Improved approximation algorithms for MAX k-CUT and MAX BISECTION

by

Alan Frieze
Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA, USA

and

Mark Jerrum
Department of Computer Science
University of Edinburgh
Edinburgh, UK

Research Report No. 94-173
October, 1994

Improved approximation algorithms for MAX k-CUT and MAX BISECTION

Alan Frieze* Mark Jerrum†

Carnegie Mellon University University of Edinburgh

October 12, 1994

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 relaxation.

*Department of Mathematics, Carnegie Mellon University, Pittsburgh PA15213, U.S.A., Supported by NSF grant CCR-9225008.

†Department of Computer Science, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, UK. The second author is a Nuffield Science Research Fellow, and is supported in part by grant GR/F 90363 of the UK Science and Engineering Research Council, and Esprit Working Group "RAND;" this work was done while visiting Carnegie Mellon University.

1 Introduction

Goemans and Williamson [5] 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 some time ago by Wolsey [11]. (One significant exception to this general rule has been the case of Bin Packing.)

The main novelty of [5] is that it uses a Semi-Definite Program (SDP) as a relaxation. To be more precise let us consider the problem MAX-CUT studied (among others) in [5]: we are given a vertex set $V = \{1, \ldots, n\}$ and non-negative weights $w_{i,j}$, $1 \leq i, j \leq n$, where $w_{i,j} = w_{j,i}$ and $w_{i,i} = 0$ for all i, j. If $S \subseteq V$ and $\overline{S} = V \setminus S$ then the *weight* of the *cut* $(S : \overline{S})$ is

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

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

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

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

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

Here $S_n = \{x \in \mathbf{R}^n : ||x|| = 1\}$ is the unit sphere in n dimensions. SDP's are a special class of convex program (see Alizadeh [1] for a detailed exposition). In particular the above problem can be replaced by

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

Here $Y_{i,j}$ replaces $v_i \cdot v_j$ and the notation $Y \succ 0$ indicates that the matrix Y is constrained to be positive semi-definite; this constraint defines a convex subset of $\mathbf{R}^{n^2}$. 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

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

The problem is then

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

Note that MAX k-CUT has an important interpretation as the search for a ground state in the anti-ferromagnetic k-state Potts model: see Welsh [10].

To attack this problem we need to be able to handle variables which can take on one of k values as opposed to just two, a similar problem to that faced in trying to colour graphs. Our solution is a natural extension of the existing solution for the case $k = 2$, but the performance analysis presents greater technical difficulties. Karger, Motwani and Sudan [6]have indepenently used the same partitioning heuristic as us to get improved bounds on approximate graph colouring.

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

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

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 which produces a partition $\mathcal{P}_k$. We prove the existence of a sequence of constants $\alpha_k, k \geq 2$ 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 α_k *satisfy*

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

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

(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 $k = 2$ is the same as that achieved by Goemans and Williamson, as our heuristic is a generalisation of theirs.

Our next 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.

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

A random bisection produces an expected guarantee of $\frac{1}{2}$. We describe a heuristic BISECT which produces a partition $\mathcal{P}_B$ such that if $\mathcal{P}_B^*$ denotes the optimal bisection,

Theorem 2 *Let ϵ be a small positive constant. Then $\mathbf{E}(w(\mathcal{P}_B)) \geq \beta\, w(\mathcal{P}_B^*)$ where $\beta = 2(\sqrt{2(1-\epsilon)\alpha_2} - 1)$, which is greater than 0·65 for ϵ sufficiently small.*

Note that $\alpha_2 = 0{\cdot}878567\ldots$, as in Theorem 1. The difficulty with generalising Goemans and Williamson's heuristic to MAX BISECTION is that their heuristic does not generally give a bisection of V. We prove that a simple modification of their basic algorithm beats the trivial $\frac{1}{2}$ lower bound.

