On Quadratic Programming with a Ratio Objective

Aditya Bhaskara* Moses Charikar† Rajsekar Manokaran‡
Aravindan Vijayaraghavan§

Abstract

Quadratic Programming (QP) is the well-studied problem of maximizing over $\{-1,1\}$ values the quadratic form $\sum_{i \neq j} a_{ij} x_i x_j$. QP captures many known combinatorial optimization problems, and assuming the unique games conjecture, semidefinite programming techniques give optimal approximation algorithms. We extend this body of work by initiating the study of Quadratic Programming problems where the variables take values in the domain $\{-1,0,1\}$. The specific problems we study are

$$\textsf{QP-Ratio}: \quad \max_{\{-1,0,1\}^n} \frac{\sum_{i \neq j} a_{ij} x_i x_j}{\sum x_i^2}, \quad \text{and} \quad \textsf{Normalized QP-Ratio}: \quad \max_{\{-1,0,1\}^n} \frac{\sum_{i \neq j} a_{ij} x_i x_j}{\sum d_i x_i^2}.$$

$$\text{where } d_i = \sum_j |a_{ij}|$$

These are natural relatives of several well studied problems (in fact Trevisan introduced the latter problem as a stepping stone towards a spectral algorithm for Max Cut Gain). These quadratic ratio problems are good testbeds for both algorithms and complexity because the techniques used for quadratic problems for the $\{-1,1\}$ and $\{0,1\}$ domains do not seem to carry over to the $\{-1,0,1\}$ domain. We give approximation algorithms and evidence for the hardness of approximating these problems.

We consider an SDP relaxation obtained by adding constraints to the natural eigenvalue (or SDP) relaxation for this problem. Using this, we obtain an $\tilde{O}(n^{1/3})$ algorithm for QP-ratio. We also obtain an $\tilde{O}(n^{1/4})$ approximation for bipartite graphs, and better algorithms for special cases.

As with other problems with ratio objectives (e.g. uniform sparsest cut), it seems difficult to obtain inapproximability results based on $\mathbf{P} \neq \mathbf{NP}$. We give two results that indicate that QP-Ratio is hard to approximate to within any constant factor: one is based on the assumption that random instances of Max k-AND are hard to approximate, and the other makes a connection to a ratio version of Unique Games.

There is an embarrassingly large gap between our upper bounds and lower bounds. In fact, we give a natural distribution on instances of QP-Ratio for which an n^{ε} approximation (for small ε) seems out of reach of current techniques. On the one hand, this distribution presents a concrete barrier for algorithmic progress. On the other hand, it is a challenging question to develop lower bound machinery to establish a hardness result of n^{ε} for this problem.

*Department of Computer Science, Princeton University, and Center for Computational Intractability. Supported by NSF CCF 0832797. Email: bhaskara@cs.princeton.edu

†Department of Computer Science, Princeton University, and Center for Computational Intractability. Supported by NSF CCF 0832797. Email: moses@cs.princeton.edu

‡Department of Computer Science, Princeton University, and Center for Computational Intractability. Supported by NSF CCF 0832797. Email: rajsekar@cs.princeton.edu

§Department of Computer Science, Princeton University, and Center for Computational Intractability. Supported by NSF CCF 0832797. Email: aravindv@cs.princeton.edu

1 Introduction

Semidefinite programming techniques have proved very useful for quadratic optimization problems (i.e. problems with a quadratic objective) over $\{0,1\}$ variables or $\{\pm 1\}$ variables. Such problems admit natural SDP relaxations and beginning with the seminal work of Goemans and Williamson [GW95], sophisticated techniques have been developed for exploiting these SDP relaxations to obtain approximation algorithms. For a large class of constraint satisfaction problems, a sequence of exciting results[KKMO07, KO06, KV05] culminating in the work of Raghavendra[Rag08], shows that in fact, such SDP based algorithms are optimal (assuming the Unique Games Conjecture).

In this paper, we initiate a study of quadratic programming problems with variables in $\{0, \pm 1\}$. In contrast to their well studied counterparts with variable values in $\{0,1\}$ or $\{\pm 1\}$, to the best of our knowledge, such problems have not been studied before. These problems admit natural SDP relaxations similar to problems with variable values in $\{0,1\}$ or $\{\pm 1\}$, yet we know very little about how (well) these problems can be approximated. We focus on some basic problems in this class:

$$\textsf{QP-Ratio}: \quad \max_{\{-1,0,1\}^n} \frac{\sum_{i \neq j} a_{ij} x_i x_j}{\sum x_i^2}, \quad \text{and} \quad \textsf{Normalized QP-Ratio}: \quad \max_{\{-1,0,1\}^n} \frac{\sum_{i \neq j} a_{ij} x_i x_j}{\sum d_i x_i^2}. \quad (1)$$
$$\text{where } d_i = \sum_j |a_{ij}|$$

Note that the numerator is the well studied quadratic programming objective $\sum_{i<j} a_{i,j} x_i x_j$. Ignoring the value of the denominator for a moment, the numerator can be maximized by setting all variables to be ± 1. However, the denominator term in the objective makes it worthwhile to set variables to 0. An alternate phrasing of the ratio-quadratic programming problems is the following: the goal is to select a subset of non-zero variables S and assign them values in $\{\pm 1\}$ so as to maximize the ratio of the quadratic programming objective $\sum_{i<j \in S} a_{i,j} x_i x_j$ to the (normalized) size of S.

This problem is a variant of well studied problems: Eliminating 0 as a possible value for variables gives rise to the problem of maximizing the numerator over $\{\pm 1\}$ variables – a well studied problem with an $O(\log n)$ approximation [NRT99, Meg01, CW04]. On the other hand, eliminating -1 as a possible value for variables (when the $a_{i,j}$ are non-negative) results in a polynomial time solvable problem. Another closely related problem to QP-Ratio is a *budgeted* variant where the goal is to maximize the numerator (for the QP-Ratio objective) subject to the denominator being at most k. This is harder than QP-Ratio in the sense that an α-approximation for the budgeted version translates to an α-approximation for QP-Ratio (but not vice versa). The budgeted version is a generalization of k-Densest Subgraph, a well known problem for which there is a huge gap between current upper[BCC+10] and lower bounds[Kho04, Fei02]. In this paper, we chose to focus on the "easier" class of ratio problems.

Though it is a natural variant of well studied problems, QP-Ratio seems to fall outside the realm of our current understanding on both the algorithmic and inapproximability fronts. One of the goals of our work is to enhance (and understand the limitations of) the SDP toolkit for approximation algorithms by applying it to this natural problem. On the hardness side, the issues that come up are akin to those arising in other problems with a ratio/expansion flavor, where conventional techniques in inapproximability have been ineffective.

The Normalized QP-Ratio objective arose in recent work of Trevisan[Tre09] on computing Max Cut Gain using eigenvalue techniques. The idea here is to use the eigenvector to come up with a 'good' partial assignment, and recurse. Crucial to this procedure is a quantity called the *GainRatio* defined for a graph; this is a special case of Normalized QP-Ratio where $a_{ij} = -1$ for edges, and 0 otherwise.

1.1 Our results

We first study mathematical programming relaxations for QP-Ratio. The main difficulty in obtaining such relaxations is imposing the constraint that the variables take values $\{-1, 0, 1\}$. Capturing this using convex constraints is the main challenge in obtaining good algorithms for the problem.

We consider a semidefinite programming (SDP) relaxation obtained by adding constraints to the natural eigenvalue relaxation, and round it to obtain an $\widetilde{O}(n^{1/3})$ approximation algorithm. An interesting special case is bipartite instances of QP-Ratio, where the support of a_{ij} is the adjacency matrix of a bipartite graph (akin to bipartite instances of quadratic programming, also known as the Grothendieck problem). For bipartite instances, we obtain an $\tilde{O}(n^{1/4})$ approximation and an almost matching SDP integrality gap of $\Omega(n^{1/4})$.

Our original motivation to study quadratic ratio problems was the GainRatio problem studied in Trevisan [Tre09]. We give a sharp contrast between the strengths of different relaxations for the problem and disprove Trevisan's conjecture that the eigenvalue approach towards Max Cutgain matches the bound achieved by an SDP-based approach[CW04]. See Section 3 for details.

Complementing our algorithmic result for QP-Ratio, we show hardness results for the problem. We first show that there is no PTAS for the problem assuming $P \neq NP$. We also provide evidence that it is hard to approximate to within any constant factor. We remark that current techniques seem insufficient to prove such a result based on standard assumptions (such as $P \neq NP$) – a similar situation exists for other problems with a ratio objective such as sparsest cut.

In Section 4.2 we rule out constant factor approximation algorithms for QP-Ratio assuming that random instances of k-AND are hard to distinguish from 'well-satisfiable' instances. This hypothesis was used as a basis to prove optimal hardness for the so called 2-Catalog problem (see [Fei02]) and has proven fruitful in ruling out $O(1)$-approximations for the densest subgraph problem (see [AAM+11]). It is known that even very strong SDP relaxations (in particular, $\Omega(n)$ rounds of the Lasserre hierarchy) cannot refute this conjecture [Tul09].

We also show a reduction from Ratio UG (a ratio version of the well studied unique games problem), to QP-Ratio. We think that ratio version of Unique Games is an interesting problem worthy of study that could shed light on the complexity of other ratio optimization questions. The technical challenge in our reduction is to develop the required fourier-analytic machinery to tackle PCP-based reductions to ratio problems.

There is a big gap in the approximation guarantee of our algorithm and our inapproximability results. We suspect that the problem is in fact hard to approximate to an n^{ε} factor for some $\varepsilon > 0$. In Section 4.1, we decribe a natural distribution over instances which we believe are hard to approximate up to polynomial factors. Our reduction from k-AND in fact generates hard instances of a similar structure albeit ruling out only constant factor approximations.

2 Algorithms for QP-Ratio

We start with the most natural relaxation for QP-Ratio (1) :

$$\max \frac{\sum_{i,j} A_{ij} x_i x_j}{\sum_i x_i^2} \text{ subject to } x_i \in [-1, 1]$$

(instead of $\{0, \pm 1\}$). The solution to this is precisely the largest eigenvector of A (scaled such that entries are in $[-1, 1]$). However it is easy to construct instances for which this relaxation is bad: if A were the adjacency matrix of a $(n+1)$ vertex star (with v_0 as the center of the star), the relaxation can cheat by setting $x_0 = \frac{1}{2}$ and $x_i = \frac{1}{\sqrt{2n}}$ for $i \in [n]$ to give a gap of $\Omega(\sqrt{n})$ (the integer optimum is at most 1).

We show that SDP relaxations give more power in expressing the constraints $x_i \in \{0, \pm 1\}$? Consider the following relaxation:

$$\max \sum_{i,j} A_{ij} \cdot \langle \mathbf{w}_i, \mathbf{w}_j \rangle \;\text{ subject to }\; \sum_i \mathbf{w}_i^2 = 1, \text{ and}$$
$$|\langle \mathbf{w}_i, \mathbf{w}_j \rangle| \le \mathbf{w}_i^2 \text{ for all } i, j \tag{2}$$

It is easy to see that this is indeed a relaxation: start with an integer solution $\{x_i\}$ with k non-zero x_i, and set $\mathbf{v}_i = (x_i/\sqrt{k}) \cdot \mathbf{v}_0$ for a fixed unit vector $\mathbf{v}_0$.

Without constraint (2), the SDP relaxation is equivalent to the eigenvalue relaxation given above. Roughly speaking, equation (2) tries to impose the constraint that non-zero vectors are of equal length. In the example of the $(n+1)$-vertex star, this relaxation has value equal to the true optimum. In fact, for any instance with $A_{ij} \ge 0$ for all i, j, this relaxation is exact [Cha00].

There are other natural relaxations one can write by viewing the $\{0, \pm 1\}$ requirement like a 3-alphabet CSP. We consider one of these in section 2.5, and show an $\Omega(n^{1/2})$ integrality gap. It is interesting to see if lift and project methods starting with this relaxation can be useful.

In the remainder of the section, we describe a simple $\widetilde{O}(n^{1/3})$ rounding algorithm, which shows that the additional constraints (2) indeed help. We first describe an integrality gap of roughly $n^{1/4}$, as it highlights the issues that arise in rounding the SDP solution.

