

A mildly exponential time algorithm for
approximating the number of solutions to a
multidimensional knapsack problem

Martin Dyer
University of Leeds
Leeds LS2 9JT
UK

Alan Frieze Ravi Kannan
Ajai Kapoor Ljubomir Perkovic
Carnegie Mellon University
Pittsburgh PA15213
USA

Umesh Vazirani
University of California
Berkeley CA94320
USA

May 22, 2006

Abstract

We describe a $2^{O(r\sqrt{n}(\log n)^{5/2})}\epsilon^{-2}$ time randomized algorithm which estimates the number of feasible solutions of a multidimensional knapsack problem within $1 \pm \epsilon$ of the exact number. (Here r is the number of constraints and n is the number of integer variables.) The algorithm

uses a Markov chain to generate an almost uniform random solution to the problem.

1 Introduction

An instance of the multidimensional knapsack problem in n variables is defined by a non-negative $r \times n$ integer matrix A , a vector $d \in \mathbb{Z}^n$, and a vector $b \in \mathbb{Z}^r$. We assume, to avoid trivialities, that all components of b and d are positive. The set of feasible solutions, K , is then

$$K = \{x \in \mathbb{Z}^n \mid Ax \leq b, 0 \leq x \leq d\}. \quad (1)$$

We will assume, without loss of generality that $d_j \leq \min_i b_i/a_{ij}$. Let $d_{\max} = \max_j d_j$. In particular, if $d_{\max} = 1$, we have the *zero-one* multidimensional knapsack problem, and if $r = 1$ we have the knapsack problem. The *optimization* problem consists of finding the maximum value of a given linear function over K . This is known to be NP-Complete, even for $r = 1$ and zero-one variables. The *counting* problem seems to be even harder. This asks for $|K|$. This problem is #P-complete, again even for $r = 1$ and zero-one variables.

Markov chains have been successfully used to approximately solve several #P-complete problems [1, 2, 3, 4, 6, 7, 10, 12, 14]. In all of these problems the running time of the algorithm is polynomial in problem size and relative error. On the other hand, the general zero-one permanent still resists polynomial time approximation, but Jerrum and Vazirani [9] have reduced the time complexity to $2^{O(\sqrt{n}(\log n)^2)}\epsilon^{-2}$ for computing an ϵ -approximation i.e. an approximation with relative error in the range $1 \pm \epsilon$.

In this paper we make similar progress in approximating $|K|$. We show that there is an $2^{O(r\sqrt{n}(\log n)^{5/2})}\epsilon^{-2}$ time algorithm for computing an ϵ -approximation to $|K|$. Furthermore when $d_j \geq n^2$ for $j \in [n] = \{1, 2, \dots, n\}$ we can make the scheme polynomial.

Note that there is an easy randomized approximation algorithm, given by generating random points as follows. Suppose we simply choose x_j randomly

in $[0, d_j]$. It is obvious that

$$\Pr(x_j \leq \lfloor d_j/n \rfloor) \geq 1/(2n).$$

However, if $x_j \leq \lfloor d_j/n \rfloor$ ($j \in [n]$) we have

$$Ax \leq \frac{1}{n}Ad \leq b,$$

since $a_{ij}d_j \leq b_i$ for all $i \in [r]$, $j \in [n]$. Therefore $x \in K$. Thus, by randomly generating x in this fashion, we have probability $(2n)^{-n}$ of having $x \in K$. Hence a uniformly distributed point in K can be generated in $(2n)^{n+1}$ time with very high probability. This easily gives an approximation algorithm with the time complexity $n^{O(n)}r\epsilon^{-2}$. As previously stated our approximation algorithm runs in time $2^{O(r\sqrt{n}(\log n)^{5/2})}$ and so we will assume here that $r = O(n^\kappa)$ for some fixed $\kappa < \frac{1}{2}$. (In the zero-one case, an $O(rn2^n)$ time *exact deterministic* algorithm is, of course, trivial.) It may be worth pointing out that while there are fully polynomial time approximation schemes for the optimisation version of the knapsack problem, there are none for the multidimensional problem unless $P=NP$. Furthermore, we have not been able to make any use of these optimisation results for knapsack in the associated counting problem.

We develop the argument in three steps. First we examine the zero-one knapsack problem in section 3, then move to the general zero-one case in section 4. Finally, we consider the general problem in section 5.

As a tool in our proof we prove a general fact about random sampling (Theorem 6) that may be of independent interest and so we mention it here: suppose we have a population of n non-negative reals c_1, c_2, \dots, c_n with mean μ and $c_i \leq \alpha$, for $i = 1, 2, \dots, n$, and we pick without replacement a sample of t c_i 's; let S_t be their sum. The Chernoff-Hoeffding Theorems assert

$$\Pr(S_t - t\mu \geq x) \leq \exp \left\{ -\frac{2x^2}{t\alpha^2} \right\} \quad \text{for all } x \geq 0.$$

We prove a similar bound, but where α is replaced by a quantity that we have "observed" from the samples. Let S_m be the sum of the m largest samples among the t we have picked. Then we prove

$$\Pr(S_t - t\mu > S_m) = O(\exp\{-m^2(t-m)^2/t^3\}) \quad \text{for } t = m+1, \dots, n.$$

2 Rapidly Mixing Markov Chains

It is well known that the approximate counting of combinatorial objects is intimately related to their near uniform generation—see Jerrum, Valiant and Vazirani [8] for a formal statement of this relation.