Note that there is a natural generalisation of this problem 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 $\mathbf{R}^{k-1}$ 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 1

$$a_i \cdot a_j = -1/(k-1), \quad \textit{for } i \neq j. \tag{3}$$

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 $(b_k - c_k) \cdot (b_1 - c_k) = -(k-1)x^2 = -1/(k-1)$. □

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

Lemma 2 *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

$$0 \leq (u_1 + u_2 + \cdots + u_k)^2$$

$$\leq \; k + k(k-1)\gamma.$$

□

Given Lemma 1 we can formulate MAX k-CUT as follows:

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

Here we use the fact that

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

To obtain our SDP relaxation we replace y_i by v_i, where v_i can now be any vector in S_n. 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

$$\begin{array}{lll} \mathrm{SDP}_k\text{:} & \text{maximise} & \frac{k-1}{k}\sum_{i<j} w_{i,j}(1 - v_i \cdot v_j) \\ & \text{subject to} & v_j \in S_n, \quad \forall j \\ & & v_i \cdot v_j \geq -1/(k-1), \quad \forall i \neq j \end{array} \qquad (4)$$

Note that (4) reduces to the linear constraint $Y_{i,j} \geq -1/(k-1)$ 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 $v_1, v_2, \ldots, v_n \in S_n$.

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

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

(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. 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 the heuristic of Goemans and Williamson, although they define it in terms of cutting S_n by a random hyperplane through the origin.

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 $j \in P_i$, $1 \leq i \leq k$ we see that

$$\mathbf{E}(W_k) = \sum_{i<j} w_{i,j} \Pr(y_i \neq y_j). \tag{5}$$

Now by symmetry $\Pr(y_i \neq y_j)$ depends only on the angle θ between v_i and v_j, and hence on $\rho = \cos\theta = v_i \cdot v_j$. Let this separation probability be denoted by $\Phi_k(\rho)$. It then follows from (5) that

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

$$\geq \alpha_k,$$

where

$$\alpha_k = \min_{-1/(k-1)\leq\rho\leq 1} \frac{k\,\Phi_k(\rho)}{(k-1)(1-\rho)}.$$

We leave the estimation of the α_k to an appendix (see Corollaries 1, 2, and 3). Suffice it to say that they satisfy the claims of Theorem 1.

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

$$\begin{array}{lll} \mathrm{IP_B}: & \text{maximise} & \frac{1}{2}\sum_{i<j} w_{i,j}(1-y_iy_j) \\ & \text{subject to} & \sum_{i<j} y_iy_j \leq -n/2 \\ & & y_j \in \{-1,1\} \quad \forall j \in V \end{array} \tag{6}$$

Constraint (6) expresses the fact that we force $|S| = n/2$ by maximising the number of pairs i, j where $i \in S, j \notin S$. It has the advantage of being easily relaxed to give an SDP problem:

$$\begin{array}{lll} \mathrm{SDP_B}: & \text{maximise} & \frac{1}{2}\sum_{i<j} w_{i,j}(1-v_i\cdot v_j) \\ & \text{subject to} & \sum_{i<j} v_i\cdot v_j \leq -n/2 \\ & & v_j \in S_n, \quad \forall j \in V \end{array} \tag{7}$$

We can now describe our heuristic: ϵ is a small positive constant, $\epsilon = 1/100$ is small enough.

BISECT

Step 1 solve the problem $\mathrm{SDP_B}$ to obtain vectors $v_1, v_2, \ldots, v_n \in S_n$.

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

Step 2 choose 2 *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 $\zeta(i) = \sum_{j \notin S_t} w_{i,j}$ and let $S_t = \{x_1, x_2, \ldots, x_\ell\}$ where $\zeta(x_1) \geq \zeta(x_2) \geq \cdots \geq \zeta(x_\ell)$. Let $\tilde{S}_t = \{x_1, \ldots, x_{n/2}\}$.

Clearly the construction in Step 4 satisfies

$$w(\tilde{S}_t : V \setminus \tilde{S}_t) \geq \frac{n\, w(S_t : V \setminus S_t)}{2\ell}. \tag{8}$$

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

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

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 [5]),

$$\mathbf{E}(X_t) \geq \alpha_2 W^*. \tag{9}$$

Also

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

where $N = n^2/4$ (note the use of (7) here.)

