Improved Approximation Algorithms for MAX k-CUT and MAX BISECTION

Alan Frieze*
Carnegie Mellon University

Mark Jerrum†
University of Edinburgh

June 1994; revised June 1995

Abstract

Polynomial-time approximation algorithms with non-trivial performance guarantees are presented for the problems of (a) partitioning the vertices of a weighted graph into k blocks so as to maximise the weight of crossing edges, and (b) partitioning the vertices of a weighted graph into two blocks of equal cardinality, again so as to maximise the weight of crossing edges. The approach, pioneered by Goemans and Williamson, is via a semidefinite programming relaxation.

1 Introduction

Goemans and Williamson [8] have significantly advanced the theory of approximation algorithms. Previous work on approximation algorithms was largely dependent on comparing heuristic solution values to that of a Linear Program (LP) relaxation, either implicitly or explicitly. This was recognised

*Department of Mathematics, Carnegie Mellon University, Pittsburgh PA15213, U.S.A.
The first author was supported in part by NSF grant CCR-9225008.

†Department of Computer Science, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, United Kingdom. The work described here was undertaken while the second author was visiting Carnegie Mellon University; at that time, he was a Nuffield Science Research Fellow, and was supported in part by grant GR/F 90363 of the UK Science and Engineering Research Council, and Esprit Working Group 7097 "RAND."

some time ago by Wolsey [16]. (One significant exception to this general rule has been the case of Bin Packing.)

The main novelty of [8] is the use of a SemiDefinite Program (SDP) as a relaxation. To be more precise let us consider the problem MAX CUT studied (among others) in [8]: we are given a vertex set $V = \{1, \ldots, n\}$ and non-negative weights w_{ij}, for $1 \le i, j \le n$, where $w_{ij} = w_{ji}$ and $w_{ii} = 0$ for all i, j. If $S \subseteq V$ and $\overline{S} = V \setminus S$ then the *weight* of the *cut* $(S : \overline{S}\,)$ is

$$w(S : \overline{S}\,) = \sum_{i \in S, j \in \overline{S}} w_{ij}.$$

The aim is to find a cut of maximum weight.

Introducing integer variables $y_j \in \{-1, 1\}$ for $j \in V$ we can formulate the MAX CUT problem as

$$\begin{array}{rll} \text{IP:} & \text{maximise} & \frac{1}{2}\sum_{i<j} w_{ij}(1 - y_i y_j) \\ & \text{subject to} & y_j \in \{-1, 1\}, \quad \forall j \in V. \end{array}$$

The key insight of Goemans and Williamson is that instead of converting this to an integer linear program and then considering the LP relaxation, it is possible to relax IP directly to the following:

$$\begin{array}{rll} \text{SDP:} & \text{maximise} & \frac{1}{2}\sum_{i<j} w_{ij}(1 - v_i \cdot v_j) \\ & \text{subject to} & v_j \in S_{n-1}, \quad \forall j \in V. \end{array}$$

Here $S_{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ is the unit sphere in n dimensions. SDP is an example of special kind of convex program, called a *semidefinite program* for reasons that will become apparent presently, which is efficiently solvable in both theoretical and practical senses. In particular, an optimal solution within given additive error ε may be computed in time polynomial in n and $\log \varepsilon^{-1}$. (See Alizadeh [1] for a detailed exposition.) More explicitly, the optimisation problem SDP is equivalent to:

$$\begin{array}{rll} \text{CP:} & \text{maximise} & \frac{1}{2}\sum_{i<j} w_{ij}(1 - Y_{ij}) \\ & \text{subject to} & Y_{jj} = 1, \quad \forall j \in V \\ & & Y = [Y_{ij}] \succ 0. \end{array} \tag{1}$$

Here Y_{ij} replaces $v_i \cdot v_j$, and the notation $Y \succ 0$ indicates that the matrix Y is constrained to be positive semidefinite; this constraint defines a convex subset of $\mathbb{R}^{n^2}$. The idea of Goemans and Williamson is to solve SDP and then use the following simple (randomised rounding) heuristic to obtain a remarkably good solution to MAX CUT: choose a random hyperplane through the origin, and partition the vectors v_i (and hence the vertex set V) according to which side of the hyperplane they fall.

This is an exciting new idea and it is important to see in what directions it can be generalised. In this paper we do so in two ways. First we consider MAX k-CUT where the aim is to partition V into k subsets: for a partition $\mathcal{P} = P_1, P_2 \ldots, P_\ell$ of V we let $|\mathcal{P}| = \ell$ and

$$w(\mathcal{P}) = \sum_{1 \le r < s \le \ell} \sum_{i \in P_r, j \in P_s} w_{ij}.$$

The problem is then

$$\text{MAX } k\text{-CUT:} \quad \begin{array}{ll} \text{maximise} & w(\mathcal{P}) \\ \text{subject to} & |\mathcal{P}| = k. \end{array}$$

Note that MAX k-CUT may be interpreted as the search for a ground state in the anti-ferromagnetic k-state Potts model: see Welsh [15].

Papadimitriou and Yannakakis [11] studied an unweighted ($w_{ij} \in \{0, 1\}$) version of MAX k-CUT in the guise of "MAX k-COLORABLE SUBGRAPH," and showed it to be MAX SNP-complete. In the light of Arora, Lund, Motwani, Sudan and Szegedy's characterisation of the class NP in terms of probabilistically checkable proofs [3], this result implies that there can be no polynomial-time approximation scheme for MAX k-CUT, for any $k \ge 2$, unless P = NP. The question, then, is how closely may MAX k-CUT be approximated in polynomial time?

