On Approximating Quadratic Programs

Sanjeev Arora, Eli Berger, Elad Hazan, Guy Kindler, Muli Safra

Abstract

This paper studies the computational complexity of the following type of quadratic programs: given an arbitrary matrix whose diagonal elements are zero, find $x \in \{-1, +1\}^n$ that maximizes $x^T Ax$. This problem recently attracted attention due to its application in various clustering settings (Charikar and Wirth, 2004) as well as an intriguing connection to the famous *Grothendieck inequality* (Alon and Naor, 2004). It is approximable to within a factor of $O(\log n)$ [Nes98, NRT99, Meg01, CW04], and known to be NP-hard to approximate within any factor better than $13/11 - \epsilon$ for all $\epsilon > 0$ [CW04]. We show that it is quasi-NP-hard to approximate to a factor better than $O(\log^\gamma n)$ for some $\gamma > 0$.

The integrality gap of the natural semidefinite relaxation for this problem is known as the *Grothendieck constant* of the complete graph, and known to be $\Theta(\log n)$ (Alon, K. Makarychev, Y. Makarychev and Naor, 2005 [AMMN]). The proof of this fact was *nonconstructive*, and did not yield an explicit problem instance where this integrality gap is achieved. Our techniques yield an explicit instance for which the integrality gap is $\Omega(\frac{\log n}{\log \log n})$, essentially answering one of the open problems of [AMMN].

1 Introduction

This paper deals with the following class of quadratic programs (henceforth denoted MaxQP):

$$\begin{aligned} \text{Maximize} \quad & x^T A x \\ \text{Subject to} \quad & x_i \in \{-1,1\} \qquad \forall i \in [n] \end{aligned}$$

Here the matrix A is arbitrary, except that the trace (sum of all diagonal entries) is zero. Without loss of generality it is also symmetric, as A it can be replaced by the average of A and A^T retaining the same value for every assignment. Replacing the constraint $x_i \in \{-1,1\}$ by $x_i^2 \leq 1$ does not change the optimum, as is easy to verify. Also without loss of generality we may assume all diagonal entries are zero.

This subcase of quadratic programming has attracted a lot of attention recently thanks to a surprising web of connections. First, it is an attractive subcase to begin with, being a generalization of problems such as MAX-CUT, in which the constraints involve *pairs* of vertices. Second, the obvious generalization of the seminal MAX-CUT algorithm of Goemans and Williamson fails already for this problem — the mixed *signs* of the entries of A cause problems for the GW rounding algorithm. One would hope that investigating this problem would lead to new techniques for analyzing SDP relaxations for other problems. Third, it seems to capture the essential difficulty of a natural optimization problem called *correlation clustering* introduced by Bansal, Blum, and Chawla [BBC], and this was the motivation for its study in Charikar and Wirth [CW04]. (It is also studied in physics in context of *spin glass models*, see [Tal03]) And finally, the integrality gap of the obvious SDP relaxation seems related to questions studied in analysis. In particular, the famous *Grothendieck's inequality* implies an $O(1)$-approximation to the *bipartite* case of this problem where the objective is $x^T Ay$ where x, y are vectors in $\{-1,1\}$. This was pointed out by Alon and Naor [AN04], who gave an algorithmic version of Grothendieck's inequality (in other words, a *rounding algorithm* for the obvious SDP relaxation). They used this algorithm to derive an $O(1)$-approximation to the *cut norm* of a matrix, which plays an important role in approximation algorithms for dense graph problems [FK99].

Motivated by the Goemans-Williamson work, Nesterov and Nemirovskii had independently [Nes98, NRT99] obtained $O(\log n)$-approximations to MaxQP. This algorithm was later rediscovered in the clustering context by Charikar and Wirth, who also pointed that the known hardness results for MAX-CUT implied that $13/11 - \epsilon$ approximation is NP-hard. They raised the obvious question, whether the approximation ratio can be improved from $\log n$ to $O(1)$.

In this paper we resolve this question on the negative side, and prove the following:

Theorem 1. *There exists a constant $\gamma > 0$ such that if $NP \not\subseteq DTIME(n^{\log^3 n})$, then MaxQP cannot be approximated in polynomial time to a factor smaller then $O(\log^\gamma n)$.*

Furthermore, assuming a sufficiently strong version of Khot's *unique games conjecture* [Kho02] we prove that computing a $O(\log n)$-approximation is also hard. Note that though the unique games conjecture has become very popular in PCP work lately, the version used here is stronger than any used before.

Independently, Khot and O'Donnell [KO] have proved that MaxQP cannot be approximated in polynomial time up to a factor smaller then $O(\log \log n)$. Their proof assumes the (standard version of) unique games conjecture.

The second aspect of our work is a better understanding of the standard SDP relaxation for the MaxQP problem, which is used both in the above-mentioned $O(\log n)$-approximation, as well as in a formal study by Alon et al [AMMN] of the *Grothendieck constant* of a graph. The Grothendieck constant of an n-node graph $G = (V, E)$ is the maximum integrality gap of the above SDP among all matrices A whose entries are non-zero precisely for $\{i, j\}$ that are edges in E. Alon et al. proved that this integrality gap, the Grothendieck constant, is $\Omega(\log n)$ for the complete graph. This improves upon Kashin and Szarek [KS03], who obtained a bound of $\Omega(\sqrt{\log n})$. However, both proofs are non-constructive, in the sense that they do not generate an explicit matrix for which the integrality gap is achieved. We essentially answer this question and provide an explicit matrix for which the integrality gap is $\Omega(\frac{\log n}{\log\log n})$.

The rest of the paper is organized as follows. First we present a few definitions and previous results & conjectures in Section 2. Then we prove Theorem 1 and the stronger hardness result assuming the strong version of the unique games conjecture in section 3. Section 4 contains the explicit construction of a matrix that achieves integrality gap of $\Omega(\frac{\log n}{\log\log n})$.

2 Preliminaries

The MaxQP problem we consider is defined as follows

Definition 1 **(MaxQP).** *An instance of the* MaxQP *problem is a matrix $M \in \mathbb{R}^{n\times n}$ with non-negative trace and a set of variables $\{x_1, ..., x_n\}$. The objective is to find an assignment $A : \{x_i\} \mapsto \{-1, 1\}$ that maximizes the quadratic form $x^T Mx$. The objective value of an instance I under assignment A is denoted by $I(A)$.*

The natural semi-definite relaxation for MaxQP is defined as

Definition 2 **(MaxQP relaxed version).** *Given a matrix $M \in \mathbb{R}^{n\times n}$ with non-negative trace, assign unit vectors (i.e. vectors of l_2 norm 1) $v_i \in R^n$ such to maximize the expression $\sum_{ij} M_{ij} \cdot \langle v_i, v_j\rangle$.*

A common starting point for our hardness results is the label cover problem defined below.