In our problem this general reduction implies that if in time $T(n)$ we can generate a solution $x \in K$ with probability $p(x)$ so that say, $p(x)/p(y) \leq 2$ for all $x, y \in K$ then it takes $O(T(n)s(n)\epsilon^{-2} \log \delta^{-1})$ time to compute an ϵ -approximation with probability at least $1 - \delta$, where $s(n)$ is a polynomial in n . As our $T(n)$ dominates the remaining terms, we concentrate on estimating its value from now on.

For this purpose we define an ergodic reversible Markov chain with set of states K , whose steady state distribution is uniform. We then run this chain for sufficient number of steps that we can be sure we are close enough to the steady state. We then take the current state as our near uniform sample.

In detail, let K be the set of states of Markov chain \mathcal{M} . If $x, y \in \mathbb{Z}^n$, define the Hamming distance $h(x, y) = |\{j : x_j \neq y_j\}|$. (Normally this is defined only for zero-one vectors, and this will be the principal use here, but we also use the definition for general integer vectors.) The transition probabilities $p(x, y)$ are then given by

$$p(x, y) = \begin{cases} 0 & \text{if } h(x, y) > 1 \\ \frac{1}{2nd_{\max}} & \text{if } h(x, y) = 1 \\ 1 - \sum_{z \neq x} p(x, z) & \text{if } y = x \end{cases}$$

Note that \mathcal{M} can be simulated easily, as follows. Start at any point of K (e.g. the origin). At the “current” point $x \in K$, repeat the following step. Do nothing with probability $\frac{1}{2}$, otherwise choose a coordinate direction $k \in [n]$ uniformly at random. Choose an integer $x'_k \neq x_k$ from $[0, d_{\max}]$ uniformly at random. Let $x'_j = x_j$ ($j \in [n] \setminus \{k\}$). If $x' \in K$, x' becomes the new current point, else remain at x . Note also that, since \mathcal{M} is symmetric, aperiodic and irreducible, its stationary distribution π is uniform over its states.

Now let $p^{(t)}$ be the distribution of the current point after t steps, starting at $x = 0$. The variational distance between $p^{(t)}$ and π can be bounded in terms

of the *conductance*, Φ :

$$\Phi = \min \left\{ \Phi_S : \pi(S) \leq \frac{1}{2} \right\}$$

where for $S \neq \emptyset$,

$$\Phi_S = \frac{\sum_{x \in S, y \in \bar{S}} \pi(x) p(x, y)}{\pi(S)},$$

and

$$\pi(S) = \sum_{x \in S} \pi(x).$$

It will be useful to think in terms of a digraph $\Gamma = (K, E)$ where $E = \{(x, y) \in K^2 : h(x, y) = 1\}$. If we let $E_{S, \bar{S}} = \{(x, y) \in E : x \in S, y \notin S\}$ and $\rho(S) = |E_{S, \bar{S}}|/|S|$ then

$$\Phi = \min_{|S| \leq N/2} \left\{ \frac{\rho(S)}{2nd_{\max}} \right\} \quad (2)$$

Now, letting $N = |K|$, it follows from results in Sinclair and Jerrum [15] that for all $x \in K$

$$|p^{(t)}(x) - \pi(x)| \leq \sqrt{N}(1 - \Phi^2/2)^t.$$

Thus our claim about near uniform generation can be justified by showing that the conductance Φ of \mathcal{M} satisfies

$$\Phi = 2^{-O(r\sqrt{n}(\log n)^{5/2})}. \quad (3)$$

In view of (2), assuming that d_{\max} is suitably bounded as a function of n , this will then follow from

$$\rho(S) = 2^{-O(r\sqrt{n}(\log n)^{5/2})} \quad \text{for all } S \subseteq K, |S| \leq N/2. \quad (4)$$

We will use a variation on the canonical path argument introduced by Jerrum and Sinclair [6]. Fix $S \subseteq K, |S| \leq N/2$. We will define a set $T = T(S) \subseteq S \times \bar{S}$ with

$$|T| \geq \frac{|S| \cdot |\bar{S}|}{2}. \quad (5)$$

We will also define a *canonical* path $P_{x,y}$ in Γ from x to y for every $(x, y) \in T$. For each $(v, w) \in E_{S, \bar{S}}$ we let $W_{v,w}$ denote the set of canonical paths using the

edge (v, w) . Let $\lambda_{v,w} = |W_{v,w}|$ and let $\lambda = \lambda(S) = \max\{\lambda_{v,w} : (v, w) \in E_{S,\bar{S}}\}$. Clearly

$$\begin{aligned} |E_{S,\bar{S}}| &\geq \frac{|T|}{\lambda} \\ &\geq \frac{|S| \cdot |\bar{S}|}{2\lambda} \end{aligned}$$

and so

$$\rho(S) \geq \frac{N}{4\lambda}.$$

We will complete the proof of (4) by showing that for any S

$$\lambda = 2^{O(r\sqrt{n}(\log n)^{5/2})} N. \quad (6)$$

3 The Zero-One Knapsack Problem

We concentrate first on the zero-one case. The case $r = 1$ has a special feature which allows a slightly simpler proof. We will give this and then analyse the general zero-one case later. For simplicity of notation we now let $K = \{x \in \{0, 1\}^n : a_1x_1 + a_2x_2 + \dots + a_nx_n \leq b\}$ where a_1, a_2, \dots, a_n, b are positive integers.