2.1 Integrality gap instance

Consider a complete bipartite graph on L, R, with $|L| = n^{1/2}$, and $|R| = n$. The edge weights are set to ± 1 uniformly at random. Denote by B the $n^{1/2} \times n$ matrix of edge weights (rows indexed by L and columns by R). A standard Chernoff bound argument shows

Lemma 2.1. *With high probability over the choice of B, we have $\mathsf{opt} \le \sqrt{\log n} \cdot n^{1/4}$.*

Proof. Let $S_1 \subseteq L$, $S_2 \subseteq R$ be of sizes a, b respectively. Consider a solution in which these are the only variables assigned non-zero values (thus we fix some ± 1 values to these variables). Let val denote the value of the numerator. By the Chernoff bound, we have

$$\mathbf{P}[\mathsf{val} \ge c\sqrt{ab}] \le e^{-c^2/3},$$

for any $c > 0$. Now choosing $c = 10\sqrt{(a+b)\log n}$, and taking union bound over all choices for S_1, S_2 and the assignment (there are $\binom{\sqrt{n}}{a}\binom{n}{b}2^{a+b}$ choices overall), we get that w.p. at least $1 - 1/n^3$, no assignment with this choice of a and b gives val bigger than $\sqrt{ab(a+b)\log n}$. The ratio in this case is at most $\sqrt{\log n \cdot \frac{ab}{a+b}} \le \sqrt{\log n} \cdot n^{1/4}$. Now we can take union bound over all possible a and b, thus proving that $\mathsf{opt} \le n^{1/4}$ w.p. at least $1 - 1/n$. ∎

Let us now exhibit an SDP solution with value $n^{1/2}$. Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\sqrt{n}}$ be mutually orthogonal vectors, with each $\mathbf{v}_i^2 = 1/2n^{1/2}$. We assign these vectors to vertices in L. Now to the jth vertex in R, assign the vector $\mathbf{w}_j$ defined by $\mathbf{w}_j = \sum_i B_{ij} \frac{\mathbf{v}_i}{\sqrt{n}}$.

It is easy to check that $\mathbf{w}_j^2 = \sum_i \frac{\mathbf{v}_i^2}{n} = \frac{1}{2n}$. Further, note that for any i, j, we have (since all $\mathbf{v}_i$ are orthogonal) $B_{ij}\langle \mathbf{v}_i, \mathbf{w}_j \rangle = B_{ij}^2 \cdot \frac{\mathbf{v}_i^2}{\sqrt{n}} = \frac{1}{2n}$. This gives $\sum_{i,j} B_{ij}\langle \mathbf{v}_i, \mathbf{w}_j \rangle = n^{3/2} \cdot (1/2n) = n^{1/2}/2$.

From these calculations, we have $\forall i, j$, $|\mathbf{v}_i \cdot \mathbf{w}_j| \le \mathbf{w}_j^2$ (thus satisfying (2); other inequalities of this type are trivially satisfied). Further we saw that $\sum_i \mathbf{v}_i^2 + \sum_j \mathbf{w}_j^2 = 1$. This gives a feasible solution of value $\Omega(n^{1/2})$. Hence the SDP has an $\widetilde{\Omega}(n^{1/4})$ integrality gap.

Connection to the star example. This gap instance can be seen as a collection of $n^{1/2}$ stars (vertices in L are the 'centers'). In each 'co-ordinate' (corresponding to the orthogonal $\mathbf{v}_i$), the assigment looks like a star. $O(\sqrt{n})$ different co-ordinates allow us to satisfy the constraints (2).

This gap instance is bipartite. In such instances it turns out that there is a better rounding algorithm with a ratio $\widetilde{O}(n^{1/4})$ (Section 2.3). Thus to bridge the gap between the algorithm and the integrality gap we need to better understand non-bipartite instances.

2.2 An $O(n^{1/3})$ rounding algorithm

Consider an instance of QP-Ratio defined by $A_{(n\times n)}$. Let $\mathbf{w}_i$ be an optimal solution to the SDP, and let the objective value be denoted sdp. We will sometimes be sloppy w.r.t. logarithmic factors in the analysis.

Since the problem is the same up to scaling the A_{ij}, let us assume that $\max_{i,j}|A_{ij}| = 1$. There is a trivial solution which attains a value 1/2 (if i, j are indices with $|A_{ij}| = 1$, set x_i, x_j to be ± 1 appropriately, and the rest of the x's to 0). Now, since we are aiming for an $\widetilde{O}(n^{1/3})$ approximation, we can assume that $\mathsf{sdp} > n^{1/3}$.

As alluded to earlier (and as can be seen in the gap example), the difficulty is when most of the contribution to sdp is from non-zero vectors with very different lengths. The idea of the algorithm will be to move to a situation in which this does not happen. First, we show that if the vectors indeed have roughly equal length, we can round well. Roughly speaking, the algorithm uses the lengths $\|\mathbf{v}_i\|$ to determine whether to pick i, and then uses the ideas of [CW04] (or the earlier works of [NRT99, Meg01]) applied to the vectors $\frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$.

Lemma 2.2. *Given a vector solution $\{\mathbf{v}_i\}$, with $\mathbf{v}_i^2 \in [\tau/\Delta, \tau]$ for some $\tau > 0$ and $\Delta > 1$, we can round it to obtain an integer solution with cost at least* $\mathsf{sdp}/(\sqrt{\Delta}\log n)$.

Proof. Starting with $\mathbf{v}_i$, we produce vectors $\mathbf{w}_i$ each of which is either 0 or a unit vector, such that

$$\text{If } \frac{\sum_{i,j} A_{ij}\langle \mathbf{v}_i, \mathbf{v}_j\rangle}{\sum_i \mathbf{v}_i^2} = \mathsf{sdp}, \text{ then } \frac{\sum_{i,j} A_{ij}\langle \mathbf{w}_i, \mathbf{w}_j\rangle}{\sum_i \mathbf{w}_i^2} \geq \frac{\mathsf{sdp}}{\sqrt{\Delta}}.$$

Stated this way, we are free to re-scale the $\mathbf{v}_i$, thus we may assume $\tau = 1$. Now note that once we have such $\mathbf{w}_i$, we can throw away the zero vectors and apply the rounding algorithm of [CW04] (with a loss of an $O(\log n)$ approximation factor), to obtain a $0, \pm 1$ solution with value at least $\mathsf{sdp}/(\sqrt{\Delta}\log n)$.

So it suffices to show how to obtain the $\mathbf{w}_i$. Let us set (recall we assumed $\tau = 1$)

$$\mathbf{w}_i = \begin{cases} \mathbf{v}_i/\|\mathbf{v}_i\|, \text{ with prob. } \|\mathbf{v}_i\| \\ 0 \text{ otherwise} \end{cases}$$

(this is done independently for each i). Note that the probability of picking i is proportional to the length of $\mathbf{v}_i$ (as opposed to the typically used square lengths, [CMM06] say). Since $A_{ii} = 0$, we have

$$\frac{\mathbf{E}\left[\sum_{i,j} A_{ij}\langle \mathbf{w}_i, \mathbf{w}_j\rangle\right]}{\mathbf{E}\left[\sum_i \mathbf{w}_i^2\right]} = \frac{\sum_{i,j} A_{ij}\langle \mathbf{v}_i, \mathbf{v}_j\rangle}{\sum_i |\mathbf{v}_i|} \geq \frac{\sum_{i,j} A_{ij}\langle \mathbf{v}_i, \mathbf{v}_j\rangle}{\sqrt{\Delta}\sum_i \mathbf{v}_i^2} = \frac{\mathsf{sdp}}{\sqrt{\Delta}}. \tag{3}$$

The above proof only shows the existence of vectors $\mathbf{w}_i$ which satisfy the bound on the ratio. The proof can be made constructive using the method of conditional expectations, where we set variables one by one, i.e. we first decide whether to make $\mathbf{w}_1$ to be a unit vector along it or the 0 vector, depending on which maintains the ratio to be $\geq \theta = \frac{\mathsf{sdp}}{\sqrt{\Delta}}$. Now, after fixing $\mathbf{w}_1$, we fix $\mathbf{w}_2$ similarly etc., while always maintaining the invariant that the ratio $\geq \theta$.

At step i, let us assume that $\mathbf{w}_1, \ldots, \mathbf{w}_{i-1}$ have already been set to either unit vectors or zero vectors. Consider $\mathbf{v}_i$ and let $\tilde{\mathbf{v}}_i = \mathbf{v}_i/\|\mathbf{v}_i\|$. $\mathbf{w}_i = \tilde{\mathbf{v}}_i$ w.p. $p_i = \|\mathbf{v}_i\|$ and 0 w.p $(1-p_i)$.

In the numerator, $B = \mathbf{E}[\sum_{j\neq i, k\neq i} a_{jk}\langle w_j, w_k\rangle]$ is contribution from terms not involving i. Also let $\mathbf{c}_i = \sum_{k\neq i} a_{ik}\mathbf{w}_k$ and let $\mathbf{c}'_i = \sum_{j\neq i} a_{ji}\mathbf{w}_j$. Then, from equation 3

$$\theta \leq \frac{\mathbf{E}[\sum_{j,k} a_{jk}\langle \mathbf{w}_j, \mathbf{w}_k\rangle]}{E[\sum_j |\mathbf{w}_j|^2]} = \frac{p_i\big(\langle \tilde{\mathbf{v}}_i, \mathbf{c}_i\rangle + \langle \mathbf{c}'_i, \tilde{\mathbf{v}}_i\rangle + B\big) + (1-p_i)B}{p_i\big(1 + \sum_{j\neq i}\|\mathbf{w}_j\|^2\big) + (1-p_i)\big(\sum_{j\neq i}\|\mathbf{w}_j\|^2\big))}$$

Hence, by the simple fact that if c, d are positive and $\frac{a+b}{c+d} > \theta$, then either $\frac{a}{c} > \theta$ or $\frac{b}{d} > \theta$, we see that either by setting $\mathbf{w}_i = \tilde{\mathbf{v}}_i$ or $\mathbf{w}_i = 0$, we get value at least θ. ■

Let us define the 'value' of a set of vectors $\{\mathbf{w}_i\}$ to be $\mathsf{val} := \frac{\sum A_{ij}\langle \mathbf{w}_i, \mathbf{w}_j\rangle}{\sum_i \mathbf{w}_i^2}$. The $\mathbf{v}_i$ we start will have $\mathsf{val} = \mathsf{sdp}$.

Claim 2.3. *We can move to a set of vectors such that (a)* val *is at least* $\mathsf{sdp}/2$*, (b) each non-zero vector* $\mathbf{v}_i$ *satisfies* $\mathbf{v}_i^2 \geq 1/n$*, (c) vectors satisfy (2), and (d)* $\sum_i \mathbf{v}_i^2 \leq 2$.

The proof is by showing that very small vectors can either be enlarged or thrown away .

Proof. Suppose $0 < \mathbf{v}_i^2 < 1/n$ for some i. If $S_i = \sum_j A_{ij}\mathbf{v}_i \cdot \mathbf{v}_j \leq 0$, we can set $\mathbf{v}_i = 0$ and improve the solution. Now if $S_i > 0$, replace $\mathbf{v}_i$ by $\frac{1}{\sqrt{n}} \cdot \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$ (this only increases the value of $\sum_{i,j} A_{ij}\langle \mathbf{v}_i, \mathbf{v}_j\rangle$), and repeat this operation as long as there are vectors with $\mathbf{v}_i^2 < 1/n$. Overall, we would only have increased the value of $\sum_{i,j} A_{ij}\mathbf{v}_i \cdot \mathbf{v}_j$, and we still have $\sum_i \mathbf{v}_i^2 \leq 2$. Further, it is easy to check that $|\langle \mathbf{v}_i, \mathbf{v}_j\rangle| \leq \mathbf{v}_i^2$ also holds in the new solution (though it might not hold in some intermediate step above). ■

The next lemma also gives an upper bound on the lengths – this is where the constraints (2) are crucial.

Lemma 2.4. *Suppose we have a solution of value* Bn^ρ *and* $\sum_i \mathbf{v}_i^2 \leq 2$*. We can move to a solution with value at least* $Bn^\rho/2$*, and* $\mathbf{v}_i^2 < 16/n^\rho$ *for all* i.