Definition 3. *The Label Cover problem $\mathcal{L}(V, W, E, [R], \{\sigma_{v,w}\}_{(v,w)\in E})$ is defined as follows. We are given a regular bipartite graph with left side vertices V, right side vertices W, and a set of edges E. In addition, for every edge $(v, w) \in E$ we are given a map $\sigma_{v,w} : [R] \to [R]$. A* labelling *of the instance is a function ℓ assigning one label to each vertex of the graph, namely $\ell\colon V \cup W \to [R]$. A labelling ℓ satisfies an edge (v, w) if*

$$\sigma_{v,w}(\ell(w)) = \ell(u) \ .$$

The value *of the label cover problem is defined to be the maximum, over all labellings, of the fraction of edges satisfied.*

The PCP Theorem [AS98, ALM+98] combined with Raz's parallel repetition theorem [Raz98] yields the following theorem, which will be used in the proof of Theorem 1

Theorem 2 **(Quasi-NP-hardness).** *There exists a constant $\gamma > 0$ so that for any language L in NP, any input w and any $R > 0$, one can construct a labeling instance $\mathcal{L}$, with $|w|^{O(\log R)}$ vertices, and label set of size R, so that: If $w \in L$, $\ell(\mathcal{L}) = 1$ and otherwise $\ell(\mathcal{L}) < R^{-\gamma}$. Furthermore, $\mathcal{L}$ can be constructed in time polynomial in its size.*

A stronger better hardness result can be achived if we assume a stronger version of a conjecture by Khot, which is described next.

Definition 4. *The Unique Label Cover problem is Label Cover restricted to instances in which the labelling constraints on the edges are given by* **bijective** *maps* $\sigma_{u,w} : [R] \to [R]$.

Khot [Kho02] has conjectured that the unique label cover problem is hard to approximate in the following sense. (Note that for convenience this conjecture is often stated using the terminology of NP-hardness, but technically one usually only needs to assume that the problem does not have an efficient algorithm.)

Conjecture 1 (Unique Game Conjecture). *For every* $\varepsilon, \delta > 0$ *there exists a number* $R(\varepsilon, \delta)$*, such that any SAT instance* Φ *of size* n *can be reduced in time* $O(n^{\log R})$ *to a unique label cover instance* $\mathcal{L}$ *with* n *vertices and label set size* R*, and the following properties: If* Φ *is satisfiable, then* $\mathcal{L}$ *has objective value at least* $1 - \varepsilon$*. Otherwise,* $\mathcal{L}$ *has objective value of at most* δ*.*

The minimum possible value for $R(\varepsilon, \delta)$ in the above conjecture is $\frac{1}{\delta}$. This follows from the fact that a random label assignment satisfies an expected $\frac{1}{R}$ fraction of the constraints, and thus $\delta > \frac{1}{R}$. Recently Hastad [Has05] showed a polynomial time approximation algorithm for unique label cover that implies $R \geq \frac{1}{\delta} + o(\frac{1}{\delta})$. The stronger conjecture we shall use is obtained by taking $R(O(1), \delta) = O(\frac{1}{\delta})$.

Conjecture 2 (Unique Game Conjecture - Strong Version). *For every* $\delta > 0$ *there exists a number* $R(\delta) = O(\frac{1}{\delta})$*, such that any SAT instance* Φ *of size* n *can be reduced in time* $O(n^{\log R})$ *to a unique label cover instance* $\mathcal{L}$ *with* n *vertices and label set size* R*, and the following properties: If* Φ *is satisfiable, then* $\mathcal{L}$ *has objective value at least* $\Omega(1)$*. Otherwise,* $\mathcal{L}$ *has objective value of at most* δ*.*

Assuming this strong version we shall prove:

Theorem 3. *Assuming the strong version of the unique games conjecture, it is NP-hard to approximate* MaxQP *up to within a factor smaller than* $O(\log n)$*.*

We do not have compelling evidence in favor or against conjecture (2). Nevertheless, we choose to pose it since resolving it to both side has implications. If the conjecture is indeed true then the optimal hardness of approximation result for MaxQP is obtained by Theorem (4). Of course, in this case the standard version of UCG is true, implying, what is by now a host of hardness of approximation results. On the other hand, resolving this conjecture on the negative side is a less ambitious goal then refuting the standard UGC, implies a non-trivial approximation algorithm for finding an assignment to a systems of equations over a finite field that maximizes the number of equations satisfied.

2.1 Analytic notions

In this paper we consider properties of boolean functions over n variables, namely functions over n variables that admit only two values. We consider functions $f\colon \{-1,1\}^n \mapsto \mathbb{R}$ and say a function is *boolean-valued* if its range is $\{-1,1\}$. The domain $\{-1,1\}^n$ is viewed as a probability space under the uniform measure and the set of all functions $f\colon \{-1,1\}^n \mapsto \mathbb{R}$ as an inner product space under $\langle f, g \rangle = \mathbb{E}[fg]$. The associated norm in this space is given by $\|f\|_2 = \sqrt{\mathbb{E}[f^2]}$. We also define the r-norm for every $1 \leq r < \infty$, by $\|f\|_r = (E[|f|^r])^{1/r}$. In addition, let $\|f\|_\infty = \max\{|f(x)|\}$.

