obtain a PTAS-like approximation guarantee [9]. This motivates the quest for finding more efficient embeddings of low-dimensional or *approximately* low-dimensional ℓ_2^2 metrics into ℓ_1.

Goemans (unpublished, appears in [15]) showed that if the points satisfying ℓ_2^2 triangle inequalities lie in d dimensions, then they can be embedded into ℓ_2 (and hence into ℓ_1, since there is an isometry from $\ell_2 \hookrightarrow \ell_1$ [16]) with $\sqrt{d}$ distortion.

▶ **Theorem 1.1** (Goemans [15, Appendix B]). *Let $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ be n points satisfying ℓ_2^2 triangle inequalities. Then there exists an $\ell_2^2 \hookrightarrow \ell_2$ embedding $x_i \mapsto f(x_i)$ with distortion $\sqrt{d}$, that is,*

$$\frac{1}{\sqrt{d}}\ \|x_i - x_j\|_2^2 \leq \|f(x_i) - f(x_j)\|_2 \leq \|x_i - x_j\|_2^2, \quad \forall\ i, j \in V.$$

Comparison of Goemans and ARV

Since n points satisfying ℓ_2^2 triangle inequalities must lie in $d = \Omega(\log n)$ dimensions (Chapter 15, [1]), the ARV algorithm [3] implies an $\ell_2^2 \hookrightarrow \ell_1$ embedding with *average* distortion $O(\sqrt{d})$, and Arora-Lee-Naor [2] improve it to $\widetilde{O}(\sqrt{d})$ worst-case distortion. In the other direction, is it possible to extend Theorem 1.1 to give ARV-like guarantees? Here are two immediate ideas that come to mind.

- Combine Theorem 1.1 with a dimension reduction to $O(\log n)$ dimensions for ℓ_2^2 metrics, similar to the Johnson-Lindenstrauss lemma for ℓ_2. Such a dimension reduction for ℓ_2^2 that approximately preserves all pairwise ℓ_2^2 distances is ruled out by Magen and Moharrami [15], although their proof does not rule out dimension reduction for *average* distortion.
- Extend Theorem 1.1 to work with approximate ℓ_2^2 triangle inequalities, and then combine it with the Johnson-Lindenstrauss lemma. The Johnson-Lindenstrauss lemma, when applied to points satisfying ℓ_2^2 triangle inequalities, preserves their ℓ_2^2 triangle inequalities only approximately. That is, the points after the Johnson-Lindenstrauss random projection satisfy

  $$\|x_i - x_j\|_2^2 + \|x_j - x_k\|_2^2 \geq (1 - O(\epsilon))\, \|x_i - x_k\|_2^2 \qquad \forall\ i, j, k \in [n].$$

  We note that a generalization of Theorem 1.1 that accommodates approximate ℓ_2^2 triangle inequalities (in the additive sense not multiplicative as above) does hold, but its only proof (due to Trevisan [personal communication]) that we are aware of uses the technical core of the analysis of the ARV algorithm.

Here we seek a robust generalization of Goemans' theorem that avoids the above caveats. Our version of Goemans' theorem uses *average* distortion instead of worst-case. It is robust in the sense that it works with *approximate* dimension instead of the actual dimension. Such a robust version opens up another possible approach to the general Sparsest Cut problem: reduce the *approximate* dimension while preserving the pairwise distances on *average*, and then apply the robust version of Goemans' theorem. Moreover, our definition of the approximate dimension is spectral, and our results can be easily compared to those of Guruswami-Sinop [9] on Lasserre SDP hierarchies and Kwok et al. [13] on higher order Cheeger inequalities (see Sections 1.1 and 1.2 for comparisons).

1.1 Our Results

We consider a robust version of Goemans' theorem, when the points $x_1, x_2, \ldots, x_n$ are only *approximately* low-dimensional. We quantify this *approximate* dimension by the *stable rank* of the difference matrix $M \in \mathbb{R}^{d \times \binom{n}{2}}$ having columns $\{x_i - x_j\}_{i<j}$. Stable rank of the difference matrix is a natural choice because (a) stable rank is a continuous proxy for rank or dimension arising naturally in many applications [5, 17], (b) the difference matrix M is invariant under any shift of origin, and (c) the difference matrix of the SDP solution for the Sparsest Cut problem on *low threshold-rank* graphs indeed has low stable rank (implicit in [9]).

▶ **Definition 1.2** (Stable Rank). Given $x_1, \ldots, x_n \in \mathbb{R}^d$, let $M \in \mathbb{R}^{d \times \binom{n}{2}}$ be the matrix with columns $\{x_i - x_j\}_{i<j}$. The stable rank of the points is defined as the stable rank of M, given by $\operatorname{sr}(M) := \|M\|_F^2 / \|M\|_2^2$, where $\|M\|_F$ and $\|M\|_2$ are the Frobenius and spectral norm of M respectively.

Note that $\operatorname{sr}(M) \leq \operatorname{rank}(M) \leq d$, when the points $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$. Our robust version of Goemans' theorem is as follows.