Proof. Let $\mathbf{v}_i^2 > 16/n^\rho$ for some index i. Since $|\langle \mathbf{v}_i, \mathbf{v}_j\rangle| \leq \mathbf{v}_j^2$, we have that for each such i,

$$\sum_j A_{ij}\langle \mathbf{v}_i, \mathbf{v}_j\rangle \leq B\sum_j \mathbf{v}_j^2 \leq 2B$$

Thus the contribution of such i to the sum $\sum_{i,j} A_{ij}\langle \mathbf{v}_i, \mathbf{v}_j\rangle$ can be bounded by $m \times 4B$, where m is the number of indices i with $\mathbf{v}_i^2 > 16/n^\rho$. Since the sum of squares is ≤ 2, we must have $m \leq n^\rho/8$, and thus the contribution above is at most $Bn^\rho/2$. Thus the rest of the vectors have a contribution at least $\mathsf{sdp}/2$ (and they have sum of squared-lengths ≤ 2 since we picked only a subset of the vectors) ■

Theorem 2.5. *Suppose A is an $n \times n$ matrix with zero's on the diagonal. Then there exists a polynomial time* $\widetilde{O}(n^{1/3})$ *approximation algorithm for the* QP-Ratio *problem defined by A.*

Proof. As before, let us rescale and assume $\max i, j|A_{ij}| = 1$. Now if $\rho > 1/3$, Lemmas 2.3 and 2.4 allow us to restrict to vectors satisfying $1/n \leq \mathbf{v}_i^2 \leq 4/n^\rho$, and using Lemma 2.2 gives the desired $\widetilde{O}(n^{1/3})$ approximation; if $\rho < 1/3$, then the trivial solution of 1/2 is an $\widetilde{O}(n^{1/3})$ approximation. ■

2.3 The bipartite case

In this section, we prove the following theorem:

Theorem 2.6. *When A is bipartite (i.e. the adjacency matrix of a weighted bipartite graph), there is a (tight upto logarithmic factor) $\widetilde{O}(n^{1/4})$ approximation algorithm for* QP-Ratio *.*

Bipartite instances of QP-Ratio can be seen as the ratio analog of the Grothendieck problem [AN06]. The algorithm works by rounding the semidefinite program relaxation from section 2. As before, let us assume $\max_{i,j} a_{ij} = 1$ and consider a solution to the SDP (2). To simplify the notation, let u_i and v_j denote the vectors on the two sides of the bipartition. Suppose the solution satisfies:

$$(1) \sum_{(i,j)\in E} a_{ij}\langle u_i, v_j\rangle \geq n^{\alpha}, \qquad (2) \sum_i u_i^2 = \sum_j v_j^2 = 1.$$

If the second condition does not hold, we scale up the vectors on the smaller side, losing at most a factor 2. Further, we can assume from Lemma 2.3 that the squared lengths u_i^2, v_j^2 are between $\frac{1}{2n}$ and 1. Let us divide the vectors $\{u_i\}$ and $\{v_j\}$ into $\log n$ groups based on their squared length. There must exist two levels (for the u and v's respectively) whose contribution to the objective is at least $n^{\alpha}/\log^2 n$.[1] Let L denote the set of indices corresponding to these u_i, and R denote the same for v_j. Thus we have $\sum_{i\in L, j\in R} a_{ij}\langle u_i, v_j\rangle \geq n^{\alpha}/\log^2 n$. We may assume, by symmetry that $|L| \leq |R|$. Now since $\sum_j v_j^2 \leq 1$, we have that $v_j^2 \leq 1/|R|$ for all $j \in R$. Also, let us denote by A_j the $|L|$-dimensional vector consisting of the values a_{ij}, $i \in L$. Thus

$$\frac{n^{\alpha}}{\log^2 n} \leq \sum_{i\in L, j\in R} a_{ij}\langle u_i, v_j\rangle \leq \sum_{i\in L, j\in R} |a_{ij}| \cdot v_j^2 \leq \frac{1}{|R|}\sum_{j\in R}\|A_j\|_1. \tag{4}$$

We will construct an assignment $x_i \in \{+1, -1\}$ for $i \in L$ such that $\frac{1}{|R|} \cdot \sum_{j\in R} \big|\sum_{i\in L} a_{ij}x_i\big|$ is 'large'. This suffices, because we can set $y_j \in \{+1, -1\}$, $j \in R$ appropriately to obtain the value above for the objective (this is where it is crucial that the instance is bipartite – there is no contribution due to other y_j's while setting one of them).

Lemma 2.7. *There exists an assignment of $\{+1, -1\}$ to the x_i such that*

$$\sum_{j\in R} \big|\sum_{i\in L} a_{ij}x_i\big| \geq \frac{1}{24}\sum_{j\in R}\|A_j\|_2$$

Furthermore, such an assignment can be found in polynomial time.

Proof. The intuition is the following: suppose $X_i, i \in L$ are i.i.d. $\{+1, -1\}$ random variables. For each j, we would expect (by random walk style argument) that $\mathbf{E}\left[\big|\sum_{i\in L} a_{ij}X_i\big|\right] \approx \|A_j\|_2$, and thus by linearity of expectation, $\mathbf{E}\left[\sum_{j\in R}\big|\sum_{i\in L} a_{ij}X_i\big|\right] \approx \sum_{j\in R}\|A_j\|_2$. Thus the existence of such x_i follows. This can in fact be formalized:

$$\mathbf{E}\left[\big|\sum_{i\in L} a_{ij}X_i\big|\right] \geq \|A_j\|_2/12 \tag{5}$$

This equation is seen to be true from the following lemma

[1] Such a clean division into levels can only be done in the bipartite case – in general there could be negative contribution from 'within' the level.

Lemma 2.8. *Let $b_1, \dots, b_n \in \mathbb{R}$ with $\sum_i b_i^2 = 1$, and let $X_1, \dots, X_n$ be i.i.d. $\{+1, -1\}$ r.v.s. Then*

$$\mathbf{E}[|\sum_i b_i X_i|] \geq 1/12.$$

Proof. Define the r.v. $Z := \sum_i b_i X_i$. Because the X_i are i.i.d. $\{+1, -1\}$, we have $\mathbf{E}[Z^2] = \sum_i b_i^2 = 1$. Further, $\mathbf{E}[Z^4] = \sum_i b_i^4 + 6 \sum_{i<j} b_i^2 b_j^2 < 3(\sum_i b_i^2)^2 = 3$. Thus by Paley-Zygmund inequality,

$$\mathbf{P}[Z^2 \geq \frac{1}{4}] \geq \frac{9}{16} \cdot \frac{(\mathbf{E}[Z^2])^2}{\mathbf{E}[Z^4]} \geq \frac{3}{16}.$$

Thus $|Z| \geq 1/2$ with probability at least $3/16 > 1/6$, and hence $\mathbf{E}[|Z|] \geq 1/12$. ■

We can also make the above constructive. Let r.v. $S := \sum_{j \in R} \left| \sum_{i \in L} a_{ij} X_i \right|$. It is a non-negative random variable, and for every choice of X_i, we have

$$S \leq \sum_{j \in R} \sum_{i \in L} |a_{ij}| \leq L^{1/2} \sum_{j \in R} \|A_j\|_2 \leq n^{1/2}\, \mathbf{E}[S]$$

Let p denote $\mathbf{P}[S < \frac{\mathbf{E}[S]}{2}]$. Then from the above inequality, we have that $(1-p) \geq \frac{1}{2n^{1/2}}$. Thus if we sample the X_i say n times (independently), we *hit* an assignment with a large value of S with high probability. ■

Proof of Theorem 2.6. By Lemma 2.7 and Eq (4), there exists an assignment to x_i, and a corresponding assignment of $\{+1, -1\}$ to y_j such that the value of the solution is at least

$$\frac{1}{|R|} \cdot \sum_{j \in R} \|A_j\|_2 \geq \frac{1}{|R|\ |L|^{1/2}} \sum_{j \in R} \|A_j\|_1 \geq \frac{n^\alpha}{|L|^{1/2} \log^2 n}. \qquad \text{[By Cauchy Schwarz]}$$

Now if $|L| \leq n^{1/2}$, we are done because we obtain an approximation ratio of $O(n^{1/4} \log^2 n)$. On the other hand if $|L| > n^{1/2}$ then we must have $\|u_i\|_2^2 \leq 1/n^{1/2}$. Since we started with u_i^2 and v_i^2 being at least $1/2n$ (Lemma 2.3) we have that all the squared lengths are within a factor $O(n^{1/2})$ of each other. Thus by Lemma 2.2 we obtain an approximation ratio of $O(n^{1/4} \log n)$. This completes the proof. ■

2.4 Algorithms for special cases

2.4.1 Poly-logarithmic approximations for positive semidefinite matrices

The MaxQP problem has a better approximation guarantee (of $2/\pi$) when A is psd. Even for the QP-Ratio problem, we can do better in this case than for general A. In fact, it is easy to obtain a polylog(n) approximation.

This proceeds as follows: start with a solution to the eigenvalue relaxation (call the value ρ). Since A is psd, the numerator can be seen as $\sum_i (B_i x)^2$, where B_i are linear forms. Now divide the x_i into $O(\log n)$ levels depending on their absolute value (need to show that x_i are not too small – poly in $1/n$, $1/|A|_\infty$). We can now see each term $B_i x_i$ a sum of $O(\log n)$ terms (grouping by level). Call these terms $C_i^1, \dots, C_i^\ell$, where ℓ is the number of levels. The numerator is upper bounded by $\ell(\sum_i \sum_j (C_i^j)^2)$, and thus there is some j such that $\sum_i (C_i^j)^2$ is at least $1/\log^2 n$ times the numerator. Now work with a solution y which sets $y_i = x_i$ if x_i is in the jth level and 0 otherwise. This is a solution to the ratio question with value at least ρ/ℓ^2. Further, each $|y_i|$ is either 0 or in $[\rho, 2\rho]$, for some ρ.

From this we can move to a solution with $|y_i|$ either 0 or 2ρ as follows: focus on the numerator, and consider some $x_i \neq 0$ with $|x_i| < 2\rho$ (strictly). Fixing the other variables, the numerator is a convex function of x_i in the interval $[-2\rho, 2\rho]$ (it is a quadratic function, with non-negative coefficient to the x_i^2 term, since A is psd). Thus there is a choice of $x_i = \pm 2\rho$ which only increases the numerator. Perform this operation until there are no $x_i \neq 0$ with $|x_i| < 2\rho$. This process increases each $|x_i|$ by a factor at most 2. Thus the new solution has a ratio at least half that of the original one. Combining these two steps, we obtain an $O(\log^2 n)$ approximation algorithm.

2.4.2 Better approximations when the optimum is large.

We can also obtain a much better approximation algorithm for QP-Ratio when the maximum value of the instance is large, say εd_{max}, where $d_{max} = \max_i \sum_i |a_{ij}|$. For QP-Ratio instances A with $OPT(A) \geq \varepsilon d_{max}$, we can find a solution of value $e^{-O(1/\varepsilon)} d_{max}$ using techniques from section 3.

This is because when all the degrees d_i are roughly equal (say $\gamma d_{max} \leq d_i \leq d_{max}$ for some constant $\gamma > 0$), then it is easy to check that an $O(\alpha)$ approximation to Normalized QP-Ratio (defined in section 3) is an $O(\alpha/\gamma)$ approximation to the same instance of QP-Ratio. Further, when $OPT(\text{QP-Ratio}) \geq \varepsilon d_{max}$, we can throw away vertices i of degree $d_i < \frac{\varepsilon}{2} d_{max}$ without losing in the objective. Hence, for a QP-Ratio instance A when $OPT(A) \geq \varepsilon d_{max}$, we can find a solution to QP-Ratio of value $e^{-O(1/\varepsilon)} d_{max}$.

2.5 Other Relaxations for QP-Ratio

For problems in which variables can take more than two values (e.g. CSPs with alphabet size $r > 2$), it is common to use a relaxation where for every vertex u (assume an underlying graph), we have variables $x_u^{(1)}, .., x_u^{(r)}$, and constraints such as $\langle x_u^{(i)}, x_u^{(j)} \rangle = 0$ and $\sum_i \langle x_u^{(i)}, x_u^{(i)} \rangle = 1$ (intended solution being one with precisely one of these variables being 1 and the rest 0).