Fourier expansion. For $S \subseteq [n]$, let χ_S denote the parity function on S, $\chi_S(x) = \prod_{i \in S} x_i$. It is well known that the set of all such functions forms an orthonormal basis for our inner product space and thus every function $f : \{-1,1\}^n \to \mathbb{R}$ can be expressed as

$$f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S.$$

Here the real quantities $\hat{f}(S) = \langle f, \chi_S \rangle$ are called the *Fourier coefficients* of f and the above is called the *Fourier expansion* of f. *Plancherel's identity* states that $\langle f, g \rangle = \sum_S \hat{f}(S)\hat{g}(S)$ and in particular, $\|f\|_2^2 = \sum_S \hat{f}(S)^2$. Thus if f is boolean-valued then $\sum_S \hat{f}(S)^2 = 1$, and if $f : \{-1,1\}^n \to [-1,1]$ then $\sum_S \hat{f}(S)^2 \leq 1$. We speak of f's squared Fourier coefficients as *weights*, and we speak of the sets S being stratified into *levels* according to $|S|$. So for example, by the *weight of f at level 1* we mean $\sum_{|S|=1} \hat{f}(S)^2$.
For a function f as above we denote its linear part by

$$f^{=1} = \sum_{S \subset [n], |S|=1} \hat{f}(S)\chi_S$$

and similarly its non-linear part by

$$f^{\neq 1} = \sum_{S \subseteq [n], |S| \neq 1} \hat{f}(S)\chi_S \ .$$

Vector functions. In the last part of the paper we consider functions $f\colon \{-1,1\}^n \mapsto \mathbb{S}^{n-1}$, i.e. functions that map into vectors of l_2 norm 1 (vectors that lie on the unit sphere). Such functions can also be represented in the same Fourier basis as

$$f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S.$$

Where now the *Fourier coefficients* of f are allowed to be vectors of norm at most 1, that is, lie inside the unit n-dimensional ball $\hat{f}(S) \in \mathbb{B}^{n-1}$.

3 Hardness of QP

In this section we prove two hardness results for the QP problem, Theorems 1 and 4, as stated before. The proof reduces a label cover instance to an instance of MaxQP by encoding an assignment to the label cover instance using the long code. An assignment over the long code variables is regarded as a boolean function, and the objective value can easily be expressed in terms of the Fourier coefficients of these functions.

The crucial property exploited by our proofs is that the quadratic forms of MaxQP instances can have arbitrary (and in particular negative) coefficients (except for the non-negative trace constraint). Whereas previous reductions of similar nature were geared towards positive coefficient only, we show that arbitrary coefficients allow us to impose strong constraints on the functions derived from assignments. In particular, we can impose the constraints that these functions are very close to being linear - usually the main obstacle in hardness results of this nature.

3.1 The reduction

Given an instance of label cover $\mathcal{L} = \mathcal{L}(V, W, E, [R], \{\sigma_{v,w}\}_{(v,w) \in E})$, we describe a reduction which constructs an instance of MaxQP denoted $\mathcal{Q}_{\mathcal{L}}$. The trace of our initial construction will not be zero, however in Subsection 3.4 we eliminate all non-zero diagonal entries in $\mathcal{Q}_{\mathcal{L}}$.

Parameters. Let $\mathcal{L} = \mathcal{L}(V, W, E, [R], \{\sigma_{v,w}\}_{(v,w)\in E})$ be an instance of label cover, where the size of the instance is $n = |V| + |W|$. The reduction uses three parameters, ν, b, and d, which are set by

$$\nu \stackrel{def}{=} \min\left\{\frac{1}{2n}, \frac{\epsilon}{100R}\right\}, \quad \text{and} \quad b, d \stackrel{def}{=} e^{10R} + 4\nu^{-6} \ .$$

The variables. For every vertex $u \in V \cup W$ of the original instance $\mathcal{L}$, the reduction generates d sets of new variables, denoted $\left\{C_u^i\right\}_{i\in[d]}$. There will be a variable $C_u^i(x) \in C_u^i$ for every element $x \in \{-1,1\}^R$ of the R-dimensional discrete hypercube. The the QP instance $\mathcal{Q}_{\mathcal{L}}$ will therefore be defined over $N \stackrel{def}{=} d(|V| + |W|)2^R$ variables.

The quadratic form. When restricted to a subset C_u^i, an assignment f to the variables of the QP instance can be viewed as a Boolean function f_u^i, defined by $f_u^i(x) \stackrel{def}{=} f(C_u^i(x))$. Let $f_u \stackrel{def}{=} E_{i\in[d]}[f_u^i]$. We write our quadratic form as a convex combination of bilinear forms, defined over the functions f_u^i. We have two kinds of forms: the *internal forms*, and the *external forms*.

- **Internal Forms.** For every $u \in V \cup W$ and every $i, j \in [d]$ we write

  $$T_{u,i,j}(f) \stackrel{def}{=} -b \sum_{S\subseteq[R],\ |S|\neq 1} \widehat{f_u^i}(S)\widehat{f_u^j}(S) \ .$$

  In addition, let

  $$T_u(f) = \mathbb{E}_{i,j\in[d]}[T_{u,i,j}(f)] = -b \sum_{S\subseteq[R],\ |S|\neq 1} \widehat{f_u}^2(S).$$

- **External Forms.** For every edge $(v, w) \in E$ and every $i, j \in [d]$ we write

  $$T_{v,w,i,j}(f) = \sum_{k\in[R]} \widehat{f_v^i}(k)\widehat{f_w^j}(\sigma_{v,w}(k)) \ ,$$

  and let

  $$T_{vw}(f) = \mathbb{E}_{i,j\in[d]}[T_{v,w,i,j}(f)] = \sum_{k\in[R]} \widehat{f_v}(k)\widehat{f_w}(\sigma_{v,w}(k)).$$

Our QP instance is given by the following quadratic form.

$$\mathcal{Q}_{\mathcal{L}}(f) \stackrel{def}{=} \nu\mathbb{E}_{u\in V\cup W}[T_u(f)] + (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \tag{1}$$

This concludes our reduction, except for the removal of diagonal entries. In the next two subsections we continue by proving completeness and soundness properties for the reduction (Lemma 1 and Lemma 2 respectively. We then show in Subsection 3.4 that removing diagonal entries does not change the properties of $\mathcal{Q}_{\mathcal{L}}$ significantly, and finally in Subsection 3.5 we conclude the proof of Theorem 1 and of Theorem 4.

3.2 Completeness

Let $\mathcal{L}$ and $\mathcal{Q}_{\mathcal{L}}$ be as above. Recall[guy: **this was not defined before.. it should be defined in preliminaries**] that the value of $\mathcal{L}$ is the maximal fraction of edges that can be satisfied by a labelling, and that the value of $\mathcal{Q}_{\mathcal{L}}$, $val(\mathcal{Q}_{\mathcal{L}})$, is the maximal value that it can obtain for a Boolean assignment. The following lemma states that the value of $\mathcal{L}$ is a lower bound for the value of $\mathcal{Q}_{\mathcal{L}}$. *

Lemma 1. *If $val(\mathcal{L}) \geq 1 - \varepsilon$, then $val(\mathcal{Q}_{\mathcal{L}}) \geq (1-\varepsilon)(1-\nu)$.*

PROOF: According to the assumption, $\mathcal{L}$ has some labelling $l : V \cup W \to [R]$ satisfying at least $1 - \varepsilon$ of its constraints. We define an assignment f for the QP instance by $f_u^i(x) \stackrel{def}{=} x_{l(u)}$.