▶ **Theorem 1.3** (Embedding almost low-dimensional vectors). *Let $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ be n points satisfying ℓ_2^2 triangle inequalities. Then there exists an $\ell_2^2 \hookrightarrow \ell_2$ embedding $x_i \mapsto h(x_i)$ with* average *distortion bounded by the stable rank of M, that is,*

$$\|h(x_i) - h(x_j)\|_2 \leq \|x_i - x_j\|_2^2, \quad \forall i, j \in V,$$

and

$$\frac{1}{\mathrm{sr}\,(M)} \sum_{i<j} \|x_i - x_j\|_2^2 \leq \sum_{i<j} \|h(x_i) - h(x_j)\|_2 \,.$$

We note that the above theorem is not a strict generalization of Goemans' theorem to the approximate dimension case. To obtain a truly robust version of Goemans' theorem quantitatively, one might ask if the dependence on $\mathrm{sr}\,(M)$ in the above theorem can be improved from $\mathrm{sr}\,(M)$ to $\sqrt{\mathrm{sr}\,(M)}$.

Our proof technique gives a new perspective on ℓ_2^2 metric, an alternate proof of Goemans' theorem, and a simpler *algorithmic* proof for average distortion $\sqrt{d}$ based on a squared-length distribution (see Section 4, and the remark following the proof of Theorem 4.1). Also, the result can be quantitatively compared to guarantees given by higher-order Cheeger inequalities [13]; we discuss this in more detail at the end of this section. While most known embeddings from ℓ_2^2 to ℓ_1 are Frechet embeddings, our embedding is projective (similar in spirit to [9, 7]).

Theorem 1.3 immediately implies an $\mathrm{sr}\,(M)$-approximation to the Uniform Sparsest Cut problem. In fact, with a slight modification, we obtain a similar result for the general Sparsest Cut problem (see theorem below).

► **Theorem 1.4.** *There is an r/δ-approximation algorithm for* Sparsest Cut *instances C, D satisfying $\lambda_r(C, D) \geq \Phi_{SDP}/(1-\delta)$, where $\lambda_r(C, D)$ is the r-th smallest generalized eigenvalue (see Section 2) of the Laplacians of the cost and demand graphs.*

The precondition on $\lambda_r(C, D)$ is the same as in previous works [9, 7], and we improve the $O(r/\delta^2)$-approximation of [7] by a factor of $1/\delta$. Our proof follows from the robust version of Goemans' embedding into ℓ_2 whereas these previous works gave embeddings directly into ℓ_1 by either using higher levels of Lasserre explicitly [9] or using only the basic SDP solution but inspired by the properties of Lasserre vectors [7]. We can infer the following corollary almost immediately:

► **Corollary 1.5.** *For any $\epsilon > 0$ and a d-regular cost graph C satisfying $\lambda_r(C) \geq \epsilon d$, there is a* $\max\left\{O(r), \frac{1}{\sqrt{\epsilon}}\right\}$ *approximation to* Uniform Sparsest Cut.

Proof. The implicit demand graph here is K_n, the complete graph on n vertices, and thus the generalized eigenvalues are $\lambda_r(C, K_n) = \lambda_r/n$. Consider two cases: If $\Phi_{SDP} \leq \epsilon d/100n$ then $\lambda_r/n \geq 100\Phi_{SDP}$ yielding an $O(r)$ approximation by Theorem 1.4. Otherwise, if $\Phi_{SDP} \geq \epsilon d/100n$, then running a basic Cheeger rounding and analysis on (one co-ordinate of) the SDP solution would itself give a cut of sparsity $O(d\sqrt{\epsilon}/n) \leq \Phi_{SDP}/\sqrt{\epsilon}$. Thus, using the minimum of these gives a cut within a factor $\max\{O(r), 1/\sqrt{\epsilon}\}$ of the optimum. ◄

1.2 Related work

We recall that the best known upper bound for the worst-case distortion of embedding $\ell_2^2 \hookrightarrow \ell_1$ is $O(\sqrt{\log n} \cdot \log\log n)$ [3, 2], while the best known lower bound is $(\log n)^{\Omega(1)}$ for worst-case distortion [6], and $\exp(\Omega(\sqrt{\log\log n}))$ for average distortion [11]. Guarantees to Sparsest Cut on *low threshold-rank* graphs were obtained using higher levels of the Lasserre hierarchy for SDPs [4, 9]. In contrast, a previous work of the first and third author [7] showed weaker guarantees, but using just the basic SDP relaxation. Oveis Gharan and Trevisan [8] also give a rounding algorithm for the basic SDP relaxation on low-threshold rank graphs, but require a stricter pre-condition on the eigenvalues ($\lambda_r \gg \log^{2.5} r \cdot \Phi(G)$), and leverage it to give a stronger $O(\sqrt{\log r})$-approximation guarantee. Their improvement comes from a new structure theorem on the SDP solutions of low threshold-rank graphs being clustered, and using the techniques in ARV for analysis.

