Algorithms and Hardness for Subspace Approximation

Amit Deshpande* Madhur Tulsiani† Nisheeth K. Vishnoi‡

November 18, 2010

Abstract

The subspace approximation problem $\mathsf{Subspace}(k,p)$ asks for a k-dimensional linear subspace that fits a given set of m points in $\mathbb{R}^n$ optimally. The error for fitting is a generalization of the least squares fit and uses the ℓ_p norm of the (ℓ_2) distances of the points from the subspace. Most of the previous work on subspace approximation has focused on small or constant k and $p = 1$ or ∞, using coresets and sampling techniques from computational geometry.

In this paper, extending another line of work based on convex relaxation and rounding, we give a polynomial time algorithm, *for any k and any $p \geq 2$*.This extends a result of Varadarajan, Venkatesh, Ye and Zhang [23], who gave an $O(\sqrt{\log m})$ approximation for all k and $p = \infty$. The approximation guarantee of our algorithm is roughly $\sqrt{2}\gamma_p$ where $\gamma_p \approx \sqrt{p/e}(1 + o(1))$ is the p^th norm of a standard normal random variable. The approximation ratio improves to γ_p in the interesting special case when $k = n - 1$. We also show that the convex relaxation we use has an integrality gap (or "rank gap") of $\gamma_p(1 - \varepsilon)$, for any constant $\varepsilon > 0$.

We also study the hardness of approximating this problem. We show that assuming the Unique Games Conjecture, the subspace approximation problem is hard to approximate within a factor better than $\gamma_p(1 - \varepsilon)$, for any constant $\varepsilon > 0$. The hardness reduction involves a dictatorship test which is somewhat different from "long-code" based tests used in reductions from Unique Games, and seems better suited for problems of a continuous nature.

Keywords: approximation algorithms, convex programming, unique games

*Microsoft Research India. `amitdesh@microsoft.com`.

†Institute for Advanced Study. `madhurt@math.ias.edu`. This material is based upon work supported by the National Science Foundation under grant No. CCF-0832797 and IAS Sub-contract no. 00001583. Work partly done while visiting Microsoft Research India.

‡Microsoft Research India. `nisheeth.vishnoi@gmail.com`

1 Introduction

Large data sets that arise in data mining, machine learning, statistics and computational geometry problems are naturally modeled as sets of points in a high-dimensional Euclidean space. Even though these points live in a high-dimensional space, in practice they are observed to have low intrinsic dimension and it is an algorithmic challenge to capture their underlying low-dimensional structure. The subspace approximation problem described below generalizes several problems formulated in this context.

Subspace(k,p): Given points $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$, an integer k, with $0 \leq k \leq n$, and $p \geq 1$, find a k-dimensional linear subspace that minimizes the sum of p-th powers of Euclidean distances of these points to the subspace, or equivalently,

$$\underset{V \,:\, \dim(V)=k}{\text{minimize}} \quad \left(\sum_{i=1}^{m} d(a_i, V)^p \right)^{1/p}.$$

Note that, here, ℓ_p norm is used as a function of $(d(a_1, V), d(a_2, V), \ldots, d(a_m, V))$; the individual distances $d(a_i, V)$ are the usual ℓ_2 distances.

We describe below the special cases of the subspace approximation problem which have been studied previously and the known results about them.

1. **Low-rank matrix approximation problem or the least squares fit ($p = 2$)**: Given a matrix $A \in \mathbb{R}^{m\times n}$ and $0 \leq k \leq n$, the matrix approximation problem is to find another matrix $B \in \mathbb{R}^{m\times n}$ of rank at most k that minimizes the Frobenius (also known as Hilbert-Schmidt) norm of their difference $\|A - B\|_F \stackrel{\text{def}}{=} \left(\sum_{ij}(A_{ij} - B_{ij})^2\right)^{1/2}$. Taking the rows of A to be points $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$, the above problem is equivalent to the problem Subspace(k,2). Elementary linear algebra shows that the optimal subspace is spanned by the top k right singular vectors of A, which can be found in time $O(\min\{mn^2, m^2n\})$ using Singular Value Decomposition (SVD) [11].

2. **Computing radii of point sets ($p = \infty$)**: Given points $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$, their *outer* $(n-k)$*-radius* is defined as the minimum, over all k-dimensional linear subspaces, of the maximum Euclidean distance of these points to the subspace (which is equivalent to Subspace(k,∞)). Gritzmann and Klee initiated the study of this quantity in computational convex geometry [12] and gave a polynomial time algorithm for the minimum enclosing ball problem (or the problem Subspace(0,∞)).

   (a) *For small k:* Bădoiu, Har-Peled and Indyk [4] gave a $(1+\varepsilon)$-approximation algorithm running in polynomial time for the minimum enclosing cylinder problem (equivalent to Subspace(1,∞)), which was further extended by Har-Peled and Varadarajan [13] to Subspace(k,∞) for constant k.

   (b) *For large k:* Brieden, Gritzmann and Klee [3] showed that it is NP-hard to approximate the width of a point set (equivalent to Subspace($n-1$,∞)) within any constant factor. From the algorithmic side, the results by Nesterov [18] and Nemirovski, Roos and Tarlaky [17] on quadratic optimization imply $O(\sqrt{\log m})$-approximation for Subspace($n-1$,∞) in polynomial time. Building on these techniques, Varadarajan, Venkatesh, Ye and Zhang [23] gave a polynomial time $O(\sqrt{\log m})$-approximation algorithm for Subspace(k,∞), for any k. On the hardness side, they proved that there exists a constant $\delta > 0$ such that, for any $0 < \varepsilon < 1$ and $k \leq n - n^{\varepsilon}$, there is no polynomial time algorithm that gives $(\log m)^{\delta}$-approximation for Subspace(k,∞) unless NP $\subseteq$ DTIME $\left(2^{\mathrm{polylog}(n)}\right)$.

3. **Other values of** p: For general p and constant k, a result of Shyamalkumar and Varadarajan [20] and subsequent work by Deshpande and Varadarajan [7] gave a $(1+\varepsilon)$-approximation algorithm with running time $O(mn \cdot \exp(k, p, 1/\varepsilon))$. The running time was recently improved to $O(mn \cdot \text{poly}(k, 1/\varepsilon) + (m+n) \cdot \exp(k, 1/\varepsilon))$ by Feldman, Monemizadeh, Sohler and Woodruff [10], for the case $p = 1$.

For $p \neq 2$, we do not know any suitable generalization of SVD, and therefore, have no exact characterization of the optimal subspace. The approximation techniques used so far to overcome this are: (i) coresets and sampling-based techniques: which give nearly optimal approximations but only for small or constant k and p. (ii) convex relaxations and rounding: which give somewhat sub-optimal approximations mostly for large values of k; the only exception is the result of Varadarajan, Venkatesh, Ye and Zhang [23] which works for any k (but only for $p = \infty$).

Our work

In this paper, we study the problem $\mathsf{Subspace}(k,p)$ for $p < \infty$, about which little is known in general. One motivation for doing so is that often the case $p < \infty$ gives significantly better approximation guarantees and requires somewhat different techniques to analyze than $p = \infty$. This is evident in the work for subspace approximation for small k ([7] and [10] for $p < \infty$ versus [4] and [13] for $p = \infty$) and in the work on regression ([5] and [6] versus the $p = \infty$ case which is solvable by fixed dimensional linear programming). Also, in the study of hardness of approximation, the case $p = \infty$ can often be reduced to a discrete problem; while the case $p < \infty$ is inherently of a more continuous nature, and requires somewhat different techniques.

On the algorithmic side, we give a factor $\gamma_p \cdot \sqrt{2 - (1/n-k)}$ approximation algorithm for the problem $\mathsf{Subspace}(k,p)$ in $\mathbb{R}^n$, where $\gamma_p \approx \sqrt{p/e}(1 + o(1))$ is the p^{th} norm of a standard Gaussian. Our algorithm is based on a convex relaxation, similar to the semi-definite relaxations used in [17] and [23] for $p = \infty$. We give a tighter analysis for general p. We also exhibit gap instance for the convex program. We show a gap of factor γ_p for $\mathsf{Subspace}(k,p)$ (when k is superconstant) showing that our analysis is tight up to the factor of $\sqrt{2 - (1/n-k)}$.

We also investigate the hardness of approximation for $\mathsf{Subspace}(k,p)$. We give a reduction from the Unique Label Cover problem of Khot [14] to the problem of approximating $\mathsf{Subspace}(n-1,p)$ within a factor γ_p (which can trivially be extended to a reduction to $\mathsf{Subspace}(k,p)$ for $k = n^{\Omega(1)}$). The reduction is related to the ones used for similar geometric problems in [16], [15] and [2]. However, an interesting difference here in comparison to usual reductions is that we use a different (real-valued) encoding of the assignment to Unique Label Cover (in terms of the Fourier coefficients of the long-code instead of the truth table) which is more natural in our context. This may also be useful for other problems of a continuous nature.

Other related problems

L_p**-Grothendieck problem.** In the $k = n-1$ case, subspace approximation problem can be rewritten as $\min_{\|z\|_2=1} \|Az\|_p$, where the rows of $A \in \mathbb{R}^{m \times n}$ represent the points $a_1, a_2, \ldots, a_m$ and $z \in \mathbb{R}^n$ represents the unit normal to the subspace we are asked to find. When A is invertible, this problem can be shown (using duality in Banach spaces) to be equivalent to a special case of the L_p-Grothendieck problem (introduced by Kindler, Naor and Schechtman [16]) which asks for maximizing $x^T M x$ subject to $\|x\|_p \leq 1$. $\mathsf{Subspace}(n-1, p)$ with invertible A, reduces to this problem with $M = (A^{-1})^T A^{-1}$.

In this special case, using Grothendieck's inequality and a technique by Alon and Naor [1], one can get $O(1)$-approximation. Moreover, in this case, the above problem is also equivalent to finding diameters of convex bodies given by $\|Ax\|_p \leq 1$ and computing $p \mapsto 2$ norm of the matrix A^{-1}.

l_p-regression problem. In the l_p regression problem, we are given an $m \times n$ matrix A and a vector $b \in \mathbb{R}^m$, and the goal is to minimize $||Az - b||_p$ over all $z \in \mathbb{R}^n$. This is clearly related to subspace approximation with $k = n - 1$, but the fact that z is unconstrained makes it a convex optimization problem. Efficient approximation algorithms for the regression problem are given by Clarkson [5] for $p = 1$, Drineas, Mahoney, and Muthukrishnan [8] for $p = 2$, and Dasgupta et al. [6] for $p \geq 1$. It is not clear that these results can be employed fruitfully for the subspace approximation problem for $k = n - 1$ where it is required that $\|z\| \geq 1$.

2 Preliminaries and Notation

Throughout this paper, $\|\cdot\|_p$ denotes the p-norm. Norms of vectors are taken with respect to the counting measure and of functions are taken with respect to the uniform probability measure on their domain. When the subscript is unspecified, $\|\cdot\|$ denotes $\|\cdot\|_2$.