The Fourier coefficients of f_u^i are $\widehat{f_u^i}(\{l(u)\}) = 1$, and $\widehat{f_u^i}(S) = 0$ whenever $S \neq \{l(u)\}$. Hence for every $u \in V \cup W$ and $i, j \in [d]$ we have $T_{u,i,j}(f) = 0$, and therefore $T_u = 0$. Next, let $(v, w) \in E$ and $i, j \in [d]$. If the edge (v, w) is satisfied by the labelling, namely $\sigma_{vw}(l(v)) = l(w)$ (this is true for at least a $(1-\varepsilon)$-fraction of the edges), then

$$T_{v,w,i,j}(f) = \sum_{k \in [R]} \delta_{k,l(v)} \delta_{\sigma_{vu}(k),l(w)} = 1.$$

If the edge (v, w) is not satisfied by the labelling then the expression about yields 0. Hence the overall value of the QP instance is

$$\mathcal{Q}_{\mathcal{L}}(f) = \nu \mathbb{E}_{u \in V \cup W}[T_u(f)] + (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \geq (1-\nu)(1-\varepsilon) \geq (1-\varepsilon)(1-\nu) \ .$$

□

3.3 Soundness

Let us state the soundness property of $\mathcal{Q}_{\mathcal{L}}$.

Lemma 2. *If $\mathcal{Q}_{\mathcal{L}}(f) \geq \varepsilon$ for assignment f, then there exists a labelling for $\mathcal{L}$ which satisfies at least a fraction an $\Omega(\varepsilon)$-fraction of the edges.*

Proof. Consider any assignment with $\mathcal{Q}_{\mathcal{L}}(f) \geq \varepsilon$. As a first step, we show that the functions f_u induced by such an assignment are extremely close to being linear functions.

Claim 1. *For all vertices $u \in V \cup W$ it holds that $\|T_v(f)\|_2^2 \leq \frac{1}{\sqrt{b}}$.*

PROOF: Note that, being averages of Boolean functions, the functions f_u take values in $[-1, 1]$. Their L_2 norm is thus bounded by 1. In particular, their Fourier coefficients are bounded by 1 in absolute value.

According to the construction, the absolute value of every T_{vw} form is bounded by:

$$|T_{vw}(f)| \quad = \quad \left|\mathbb{E}_{i,j\in[d]}[T_{vw}(i,j)]\right| = \sum_{k=1}^{R} \left|\widehat{f}_v(\{k\})\right| \left|\widehat{f}_w(\{\sigma_{v,w}(k)\})\right| \leq R$$

For a T_v form we have

$$T_v(f) \quad = \quad -b \sum_{|S| \neq 1} \widehat{f}_v(S)^2 = -b\|f_v^{\neq 1}\|_2^2 \ ,$$

By equation 1 and the assumption $\mathcal{Q}_{\mathcal{L}}(f) \geq \varepsilon$ we have:

$$\varepsilon \leq \mathcal{Q}_{\mathcal{L}}(f) = \nu \mathbb{E}_{u \in V \cup W}[T_u(f)] + (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \leq -\nu b \mathbb{E}_{u \in V \cup W}[\|f_u^{\neq 1}\|_2^2] + R \tag{2}$$

Which implies $\mathbb{E}_{u \in V \cup W}[\|f_u^{\neq 1}\|_2^2 \leq \frac{2R}{\nu b}$. Now suppose that there exists an f such that $\|A_f^{\neq 1}\|_2^2 > \frac{1}{\sqrt{b}}$. This implies:

$$E_f\left[\|A_f^{\neq 1}\|_2^2\right] \geq \frac{1}{n}\left[1 \cdot \frac{1}{\sqrt{b}} + (n-1)\cdot 0\right] = \frac{1}{n\sqrt{b}} > \frac{2R}{\nu b}$$

In contradiction to the previous conclusion.

□

Claim 2. *For all vertices $u \in V \cup W$ it holds that $\sum_{k=1}^{R} |\widehat{f}_v(\{k\})| \leq 2$.*

Proof: By the previous Lemma, $\|f_v^{\neq 1}\|_2^2 \leq \frac{1}{\sqrt{b}} \leq e^{-5R}$.

Now suppose that $\sum_{k=1}^{R} |\widehat{f}_v(\{k\})| > 2$. Since $f_v^{=1}$ is a linear function with coefficients $\{\widehat{f}_v(\{k\}) | k \in [R]\}$, there exists a value $y \in \{+1,-1\}^R$ for which $f_v^{=1}(y) = \sum_{k=1}^{R} |\widehat{f}_v(\{k\})| > 2$. For this y we have $f_v^{\neq 1}(y) = f_v(y) - f_v^{=1}(y) \leq -1$. Therefore,

$$\|f_v^{\neq 1}\|_2^2 \geq 2^{-R} ,$$

and this is a contradiction. □

The following simple argument shows that the expected value of T_{vw} is large for the assignment f.

Claim 3. $\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \geq \frac{1}{2}\varepsilon$.

Proof: We are assuming that $\mathcal{Q}_{\mathcal{L}}(f) = \nu \mathbb{E}_u[T_u(f)] + (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \geq \varepsilon$. Note that $T_u(f) \leq 0$. Hence, $\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \geq \frac{\varepsilon}{1-\nu} \geq \frac{1}{2}\varepsilon$. □

Using the previous claims, we now define a random label assignment as follows. The assignment to every $v \in V \cup W$ is randomly and independently chosen to be k with probability $\frac{1}{2}|\widehat{f}_v(\{k\})|$ (the sum of these probabilities is at most one by Claim 2), and with probability $1 - \frac{1}{2}\sum_k |\widehat{f}_v(\{k\})|$ we leave v un-assigned.

Let c_{vw} be an indicator random variable that is set to 1 if and only if the label assignment above satisfies the label-cover constraint on the edge (v,w).