Kwok et al. [13] showed that a better analysis of Cheeger's inequality gives a $O(r \cdot \sqrt{d/\lambda_r})$ approximation to Uniform Sparsest Cut on d-regular graphs. In particular, when $\lambda_r(G) \geq \epsilon d$, this gives a $O(r/\sqrt{\epsilon})$ approximation for the Uniform Sparsest Cut problem. Note that Corollary 1.5 gives a slightly better approximation in this setting.

Further, while the Kwok et al. result is tight with respect to the spectral solution, our approach allows for an improvement in terms of the dependence on r to $\sqrt{r}$, since it uses the SDP relaxation rather than a spectral solution.

2 Preliminaries and Notation

Sets, Matrices, Vectors

We use $[n] = \{1, \ldots, n\}$. For a matrix $X \in \mathbb{R}^{d \times d}$, we say $X \succeq 0$ or X is positive-semidefinite (psd) if $y^T X y \geq 0$ for all $y \in \mathbb{R}^d$. The Gram-matrix of a matrix $M \in \mathbb{R}^{d_1 \times d_2}$ is the matrix $M^T M$, which is psd.

Every matrix M has a singular value decomposition $M = \sum_i \sigma_i u_i v_i^T = UDV^T$. Here, the matrices U, V are Unitary, and D is the diagonal matrix of the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$, in non-increasing order. When not clear from context, we denote the singular values of M by $\sigma_i(M)$.

The *Frobenius* norm of M is given by $\|M\|_F := \sqrt{\sum_i \sigma_i^2(M)} = \sqrt{\sum_{i \in [d_1], j \in [d_2]} M(i,j)^2}$. In our analysis, we will sometimes view a matrix M as a collection of its columns viewed as vectors; $M = (m_j)_{j \in [d_2]}$. In this case, $\|M\|_F^2 = \sum_j \|m_j\|_2^2$. The spectral norm of M is $\|M\|_2 := \sigma_1$.

Generalized Eigenvalues

Given two symmetric matrices $X, Y \in \mathbb{R}^d \times d$ with $Y \succeq 0$, and for $i \leq \text{rank}(Y)$, we define their i-th smallest generalized eigenvalue as the following:

$$\lambda_i = \max_{\text{rank}(Z) \leq i-1} \min_{w \perp Z; w \neq 0} \frac{w^T X w}{w^T Y w}$$

Rank and Stable Rank

The rank of the matrix M (denoted by $\text{rank}\,(M)$) is the number of non-zero singular values. Recall that the *stable rank* of the matrix M, $\text{sr}\,(M) = \frac{\|M\|_F^2}{\sigma_1(M)^2}$. Note that $\text{sr}\,(M) = \sum_{i=1}^{\text{rank}(M)} \sigma_i^2(M)/\sigma_1^2(M) \leq \text{rank}\,(M)$.

Metric spaces and embeddings

For our purposes, a (semi-)metric space $(\mathcal{X}, d)$ consists of a finite set of points $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and a distance function $d : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}_{\geq 0}$ satisfying the following three conditions:

1. $d(x, x) = 0$, $\forall x \in \mathcal{X}$.
2. $d(x, y) = d(y, x)$.
3. (Triangle inequality) $d(x, y) + d(y, z) \geq d(x, z)$.

An *embedding* from a metric space $(\mathcal{X}, d)$ to a metric space $(\mathcal{Y}, d')$ is a mapping $f : \mathcal{X} \to \mathcal{Y}$. The embedding is called a *contraction*, if

$$d'(f(x_i), f(x_j)) \le d(x_i, x_j), \qquad \forall x_i, x_j \in \mathcal{X}.$$

For convenience, we will only deal with contractive mappings in this paper. A contractive mapping is said to have (worst-case) distortion Δ, if

$$\sup_{i,j} \frac{d(x_i, x_j)}{d'(f(x_i), f(x_j))} \le \Delta.$$

It is said to have *average* distortion β, if

$$\frac{\sum_{i<j} d(x_i, x_j)}{\sum_{i<j} d(f(x_i), f(x_j))} \le \beta.$$

Note that a mapping with worst-case distortion Δ also has average distortion Δ, but not necessarily vice-versa.

The ℓ_2^2 space

A set of points $\{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^d$ are said to satisfy ℓ_2^2 triangle inequality constraints, or said to be in ℓ_2^2 space, if it holds that

$$\|x_i - x_j\|_2^2 + \|x_j - x_k\|_2^2 \ge \|x_i - x_k\|_2^2 \qquad \forall i, j, k \in [n].$$

These satisfy the triangle inequalities on the *squares* of their ℓ_2 distances. The corresponding metric space is $(\mathcal{X}, d)$, where $d(i, j) := \|x_i - x_j\|_2^2$.