To attack this problem we need to be able to handle variables that can take on one of k values as opposed to just two, a similar problem to that faced by Karger, Motwani and Sudan in trying to colour 3-colourable graphs with relatively few colours [10]. Our solution (and theirs) is a natural extension of the existing solution for the case $k = 2$, but the performance analysis presents greater technical difficulties.

The simplest heuristic for MAX k-CUT is just to randomly partition V into k sets. If $\widehat{\mathcal{P}}$ denotes the (random) partition produced and $\mathcal{P}^*$ denotes the optimum partition then it is easy to see that

$$\mathbf{E}(w(\widehat{\mathcal{P}})) \ge \left(1 - \frac{1}{k}\right) w(\mathcal{P}^*),$$

since each edge (i,j) has probability $(1-k^{-1})$ of joining vertices in different sets of the partition.

We describe a (randomised) heuristic "k-CUT" that produces a k-partition, say $\mathcal{P}_k$, which is provably better on average than the one produced by oblivious random partitioning. This heuristic is a natural extension of that of Goemans and Williamson, and is similar to the one discovered earlier and independently by Karger, Motwani and Sudan for the related problem of finding "semicolorings" of a graph. We prove the existence of a sequence of constants α_k, for $k \geq 2$, such that if $\mathcal{P}_k^*$ denotes the optimal partition in MAX k-CUT then:

Theorem 1 $\mathbf{E}(w(\mathcal{P}_k)) \geq \alpha_k\, w(\mathcal{P}_k^*)$, *where the constants* α_k *satisfy*

(i) $\alpha_k > 1 - k^{-1}$;

(ii) $\alpha_k - (1-k^{-1}) \sim 2k^{-2} \ln k$;[1]

(iii) $\alpha_2 \geq 0{\cdot}878567$, $\alpha_3 \geq 0{\cdot}800217$, $\alpha_4 \geq 0{\cdot}850304$, $\alpha_5 \geq 0{\cdot}874243$, $\alpha_{10} \geq 0{\cdot}926642$, *and* $\alpha_{100} \geq 0{\cdot}990625$.

The performance ratio for $k = 2$ in the above theorem is the same as that quoted by Goemans and Williamson, as in this special case we are able to carry across their analysis unchanged.

It will be seen that we achieve an improvement over the random partitioning heuristic for all k, and this is the main contribution of the article. However, it must be admitted that the improvement for large k is rather small. Kann, Khanna, Lagergren and Panconesi show, by presenting a more refined approximation-preserving reduction that the one employed by Papadimitriou and Yannakakis, that there can be no polynomial-time approximation algorithm for MAX k-CUT with performance ratio $1-1/239k$, unless P = NP [9]. (Note, however, that polynomial-time approximation schemes are known for the case of dense graphs: see de la Vega [6] or Arora, Karger and Karpinski [2].) This leaves open the possibility of an algorithm with performance ratio bounded below by $1-\alpha k^{-1}$, uniformly over k, for some $1/239 < \alpha < 1$. We say something about the theoretical limitations of our chosen semidefinite program relaxation in the following section.

[1]Throughout this article, the relation $\sim$ indicates two expressions whose ratio tends to 1 as $k \to \infty$.

Our second result concerns the problem MAX BISECTION. Here we have to partition V into two subsets of equal size (assuming that n is even) so as to maximise w.

$$\begin{array}{lll} \text{MAX BISECTION:} & \text{maximise} & w(\mathcal{P}) \\ & \text{subject to} & \mathcal{P} = S, V \setminus S \\ & & |S| = n/2. \end{array}$$

A problem of this type might arise as follows. Postulate a set of m people each of whom select a pair of "activities" from a set of n activities. Assume that n is even. We are required to split the activities evenly between two timetable slots so as to maximise the number of people who are able participate in both their chosen activities.

As with MAX CUT, a random bisection produces an expected perfomance ratio of $\frac{1}{2}$. Let ε be a small positive constant. We describe a heuristic "BISECT" which produces a partition $\mathcal{P}_B$, such that if $\mathcal{P}_B^*$ denotes the optimal bisection:

Theorem 2 $\mathbf{E}(w(\mathcal{P}_B)) \geq \beta\, w(\mathcal{P}_B^*)$, *where* $\beta = 2(\sqrt{2\alpha_2} - 1) - \varepsilon$.

Note that $\alpha_2 = 0{\cdot}878567\ldots$, as in Theorem 1, and $\beta > 0{\cdot}651$ for ε sufficiently small. The difficulty with generalising Goemans and Williamson's heuristic to MAX BISECTION is that it does not generally give a bisection of V. We prove that a simple modification of their basic algorithm is adequate to achieve the improved performance ratio claimed in Theorem 2.

Note that there is a natural generalisation of this problem to MAX k-SECTION, where we seek to partition V into k equal pieces. Unfortunately we cannot prove that the natural generalisation of our bisection heuristic beats the $1 - k^{-1}$ lower bound of the simple random selection heuristic when $k \geq 3$.

2 MAX k-CUT

In this section we describe our heuristic "k-CUT." We first describe a suitable way of modelling variables which can take one of k values. Just allowing $y_j = 1, 2, \ldots, k$ does not easily yield a useful integer program. Instead we allow y_j to be one of k *vectors* $a_1, a_2, \ldots, a_k$ defined as follows: take an equilateral simplex Σ_k in $\mathbb{R}^{k-1}$ with vertices $b_1, b_2, \ldots, b_k$. Let $c_k = (b_1 + b_2 + \cdots + b_k)/k$ be the centroid of Σ_k and let $a_i = b_i - c_k$, for $1 \leq i \leq k$. Finally assume that Σ_k is scaled so that $|a_i| = 1$ for $1 \leq i \leq k$.