We can use such a relaxation for our problem as well: for every x_i, we have three vectors a_i, b_i, and c_i, which are supposed to be 1 if $x_i = 0, 1$, and -1 respectively (and 0 otherwise). In these terms, the objective becomes

$$\sum_{i,j} A_{ij} \langle b_i, b_j \rangle - \langle b_i, c_j \rangle - \langle c_i, b_j \rangle + \langle c_i, c_j \rangle = \sum_{i,j} A_{ij} \langle b_i - c_i, b_j - c_j \rangle.$$

The following constraints can be added

$$\sum_i b_i^2 + c_i^2 = 1 \tag{6}$$

$$\langle a_i, b_j \rangle, \langle b_i, c_j \rangle, \langle a_i, c_j \rangle \geq 0 \text{ for all } i, j \tag{7}$$

$$\langle a_i, a_j \rangle, \langle b_i, b_j \rangle, \langle c_i, c_j \rangle \geq 0 \text{ for all } i, j \tag{8}$$

$$\langle a_i, b_i \rangle = \langle b_i, c_i \rangle = \langle a_i, c_i \rangle = 0 \tag{9}$$

$$a_i^2 + b_i^2 + c_i^2 = 1 \text{ for all } i \tag{10}$$

Let us now see why this relaxation does not perform better than the one in (2). Suppose we start with a vector solution $\mathbf{w}_i$ to the earlier program. Suppose these are vectors in $\mathbb{R}^d$. We consider vectors in $\mathbb{R}^{n+d+1}$, which we define using standard direct sum notation (to be understood as concatenating co-ordinates). Here e_i is a vector in $\mathbb{R}^n$ with 1 in the ith position and 0 elsewhere. Let 0_n denote the 0 vector in $\mathbb{R}^n$.

We set (the last term is just a one-dim vector)

$$b_i = 0_n \oplus {}^{\mathbf{w}_i}/_2 \oplus ({}^{|\mathbf{w}_i|}/_2)$$
$$c_i = 0_n \oplus -{}^{\mathbf{w}_i}/_2 \oplus ({}^{|\mathbf{w}_i|}/_2)$$
$$a_i = \sqrt{1 - \mathbf{w}_i^2} \cdot e_i \oplus 0_d \oplus (0)$$

It is easy to check that $\langle a_i, b_j \rangle = \langle a_i, c_j \rangle = 0$, and $\langle b_i, c_j \rangle = {}^1/_4 \cdot (-\langle \mathbf{w}_i, \mathbf{w}_j \rangle + |\mathbf{w}_i||\mathbf{w}_j|) \geq 0$ for all i, j (and for $i = j$, $\langle b_i, c_i \rangle = 0$). Also, $b_i^2 + c_i^2 = \mathbf{w}_i^2 = 1 - a_i^2$. Further, $\langle b_i, b_j \rangle = {}^1/_4 \cdot (\langle \mathbf{w}_i, \mathbf{w}_j \rangle + |\mathbf{w}_i||\mathbf{w}_j|) \geq 0$. Last but not least, it can be seen that the objective value is

$$\sum_{i,j} A_{ij} \langle b_i - c_i, b_j - c_j \rangle = \sum_{i,j} A_{ij} \langle \mathbf{w}_i, \mathbf{w}_j \rangle,$$

as desired. Note that we never even used the inequalities (2), so it is only as strong as the eigenvalue relaxation (and weaker than the sdp relaxation we consider).

Additional valid constraints of the form $a_i + b_i + c_i = v_0$ (where v_0 is a designated fixed vector) can be introduced – however it it can be easily seen that these do not add any power to the relaxation.

3 Normalized QP-Ratio

Given any symmetric matrix A, the normalized QP-Ratio problem aims to find the best $\{-1, 0, 1\}$ assignment which maximizes the following:

$$\max_{\mathbf{x} \in \{-1,0,1\}^n} \frac{\sum_{i \neq j} 2 a_{ij} x_i x_j}{\sum_{i \neq j} |a_{ij}|(x_i^2 + x_j^2)} \tag{11}$$
$$= \frac{x^t A x}{\sum_i d_i x_i^2} \qquad \text{where } d_i = \sum_j |a_{ij}| \text{ are "the degrees"}$$

Note that when the degrees d_i are all equal ($d_i = d \quad \forall i$), this is the same as QP-Ratio upto a scaling. Though the two objectives have a very similar flavor, the normalized objective tends to penalize picking vertices of high degree in the solution.

This problem was recently considered by Trevisan [Tre09] in the special case when $A = -W(G)$ where $W(G)$ are the matrix of edge weights (0 if there is no edge) and called this quantity the *GainRatio* of G. He gave an algorithm for Max Cut-Gain which uses GainRatio as a subroutine, based purely on an eigenvalue relaxation (as opposed to the SDP-based algorithm of [CW04]). His algorithm for GainRatio can also be adapted to give an algorithm for Normalized QP-Ratio with a similar guarantee. We sketch it below.

3.1 Algorithm based on [Tre09]

Consider the natural relaxation

$$\max_{\mathbf{x} \in [-1,1]^n} \frac{x^t A x}{\sum_i d_i x_i^2} \tag{12}$$

This is also the maximum eigenvalue of $D^{-1/2} A D^{1/2}$ where D is the diagonal matrix of degrees. Trevisan [Tre09] gave a randomized rounding technique which uses just threshold cuts to give the following guarantee.

Lemma 3.1. *[Tre09] In the notation stated above, for every $\gamma > 0$, there exists $c_1, c_2 > 0$ with $c_1 c_2 \leq \gamma e^{1/\gamma}$, such that given any $\boldsymbol{x} \in \mathbb{R}^n$ s.t. $\boldsymbol{x}^t A \boldsymbol{x} \geq \varepsilon \boldsymbol{x}^t D \boldsymbol{x}$, outputs a distribution over discrete vectors $\{-1, 0, 1\}^n$ (using threshold cuts) with the properties:*

1. $|c_1 \mathbf{E}\, Y_i Y_j - x_i x_j| \leq \gamma(x_i^2 + x_j^2)$
2. $\mathbf{E}\, |Y_i| \leq c_2 x_i^2$

Proposition 3.2. *Given a Normalized QP-Ratio instance A with value at least ε finds a solution $y \in \{-1, 0, 1\}^n$ of value $e^{-O(1/\varepsilon)}$.*

Proof. For the eigenvalue relaxation equation 12, there is a feasible solution $\boldsymbol{x}$ such that $\boldsymbol{x}^t A \boldsymbol{x} \geq \varepsilon \boldsymbol{x}^t D \boldsymbol{x}$ where $D_i i = \sum_j |a_{ij}|$ and $D_{ij} = 0$ for $i \neq j$. Now applying Lemma 3.1, we have

$$\begin{aligned}
\mathbf{E}[a_{ij} Y_i Y_j] &\geq \frac{1}{c_1}\big(a_{ij} x_i x_j - \gamma |a_{ij}|(x_i^2 + x_j^2)\big) \\
\mathbf{E}[\sum_{ij} a_{ij} Y_i Y_j] &\geq \frac{1}{c_1}\big(\boldsymbol{x}^t A \boldsymbol{x} - 2\gamma \boldsymbol{x}^t D \boldsymbol{x}\big) \\
&\geq \frac{(\varepsilon - 2\gamma)}{c_1}(\boldsymbol{x}^t D \boldsymbol{x})
\end{aligned}$$

Also, $\mathbf{E}[\sum_i d_i |Y_i|] \leq c_2 \boldsymbol{x}^t D \boldsymbol{x}$. Hence, there exists some vector $\boldsymbol{y} \in \{-1, 0, 1\}^n$ of value $(\varepsilon - 2\gamma)/c_1 c_2$, which shows what we need for sufficiently small $\gamma < \varepsilon/2$. As in previous section 2, this can also be derandomized using the method of conditional expectations (in fact, since the distribution is just over threshold cuts, it suffices to run over all n threshold cuts to find the vector $\boldsymbol{y}$). ∎

3.2 Eigenvalue relaxation for Max Cut-Gain

As mentioned earlier Trevisan [Tre09] shows that if the eigenvalue is ε, the GainRatio is at least $e^{-O(1/\varepsilon)}$. He also conjectures that there could a better dependence: that the GainRatio is at least $\varepsilon/\log(1/\varepsilon)$, whenever eigenval $= \varepsilon$. This would give an eigenvalue based algorithm which matches the SDP-based algorithm of [CW04]. We show that this conjecture is false, and describe an instance for which eigenval is ε, but the GainRatio is at most $\exp(-1/\varepsilon^{1/4})$. This shows that the eigenvalue based approach is necessarily 'exponentially' weaker than an SDP-based one. Roughly speaking, SDPs are stronger because they can enforce vectors to be all of equal length, while this cannot be done in an eigenvalue relaxation. First, let us recall the eigenvalue relaxation for Max CutGain

$$\text{Eig} = \max_{x_u \in [-1,1]} \frac{\sum_{\{u,v\} \in E(G)} -w_{uv} x_u x_v}{\sum_{\{u,v\} \in E(G)} |w_{uv}|(x_u^2 + x_v^2)}.$$

Description of the instance. In what follows, let us fix ε to be a small constant, and write $M = 1/\varepsilon$ (thought of as an integer), and $m = 2/\varepsilon$.

The vertex set is $V = V_1 \cup V_2 \cup \cdots \cup V_m$, where $|V_i| = M^i$. We place a clique with edge weight 1 on each set V_i. Between V_i and V_{i+1}, we place a complete bipartite graph with edge-weight $(1/2+\varepsilon)$. We will call V_i the ith level. Thus the total weight of the edges in the ith level is roughly $M^{2i}/2$, and the weight of edges between levels i and $(i+1)$ is $(1/2 + \varepsilon)M^{2i+1}$.

Lemma 3.3. *There exist $x_i \in [-1, 1]$ such that $\frac{\sum_{\{u,v\} \in E(G)} -w_{uv} \cdot x_u x_v}{\sum_{\{u,v\} \in E(G)} |w_{uv}|(x_u^2 + x_v^2)} = \Omega\big(\varepsilon^2\big)$.*

Proof. Consider a solution in which vertices u in level i have $x_u = (-1)^i \varepsilon^i$. We have

$$\begin{aligned}
\frac{\sum_{\{u,v\}\in E(G)} -w_{uv}\cdot x_u x_v}{\sum_{\{u,v\}\in E(G)} x_u^2 + x_v^2} &= \frac{-N_0^2\sum_{i=1}^{m}(M^{2i}/2)\varepsilon^{2i} + N_0^2\sum_{i=1}^{m-1}\varepsilon^{2i+1}(\frac{1}{2}+\varepsilon)M^{2i+1}}{N_0^2\sum_{i=1}^{m}(M^{2i}/2)\cdot 2\varepsilon^{2i} + N_0^2\sum_{i=1}^{m-1}(\frac{1}{2}+\varepsilon)M^{2m+1}(\varepsilon^{2i}+\varepsilon^{2i+2})} \\
&\geq \frac{-\frac{m}{2} + (m-1)(\frac{1}{2}+\varepsilon)}{\frac{3}{\varepsilon}m} \qquad \text{(noting } M\varepsilon = 1) \\
&= \Omega(\varepsilon^2) \qquad \text{(setting } m \approx \frac{2}{\varepsilon})
\end{aligned}$$

■

Let us now prove an upper bound on the GainRatio of G. Consider the optimal solution Y. Let the fraction of vertices u in level i with non-zero Y_u be λ_i. Of these, suppose $(\frac{1}{2}+\eta_i)$ fraction have $Y_u = +1$ and $(\frac{1}{2}-\eta_i)$ have $Y_u = -1$. It is easy to see that we may assume η_i's alternate in sign Thus, for convenience, we will let η_i denote the negated values for the alternate levels and treat all η_i's as positive.