Graphs and Laplacians

All graphs will be defined on a vertex set V of size n. The vertices will usually be referred to by indices $i, j, k, l \in [n]$. Given a graph with weights on pairs $W : \binom{V}{2} \mapsto \mathbb{R}^+$, the graph Laplacian matrix is defined as:

$$L_W(i, j) := \begin{cases} -W(i, j) & \text{if } i \ne j \\ \sum_k W(i, k) & \text{if } i = j. \end{cases}$$

Note that $L_W \succeq 0$. We will denote the eigenvalues of (the Laplacian of) G by $0 = \lambda_1 \le \lambda_2 \ldots \le \lambda_n$, in *increasing* order.

Sparsest Cut SDP

The SDP we use for Sparsest Cut on the vertex set V with costs and demands $c_{ij}, d_{kl} \ge 0$ and corresponding cost and demand graphs $C : \binom{V}{2} \mapsto \mathbb{R}^+$ and $D : \binom{V}{2} \mapsto \mathbb{R}^+$, is effectively the following:

$$\textbf{SDP:} \quad \Phi_{SDP} := \min \sum_{i<j} c_{ij} \|x_i - x_j\|_2^2$$

$$\text{subject to} \begin{cases} \|x_i - x_j\|_2^2 + \|x_j - x_k\|_2^2 \ge \|x_i - x_k\|_2^2 & \forall i, j, k \in [n]. \\ \sum_{k<l} d_{kl} \|x_k - x_l\|_2^2 = 1. \end{cases}$$

Note that the solution to the above SDP is in ℓ_2^2 space.

ℓ_1 embeddings and cuts

Since ℓ_1 metrics are exactly the cone of cut-metrics, it follows from the previous discussion on embeddings, that producing an embedding of the SDP solutions $\mathcal{X} = \{x_1, \ldots .x_n\}$ in ℓ_2^2 space to ℓ_1 space with distortion α would give an α-approximation to SPARSEST CUT. Producing one with *average* distortion α would give an α-approximation to UNIFORM SPARSEST CUT. Furthermore, since ℓ_2 embeds isometrically (distortion 1) into ℓ_1, it suffices to show embeddings into ℓ_2 for the above purposes.

Key Lemma

The following lemma about ℓ_2^2 spaces was observed by Deshpande and Venkat [7]. We will reuse this in the rest of the paper.

▶ **Lemma 2.1** ([7, Proposition 1.3]). *Let $x_1, x_2, \ldots, x_n$ be n points satisfying ℓ_2^2 triangle inequalities. Then*

$$\left\langle x_i - x_j, \frac{x_k - x_l}{\|x_k - x_l\|_2} \right\rangle^2 \leq |\langle x_i - x_j, x_k - x_l\rangle| \leq \|x_i - x_j\|_2^2, \quad \forall i, j, k, l \in V.$$

An immediate consequence of this lemma is that we can show that a large class of naturally defined $\ell_2^2 \hookrightarrow \ell_2$ embeddings are contractions.

▶ **Lemma 2.2** (Contraction). *Let $x_1, x_2, \ldots, x_n$ be n points satisfying ℓ_2^2 triangle inequalities. For any probability distribution $\{p_{kl}\}_{k<l}$, let P be the symmetric psd matrix defined as $P := \sum_{k<l} p_{kl}\ (x_k - x_l)(x_k - x_l)^T$. Then the $\ell_2^2 \hookrightarrow \ell_2$ embedding given by $x_i \mapsto P^{1/2} x_i$ is a contraction, that is,*

$$\left\| P^{1/2}(x_i - x_j) \right\|_2 \leq \|x_i - x_j\|_2^2, \quad \forall i, j \in V.$$

Proof. The following holds for all i, j:

$$\begin{aligned}
\left\| P^{1/2}(x_i - x_j) \right\|_2 &= \left((x_i - x_j)^T P (x_i - x_j) \right)^{1/2} \\
&= \left(\sum_{k<l} p_{kl} \langle x_i - x_j, x_k - x_l \rangle^2 \right)^{1/2} \\
&\leq \left(\sum_{k<l} p_{kl}\ \|x_i - x_j\|_2^4 \right)^{1/2} && \text{[By Lemma 2.1]} \\
&= \|x_i - x_j\|_2^2. && \text{[Since } \textstyle\sum_{k<l} p_{kl} = 1\text{]}
\end{aligned}$$

◀

3 Embedding almost low-dimensional vectors

We now prove the robust version of Goemans' theorem in terms of stable rank. We give two proofs, and show an application to round solutions to SPARSEST CUT on low-threshold-rank graphs. As before, given a set of points $x_1, \ldots, x_n$ in $\mathbb{R}^d$, define their difference matrix $M \in \mathbb{R}^{d \times \binom{n}{2}}$ as the matrix with columns as $\{x_i - x_j\}_{i<j}$.

