A Randomized Algorithm to Optimize over Certain Convex Sets

Ravi Kannan*
School of Computer Science
Carnegie Mellon University
kannan+@cs.cmu.edu

John Mount†
School of Computer Science
Carnegie Mellon University
jmount+@cs.cmu.edu

Sridhar Tayur
Graduate School of Industrial Administration
Carnegie Mellon University
tayur@kanban.gsia.cmu.edu

November 3, 1994

Abstract

This paper presents a randomized polynomial time algorithm to nearly minimize a linear function over an up-monotone convex set in the positive orthant given only by a membership oracle. Our original motivation for this is a stochastic optimization problem called the component commonality problem in the literature.

1 INTRODUCTION.

This paper presents a randomized polynomial time algorithm to nearly minimize a linear function over an up-monotone convex set (i.e., a convex set with the property that if x belongs to the set and y is component wise greater than or equal to x, then y belongs to the set too) in the positive orthant given only by a membership oracle (i.e., an oracle which given a point x in space, returns the correct answer to the question of whether x is in the convex set). We also assume that we are given some point x^f in the convex set to start with. Our original motivation for this is a stochastic optimization problem called the component commonality problem [9]. Our randomized algorithm is based on a random walk. While similar algorithms are used for other optimization problems, for example, in simulated annealing, this seems to be the first provably polynomial time algorithm to achieve near-optimality with high probability.

The Component Commonality (CC) problem in discrete time arises as follows. There are m products (indexed by j) with correlated random demands ($(d^1, \ldots, d^m)$ with distribution $H(\cdot)$

*Ravi's NSF grant.
†Ravi's NSF grant.

and density $h(\cdot)$) in any period, that require some or all of n components (indexed by i) in the ratio u_{ij}. The key feature in this setting is that as many of the n components are shared by many products, and as the assembly time is small relative to the procurement time of the components, the assembly is done after the demand has realized, thus taking advantage of commonality of components. The CC problem, then, is to find the quantities q_i of each component (at unit cost $c_i > 0$) that should be on stock at the beginning of each period so as to minimize total costs of purchase while satisfying the demands with probability at least a given fraction γ. Formally, for a stock level $q = (q_1, q_2, \dots q_n)$, define $F(q) = \{d : \sum_{j=1}^{m} u_{ij} d^j \leq q_i \text{ for } i = 1, 2 \dots n\}$ to be the set of demands that can be met with q on stock. Let

$$g(q) = \int_{F(q)} h(d)$$

be the probability that we can meet the demands with stock level q. Then the problem is :

$$\text{Minimize } \sum_{i=1}^{n} c_i q_i \ : \ g(q) \geq \gamma.$$

With some general assumption on the density $h(\cdot)$, (for example quasi-concavity or log-concavity, cf section 3) it can be seen that the feasible set of stock levels (i.e., $\{q : g(q) \geq \gamma\}$) is convex.

The CC problem falls into a class of problems called probabilistic constrained programming (PCP), a type of stochastic program. Our methods apply to other problems as well in this class; we have given only the example of the CC problem here. Previous efforts to solve this class of problems have been by non-linear programming (NLP) methods. A standard non-linear programming approach would involve as subroutines : (i) checking if the constraint $g(q) \geq \gamma$ is valid and (ii) estimating the gradient of $g(q)$. Theoretically, we may also apply algorithms like the ellipsoid algorithm. In this case we need a "separation subroutine" which given an x says whether it is feasible and if not returns a hyperplane separating x from the feasible set. It can be seen that the task of this subroutine is essentially (ii).

The first task (i) may be solved by sampling, if we assume that we can draw samples according to the density $h(\cdot)$; in practice, γ is likely to be close to 1 (in any case at least 1/2) and so the number of samples required is not enormous. However, the second task is considerably more onerous since to compute each of the n directional derivatives, we need very accurate values of $g(q)$. Because of these and other difficulties, there are no provable polynomial run-time bounds known for gradient descent based algorithms. (Wets [13] discusses these and other related difficulties in greater detail.)