Lemma 3

$$a_i \cdot a_j = -1/(k-1), \quad \textit{for } i \neq j. \tag{2}$$

Proof Since $a_1, a_2, \ldots, a_k$ are of unit length we have to show that the angle between a_i and a_j is $\arccos(-1/(k-1))$ for $i \neq j$. Let $b_1, b_2, \ldots, b_{k-1}$ lie in the plane $x_{k-1} = 0$ and form an equilateral simplex of dimension $k-2$. Let $b_i = (b'_i, 0)$ for $1 \leq i \leq k-1$, where b'_i has dimension $k-2$, and assume $b'_1 + b'_2 + \cdots + b'_{k-1} = 0$. Then $c_k = (0, 0, \ldots, 0, x)$ and $b_k = (0, 0, \ldots, 0, kx)$ for some $x > 0$. But $|b_k - c_k| = 1$ and so $x = 1/(k-1)$. But then $(b_k - c_k) \cdot (b_1 - c_k) = -(k-1)x^2 = -1/(k-1)$. $\square$

Note that $-1/(k-1)$ is the best angle separation we can obtain for k vectors as we see from:

Lemma 4 *If $u_1, u_2, \ldots, u_k$ satisfy $|u_i| = 1$ for $1 \leq i \leq k$, and $u_i \cdot u_j \leq \gamma$ for $i \neq j$, then $\gamma \geq -1/(k-1)$.*

Proof $0 \leq (u_1 + u_2 + \cdots + u_k)^2 \leq k + k(k-1)\gamma$. $\square$

Given Lemma 3 we can formulate MAX k-CUT as follows:

$$\begin{array}{lll} \mathrm{IP}_k: & \text{maximise} & \frac{k-1}{k} \sum_{i<j} w_{ij}(1 - y_i \cdot y_j) \\ & \text{subject to} & y_j \in \{a_1, a_2, \ldots, a_k\}. \end{array}$$

Here we use the fact that

$$1 - y_i \cdot y_j = \begin{cases} 0, & \text{if } y_i = y_j \\ k/(k-1), & \text{if } y_i \neq y_j \end{cases}$$

To obtain our SDP relaxation we replace y_i by v_i, where v_i can now be any vector in S_{n-1}. There is a problem in that we can have $v_i \cdot v_j = -1$ whereas $y_i \cdot y_j \geq -1/(k-1)$. We need therefore to add the constraint $v_i \cdot v_j \geq -1/(k-1)$. We obtain

$$\begin{array}{lll} \mathrm{SDP}_k: & \text{maximise} & \frac{k-1}{k} \sum_{i<j} w_{ij}(1 - v_i \cdot v_j) \\ & \text{subject to} & v_j \in S_{n-1}, \quad \forall j \\ & & v_i \cdot v_j \geq -1/(k-1), \quad \forall i \neq j \end{array} \tag{3}$$

Note that (3) reduces to the linear constraint $Y_{ij} \geq -1/(k-1)$ if we go to the convex programming form CP. We can now describe our heuristic:

k**-CUT**

Step 1. Solve the problem SDP_k to obtain vectors $v_1, v_2, \ldots, v_n \in S_{n-1}$.

Step 2. Choose k *random* vectors $z_1, z_2, \ldots, z_k$.

Step 3. Partition V according to which of $z_1, z_2, \ldots, z_k$ is closest to each v_j, i.e., let $\mathcal{P} = P_1, P_2, \ldots, P_k$ be defined by

$$P_i = \{j : v_j \cdot z_i \geq v_j \cdot z_{i'} \text{ for all } i' \neq i\}, \quad \text{for } 1 \leq i \leq k.$$

(Break ties for the minimum arbitrarily: they occur with probability zero!)

The most natural way of choosing $z_1, z_2, \ldots, z_k$ is to choose them independently at random from S_{n-1}. Forcing $|z_i| = 1$ complicates the analysis marginally and so we let $z_j = (z_{1,j}, z_{2,j}, \ldots, z_{n,j})$, $1 \leq j \leq k$ where the $z_{i,j}$ are kn independent samples from a (standard) normal distribution with mean 0 and variance 1. When $k = 2$ we have (modulo the normalisation $|z_i| = 1$) the heuristic of Goemans and Williamson, although they define it in terms of cutting S_{n-1} by a random hyperplane through the origin. Karger et al. [10] also use the above partitioning heuristic in their approach to approximate colouring, though applied to a slightly different semidefinite relaxation. However, we must diverge at this point, as their analysis — though adequate for the colouring application — is not sharp enough to yield Theorem 1.

Let W_k denote the weight of the partition produced by the heuristic, let W_k^* be the weight of the optimal partition and let $\widetilde{W}_k$ denote the maximum value of SDP_k. Putting $y_j = a_i$, for $j \in P_i$ and $1 \leq i \leq k$, we see that

$$\mathbf{E}(W_k) = \sum_{i<j} w_{ij} \Pr(y_i \neq y_j). \tag{4}$$

Now by symmetry $\Pr(y_i \neq y_j)$ depends only on the angle θ between v_i and v_j, and hence on $\varrho = \cos\theta = v_i \cdot v_j$. Let this separation probability be denoted by $\Phi_k(\varrho)$. It then follows from (4) that

$$\begin{aligned} \frac{\mathbf{E}(W_k)}{W_k^*} &\geq \frac{\mathbf{E}(W_k)}{\widetilde{W}_k} \\ &= \frac{\sum_{i<j} w_{ij}\, \Phi_k(v_i \cdot v_j)}{\frac{k-1}{k} \sum_{i<j} w_{ij}(1 - v_i \cdot v_j)} \\ &\geq \alpha_k, \end{aligned}$$

where

$$\alpha_k \; = \; \min_{-1/(k-1)\le\varrho\le 1} \frac{k\,\Phi_k(\varrho)}{(k-1)(1-\varrho)}.$$

The main work lies in bounding the quantity α_k. As Goemans and Williamson showed, the computation can be done exactly in the case $k = 2$, save for a final step involving the optimisation of a simple trigonometric function. When $k > 2$, it appears that we must work much harder, and the remainder of the section is devoted to obtaining lower bounds in this case. The results are summarised in Corollaries 6, 7, and 8, which, taken together, establish Theorem 1. First, some definitions and a technical lemma.