Thus if

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

then

$$\mathbf{E}(Z_t) \geq 2\alpha_2. \tag{10}$$

On the other hand

$$Z_t \leq 2, \tag{11}$$

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

Define $Z_\tau = \max_{1 \leq t \leq K}\{Z_t\}$. Now (10) and (11) imply that for any $\epsilon > 0$

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

and so

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

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

$$Z_\tau \geq 2(1-\epsilon)\alpha_2 \tag{12}$$

and suppose

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

which from (10) and (12) implies

$$Y_\tau \geq (2(1-\epsilon)\alpha_2 - \lambda)N. \tag{13}$$

Suppose $|S_\tau| = \delta n$; then (13) implies

$$\delta(1-\delta) \geq (2(1-\epsilon)\alpha_2 - \lambda)/4. \tag{14}$$

Applying (8) and (14) we see that

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

The last inequality follows from simple calculus.

Thus

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

Finally note that the partition output by BISECT is at least as good as $\widetilde{S}_\tau$. We divide ϵ above by 3 to get the precise result.

4 Appendix

Let u, v be vectors, and $r_1, \ldots, r_k$ be a sequence of vectors, all in $\mathbf{R}^n$. We say that u and v are *separated by* $r_1, \ldots, r_k$ if the vector r_i minimising $u \cdot r_i$ is distinct from the vector r_j minimising $v \cdot r_j$. 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

$$(-1)^i \sqrt{i!}\, \phi_i(x) g(x) = \frac{d^i g(x)}{dx^i}. \tag{15}$$

Let h_i 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 3 *Suppose $u, v \in \mathbf{R}^n$ are unit vectors at angle θ, and $r_1, \ldots, r_k$ is a sequence of random vectors. Let $\rho = \cos\theta = u \cdot v$, and denote by $N_k(\rho) = 1 - \Phi_k(\rho)$ the probability that u and v are* not *separated by $r_1, \ldots, r_k$. Then the Taylor series expansion*

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

of $N_k(\rho)$ about the point $\rho = 0$ converges for all ρ in the range $|\rho| \le 1$. 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 $a_2 = kh_2^2/(k-1)(k-2)$.

Proof We begin by computing the joint distribution of $x = u \cdot r$ and $y = v \cdot r$, where $r = (\xi_1, \ldots, \xi_n)$ is a random vector. 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 $u = (1, 0, \ldots, 0)$ and

$v = (\cos\theta, \sin\theta, 0, \ldots, 0)$. Then

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

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

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

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

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

the corresponding cumulative distribution function.

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

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

where

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

There is no expression for the integral $I(\rho)$ in closed form, so we compute instead a Taylor series expansion for $I(\rho)$ about $\rho = 0$ using ideas (and notation) from Bofinger and Bofinger [2]. The *Mehler expansion* [9] of the bivariate normal probability density function

$$f(x, y; \rho) = g(x)g(y)\Big(1 + \rho\phi_1(x)\phi_1(y) + \rho^2\phi_2(x)\phi_2(y) + \cdots\Big), \tag{16}$$

converges uniformly for $|\rho| < 1$. Three facts that follow easily from the Mehler expansion and definition (15) of the Hermite polynomials are:

$$\frac{d}{dx}\, g(x)\phi_{i-1}(x) = -\sqrt{i}\, g(x)\phi_i(x), \tag{17}$$

$$\frac{\partial F(x, y; \rho)}{\partial \rho} = f(x, y; \rho) \tag{18}$$

and

$$\left.\frac{\partial^i f}{\partial \rho^i}\right|_{\rho=0} = i!\, g(x)g(y)\phi_i(x)\phi_i(y). \tag{19}$$

We now evaluate $I(\rho)$ and its successive derivatives with respect to ρ at the point $\rho = 0$ 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 $I(\rho)$ about the point $\rho = 0$. 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

$$I(0) = \left(\int g(x)G(x)^{k-1}dx\right)^2 = \frac{1}{k^2}, \tag{20}$$

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 is achieved by the first variable.[1]