3.1 Defining Canonical Paths

Consider a permutation σ on $[n]$, an *offset* u ($0 \leq u < n$) and states $x \in S$ and $y \in \bar{S}$. We define the sequence $Q_{x,y} = Q_{x,y}(\sigma, u) = v^0, v^1, \dots, v^n$ as follows: $v^0 = x, v^n = y$ and $v_j^i = v_j^{i-1}$ for $j \neq \sigma(i \oplus u)$ and $v_{\sigma(i \oplus u)}^i = y_{\sigma(i \oplus u)}$, $i \oplus u = (i + u - 1 \bmod n) + 1$. Thus, we go through the components of x in the order $\sigma(1 \oplus u), \sigma(2 \oplus u), \dots, \sigma(n \oplus u)$, changing one component from x_j to y_j at each step. If all intermediate points are in K then we say $Q_{x,y}$ is feasible. (We may remove the loops caused by $v^i = v^{i+1}$.)

Assume we have chosen σ, u and for notational convenience assume that σ is the identity permutation and $u = 0$. Fix $(v, w) \in E_{S,\bar{S}}$ and consider those

$(x, y) \in S \times \bar{S}$ for which the path $Q_{x,y}$ uses (v, w) . Then for some $t \geq 1$ we can write

$$v = (y_1 \dots y_{t-1} x_t x_{t+1} \dots x_n) \quad \text{and} \quad w = (y_1 \dots y_{t-1} y_t x_{t+1} \dots x_n).$$

Then σ , u , and (v, w) together fix y_1, \dots, y_t and x_t, \dots, x_n and $\lambda_{v,w}$ is bounded by the number of possibilities for $x_1, \dots, x_{t-1}, y_{t+1}, \dots, y_n$. This leads to the idea of the *complementary point* $w' = (x_1, \dots, x_t, y_{t+1}, \dots, y_n)$ of (v, w) from Jerrum and Sinclair [15]. Let $W' = W'(x, y)$ be the set of complementary points w' such that (v, w) is on $Q_{x,y}$. Clearly, if we use the same σ for each x, y then the number of (x, y) pairs in $S \times \bar{S}$ for which $Q_{x,y}$ uses (v, w) is at most n times the total possible number of complementary points for w . (Since n is the number of choices for u .)

Now let $D_\beta = \{\xi : h(\xi, \eta) \leq \beta \text{ for some } \eta \in K\}$. Then, the following lemma holds:

Lemma 1 *There exist σ^* and $T^* \subset S \times \bar{S}$ such that $|T^*| \geq \frac{|S||\bar{S}|}{2}$ and*

- (i) $W' \subseteq D_{4m}$, where $m = \lceil 4\sqrt{n \log_2 n} \rceil$.
- (ii) For all $(x, y) \in T^*$, there exists $u^* = u^*(x, y)$ such that $P_{x,y} = Q_{x,y}(\sigma^*, u^*)$ is feasible.

Proof Fix $x \in S$, $y \in \bar{S}$. Choose σ randomly, u arbitrarily and an intermediate point v of $Q_{x,y}$, say the ℓ -th point in the sequence. For $\eta \in \{0, 1\}^n$, let its support $I_\eta = \{j : \eta_j = 1\}$. Let $I = \{\sigma(1 \oplus u), \sigma(2 \oplus u), \dots, \sigma(\ell \oplus u)\}$. Then

$$I_{w'} = (I_x \cap I) \cup (I_y \setminus I)$$

where w' is the complementary point corresponding to the edge (v, w) of $Q_{x,y}$ which starts with v .

Now for $X \subseteq [n]$ let $a(X) = \sum_{j \in X} a_j$. Then $a(I_x \cap I)$ is the sum of the elements of a random ℓ -subset of the *multiset* $A_x = \{a_j x_j : j \in [n]\}$. Assume next that $|I_x \cap I| \geq 2m$ and let Δ_x denote the sum of the m largest elements of the multiset $\{a_j : j \in I_x \cap I\}$. Now, where $ax = \sum_{j \in [n]} a_j x_j$,

$$E(a(I_x \cap I)) = \mu_x = \frac{\ell}{n} ax$$

and so from Theorem 6 (see Section 7)

$$\begin{aligned}\Pr(a(I_x \cap I) \geq \mu_x + \Delta_x) &= O(e^{-m^2/4\ell}) \\ &= O(n^{-4}).\end{aligned}$$

Thus, regardless of the size of $I_x \cap I$ we can assert that with probability $1 - O(n^{-4})$ there exists a set $J_x \subseteq I_x \cap I$ such that

$$|(I_x \cap I) \setminus J_x| \leq 2m \quad (7)$$

$$a(J_x) \leq \mu_x. \quad (8)$$

Now consider $I_y \setminus I$.

$$E(a(I_y \setminus I)) = \mu_y = \frac{n - \ell}{n} ay.$$

Then, in an analogous fashion to $I_x \cap I$ we have that with probability $1 - O(n^{-4})$ there exists a set $J_y \subseteq I_y \setminus I$ such that

$$|(I_y \setminus I) \setminus J_y| \leq 2m \quad (9)$$

$$a(J_y) \leq \mu_y. \quad (10)$$

Now define w'' by $I_{w''} = J_x \cup J_y$. Then (7) and (9) imply

$$h(w', w'') \leq 4m.$$

We find that $w'' \in K$. Indeed, using (8) and (10)

$$\begin{aligned}aw'' &= a(J_x) + a(J_y) \\ &\leq \frac{\ell}{n} ax + \frac{n - \ell}{n} ay \\ &\leq \frac{\ell}{n} b + \frac{n - \ell}{n} b \\ &= b.\end{aligned} \quad (11)$$

So for each of the pairs (x, y) such that $x \in S$ and $y \in \bar{S}$ we have that the probability a random permutation satisfies

$$w' \in D_{4m} \text{ for all complementary points } w' \text{ arising from all offsets,} \quad (12)$$

is at least $1 - O(\frac{1}{n^2})$. Then by a simple counting argument we have that there must exist some permutation σ^* with at least a $1 - O(\frac{1}{n^2})$ fraction of x, y pairs satisfying (12).