Let u, v be vectors, and $r_1, \ldots, r_k$ be a sequence of vectors, all in $\mathbb{R}^n$. We say that u and v are *separated by* $r_1, \ldots, r_k$ if the vector r_i maximising $u\cdot r_i$ is distinct from the vector r_j maximising $v\cdot r_j$. When we speak of a random vector, we mean a vector $r = (\xi_1, \ldots, \xi_n)$ whose coordinates ξ_i are independent, normally distributed random variables with mean 0 and variance 1. Note that the probability density function of r is $(2\pi)^{-n/2}\exp(-(\xi_1^2+\cdots+\xi_n^2)/2)$, and in particular is spherically symmetric.

Denote by $g(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ the probability density function of the univariate normal distribution, and by $G(x) = \int_{-\infty}^{x} g(\xi)\,d\xi$ the corresponding cumulative distribution function. For $i = 1, 2, \ldots$, the normalised *Hermite polynomials* $\phi_i(\,\cdot\,)$ are defined by

$$(-1)^i\sqrt{i!}\,\phi_i(x)g(x) = \frac{d^i g(x)}{dx^i}. \tag{5}$$

Let $h_i = h_i(k)$ denote the expectation of $\phi_i(x_{\max})$, where $x_{\max}$ is distributed as the maximum of a sequence of k independent normally distributed random variables.

Lemma 5 *Suppose $u, v \in \mathbb{R}^n$ are unit vectors at angle θ, and $r_1, \ldots, r_k$ is a sequence of random vectors. Let $\varrho = \cos\theta = u\cdot v$, and denote by $N_k(\varrho) = 1-\Phi_k(\varrho)$ the probability that u and v are* not *separated by $r_1, \ldots, r_k$. Then the Taylor series expansion*

$$N_k(\varrho) = a_0 + a_1\varrho + a_2\varrho^2 + a_3\varrho^3 + \cdots$$

of $N_k(\varrho)$ about the point $\varrho = 0$ converges for all ϱ in the range $|\varrho| \le 1$. The coefficients a_i of the expansion are all non-negative, and their sum converges to $N_k(1) = 1$. The first three coefficients are $a_0 = 1/k$, $a_1 = h_1^2/(k-1)$ and $a_2 = kh_2^2/(k-1)(k-2)$.

The main work lies in the proof of the above lemma, which we defer to the end of the section in order to press on with establishing the claims about α_k made in Theorem 1.

Corollary 6 *$\alpha_k > 1 - k^{-1}$, for all $k \geq 2$.*

Proof Denote by $A_k(\varrho)$ the function

$$A_k(\varrho) = \frac{k(1 - N_k(\varrho))}{(k-1)(1-\varrho)},$$

so that the performance ratio of the k-CUT heuristic can be expressed as

$$\alpha_k = \min_{-1/(k-1) \leq \varrho < 1} A_k(\varrho).$$

At $\varrho = 0$, the numerator and denominator of $A_k(\varrho)$ are both $k-1$; at $\varrho = 1$ they are both 0. Since the power series expansion of $N_k(\varrho)$ has only positive terms, the numerator is a concave function in the range $0 \leq \varrho \leq 1$, and hence $A_k(\varrho) \geq 1$ in that range.

Turning to the case $\varrho < 0$, note that $N_k(1) = 1$ and $N_k(-1) = 0$ implies $\sum_{i \text{ even}} a_i = \frac{1}{2}$; furthermore, since $h_1(k)$ increases with k and $h_1(3) = 3/2\sqrt{\pi}$ (using calculations described by David in [5, Section 3.2]), we have $a_1 \geq 9/4\pi(k-1)$. Therefore,

$$N_k(\varrho) \leq \frac{1}{k} - \frac{9(-\varrho)}{4\pi(k-1)} + \frac{\varrho^2}{2} \leq \frac{1}{k} - \frac{(-\varrho)}{5(k-1)},$$

where the second inequality is valid over the range $-1/(k-1) \leq \varrho \leq 0$, since $9/4\pi - 1/2 \geq 1/5$, and hence

$$A_k(\varrho) \geq \frac{1}{1-\varrho}\left(1 + \frac{k(-\varrho)}{5(k-1)^2}\right).$$

It is easily verified that the above expression is strictly greater than $1 - k^{-1}$ over the closed interval $-1/(k-1) \leq \varrho \leq 0$. $\square$

Corollary 7 *$\alpha_k - (1 - k^{-1}) \sim 2k^{-2} \ln k$.*

Proof Galambos [7, Section 2.3.2], gives the asymptotic distribution of the maximum of k independent, normally distributed random variables. In particular the quantity $h_1(k)$, which is just the expectation of the maximum, satisfies $h_1(k) \sim \sqrt{2 \ln k}$. Thus we have the asymptotic estimate

$$N_k(\varrho) = \frac{1}{k} + \bigl(1 + \varepsilon(k)\bigr)\frac{2 \ln k}{k}\varrho + O(\varrho^2),$$

where $\varepsilon(k)$ is a function tending to 0, as $k \to \infty$. The result follows by arguments used in the proof of the previous corollary. As before, we need only concern ourselves with negative ϱ; then, plugging the above estimate for $N_k(\varrho)$ into the formula for $A_k(\varrho)$, one finds that $A_k(\varrho)$ is bounded below by $1 - k^{-1} + (1 + \varepsilon(k))\, 2k^{-2} \ln k$ for ϱ is the range $-1/(k-1) \leq \varrho \leq 0$. $\square$