Using well-known methods from Stochastic Optimization, it will be easy to see that the CC problem can be formulated as the problem of minimizing a linear function over an up-monotone convex set given by a membership oracle. This is done in section 3. To check whether a particular q is a feasible stock level, one needs only to answer a question of the form (i), so a membership oracle is easily available. [This property that questions of the form (i), i.e., membership queries, are easier to answer than queries of the form (ii) is shared by several Stochastic Optimization problems. This is the reason for the interest in this model.]

2 ALGORITHM OUTLINE AND TIME BOUNDS.

Using the membership oracle, we approximately minimize a linear function over an up-monotone convex feasible set int the positive orthant as follows. We may assume a suitable upper bound an the variables so we can enclose this feasible region in a rectangle (in n dimensions). At the heart of our approach is a positive real-valued logarithmically concave function F on the rectangle with the following properties: (1) the integral of F over a region consisting of near-optimal solutions is at least a constant fraction of the integral of F over the whole feasible set, and (2) the integral of F over the feasible set is at least a constant fraction of the integral of F over the entire rectangle. Thus, if we pick a "random sample" from the rectangle with probability density proportional to F (we refer to this throughout as "sampling according to F"), we would get a near-optimal solution with constant probability; this probability can be boosted by repeated sampling. Our algorithm, then, is simply a choice of F (determined by two parameters α, a damping factor favoring feasible points, and β, a bias favoring points with better objective values) and a method to obtain a sample according to F. We show that a certain biased random walk (on the uniform grid of size δ, to be determined), starting from a feasible solution (x^f), is indeed able to pick a random sample from the feasible set with probability (approximately) proportional to F. While it is relatively easy to argue that in the steady state, this random walk picks a sample with density proportional to F, it is nontrivial to show that this steady state is approached in a polynomial number of steps. To accomplish this central result, we draw on recently developed results in the theory of rapidly mixing Markov Chains as well as on random walks in convex sets [5] , [1]. The latter paper gives a technique for sampling from log-concave distributions which we use here, although, we have tried to make this paper self-contained by giving as many details as possible. Our random walk can be executed with only local knowledge of F as well as a membership (not a separation) oracle for the feasible set.

Given an instance of the problem (a membership oracle for K and objective function $c > 0$), $\epsilon > 0$ (relative error), $\kappa > 0$ (failure probability) and an initial feasible point x^f, the algorithm succeeds with probability at least $1 - \kappa$ in finding a $q^{alg} \in \mathcal{R}^n$ which is feasible and such that

$c \cdot q^{alg} \le (1+\epsilon)(c \cdot q^{opt})$. Rather than come within ϵ of the optimal in one long random walk, we develop an adaptive algorithm which improves the feasible solution in stages (by iteratively refining the gap between a known feasible solution and a probabilistic pointwise lower bound lower bound L on optimal cost).

Each stage begins with a feasible solution x^f and a probabilistic "lower bound" L on hand where if there is a feasible x with $c \cdot x < L$, then the algorithm has failed. (We will of course ensure that the probability of failure is low.) We refer to $c \cdot x^f - L$ as the "gap" (at the beginning of the stage). At the end of the stage, we have a new lower bound and a new feasible solution; we ensure that the gap at the end of a stage is at most 1/2 of the gap at the beginning. We stop when the gap is at most ϵ times the value of the cost lower bound. The algorithm is described in detail in figure 1 and justified in section 7, but we give here a short verbal description.

Each stage proceeds as follows : the feasible set is enclosed in a rectangular solid. We devise a log-concave function F on the rectangle with the two properties described above. [The function F has two components : a "penalty" (called the gauge function in what follows) for going out of the feasible set which increases as we become "more infeasible" and a bias (drift) which favors low objective function value. In some vague sense, this is similar to Lagrangian relaxation with both feasibility and optimality represented by one function.]

Then we discretize by dividing the rectangle into small cubes. We perform a random walk on the cubes with transition probabilities depending on F. It will be easy to see that the steady state probabilities of this random walk will be proportional to F. We will also show fast convergence to the steady state, so that after a polynomial number of steps, we are "close" to the steady state probabilities.

After doing the random walk for this number of steps, one of the following two scenarios occurs :

(i) We have found a feasible solution whose value cuts down the gap by a factor of at least 1/2. In this case, we replace our old feasible solution by this and go to the next stage.