For each x, y there is an offset $u^* = u^*(x, y)$ such that $Q_{x,y}(\sigma^*, u^*)$ is feasible (see Lovász [11], Problem 3.21, the *gas station* problem). Thus there exists a set $T^* \subseteq S \times \bar{S}$ such that

$$|T^*| \geq |S| \cdot |\bar{S}|(1 - O(n^{-2})) \quad (13)$$

and $P_{x,y} = Q_{x,y}(\sigma^*, u^*)$ is feasible for all $(x, y) \in T^*$. This completes the proof of the lemma. \square

Finally observe that for $(v, w) \in E_{S;\bar{S}}$ we have

$$\begin{aligned} \lambda_{v,w} &\leq n|D_{4m}| \\ &\leq nN \binom{n}{4m} \\ &= 2^{O(\sqrt{n}(\log n)^{3/2})} N \end{aligned}$$

and (6) follows immediately.

4 The General Zero-One Problem

When $r \geq 2$ the argument that the complementary points are almost always “close” to K is still valid, but it is no longer possible to use an offset to make the intermediate points of $Q_{x,y}$ satisfy all the constraints simultaneously. Since offsets are no longer in use we drop the associated parameter in $Q_{x,y}$.

For $R \subseteq [n]$ and $x \in \{0, 1\}^n$ we define $z = z(x, R)$ by $I_z = I_x \setminus R$ i.e. to obtain z , simply set components j to zero for $j \in R$.

Given $x, y \in K$, permutation σ and $R \subseteq J(x, y) = (I_x \setminus I_y) \cup (I_y \setminus I_x)$ we define a new permutation $\tau = \tau(x, y, \sigma, R)$ and then take $Q_{x,y}(\tau)$ as our path. Here τ is defined as follows: let $R_\alpha = R \cap I_\alpha$ and $r_\alpha = |R_\alpha|$ for $\alpha = x, y$. Starting with the sequence $\sigma(1), \sigma(2), \dots, \sigma(n)$ we move the elements of R_x to the front (in natural order) and then move the elements of R_y to the back. τ is defined

by the sequence obtained e.g. if $\sigma = (10, 9, 1, 3, 7, 4, 8, 5, 6, 2)$ and $R_x = \{7, 9\}, R_y = \{3, 5\}$ then $\tau = (7, 9, 10, 1, 4, 8, 6, 2, 3, 5)$.

Lemma 2 *There exists σ^* and $T^* \subseteq S \times \bar{S}$ such that $|T^*| \geq \frac{|S||\bar{S}|}{2}$ and*

- (i) $W' \subseteq D_{16mr \log_2 n + 4mr}, \quad (m = \lceil 4\sqrt{n \log_2 n} \rceil)$
- (ii) *For all $(x, y) \in T^*$, there exists $R = R(x, y), |R| \leq 16mr \log_2 n$ such that $Q_{x,y}(\tau)$ is feasible.*

Proof Fix $x \in S, y \in \bar{S}$ and choose σ randomly. Observe first that no matter what choice we make for R it is the case that each complementary point of $Q_{x,y}(\tau)$ is within Hamming distance $|R|$ of a complementary point of $Q_{x,y}(\sigma)$. Consider a complementary point w' of $Q_{x,y}(\sigma)$. The analysis of Lemma 1 shows that with probability $1 - O(r/n^4)$ we can find sets X_1, X_2, \dots, X_r all of size at most $4m$ such that zeroising the X_i -components of w' produces a vector which satisfies the i th constraint (here X_i is equal to $J_x \cup J_y$ of the lemma). Hence zeroising all of the $\bigcup_{i=1}^r X_i$ -components produces a member of K . Thus in this case all complementary points of $Q_{x,y}(\tau)$ are within Hamming distance $|R| + 4mr$ of K and the first part follows (modulo a definition of R .)

Our next observation is that if $R' \subseteq R \subseteq J(x, y)$ and all intermediate points in $Q_{x,y}(\tau(x, y, \sigma, R'))$ satisfy the i th constraint then the same is true for $Q_{x,y}(\tau(x, y, \sigma, R))$. This is simply because the intermediate points in the latter path have supports which are subsets of those in the former.

We will now define $R = R_1 \cup R_2 \cup \dots \cup R_r$ where all we claim is that the intermediate points of $Q_{x,y}(\tau(x, y, \sigma, R_i))$ satisfy the i th constraint with probability $1 - O(\log_2 n/n^4)$.

We will now concentrate on R_1 which we decompose as $\bigcup_{j=1}^p R_{1,j}$, where $p \leq \log_2 n$. We will use the notation $a^1(I) = \sum_{j \in I} a_{1,j}$ for $I \subseteq [n]$.

Let v be one of the first $n/2$ points of $Q_{x,y}(\sigma)$. Let $B_x = I_v \cap (I_x \setminus I_y)$ and $B_y = I_v \cap (I_y \setminus I_x)$ so that $I_v = B_x \cup B_y \cup (I_x \cap I_y)$. We can argue as in the proof of Lemma 1 that with probability $1 - O(n^{-4})$ there exist $B''_x \subseteq B_x, B''_y \subseteq B_y$ with $|B_x - B''_x|, |B_y - B''_y| \leq 2m$ and $a^1(I_{v''}) \leq b_1$ where

$I_{v''} = B_x'' \cup B_y'' \cup (I_x \cap I_y)$. If $|I_x \setminus I_y| > 8m$ then L_x is the index set of the $8m$ largest elements in $I_x \setminus I_y$. Otherwise $L_x = I_x \setminus I_y$. Define L_y similarly with respect to $I_y \setminus I_x$. It follows from Theorem 5 (below) that if $|L_x| \geq 8m$ then $|L_x \cap I_v| \geq 2m$ with probability $1 - O(n^{-4})$. (An element of L_x is in I_v with probability at least one half.) Similarly, $|L_y| \geq 8m$ implies $|L_y \cap I_v| \geq 2m$ with probability $1 - O(n^{-4})$. Hence with high probability if \tilde{v} is defined by $I_{\tilde{v}} = I_v \setminus (L_x \cup L_y)$ then $a^1(I_{\tilde{v}}) \leq a^1(I_{v''}) \leq b_1$.