Corollary 8 *$\alpha_3 \geq 0{\cdot}800217$, $\alpha_4 \geq 0{\cdot}850304$, $\alpha_5 \geq 0{\cdot}874243$, $\alpha_{10} \geq 0{\cdot}926642$, and $\alpha_{100} \geq 0{\cdot}990625$.*

Proof We use the bound $N_k(\varrho) \leq 1/k + a_1\varrho + a_2\varrho^2 + \varrho^4/2$, valid for $-1 < \varrho < 0$, and evaluate a_1 and a_2 numerically. The bound follows from two observations: (i) all coefficients a_i are non-negative, and hence the odd terms make a negative contribution to the sum, and (ii) the even coefficients sum to $\frac{1}{2}$, and hence the sum of the even terms from the fourth power upwards is bounded above by $\frac{1}{2}\varrho^4$. $\square$

Note that by computing further terms in the Taylor expansion of $N_k(\varrho)$ it is possible to obtain better bounds on α_k; e.g., expanding to the term in ϱ^4 yields the bound $\alpha_3 \geq 0{\cdot}832718$. The calculations, though routine (more integration by parts, à la proof of Lemma 5), are lengthy, and we shall not try the reader's patience by repeating them here. The returns from this additional computation in any case decline rapidly as k increases.

As remarked in the introduction, the performance ratio for large k, though better than random partitioning heuristic, is hardy impressive. One may obtain upper bounds on the performance ratio that can be achieved using our approach by exhibiting graphs G for which the optimum solution to the relaxation SDP_k is large in relation to the size of the maximum cut in G. Assume $k \geq 2$ is even, and let $G = K_n$ be the complete graph on $n = 3k/2$ vertices. A feasible solution to SDP_k is obtained by placing the vectors v_i at the corners of a $(n-1)$-dimensional equilateral simplex. Then, by routine calculation, the ratio between the actual maximum k-cut in G, and the optimal solution to SDP_k, is at most $1-1/9(k-1)$. This upper bound is much larger that the performance ratios quoted in Theorem 1: perhaps there are much worse instances than the complete graph K_n, or perhaps our heuristic is not extracting as much information from the relaxation as it might.

Finally, as promised, we present the proof of the technical lemma.

Proof of Lemma 5 We begin by computing the joint distribution of $x = u\cdot r$ and $y = v \cdot r$, where $r = (\xi_1, \dots, \xi_n)$ is a random vector. Each of the random vectors r_i in the statement of the lemma induces an independent sample from this distribution. The quantity $N_k(\varrho)$ we wish to estimate is the probability that one sample point dominates the other $k-1$, coordinatewise. Since the density function of r is spherically symmetric, this joint distribution is dependent on θ only, and not on the particular choice of u and v; for convenience let $u = (1, 0, \dots, 0)$ and $v = (\cos\theta, \sin\theta, 0, \dots, 0)$. Then

$$
\begin{aligned}
&\Pr(u \cdot r \le x \ \text{ and } \ v \cdot r \le y) \\
&\quad = \Pr(\xi_1 \le x \ \text{ and } \ \xi_1 \cos\theta + \xi_2 \sin\theta \le y) \\
&\quad = \frac{1}{2\pi} \int_{\xi_1=-\infty}^{x} \int_{\xi_2=-\infty}^{(y-\xi_1\cos\theta)/\sin\theta} \exp\left(-\frac{\xi_1^2+\xi_2^2}{2}\right) d\xi_2\, d\xi_1 \\
&\quad = \frac{1}{2\pi\sin\theta} \int_{\zeta_1=-\infty}^{x} \int_{\zeta_2=-\infty}^{y} \exp\left(-\frac{\zeta_1^2 - 2\cos(\theta)\zeta_1\zeta_2 + \zeta_2^2}{2(\sin\theta)^2}\right) d\zeta_2\, d\zeta_1,
\end{aligned}
$$

where we have applied the change of coordinates $\zeta_1 = \xi_1$ and $\zeta_2 = \xi_1 \cos\theta + \xi_2 \sin\theta$. The joint probability density function of $x = u \cdot r$ and $y = v \cdot r$ is thus

$$
f(x, y; \varrho) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{x^2 - 2\varrho xy + y^2}{2(1-\varrho^2)}\right),
$$

where $\varrho = \cos\theta$; this is the probability density function of the bivariate normal distribution in standard form, with correlation $\varrho = \cos\theta$. Denote by

$$
F(x, y; \varrho) = \int_{\xi=-\infty}^{x} \int_{\eta=-\infty}^{y} f(\xi, \eta; \varrho)\, d\eta\, d\xi
$$

the corresponding cumulative distribution function.