The expected number of constraints satisfied by our assignment is:

$$\begin{aligned}
\mathbb{E}_{(v,w)\in E(\mathcal{L})}[c_{v,w}] &= \mathbb{E}_{v,w}\left[\sum_{k\in[R]} \frac{1}{2}|\widehat{f}_v(k)| \cdot \frac{1}{2}|\widehat{f}_w(\sigma_{vw}(k))|\right] \\
&\geq \frac{1}{2}\mathbb{E}_{v,w}[\sum_{k\in[R]} \widehat{f}_v(k)\widehat{f}_w(\sigma_{vw}(k))] \\
&= \frac{1}{4}\mathbb{E}_{v,w}\left[T_{vw}\right] \geq \frac{1}{8}\varepsilon
\end{aligned}$$

This completes the proof of Lemma 2.

3.4 Removing the diagonal

The instance $\mathcal{Q}_{\mathcal{L}}$ constructed in the previous section has non-zero trace. However since we took care to have d "copies" of every set of variables, the interaction of any variable set C_u^i with itself, both in T_v and T_{vw}, is negligible. More formally, consider the QP instance $B_{\mathcal{L}}$, that is obtained from $\mathcal{Q}_{\mathcal{L}}$ by removing all terms of the form $T_{u,i,i}(f)$.[guy: **what about terms of the form** $T_{u,u,i,i}$**?**] *

Recall that $T_{u,i,j}(f) = \sum_{|S|=1} \widehat{f_u^i}(S)\widehat{f_u^j}(S) - b\sum_{|S|\neq 1} \widehat{f_u^i}(S)\widehat{f_u^j}(S)$, and therefore

$$|T_{u,i,i}(f)| \leq b \sum_S \widehat{f_u^i}^2(S) = b\ .$$

Hence, for any specific assignment f, the difference in value of $\mathcal{Q}_{\mathcal{L}}$ and $B_{\mathcal{L}}$ is bounded by

$$|\mathcal{Q}_{\mathcal{L}}(f) - B_{\mathcal{L}}(f)| \quad \leq \quad \frac{\nu}{(|V|+|W|)d^2} \sum_{u \in V \cup W} T_{u,i,i} \leq \frac{\nu b}{d^2} = \frac{\nu}{b} \leq e^{-10R}$$

3.5 Concluding the hardness proofs

We can now prove Theorem 1 and Theorem 4 as Lemma 1 and Lemma 2.

Proof of Theorem 1. Given an instance of label cover $\mathcal{L}$ as in theorem 2, construct $B_{\mathcal{L}}$ as described above. The QP instance has the following properties:

1. The size of the instance is $N = O(n^{\log R} \cdot 2^R)$.
2. By lemma 1, if there exists an assignment A satisfying more than a $1-\varepsilon$ fraction of the equations of Ψ, then the value of the QP is at least $1-\varepsilon-o(\varepsilon)$.
3. By lemma 2, if the value of $B_{\mathcal{L}}$ is at least δ, then there exists an assignment that satisfies $\Omega(\delta)$ of the constraints of Ψ.

Set $R = \log^2 n$. Suppose that we could approximate $val(B_{\mathcal{L}})$ in[guy: **opt used to be val, and in any case it was not defined**] polynomial time to a factor better then $O(\log^\gamma N)$. Then if the best assignment for Ψ satisfies fraction 1 of the equations, we can find a solution to the QP instance of value $1 \cdot \log^\gamma(N) = \log^\gamma(n^{\log R} 2^R) = \Omega(\log^\gamma(2^{\log^2 n})) = \Omega(R^\gamma)$. *

On the other hand, if every assignment satisfies at most R^γ of the constraints, then any QP solution will have value at most R^γ. Thus in time $poly(N) = n^{O(\log^2 n)}$ we can distinguish between the two case of the label cover instance. By theorem 2, this implies $NP \subseteq DTIME(n^{\log^3 n})$. □

By applying our reduction to the label cover problem as in Conjecture 4 we can similarly prove the following

Theorem 4. *Assuming the UGC (strong version), it is NP-hard to approximate QP up to a factor of* $\Omega(\log n)$.

4 Explicit Integrality Gap

In this section we prove Theorem 5, showing an explicit family of MaxQP instances with increasing integrality gap. Our construction was inspired by the recent embedding lower bound of Khot and Vishnoi [KV05].

Theorem 5. *There exists a family of* MaxQP *instances of unbounded size, where the integrality gap of instances over n variables is* $\Omega(\frac{\log n}{\log \log n})$.

Notation. For any $n \in \mathbb{N}$ we define an explicit quadratic form as follows. Let $\mathcal{F} \stackrel{def}{=} \{f | f : \{1,-1\}^n \mapsto \{1,-1\}\}$ be the set of all Boolean functions on n bits. Let $R \stackrel{def}{=} 2^n$ and $N = 2^R \stackrel{def}{=} 2^{2^n}$. For any $f \in \mathcal{F}$ and $T \subseteq [n]$, let $f \circ T \in \mathcal{F}$ denote the function defined by $f \circ T(x) = f(x \oplus T)$, where $x \oplus T$ denotes the vector obtained from x by flipping the value of x_k for every $k \in T$.

Let $f \sim_\eta f'$ the distribution on pairs of functions $f, f' \in \mathcal{F}$ where f is chosen uniformly at random and f' is obtained by flipping each value of f independently with probability η. Denote by $\rho \sim_\eta \{\pm 1\}^n$ the distribution on n-bit strings such that each entry is chosen independently to be -1 with probability η and 1 otherwise.

4.1 The construction

Our construction makes use of three parameters, that we fix as follows. Let $\nu = \frac{1}{R^2}$, $b = N^{10}$, $d = b^2 = N^{20}$

Variables. We generate an instance of QP, denoted $I_n = (V, M)$ where V is the set of variables and M a matrix of dimension $|V| \times |V|$. It will be more convenient for us to have more than one label for each variable. That is, we first define a quadratic form over a larger number of variables, and then identify some of them, thereby obtaining a form over a smaller number of variables each having more than one label. The initial set of variables is $V = \{\langle f, g, i\rangle | f, g \in \mathcal{F}, i \in [d]\}$. We define an equivalence relation over the variables by setting $\langle f, g \circ T, i\rangle \equiv \langle f\chi_T, g, i\rangle$ for every subset $T \subseteq [n]$, and identify all the variables that belong to the same equivalence class.