Thus with probability $1 - O(n^{-4})$ the first $n/2$ points of $Q_{x,y}(\tau(x, y, \sigma, L_x \cup L_y))$ satisfy the first inequality. We therefore take $R_{1,1} = L_x \cup L_y$. Let x' denote the $\frac{1}{2}n$ th point of $Q_{x,y}(\tau(x, y, \sigma, L_x \cup L_y))$ and condition on its value. The remaining $n/2$ components are changed by σ in random order in going from x' to y . To define $R_{1,2}$ repeat the above argument with x' in place of x and $n/4$ in place of $n/2$. In this way we obtain a set $R_{1,2}$ such that with probability $1 - O(n^{-4})$, the first $3n/4$ points of $Q_{x,y}(\tau(x, y, \sigma, R_{1,1} \cup R_{1,2}))$ satisfy the inequality. The remaining sets in the partition of R_1 are obtained similarly. We can stop when we have at most $16m$ components to change in order to get to y .

The existence of T^* is now inferred as at the end of the proof of Lemma 1. \square

We can now prove (6) fairly easily. Fix $(v, w) \in E_{S;\bar{S}}$. If $Q_{x,y}$ uses (v, w) then we can fix x, y by fixing the permutation τ and the complementary point. But there are at most $\binom{n}{16mr \log_2 n}$ choices for τ , given σ^* . Hence

$$\lambda_{v,w} \leq n |D_{16mr \log_2 n + 4mr}| \binom{n}{16mr \log_2 n} \quad (14)$$

$$\leq n \binom{n}{16mr \log_2 n + 4mr} N \binom{n}{16mr \log_2 n} \quad (15)$$

$$= O(2^{O(r\sqrt{n}(\log n)^{5/2})} N) \quad (16)$$

and (6) follows immediately.

5 The General Problem

We now consider the general case (1). Observe from (2) that Φ will be very small if d_{\max} is very large. Note also that the common device of replacing a general integer variable x_j by a sum of zero-one variables $\sum_{s=0}^{\ell_j} 2^s z_{sj}$, where ℓ_j is suitably chosen, does not seem to work here for two reasons:

- (a) If $d_j \neq 2^{\ell_j+1} - 1$, then the number of solutions is not preserved (unless we add a further constraint).
- (b) The number of variables n is increased by a factor $\sum_j \log_2(d_j + 1)$. If any d_j is nonpolynomially large in n , this leads to a significant increase in the algorithm's time complexity.

Thus we will adopt a different strategy. We need to be able to deal with the possibility of very large d_j , but first consider the case where $d_{\max} < n^3$. If canonical paths and complementary points are defined exactly as before, then the arguments of Section 4 require only minor modification.

Thus, given $x, y \in K$, consider two zero-one vectors $\xi^{(1)}, \xi^{(2)}$, where $\xi_j^{(1)} = 1$, $\xi_j^{(2)} = 0$ if $x_j > y_j$ and $\xi_j^{(1)} = 0$, $\xi_j^{(2)} = 1$ if $y_j > x_j$. (If $x_j = y_j$, the values of $\xi_j^{(1)}, \xi_j^{(2)}$ can be arbitrary.) Now, the canonical path from x to y can be chosen using that from $\xi^{(1)}$ to $\xi^{(2)}$ in the zero-one polytope defined by the system

$$\sum_{j=1}^n (a_{ij}|x_j - y_j|)\xi_j \leq (b_i - \sum_{j=1}^n a_{ij} \min\{x_j, y_j\}) \quad (i \in [r]),$$

where x, y are to be treated as constant vectors. We simply interpret changing a component from $\xi_j^{(1)}$ to $\xi_j^{(2)}$ as requiring the change of a component from x_j to y_j . We can apply Lemma 2 to this system and the bound on Hamming distance still applies. Now the number of points within Hamming distance β of K is clearly at most $n^{3\beta}$ times the estimate (15). This additional factor does not, however, inflate the estimate (16).

To finish this case we must check that the intermediate points of the path $Q_{x,y}$ so defined are feasible. Consider the k 'th point and for notational convenience

assume that $\tau = (1, 2, \dots, n)$. Then we know that

$$\sum_{i=1}^k ((a_{ij}|x_j - y_j|)\xi_j^{(2)} + \sum_{j=k+1}^n (a_{ij}|x_j - y_j|)\xi_j^{(1)}) \leq (b_i - \sum_{j=1}^n a_{ij} \min\{x_j, y_j\}) \quad (i \in [r]).$$

But

$$\min\{x_j, y_j\} + |x_j - y_j|\xi_j^{(2)} \geq y_j$$

and

$$\min\{x_j, y_j\} + |x_j - y_j|\xi_j^{(1)} \geq x_j.$$

Thus

$$\sum_{i=1}^k a_{ij} y_j + \sum_{j=k+1}^n a_{ij} x_j \leq b_i \quad (i \in [r]).$$

and the path $Q_{x,y}$ is feasible. Thus we may conclude that (6) remains valid when $d_{\max} < n^3$ (and, in fact, for any polynomial bound on d_{\max}).