(ii) Otherwise, we have (probabilistic) proof of a greater lower bound and we go to the next stage with this new lower bound (again cutting the gap down by a factor of 1/2).

Although F has been devised accurately to have the desired properties, several errors are introduced in the sampling procedure which are tackled in the paper. There are errors due to discretizing into small cubes, due to the inexact computation of the gauge function and due to the fact that the lower bounds are only probabilistic. The management of these errors is the main focus of section 5.

Our main result is two bounds on the running time of the algorithm. [The running time is bounded above by the minimum of the two.]

If ν is the ratio of the value of the given initial feasible solution to the optimal value, ϵ is the required relative error, $1-\kappa$ the required success probability and n, the dimension of the up-monotone convex feasible set K, the first bound is

$$O\left(\frac{n^7\left(\log\left(\frac{n}{\epsilon}\right)\right)^2\log\left(\frac{\log\left(\frac{1}{\epsilon}\right)}{\kappa}\right)\log\left(\frac{\nu}{\epsilon}\right)}{\epsilon^2}\right).$$

If in addition, we are given $x^l > 0$ such that $\forall y \in K$, we have $x^l \leq y$ and an x^u such that there is an optimal solution z with $z \leq x^u$ (so we can replace K by $K \cap \{x : x \leq x^u\}$), then we also have a bound

$$O\left(n^5\left(\frac{\left\|x^u - x^l\right\|_\infty}{\epsilon \min_i x_i^l}\right)^2\left(\ln\left(\frac{n\left\|x^u - x^l\right\|_\infty}{\epsilon \min_i x_i^l}\right)\right)^2\ln\left(\frac{1}{\kappa}\right)\right).$$

The rest of the paper is organized as follows. Section 3 notes that the CC problem falls into the framework of up-monotone convex sets and has some general remarks. Section 4 constructs the appropriate log concave and gauge functions that are to be used in each stage of the algorithm. Section 5 describes the random walk to be performed at any stage of the adaptive algorithm, and contains the analysis of the errors introduced due to discretization, gauge function approximation and walking for finite number of steps. In section 6, we find a bound on the spectral gap of the Markov chain, which allows us to use results on rapid mixing to provide a bound on the number of steps required in any stage. In section 7, we prove the adaptive algorithm's correctness and run time, and present two variants and prove their run times. Proofs for certain lemmas have been moved to the appendix.

3 THE COMPONENT COMMONALITY PROBLEM.

The problem was described in detail in the Introduction. If U denotes the matrix of the u_{ij} 's there, let $y = y(d) = Ud$. Under the assumption that $h(\cdot)$ is log-concave, it is easy to see that y has a log-concave density. Let D denote the density of the y. Let μ_D denote the corresponding measure. (So for any measurable set S, $\mu_D(S) = \int_S D$.) Then the feasible set K of stock-levels can be expressed as

$$K = \{x \in \mathcal{R}^n : \mu_D(\mathrm{dom}(x)) \geq \gamma\}$$

where $\mathrm{dom}(x)$ is $\{y : y \leq x\}$. It is easy to show that K is convex ([13]). It is also clearly up-monotone.

3.1 General Remarks

While we do assume that a membership oracle is available for K, we do not assume that a separation oracle is available. By a theorem of Yudin and Nemirovskiĭ it is known that membership and separation are polynomial time equivalent for convex programming problems like this one [7]. But the conversion has a large exponent, and the approach here is more efficient.

Instead of computing the integral each time the membership oracle is called, we could instead pick m samples $y^1, y^2, \ldots y^m$ according to D at the outset and then for each query x to the membership oracle for K, say yes to the query if and only if for at least γm of the y^i, we have $x \geq y^i$. It is easy to show by standard techniques that for m large enough, this suffices as a an approximate test of membership in K. We do not go into the details here. Also, in the actual situation, either the y^i may be available from past data or if one hypothesizes a particular log-concave density D, we may draw samples according to the density using the techniques of [1].

One may be tempted to solve the discrete problem : given m points $y^1, y^2, \ldots y^m$ in $\mathcal{R}^n$, a fraction γ, a positive vector c and a real number ζ, does there exist an $x \in \mathcal{R}^n$ such that $x \geq y^i$ is satisfied for at least γm different i 's and we also have $c \cdot x \leq \zeta$? In this generality, we show (in the appendix) that the problem is NP-hard. So one needs to exploit the special nature of K, namely its convexity.