With these parameters, we see that

$$\begin{aligned}
\text{Numerator} = \sum_{i=1}^{m} M^{2i}\lambda_i^2\Big[-\frac{1}{2}\big(\frac{1}{2}+\eta_i\big)^2 - \frac{1}{2}\big(\frac{1}{2}-\eta_i\big)^2 + \big(\frac{1}{2}+\eta_i\big)\big(\frac{1}{2}-\eta_i\big)\Big] \\
+ \sum_{i=1}^{m-1}\big(\frac{1}{2}+\varepsilon\big)M^{2i+1}\lambda_i\lambda_{i+1}\Big[\big(\frac{1}{2}+\eta_i\big)\big(\frac{1}{2}+\eta_{i+1}\big) + \big(\frac{1}{2}-\eta_i\big)\big(\frac{1}{2}-\eta_{i+1}\big) \\
- \big(\frac{1}{2}+\eta_i\big)\big(\frac{1}{2}-\eta_{i+1}\big) - \big(\frac{1}{2}-\eta_i\big)\big(\frac{1}{2}+\eta_{i+1}\big)\Big] \\
= 2\Big(\sum_{i=1}^{m} -M^{2i}\lambda_i^2\eta_i^2 + (1+2\varepsilon)\sum_{i=1}^{m-1} M^{2i+1}\lambda_i\lambda_{i+1}\eta_i\eta_{i+1}\Big)
\end{aligned}$$

Hence the numerator is

$$\text{Numerator} = 2\Big(\sum_{i=1}^{m} -M^{2i}\lambda_i^2\eta_i^2 + (1+2\varepsilon)\sum_{i=1}^{m-1} M^{2i+1}\lambda_i\lambda_{i+1}\eta_i\eta_{i+1}\Big) \tag{13}$$

Note that the denominator is at least $\sum_i \lambda_i M^{2i}$ (there is a contribution from every edge at least one end-point of which has Y nonzero). We will in fact upper bound the quantity Numerator$/2\sum_i \lambda_i\eta_i M^i$. This clearly gives an upper bound on the ratio we are interested in (as the η_i are smaller than 1/2). Let us write $\gamma_i = \lambda_i\eta_i$. We are now ready to prove the theorem which implies the desired gap. A simple inequality useful in the proof is the following (it follows from the well-known fact that the largest eigenvalue of the length n path is $\cos(\frac{\pi}{n+1}) \approx 1 - \frac{1}{n^2}$):

$$\forall n > 1 \text{ and } x_i \in \mathbb{R},\ x_1^2 + x_2^2 + \dots x_n^2 \geq \big(1 + \frac{1}{n^2}\big)(x_1x_2 + x_2x_3 + \dots x_{n-1}x_n) \tag{14}$$

Theorem 3.4. *Let $\gamma_i \geq 0$ be real numbers in $[0,1]$, and let ε, M, m be as before. Then*

$$\frac{-\sum_{i=1}^{m}\gamma_i^2 M^{2i} + (1+2\varepsilon)\sum_{i=1}^{m-1}\gamma_i\gamma_{i+1}M^{2i+1}}{\sum_{i=1}^{m}\gamma_i M^{2i}} < \frac{1}{M^{\sqrt{m}/4}} \tag{15}$$

Proof. Consider the numbers $\gamma_i M^i$ and let r be the index where it is maximized. Denote this maximum value by D.

Claim. Suppose $1 \le j \le m$ and $j \notin [r - \frac{m^{1/2}}{2}, r + \frac{m^{1/2}}{2}]$ and $\gamma_j M^j \ge \frac{D}{M^{\sqrt{m}/4}}$. Then (15) holds.

Suppose first that $j > r + \frac{m^{1/2}}{2}$. The numerator numerator of (15) is at most $D^2 \times 3m$ while the denominator is at least $\gamma_j M^{2j} > \frac{D}{M^{\sqrt{m}/4}} \times M^j > \frac{D}{M^{\sqrt{m}/4}} \times M^{r+\frac{1}{2}\sqrt{m}} > D^2 \times M^{\sqrt{m}/4}$. (the last inequality is because $D < M^r$, since $\gamma_r < 1$). This implies that the ratio is at most $\frac{1}{M^{\sqrt{m}/4}}$, and hence (15) holds.

Next, suppose $j < r - \frac{m^{1/2}}{2}$. This means, since $\gamma_j < 1$, that $\gamma_r < M^{-\sqrt{m}/4}$. Thus the numerator of (15) is bounded from above by $D \times 3m$ as above, while the denominator is at least $\gamma_r M^{2r} = D^2/\gamma_r > D^2 M^{\sqrt{m}/4}$. Thus the ratio is at most $\frac{1}{M^{\sqrt{m}/4}}$, proving the claim.

Thus for all indices $j \notin [r - \frac{1}{2}\sqrt{m}, r + \frac{1}{2}\sqrt{m}]$ (let us call this interval $\mathcal{I}$), we have $\gamma_j M^j < \frac{D}{M^{\sqrt{m}/4}}$. We thus split the numerator of (15) as

$$\Big(- \sum_{i \in \mathcal{I}} \gamma_i^2 M^{2i} + (1 + 2\varepsilon) \cdot \sum_{i,(i+1) \in \mathcal{I}} \gamma_i \gamma_{i+1} M^{2i+1} \Big) + \text{remaining terms}$$

Note that the part in the parenthesis is ≤ 0 by suitable application of Eq. (14), while the remaining terms are each smaller than $\frac{D^2}{M^{\sqrt{m}/2}}$. Thus the numerator is $< m \frac{D^2}{M^{\sqrt{m}/2}}$. Note that the denominator (of (15)) is at least D^2. These two together complete the proof of the theorem. ∎

Moving to an unweighted instance Now we will show that by choosing N_0 large enough (recall that we chose V_i of size $N_0 M^i$), we can bound how far cuts are from expectation. Let us start with a simple lemma.

Lemma 3.5. *Suppose A and B are two sets of vertices with m and n vertices resp. Suppose each edge is added independently at random w.p. $(\frac{1}{2} + \varepsilon)$. Then*

$$\mathbf{P}\left[|\#\ \mathit{Edges} - (\frac{1}{2} + \varepsilon) mn| > t\sqrt{mn} \right] < e^{-t^2/2}$$

Proof. Follows from Chernoff bounds (concentration of binomial r.v.s) ∎

Next, we look at an arbitrary partitioning of vertices with λ_i and ε_i values as defined previously (λ_i of the Y's nonzero, and $(\frac{1}{2} + \varepsilon_i)$ of them of some sign).

Let us denote $n_i = N_0 \lambda_i M^i$. Between levels i and $i+1$, the probability that the number of edges differs from expectation by $(n_i n_{i+1})^{1/3}$ is at most (by the lemma above) $e^{-(n_i n_{i+1})^{2/3}}$. By choosing N_0 big enough (say M^{10m}) we can make this quantity smaller than e^{-12m} (since each n_i is at least M^{9m}). Thus the probability that the sum of the 'errors' over the m levels is larger than $\mathrm{Err} = \sum (n_i n_{i+1})^{1/3}$ is at most me^{-12m}.

The total number of vertices in the subgraph (with Y nonzero) is at most $N_0 M^m$. Thus the number of cuts is $2^{N_0 M^m} < e^{11m}$. Thus there exists a graph where none of the cuts have sum of the 'errors' as above bigger than Err.

Now it just remains to bound $\frac{\sum (n_i n_{i+1})^{1/3}}{\sum \lambda_i N_0^2 M^{2i}}$. Here again the fact that N_0 is big comes to the rescue (there is only a $N_0^{2/3}$ in the numerator) and hence we are done.

4 Hardness of Approximating QP-Ratio

Given that our algorithmic techniques give only an $n^{1/3}$ approximation in general, and the natural relaxations do not seem to help, it is natural to ask how hard we expect the problem to be. Our results in this direction are as follows: we show that the problem is APX-hard (i.e., there is no PTAS unless $P = NP$). Next, we show that there cannot be a constant factor approximation assuming that Max k-AND is hard to approximate 'on average' (related assumptions are explored in [Fei02]). Our reduction therefore gives a (fairly) natural *hard distribution* for the QP-Ratio problem.

4.1 Candidate Hard Instances

As can be seen from the above, there is an embarrassingly large gap between our upper bounds and lower bounds. We attempt to justify this by describing a natural distribution on instances we do not know how to approximate to a factor better than n^δ (for some fixed $\delta > 0$).

Let $\mathcal{G}$ denote a bipartite random graph with vertex sets V_L of size n and V_R of size $n^{2/3}$, left degree n^δ for some small δ (say 1/10) [i.e., each edge between V_L and V_R is picked i.i.d. with prob. $n^{-(9/10)}$]. Next, we pick a random (planted) subset P_L of V_L of size $n^{2/3}$ and random assignments $\rho_L : P_L \mapsto \{+1, -1\}$ and $\rho_R : V_R \mapsto \{+1, -1\}$. For an edge between $i \in P_L$ and $j \in V_R$, the weight $a_{ij} := \rho_L(i)\rho_R(j)$. For all other edges we assign $a_{ij} = \pm 1$ independently at random.

The optimum value of such a *planted* instance is roughly n^δ, because the assignment of ρ_L, ρ_R (and assigning 0 to $V_L \setminus P_L$) gives a solution of value n^δ. However, for $\delta < 1/6$, we do not know how to find such a planted assignment: simple counting and spectral approaches do not seem to help.

Making progress on such instances appears to be crucial to improving the algorithm or the hardness results. We remark that the instances produced by the reduction from Random k-AND are in fact similar in essence. We also note the similarity to other problems which are beyond current techniques, such as the Planted Clique and Planted Densest Subgraph problems [BCC+10].

4.2 Reduction from Random k-AND

We start out by quoting the assumption we use.

Conjecture 4.1 (Hypothesis 3 in [Fei02])**.** *For some constant $c > 0$, for every k, there is a Δ_0, such that for every $\Delta > \Delta_0$, there is no polynomial time algorithm that, on most k-AND formulas with n-variables and $m = \Delta n$ clauses, outputs '`typical`', but never outputs '`typical`' on instances with $m/2^{c\sqrt{k}}$ satisfiable clauses.*

The reduction to QP-Ratio is then as follows: Given a k-AND instance on n variables $X = \{x_1, x_2, \ldots x_n\}$ consisting of m clauses $C = \{C_1, C_2, \ldots C_m\}$, and a parameter $0 < \alpha < 1$, let $A = \{a_{ij}\}$ denote the $m \times n$ matrix such that a_{ij} is $1/m$ if variable x_j appears as is in clause C_i, a_{ij} is $-1/m$ if it appears negated and 0 otherwise.

Let $f : X \to \{-1, 0, 1\}$ and $g : C \to \{-1, 0, 1\}$ denote functions which are supposed to correspond to assignments. Let $\mu_f = \sum_{i \in [n]} |f(x_i)|/n$ and $\mu_g = \sum_{j \in m} |g(C_j)|/m$. Let

$$\vartheta(f, g) = \frac{\sum_{ij} a_{ij} f(x_i) g(C_j)}{\alpha \mu_f + \mu_g}. \tag{16}$$

Observe that if we treat $f(), g()$ as variables, we obtain an instance of QP-Ratio [as described, the denominator is *weighted*; we need to replicate the variable set X $\alpha\Delta$ times (each copy has same set of neighbors in C) in order to reduce to an unweighted instance – see Appendix C for details]. We pick $\alpha = 2^{-c\sqrt{k}}$ and Δ a large enough constant so that Conjecture 4.1 and Lemmas 4.3 and 4.4 hold. The completeness follows from the natural assignment .

Lemma 4.2 (Completeness)**.** *If the k-AND instance is such that an α fraction of the clauses can be satisfied, then there exists function f, g such that θ is at least $k/2$.*

Proof. Consider an assignment that satisfies an α fraction of the constraints. Let f be such that $f(x_i) = 1$ if x_i is true and -1 otherwise. Let g be the indicator of (the α fraction of the) constraints that are satisfied by the assignment. Since each such constraint contributes k to the sum in the numerator, the numerator is at least αk while the denominator 2α. ■

Soundness: We will show that for a typical random k-AND instance (i.e., with high probability), the matrix A is such that the maximum value $\vartheta(f, g)$ can take is at most $o(k)$.

Let the maximum value of ϑ obtained be ϑ_{max}. We first note that there exists a solution f, g of value $\vartheta_{max}/2$ such that the equality $\alpha\mu_f = \mu_g$ holds[2] – so we only need consider such assignments.

Now, the soundness argument is two-fold: if only a few of the vertices (X) are picked ($\mu_f < \frac{\alpha}{400}$) then the expansion of small sets guarantees that the value is small (even if each picked edge contributes 1). On the other hand, if many vertices (and hence clauses) are picked, then we claim that for every assignment to the variables (every f), only a small fraction ($2^{-\omega(\sqrt{k})}$) of the clauses contribute more than $k^{7/8}$ to the numerator.