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

By identities (18) and (19),

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

(Passing the derivative through the integral is justified by Section 1.88 of Titchmarsh's text on analysis of functions [8].) The first integral is simply h_1/k; the second may be simplified using integration by parts, and identity (17):

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

Substituting these expressions for the two integrals yields

$$\left.\frac{\partial I}{\partial \rho}\right|_{\rho=0} = \frac{h_1^2}{k(k-1)}. \tag{21}$$

Differentiating with respect to ρ a second time, we obtain

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

The first integral is just h_2/k. The second, using integration by parts and identity (17), is

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

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

$$\left.\frac{\partial^2 I}{\partial \rho^2}\right|_{\rho=0} = \left(\frac{2}{k^2} + \frac{6}{k^2(k-1)} + \frac{8}{k^2(k-1)(k-2)}\right) h_2^2 = \frac{2h_2^2}{(k-1)(k-2)}. \tag{22}$$

In principle the process of repeated differentiation by ρ could be continued indefinitely; for any i, the ith derivative of $I(\rho)$ evaluated at $\rho = 0$ is a positive linear combination of squares of one-dimensional integrals. This observation, combined with (20), (21), and (22) establishes the claims concerning the Taylor expansion of $I(\rho)$.

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

$$I(\rho) = \iint \sum_{i=0}^{\infty} \rho^i s_i(x, y)\, dx\, dy, \tag{23}$$

where $s_i(x, y) = \sum_{j=0}^{n_i - 1} t_{ij}(x, y)$ is a sum of terms $t_{ij}(x, y)$, and each term $t_{ij}(x, y)$ 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 [7], 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.) Noting that $n_i = O(i^{k-1})$, we see that the sum

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

converges, provided $|\rho| < 1$. Thus — by uniform convergence of the Mehler expansion, and the theorems contained in Sections 1.71 and 1.77 of Titchmarsh [8] — it is permissible to integrate (23) term by term, yielding

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

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

Denote by $A_k(\rho)$ the function

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

and recall that the performance ratio of the k-CUT heuristic is given by

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

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

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

Turning to the case $\rho < 0$, 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 [3, Section 3.2]), we have $a_1 \geq 9/4\pi(k-1)$. Therefore,

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

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

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

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

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

Proof Galambos [4, 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

$$N_k(\rho) = 1 - \frac{1}{k} + \left(1 + \epsilon(k)\right)\frac{2 \ln k}{k}\rho + O(\rho^2),$$

where $\epsilon(k)$ is a function tending to 0, as $k \to \infty$. The result follows by arguments used in the proof of the previous corollary. □

Corollary 3 *$\alpha_2 \geq 0{\cdot}878567$, $\alpha_3 \geq 0{\cdot}800217$, $\alpha_4 \geq 0{\cdot}850304$, $\alpha_5 \geq 0{\cdot}874243$, $\alpha_{10} \geq 0{\cdot}926642$, and $\alpha_{100} \geq 0{\cdot}990625$.*

Proof The value of α_2 was obtained by Goemans and Williamson. For $k \geq 3$, we use the bound $N_k(\rho) \leq 1 - 1/k + a_1\rho + a_2\rho^2 + \rho^4/2$, valid for

$-1 < \rho < 0$, and evaluate a_1 and a_2 numerically. (Observe that the coefficient of ρ^3 is positive, and hence the term itself makes a negative contribution.) Note that by computing further terms in the Taylor expansion of $N_k(\rho)$ it is possible to give better bounds on α_k; e.g., by expanding to the term in ρ^4, we obtain $\alpha_3 \geq 0{\cdot}832718$. □

References

[1] F. Alizadeh, *Interior point methods in Semi-Definite Programming with applications to Combinatorial Optimisation.* (To appear.)

[2] 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.

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

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

[5] 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* (1994) pp 422–431.

[6] D. Karger, R. Motwani, and M. Sudan, *Improved graph coloring by semidefinite programming.* (In preparation.)

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

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

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

[10] D. J. A. Welsh, Complexity: Knots, Colourings and Counting, *London Mathematical Society Lecture Notes* **186**, Cambridge University Press, 1993.

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