Let $r_1, \ldots, r_k$ be independent random vectors; then

$$\begin{aligned}
&\Pr(u \text{ and } v \text{ are } \textit{not} \text{ separated by } r_1, \ldots, r_k) \\
&\quad = k \times \Pr(u \cdot r_1 = \max_i u \cdot r_i \text{ and } v \cdot r_1 = \max_j v \cdot r_j) \\
&\quad = k\, I(\varrho),
\end{aligned}$$

where

$$I(\varrho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y; \varrho)\, F(x, y; \varrho)^{k-1}\, dx\, dy.$$

There is no expression for the integral $I(\varrho)$ in closed form, so we compute instead a Taylor series expansion for $I(\varrho)$ about $\varrho = 0$ using ideas (and notation) from Bofinger and Bofinger [4]. The *Mehler expansion* [14] of the bivariate normal probability density function

$$f(x, y; \varrho) = g(x)g(y)\bigl(1 + \varrho\phi_1(x)\phi_1(y) + \varrho^2\phi_2(x)\phi_2(y) + \cdots\bigr), \tag{6}$$

converges uniformly for $|\varrho| < 1$. Three facts that follow easily from the Mehler expansion and definition (5) of the Hermite polynomials are:

$$\frac{d}{dx}\, g(x)\phi_{i-1}(x) = -\sqrt{i}\, g(x)\phi_i(x), \tag{7}$$

$$\frac{\partial F(x, y; \varrho)}{\partial \varrho} = f(x, y; \varrho) \tag{8}$$

and

$$\left.\frac{\partial^i f}{\partial \varrho^i}\right|_{\varrho=0} = i!\, g(x)g(y)\phi_i(x)\phi_i(y). \tag{9}$$

We now evaluate $I(\varrho)$ and its successive derivatives with respect to ϱ at the point $\varrho = 0$ by noting that $F(x, y; 0)$ and $f(x, y; 0)$ factorise into $G(x)G(y)$ and $g(x)g(y)$, respectively. In this way we obtain a Taylor series expansion for $I(\varrho)$ about the point $\varrho = 0$. We defer an examination of the radius of convergence of this Taylor expansion to the end of the proof.

Starting with I itself, we have

$$I(0) = \left(\int g(x)G(x)^{k-1} dx\right)^2 = \frac{1}{k^2}, \tag{10}$$

where the second equality can be seen by interpreting the integral as the probability that the maximum of a sequence of k independent, normally distributed random variables is achieved by the first variable.[2]

By identities (8) and (9),

$$\left.\frac{\partial I}{\partial \varrho}\right|_{\varrho=0} = \left(\int g(x)\phi_1(x)G(x)^{k-1}\,dx\right)^2 + (k-1)\left(\int g(x)^2 G(x)^{k-2}\,dx\right)^2.$$

(Passing the derivative through the integral is justified by Section 1.88 of Titchmarsh's text on analysis of functions [13].) The first integral is simply h_1/k; the second may be simplified using integration by parts, and identity (7):

$$\begin{aligned}\int g(x)(g(x)G(x)^{k-2})\,dx &= \left[\frac{g(x)G(x)^{k-1}}{k-1}\right]_{-\infty}^{\infty} - \frac{1}{k-1}\int g'(x)G(x)^{k-1}\,dx \\ &= \frac{1}{k-1}\int g(x)\phi_1(x)G(x)^{k-1}\,dx \\ &= \frac{h_1}{k(k-1)}.\end{aligned}$$

Substituting these expressions for the two integrals yields

$$\left.\frac{\partial I}{\partial \varrho}\right|_{\varrho=0} = \frac{h_1^2}{k(k-1)}. \tag{11}$$

Differentiating with respect to ϱ a second time, we obtain

$$\begin{aligned}\left.\frac{\partial^2 I}{\partial \varrho^2}\right|_{\varrho=0} = 2&\left(\int g(x)\phi_2(x)G(x)^{k-1}\,dx\right)^2 \\ &+ 3(k-1)\left(\int g(x)^2\phi_1(x)G(x)^{k-2}\,dx\right)^2 \\ &+ (k-1)(k-2)\left(\int g(x)^3 G(x)^{k-3}\,dx\right)^2.\end{aligned}$$

[2] Integration will be assumed to be over the infinite line when the limits of integration are omitted.

The first integral is just h_2/k. The second, using integration by parts and identity (7), is

$$\int (g(x)\phi_1(x))(g(x)G(x)^{k-2})\,dx = -\frac{1}{k-1}\int(-\sqrt{2}\,g(x)\phi_2(x))G(x)^{k-1}\,dx$$
$$= \frac{\sqrt{2}\,h_2}{k(k-1)}.$$

A further application of integration by parts reduces the third integral to the second, from which

$$\int g(x)^3 G(x)^{k-3} = \frac{2\sqrt{2}\,h_2}{k(k-1)(k-2)}.$$

Substituting these expressions for the three integrals yields

$$\left.\frac{\partial^2 I}{\partial \varrho^2}\right|_{\varrho=0} = \left(\frac{2}{k^2}+\frac{6}{k^2(k-1)}+\frac{8}{k^2(k-1)(k-2)}\right)h_2^2 = \frac{2h_2^2}{(k-1)(k-2)}. \tag{12}$$

In principle the process of repeated differentiation by ϱ could be continued indefinitely; for any i, the ith derivative of $I(\varrho)$ evaluated at $\varrho = 0$ is a positive linear combination of squares of one-dimensional integrals. This observation, combined with (10), (11), and (12) establishes the claims concerning the Taylor expansion of $I(\varrho)$.