Thus suppose, for some k ($0 \leq k < n$), that $d_j < n^3$ for $j = 1, 2, \dots, k$, and $d_j \geq n^3$ for $j = k+1, k+2, \dots, n$. Define $\sigma_j = \lceil (d_j + 1)/n^3 \rceil$, and let $x'_j = \lfloor x_j/\sigma_j \rfloor$, $d'_j = \lfloor d_j/\sigma_j \rfloor$ for $j \in [n]$. Also let $a'_{ij} = \sigma_j a_{ij}$ for $i \in [r]$, $j \in [n]$. We can generate a point (near) uniformly in $K' = \{x' \in \mathbb{Z}^n : A'x' \leq b, 0 \leq x' \leq d'\}$ by the above method, since

$$d'_j \leq d_j/\sigma_j \leq n^3 d_j/(d_j + 1) < n^3.$$

After generating such an x' we then choose x by

$$x_j = \sigma_j x'_j + Z_j, \quad (j \in [n]), \tag{17}$$

where Z_j is an integer chosen uniformly from $[0, \sigma_j - 1]$. We accept the point x if it lies in K , otherwise we try again. Now (17) determines a bijection between the integer x 's in some set H and the pairs (x', Z) . The generated x 's are evidently uniform on H , a union of hyper-rectangles, one corresponding to each x' . Note that $K \subseteq H$ since, for each $x \in K$, there is a unique pair (x', Z) with $x' \in K'$ given by $x'_j = \lfloor x_j/\sigma_j \rfloor$, $Z_j = x_j \bmod \sigma_j$, $j \in [n]$. The accepted points will therefore be (near) uniform in K , as required. However, the generated x may be rejected, so we must check that acceptance occurs with sufficiently high probability. It is clearly enough that the acceptance probability be large in comparison with the estimate of conductance (3). In

that case the number of repetitions required for acceptance will not significantly affect the total time to generate a (near) uniform point in K .

Suppose we have some “good” set $G \subseteq K'$ such that

- (i) $\Pr(x \in K \mid x' \in G) \geq \frac{1}{2}$,
- (ii) $\Pr(x' \in G) \geq n^{-4r}$.

Then it is easy to see that

$$\Pr(x \in K) \geq \Pr(x \in K \mid x' \in G) \Pr(x' \in G) \geq \frac{1}{2} n^{-4r},$$

which is large compared to (3) and we are done.

Thus it remains only to define the set G and prove that it has properties (i) and (ii). To this end, for $y \in \mathbb{R}^n$, let $y = (\hat{y}, \bar{y})$, where $\hat{y} \in \mathbb{R}^k$, $\bar{y} \in \mathbb{R}^{n-k}$, and k is as defined above. We will use the same notation for the corresponding partition of a matrix by its first k and last $n - k$ columns. Note that $\hat{x} = \hat{x}'$, $\hat{A} = \hat{A}'$, since $\sigma_j = 1$ for $j \leq k$. Now define

$$G = \{x' \in K' : \hat{A}\hat{x} \leq \frac{k}{k+1}b\} \text{ and } \hat{G} = \{x' : x' \in G\}.$$

Lemma 3 $\Pr(x \in K \mid x' \in G) \geq \frac{1}{2}$.

Proof For $x \in H$ let

$$D(x) = \{y \in \mathbb{R}^{n-k} : x_{k+j}/\sigma_{k+j} \leq y_j < (x_{k+j} + 1)/\sigma_{k+j}, (j \in [n - k])\}.$$

Fix $\hat{g} \in \hat{G}$. Clearly $(\hat{g}, \xi) \in K'$ if and only if ξ lies in the polytope

$$\bar{P} = \{\xi \in \mathbb{R}^{n-k} : \bar{A}'\xi \leq b - \hat{A}\hat{g} = b^*, 0 \leq \xi \leq \bar{d}\},$$

where $b^* \geq \frac{1}{k+1}b$. Now

$$\Pr(x \in K \mid \hat{x}' = \hat{g}) = \frac{\text{vol}(H')}{\text{vol}(H'')}, \quad (18)$$

where

$$H' = \bigcup_{\substack{x \in K \\ \hat{x}' = \hat{g}}} D(x) \text{ and } H'' = \bigcup_{\substack{x \in H \\ \hat{x}' = \hat{g}}} D(x).$$

Now $H' \supseteq \bar{P}$ since if $\xi \in \bar{P}$ then $\xi \in D(x)$ where $\hat{x} = \hat{g}$, $x_{j+k} = \lfloor \sigma_{j+k} \xi_{j+k} \rfloor$ for $j \in [n-k]$ and x is a member of K .

On the other hand the assumptions on d_j imply

$$a'_{ij} \leq a_{ij}(2d_j/n^3) \leq 2b_i/n^3 \leq 2(k+1)b_i^*/n^3. \quad (19)$$

But then $z \in H''$ implies

$$\bar{A}'z \leq b^* + \bar{A}'\mathbf{1} \leq (1 + \frac{2(k+1)(n-k)}{n^3})b^*,$$

where $\mathbf{1}$ is the vector of all 1's, using (19). So

$$H'' \subseteq \left(1 + \frac{2(k+1)(n-k)}{n^3}\right) \bar{P} \quad (20)$$

Hence

$$\Pr(x \in K \mid \hat{x}' = \hat{g}) \geq (1 + \frac{2(k+1)(n-k)}{n^3})^{-(n-k)} \geq \exp\{-2(k+1)(n-k)^2/n^3\} \geq \frac{1}{2},$$

for $n \geq 4$, completing the proof. \square

Lemma 4 $\Pr(x' \in G) \geq n^{-4r}$.