We partition the vertices into disjoint sets by setting

$$V_{f,i} = \{\langle f, g, i\rangle | g \in \mathcal{F}\}.$$

Given an assignment A to the variables (whether a Boolean or a vector assignment), its restriction to $V_{f,i}$ can be viewed as a function over $\mathcal{F}$. We denote this function by A^i_f.

The quadratic form. Our final quadratic form is a convex combination of bilinear forms over the functions A^i_f, which are defined in terms of their Fourier representation.[guy: **add definition of Fourier representation over $\mathcal{F}$ in preliminaries**] As in the case of the hardness reduction, we have *internal forms* and *external forms.* *

- **Internal Forms.** For every $f \in \mathcal{F}$ we let M_f be defined by

$$M_f(A) \stackrel{def}{=} \mathbb{E}_{i,j\in[d]}\left[-b \sum_{|\alpha|\neq 1} \widehat{A^i_f}(\alpha)\widehat{A^j_f}(\alpha)\right] .$$

Note that if we define $A_f \stackrel{def}{=} \mathbb{E}_{i\in[d]}[A^i_f]$, then

$$M_f(A) = -b \sum_{|\alpha|\neq 1} \widehat{A_f}^2(\alpha) \ . \tag{3}$$

- **External forms.** For every $f, f' \in \mathcal{F}$, let $M_{f,f'}$ be defined by

$$M_{f,f'}(A) \stackrel{def}{=} \mathbb{E}_{i,j\in[d]}\left[\sum_{|\alpha|=1} \widehat{A^i_f}(\alpha)\widehat{A^j_f}(\alpha)\right] = \sum_{|\alpha|=1} \widehat{A_f}(\alpha)\widehat{A_{f'}}(\alpha) \tag{4}$$

The final quadratic form is given by the following convex combination of the internal and external forms:

$$M_n(A) \stackrel{def}{=} \nu \cdot \mathbb{E}_{f\in\mathcal{F}}[M_f] + (1-\nu)\cdot \mathbb{E}_{f\sim_\eta f'}\left[M_{f,f'}\right]$$

We now state the main lemma of this section,

Lemma 3. *For every $0<\eta<\frac{1}{2}$, for every large enough n, the* MaxQP *instance I_n satisfies the following properties:*

1. *For every Boolean assignment A, we have $M_n(A) \leq \frac{1}{R^{\frac{\eta}{1-\eta}}}$*
2. *There exists a vector assignment A_v for which $M_n(A) \geq 1-2\eta$.*

Before we prove Lemma 3, let us show how it implies Theorem 5.

Proof:[Proof of Theorem 5.] The number of variables in the instance I_n is $\frac{N^2\cdot d}{R} = O(N^{22})$. According to Lemma 3, the integrality gap is $R^{\frac{\eta}{1-\eta}}\cdot(1-2\eta)$. Fix $\eta = \frac{1}{2}-\frac{1}{\log R}$, then the integrality gap becomes:

$$R^{\frac{\eta}{1-\eta}}\cdot(1-2\eta) = \Omega(R^{1-\frac{2}{\log R}}\cdot\frac{1}{\log R}) = \Omega(\frac{R}{\log R}) = \Omega(\frac{\log N}{\log\log N})$$

□

4.2 Integral solution

In this subsection we prove the first part of Lemma 3.

Lemma 4. *For any Boolean assignment A, the value of the QP instance I_n satisfies $M_n(A) \leq \frac{1}{R^{\frac{\eta}{1-\eta}}}$.*

To prove this lemma, we start by examining a few properties of the boolean functions $\{A_f | f \in \mathcal{F}\}$. The fact that every variable of the instance I_n has several labels implies a certain relationship between the Fourier coefficients of the functions A_f. This is formalized in the following claim.

Claim 4. *For any $T\subseteq[n]$ and any $f\in\mathcal{F}$ it holds $\forall x\subseteq[n] \quad \widehat{A}_f(x) = \widehat{A}_{f\chi_T}(x\circ T)$.*

Proof:

Consider a certain function $f\in\mathcal{F}$ and subset $T\subseteq[n]$. Since for every function $g\in\mathcal{F}$ the vertices $\langle f, g\circ T, i\rangle \equiv \langle f\chi_T, g, i\rangle$ were identified, the assignment A must satisfy:

$$\forall g \quad A_f(g\circ T) = A_{f\chi_T}(g)$$

Writing these equations in Fourier basis we have:

$$\forall g \quad \sum_{x\subseteq[n]} \widehat{A}_f(x)\chi_x(g\circ T) = \sum_{x\subseteq[n]} \widehat{A}_{f\chi_T}(x)\chi_x(g)$$

If the values $\widehat{A}_f(x)$ are fixed, this system of linear equations has one possible solution is $\forall x \quad \widehat{A}_f(x) = \widehat{A}_{f\chi_T}(x\circ T)$. In fact this is the only possible solution, as the linear system of equations above has full rank.

□

Another property of any assignment with $M_n(A) > 0$ is that each A_f is extremely close to being a linear function. The following two claims prove this fact for two different measures of distance - the l_1 and l_2 norms.

Claim 5. *For any assignment such that $M_n(A) > 0$ it holds that $\forall f \in \mathcal{F} \quad \|A_f^{\neq 1}\|_2^2 \leq \frac{1}{N^6}$*

PROOF: By equations 3 and 4 we have

$$M_n(A) = \nu \mathbb{E}_f \left[-b \|A_f^{\neq 1}\|_2^2 \right] + (1-\nu) \mathbb{E}_{f \sim_\eta f' \in \mathcal{F}} \left[\sum_{|\alpha|=1} \widehat{A}_f(\alpha) \widehat{A}_{f'}(\alpha) \right] \tag{5}$$

Assuming $M_n(A) > 0$ this translates to

$$0 < M_n(A) \quad \leq \quad 1 - \nu - b\mathbb{E}_f \left[\|A_f^{\neq 1}\|_2^2 \right] \leq 2 - \nu b \mathbb{E}_f \left[\|A_f^{\neq 1}\|_2^2 \right]$$

According to the choice of parameters we obtain

$$\mathbb{E}_f \left[\|A_f^{\neq 1}\|_2^2 \right] \leq \frac{2}{\nu b} < \frac{1}{N^8}$$

Now suppose that there exists an f such that $\|A_f^{\neq 1}\|_2^2 > \frac{1}{N^6}$. This implies:

$$\mathbb{E}_f \left[\|A_f^{\neq 1}\|_2^2 \right] \geq \frac{1}{N} \left[1 \cdot \frac{1}{N^6} + (N-1) \cdot 0 \right] > \frac{1}{N^8}$$

In contradiction to the previous conclusion.