Proof of Theorem 1.3. Let u and v be the top left and right singular vector of M, respectively, and $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_d$ be the singular values of M. Then $Mv = \sigma_1 u$, or in other words, $\sigma_1 u = \sum_{k<l} v_{kl}(x_k - x_l)$. Now consider the embedding $x_i \mapsto h(x_i) = P^{1/2} x_i$, where the probability distribution $p_{kl} \propto |v_{kl}|$, that is

$$P = \sum_{k<l} \frac{|v_{kl}|}{\|v\|_1} (x_k - x_l)(x_k - x_l)^T.$$

This embedding is a contraction by Lemma 2.2. Now let's bound its average distortion.

$$\begin{aligned}
\sum_{i<j} \|h(x_i) - h(x_j)\|_2 &= \sum_{i<j} \left\| P^{1/2}(x_i - x_j) \right\|_2 \\
&= \sum_{i<j} \left((x_i - x_j)^T P (x_i - x_j) \right)^{1/2} \\
&= \sum_{i<j} \left(\sum_{k<l} \frac{|v_{kl}|}{\|v\|_1} \langle x_i - x_j, x_k - x_l \rangle^2 \right)^{1/2} \\
&\geq \sum_{i<j} \sum_{k<l} \frac{|v_{kl}|}{\|v\|_1} |\langle x_i - x_j, x_k - x_l \rangle| && \text{[By Jensen's inequality]} \\
&\geq \sum_{i<j} \frac{1}{\|v\|_1} \left| \left\langle x_i - x_j, \sum_{k<l} v_{kl}(x_k - x_l) \right\rangle \right| && \text{[By triangle inequality]} \\
&= \frac{1}{\|v\|_1} \sum_{i<j} |\langle x_i - x_j, \sigma_1 u \rangle| \\
&= \frac{1}{\|v\|_1} \sum_{i<j} \sigma_1^2 |v_{ij}| \\
&= \sigma_1^2 = \frac{\|M\|_F^2}{\mathrm{sr}(M)} \\
&= \frac{1}{\mathrm{sr}(M)} \sum_{i<j} \|x_i - x_j\|_2^2 .
\end{aligned}$$

◀

3.1 An alternative proof

We can alternatively get the same guarantee as in Theorem thm:stable-rank, by giving a one-dimensional ℓ_2 embedding (and hence also ℓ_1 embedding without any extra effort) along the top singular vector of the difference matrix M. This gives an interesting "spectral" algorithm that uses spectral information about the point set, akin to spectral algorithms in graphs that use the spectrum of the graph Laplacian.

► **Theorem 3.1.** *Let $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ be n points satisfying ℓ_2^2 triangle inequalities with M as their difference matrix. Let $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^{\binom{n}{2}}$ be its top left and right singular vectors, respectively. Then $x_i \mapsto \frac{\sigma_1}{\|v\|_1} \langle x_i, u \rangle$ is an $\ell_2^2 \hookrightarrow \ell_2$ embedding with average distortion bounded by the stable rank of M.*

Proof. We have $Mv = \sigma_1 u$, or equivalently, $\sigma_1 u = \sum_{k<l} v_{kl}(x_k - x_l)$. Our embedding is a contraction since

$$\begin{aligned}\frac{\sigma_1}{\|v\|_1}\left|\langle x_i - x_j, u\rangle\right| &= \frac{1}{\|v\|_1}\left|\left\langle x_i - x_j, \sum_{k<l} v_{kl}(x_k - x_l)\right\rangle\right| \\ &\leq \frac{1}{\|v\|_1}\sum_{k<l}|v_{kl}|\left|\langle x_i - x_j, x_k - x_l\rangle\right| \\ &\leq \frac{1}{\|v\|_1}\sum_{k<l}|v_{kl}|\,\|x_i - x_j\|_2^2 && \text{[By Lemma 2.1]} \\ &= \|x_i - x_j\|_2^2 .\end{aligned}$$

Now let's bound the average distortion.

$$\begin{aligned}\sum_{i<j}\frac{\sigma_1}{\|v\|_1}\left|\langle x_i - x_j, u\rangle\right| &= \sum_{i<j}\frac{\sigma_1}{\|v\|_1}|\sigma_1 v_{ij}| && \text{[Since } u^T M = \sigma_1 v^T\text{]} \\ &= \sigma_1^2 = \frac{\|M\|_F^2}{\operatorname{sr}(M)} \\ &= \frac{1}{\operatorname{sr}(M)}\sum_{i<j}\|x_i - x_j\|_2^2 .\end{aligned}$$

◀

3.2 Application to Sparsest Cut on low-threshold rank graphs

We first state a property of SDP solutions on low threshold-rank graphs, proved by Guruswami and Sinop [9] using the Von-Neumann inequality.