2.1 The Subspace Approximation Problem

We will use a formulation of the problem $\mathsf{Subspace}(k,p)$ for points $a_1, \ldots, a_m$, in terms of the orthogonal complement of the desired subspace V. Let $z_1, \ldots, z_{n-k}$ be an orthonormal basis for the orthogonal complement and let $Z \in \mathbb{R}^{n\times(n-k)}$ denote the matrix with the j^{th} column $Z^{(j)} = z_j$. Then $d(a_i, V) = \left\|a_i^T Z\right\|_2$ and the problem of finding (the orthogonal complement of) the subspace can be stated as

$$\begin{array}{ll} \text{minimize} & \left(\displaystyle\sum_{i=1}^{m} \left\|a_i^T Z\right\|_2^p\right)^{1/p} \\ \text{subject to:} & \left\|Z^{(j)}\right\| \geq 1 \ \forall j \in \{1, \ldots, n-k\} \\ & \left\langle Z^{(j_1)}, Z^{(j_2)}\right\rangle = 0 \ \forall j_1 \neq j_2, \quad Z \in \mathbb{R}^{n\times(n-k)} \end{array}$$

For the hardness results we shall be concerned with the special case of the problem with $k = n-1$. For points $a_1, \ldots, a_m \in \mathbb{R}^n$, let A be $m \times n$ matrix with $A_i = a_i^T$. The problem $\mathsf{Subspace}(n-1,p)$ is then simply to minimize $\|Az\|_p$ for $z \in \mathbb{R}^n$, subject to $\|z\|_2 \geq 1$.

Remark 2.1 It is easy to check that (by a change of variable and suitable modification of A) both the norms can be taken to be with respect to an arbitrary measure instead of the counting measure. In particular, if $A \in \mathbb{R}^{m\times n}$, the p-norm is taken with respect to a measure μ on $[m]$ and the 2-norm with respect a measure ν on $[n]$, then we change variables to $\tilde{z}$ with $\tilde{z}_j = \sqrt{\nu(j)}z_j$ and modify A_{ij} to $A_{ij}(\mu(i))^{1/p}/\sqrt{\nu(j)}$ to get an equivalent problem with norms according to the counting measure.

2.2 Bernoulli and Gaussian Random Variables

A Bernoulli random variable is a discrete random variable taking values in $\{-1, 1\}$ with probability $1/2$ each. A standard normal random variables (or 1-dimensional Gaussian) is a continuous random variable with probability density function $1/\sqrt{2\pi}\cdot\exp(-x^2/2)$. We use γ_p to denote the p^{th} moment of $N(0,1)$,

$$\gamma_p \stackrel{\text{def}}{=} \left(\int_{-\infty}^{\infty} |x|^p \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx\right)^{1/p} = \left(\frac{2^{p/2}\cdot\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}\right)^{1/p} \approx \sqrt{\frac{p}{e}}(1+o(1)).$$

We shall require both upper and lower bounds on moments of a sum of Bernoulli random variables by the moment of an appropriate Gaussian. The following upper bound is one direction of the Khintchine inequality (see [19]) well-known in functional analysis.

Claim 2.2 *Let $x_1, \ldots, x_R$ be independent Bernoulli random variables and let $c_1, \ldots, c_R \in \mathbb{R}$ and $\|\mathbf{c}\| = \sqrt{c_1^2 + \cdots + c_R^2}$. Then for any positive $p > 0$,*

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_R}\left[\left(\sum_{i=1}^{R} c_i \cdot x_i\right)^p\right] \quad \leq \quad \gamma_p^p \cdot \|\mathbf{c}\|^p$$

The following version of the reverse direction, when all c_i's are much smaller than $\|\mathbf{c}\|$, can be derived using the Berry-Esseen Theorem (as in [22]). A proof of the statement below appears in [16] (as Lemma 2.5).

Claim 2.3 *Let $x_1, \ldots, x_R$ be independent Bernoulli random variables and let $c_1, \ldots, c_R \in \mathbb{R}$ be such that for all $i \in [R]$, $|c_i| \leq \tau \cdot \|\mathbf{c}\|$ for $\tau \in (0, e^{-4})$. Then, for any $p \geq 1$,*

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_R}\left[\left|\sum_{i=1}^{R} c_i \cdot x_i\right|^p\right] \quad \geq \quad \gamma_p^p \cdot \|\mathbf{c}\|^p \cdot \left(1 - 10\tau \cdot (\log(1/\tau))^{p/2}\right).$$

3 Technical Overview

In this section we describe our results and give a general outline of the sections that follow.

3.1 Algorithm for Subspace(k,p)

We formulate problem Subspace(k,p) for points $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$ in terms of the orthogonal complement of the desired subspace V. Let $Z \in \mathbb{R}^{n \times (n-k)}$ be the matrix whose columns form an orthonormal basis for the orthogonal complement of V, then the distance of a point a_i from V can be written as $d(a_i, V) = \left\|a_i^T Z\right\|_2 = (a_i^T Z Z^T a_i)^{1/2}$. Note that ZZ^T is a positive semidefinite (p.s.d.) matrix of rank $n-k$, all of whose nonzero singular values are 1 and whose singular vectors (the columns of Z) specify the (complement of the) subspace V.

A convex relaxation of Subspace(k,p) is then obtained by optimizing over arbitrary positive semidefinite matrices X and replacing the requirement that the matrix have rank $n-k$ by a condition on the trace of X (see Figure 1). This is similar to the relaxations used in [17, 23]. The problem then reduces to giving a "rounding algorithm" which reduces the rank of the matrix X (which might be as large as n) to $n-k$, and achieves a good approximation of the objective value of the convex program.

In keeping with the intuition that the singular vectors of ZZ^T span the orthogonal complement of V, our algorithm looks at the singular vectors of the matrix X obtained by solving the convex relaxation. It then divides the singular vectors into $n-k$ "bins", and constructs one vector for each bin by taking a random linear combination of vectors within each bin.

Our algorithm described in Section 4 achieves an approximation ratio of $\gamma_q \cdot (2 - \frac{1}{n-k})^{1/2}$ for Subspace(k,p), where $q = 2 \cdot \lceil p/2 \rceil$. (See Theorem 4.4.)

We remark that the problem of obtaining low-rank solutions to a semidefinite program was also considered by [21], and was addressed by simply taking random (chosen according to a Gaussian) linear combinations of the singular vectors of the relevant matrix. However, in their case, they were' only interested in satisfying the constraints, with an error depending inversely on the rank parameter. In our case, we require a rank $n-k$ positive semidefinite matrix, all of whose eigenvalues are exactly 1. Since the only constraint enforcing this is a constraint on the trace of the matrix, even a small multiplicative error in satisfying the constraint can make some singular values quite small. To resolve this, we proceed by dividing the singular vectors in various bins and take Bernoulli linear combinations, do directly generate the orthogonal singular vectors.

3.2 A gap instance

In Section 5, we show that the convex relaxation we use has an integrality gap, or more correctly "rank gap", of $\gamma_p(1-\varepsilon)$, for any constant $\varepsilon > 0$. Given any constant $\varepsilon > 0$, we construct points $b_1, b_2, \ldots, b_m \in \mathbb{R}^n$ such that the optimum for $\mathsf{Subspace}(n-1,p)$ on these points (a rank 1 p.s.d. matrix) and the optimum for its corresponding convex relaxation (a rank n p.s.d. matrix) are at least a factor of $\gamma_p(1-\varepsilon)$ apart. We first show such a gap for the continuous analog of $\mathsf{Subspace}(n-1,p)$ where the point set is the entire $\mathbb{R}^n$ equipped with Gaussian measure (Theorem 5.1). We then discretize this example to get our final integrality gap construction (Theorem 5.2).

This also gives a gap of factor $\gamma_p(1-\varepsilon)$ for $\mathsf{Subspace}(k,p)$ for any super-constant $k = k(n)$, since an instance of $\mathsf{Subspace}(n-1,p)$ in $\mathbb{R}^n$ can be trivially converted (by adding extra zero coordinates) to an instance of $\mathsf{Subspace}(k,p)$ in $\mathbb{R}^{n'}$ with $k(n') = n-1$.

3.3 Unique-Games hardness

In Section 6, we describe a reduction from Unique Label Cover to the problem of approximating $\mathsf{Subspace}(n-1,p)$ within a factor better than γ_p (for a constant $p \geq 1$). By a trivial reduction from $\mathsf{Subspace}(n-1,p)$ to $\mathsf{Subspace}(k,p)$ for any $k = n^{\Omega(1)}$, this gives the hardness of approximating $\mathsf{Subspace}(k,p)$ better than γ_p, assuming the Unique Games Conjecture.

To understand the intuition for the reduction, let us consider the simpler problem of testing whether a given function $f : \{-1,1\}^R \to \{-1,1\}$ is a "dictator" i.e. $f(x_1, \ldots, x_R) = x_i$ for some $i \in [R]$, which is a useful primitive in such reductions. The problem is to design an instance $\mathcal{I}$ of $\mathsf{Subspace}(n-1,p)$ and interpret the description of f as a solution to $\mathcal{I}$. The required property is that if f is a dictator then the corresponding subspace fits the points in $\mathcal{I}$ with small error. On the other hand, if f is "far from being a dictator", the error is required to be larger by a factor of γ_p.

In most reductions, f is assumed to be described by its truth table. However, if we want to interpret the input simply as the coordinates of a vector z, there is no way to enforce that the coordinates be boolean. It turns out to be more convenient if we require f as a list of its Fourier coefficients which can be thought of as a vector with arbitrary real numbers coordinates and norm 1 (since $\mathbb{E}[f^2] = 1$). Also, since we are only interested in dictator functions, it is sufficient to ask for the "level 1" Fourier coefficients $\hat{f}(\{1\}), \ldots, \hat{f}(\{R\})$.

In particular, consider the input being described by R real numbers $b_1, \ldots, b_R$ such that $\sum_i b_i^2 = 1$ and we think of it as describing the function $f_b(x_1, \ldots, x_R) = b_1 \cdot x_1 + \cdots + b_R \cdot x_R$. We also think of $b_1, \ldots, b_R$ as the normal to some $R-1$ dimensional subspace of $\mathbb{R}^R$. Let the points be given by $(1/2^{R/p}) \cdot x$ for each vector $x \in \{-1,1\}^R$, so that the objective of the subspace approximation problem is exactly $\|f_b\|_p$. If f is a dictator, i.e., one of the b_i's is 1 and others 0, then $\|f_b\|_p = 1$. Also, if it is far from a dictator in the sense that $\max_i b_i \leq \tau$ for a small constant τ, then $\|f_b\|_p \approx \gamma_p$ by Claim 2.3.

Translating this intuition to a reduction from Unique Label Cover turns out to be slightly technical due to the fact that we need to consider one function for each vertex of Unique Label Cover and all bounds on norms do not hold for individual functions but only on average. Similar technicalities arise when working with the ℓ_p norm in [16] (though they specify functions by their truth tables).