The following lemma handles the first case when $\mu_f < \alpha/400$.

Lemma 4.3. *Let k be an integer, $0 < \delta < 1$, and Δ be large enough. If we choose a bipartite graph with vertex sets X, C of sizes $n, \Delta n$ respectively and degree k (on the C-side) uniformly at random, then w.h.p., for every $T \subset X, S \subset C$ with $|T| \le n\alpha/400$ and $|S| \le \alpha|T|$, we have $|E(S,T)| \le \sqrt{k}|S|$.*

Proof. Let $\mu := |T|/|X|$ (at most $\alpha/400$ by choice), and $m = \Delta n$. Fix a subset S of C of size $\alpha\mu m$ and a subset T of X of size μn. The expected number of edges between S and T in G is $\mathbf{E}[E(S,T)] = k\mu \cdot |S|$. Thus, by Chernoff-type bounds (we use only the *upper tail*, and we have negative correlation here),

$$\mathbf{P}[E(S,T) \ge \sqrt{k}|S|] \le \exp\big(-\frac{(\sqrt{k}|S|)^2}{k\mu \cdot |S|}\big) \le \exp\left(-\alpha m/10\right)$$

The number of such sets S, T is at most $2^n \times \sum_{i=1}^{\alpha^2 m/400} \binom{m}{i} \le 2^n 2^{H(\alpha^2/400)m} \le 2^{n+\alpha m/20}$. Union bounding and setting $m/n > 20/\alpha$ gives the result. ■

Now, we need to bound $\vartheta(f, g)$ for solutions such that $\alpha\mu_f = \mu_g \ge \alpha^2/400$. We use the following lemma about random instances of k-AND .

Lemma 4.4. *For large enough k and Δ, a random k-AND instance with Δn clauses on n variables is such that: for any assignment, at most a $2^{\frac{-k^{3/4}}{100}}$ fraction of the clauses have more than $k/2 + k^{7/8}$ variables 'satisfied' [i.e. the variable takes the value dictated by the AND clause] w.h.p.*

Proof. Fix an assignment to the variables X. For a single random clause C, the expected number of variables in the clause that are satisfied by the assignment is $k/2$. Thus, the probability that the assignment satisfies more than $k/2(1 + \delta)$ of the clauses is at most $\exp(-\delta^2 k/20)$. Further, each k-AND clause is chosen independently at random. Hence, by setting $\delta = k^{-1/8}$ and taking a union bound over all the 2^n assignments gives the result (we again use the fact that $m \gg n/\alpha$). ■

Lemma 4.4 shows that for every $\{\pm 1\}^n$ assignment to the variables x, at most $2^{-\omega(\sqrt{k})}$ fraction of the clauses contribute more than $2k^{7/8}$ to the numerator of $\vartheta(f, g)$. We can now finish the proof of the soundness part above.

[2]For instance, if $\alpha\mu_f > \mu_g$, we can always pick more constraints such that the numerator does not decrease (by setting $g(C_j) = \pm 1$ in a greedy way so as to not decrease the numerator) till $\mu_{g'} = \alpha\mu_f$, while losing a factor 2. Similarly for $\alpha\mu_f < \mu_g$, we pick more variables.

Proof of Soundness. Lemma 4.3 shows that when $\mu_f < \alpha/400$, $\vartheta(f,g) = O(\sqrt{k})$. For solutions such that $\mu_f > \alpha/400$, i.e., $\mu_g \geq \alpha^2/400 = 2^{-2\sqrt{k}}/400$, by Lemma 4.4 at most $2^{-\omega(\sqrt{k})}$ ($\ll \mu_g/k$) fraction of the constraints contribute more than $k^{7/8}$ to the numerator. Even if the contribution is k [the maximum possible] for this small fraction, the value $\vartheta(f,g) \leq O(k^{7/8})$. ∎

This shows a gap of k vs $k^{7/8}$ assuming Hypothesis 4.1. Since we can pick k to be arbitrarily large, we can conclude that QP-Ratio is hard to approximate to any constant factor.

4.3 Reductions from Ratio versions of CSPs

This section is inspired by the proof of [ABH+05], who show that Quadratic Programming (QP) is hard to approximate by giving a reduction from Label Cover to QP-Ratio.

Here we ask: is there a reduction from a *ratio version* of Label Cover to QP-Ratio? For this to be useful we must also ask: is the (appropriately defined) ratio version of Label Cover hard to approximate? The answer to the latter question is yes [see section A.3 for details and proof that Ratio-LabelCover is hard to approximate to any constant factor]. Unfortunately, we do not know how to reduce from Ratio-LabelCover. However, we present a reduction starting from a ratio version of *Unique Games*. We do not know if Ratio UG is hard to approximate for the parameters we need. While it seems related to Unique Games with Small-set Expansion [RS10], a key point to note is that we do not need 'near perfect' completeness, as in typical UG reductions.

We hope the Fourier analytic tools we use to analyze the ratio objective could find use in other PCP-based reductions to ratio problems. Let us now define a ratio version of Unique Games, and a useful *intermediate* QP-Ratio problem.

Definition 4.5 (Ratio UG). Consider a unique label cover instance $\mathcal{U}\big(G(V,E),[R],\{\pi_e|e \in E\}\big)$. The value of a partial labeling $L : V \to [R] \cup \{\perp\}$ (where label $\perp$ represents it is unassigned) is

$$val(L) = \frac{|\{(u,v) \in E | \pi_{u,v}(L(u)) = L(v)\}|}{|\{v \in V | L(v) \neq \perp\}|}$$

The (s,c)-Ratio UG problem is defined as follows: given $c > s > 0$ (to be thought of as constants), and an instance $\mathcal{U}$ on a regular graph G, distinguish between the two cases:

– **YES:** There is a partial labeling $L : V \to [R] \cup \{\perp\}$, such that $val(L) \geq c$.

– **NO:** For every partial labeling $L : V \to [R] \cup \{\perp\}$, $val(L) < s$.

The main result of this section is a reduction from (s,c)-Ratio UG to QP-ratio. We first introduce the following *intermediate* problem:

Definition 4.6. QP-Intermediate: Given $A_{(n\times n)}$ with $A_{ii} \leq 0$, maximize $\frac{x^TAx}{\sum_i |x_i|}$ s.t. $x_i \in [-1,1]$.

Note that A is allowed to have diagonal entries (albeit only negative ones) , and that the variables are allowed to take values in the *interval* $[-1,1]$.

Lemma 4.7. *Let A define an instance of QP-Intermediate with optimum value opt_1. There exists an instance B of QP-Ratio on $(n \cdot m)$ variables, with $m \leq \max\{\frac{2\|A\|_1}{\varepsilon}, 2n\} + 1$, and the property that its optimum value opt_2 satisfies $\mathsf{opt}_1 - \varepsilon \leq \mathsf{opt}_2 \leq \mathsf{opt}_1 + \varepsilon$. [Here $\|A\|_1 = \sum_{i,j} |a_{ij}|$.]*

Proof. The idea is to view each variable as an average of a large number (in this case, m) of new variables: thus a fractional value for x_i is 'simulated' by setting some of the new variables to ± 1 and the others zero. See Appendix A.2 for details. ∎

Thus from the point of view of approximability, it suffices to consider QP-Intermediate. We now give a reduction from Ratio UG to QP-Intermediate.

Input: An instance $\Upsilon = (V, E, \Pi)$ of Ratio UG , with alphabet $[R]$.
Output: A QP-Intermediate instance $\mathcal{Q}$ with number of variables $N = |V| \cdot 2^R$.
Parameters: $\eta := 10^6 n^7 2^{4R}$

Construction:

- For every vertex $u \in V$, we have 2^R variables, indexed by $x \in \{-1,1\}^R$. We will denote these by $f_u(x)$, and view f_u as a function on the hypercube $\{-1,1\}^R$.
- Fourier coefficients (denoted $\widehat{f_u}(S) = \mathbf{E}_x[\chi_S(x) f_u(x)]$) are linear forms in the variables $f_u(x)$.
- For $(u,v) \in E$, define $T_{uv} = \sum_i \widehat{f_u}(\{i\}) \widehat{f_v}(\{\pi_{uv}(i)\})$.
- For $u \in V$, define $L(u) = \sum_{S:|S|\neq 1} \widehat{f_u}(S)^2$.
- The instance of QP-Intermediate we consider is

$$\mathcal{Q} := \max \ \frac{\mathbf{E}_{(u,v)\in E}\, T_{uv} - \eta\, \mathbf{E}_u\, L(u)}{\mathbf{E}_u\, |f_u|_1},$$

where $|f_u|_1$ denotes $\mathbf{E}_x[|f_u(x)|]$.

Lemma 4.8. *(Completeness) If the value of Υ is $\geq \alpha$, then the reduction gives an instance of QP-Intermediate with optimum value $\geq \alpha$.*

Proof. Consider an assignment to Υ of value α and for each u set f_u to be the corresponding dictator (or $f_u = 0$ if u is assigned $\perp$). This gives a ratio at least α (the $L(u)$ terms contribute zero for each u). ∎

Lemma 4.9. *(Soundness) Suppose the QP-Intermediate instance obtained from a reduction (starting with Υ) has value τ, then there exists a solution to Υ of value $\geq \tau^2/C$, for an absolute constant C.*

Proof. Consider an optimal solution to the instance $\mathcal{Q}$ of QP-Intermediate, and suppose it has a value $\tau > 0$. Since the UG instance is regular, we have

$$val(\mathcal{Q}) = \frac{\sum_u \mathbf{E}_{v\in\Gamma(u)}\, T_{uv} - \eta \sum_u L(u)}{\sum_u \|f_u\|_1}. \tag{17}$$

First, we move to a solution such that the value is at least $\tau/2$, and for every u, $|f_u|_1$ is either zero, or is "not too small". The choice of η will then enable us to conclude that each f_u is 'almost linear' (there are no higher level Fourier coefficients).

Lemma 4.10. *There exists a solution to $\mathcal{Q}$ of value at least $\tau/2$ with the property that for every u, either $f_u = 0$ or $\|f_u\|_1 > \frac{\tau}{n2^{2R}}$.*

Proof. Let us start with the optimum solution to the instance. First, note that $\sum_u \|f_u\|_1 \geq 1/2^R$, because if not, $|f_u(x)| < 1$ for every u and $x \in \{-1,1\}^R$. Thus if we scale all the f_u's by a factor $z > 1$, the numerator increases by a z^2 factor, while the denominator only by z; this contradicts the optimality of the initial solution. Since the ratio is at least τ, we have that the numerator of (17) (denoted num) is at least $\tau/2^R$.

Now since $|\widehat{f_u}(S)| \le \|f_u\|_1$ for any S, we have that for all u, v, $T_{uv} \le R \cdot \|f_u\|_1 \|f_v\|_1$. Thus $\mathbf{E}_{v \in \Gamma(u)} T_{uv} \le R \cdot \|f_u\|_1$. Thus the contribution of u s.t. $\|f_u\|_1 < \tau/(n2^{2R})$ to num is at most $n \times R \cdot \frac{\tau}{n2^{2R}} < \frac{\tau}{2^{R+1}} < \mathsf{num}/2$. Now setting all such $f_u = 0$ will only decrease the denominator, and thus the ratio remains at least $\tau/2$. [We have ignored the $L(u)$ term because it is negative and only improves when we set $f_u = 0$.] ∎

For a boolean function f, we define the 'linear' and the 'non-linear' parts to be

$$f^{=1} := \sum_i \widehat{f}(i)\chi(\{i\}) \quad \text{and } f^{\neq 1} := f - f^{=1} = \sum_{|S| \neq 1} \widehat{f}(S)\chi(S).$$

Our choice of η will be such that:

1. For all u with $f_u \neq 0$, $\|f_u^{\neq 1}\|_2^2 \le \|f_u\|_1^2/10^6$. Using Lemma 4.10 (and the naïve bound $\tau \ge 1/n$), this will hold if $\eta > 10^6 n^7 2^{4R}$. [A simple fact used here is that $\sum_u \mathbf{E}[T_{uv}] \le nR$.]