For an NLP approach to the CC problem, see [9] and references provided there.

4 THE FUNCTION F: BIAS AND DAMPING.

Let K be an up-monotone convex set contained in the positive orthant of $\mathcal{R}^n$, x^f a known feasible point and $c > 0$ the cost vector in the linear objective function. For real numbers $L \geq 0, T > 0$, we define a log-concave distribution $B_{(L,T)}$ on $\mathcal{R}^n$ such that if $L \leq \inf\{c \cdot y \mid y \in K\}$ then $\mu_{B_{(L,T)}}(\{y \in K \mid c \cdot y \leq \inf\{c \cdot y \mid y \in K\} + T\}) / \mu_{B_{(L,T)}}(\mathcal{R}^n) \geq \frac{1}{4}$ and show how to use the membership oracle for K to approximately sample from this distribution in an efficient manner.

Let x^l be such that $\forall y \in K,\ x^l \leq y$ and let x^u be such that $\forall y \in K,\ \exists z \in K$ such that $c \cdot z \leq c \cdot y$ and $z \leq x^u$. Note that by the up-monotone property, $x^u \in K$. If x^l and x^u are not explicitly given we can take $x^l = 0$ and x^u such that $x^u_i = (1/c_i) \sum_{i=1}^n c_i x^f_i$. Let $K_L = K \cap \{y \mid c \cdot y \geq L\}$ (we will later further restrict our attention to a "rectangle" that is roughly $\left\{x \;\middle|\; x^l \leq x \leq x^u\right\}$).

Definition 1 *For any real $L \geq 0$, $T > 0$ let "the tip" be the set $x \in K_L$ such that $c \cdot x \leq \inf\{c \cdot y \mid y \in K_L\} + T$.*

Let $\psi_{(K_L,x^u)}(x)$ denote the infimum of all real positive numbers λ such that $x^u + \frac{x-x^u}{\lambda} \in K_L$. This is the dilation of K_L about x^u needed to contain x and is called the *gauge function* associated with K_L. When the context is clear we will suppress the subscripts and use ψ instead of $\psi_{(K_L,x^u)}$.

F will be of the form

$$F(x) = e^{-\alpha \max(\psi_{(K_L,x^u)}(x)-1,0)} e^{-\beta c \cdot x} \tag{1}$$

where β and α are positive reals to be determined later. We take $F(x)$ as the unnormalized density function for $B_{(L,T)}$, a log-concave distribution on $\{x \in \mathcal{R}^n \mid x \leq x^u\}$. Note that $e^{-\beta c\cdot x}$ is the bias (in favoring better objective values), and $e^{-\alpha \max(\psi_{(K_L,x^u)}(x)-1,0)}$ is the function to damp this bias in regions that are infeasible.

The selection of α and β is done in two stages: (1) we first find β such that at least half of the probability mass in the body (according to the density $B_{(L,T)}$) is in "the tip", and then (2) find α such that at least half the mass in the entire rectangle is in the body.

4.1 Getting in the tip.

For x in K_L, the distribution F is a function of the $c \cdot x$ only (since $\psi(x) \leq 1$).

Lemma 1 *If $L \geq 0$,$T > 0$ and $\beta \geq \frac{n}{T}$, we have*

$$\int_{\text{"the tip"}} F \geq (1/2) \int_{K_L} F.$$

Proof Let $c^* = \inf\{c \cdot y \mid y \in K_L\}$, $z \in K_L$ such that $c \cdot z = c^*$ and A^* be the intersection of the hyperplane $c \cdot x = c^* + T$ and K. Clearly, the convexity of K_L implies that $\int_{K_L \cap \{x: c\cdot x \leq c^*+T\}} F$ is not increased if we replace $K_L \cap \{x : c\cdot x \leq c^* + T\}$ by the convex hull of z and A^*. Also, replacing $K_L \cap \{x : c \cdot x \geq c^* + T\}$ by the truncated cone formed by intersecting $\{x : c \cdot x \geq c^* + T\}$ with the minimal pointed cone with vertex z containing A^* cannot decrease $\int F$ over this set. So it suffices to prove the lemma with K_L equal to the infinite cone. Then the ratio of integrals in the lemma is