4 Approximation Algorithm via Convex Programming

To relax the minimization problem for Subspace(k,p) to a convex problem, we rewrite the distances $\left\|a_i^T Z\right\|$ in the objective as $(a_i^T Z Z^T a_i)^{1/2}$. Noting that ZZ^T is a positive semidefinite matrix of rank $n-k$, we get the following natural relaxation similar to the one used in [17, 23].

Minimization Problem	Convex Relaxation
minimize $\left(\sum_{i=1}^{m} \left\|a_i^T Z Z^T a_i\right\|^{p/2}\right)^{1/p}$	minimize $\left(\sum_{i=1}^{m} \left\|a_i^T X a_i\right\|^{p/2}\right)^{1/p}$
subject to: $\left\|Z^{(j)}\right\| \geq 1 \; \forall j \in \{1,\ldots,n-k\}$	subject to: $\mathsf{Tr}(X) \geq n-k$
$\left\langle Z^{(j_1)}, Z^{(j_2)} \right\rangle = 0 \; \forall j_1 \neq j_2$	$I \succcurlyeq X \succcurlyeq 0$
$Z \in \mathbb{R}^{n \times (n-k)}$	$X \in \mathbb{R}^{n \times n}$

Figure 1: The problem Subspace(k, p) and its convex relaxation

Note that this relaxation removes the constraint on the rank and relaxes the constraint on the length of the individual vectors $Z^{(j)}$ to the trace of entire matrix X. Also, the objective function is written as $\left(\sum_i \left|a_i^T X a_i\right|^{p/2}\right)^{1/p}$ which is not convex. However, for solving the convex program, we can work with $\sum_i \left|a_i^T X a_i\right|^{p/2}$, which is convex for $p \geq 2$.

In Figure 2, we give a "rounding algorithm" for the relaxation. Note that the problem here is not really to round the solution to an integer solution as with most convex relaxations, but instead to *reduce the rank* of the solution to the program, while obtaining a good approximation of the objective.

<u>Input</u>: A matrix $X \in \mathbb{R}^{n \times n}$ satisfying $I \succeq X \succeq 0$ and $\mathsf{Tr}(X) \geq n-k$.

1. Express X in terms of its singular vectors as $X = \sum_{t=1}^{r} \lambda_t x_t x_t^T$ where the vectors $x_1, \ldots, x_r$ form an orthonormal set and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$.
2. Partition $[r]$ into $n-k$ subsets $S_1, \ldots, S_{n-k}$. Start with $S_1 = \cdots = S_{n-k} = \emptyset$. Then for t from 1 to r do:
   (a) Find the set S_j for which $\sum_{t' \in S_j} \lambda_{t'}$ is minimum.
   (b) Set $S_j := S_j \cup \{t\}$.
3. Pick r independent Bernoulli variables $b_1, \ldots, b_r \in_R \{-1, 1\}$. For each $j \in [n-k]$, let $y_j \stackrel{\text{def}}{=} \sum_{t \in S_j} b_t \cdot \sqrt{\lambda_t} \cdot x_t$.
4. Output the matrix $Z \in \mathbb{R}^{n \times (n-k)}$ with $Z^{(j)} \stackrel{\text{def}}{=} \frac{y_j}{\|y_j\|}$.

Figure 2: The rank reduction algorithm

We shall show the algorithm outputs a matrix Z of rank $n-k$ which achieves an approximation ratio of $\gamma_p \cdot \sqrt{2 - 1/n-k}$ in expectation, for even integers $p \geq 2$. An approximation guarantee for other values of p can be obtained via Jensen's inequality. We state the dependence on $n-k$ precisely as we shall be interested in the case $n - k = 1$. For notational convenience, we shall use $\alpha_{n,k}$ to denote the quantity $\sqrt{2 - 1/n-k}$ in the rest of this section.

It is clear that the columns of the matrix Z given by the algorithm form an orthonormal set since they are all in the span of distinct eigenvectors of X, and are normalized to have length 1. However, this assumes that the lengths of the vectors y_j are nonzero. Since a vector y_j is a weighted sum of orthogonal vectors, $\|y_j\|^2 = \sum_{t \in S_j} \lambda_t$. The following claim gives a lower bound on this quantity which is also useful in bounding the approximation ratio.

Claim 4.1 *Let $S_1, \ldots, S_{n-k}$ be the partition constructed by the algorithm in step 2. Then*

$$\forall j \in [n-k], \quad \sum_{t \in S_j} \lambda_t \quad \geq \quad \frac{1}{\alpha_{n,k}^2}.$$

Proof: Let $j_0 \stackrel{\text{def}}{=} \operatorname{argmin}_j \left\{ \sum_{t \in S_j} \lambda_t \right\}$ and let $s^* \stackrel{\text{def}}{=} \sum_{t \in S_{j_0}} \lambda_t$. Let $Q \stackrel{\text{def}}{=} \{j_0\} \cup \{j \mid |S_j| > 1\}$. Note that the algorithm ensures that $|S_j| > 0$ for all j but in T we discard the singleton sets. We will show that $s^* \geq 1/\left(2 - 1/|Q|\right)$, which will prove the claim since $|Q| \leq n - k$.

We argue that for each $j \in Q$, $j \neq j_0$, $\sum_{t \in S_j} \lambda_t \; \leq \; 2s^*$. To see this, let t_j be the maximal index in S_j. At step $t = t_j$, t_j was added to set S_j and not to the set S_{j_0}. Hence,

$$\sum_{t \in S_j, t < t_j} \lambda_t \;\leq\; \sum_{t \in S_{j_0}, t < t_j} \lambda_t \;\leq\; s^*.$$

Also, there exists at least one $t_0 \in S_{j_0}$ such that $t_0 < t_j$. This is because S_j was non-empty at step t_j (otherwise it would be a singleton). But then $\lambda_{t_j} \leq \lambda_{t_0} \leq s^*$ and, hence, $\sum_{t \in S_j} \lambda_t \leq 2s^*$.

Finally, we note that for each $j \notin Q$, S_j contains exactly one element t, the eigenvalue λ_t corresponding to which is at most 1. Thus,

$$(|Q| - 1) \cdot 2s^* + s^* + (n - k - |Q|) \cdot 1 \quad \geq \quad \sum_{t \in [r]} \lambda_t \quad \geq \quad n - k,$$

which completes the proof. ■

The following lemma proves the required approximation guarantee for the expected p^{th} moment of the distance a single point a_i from the orthogonal complement of the column span of Z.

Lemma 4.2 *Let X be the solution of the convex relaxation and let Z be the matrix returned by the algorithm. Also, let p be even. Then, for each $i \in [m]$*

$$\mathop{\mathbb{E}}_{Z} \left[\left\| a_i^T Z \right\|_2^p \right] \quad \leq \quad \gamma_p^p \cdot \alpha_{n,k}^p \cdot \left(a_i^T X a_i \right)^{p/2}.$$

Proof: We can expand $\left\| a_i^T Z \right\|$, using W_j to denote $\left\langle a_i, Z^{(j)} \right\rangle$, as

$$\mathop{\mathbb{E}}_{Z} \left[\left\| a_i^T Z \right\|_2^p \right] \;=\; \mathop{\mathbb{E}}_{Z} \left[\left(\sum_{j=1}^{n-k} \left\langle a_i, Z^{(j)} \right\rangle^2 \right)^{p/2} \right] \;=\; \mathbb{E} \left[\left(\sum_{j=1}^{n-k} W_j^2 \right)^{p/2} \right]$$

Note that the W_j-s are independent random variables since each W_j only depends on b_t such that $t \in S_j$, and the sets are disjoint. Using the multinomial expansion and the fact that p is even, the above can be written as

$$\mathbb{E}\left[\left(\sum_{j=1}^{n-k} W_j^2\right)^{p/2}\right] = \sum_{p_1,\ldots,p_{n-k}} \binom{p/2}{p_1,\ldots,p_{n-k}} \mathbb{E}\left[\prod_j W_j^{2p_j}\right]$$
$$= \sum_{p_1,\ldots,p_{n-k}} \binom{p/2}{p_1,\ldots,p_{n-k}} \left(\prod_j \mathbb{E}[W_j^{2p_j}]\right).$$

The following claim then finishes the proof.

Claim 4.3 $\mathbb{E}\left[W_j^{2p_j}\right] \leq \gamma_p^{2p_j} \cdot \left(\frac{\sum_{t\in S_j} \lambda_t \langle a_i, x_t\rangle^2}{\sum_{t\in S_j}\lambda_t}\right)^{p_j}$.

Proof: The proof follows an application of upper bound on a sum Bernoulli variables derived in Claim 2.2. We expand $\mathbb{E}[W_j^{2p_j}]$ as

$$\mathbb{E}\left[W_j^{2p_j}\right] = \mathbb{E}\left[\left(\frac{\left\langle a_i, \sum_{t\in S_j} b_t\cdot\sqrt{\lambda_t}\cdot x_t\right\rangle}{\left\|\sum_{t\in S_j} b_t\cdot\sqrt{\lambda_t}\cdot x_t\right\|}\right)^{2p_j}\right] = \frac{\mathbb{E}\left[\left(\sum_{t\in S_j} b_t\cdot\sqrt{\lambda_t}\cdot\langle a_i, x_t\rangle\right)^{2p_j}\right]}{\left(\sum_{t\in S_j}\lambda_t\right)^{p_j}}.$$

Claim 2.2 gives that $\mathbb{E}\left[\left(\sum_{t\in S_j} b_t\cdot\sqrt{\lambda_t}\cdot\langle a_i, x_t\rangle\right)^{2p_j}\right] \leq \gamma_{2p_j}^{2p_j}\cdot\left(\sum_{t\in S_j}\lambda_t\langle a_i, x_t\rangle^2\right)^{p_j}$ and noting that $\gamma_{2p_j} \leq \gamma_p$ (since $2p_j \leq p$) proves the claim. ■

For each j, let D_j denote $\sum_{t\in S_j}\lambda_t\langle a_i, x_t\rangle^2$ and let Λ_j denote $\sum_{t\in S_j}\lambda_t$. Using the above claim we get that

$$\mathop{\mathbb{E}}_Z\left[\left\|a_i^T Z\right\|_2^p\right] \leq \sum_{p_1,\ldots,p_{n-k}} \binom{p/2}{p_1,\ldots,p_{n-k}} \cdot \prod_j \left(\frac{D_j}{\Lambda_j}\right)^{p_j} \cdot \gamma_p^p = \left(\sum_j \frac{D_j}{\Lambda_j}\right)^{p/2} \cdot \gamma_p^p.$$

Claim 4.1 gives that ${}^1/_{\Lambda_j} \leq \alpha_{n,k}^2$. Also, we have that $\sum_j D_j = \sum_t \lambda_t\langle a_i, x_t\rangle^2 = a_i^T X a_i$. Combining these gives $\mathbb{E}_Z\left[\left\|a_i^T Z\right\|_2^p\right] \leq \gamma_p^p\cdot\alpha_{n,k}^p\cdot\left(a_i^T X a_i\right)^{p/2}$ which proves the lemma. ■

An approximation guarantee for other values of p can be obtained via a standard application of Jensen's Inequality. We state the dependence on $n-k$ precisely as we shall be interested in the case $n-k=1$ in the later sections. Notice that the approximation factor is γ_q, where $q = 2\cdot\lceil p/2\rceil$, in the case $n-k=1$, and thus matches the integrality gap and unique-games hardness that appear in the later sections.