□

Claim 6.

$$\forall f \in \mathcal{F} \quad \sum_{x \subseteq [n]} |A_f(x)| \leq 2$$

PROOF: By Claim 5, for every $f \in \mathcal{F}$ we have $\|A_f^{\neq 1}\|_2^2 < \frac{1}{N^8} \leq e^{-4R}$.

The rest of the argument is the same as in claim 2.

□

We can now proceed with the proof of Lemma 4.

PROOF:[Lemma 4] Consider any Boolean assignment A to the variables of I. Suppose that $M_n(A) > 0$ (the assignment that achieves the maximum over $M_n(A)$ definitely satisfies this). From equation 5 we have

$$M_n(A) \leq \mathbb{E}_{f \sim_\eta f' \in \mathcal{F}} [\sum_{x \subseteq [n]} \widehat{A}_f(x) \widehat{A}_{f'}(x)]$$

Using the assignment A, we proceed to define a random function $\Phi = \Phi_A : \mathcal{F} \mapsto [R]$ as follows. For every set of the form $\{f\chi_S | S \subseteq [n]\}$ we pick an arbitrary representative f, and set $\Phi_A(f)$ to be x with probability $\frac{1}{2}|\widehat{A}_f(x)|$, and with probability $1 - \frac{1}{2}\sum_x |\widehat{A}_f(x)|$ we set it arbitrarily to zero (note that the above probabilities are indeed non-negative, and that they sum up to 1 by Claim 6).

Once an assignment for f has been chosen, we set $\Phi(f\chi_T) \stackrel{def}{=} \Phi(f) \oplus T$ for every $T \subseteq [n]$. Note that the resulting function Φ must be balanced, that is, it must satisfy

$$\forall x \subseteq [n] \ . \ \Pr_{f \in \mathcal{F}} [\Phi(f) = x] = \frac{1}{R} \tag{6}$$

This implies the following bound on the stability of Φ_A, proven by Khot [Kho] (see proof in the Appendix)

Lemma 5 (Khot). *For any function* $\Psi : \{\pm 1\}^t \mapsto [R]$ *such that* $\forall i \in [R]$. $\Pr_{x\in\{\pm1\}^t}[\Psi(x) = i] = \frac{1}{R}$ *it holds that:*

$$\Pr_{x\sim_\eta x'\in\{\pm1\}^t}[\Psi(x) = \Psi(x')] \leq \frac{1}{R^{\frac{\eta}{1-\eta}}} \ .$$

The usefulness of Φ_A is given by the following claim, in which it is shown to bound $M_n(A)$.

Claim 7.

$$M_n(A) \leq 4 \Pr_{f\sim_\eta f'\in\mathcal{F}}[\Phi(f) = \Phi(f')] + \nu$$

Proof: Let $N(f) = \{f\chi_T | T \subseteq [x]\}$, and $\Pr[f, f'] = \Pr_{\rho\sim_\eta\{\pm1\}^R}[f' = f\rho]$ be the probability of obtaining f' from f under η noise. Let $I(f, f')$ be the indicator random variable that is 1 if and only if $\Phi(f) = \Phi(f')$. By definition of Φ we have:

$$\begin{aligned}
\Pr_{f\sim_\eta f'\in\mathcal{F}}[\Phi(f) = \Phi(f')] &= \frac{1}{|F|^2}\sum_{f,f'\in\mathcal{F}} \Pr[f, f'] \cdot I(f, f') \\
&= \frac{1}{|F|^2}\sum_{f\in\mathcal{F}}\left(\sum_{f'\notin N(f)} \Pr[f, f']I(f, f') + \sum_{T\subseteq[x]} \Pr[f, f\chi_T]I(f, f')\right) \\
&\geq \frac{1}{|F|^2}\sum_{f\in\mathcal{F}}\sum_{f'\notin N(f)} \Pr[f, f']I(f, f') \\
&\qquad \text{since for } f \notin N(f) \text{ the values of } \Phi(f) \text{ and } \Phi(f') \text{ are independent} \\
&= \frac{1}{|F|^2}\sum_{f\in\mathcal{F}}\sum_{f'\notin N(f)} \Pr[f, f'] \sum_{x\subseteq[n]} \frac{1}{2}|\widehat{A}_f(x)|\frac{1}{2}|\widehat{A}_{f'}(x)| \\
&= \frac{1}{4|F|^2}\Bigg(\sum_{f,f'\in\mathcal{F}} \Pr[f, f'] \sum_{x\subseteq[n]} |\widehat{A}_f(x)||\widehat{A}_{f'}(x)| \\
&\qquad - \sum_{f\in\mathcal{F}}\sum_{T\subseteq[n]} \Pr[f, f\chi_T] \sum_{x\subseteq[n]} |\widehat{A}_f(x)||\widehat{A}_{f\chi_T}(x)|\Bigg) \\
&\geq \frac{1}{4}M_n(A) - \frac{2R}{N} \geq \frac{1}{4}M_n(A) - \frac{1}{4}\nu
\end{aligned}$$

□

This lemma together with the bound of Lemma 5 yields

$$M_n(A) \leq \frac{4}{R^{\frac{\eta}{1-\eta}}} + \nu = O\left(\frac{1}{R^{\frac{\eta}{1-\eta}}}\right) \ , \tag{7}$$

Proving Lemma (4).

□

4.3 Vector solution

Consider the vector assignment, given by the Fourier coefficients:

$$\forall\alpha \ \ \widehat{A}_f(\alpha) = \widehat{A}^i_f(\alpha) = \begin{cases} \frac{1}{R} f\chi_T & |\alpha| = 1, \alpha = T \subseteq [n] \\ \\ 0 & o/w \end{cases}$$

Notice that the vectors $\widehat{A}_f(\alpha)$ are orthogonal, and their norms satisfy

$$\forall\alpha \ \ \|\widehat{A}_f(\alpha)\|_2 = \begin{cases} \frac{1}{R} & |\alpha| = 1 \\ \\ 0 & o/w \end{cases} \tag{8}$$

In the standard basis, these vectors can be written as:

$$A_f(g) = A^i_f(g) = \frac{1}{R}\sum_{T\subseteq[n]} g(1 \circ T) \cdot f\chi_T$$

Khot and Vishnoi observed that the above vector assignment assigns the same vector to all vertices in the equivalence classes $\{f\chi_T | T \subseteq [n]\}$.