$$\frac{\int_{c^*}^{c^*+T} \left(\frac{\lambda - c^*}{T}\right)^{n-1} \mathrm{area}(\mathrm{A}^*)\, e^{-\beta\lambda}\, \mathrm{d}\,\lambda}{\int_{c^*}^{\infty} \left(\frac{\lambda - c^*}{T}\right)^{n-1} \mathrm{area}(\mathrm{A}^*)\, e^{-\beta\lambda}\, \mathrm{d}\,\lambda} \quad = \quad \frac{\int_{c^*}^{c^*+T} (\lambda - c^*)^{n-1}\, e^{-\beta\lambda}\, \mathrm{d}\,\lambda}{\int_{c^*}^{\infty} (\lambda - c^*)^{n-1}\, e^{-\beta\lambda}\, \mathrm{d}\,\lambda}.$$

We change variables (and consult standard integral tables) to get:

$$\frac{\int_0^T \lambda^{n-1} e^{-\beta\lambda}\, \mathrm{d}\,\lambda}{\int_0^\infty \lambda^{n-1} e^{-\beta\lambda}\, \mathrm{d}\,\lambda} \quad = \quad 1 - e^{-\beta T} \sum_{k=0}^{n-1} \frac{(\beta T)^k}{k!}.$$

By picking $\beta = \frac{n}{T}$ the ratio is $1 - e^{-n} \sum_{k=0}^{n-1} \frac{n^k}{k!}$, which is $\geq \frac{1}{2}$ for $n \geq 1$, so any

$$\beta \geq \frac{n}{T}, \tag{2}$$

will do.

$\square$

4.2 Staying in the body.

We now show that at least $\frac{1}{2}$ of the mass is in the body K_L, which would imply that at least $\frac{1}{4}$ of the mass of $B_{(L,T)}$ is in the tip (the near optimal feasible region), for a suitable choice of α.

Let ∂K_L be the set of y in the boundary of K_L. For any $\zeta > 0$ all of K_L can be covered by a collection C of disjoint cones such that for each cone $r \in C$, we have $\|x - y\|_2 \leq \zeta$ for $x, y \in r \cap \partial K_L$.

Now, if less than half of the mass according to $B_{(L,T)}$ is in K_L then there must be a cone $r \in C$ such that less than half the mass according to $B_{(L,T)}$ restricted to r is in $r \cap K_L$. By the arbitrary nature of ζ we see the same must hold for some "infinitely" thin cone. Chose such a cone and let y be the point at which the cone intersects ∂K and set $\lambda_0 = \|x^u - y\|_2$.

Lemma 2 *Suppose*

$$\alpha \geq \max\Big(3\beta(c \cdot x^u - L) + 3n - 5, n(e^2+1) + 1\Big) \tag{3}$$

Then the mass of any infinitesimal cone outside of K_L can be shown to be no more than the mass of the same cone inside K_L , thus yielding a ratio of feasible to total of at least $\frac{1}{2}$.

Proof For the proof of this lemma only, it will be convenient to multiply masses by $e^{\beta c \cdot x^u}$; so for any set S, we mean by the mass of S, the quantity $e^{\beta c \cdot x^u} \int_S F$. The mass of the cone outside K_L is given by:

$$\begin{aligned}
& \int_{\lambda_0}^{\infty} \lambda^{n-1} e^{\beta(c \cdot x^u - c \cdot y)\frac{\lambda}{\lambda_0} - \alpha(\frac{\lambda}{\lambda_0} - 1)} \, \mathrm{d}\,\lambda \\
= \;& \lambda_0^n \int_1^{\infty} t^{n-1} e^{\beta(c \cdot x^u - c \cdot y)t - \alpha(t-1)} \, \mathrm{d}\,t \\
\leq \;& \lambda_0^n \int_1^{\infty} e^{(t-1)(n-1)} e^{\beta(c \cdot x^u - c \cdot y)t - \alpha(t-1)} \, \mathrm{d}\,t \\
= \;& \lambda_0^n e^{\beta(c \cdot x^u - c \cdot y)} / (\alpha - \beta(c \cdot x^u - c \cdot y) - (n-1)).
\end{aligned}$$