▶ **Proposition 3.2** (Von-Neumann inequality [9, Theorem 3.3]). *Let $0 \leq \lambda_1 \leq \ldots \leq \lambda_m$ be the generalized eigenvalues of the Laplacian matrices of the cost and demand graphs. Let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ be the singular vectors of the matrix M with columns $\{\sqrt{d_{ij}}(x_i - x_j)\}_{i<j}$. Then*

$$\frac{\sum_{t \geq r+1}\sigma_j^2}{\sum_{t=1}^{n}\sigma_j^2} \leq \frac{\Phi_{SDP}}{\lambda_{r+1}}.$$

In particular, note that on graphs where $\lambda_r \geq \Phi_{SDP}/(1-\delta)$, $\sum_{i \leq r}\sigma_i^2 \geq \delta\sum_i \sigma_i^2$. This implies that $\operatorname{sr}(M) = \sum_i \sigma_i^2/\sigma_1^2 \leq r \cdot \sum_i \sigma_i^2/\sum_{i \leq r}\sigma_i^2 \leq r/\delta$.

We can now modify the proof of Theorem 3.1 to prove Theorem 1.4.

Proof of Theorem 1.4. Let $x_1, \ldots, x_n$ be the SDP solution on given instance C, D. We now let M be the matrix with columns $\{\sqrt{d_{kl}}(x_k - x_l)\}_{k<l}$, and u, v, σ_1 to be the top left singular vector, top right singular vector, and the maximum singular value respectively of M. By the preceding remark, $\operatorname{sr}(M) \leq r/\delta$. The mapping we use is as follows

$$x_i \mapsto \frac{1}{\sum_{kl}\sqrt{d_{kl}}v_{kl}}\langle x_i, u\rangle .$$

The proofs to show contraction and bound the distortion follow exactly as in the proof of Theorem 3.1. Note that while looking at the distortion, we need to lower bound the quantity $\sum_{ij} d_{ij}\|g(x_i) - g(x_j)\|_2$. ◀

As in Deshpande and Venkat [7], the above algorithm is a fixed polynomial time algorithm and does not grow with the threshold rank unlike the algorithm of Guruswami and Sinop [9] where they use r-levels of the Lasserre SDP hierarchy to secure the guarantee. Furthermore, the above analysis improves the guarantee of Deshpande and Venkat [7] by a factor of $O(1/\delta)$.

4 Embedding low-dimensional vectors à la Goemans

In this section, we first view the proof of Goemans' theorem in the framework of Lemma 2.2 by giving a probability distribution using the minimum volume enclosing elliposid of the difference vectors $(x_i - x_j)$'s. We then give a simpler proof, albeit for the *average* distortion case, based on a probability distribution arising from a squared-length distribution. Via a well-known duality statement, this technique recovers Goemans' theorem for *worst-case* distortion for embeddings into ℓ_1, although non-constructively.

4.1 An alternate proof of Goemans' theorem

Here is an adaptation of the proof from [15] re-stated in our framework. The following proof is arguably simpler and more straightforward as it works with the difference vectors instead of the original vectors and their negations.

▶ **Theorem 1.1** (restated – Goemans [15, Appendix B])). Let $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ be n points satisfying ℓ_2^2 triangle inequalities. Then there exists an $\ell_2^2 \hookrightarrow \ell_2$ embedding $x_i \mapsto f(x_i)$ with distortion $\sqrt{d}$, that is,

$$\frac{1}{\sqrt{d}} \; \|x_i - x_j\|_2^2 \leq \|f(x_i) - f(x_j)\|_2 \leq \|x_i - x_j\|_2^2, \quad \forall\, i, j \in V.$$

Proof. Consider all the difference vectors $(x_i - x_j)$'s, and let their minimum volume enclosing ellipsoid be given by $E := \{x \;:\; x^T Q x \leq 1\}$, for some psd matrix $Q \in \mathbb{R}^{d\times d}$. By John's theorem (or Lagrangian duality for the corresponding convex program), we have $Q^{-1} = \sum_{k<l} \alpha_{kl}\, (x_k - x_l)(x_k - x_l)^T$, with all $\alpha_{kl} \geq 0$. Moreover, $\alpha_{kl} \neq 0$ iff $(x_k - x_l)^T Q (x_k - x_l) = 1$. Notice that $d = \mathrm{Tr}\,(\mathbb{I}_d) = \mathrm{Tr}\,\left(Q^{1/2} Q^{-1} Q^{1/2}\right) = \sum_{k<l} \alpha_{kl}$. We define the embedding as

$$f(x_i) := \frac{1}{\sqrt{d}} Q^{-1/2} x_i.$$

This embedding is a contraction by Lemma 2.2. We now bound the distortion:

$$\begin{aligned} \|f(x_i) - f(x_j)\|_2 &= \frac{1}{\sqrt{d}} \; \left\|Q^{-1/2}(x_i - x_j)\right\|_2 \\ &\geq \frac{1}{\sqrt{d}} \; \frac{\|x_i - x_j\|_2^2}{\left\|Q^{1/2}(x_i - x_j)\right\|_2} && \text{[By Cauchy-Schwarz inequality]} \\ &\geq \frac{1}{\sqrt{d}} \; \|x_i - x_j\|_2^2. && \text{[Since } (x_i - x_j)^T Q (x_i - x_j) \leq 1, \text{ for all } i, j] \end{aligned}$$

◀

4.2 A simpler proof for average distortion embedding

We now give an average distortion version of Goemans' theorem using a simple squared-length distribution on the difference vectors $(x_i - x_j)$'s in the Lemma 2.2. Interestingly, this can be modified to weighted averages and gives yet another proof of Goemans' worst-case distortion result, although non-constructively.