Proof If $x' \in K' \setminus G$, consider the rows $i \in [r]$ such that $(\hat{A}\hat{x})_i > \frac{k}{k+1}b_i$. For each such i , there exists a smallest $t = t(i)$ ($1 \leq t \leq k$) such that $a_{it}x_t > b_i/(k+1)$. Thus define a function $g : K' \rightarrow G$ by

$$g_j(x') = \begin{cases} x'_j & \text{if } j \neq t(i) \text{ for any } i, \\ 0 & \text{otherwise} \end{cases}$$

Clearly $g(x') \in G$, as required. Since $g(x')$ differs from x' in at most $\min\{r, k\}$ coordinates, the point-to-set map g^{-1} clearly partitions K' into classes of size at most

$$\binom{n}{r} n^{3\min\{r, k\}} \leq n^{4r},$$

since $x' < n^3\mathbf{1}$. Hence $|K'| \leq n^{4r}|G|$ and

$$\Pr(x' \in G) = |G|/|K'| \geq n^{-4r},$$

as claimed. \square

6 A polynomially solvable case

In this section we prove a stronger result in a special case. If $d_j \geq n^2, j = 1, 2, \dots, n$, then we can construct a *fully polynomial* randomized approximation scheme for $|K|$ i.e. our algorithm is polynomial in n, ϵ^{-1} . Thus such cases are “easy”, the reason being that the counting closely resembles volume computation [4].

We can use the scaling of the previous section to “round” the feasible polytope K to K' . Then we have $B(n/2) \subseteq K' \subset B(n^3)$, where $B(\nu) = \{0 \leq x' \leq \nu \mathbf{1}\}$. The first inclusion follows from $a_{ij}\sigma_j \leq 2a_{ij}d_j/n^2 \leq 2b_i/n^2$ for $j \in [n]$, while the second is a restatement of $d'_j < n^3$. We perform a *biased* random walk on the integer lattice points $x' \in B(n^3)$ with the weighting function $T(x') = e^{-2nt(x')}$, where

$$t(x') = \min\{t \geq 0 : A'x' \leq (1+t)b\}.$$

At a point x' we do nothing with probability $1/2$. Otherwise we randomly choose $\pm j$ from $\{\pm 1, \pm 2, \dots, \pm n\}$. We then let $y' = x' \pm e_j$, where e_j denotes the j 'th unit vector. If $y' \in B(n^3)$ we move there with probability $\min\{1, T(y')/T(x')\}$. This completes a step of the walk.

It is easily shown that t varies by only a constant factor over any unit cube in $B(n^3)$, using the rounding property. Indeed if $A'x' \leq (1+t)b$ then

$$A'(x' + \mathbf{1}) \leq \left(1 + t + \frac{2}{n}\right) b. \quad (21)$$

Walks of this type were studied in Applegate and Kamman [1] (see also Dyer and Frieze [3] or Lovász and Simonovits [13] for some improvements.) The function T is log-concave and $\log T$ has a Lipschitz constant of $O(n)$. It follows that the walk mixes rapidly i.e. the conductance is $1/p(n)$ for some polynomial p and we can obtain a near uniform random point in K' in polynomial time.

A point x is generated from a random $x' \in K'$ as in the previous section - see (17). The rounding property ensures a probability of at least e^{-2} that $x \in K$.

$$\Pr(x \in K) \geq \text{vol}(K')/\text{vol}(K' + B(1))$$

$$\geq \left(1 + \frac{2}{n}\right)^{-n},$$

by (21). We can therefore sample nearly uniformly from K in polynomial time and we obtain a fully polynomial time approximation scheme as stated.

7 Modification of a Theorem of Hoeffding

Let a population C consist of n real values $c_1 \geq c_2 \geq \dots \geq c_n \geq 0$. Let $c_{i_1}, c_{i_2}, \dots, c_{i_t}, i_1 > i_2 > \dots > i_t$ denote a random sample drawn without replacement from C . Let $S_k = c_{i_1} + c_{i_2} + \dots + c_{i_k}$ for $k = 1, 2, \dots, t$. Let $\mu = \frac{1}{n} \sum_{i=1}^n c_i$. Hoeffding proved

Theorem 5 *If $0 \leq c_i \leq \alpha$ for $i = 1, 2, \dots, n$ then*

$$\Pr(S_t - t\mu > x) \leq \exp\left\{-\frac{2x^2}{t\alpha^2}\right\} \quad \text{for all } x \geq 0.$$

□

This is not suitable for us because we do not have enough control over the size of α . We prove the following theorem which we believe is of interest in its own right.

Theorem 6 *With the above notation*

$$\Pr(S_t - t\mu > S_m) = O(\exp\{-m^2(t-m)^2/t^3\}) \quad \text{for } t = m+1, \dots, n.$$