For the mass of the ray inside K_L we will break into two cases depending if $\beta(c \cdot x^u - c \cdot y)$ is ≥ 2 or not. The mass of the ray inside K_L is at least

$$
\begin{aligned}
& \int_0^{\lambda_0} \lambda^{n-1} e^{\beta \frac{\lambda}{\lambda_0}(c \cdot x^u - c \cdot y)} \, \mathrm{d}\, \lambda \\
= \;& \lambda_0^n \int_0^1 t^{n-1} e^{\beta t(c \cdot x^u - c \cdot y)} \, \mathrm{d}\, t.
\end{aligned}
$$

We now consider the two cases.

Case 1: $\beta(c \cdot x^u - c \cdot y) \geq 2$: An integration by parts gives the mass of the ray inside K_L equals

$$
\begin{aligned}
& \frac{\lambda_0^n}{\beta(c \cdot x^u - c \cdot y)} (e^{\beta(c \cdot x^u - c \cdot y)} - (n-1) \int_0^1 t^{n-2} e^{\beta t(c \cdot x^u - c \cdot y)} \, \mathrm{d}\, t) \\
\geq \;& \frac{\lambda_0^n}{\beta(c \cdot x^u - c \cdot y)} (e^{\beta(c \cdot x^u - c \cdot y)} - (n-1) \int_0^1 e^{(t-1)(n-2)} e^{\beta t(c \cdot x^u - c \cdot y)} \, \mathrm{d}\, t) \\
\geq \;& \frac{\lambda_0^n e^{\beta(c \cdot x^u - c \cdot y)}}{\beta(c \cdot x^u - c \cdot y)} (1 - \frac{(n-1)}{\beta(c \cdot x^u - c \cdot y) + n - 2}).
\end{aligned}
$$

The ratio of mass inside K_L to outside is at least

$$
\frac{\beta(c \cdot x^u - c \cdot y) - 1}{\beta(c \cdot x^u - c \cdot y) + n - 2} \frac{\alpha - \beta(c \cdot x^u - c \cdot y) - n + 1}{\beta(c \cdot x^u - c \cdot y)}.
$$

Since $\beta(c \cdot x^u - c \cdot y) \geq 2$ and $\alpha \geq 3\beta(c \cdot x^u - c \cdot y) + 3n - 5$, the ratio is at least 1.

Case 2: $\beta(c \cdot x^u - c \cdot y) < 2$: The mass of the ray inside K_L is at least

$$
\lambda_0^n \int_0^1 t^{n-1} \, \mathrm{d}\, t
$$

yielding a ratio of

$$
\frac{\alpha - \beta(c \cdot x^u - c \cdot y) - (n-1)}{n e^{\beta(c \cdot x^u - c \cdot y)}}
$$

which, by our assumption on β, is at least

$$
\frac{\alpha - 2 - (n-1)}{n e^2}
$$

and since $\alpha \geq n(e^2 + 1) + 1$ is at least 1.

$\square$

Summarizing, we have the following:

Theorem 1 *For any $L \geq 0$, $T > 0$ and F as in (1) with α, β satisfying*

$$\begin{aligned} \beta &\geq \frac{n}{T} \\ \alpha &\geq \max\Big(3\beta(c \cdot x^u - L) + 3n - 5, n(e^2+1)+1\Big) \end{aligned}$$

then

$$\int_{\text{"the tip"}} F \geq (1/4) \int_{\mathcal{R}^n} F.$$

5 SAMPLING PROCEDURE.

We now show how to approximately sample according to F (namely, $B_{(L,T)}$). First, we discretize the rectangle $x^l \leq x \leq x^u$. Next, we find an approximation for F in regions not in K_L. Third, we devise the transition matrix of a Markov chain which realizes the desired random walk.

Let E^i denote the unit vector directed in the ith coordinate direction. Let $C_\delta(p)$ denote the cube of side 2δ centered at p.