▶ **Theorem 4.1.** *Let $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ be points satisfying ℓ_2^2 triangle inequalities. Then there exists an ℓ_2^2-to-ℓ_2 embedding $x_i \mapsto g(x_i)$ with* average *distortion $\sqrt{d}$, that is,*

$$\|g(x_i) - g(x_j)\|_2 \leq \|x_i - x_j\|_2^2, \quad \textit{for all } i, j,$$

$$\textit{and} \qquad \frac{1}{\sqrt{d}} \; \sum_{i<j} \|x_i - x_j\|_2^2 \leq \sum_{i<j} \|g(x_i) - g(x_j)\|_2$$

Proof. Let $\{p_{kl}\}_{k<l}$ define a probability distribution with $p_{kl} \propto \|x_k - x_l\|_2^2$. Given this distribution, let P be the symmetric psd matrix defined as $P := \sum_{k<l} p_{kl}\,(x_k - x_l)(x_k - x_l)^T \in \mathbb{R}^{d\times d}$. Consider the embedding that maps x_i to $g(x_i) := P^{1/2}x_i$. The embedding is a contraction by Lemma 2.2.

Now let's bound the average distortion. First, note that:

$$\|g(x_i) - g(x_j)\|_2 = \left\|P^{1/2}(x_i - x_j)\right\|_2 \geq \frac{\|x_i - x_j\|_2^2}{\left\|P^{-1/2}(x_i - x_j)\right\|_2},$$

where the inequality follows from the Cauchy-Schwarz inequality.

Summing over all pairs i, j and using the definition of p_{ij} we have

$$\begin{aligned}
\sum_{i<j} \|g(x_i) - g(x_j)\|_2 &\geq \left(\sum_{k<l} \|x_k - x_l\|_2^2\right) \sum_{i<j} \frac{p_{ij}}{\sqrt{(x_i - x_j)^T P^{-1}(x_i - x_j)}} \\
&\geq \left(\sum_{k<l} \|x_k - x_l\|_2^2\right) \left(\sum_{i<j} p_{ij}\ (x_i - x_j)^T P^{-1}(x_i - x_j)\right)^{-1/2} \\
&\qquad\qquad\qquad\qquad\text{[by Jensen's inequality]} \\
&= \left(\sum_{k<l} \|x_k - x_l\|_2^2\right) \left(\operatorname{Tr}\left(P^{-1/2}PP^{-1/2}\right)^{-1/2}\right) \\
&= \left(\sum_{k<l} \|x_k - x_l\|_2^2\right) \operatorname{Tr}\left(\mathbb{I}_d\right)^{-1/2} \\
&= \frac{1}{\sqrt{d}} \sum_{i<j} \|x_i - x_j\|_2^2 .
\end{aligned}$$

We note that if P is not invertible then the same proof can be carried out using pseudo-inverse of P instead. ◀

▶ Remark. Although an enclosing ellipsoid of approximately optimal volume can be computed by a convex program [12], the proof of Theorem 1.1 requires a stronger, spectral approximation to the quadratic form of the minimum enclosing ellipsoid. We are not aware of any efficient algorithms for this. On the other hand, sampling (i, j) with probability $\propto \|x_i - x_j\|_2^2$ can be done in $O(nd)$ time as follows. First we compute the mean $\mu = \sum_{i=1}^n x_i/n$, and all the marginals for $(i, .)$ using

$$\sum_{j=1}^n \|x_j - x_i\|_2^2 = \sum_{j=1}^n \|x_j - \mu\|_2^2 + n\,\|\mu - x_i\|_2^2 .$$

Now we can first sample i from the marginals, and then sample j with probability $\propto \|x_i - x_j\|_2^2$. This takes $O(nd)$ time in total.

Theorem 4.1 immediately gives an efficient $\sqrt{d}$ approximation algorithm for Uniform Sparsest Cut when the SDP optimum solution resides in $\mathbb{R}^d$. Furthermore, as we point out next, the same proof can be tweaked to yield a similar result for the general Sparsest Cut problem.