Theorem 4.4 *Let X be the solution of the convex relaxation and let Z be the matrix returned by the algorithm. Let $p \geq 1$ and let $q = 2\cdot\lceil p/2\rceil$ be the smallest even integer such that $q \geq p$. Then,*

$$\mathop{\mathbb{E}}_Z\left[\left(\sum_{i=1}^m \left\|a_i^T Z\right\|_2^p\right)^{1/p}\right] \leq \gamma_q\cdot\sqrt{2-({}^1/_{n-k})}\cdot\left(\sum_{i=1}^m \left(a_i^T X a_i\right)^{p/2}\right)^{1/p}.$$

Proof: (Proof of Theorem 4.4) By the concavity of the function $f(u) = u^{1/p}$ and Jensen's Inequality we have that

$$\mathop{\mathbb{E}}_{Z}\left[\left(\sum_{i=1}^{m}\left\|a_i^T Z\right\|_2^p\right)^{1/p}\right] \quad\leq\quad \left(\mathop{\mathbb{E}}_{Z}\left[\sum_{i=1}^{m}\left\|a_i^T Z\right\|_2^p\right]\right)^{1/p},$$

and by linearity it suffices to consider a single term of the summation. Another application of Jensen's (using $p \leq q$) and Lemma 4.2 give that

$$\mathop{\mathbb{E}}_{Z}\left[\left\|a_i^T Z\right\|_2^p\right] \;=\; \mathop{\mathbb{E}}_{Z}\left[\left(\left\|a_i^T Z\right\|_2^q\right)^{p/q}\right] \;\leq\; \left(\mathop{\mathbb{E}}_{Z}\left[\left\|a_i^T Z\right\|_2^q\right]\right)^{p/q} \;\leq\; \gamma_q^p \cdot \alpha_{n,k}^p \cdot \left(a_i^T X a_i\right)^{p/2}$$

which completes the proof of the theorem. ∎

Remark 4.5 Our results are stated in terms of the expected approximation ratio achieved by the algorithm. However, one can get arbitrarily close to this ratio with high probability, simply by considering few independent runs of the algorithm and picking the best solution. In particular, one can achieve an approximation guarantee $(1+\varepsilon)\cdot\gamma_q\cdot\sqrt{2-(1/n-k)}$ with probability $1-p_e$, by using $O(1/\varepsilon \cdot \log(1/p_e))$ runs.

5 A Gap Instance for the Convex Relaxation

Here we describe an instance of $\mathsf{Subspace}(n-1,p)$ such that the value of any valid solution (which is of rank 1) is at least γ_p times the value of the convex relaxation. Note that approximation ratio of the algorithm for the case $n-k=1$ (and even p) is exactly γ_p and hence this shows that our analysis is optimal for this case.

This also gives a gap of factor γ_p for $\mathsf{Subspace}(k,p)$ for any super-constant $k = k(n)$, since an instance of $\mathsf{Subspace}(n-1,p)$ in $\mathbb{R}^n$ can be trivially converted (by adding extra zero coordinates) to an instance of $\mathsf{Subspace}(k,p)$ in $\mathbb{R}^{n'}$ with $k(n') = n-1$.

5.1 A continuous gap instance

Recall that an instance of $\mathsf{Subspace}(n-1,p)$ can be expressed as $\min_{\|z\|_2=1}\|Az\|_p$ for $A \in \mathbb{R}^{n\times m}$, where $a_1, a_2, \ldots, a_m$ form the rows of A. We consider a continuous generalization of this, where instead of points, we are given a probability distribution on $\mathbb{R}^n$ with density function $\mu(\cdot)$, and objective is:

$$\min_{\|z\|_2=1}\left(\int_{a\in\mathbb{R}^n}|\langle a,z\rangle|^p\,\mu(a)da\right)^{1/p}.$$

The corresponding convex relaxation is

$$\min_{\substack{I\succcurlyeq X\succcurlyeq 0\\ \mathsf{Tr}(X)=1}}\left(\int_{a\in\mathbb{R}^n}\left(a^T X a\right)^{p/2}\mu(a)da\right)^{1/p}.$$

We first show that Gaussian measure on $\mathbb{R}^n$, i.e., i.i.d. coordinates from $N(0,1)$, gives a gap instance for the above problem.

Theorem 5.1 *Given $\eta > 0$, there exists $n_0 \in \mathbb{Z}$ such that for all $n \geq n_0$ if μ is the Gaussian density function on $\mathbb{R}^n$ with each coordinate having mean 0 and variance 1, then*

$$\min_{\|z\|_2=1} \left(\int_{a \in \mathbb{R}^n} |\langle a, z \rangle|^p \, \mu(a) da \right)^{1/p} \;\geq\; \gamma_p (1-\eta) \cdot \min_{\substack{I \succcurlyeq X \succcurlyeq 0 \\ \mathsf{Tr}(X)=1}} \left(\int_{a \in \mathbb{R}^n} \left(a^T X a\right)^{p/2} \mu(a) da \right)^{1/p}.$$

Proof: We first consider the value of the LHS. By the rotational invariance of the Gaussian measure, the value is equal for all z and we can restrict ourselves to $z = e_1$.

$$\begin{aligned}
\min_{\|z\|_2=1} \left(\int_{a \in \mathbb{R}^n} |\langle a, z \rangle|^p \, \mu(a) da \right)^{1/p} &= \left(\int_{\mathbb{R}^n} |\langle a, e_1 \rangle|^p \, \mu(a) da \right)^{1/p} \\
&= \left(\int_{\mathbb{R}^n} |a_1|^p \frac{e^{-\|a\|^2/2}}{(2\pi)^{d/2}} da_1 \cdot da_2 \cdot \dots \cdot da_n \right)^{1/p} \\
&= \left(\int_{-\infty}^{\infty} |a_1|^p \frac{e^{-a_1^2/2}}{\sqrt{2\pi}} da_1 \cdot \prod_{j=2}^{n} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-a_j^2/2} da_j \right) \right)^{1/p} \\
&= \gamma_p.
\end{aligned}$$

In comparison, the optimum of the convex relaxation can be upper bounded by using the matrix $X = 1/n \cdot I$.

$$\begin{aligned}
\min_{\substack{I \succcurlyeq X \succcurlyeq 0 \\ \mathsf{Tr}(X)=1}} \left(\int_{a \in \mathbb{R}^n} \left(a^T X a\right)^{p/2} \mu(a) da \right)^{1/p} &\leq \frac{1}{\sqrt{n}} \left(\int_{\mathbb{R}^n} \|a\|^p \frac{e^{-\|a\|^2/2}}{(2\pi)^{n/2}} da_1 \cdot da_2 \cdot \dots \cdot da_n \right)^{1/p} \\
&= \frac{1}{\sqrt{n}} \left(\int_{\omega \in \mathbb{S}^{n-1}} \int_{r=0}^{\infty} r^p \frac{e^{-r^2/2}}{(2\pi)^{n/2}} r^{n-1} dr \cdot d\omega \right)^{1/p} \\
&= \frac{1}{\sqrt{n}} \left(\frac{1}{(2\pi)^{(n-1)/2}} \int_0^{\infty} r^{n+p-1} \frac{e^{-r^2/2}}{\sqrt{2\pi}} dr \cdot \int_{\omega \in \mathbb{S}^{n-1}} d\omega \right)^{1/p} \\
&= \frac{1}{\sqrt{n}} \left(\frac{1}{(2\pi)^{(n-1)/2}} \cdot \frac{2^{(n+p-1)/2} \Gamma(\frac{n+p}{2})}{2\sqrt{\pi}} \cdot \frac{2\pi^{n/2}}{\Gamma(n/2)} \right)^{1/p} \\
&= \left(\left(\frac{2}{n} \right)^{p/2} \cdot \frac{\Gamma(\frac{n+p}{2})}{\Gamma(n/2)} \right)^{1/p} \;\leq\; \left(1 + \frac{O(p)}{n} \right)^{1/2}
\end{aligned}$$

where the third equality used that $\int_{\omega \in \mathbb{S}^{n-1}} d\omega = \text{area}(\mathbb{S}^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}$, and $\int_0^{\infty} r^{n+p-1} \frac{e^{-r^2/2}}{\sqrt{2\pi}} dr = \gamma_{n+p-1}^{n+p-1}/2$. Choosing $n \gg p/\eta$ then proves the claim. ∎

5.2 Discretizing the gap example

A discrete analog of the above, i.e., picking sufficiently many samples from the same distribution, gives us our final integrality gap (or "rank gap") example.

Theorem 5.2 *Given any $\eta > 0$, there exist $m_0, n_0 \in \mathbb{Z}$ such that for all $m \geq m_0$ and $n \geq n_0$, if we pick i.i.d. random points $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$ with each point having i.i.d. $N(0,1)$ coordinates, then with some non-zero probability,*

$$\min_{\|z\|_2=1} \left(\frac{1}{m} \sum_{i=1}^{m} |\langle a_i, z \rangle|^p \right)^{1/p} \geq (1-\eta) \cdot \gamma_p \cdot \min_{\substack{I \succcurlyeq X \succcurlyeq 0 \\ \mathsf{Tr}(X)=1}} \left(\int_{a \in \mathbb{R}^n} \left| a_i^T X a_i \right|^{p/2} \right)^{1/p}.$$

In other words, there exist points $b_1, b_2, \ldots, b_m \in \mathbb{R}^n$, where $b_i \stackrel{\text{def}}{=} m^{-1/p} \cdot a_i$, giving the desired integrality gap example.

The theorem can be proved by using the continuous gap instance, and concentration bounds for the samples $a_1, \ldots, a_m$. We defer a full proof to the appendix.

6 Unique-Games Hardness

6.1 Khot's Unique Games Conjecture

We shall show a reduction to subspace approximation problem from the Unique Label Cover problem defined below.

Definition 6.1 *An instance of* Unique Label Cover *with alphabet size R is specified as a bipartite graph $\mathcal{U} = (V, W, E)$ with a set of permutations $\{\pi_{vw} : [R] \to [R]\}_{(v,w) \in E}$. A labeling $\mathcal{L} : V \cup W \to [R]$ is said to satisfy an edge (v, w) if $\mathcal{L}(w) = \pi_{vw}(\mathcal{L}(v))$. We denote by $\mathsf{val}(\mathcal{U})$ the maximum fraction of edges satisfied by any labeling $\mathcal{L}$.*

The Unique Games Conjecture proposed by Khot in [14] conjectures the hardness of distinguishing between the cases when the optimum to the above problem is very close to 1 and when it is very close to 0. This conjecture is an important complexity assumption as several approximation problems have been shown to be at least as hard as deciding if a given instance $\mathcal{U}$ of Unique Label Cover problem has $\mathsf{val}(\mathcal{U}) > 1 - \varepsilon$ or $\mathsf{val}(\mathcal{U}) < \delta$ for appropriate positive constants ε and δ.