2. For each u, $\|f_u^{\neq 1}\|_2^2 < \frac{1}{2^{2R}}$. This will hold if $\eta > n2^{2R}$ and will allow us to use Lemma A.1.

Also, since by Cauchy-Schwarz inequality, $|f_u|_2^2 \ge \delta_u^2$, we can conclude that 'most' of the Fourier weight of f_u is on the linear part for *every* u. We now show that the Cauchy-Schwarz inequality above must be tight up to a constant (again, for every u). The following is the key lemma in the reduction: it says that if a boolean function f is 'nearly linear', then it must also be *spread out* [which is formalized by saying $\|f\|_2 \approx \|f\|_1$]. This helps us deal with the main issue in a reduction with a ratio objective – showing we cannot have a large numerator along with a very small value of $\|f\|_1$ (the denominator). Morally, this is similar to a statement that a boolean function with a *small support* cannot have all its Fourier mass on the *linear* Fourier coefficients.

Lemma 4.11. *Let $f : \{-1,1\}^R \mapsto [-1,1]$ satisfy $\|f\|_1 = \delta$. Let $f^{=1}$ and $f^{\neq 1}$ be defined as above. Then if $\|f\|_2^2 > (10^4+1)\delta^2$, we have $\|f^{\neq 1}\|_2^2 \ge \delta^2$.*

Proof. Suppose that $\|f\|_2^2 > (10^4+1)\delta^2$, and for the sake of contradiction, that $\|f^{\neq 1}\|_2^2 < \delta^2$. Thus since $\|f\|_2^2 = \|f^{=1}\|_2^2 + \|f^{\neq 1}\|_2^2$, we have $\|f^{=1}\|^2 > (100\delta)^2$.

If we write $\alpha_i = \widehat{f}(\{i\})$, then $f^{=1}(x) = \sum_i \alpha_i x_i$, for every $x \in \{-1,1\}^R$. From the above, we have $\sum_i \alpha_i^2 > (100\delta)^2$. Now if $|\alpha_i| > 4\delta$ for some i, we have $\|f^{=1}\|_1 > (1/2) \cdot 4\delta$, because the value of $f^{=1}$ at one of $x, x \oplus e_i$ is at least 4δ for every x. Thus in this case we have $\|f^{=1}\|_1 > 2\delta$.

Now suppose $|\alpha_i| < 4\delta$ for all i, and so we can use Lemma A.2 to conclude that $\mathbf{P}_x(f^{=1}(x) > 100\delta/10) \ge 1/4$, which in turn implies that $|f^{=1}|_1 > (100\delta/10) \cdot \mathbf{P}_x(f^{=1}(x) > 100\delta/10) > 2\delta$.

Thus in either case we have $\|f^{=1}\|_1 > 2\delta$. This gives $\|f - f^{=1}\|_1 > \|f^{=1}\|_1 - \|f\|_1 > \delta$, and hence $\|f - f^{=1}\|_2^2 > \delta^2$ (Cauchy-Schwarz), which implies $\|f^{\neq 1}\|_2^2 > \delta^2$, which is what we wanted. ∎

Now, let us denote $\delta_u = |f_u|_1$. Since Υ is a unique game, we have for every edge (u,v) (by Cauchy-Schwarz),

$$T_{u,v} = \sum_i \widehat{f_u}(\{i\})\widehat{f_v}(\{\pi_{uv}(i)\}) \le \sqrt{\sum_i \widehat{f_u}(\{i\})^2}\sqrt{\sum_j \widehat{f_u}(\{j\})^2} \le |f_u|_2 |f_v|_2 \tag{18}$$

Now we can use Lemma 4.11 to conclude that in fact, $T_{u,v} \le 10^4 \delta_u \delta_v$. Now consider the following process: while there exists a u such that $\delta_u > 0$ and $\mathbf{E}_{v \in \Gamma(u)} \delta_v < \frac{\tau}{4 \cdot 10^4}$, set $f_u = 0$. We claim that this process only increases the objective value. Suppose u is such a vertex. From the bound on T_{uv}

above and the assumption on u, we have $\mathbf{E}_{v\in\Gamma(u)} T_{uv} < \delta_u \cdot \tau/4$. If we set $f_u = 0$, we remove at most twice this quantity from the numerator, because the UG instance is regular [again, the $L(u)$ term only acts in our favor]. Since the denominator reduces by δ_u, the ratio only improves (it is $\geq \tau/2$ to start with).

Thus the process above should terminate, and we must have a non-empty graph at the end. Let S be the set of vertices remaining. Now since the UG instance is regular, we have that $\sum_u \delta_u = \sum_u \mathbf{E}_{v\in\Gamma(u)} \delta_v$. The latter sum, by the above is at least $|S| \cdot \tau/(4 \cdot 10^4)$. Thus since the ratio is at least $\tau/2$, the numerator $\mathsf{num} \geq |S| \cdot \frac{\tau^2}{8 \cdot 10^4}$.

Now let us consider the following natural randomized rounding: for vertex $u \in S$, assign label i with probability $|\widehat{f}_u(\{i\})|/(\sum_i |\widehat{f}_u(\{i\})|)$. Now observing that $\sum_i |\widehat{f}_u(\{i\})| < 2$ for all u (Lemma A.1), we can obtain a solution to ratio-UG of value at least $\mathsf{num}/|S|$, which by the above is at least τ^2/C for a constant C.

This completes the proof of Lemma 4.9. ■

References

[AAM+11] Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein. Manuscript. 2011. 3

[ABH+05] Sanjeev Arora, Eli Berger, Elad Hazan, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 206–215, Washington, DC, USA, 2005. IEEE Computer Society. 16, 21

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45:501–555, May 1998. 22

[AN06] Noga Alon and Assaf Naor. Approximating the cut-norm via grothendieck's inequality. *SIAM J. Comput.*, 35:787–803, April 2006. 7

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of np. *J. ACM*, 45:70–122, January 1998. 22

[BCC+10] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an $o(n^{1/4})$ approximation for densest k-subgraph. In *STOC '10: Proceedings of the 42nd ACM symposium on Theory of computing*, pages 201–210, New York, NY, USA, 2010. ACM. 2, 14

[Cha00] Moses Charikar. Greedy approximation algorithms for finding dense components in a graph. In *APPROX '00: Proceedings of the Third International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 84–95, London, UK, 2000. Springer-Verlag. 4

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 205–214, New York, NY, USA, 2006. ACM. 5

[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*, pages 54–60, Washington, DC, USA, 2004. IEEE Computer Society. 2, 3, 5, 10, 11

[Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the 34th annual ACM Symposium on Theory of Computing (STOC'02)*, pages 534–543. ACM Press, 2002. 2, 3, 14

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. 2

[HK04] Jonas Holmerin and Subhash Khot. A new pcp outer verifier with applications to homogeneous linear equations and max-bisection. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, STOC '04, pages 11–20, New York, NY, USA, 2004. ACM. 23

[Kho04] Subhash Khot. Ruling out PTAS for graph min-bisection, densest subgraph and bipartite clique. In *Proceedings of the 44th Annual IEEE Symposium on the Foundations of Computer Science (FOCS'04)*, pages 136–145, 2004. 2

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.*, 37(1):319–357, 2007. 2

[KO06] Subhash Khot and Ryan O'Donnell. Sdp gaps and ugc-hardness for maxcutgain. In *Foundations of Computer Science, 2006. FOCS '06. 47th Annual IEEE Symposium on*, pages 217 –226, oct. 2006. 2

[KV05] Subhash Khot and Nisheeth K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *FOCS*, pages 53–62, 2005. 2

[Meg01] Alexandre Megretski. Relaxation of quadratic programs in operator theory and system analysis. In *In Systems, Approximation, Singular Integral Operators, and Related Topics*, pages 365–392, 2001. 2, 5

[NRT99] A. Nemirovski, C. Roos, and T. Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, 86:463–473, 1999. 10.1007/s101070050100. 2, 5

[O'D] Ryan O'Donnel. Analysis of boolean functions - lecture 21. In *http://www.cs.cmu.edu/ odonnell/boolean-analysis/*. 21

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC*, pages 245–254, 2008. 2

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998. 22

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *STOC '10: Proceedings of the 42nd ACM symposium on Theory of computing*, pages 755–764, New York, NY, USA, 2010. ACM. 16

[Tre09] Luca Trevisan. Max cut and the smallest eigenvalue. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 263–272, New York, NY, USA, 2009. ACM. 2, 3, 10, 11

[Tul09] Madhur Tulsiani. Csp gaps and reductions in the lasserre hierarchy. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 303–312, New York, NY, USA, 2009. ACM. 3

A Hardness of QP-Ratio

A.1 Boolean analysis

Lemma A.1. *[ABH+05] Let $f_u : \{-1,1\}^R \to [-1,1]$ be a solution to $\mathcal{Q}$ of value $\tau > 0$. Then*

$$\forall u \in V \qquad \sum_{i=1}^{R} |\widehat{f}_u(\{i\})| \leq 2.$$

Proof. Assume for sake of contradiction that $\sum_i \widehat{f}_u(\{i\}) > 2$.
Since $f_u^{=1}$ is a linear function with co-efficients $\{\widehat{f}_u(\{i\})\}$, there exists some $y \in \{-1,1\}^R$ such that $f_u^{=1}(y) = \sum_i |\widehat{f}_i(\{i\})| > 2$. For this y, we have $f^{\neq 1}(y) = f(y) - f^{=1}(y) < -1$.
Hence $|f^{\neq 1}|_2^2 > 2^{-R}$, which gives a negative value for the objective, for our choice of η. ∎

The following is the well-known Berry-Esséen theorem (which gives a quantitative version of central limit theorem). The version below is from [O'D].

Lemma A.2. *Let $\alpha_1, \dots, \alpha_R$ be real numbers satisfying $\sum_i \alpha_i^2 = 1$, and $\alpha_i^2 \leq \tau$ for all $i \in [R]$. Let X_i be i.i.d. Bernoulli (± 1) random variables. Then for all $\theta > 0$, we have*

$$\left| \mathbf{P}[\sum_i \alpha_i X_i > \theta] - N(\theta) \right| \leq \tau,$$

where $N(\theta)$ denotes the probability that $g > \theta$, for g drawn from the univariate Gaussian $\mathcal{N}(0,1)$.

A.2 Reducing QP-Intermediate to QP-Ratio

In this section we will prove Lemma 4.7. Let us start with a simple observation

Lemma A.3. *Let A be an $n \times n$ matrix (it could have arbitrary diagonal entries). Suppose $\{x_i\}$, $1 \leq i \leq n$ is the optimum solution to*

$$\max_{x_i \in [-1,1]} \frac{x^T A x}{\sum_i |x_i|}.$$

Now let $\delta < \min\{\frac{\varepsilon}{2\|A\|_1}, \frac{1}{2n}\}$ (where $\|A\|_1 = \sum_{i,j} |a_{ij}|$). Then perturbing each x_i additively by δ (arbitrarily) changes the value of the ratio by an additive factor of at most ε.