Proof Let $T_j = c_1 + c_2 + \dots + c_j$ for $j = 1, \dots, n$ and $c = c_{i_m}$. If $S_t > t\mu + S_m$ then for $x > 0$ (to be chosen later) either

$$(i) \ S_t - S_m > (t - m) \frac{T_n - T_{i_m}}{n - i_m} + cx,$$

or

$$(ii) \ (t - m) \frac{T_n - T_{i_m}}{n - i_m} + cx > t\mu.$$

We handle (i) by conditioning on $c_{i_1}, c_{i_2}, \dots, c_{i_m}$ and applying Theorem 5. This yields

$$\Pr((i)) \leq e^{-2x^2/(t-m)}. \quad (22)$$

To handle (ii) we define δ by $n - i_m = \frac{n}{t}(t - m)(1 - \delta)^{-1}$ and replace (ii) by the inequality

$$(ii') \quad cx - \frac{T_{i_m}t}{n}(1 - \delta) > t\mu\delta.$$

Observe that

$$T_{i_m}/i_m \geq \max\{\mu, c\}. \quad (23)$$

At this point we bound $|\delta|$ probabilistically. Now i_m is distributed as the m th largest index of a random t -set X chosen from $[n]$. Thus $\bar{i}_m = E(i_m) = nm/t$. Let $\theta = \bar{i}_m - i_m$ so that $\delta = \theta/(n - \bar{i}_m + \theta)$ and $\theta = \delta n \left(1 - \frac{m}{t}\right) (1 - \delta)^{-1}$. Now let $Z_k = |X \cap [k]|$. We observe that for any $y > 0$

$$\theta > y \Rightarrow Z_{[\bar{i}_m - y]} \geq m$$

and

$$\theta < -y \Rightarrow Z_{[\bar{i}_m + y]} \leq m.$$

But for any k , Z_k is distributed as the random variable S_t of Theorem 5, where C consists of k 1's and $n - k$ 0's. Thus $E(Z_k) = kt/n$ and applying the theorem gives

$$\Pr(|\theta| \geq y) \leq 2e^{-2y^2t/n^2}. \quad (24)$$

Case 1: $\delta < 0$.

It follows from (23) that

$$\begin{aligned} (ii') &\Rightarrow \frac{T_{i_m}}{i_m}x - \frac{T_{i_m}t}{n}(1 - \delta) \geq t\frac{T_{i_m}}{i_m}\delta, \\ &\Rightarrow x - \frac{i_mt}{n}(1 - \delta) \geq t\delta. \end{aligned}$$

Substituting $i_m = \frac{n(m-t\delta)}{t(1-\delta)}$ we obtain

$$(ii') \Rightarrow x \geq m.$$

We take $x = m/2$ from now on, and so

$$\Pr((ii) \text{ and } \delta < 0) = 0. \quad (25)$$

Case 2: $\delta \geq 0$.

Applying (23) we see now that (with $x = m/2$)

$$\begin{aligned}
(\text{ii}') \Rightarrow \frac{m}{2} - \frac{im}{n}t(1 - \delta) &\geq 0 \\
\Rightarrow \frac{m}{2} - i_m \frac{t-m}{n-i_m} &\geq 0 \\
\Rightarrow i_m &\leq \frac{nm}{2t-m} \\
\Rightarrow \theta &> \frac{mn}{2t} \left(1 - \frac{m}{t}\right).
\end{aligned}$$

Hence, from (24) we have

$$\Pr((\text{ii}) \text{ and } \delta \geq 0) = O(\exp\{-m^2(t-m)^2/t^3\}). \quad (26)$$

The theorem follows from (22), (25) and (26). \square

References

- [1] D. Applegate and R. Kannan, *Sampling and integration of near log-concave functions* Proc. 23rd ACM Symposium on Theory of Computing, 1991, pp. 156–163.
- [2] A. Z. Broder, *How hard is it to marry at random? (On the approximation of the permanent)*, Proceedings of the 18th Annual ACM Symposium on Theory of Computing (1986), 50–58. Erratum in Proceedings of the 20th Annual ACM Symposium on Theory of Computing (1988), 551.
- [3] M. E. Dyer and A. M. Frieze, *Computing the volume of convex bodies: a case where randomness provably helps*, in *Probabilistic Combinatorics and its Applications* (B. Bollobás, Ed.), AMS Proceedings of Symposia in Applied Mathematics 44 (1991), 123–169.
- [4] M. E. Dyer, A. M. Frieze and R. Kannan, *A random polynomial time algorithm for approximating the volume of convex bodies*, Journal of the Association for Computing Machinery 38 (1991), 1–17.

- [5] W. Hoeffding, *Probability inequalities for sums of bounded random variables*. Journal of the American Statistical Association 58 (1963), 13–30.
- [6] M. R. Jerrum and A. J. Sinclair, *Approximating the permanent*, SIAM Journal on Computing 18 (1989), 1149–1178.
- [7] M. R. Jerrum and A. J. Sinclair, *Polynomial-time approximation algorithms for the Ising model*, Department of Computer Science, Edinburgh University, 1989, to appear in SIAM Journal on Computing.
- [8] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani, *Random generation of combinatorial structures from a uniform distribution*. Theoretical Computer Science, 43 (1986), 169–188.
- [9] M. R. Jerrum and U. Vazirani, *A mildly exponential approximation algorithm for the permanent*, Proceedings of the 33rd Annual IEEE Symposium on Foundations of Computing (1992) 320–326.
- [10] A. Karzanov and L. Khachiyan, *On the conductance of order Markov chains*, Technical Report DCS TR 268, Rutgers University (1990).
- [11] L. Lovász, *Combinatorial problems and exercises*, North Holland, Amsterdam, 1979.
- [12] L. Lovász and M. Simonovits, *The mixing rate of Markov chains, an isoperimetric inequality and computing the volume*, Proceedings of the 31st Annual IEEE Symposium on Foundations of Computer Science (1990), 346–354.
- [13] L. Lovász and M. Simonovits, *Random walks in a convex body and an improved volume algorithm* Proceedings of the 33rd Annual IEEE Symposium on Foundations of Computer Science (1992).
- [14] M. Mihail and P. Winkler, *On the number of Euler orientations of a graph*, Proceedings of the 3rd Annual ACM-SIAM Symposium on Discrete Algorithms (1992), 138–145.
- [15] A. J. Sinclair and M. R. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Information and Computation 82 (1989), 93–133.