Conjecture 6.2 (Khot [14]) *Given any constants $\varepsilon, \delta > 0$, there is an integer R such that it is NP-hard to decide if for given an instance $\mathcal{U} = (V, W, E)$ of Unique Label Cover with alphabet size R, $\mathsf{val}(\mathcal{U}) \geq 1 - \varepsilon$ or $\mathsf{val}(\mathcal{U}) \leq \delta$.*

6.2 Reduction from Unique Label Cover

We will now prove Unique-Games hardness of approximating Subspace($n-1$,p) within a factor better than γ_p. As in Section 5, this also gives a hardness approximating Subspace(k,p) for k which is a sufficiently large function of k, by a trivial embedding of the given instance $\mathbb{R}^n$ into $\mathbb{R}^{n'}$ such $k(n') = n - 1$. If we want n' to be a polynomial in n, this will give a hardness for all $k = n^{\Omega(1)}$.

We describe below the reduction from an instance $\mathcal{U} = (V, W, E)$ of Unique Label Cover with alphabet size R to Subspace($n-1$,p). The variables in our reduction will be of the form $b_{w,i}$ for each $w \in W$ and $i \in [R]$. We denote the vector $(b_{w,1}, \ldots, b_{w,R})$ by $\mathbf{b}_w$ and for each $v \in V$, define $\mathbf{b}_v \stackrel{\text{def}}{=} \mathbb{E}_{w \in N(v)}[\pi_{wv}(\mathbf{b}_w)]$. For any $\mathbf{b} \in \mathbb{R}^R$, we define the function $f_b : \{-1, 1\}^R \to \mathbb{R}$ as

$$f_b(x_1, \ldots, x_R) \stackrel{\text{def}}{=} \sum_{i=1}^{R} x_i \cdot b_i$$

Norms for functions are defined as usual (over the uniform probability measure). Note that $\|f_b\|_2^2 = \|\mathbf{b}\|_2^2$. When the exponent in the norm is unspecified, $\|\cdot\|$ denotes $\|\cdot\|_2$.

Given an instance $\mathcal{U} = (V, W, E)$ of Unique Label Cover we output the following instance of subspace approximation, for a suitable constant B to be determined later:

$$\begin{array}{ll} \text{minimize} & \mathop{\mathbb{E}}_{(v,w)\in E}\left[\|f_{b_v}\|_p^p\right] + B\cdot \mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|f_{b_v} - f_{\pi_{wv}(b_w)}\right\|_p^p\right] \\ \text{subject to:} & \mathop{\mathbb{E}}_{(v,w)\in E}\left[\|f_{b_w}\|_2^2\right] \;=\; \mathop{\mathbb{E}}_{(v,w)\in E}\left[\|\mathbf{b}_w\|_2^2\right] \;\geq\; 1 \end{array}$$

Note that the variables in the problem are only the vectors $\mathbf{b}_w$ for all $w \in W$. It is easy to verify the functions f_{b_v} and f_{b_w} can be generated by application of an appropriate operator A. In the proof below we shall often drop the subscript on the permutations π_{wv} when it is clear from the context. Note that value of instance of Subspace($n-1$,p) is actually the p^{th} root of the above objective. Let $(\mathsf{opt})^p$ denote the optimal value for the above objective (so that opt is the optimal value for Subspace($n-1$,p)).

Completeness

The following claim shows that the optimum of the subspace approximation problem is low when the Unique Label Cover is instance is highly satisfiable.

Claim 6.3 *If* $\mathsf{val}(\mathcal{U}) \geq 1-\varepsilon$, *then* $(\mathsf{opt})^p \leq 1 + \varepsilon \cdot B \cdot 2^p$.

Proof: By assumption, there exists a labeling $\mathcal{L} : V \cup W \to [R]$ such that $\mathbb{P}_{(v,w)\in E}[\mathcal{L}(v) \neq \pi_{wv}(\mathcal{L}(w))] \leq \varepsilon$. We construct a solution the above instance of the subspace approximation problem, taking $b_{w,i} = 1$ if $\mathcal{L}(w) = i$ and 0 otherwise. It is easy to check that $\mathbb{E}_{(v,w)\in E}\left[\|f_{b_w}\|_2^2\right] = 1$.

We now bound the value of the objective function. First note that $f_{b_v} = \mathbb{E}_{w\in N(v)}[f_{\pi(b_w)}]$ is bounded between -1 and 1, which implies $\mathbb{E}_{(v,w)\in E}\left[\|f_{b_v}\|_p^p\right] \leq 1$. To bound the second term, we can use Jensen's Inequality to get

$$\begin{aligned} \mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|f_{b_v} - f_{\pi_{wv}(b_w)}\right\|_p^p\right] &= \mathop{\mathbb{E}}_{(v,w_1)\in E}\left[\left\|\mathop{\mathbb{E}}_{w_2\in N(v)}\left[f_{\pi_{w_2v}(b_{w_2})}\right] - f_{\pi_{w_1v}(b_{w_1})}\right\|_p^p\right] \\ &\leq \mathop{\mathbb{E}}_{v,w_1,w_2}\left[\left\|f_{\pi_{w_2v}(b_{w_2})} - f_{\pi_{w_1v}(b_{w_1})}\right\|_p^p\right]. \end{aligned}$$

Note that $\left\|f_{\pi_{w_2v}(b_{w_2})} - f_{\pi_{w_1v}(b_{w_1})}\right\|_p^p$ equals 2^{p-1} if $\pi_{w_1v}(\mathcal{L}(w_1)) \neq \pi_{w_2v}(\mathcal{L}(w_2))$ and 0 otherwise. Hence,

$$\begin{aligned} \mathop{\mathbb{E}}_{v,w_1,w_2}\left[\left\|f_{\pi_{w_2v}(b_{w_2})} - f_{\pi_{w_1v}(b_{w_1})}\right\|_p^p\right] &= 2^{p-1}\cdot \mathop{\mathbb{P}}_{v,w_1,w_2}\left[\pi_{w_1v}(\mathcal{L}(w_1)) \neq \pi_{w_2v}(\mathcal{L}(w_2))\right] \\ &\leq 2^{p-1}\left(\mathop{\mathbb{P}}_{v,w_1}\left[\pi_{w_1v}(\mathcal{L}(w_1)) \neq \mathcal{L}(v)\right] + \mathop{\mathbb{P}}_{v,w_2}\left[\pi_{w_2v}(\mathcal{L}(w_2)) \neq \mathcal{L}(v)\right]\right) \\ &\leq 2^p\cdot\varepsilon. \end{aligned}$$

Combining the two bounds above gives $(\mathsf{opt})^p \leq 1 + \varepsilon \cdot B \cdot 2^p$. ■

Soundness

For the soundness, we need to prove that if $\mathsf{val}(\mathcal{U}) \le \delta$, then $\mathsf{opt} \ge \gamma_p^p \cdot (1-\nu)$ where ν is a small constant depending on ε and δ. We first make some simple observations about the optimal solution.

Claim 6.4 *For any optimal solution $\{\mathbf{b}_w\}_{w\in W}$ to the above instance of Subspace$(n-1,p)$, it must be true that*

1. $\mathbb{E}_{(v,w)\in E}\left[\|\mathbf{b}_v\|^2\right] \;\le\; \mathbb{E}_{(v,w)\in E}\left[\|\mathbf{b}_w\|^2\right] \;=\; 1$
2. $\mathbb{E}_{(v,w)\in E}\left[\left\|f_{b_v} - f_{\pi(b_w)}\right\|_p^p\right] \le \gamma_p^p/B$.

Proof: Since scaling all vectors by a constant less than 1 can only improve the value of the objective, we can assume that for the vectors $\{\mathbf{b}_w\}_{w\in W}$ in the solution $\mathbb{E}_{(v,w)\in E}\left[\|\mathbf{b}_w\|_2^2\right] = 1$. Then Jensen's inequality gives

$$\mathop{\mathbb{E}}_{(v,w)\in E}\left[\|\mathbf{b}_v\|^2\right] \;=\; \mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|\mathop{\mathbb{E}}_{w'\in N(v)}\left[\pi_{w'v}(\mathbf{b}_{w'})\right]\right\|^2\right] \;\le\; \mathop{\mathbb{E}}_{(v,w)\in E}\left[\|\mathbf{b}_w\|^2\right] \;=\; 1.$$

To deduce the second fact, we show that there exists a feasible solution $\{\mathbf{b}_w\}_{w\in W}$ such that $\mathsf{opt} \le \gamma_p^p$. For all $w \in W$, we take $\mathbf{b}_w = (1/\sqrt{R}, \ldots, 1/\sqrt{R})$. The solution is feasible since $\|\mathbf{b}_w\| = 1$ for each $w \in W$ and also $\mathbb{E}_{(v,w)\in E}\left[\left\|f_{b_v} - f_{\pi(b_w)}\right\|_p^p\right] = 0$. Also, since f_{b_v} is a linear function of Bernoulli variables and $\|\mathbf{b}_v\| = 1$, Claim 2.2 gives that for each $v \in V$, $\|f_{b_v}\|_p \le \gamma_p$. ∎

We show that if $\mathsf{val}(\mathcal{U}) \le \delta$, then in fact the first term itself is approximately γ_p^p. As is standard in Unique Games based reductions, the proof proceeds by arguing separately about the "high-influence" and "low-influence" cases. However, since the inputs for our problem are not in the form of a long-code but the vectors $\mathbf{b}$, we will use $\max_{i\in R}\{|b_i| \,/\, \|\mathbf{b}\|\}$ as a substitute for influence of the i^{th} variable on the function f_b.

For the vertices $v \in V$ where the functions f_{b_v} have no influential coordinates, the Central Limit Theorem shows that $\|f_{b_v}\|_p$ is very close to γ_p. We then show that the contribution of the remaining vertices to the objective function is small.