We will divide the rectangle $x^l \leq x \leq x^u$ into small cubes of side 2δ where δ will be specified later.

$$\begin{aligned} U &\stackrel{\text{def}}{=} \Big\{ x \in \mathcal{R}^n \;\Big|\; \tfrac{x - x^f}{2\delta} \in Z^n \;,\; C_\delta(x) \cap \Big\{ y \in \mathcal{R}^n \;\Big|\; x^l \leq y \leq x^u \Big\} \neq \emptyset \Big\} && (4) \\ J &\stackrel{\text{def}}{=} \bigcup_{x \in U} C_\delta(x) && (5) \end{aligned}$$

We will take a random walk on the graph whose vertices are the set U of centers of cubes. Also, notice that even though there may be $x \in U$ such that $x \not\geq x^l$ we do have $x + \delta\vec{1} \geq x^l$ for all $x \in U$. It is important to notice that the l_∞ diameter of J is no more that $\left\| x^u - x^l \right\|_\infty + 4\delta$.

Many of the lemmas require that δ not be too large with respect to x^u, x^f, α and β. To formalize this we will can δ "fine" if we have

$$\delta \leq \min\Big(\frac{\min_i (x_i^u - x_i^f)}{7\alpha}, \frac{1}{7\sqrt{n}\beta c_i}\Big), \tag{6}$$

and we will often invoke the following result.

Proposition 1 *If δ is "fine" and α, β meet the conditions of Theorem 1, then $\frac{x_i^u - x_i^f}{\delta} \geq 7((e^2 + 1)n + 1)$ for all i.*

5.1 Discretization and Approximate Sampling using Membership Oracle.

Errors due to two sources need to be analyzed here. One source of error is because we discretize the region, and approximate the integral of $F(x)$ over a small rectangle by $F(p)*$(volume of the rectangle), where p is the center of the rectangle. The second source of error is in computing the gauge function for points outside the feasible region. This is because in practice we may only be able to calculate $\hat{F}(p)$, an approximation for $F(p)$ (using the membership oracle and bisection methods).

5.1.1 Estimates because of Discretization.

In this section (and in section 6) we will need for every $p \in U$ that the integral of F over any rectangle C centered at p and approximately contained in $C_\delta(p)$ is well approximated by $F(p)Vol(C)$.[1]

More precisely: For every $p \in J$ we need to determine a lower bound on ρ such that for some continuous monotone decreasing function ξ such that $\xi(0) = 0$ and all $\eta \geq 0$ in some open neighborhood of 0: if C is any rectangular region with center p contained in $C_{\delta+\eta}(p) \cap J$ then

$$\rho(1 - o(1)) \left(\frac{1}{Vol(C)} \int_C F \right) \leq F(p) \leq (\rho(1 - o(1)))^{-1} \left(\frac{1}{Vol(C)} \int_C F \right). \tag{7}$$

We will also need a lower bound on σ such that

$$\sigma F(x) \leq F(x + \lambda E^i) \tag{8}$$

for all $x \in K$, $i \in \{1 \cdots n\}$ and all $|\lambda| \leq 2\delta$.

The lemma below summarizes the errors induced here. The proof is in Appendix B.

Lemma 3 *(7) and (8) hold with*

$$\sigma \geq e^{-2\alpha\delta/(\min_i(x_i^u - x_i^f))} e^{-2\beta\delta\|c\|_\infty} \tag{9}$$

$$\rho \geq \frac{e^{-\alpha\delta/(\min_i(x_i^u - x_i^f))}}{1 + \sqrt{2\pi}\beta\delta \|c\|_2 \operatorname{erf}\left(\frac{\beta\delta\|c\|_2}{\sqrt{2}}\right) e^{\beta^2\delta^2\|c\|_2^2/2}} \tag{10}$$

[1] The approximate containment, characterized by a parameter η, is technical point used only to facilitate the proof of Lemma 4.

5.1.2 Estimates due to gauge function errors.

Let $\hat{F}(x)$ be an approximation for $F(x)$ calculated using only x and the membership oracle. In the feasible region, $\hat{F} = F$. In the infeasible region, $\hat{F}$ may not equal F because the dilation cannot be computed exactly. Note that for our analysis this must be a deterministic approximation and not one obtained by sampling; to be clear, we always calculate the same value for $\hat{F}(x)$. This is calculated to a relative accuracy of $1 \pm \frac{1}{11}$ (i.e. $\frac{10}{11}F(x) \leq \hat{F}(x) \leq \frac{12}{11}F(x)$.)