▶ **Theorem 4.2** (Sparsest Cut SDP rounding in dimension d)**.** *A Sparsest Cut instance C, D with SDP optimum solution in $\mathbb{R}^d$ has an integrality gap of at most $\sqrt{d}$.*

Proof. Let $x_1, \dots x_n$ be the optimum solution in $\mathbb{R}^d$ to the Sparsest Cut SDP. We slightly modify the embedding given in the proof of Theorem 4.1, by choosing the p_{ij}'s based on the demand graph D. Let $P = \sum_{k<l} p_{kl}\ (x_k - x_l)(x_k - x_l)^T \in \mathbb{R}^{d \times d}$, where p_{kl}'s define a probability distribution with $p_{kl} \propto d_{kl} \left\|x_k - x_l\right\|_2^2$. We define the embedding as $x_i \mapsto g(x_i) = P^{1/2} x_i$. Lemma 2.2 shows that it is a contraction. We now need to show $\sum_{i<j} d_{ij} \left\|g(x_i) - g(x_j)\right\|_2 \geq \frac{1}{\sqrt{d}} \sum_{i<j} d_{ij} \left\|x_i - x_j\right\|_2^2$. It is easy to check that the same proof goes through without any major changes. ◀

By a well-known duality (cf. [16, Proposition 15.5.2 and Exercise 4]), Theorem 4.2 also implies Goemans' worst-case distortion result (Theorem 1.1), although non-constructively.

Acknowledgements. We thank Luca Trevisan for helpful discussions and suggestions, in particular, for bringing to our attention that Goemans' Theorem was true even with approximate triangle inequalities.

References

1 Martin Aigner and Günter M. Ziegler. *Proofs from THE BOOK*. Springer, 4th edition, 2009. doi:10.1007/978-3-662-44205-0.

2 Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21:1–21, 2008. (Preliminary version in *37th STOC*, 2008). doi:10.1090/S0894-0347-07-00573-5.

3 Sanjeev Arora, Satish Rao, and Umesh V. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2), 2009. (Preliminary version in *36th STOC*, 2004). doi:10.1145/1502793.1502794.

4 Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Proc. 51st IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 472–481, 2011. arXiv:1104.4680, doi:10.1109/FOCS.2011.95.

5 Jean Bourgain and Lior Tzafriri. Invertibility of large submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel Journal of Mathematics*, 57(2):137–224, 1987. doi:0.1007/BF02772174.

6 Jeff Cheeger, Bruce Kleiner, and Assaf Naor. A $(\log n)^{\Omega(1)}$ integrality gap for the sparsest cut SDP. In *Proc. 50th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 555–564, 2009. arXiv:0910.2024, doi:10.1109/FOCS.2009.47.

7 Amit Deshpande and Rakesh Venkat. Guruswami-Sinop rounding without higher level Lasserre. In Klaus Jansen, José D. P. Rolim, Nikhil R. Devanur, and Cristopher Moore, editors, *Proc. 17th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, volume 28 of *LIPIcs*, pages 105–114. Schloss Dagstuhl, 2014. arXiv:1406.7279, doi:10.4230/LIPIcs.APPROX-RANDOM.2014.105.

8 Shayan Oveis Gharan and Luca Trevisan. Improved ARV rounding in small-set expanders and graphs of bounded threshold rank. 2013. arXiv:1304.2060.

9 Venkatesan Guruswami and Ali Kemal Sinop. Approximating non-uniform sparsest cut via generalized spectra. In *Proc. 24th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 295–305, 2013. arXiv:1112.4109, doi:10.1137/1.9781611973105.22.

10 Piotr Indyk and Jirí Matoušek. Low-distortion embeddings of finite metric spaces. In Jacob E. Goodman and Joseph O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, pages 177–196. Chapman and Hall/CRC, 2nd edition, 2004. doi:10.1201/9781420035315.ch8.

11 Daniel M. Kane and Raghu Meka. A PRG for Lipschitz functions of polynomials with applications to sparsest cut. In *Proc.* 45*th ACM Symp. on Theory of Computing (STOC)*, pages 1–10, 2013. `arXiv:1211/1109`, `doi:10.1145/2488608.2488610`.

12 Leonid G. Khachiyan. Rounding of polytopes in the real number model of computation. *Math. Oper.Res.*, 21(2):307–320, 1996. `doi:10.1287/moor.21.2.307`.

13 Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved Cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *Proc.* 45*th ACM Symp. on Theory of Computing (STOC)*, pages 11–20, 2013. `arXiv:1301.5584`, `doi:10.1145/2488608.2488611`.

14 Nathan Linial. Finite metric spaces: combinatorics, geometry and algorithms. In *Proc. of the ICM, Beijing*, volume 3, pages 573–586, 2002. `arXiv:math/0304466`.

15 Avner Magen and Mohammad Moharrami. On the nonexistence of dimension reduction for ℓ_2^2 metrics. In *Proc.* 20*th Annual Canadian Conf. on Comp. Geom.*, 2008. URL: `http://cccg.ca/proceedings/2008/paper37full.pdf`.

16 Jirí Matoušek. Embedding finite metric spaces into normed spaces. In *Lectures on Discrete Geometry*, Graduate Texts in Mathematics, chapter 5, pages 355–400. Springer, 2002. `doi:10.1007/978-1-4613-0039-7_15`.

17 Joel A. Tropp. Column subset selection, matrix factorization, and eigenvalue optimization. In *Proc.* 20*th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 978–986, 2009. URL: `http://dl.acm.org/citation.cfm?id=1496770.1496876`, `arXiv:0806.4404`.