Below, we define S_1 to be the set of vertices corresponding to low influence functions and divide the remaining vertices into three cases which we shall analyze separately. The parameters $\tau, \beta \in (0, 1/2)$ will be chosen later.

$$S_1 \;\stackrel{\text{def}}{=}\; \left\{ v \in V \;\middle|\; \max_{i\in[R]}\{|b_{v,i}|\} < \tau \cdot \|\mathbf{b}_v\| \right\}$$

$$S_2 \;\stackrel{\text{def}}{=}\; \left\{ v \in V \;\middle|\; \|\mathbf{b}_v\|^2 \le (1-\beta)\cdot \mathop{\mathbb{E}}_{w\in N(v)}\left[\|\mathbf{b}_w\|^2\right] \right\}$$

$$S_3 \;\stackrel{\text{def}}{=}\; \left\{ v \in V \setminus S_2 \;\middle|\; \exists i \text{ s.t. } \; |b_{v,i}| \ge \tau\cdot\|\mathbf{b}_v\| \text{ and } \mathop{\mathbb{P}}_{w\in N(v)}\left[\left|b_{w,\pi_{vw}(i)}\right| \ge \tau/4 \cdot \|\mathbf{b}_w\|\right] \le 1/4 \right\}$$

$$S_4 \;\stackrel{\text{def}}{=}\; V \setminus (S_1 \cup S_2 \cup S_3).$$

Since $f_{b_v}(x_1, \ldots, x_R) = b_{v,1}\cdot x_1 + \cdots + b_{v,R}\cdot x_R$ is a linear function of Bernoulli variables, Claim 2.3 gives that

$$\forall v \in S_1 \qquad \|f_{b_v}\|_p^p \;\ge\; \gamma_p^p \cdot \|\mathbf{b}_v\|_2^p \cdot \left(1 - 10\tau\cdot(\log(1/\tau))^{p/2}\right) \qquad (1)$$

Note that the norm $\|f_{b_v}\|_p$ may be unbounded for individual vertices. Hence we will use the quantity $\mathbb{E}_{(v,w)\in E}\left[\mathbb{1}_{\{S_i\}}(v)\cdot\|\mathbf{b}_v\|^2\right]$ as a measure of the contribution of the set S_i to the objective, where $\mathbb{1}_{\{S_i\}}(\cdot)$ is the indicator function of the set S_i. Claims 6.5, 6.6 and 6.7 help bound the contribution of the sets S_2, S_3 and S_4.

Claim 6.5

$$\mathop{\mathbb{E}}_{(v,w)\in E}\left[(1-\mathbb{1}_{\{S_2\}}(v))\cdot\|\mathbf{b}_v\|^2\right]\geq 1-\beta-\frac{4\gamma_p^2}{\beta B^{2/p}}.$$

Proof: Since $\mathbf{b}_w=\mathbb{E}_{w\in N(v)}[\pi_{wv}(\mathbf{b}_{b_w})]$, being in S_2 means that on average, many vectors $\mathbf{b}_w$ differ from $\mathbf{b}_v$. We use this to get a bound on the measure of S_2. We have

$$\begin{aligned}\|\mathbf{b}_v\|^2\leq(1-\beta)\cdot\mathop{\mathbb{E}}_{w\in N(v)}\left[\|\mathbf{b}_w\|^2\right] &\implies \beta\cdot\mathop{\mathbb{E}}_{w\in N(v)}\left[\|\mathbf{b}_w\|^2\right]\leq\mathop{\mathbb{E}}_{w\in N(v)}\left[\|\mathbf{b}_w\|^2\right]-\|\mathbf{b}_v\|^2\\ &\implies \beta\cdot\mathop{\mathbb{E}}_{w\in N(v)}\left[\|\mathbf{b}_w\|^2\right]\leq\mathop{\mathbb{E}}_{w\in N(v)}\left[\|\pi_{wv}(\mathbf{b}_w)-\mathbf{b}_v\|^2\right],\end{aligned}$$

as $\|\mathbf{b}_w\|=\|\pi_{wv}(\mathbf{b}_w)\|$ and that $\mathbf{b}_v$ is the mean of $\pi_{wv}(\mathbf{b}_w)$. Now, since $\|\mathbf{b}\|=\|f_b\|$, we get that

$$\begin{aligned}\beta\cdot\mathop{\mathbb{E}}_{(v,w)\in E}\left[\mathbb{1}_{\{S_2\}}(v)\cdot\|\mathbf{b}_w\|^2\right] &\leq \mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|f_{b_v}-f_{\pi(b_w)}\right\|_2^2\right]\\ &\leq \mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|f_{b_v}-f_{\pi(b_w)}\right\|_p^2\right] && \text{(since } \|f\|_2\leq\|f\|_p\text{)}\\ &\leq \left(\mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|f_{b_v}-f_{\pi(b_w)}\right\|_p^p\right]\right)^{2/p} && \text{(using Jensen's Inequality)}\\ &\leq \gamma_p^2/B^{2/p}\end{aligned}$$

where we used the assumption that $\mathbb{E}_{(v,w)\in E}\left[\left\|f_{b_v}-f_{\pi(b_w)}\right\|_p^p\right]\leq\gamma_p^p/B$. This gives that

$$\begin{aligned}\mathop{\mathbb{E}}_{(v,w)\in E}\left[(1-\mathbb{1}_{\{S_2\}}(v))\cdot\|\mathbf{b}_v\|^2\right] &\geq(1-\beta)\cdot\mathop{\mathbb{E}}_{(v,w)\in E}\left[(1-\mathbb{1}_{\{S_2\}}(v))\cdot\|\mathbf{b}_w\|^2\right]\\ &\geq(1-\beta)\cdot\left(1-\frac{\gamma_p^2}{\beta B^{2/p}}\right)\\ &\geq 1-\beta-\frac{\gamma_p^2}{\beta B^{2/p}}.\end{aligned}$$

The second inequality above used that $\mathbb{E}_{(v,w)\in E}\left[\|\mathbf{b}_w\|^2\right]=1$ from claim 6.4. ■

Claim 6.6 $\mathbb{E}_{(v,w)\in E}\left[\mathbb{1}_{\{S_3\}}(v)\cdot\|\mathbf{b}_v\|^2\right] \;\leq\; 16/\tau^2\cdot\gamma_p^2/B^{2/p}$.

Proof: Consider a vertex $v\in S_3$. Since we know that $v\notin S_2$, we get that

$$\mathop{\mathbb{P}}_{w\in N(v)}\left[\|\mathbf{b}_w\|\geq 2\,\|\mathbf{b}_v\|\right]\;\leq\;\frac{\mathbb{E}_{w\in N(v)}\left[\|\mathbf{b}_w\|^2\right]}{4\,\|\mathbf{b}_v\|^2}\;\leq\;\frac{1}{4-4\beta}.$$

Fix and $i \in [R]$ such that $|b_{v,i}| \geq \tau \cdot \|\mathbf{b}_v\|$ and $\mathbb{P}_{w \in N(v)}\left[\left|b_{w,\pi_{vw}(i)}\right| \geq \tau/4 \cdot \|\mathbf{b}_w\|\right] \leq 1/4$. By a union bound,

$$\mathop{\mathbb{P}}_{w \in N(v)}\left[\|\mathbf{b}_w\| \leq 2\|\mathbf{b}_v\| \quad \text{and} \quad \left|b_{w,\pi_{vw}(i)}\right| \leq \tau/4 \cdot \|\mathbf{b}_w\|\right] \quad \geq \quad 1 - 1/4 - 1/(4-4\beta) \quad > \quad 1/4.$$

Using this we can again say that $\|\mathbf{b}_v - \pi_{wv}(\mathbf{b}_w)\|$ must be large on average and, hence, derive a bound on the measure of S_3.

$$\begin{aligned}
\mathop{\mathbb{E}}_{w \in N(v)}\left[\|\mathbf{b}_v - \pi_{wv}(\mathbf{b}_w)\|^2\right] \quad &\geq \quad \mathop{\mathbb{E}}_{w \in N(v)}\left[\left|b_{v,i} - b_{w,\pi(i)}\right|^2\right] \\
&\geq \quad 1/4 \cdot |\tau \cdot \|\mathbf{b}_v\| - \tau/4 \cdot 2\|\mathbf{b}_v\||^2 \\
&\geq \quad \tau^2/16 \cdot \|\mathbf{b}_v\|^2 .
\end{aligned}$$

As in the previous claim, we use this to conclude that

$$\begin{aligned}
\mathop{\mathbb{E}}_{(v,w) \in E}\left[\mathbb{1}_{\{S_3\}}(v) \cdot \|\mathbf{b}_v\|^2\right] \quad &\leq \quad 16/\tau^2 \cdot \mathop{\mathbb{E}}_{(v,w) \in E}\left[\left\|f_{b_v} - f_{\pi(b_w)}\right\|_2^2\right] \\
&\leq \quad 16/\tau^2 \left(\mathop{\mathbb{E}}_{(v,w) \in E}\left[\left\|f_{b_v} - f_{\pi(b_w)}\right\|_p^p\right]\right)^{2/p} \quad \leq \quad 16/\tau^2 \cdot \gamma_p^2/B^{2/p}.
\end{aligned}$$

■

Claim 6.7 $\mathbb{E}_{(v,w) \in E}\left[\mathbb{1}_{\{S_4\}}(v)\right] \leq 64\delta/\tau^2$.

Proof: Since $v \notin S_1 \cup S_2 \cup S_3$, we know that

$$\exists i \in [R] \quad \text{such that} \quad \mathop{\mathbb{P}}_{w \in N(v)}\left[\left|b_{w,\pi_{vw}(i)}\right| \geq \tau/4 \cdot \|\mathbf{b}_w\|\right] \geq 1/4.$$

Construct a labeling for $\mathcal{U}$ by assigning to each $v \in V$, the special label i as above, and to each $w \in W$, a random label j satisfying $|b_{w,j}| \geq \tau/4 \cdot \|\mathbf{b}_w\|$. For, $w \in W$ when no such j exists or for $v \notin S_4$, we fix a label arbitrarily.

Note that there can be at most $16/\tau^2$ choices of j satisfying $|b_{w,j}| \geq \tau/4 \cdot \|\mathbf{b}_w\|$. By the condition on i, we know that, in expectation, the labeling satisfies $1/4 \cdot \tau^2/16$ fraction of the edges incident on a $v \in S_4$. Since the fraction of edges satisfied overall is at most δ, we get that

$$\mathop{\mathbb{E}}_{(v,w) \in E}\left[\mathbb{1}_{\{S_4\}}(v) \cdot \tau^2/64\right] \leq \delta \quad \Longrightarrow \quad \mathop{\mathbb{E}}_{(v,w) \in E}\left[\mathbb{1}_{\{S_4\}}(v)\right] \leq 64\delta/\tau^2.$$

■

Let ν denote $10\tau \cdot (\log(1/\tau))^{p/2}$. Using these estimates, we can now prove the soundness of the reduction.