Proof. First note that $\sum_i |x_i| \geq 1$, because otherwise we can scale all the x_i by a factor $z > 1$ and obtain a feasible solution with a strictly better value [because the numerator scales by z^2 and the denominator only by z]. Thus changing each x_i by $\delta < 1/2n$ will keep the denominator between $1/2$ and $3/2$. Now consider the numerator: it is easy to see that a term $a_{ij}x_ix_j$ changes by at most $\delta|a_{ij}|$, and thus the numerator changes by at most $\delta\|A\|_1$. Thus the ratio changes by an additive factor at most $2\delta\|A\|_1 < \varepsilon$, by the choice of δ. ∎

Proof of Lemma 4.7. Start with an instance of QP-Intermediate given by $A_{(n\times n)}$, and suppose the optimum value is opt_1. Let m be an integer which will be chosen later [think of it as sufficiently large]. Consider the quadratic form B on $n \cdot m$ variables, defined by writing $x_i = \frac{1}{m} \cdot (y_i^{(1)} + y_i^{(2)} + \cdots + y_i^{(m)})$, and expanding out $x^T Ax$. Let C be a form on (the same) $n \cdot m$ variables obtained from B by omitting the square (diagonal) terms. Now consider the QP-Ratio instance given by C. That is,

$$\text{maximize } \frac{y^T Cy}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|}, \text{ subject to } y_i^{(j)} \in \{-1, 0, 1\}. \tag{19}$$

Let us write $a_{ii} = -\alpha_i$ (by assumption $\alpha_i \geq 0$). The we observe that

$$\frac{y^T Cy}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} = \frac{y^T By}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} + \frac{\sum_i \frac{\alpha_i}{m^2} \cdot \left(\sum_j |y_i^{(j)}|\right)}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} \tag{20}$$

By the assumption on α_i, we have

$$\frac{y^T By}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} \leq \frac{y^T Cy}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} \leq \frac{y^T By}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} + \sum_i \frac{\alpha_i}{m} \tag{21}$$

We prove the two inequalities separately. First let us start with an optimum solution $\{x_i\}$ to QP-Intermediate (from A) with value opt_1. As above, define $\delta = \min\{\frac{\varepsilon}{2\|A\|_1}, \frac{1}{2n}\}$. Let us round the values x_i to the nearest integer multiple of δ [for simplicity we will assume also that $1/\delta$ is an integer]. By Lemma A.3, this will change the objective value by at most ε. We will choose m to be an integer multiple of $1/\delta$, thus if we set $y_i^{(j)} = sign(x_i)$ for $j = 1, 2, \ldots, mx_i$ and 0 for the rest, we obtain a value at least $\mathsf{opt}_1 - \varepsilon$ for the QP-Ratio problem defined by C [using the first half of (21)]. Thus $\mathsf{opt}_2 \geq \mathsf{opt}_1 - \varepsilon$.

Now consider a solution to the QP-Ratio problem defined by C, and set $x_i = \frac{1}{m} \cdot (y_i^{(1)} + \ldots y_i^{(m)})$. For this assignment, we have

$$\frac{x^T Ax}{\sum_i |x_i|} = \frac{y^T By}{\sum_i \frac{1}{m}(\sum_j |y_i^{(j)}|)} \geq \frac{y^T By}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} \geq \frac{y^T Cy}{\frac{1}{m}\sum_{i,j}|y_i^{(j)}|} - \varepsilon,$$

because we will choose $m \geq \frac{\|A\|_1}{\varepsilon}$. This implies that $\mathsf{opt}_1 \geq \mathsf{opt}_2 - \varepsilon$.

Thus we need to choose m to be the smallest integer larger than $\max\{\frac{2\|A\|_1}{\varepsilon}, 2n\}$ for all the bounds to hold. This gives the claimed bound on the size of the instance. ■

A.3 Towards NP-hardness – LabelCover with SSE

The PCP theorem, [AS98, ALM+98] combined with the parallel repetition theorem [Raz98] yields the following theorem.

Theorem A.4 (Label Cover hardness)**.** *There exists a constant $\gamma > 0$ so that any 3-SAT instance w and any $R > 0$, one can construct a Label Cover instance $\mathcal{L}$, with $|w|^{O(\log R)}$ vertices, and label set of size R, so that: if w is satisfiable, $\mathsf{val}(\mathcal{L}) = 1$ and otherwise $\mathsf{val}(\mathcal{L}) = \tau < R^{-\gamma}$. Further, $\mathcal{L}$ can be constructed in time polynomial in its size.*

Definition A.5 (((μ, η)-Expanding Label Cover)**.** An instance of the Label Cover problem is said to be (μ, η)-expanding if for every $\mu' < \mu$, and functions $f : \mathcal{A} \to [0, 1]$ and $g : \mathcal{B} \to [0, 1]$ such that $E_{a\in\mathcal{A}}[f(a)] = E_{b\in\mathcal{B}}[g(b)] = \mu'$,

$$E_{(a,b)\in E}[f(a)\ g(b)] \leq \mu'\eta.$$

Theorem A.6. *For every $\delta > 0$ and $\eta > \delta^{1/3}$, one can convert a Label Cover instance $\mathcal{L}$ into an instance (δ, η) Expanding Label Cover $\mathcal{L}'$ (with the same completeness vs soundness). Further, the size of $\mathcal{L}'$ is at most $|\mathcal{L}|^{1/\delta}$ and has a label set of size at most R/δ.*

Proof. We use one of product instances used in [HK04] (which gives Mixing but not Smoothness - Appendix A.2). They can argue about the expansion of only sets that are sufficiently large (constant fraction), while we need to work with all set sizes $< \delta n$.

Given an instance of Label Cover represented as $\Upsilon = (\mathcal{A} \cup \mathcal{B}, E, \Pi, [R])$, let $\Upsilon^k = (\mathcal{A}^k \cup \mathcal{B}, E^k, \Pi, [R])$, where $(1/\eta)^3 < k < 1/\delta$. Let $g : \mathcal{B} \to [0,1]$ be any function defined on the right hand side. For each $a \in \mathcal{A}$, let X_a denote the average value of g over the neighborhood of a. Since Υ is right-regular, we have $\mathbf{E}_{a\in\mathcal{A}}[X_a] = \mathbf{E}_{b\in\mathcal{B}}\, g(b)$.

For a vertex $a = (a_1, a_2, \ldots a_k) \in \mathcal{A}^k$, define Y_a to be the average value of g over the neighborhood of a (counting multiple edges multiple times).

$$\begin{aligned}
\mathop{\mathbf{E}}_{a\in\mathcal{A}^k}[Y_a^2] &= \mathop{\mathbf{E}}_{a_1,a_2,\ldots,a_k\in\mathcal{A}}\left[\left(\frac{\sum_i X_{a_i}}{k}\right)^2\right] \\
&= \frac{1}{k}\mathop{\mathbf{E}}_{a\in\mathcal{A}}[X_{a_i}^2] + \frac{k^2-k}{k^2}\mathop{\mathbf{E}}_{a,a'\in\mathcal{A}}[X_a X_{a'}] \\
&\leq \frac{1}{k}\delta + \delta^2 \leq \frac{2\delta}{k} && \text{(for } \delta < {}^1\!/_k)
\end{aligned}$$

Thus, by a second moment bound, the fraction of vertices in $\mathcal{A}^k$ with Y_a greater than η is at most ${}^{2\delta}\!/_{k\eta^2}$. Thus, for any $f : \mathcal{A}^k \to [0,1]$, such that $\mathbf{E}[f] = \delta$,

$$\begin{aligned}
\mathop{\mathbf{E}}_{(a,b)\in E}[f(a)g(b)] &\leq \frac{2\delta}{k\eta^2} + \eta\left(\delta - \frac{2\delta}{k\eta^2}\right) \\
&\leq 3\delta\eta \qquad \text{since } (1/\eta)^3 < k
\end{aligned}$$

■

Ratio Label Cover. Consider the ratio version of Label Cover – the goal is to find a partial assignment to a Label Cover instance, which maximizes the fraction of edges satisfied, divided by the fraction of vertices which have been assigned labels (an edge (u,v) is satisfied iff both the end points are assigned labels which satisfy the constaint $\pi_{u,v}$).

It follows by a fairly simple argument that Theorem A.6 shows that Ratio Label Cover is NP-hard to approximate within any constant factor. We sketch an argument that shows NP-hardness of 1 vs γ for any constant $\gamma > 0$. We start with Label Cover instance Υ with completeness 1 and soundess $\tau < \gamma^{4/3}$ (and appropriate label size). Our Ratio Label Cover instance is essentially obtained by applying Theorem A.6 : the instance has soundness τ and is $(\gamma^{1/3}, \gamma)$ expanding. The completeness of this instance is easily seen to be 1.

To argue the soundess, it first suffices to only consider solutions f, g which have equal measure (similar to the argument in section 4.2 – the instance first has its right side duplicated so that right and left sizes are equal, and then upto a factor 2 loss, we only pick equal number of vertices on both sides). The expansion of sets of measure $< \delta$ have value at most γ due to the expansion of the instance. Suppose there is a solution of measure $\geq \delta$ which has ratio label cover value γ, then we obtain a solution to the Label Cover instance Υ of value at least $\gamma \cdot \delta = \gamma^{4/3} > \tau$, which is a contradiction.

B APX-hardness of QP-Ratio

We proved that QP-Ratio is hard to approximate to an $O(1)$ factor assuming the small-set expansion conjecture. Here we prove a weaker hardness result – that there is no PTAS – assuming just $\mathbf{P} \neq \mathbf{NP}$. We do not go into the full details, but the idea is the following.

We reduce Max-Cut to an instance of QP-Ratio. The following is well-known (we can also start with other QP problems instead of Max-Cut)

> There exist constants $1/2 < \rho' < \rho$ such that: given a graph $G = (V, E)$ which is regular with degree d, it is NP-hard to distinguish between
> YES. $\text{MaxCut}(G) \geq \rho \cdot nd/2$, and
> NO. $\text{MaxCut}(G) \leq \rho' \cdot nd/2$.

Given an instance $G = (V, E)$ of Max-Cut, we construct an instance of QP-Ratio which has V along with some other vertices, and such that in an OPT solution to this QP-Ratio instance, *all* vertices of V would be picked (and thus we can argue about how the best solution looks).

First, let us consider a simple instance: let $abcde$ be a 5-cycle, with a cost of $+1$ for edges ab, bc, cd, de and -1 for the edge ae. Now consider a QP-Ratio instance defined on this graph (with ± 1 weights). It is easy to check that the best ratio is obtained when precisely four of the vertices are given non-zero values, and then we can get a numerator cost of 3, thus the optimal ratio is 3/4.

Now consider n cycles, $a_i b_i c_i d_i e_i$, with weights as before, but scaled up by d. Let A denote the vertex set $\{a_i\}$ (similarly $B, C, ..$). Place a clique on the set of vertices A, with each edge having a cost $10d/n$. Similarly, place a clique of the same weight on E. Now let us place a copy of the graph G on the set of vertices C.

It turns out (it is actually easy to work out) that there is an optimal solution with the following structure: (a) all a_i are set to 1, (b) all e_i are set to -1 (this gives good values for the cliques, and good value for the $a_i b_i$ edge), (c) c_i are set to ± 1 depending on the structure of G, (d) If c_i were set to $+1$, $b_i = +1$, and $d_i = 0$; else $b_i = 0$ and $d_i = -1$ (Note that this is precisely where the 5-cycle with one negative sign helps!)

Let $x_1, ..., x_n \in \{-1, 1\}$ be the optimal assignment to the Max-Cut problem. Then as above, we would set $c_i = x_i$. Let the cost of the MaxCut solution be $\theta \cdot nd/2$. Then we set $4n$ of the $5n$ variables to ± 1, and the numerator is (up to lower order terms):

$$2 \cdot (10d/n) n^2/2 + \theta \cdot nd/2 + 3nd = (\Delta + \theta) nd,$$

where Δ is an absolute constant.

We skip the proof that there is an optimal solution with the above structure. Thus we have that it is hard to distinguish between a case with ratio $(\Delta + \rho')d/4$, and $(\Delta + \rho)d/4$, which gives a small constant factor hardness.

C Reduction from a weighted to an Unweighted version

Lemma C.1. *Let A be an $n \times m$ matrix, and let $w \geq 1$ be an integer. Let opt_1 denote the optimum value of the problem*

$$\max_{x \in \{-1,0,1\}^n, y \in \{-1,0,1\}^m} \frac{x^T A y}{w\|x\|_1 + \|y\|_1}.$$

Let B be a $wn \times m$ matrix formed by placing w copies of A one below the other. [In terms of bipartite graphs, this just amounts to making w copies of the left set of vertices.] Let opt_2 denote the optimum value of the problem

$$\max_{z \in \{-1,0,1\}^{wn}, y \in \{-1,0,1\}^m} \frac{z^T B y}{\|z\|_1 + \|y\|_1}.$$

Then $opt_1 = opt_2$.

Proof. It is clear that $opt_2 \geq opt_1$: simply take a solution of value opt_1 for the first problem and form z by taking w copies of x. To see the other direction, let us view z as being formed of 'chunks' $z_1, \ldots, z_w$ of size n each. Consider a solution to the second problem of value opt_2. Then

$$opt_2 = \frac{z_1^T Ay + z_2^T Ay + \ldots z_w^T Ay}{\|z_1\|_1 + \ldots \|z_w\|_1 + |y|},$$

which implies that if we set $x = z_i$ which gives the largest value of $z_i^T Ay/\|z_i\|_1$, we obtain a value at least opt_2 to the first problem.

This completes the proof. ■