It remains to show that the Taylor expansion of $I(\varrho)$ is valid for $|\varrho| < 1$ and hence — by continuity of $N_k(\varrho)$ at $\varrho = 1$ and the fact that all terms in the expansion are positive — for $|\varrho| \leq 1$. Observe that $I(\varrho)$ is defined by an integral of the form

$$I(\varrho) = \iint \sum_{i=0}^{\infty} \varrho^i s_i(x,y)\,dx\,dy, \tag{13}$$

where $s_i(x,y) = \sum_{j=0}^{n_i-1} t_{ij}(x,y)$ is a sum of terms $t_{ij}(x,y)$, where each term is a product of factors of the form $g(x)g(y)\phi_l(x)\phi_l(y)$. Now $\iint |t_{ij}(x,y)|\,dx\,dy < 2.6$, since $\int |g(x)\phi_l(x)|\,dx < 1.6$ and $\max_x |g(x)\phi_l(x)| < 1$ for all l. (These facts follow from the key inequality on page 324 of Sansone's treatise on orthogonal functions [12], which bounds $|\phi_l(x)|$ by $c\exp(-x^2/4)$ for an absolute constant c; note, however, that the bound given by Sansone is for *un-normalised* Hermite polynomials, and must be scaled accordingly.) Noting that $n_i = O(i^{k-1})$, we see that the sum

$$\sum_{i=0}^{\infty} \varrho^i \sum_{j=0}^{n_i-1} \iint |t_{ij}(x,y)| \, dx \, dy$$

converges, provided $|\varrho| < 1$. Thus, by uniform convergence of the Mehler expansion, and the theorems contained in Sections 1.71 and 1.77 of Titchmarsh [13], it is permissible to integrate (13) term by term, yielding

$$I(\varrho) = \sum_{i=0}^{\infty} \varrho^i \iint s_i(x,y) \, dx \, dy.$$

The above expression is a power series expansion of $I(\varrho)$ valid for $|\varrho| < 1$, which must be identical to the Taylor expansion, by uniqueness. □

3 MAX BISECTION

We now describe how to ensure that the partition we obtain divides V into equal parts. As an integer program we can express MAX BISECTION as

$$\begin{array}{lll} \mathrm{IP_B}: & \text{maximise} & \frac{1}{2}\sum_{i<j} w_{ij}(1 - y_i y_j) \\ & \text{subject to} & \sum_{i<j} y_i y_j \le -n/2 \\ & & y_j \in \{-1, 1\} \quad \forall j \in V \end{array} \tag{14}$$

Inequality (14) is equivalent to $\sum_i y_i = 0$ and expresses the constraint that the sought-for partition must bisect the vertex set; this equivalence can be seen by considering the identity

$$2\sum_{i<j} y_i y_j = \Big(\sum_i y_i\Big)^2 - \sum_i y_i^2 = \Big(\sum_i y_i\Big)^2 - n.$$

The version (14) has the advantage of being easily relaxed to give an semidefinite program:

$$\begin{array}{lll} \mathrm{SDP_B}: & \text{maximise} & \frac{1}{2}\sum_{i<j} w_{ij}(1 - v_i \cdot v_j) \\ & \text{subject to} & \sum_{i<j} v_i \cdot v_j \le -n/2 \\ & & v_j \in S_{n-1}, \quad \forall j \in V \end{array} \tag{15}$$

We can now describe our heuristic. The overall strategy is to solve $\mathrm{SDP_B}$, and partition the resulting vectors v_i by a random hyperplane. The induced partition of vertices is not in general a bisection, but with reasonable probability we do not need to move many vertices to obtain balance. In swapping these relatively few vertices, we do not harm the cut to much. Let ε is a small positive constant, $\varepsilon = 1/100$ being small enough. In more detail, the procedure is as follows.

BISECT

Step 1. Solve the problem $\mathrm{SDP_B}$ to obtain vectors $v_1, v_2, \ldots, v_n \in S_{n-1}$.

Repeat Steps 2-4 below for $t = 1, 2, \ldots K = K(\varepsilon) = \lceil \varepsilon^{-1} \ln \varepsilon^{-1} \rceil$ and output the best partition $\widetilde{S}_t, V \setminus \widetilde{S}_t$ found in Step 4.

Step 2. Choose two *random* vectors z_1, z_2.

Step 3. Let $S_t = \{j : v_j \cdot z_1 \leq v_j \cdot z_2\}$.

Step 4. Suppose (w.l.o.g.) that $|S_t| \geq n/2$. For each $i \in S_t$ let $\zeta(i) = \sum_{j \notin S_t} w_{ij}$ and $S_t = \{x_1, x_2, \ldots, x_\ell\}$, where $\zeta(x_1) \geq \zeta(x_2) \geq \cdots \geq \zeta(x_\ell)$. Also, let $\widetilde{S}_t = \{x_1, \ldots, x_{n/2}\}$.

Clearly the construction in Step 4 satisfies

$$w(\widetilde{S}_t : V \setminus \widetilde{S}_t) \geq \frac{n\, w(S_t : V \setminus S_t)}{2\ell}. \tag{16}$$

In order to analyse the quality of the final partition we define two sets of random variables.

$$\begin{aligned} X_t &= w(S_t : V \setminus S_t), \quad \text{for } 1 \leq t \leq K; \\ Y_t &= |S_t|(n - |S_t|), \quad \text{for } 1 \leq t \leq K. \end{aligned}$$

Recall that $\mathcal{P}_B^*$ denotes the optimum bisection, and let $W^* \geq w(\mathcal{P}_B^*)$ denote the maximum of $\mathrm{SDP_B}$. Then, by the analysis of Theorem 1 (or [8]),

$$\mathbf{E}(X_t) \geq \alpha_2 W^*. \tag{17}$$

Also

$$\mathbf{E}(Y_t) = \sum_{i<j} \Phi_2(v_i \cdot v_j) \geq \frac{\alpha_2}{2} \sum_{i<j} (1 - v_i \cdot v_j) \geq \alpha_2 N,$$

where $N = n^2/4$. (Note the use of (15) here.)

Thus if

$$Z_t = \frac{X_t}{W^*} + \frac{Y_t}{N}$$

then

$$\mathbf{E}(Z_t) \geq 2\alpha_2. \tag{18}$$

On the other hand

$$Z_t \leq 2, \tag{19}$$

since $X_t \leq W^*$ and $Y_t \leq N$.