Lemma 6.8 *If* $\mathsf{val}(\mathcal{U}) < \delta$*, then for the reduction with parameters* B, τ *and* $\beta = \tau^2$

$$(\mathsf{opt})^p \;\geq\; \gamma_p^p \cdot \left(1 \;-\; \nu - \frac{p\tau^2}{2} \;-\; \frac{10p \cdot \gamma_p^2}{\tau^2 B^{2/p}} \;-\; \frac{p\gamma_p^2}{2}\left(\frac{64\delta}{\tau^2}\right)^{(p-2)/p}\right)$$

Proof: Using (1) we have that

$$(\mathsf{opt})^p \;\geq\; \mathop{\mathbb{E}}_{(v,w)\in E}\left[\mathbb{1}_{\{S_1\}}(v)\cdot\gamma_p^p\cdot(1-\nu)\left\|\mathbf{b}_v\right\|_2^p\right] \;\geq\; \gamma_p^p\cdot(1-\nu)\cdot\left(\mathop{\mathbb{E}}_{(v,w)\in E}\left[\mathbb{1}_{\{S_1\}}(v)\cdot\left\|\mathbf{b}_v\right\|_2^2\right]\right)^{p/2}.$$

We lower bound $\mathbb{1}_{\{S_1\}}(v)$ by $1-\mathbb{1}_{\{S_2\}}(v)-\mathbb{1}_{\{S_3\}}(v)-\mathbb{1}_{\{S_4\}}(v)$. Claims 6.5 and 6.6 and give bounds on the first two terms (with $\beta=\tau^2$).

$$\begin{aligned}\mathop{\mathbb{E}}_{(v,w)\in E}\left[(1-\mathbb{1}_{\{S_2\}}(v))\cdot\left\|\mathbf{b}_v\right\|^2\right] &\;\geq\; 1-\tau^2-\frac{4\gamma_p^2}{\tau^2B^{2/p}},\\ \mathop{\mathbb{E}}_{(v,w)\in E}\left[\mathbb{1}_{\{S_3\}}\left\|\mathbf{b}_v\right\|^2\right] &\;\leq\; \frac{16\gamma_p^2}{\tau^2B^{2/p}}\end{aligned}$$

We bound the third term using Claim 6.7 and Hölder's inequality

$$\begin{aligned}\mathop{\mathbb{E}}_{(v,w)\in E}\left[\mathbb{1}_{\{S_4\}}\left\|\mathbf{b}_v\right\|^2\right] &\;\leq\; \left(\mathop{\mathbb{E}}_{(v,w)\in E}\left[\mathbb{1}_{\{S_4\}}(v)\right]\right)^{(p-2)/p}\left(\mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|\mathbf{b}_v\right\|^p\right]\right)^{2/p}\\ &\;\leq\; \left(\frac{64\delta}{\tau^2}\right)^{(p-2)/p}\cdot\gamma_p^2,\end{aligned}$$

where the last bound used that since $\mathsf{opt}\leq\gamma_p$ (see Claim 6.4), we must have

$$\mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|\mathbf{b}_v\right\|_2^p\right] \;=\; \mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|f_{b_v}\right\|_2^p\right] \;\leq\; \mathop{\mathbb{E}}_{(v,w)\in E}\left[\left\|f_{b_v}\right\|_p^p\right] \;\leq\; \gamma_p^p.$$

Combining the bounds for the above three terms proves the lemma. ■

For a small constant η such that $\eta(\log(1/\eta))^{p/2} < 2^{-p/2}/50$, choosing parameters as

$$\begin{aligned}\tau &\stackrel{\text{def}}{=} \eta^2/p, & \delta &\stackrel{\text{def}}{=} \left(\frac{\eta}{p\gamma_p^2}\right)^{p/(p-2)}\cdot\frac{\tau^2}{64},\\ B &\stackrel{\text{def}}{=} \left(\frac{40p\gamma_p^2}{\eta\cdot\tau^2}\right)^p, &\text{and}\quad \varepsilon &\stackrel{\text{def}}{=} \frac{\eta}{2^p\cdot B}\end{aligned}$$

in Lemma 6.8 would imply that $\mathsf{opt}\leq 1+\eta$ in the completeness case and $\mathsf{opt}\geq\gamma_p\cdot(1-\eta)$ in the soundness case. This gives the following theorem.

Theorem 6.9 *For any $p\geq 2$ and sufficiently small constant η, there exist constants $\varepsilon,\delta>0$ and a reduction from* Unique Label Cover *to* Subspace$(n-1,p)$ *such that if* $\mathsf{val}(\mathcal{U})$ *is the fraction of edges satisfiable in the given instance of* Unique Label Cover *and* opt *is the optimum of the instance of* Subspace$(n-1,p)$, *then*

$$\begin{aligned}\mathsf{val}(\mathcal{U})\geq 1-\varepsilon &\implies \mathsf{opt}\leq 1+\eta \quad \textit{and}\\ \mathsf{val}(\mathcal{U})\leq\delta &\implies \mathsf{opt}\geq\gamma_p\cdot(1-\eta).\end{aligned}$$

Acknowledgments

We thank Kasturi Varadarajan for initiating the work on this problem by suggesting that we generalize the algorithm of [23] and for generous help with an early draft of this paper on which he was offered a co-authorship (but later opted out). MT would also like to thank David Steurer for helpful discussions. We thank the anonymous reviewers of this manuscript for their suggestions and references, and for pointing out an error in the previous proof of Lemma 6.8.

References

[1] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. *SIAM Journal on Computing (SICOMP)*, 35(4):787–803, 2006. 2

[2] S. Arora, E. Berger, E. Hazan, G. Kindler, and M. Safra. On non-approximability for quadratic programs. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005. 2

[3] A. Brieden, P. Gritzmann, and V. Klee. Inapproximability of some geometric and quadratic optimization problems. In *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*, 2000. 1

[4] M. Bădoiu, S. Har-peled, and P. Indyk. Approximate clustering via coresets. In *ACM Symposium on Theory of Computing (STOC)*, 2002. 1, 2

[5] K. Clarkson. Subgradient and sampling algorithms for l1 regression. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 257–266, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics. 2, 3

[6] A. Dasgupta, P. Drineas, B. Harb, R. Kumar, and M. Mahoney. Sampling algorithms and coresets for ℓp regression. In *SODA '08: Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 932–941, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics. 2, 3

[7] A. Deshpande and K. Varadarajan. Sampling-based dimension reduction for subspace approximation. In *ACM Symposium on Theory of Computing (STOC)*, 2007. 2

[8] P. Drineas, M. Mahoney, and S. Muthukrishnan. Sampling algorithms for l2 regression and applications. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1127–1136, New York, NY, USA, 2006. ACM. 3

[9] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Struct. Algorithms*, 27(2):251–275, 2005. 19

[10] D. Feldman, M. Monemizadeh, C. Sohler, and D. Woodruff. Coresets and sketches for high dimensional subspace approximation problems. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010. 2

[11] G. Golub and C. van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996. 1

[12] P. Gritzmann and V. Klee. Computational complexity of inner and outer j-radii of polytopes in finite-dimensional normed spaces. *Mathematical Programming*, 59:163–213, 1993. 1

[13] S. Har-Peled and K. Varadarajan. Projective clustering in high dimensions using coresets. In *ACM Symposium on Computational Geometry (SOCG)*, 2002. 1, 2

[14] S. Khot. On the power of unique 2-prover 1-round games. In *IEEE Conference on Computational Complexity (CCC)*, 2002. 2, 11

[15] S. Khot and A Naor. Approximate kerkel clustering. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2008. 2

[16] G. Kindler, A. Naor, and G. Schechtman. The UGC hardness threshold of the L_p Grothendieck problem. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008. 2, 4, 5

[17] A. Nemirovski, C. Roos, and T. Terlaky. On maximization of quadratic forms over intersection of ellipsoids with common center. *Mathematical Programming, Ser. A*, 86:463–473, 1999. 1, 2, 4, 6

[18] Yu. Nesterov. Global quadratic optimization via conic relaxation. working paper, CORE, Catholic University of Louvain, Belgium, 1998. 1

[19] G. Peskir and A. N. Shiryaev. The inequalities of khintchine and expanding sphere of their action. *Russian Math. Surveys*, 50(5):849–904, 1995. 4

[20] N. Shyamalkumar and K. Varadarajan. Efficient subspace approximation algorithms. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007. 2

[21] A. M. So, Y. Ye, and J. Zhang. A unified theorem on SDP rank reduction. *Mathematics of Operations Research*, 33(4):910–920, 2008. 4

[22] P. van Beek. An application of Fourier methods to the problem of sharpening the Berry-Esseen inequality. *Probability Theory and Related Fields*, 23(3):187–196, 1972. 4

[23] K. Varadarajan, S. Venkatesh, Y. Ye, and J. Zhang. Approximating the radii of point sets. *SIAM Journal on Computing (SICOMP)*, 36(6):1764–1776, 2007. 1, 2, 4, 6, 17

A Proof of Theorem 5.2

We restate the theorem below.

Theorem A.1 *Given any $\eta > 0$, there exist $m_0, n_0 \in \mathbb{Z}$ such that for all $m \geq m_0$ and $n \geq n_0$, if we pick i.i.d. random points $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$ with each point having i.i.d. $N(0,1)$ coordinates, then with some non-zero probability,*

$$\min_{\|z\|_2=1} \left(\frac{1}{m} \sum_{i=1}^{m} |\langle a_i, z \rangle|^p \right)^{1/p} \geq (1-\eta) \cdot \gamma_p \cdot \min_{\substack{I \succcurlyeq X \succcurlyeq 0 \\ \mathsf{Tr}(X)=1}} \left(\int_{a \in \mathbb{R}^n} \left| a_i^T X a_i \right|^{p/2} \right)^{1/p}.$$

In other words, there exist points $b_1, b_2, \ldots, b_m \in \mathbb{R}^n$, where $b_i \stackrel{\text{def}}{=} m^{-1/p} \cdot a_i$, giving the desired integrality gap example.