Lemma 6. *For the vector solution above we have $M_n(A) \geq 1 - 2\eta$.*

Proof: Recall by equation 5

$$\begin{aligned} M_n(A) &= \nu\mathbb{E}_f\left[-b\|A_f^{\neq 1}\|_2^2\right] + (1-\nu)\mathbb{E}_{f\sim_\eta f'\in\mathcal{F}}\left[\sum_\alpha \widehat{A}_f(\alpha)\widehat{A}_{f'}(\alpha)\right] \\ &\quad \text{by equation 8} \\ &= (1-\nu)\mathbb{E}_{f\sim_\eta f'\in\mathcal{F}}\left[\sum_\alpha \widehat{A}_f(\alpha)\widehat{A}_{f'}(\alpha)\right] \\ &= \mathbb{E}_{f\sim_\eta f'\in\mathcal{F}}\left[\sum_{T\subseteq[n]} \frac{1}{R} f\chi_T \cdot \frac{1}{R} f'\chi_T\right] = \mathbb{E}_{f\sim_\eta f'\in\mathcal{F}}\left[\frac{1}{R}\langle f, f'\rangle\right] \\ &= (1-\nu)\cdot(1-2\eta) \geq 1 - 2\eta \end{aligned}$$

□

Lemma 3 now follows from Lemmas 4 and 6.

4.4 Removing the diagonal

As the parameter d is much larger b, we can apply a modification very similar to the corresponding modification in the hardness of approximation result (Subsection 3.4), to obtain a matrix with zero diagonal entries. The details are omitted for brevity.

A Stability of balanced multi-valued functions

For completeness, we provide the proof of Khot's Lemma 5:
PROOF:[Lemma 5] Given Ψ, define $\forall j \in [R]\ \ \Phi_j(x) : \{\pm 1\}^t \mapsto \{0,1\}$ as follows:

$$\Phi_j(x) \stackrel{def}{=} \begin{cases} 1 & \Psi(x) = j \\ \\ 0 & o/w \end{cases}$$

Then:

$$\begin{aligned}
\Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x') = j] &= \Pr_{x \sim_\eta x' \in D}[\Phi_j(x) = \Phi_j(x') = 1] = \mathbb{E}_{x \sim_\eta x' \in \{\pm 1\}^t}[\Phi_j(x)\Phi_j(x')] \\
&= \mathbb{E}_{x \in \{\pm 1\}^t, \rho \sim_\eta \{\pm\}^t}[(\sum_{\alpha \subseteq [t]} \widehat{\Phi_j}(\alpha)\chi_\alpha(x))(\sum_{\beta \subseteq [t]} \widehat{\Phi_j}(\beta)\chi_\beta(x \oplus \rho))] \\
&= \sum_{\alpha \subseteq [t]} \widehat{\Phi_j}^2(\alpha)\mathbb{E}_{\rho \sim_\eta \{\pm\}^t}[\chi_\alpha(\rho)))] = \sum_{\alpha \subseteq [t]} \widehat{\Phi_j}^2(\alpha)(1 - 2\eta)^{|\alpha|} \\
&= \|T_{\sqrt{1-2\eta}}[\Phi_j]\|_2^2
\end{aligned}$$

Where $T_\delta[f]$ is the Beckner operator:

$$T_\delta[f] = \sum_S \delta^{|S|} \widehat{f}(S)\chi_S$$

Now using the Beckner inequality (which states $\|T_\delta[f]\|_p \leq \|f\|_r$ for $r \leq p$, $\delta \leq \sqrt{\frac{r-1}{p-1}}$):

$$\begin{aligned}
\Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x') = j] &= \|T_{\sqrt{1-2\eta}}[\Phi_j]\|_2^2 \\
&\leq \|\Phi_j\|_{2-2\eta}^2 && \text{using Beckner} \\
&= \mathbb{E}_{x \in \{\pm 1\}^t}[\Phi_j(x)^{2-2\eta}]^{2/2-2\eta} \\
&= (\tfrac{1}{R})^{\frac{1}{1-\eta}} && \text{by properties of } \Phi_j
\end{aligned}$$

Therefore:

$$\Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x')] = \sum_{j \in [R]} \Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x') = j] \leq R \cdot \frac{1}{R^{1/1-\eta}} = \frac{1}{R^{\frac{\eta}{1-\eta}}}$$

□

References

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. *Journal of the ACM*, 45:501–555, 1998.

[AMMN] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. Quadratic forms on graphs. *To appear STOC 2005.*

[AN04] Noga Alon and Assaf Naor. Approximating the cut-norm via grothendieck's inequality. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, New York, NY, USA, 2004. ACM Press.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45:70–122, 1998.

[BBC] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Mach. Learn.*, 56(1-3):89–113.

[CW04] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: Extending grothendieck's inequality. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 54–60, Washington, DC, USA, 2004. IEEE Computer Society.

[FK99] A. M. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19:175–200, 1999.

[Has05] J. Hastad. Every 2-csp allows non-trivial approximation. *manuscript*, 2005.

[Kho] Subhash Khot. personal communications, march 2005.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, New York, NY, USA, 2002. ACM Press.

[KO] S. Khot and R. O'Donnell. personal communications, march 2005.

[KS03] B. S. Kashin and S. J. Szarek. On the gram matrices of systems of uniformly bounded functions. *Proceedings of the Steklov Institute of Mathematics*, 243:227–233, 2003.

[KV05] S. Khot and N. Vishnoi. On embeddability of negative type metrics into l_1. *manuscript*, 2005.

[Meg01] A. Megretski. Relaxation of quadratic programs in operator theory and system analysis. *Systems, Approximation, Singular Integral Operators, and Related Topics (Bordeaux, 2000)*, (3):365–392, 2001.

[Nes98] Y. Nesterov. Global quadratic optimization via conic relaxation. *Working paper CORE*, 1998.

[NRT99] A. Nemirovski, C. Roos, and T. Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, 86(3):463–473, 1999.

[Raz98] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.

[Tal03] Michel Talagrand. *Spin Glasses: a Challenge to Mathematicians*, volume 46 of *Ergbnisse der Mathematik und ihrer Grenzgebiete.* New York, 2003.