Define $Z_\tau = \max_{1 \leq t \leq K}\{Z_t\}$. Now (18) and (19) imply that, for any $\varepsilon > 0$,

$$\Pr(Z_1 \leq 2(1-\varepsilon)\alpha_2) \leq \frac{1-\alpha_2}{1-(1-\varepsilon)\alpha_2}$$

and so

$$\Pr(Z_\tau \leq 2(1-\varepsilon)\alpha_2) \leq \left(\frac{1-\alpha_2}{1-(1-\varepsilon)\alpha_2}\right)^K \leq \varepsilon,$$

for the given choice of $K(\varepsilon)$. Assume that

$$Z_\tau \geq 2(1-\varepsilon)\alpha_2 \tag{20}$$

and suppose

$$X_\tau = \lambda W^*,$$

which from (18) and (20) implies

$$Y_\tau \geq (2(1-\varepsilon)\alpha_2 - \lambda)N. \tag{21}$$

Suppose $|S_\tau| = \delta n$; then (21) implies

$$\delta(1-\delta) \geq (2(1-\varepsilon)\alpha_2 - \lambda)/4. \tag{22}$$

Applying (16) and (22) we see that

$$\begin{aligned} w(\widetilde{S}_\tau : V \setminus \widetilde{S}_\tau) &\geq w(S_\tau : V \setminus S_\tau)/(2\delta) \\ &\geq \lambda W^*/(2\delta) \\ &\geq (2(1-\varepsilon)\alpha_2 - 4\delta(1-\delta))W^*/(2\delta) \\ &\geq 2(\sqrt{2(1-\varepsilon)\alpha_2} - 1)W^*. \end{aligned}$$

The last inequality follows from simple calculus. Thus

$$\begin{aligned}\mathbf{E}(w(\widetilde{S}_\tau)) &\geq 2(\sqrt{2(1-\varepsilon)\alpha_2}-1)\left(1-\left(\frac{1-\alpha_2}{1-(1-\varepsilon)\alpha_2}\right)^K\right)W^* \\ &\geq 2(\sqrt{2(1-3\varepsilon)\alpha_2}-1)W^*.\end{aligned}$$

Finally note that the partition output by BISECT is at least as good as $\widetilde{S}_\tau$. We divide ε above by $3\sqrt{2\alpha_2}$ to obtain the precise result presented in Theorem 2. Sanjeev Mahajan has pointed out that the algorithm extends easily to "MAX $(c, 1-c)$-CUT" where the two blocks of the partition are required to have between cn and $(1-c)n$ vertices. For example, one may obtain an 0·81 approximation for MAX $(\frac{1}{3}, \frac{2}{3})$-CUT.

As with MAX k-cut, we can look for upper bounds on the performance ratio that may be obtained using this approach. Let $G = K_{2,2,2}$ be the complete tripartite graph on $2+2+2$ vertices. The maximum bisection of G has 8 edges, whereas the optimum solution to the relaxation $\mathrm{SDP_B}$ is at least 9. (Arrange the vectors v_i in pairs at the corners of an equilateral triangle.) Thus $8/9 < 0{\cdot}889$ is an upper bound on performance ratio that is achievable using this relaxation. Given the somewhat crude nature of the approach, the large gap between this and the provable lower bound on performance ratio is perhaps not surprising.

References

[1] F. Alizadeh, Interior point methods in semidefinite programming with applications to combinatorial optimisation, *Technical Report TR-93-050*, International Computer Science Institute, Berkeley CA, September 1993.

[2] S. Arora, D. Karger and M. Karpinski, Polynomial time approximation schemes for dense instances of NP-hard problem, *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, ACM Press, 1995, pp. 284-293.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy, Proof verification and hardness of approximation problems, Proceedings of the 33rd IEEE Symposium on Foundations of Computer Science, IEEE Computer Society Press, 1992, pp. 14–23.

[4] E. Bofinger and V. J. Bofinger, The correlation of maxima in samples drawn from a bivariate normal distribution, *The Australian Journal of Statistics* **7** (1965), pp. 57–61.

[5] H. A. David, *Order Statistics*, Wiley, New York, 1980.

[6] W. F. de la Vega, *MAXCUT has a randomised approximation scheme in dense graphs*, Manuscript, December 1994.

[7] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York, 1978.

[8] M. X. Goemans and D. P. Williamson, .878-Approximation algorithms for MAX CUT and MAX 2SAT, *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, ACM Press, 1994, pp. 422–431.

[9] V. Kann, S. Khanna, J. Lagergren, and A. Panconesi, On the hardness of approximating MAX k-CUT and its dual, *Technical Report TRITA-NA-9505*, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1995.

[10] D. Karger, R. Motwani, and M. Sudan, Approximate graph coloring by semidefinite programming, *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science*, Computer Society Press, 1994, pp. 2–13.

[11] C. H. Papadimitriou and M. Yannakakis, Optimization, approximation and complexity classes, *Journal of Computer and System Sciences* **43** (1991), pp. 425–440.

[12] G. Sansone, *Orthogonal Functions*, (translated from the Italian by A. H. Diamond), Interscience Publishers, New York, 1959.

[13] E. C. Titchmarsh, *The Theory of Functions* (second edition), Oxford University Press, 1939.

[14] G. N. Watson, Notes on generating functions of polynomials: Hermite polynomials, *Journal of the London Mathematical Society* **8** (1933), pp. 194–199.

[15] D. J. A. Welsh, Complexity: knots, colourings and counting, *London Mathematical Society Lecture Notes* **186**, Cambridge University Press, 1993.

[16] L. A. Wolsey, Heuristic analysis, linear programming and branch and bound, *Mathematical Programming Study* **13:** *Combinatorial Optimization II*, North-Holland, 1980, pp. 121–134.