Proof: Let $a_1, a_2, \ldots, a_m$ be i.i.d. random points in $\mathbb{R}^n$, where each point has i.i.d. $N(0,1)$ coordinates. Then, as we have seen above

$$\mathbb{E}\left[|\langle a_i, y\rangle|^p\right] = \int_{\mathbb{R}^n} |\langle a, y\rangle|^p \, \mu(a) da = \gamma_p^p, \qquad \text{for } y \in \mathbb{S}^{n-1},$$
$$\mathsf{Var}\left[|\langle a_i, y\rangle|^p\right] = \mathbb{E}\left[|\langle a_i, y\rangle|^{2p}\right] - \mathbb{E}\left[|\langle a_i, y\rangle|^p\right]^2 = \gamma_{2p}^{2p} - \gamma_p^{2p}, \qquad \text{for } y \in \mathbb{S}^{n-1}.$$

By Chebyshev's Inequality,

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^m |\langle a_i, y\rangle|^p \le (1-\varepsilon)\gamma_p^p\right] \le \frac{(\gamma_{2p}^{2p} - \gamma_p^{2p})}{m\varepsilon^2\gamma_p^{2p}}.$$

Let $\mathcal{N}$ be any δ-net of the unit sphere (i.e., $\mathcal{N} \subseteq \mathbb{S}^{n-1}$ such that for any $z \in \mathbb{S}^{n-1}$, there exists some $y \in \mathcal{N}$ such that $\|y - z\|_2 \le \delta$), where δ is a parameter that will be picked later. It is known (e.g. see Claim 2.9 in [9]) how to construct such δ-nets of $\mathbb{S}^{n-1}$ with size as small as $|\mathcal{N}| \le (\frac{9}{\delta})^n$. Now using union bound over $\mathcal{N}$

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^m |\langle a_i, y\rangle|^p \ge (1-\varepsilon)\gamma_p^p, \text{ for all } y \in \mathcal{N}\right] \ge 1 - \frac{(\frac{9}{\delta})^n \cdot (\gamma_{2p}^{2p} - \gamma_p^{2p})}{m\varepsilon^2\gamma_p^{2p}} > \frac{3}{4},$$

as long as we choose m large enough so that

$$m > \frac{4 \cdot (\frac{9}{\delta})^n \cdot (\gamma_{2p}^{2p} - \gamma_p^{2p})}{\varepsilon^2\gamma_p^{2p}}.$$

For any $z \in \mathbb{S}^{n-1}$, using $y \in \mathcal{N}$ closest to it

$$\begin{aligned}\sum_{i=1}^m |\langle a_i, z\rangle|^p &= \sum_{i=1}^m |\langle a_i, y\rangle + \langle a_i, z - y\rangle|^p \\ &\ge \sum_{i=1}^m \langle a_i, y\rangle^p - p\delta \sum_{i=1}^m \|a_i\|_2^{p-1}.\end{aligned}$$

Therefore,

$$\mathbb{P}\left[\min_{\|z\|_2=1} \frac{1}{m}\sum_{i=1}^m |\langle a_i, z\rangle|^p \ge (1-\varepsilon)\gamma_p^p - \frac{p\delta}{m}\sum_{i=1}^m \|a_i\|_2^{p-1}\right] > \frac{3}{4}.$$

But we also know that

$$\mathbb{E}\left[\|a_i\|_2^{p-1}\right] = \int_{a \in \mathbb{R}^n} \|a\|_2^{p-1} \mu(a) da = n^{(p-1)/2}(1 + o(1))$$
$$\mathsf{Var}\left[\|a_i\|_2^{p-1}\right] = \mathbb{E}\left[\|a_i\|_2^{2p-2}\right] - \mathbb{E}\left[\|a_i\|_2^{p-1}\right]^2 = n^{(p-1)}(1 + o(1))$$

By Chebyshev's Inequality,

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^m \|a_i\|_2^{p-1} \ge (1+\varepsilon)n^{(p-1)/2}(1 + o(1))\right] \le \frac{1 + o(1)}{m\varepsilon^2}.$$

Hence, choosing $m > {}^5\!/\varepsilon^2$, we have

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^{m}\|a_i\|_2^{p-1} \le (1+\varepsilon)n^{(p-1)/2}(1+o(1))\right] \ge 1 - \frac{1+o(1)}{m\varepsilon^2} > \frac{3}{4}.$$

Putting these together,

$$\mathbb{P}\left[\min_{\|z\|_2=1}\frac{1}{m}\sum_{i=1}^{m}|\langle a_i, z\rangle|^p \ge (1-\varepsilon)\gamma_p^p - p\delta(1+\varepsilon)n^{(p-1)/2}(1+o(1))\right] > 3/4.$$

Overall, choosing

$$\varepsilon \stackrel{\text{def}}{=} \frac{\eta^2}{8}, \quad \delta \stackrel{\text{def}}{=} \frac{\eta^2\gamma_p^p}{(8+\eta^2)pn^{(p-1)/2}}, \quad \text{and} \quad m > \max\left\{\frac{4\delta^{-n}(\gamma_{2p}^{2p}-\gamma_p^{2p})}{\varepsilon^2\gamma_p^{2p}}, \frac{5}{\varepsilon^2}\right\},$$

we get

$$\mathbb{P}\left[\min_{\|z\|_2=1}\frac{1}{m}\sum_{i=1}^{m}|\langle a_i, z\rangle|^p \ge \left(1-\frac{\eta^2}{4}\right)\gamma_p^p\right] > \frac{1}{2}.$$

On the other hand to analyze the value of the corresponding convex relaxation, we use

$$\mathbb{E}\left[\|a_i\|_2^p\right] = \int_{a\in\mathbb{R}^n}\|a\|_2^p\,\mu(a)da = n^{p/2}(1+o(1))$$
$$\mathsf{Var}\left[\|a_i\|_2^p\right] = \mathbb{E}\left[\|a_i\|_2^{2p}\right] - \mathbb{E}\left[\|a_i\|_2^p\right]^2 = n^p(1+o(1))$$

Again by Chebyshev's Inequality,

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^{m}\|a_i\|_2^p \ge (1+{}^\eta\!/_2)\,n^{p/2}(1+o(1))\right] \le \frac{4(1+o(1))}{m\eta^2}.$$

Choosing $m > {}^9\!/\eta^2$, we get

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^{m}\|a_i\|_2^p \le (1+{}^\eta\!/_2)\,n^{p/2}(1+o(1))\right] \ge 1 - \frac{4(1+o(1))}{m\eta^2} > \frac{1}{2}.$$

Therefore, the convex relaxation satisfies

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^{m}\left|{}^1\!/_n \cdot a_i^T I a_i\right|^{p/2} \le (1+{}^\eta\!/_2)\,(1+o(1))\right] > \frac{1}{2}.$$

Hence,

$$\mathbb{P}\left[\min_{\|z\|_2=1}\left(\frac{1}{m}\sum_{i=1}^{m}|\langle a_i, z\rangle|^p\right)^{1/p} \ge (1-\eta)\cdot\gamma_p\cdot\min_{\substack{I\succcurlyeq X\succcurlyeq 0\\ \mathsf{Tr}(X)=1}}\left(\frac{1}{m}\sum_{i=1}^{m}\left|a_i^T X a_i\right|^{p/2}\right)^{1/p}\right]$$
$$\ge \mathbb{P}\left[\min_{\|z\|_2=1}\frac{1}{m}\sum_{i=1}^{m}|\langle a_i, z\rangle|^p \ge (1-{}^{\eta^2}\!/_4)\cdot\gamma_p^p \quad \text{and} \quad \frac{1}{m}\sum_{i=1}^{m}\left|{}^1\!/_n\cdot a_i^T I a_i\right|^{p/2} \le (1+{}^\eta\!/_2)\,(1+o(1))\right]$$
$$> 0.$$

■

B NP-hardness of Subspace Approximation

In this section, we show unconditionally that the problem Subspace($n-1$,p) is NP-hard, for $p > 2$, using a reduction from the Min-Uncut problem on graphs. Such a result was also obtained independently by Gibson and Xiao (personal communication).

Min-Uncut problem: Given a graph $G = (V, E)$, find a bipartition of its vertices $V = S \cup T$ that minimizes the number of edges with both endpoints on the same side of the bipartition.

Let $|V| = n$ and $|E| = m$. Min-Uncut problem is known to be NP-hard, i.e., for some $1 \leq t \leq m$ it is NP-hard to find if the Min-Uncut has at most t edges. We give a polynomial time reduction from Min-Uncut to subspace approximation as follows: Given an instance of Min-Uncut, construct a matrix $A \in \mathbb{R}^{(m+n)\times n}$ such that

$$\min_{\|y\|_2=\sqrt{n}} \|Ay\|_p^p = \min_{\|y\|_2=\sqrt{n}} \sum_{ij\in E} (y_i + y_j)^p + N \sum_{i=1}^{n} y_i^p,$$

where N is an integer polynomially large in n and m which will be chosen later.

Yes case: The Min-Uncut has at most t edges. Define $x_i = 1$ if $i \in S$ and $x_i = 1$ if $i \in T$. Using this $x \in \{-1, 1\}^n$ we get $\text{OPT} \leq \|Ax\|_p^p = t2^p + Nn$.

No case: Otherwise, for any bipartition the Min-Uncut has at least $t + 1$ edges, i.e., for any $x \in \{-1, 1\}^n$ we have $\sum_{ij\in E}(x_i + x_j)^p \geq (t + 1)2^p$. Now divide the sphere of radius $\sqrt{n}$ into two parts as follows:

$$S = \{y \ : \ \|y\|_2 = \sqrt{n} \text{ and } |y_j| \in (1-\varepsilon, 1+\varepsilon) \text{ for all } j \in [n]\},$$
$$T = \{y \ : \ \|y\|_2 = \sqrt{n} \text{ and } y \notin S\},$$

where $\varepsilon < {}^1\!/_p \cdot (m + 1)$. For any $y \in T$,

- Case 1: $|y_i| = 1 + \varepsilon_i \geq 1 + \varepsilon$ for some i. Then,

$$\sum_{j=1}^{n} y_j^p \geq (1+\varepsilon_i)^p + (n-1)\left(\frac{n-(1+\varepsilon_i)^2}{n-1}\right)^{p/2}$$
$$\geq (1+\varepsilon_i)^p + (n-1)\left(1 - \frac{2\varepsilon_i + \varepsilon_i^2}{n-1}\right)^{p/2}$$
$$\geq 1 + p\varepsilon_i + \binom{p}{2}\varepsilon_i^2 + (n-1)\left(1 - {}^p\!/_2 \cdot \frac{2\varepsilon_i + \varepsilon_i^2}{n-1}\right)$$
$$\geq n + \frac{p^2\varepsilon^2}{4} \qquad \text{using } p > 2\left(1 + \frac{1}{n-1}\right) \text{ for large enough } n.$$

- Case 2: $|y_i| = 1 - \varepsilon_i \leq 1 - \varepsilon$ for some i. Then,

$$\sum_{j\neq i} y_j^2 = n - (1 - \varepsilon_i^2).$$

  Hence, there exists some k such that

$$y_k^2 \geq \frac{n-(1-\varepsilon_i)^2}{n-1} \geq 1 + \frac{2\varepsilon_i - \varepsilon_i^2}{n-1} \geq 1 + \frac{\varepsilon}{n} \Rightarrow |y_k| \geq 1 + \frac{\varepsilon}{2n}.$$

  Therefore, using the same analysis as in the previous case, we get

$$\sum_{j=1}^{n} y_j^p \geq n + \frac{p^2\varepsilon^2}{16n^2}.$$

Using the above property of $y \in T$, we get

$$\begin{aligned}
\sum_{ij\in E}(y_i+y_j)^p + N\sum_{j=1}^{n} y_j^p &\geq N\sum_{j=1}^{p} y_j^p \\
&\geq Nn + \frac{Np^2\varepsilon^2}{16n^2} \\
&> t2^p + Nn \qquad \text{using } N > 2^{p+4}n^2m(m+1)^2.
\end{aligned}$$

For any $y \in S$,

$$\begin{aligned}
\sum_{ij\in E}(y_i+y_j)^p + N\sum_{j=1}^{n} y_j^p &\geq (1-\varepsilon)^p(t+1)2^p + Nn \\
&\geq (1-p\varepsilon)(t+1)2^p + Nn \\
&> t2^p + Nn \qquad \text{using } \varepsilon < \frac{1}{p(t+1)}.
\end{aligned}$$