To calculate $\hat{F}(x)$ to a relative accuracy of $1/11$, it is sufficient to calculate the gauge function to an absolute error of $\pm\frac{\ln(12/11)}{\alpha}$. To achieve this, it is sufficient to calculate the distance of the point in K on the line segment from x to x^u that is farthest from x^u to an absolute accuracy of $\frac{\ln(12/11)(\min_i(x_i^u - x_i^f))^2}{2\left\|x^u - x^f\right\|_2 \alpha}$. This can be done very quickly by bisection search using our membership oracle.

5.2 The Markov Chain.

For each $x \in U$ let $N(x)$ be the "neighborhood" of x which is the set of all vertices in U that differ from x in exactly one coordinate by $\pm 2\delta$. The transition probabilities $P(x, y)$ will be :

$$P(x,y) = \begin{cases} \frac{1}{2n}\min(1, \frac{\hat{F}(y)}{\hat{F}(x)}) & y \in N(x) \\ 1 - \sum_{z \in N(x)} P(x,z) & x = y \\ 0 & y \notin N(x) \end{cases} \tag{11}$$

where $\hat{F}$ is a deterministic estimate for F. It is easy to see that $P(x, y)$ induces a time reversible irreducible aperiodic Markov chain with steady state probabilities $\pi(\cdot)$ proportional to $\hat{F}(\cdot)$. We will show, in the next section, that after a sufficient number of steps, we are fairly close to the steady state. Let π be the unique steady state distribution for our chain (approximately $B_{(L,T)}$). But first we show that in the steady state there is a reasonable chance of observing states corresponding to cubes covering the near optimal feasible portions of K_L.

Theorem 2 *Let $\pi(\cdot)$ be the steady state probabilities of the above Markov chain, then if α, β satisfy the conditions of Theorem 1 and δ is "fine" and "the tip" is contained in J then*

$$\sum_{x \in U, C_\delta(x) \cap \text{"the tip"} \neq \emptyset} \pi(x) \geq \frac{1}{6}.$$

Proof Accounting for the errors due to discretization and in gauge function computation, $\pi(x)$ $(= D\hat{F}(x))$ satisfies

$$D(1-\frac{1}{11})\rho\left(\frac{1}{(2\delta)^n}\int_{C_\delta(x)}F\right) \leq \pi(x) \leq D(1+\frac{1}{11})\rho^{-1}\left(\frac{1}{(2\delta)^n}\int_{C_\delta(x)}F\right). \tag{12}$$

where $D = \left(\sum_{x\in U}\hat{F}\right)^{-1}$.

Now, since the "tip" probability is at least $\frac{1}{4}$ (by construction of F), we have the steady state probability of the cubes covering the tip is:

$$\sum_{x\in U, C_\delta(x)\cap\text{“the tip”}\neq\emptyset} \pi(x) \geq \frac{1\times(1-\frac{1}{11})\rho}{3\times(1+\frac{1}{11})\rho^{-1}+1\times(1-\frac{1}{11})\rho}.$$

δ is fine, so:

$$\frac{\alpha\delta}{\min_i(x_i^u - x_i^f)} \leq \frac{1}{7} \tag{13}$$

and

$$\beta\delta\,\|c\|_2 \leq \frac{1}{7}, \tag{14}$$

and the theorem follows from inequality (10).

□

6 SPECTRAL GAP OF THE MARKOV CHAIN.

In this section, we determine how many steps are necessary for the Markov chain to "mix". We need to find a relationship between the number of steps walked and how close we are to the steady state. For this we need a central result of Sinclair and Jerrum [12]. We also need several well-know facts that are collected in [3]. Let X be the set of states in our Markov chain. For $x, y \in X$, let $P^t(x,y)$ be the probability that starting in state x we are in state y after t steps. As before, let π be the unique steady state distribution for our chain (approximately $B_{(L,T)}$). Recall that $P(x,y)$ induces a time reversible irreducible aperiodic Markov chain; however, it is not strongly aperiodic [12]. This is because we do not insist that we have $P(x,x) \geq \frac{1}{2}$ for all x. Because of this, we not only need an upper bound for the second largest eigenvalue but also need a lower bound on the smallest eigenvalue [3].

We will use Proposition 3 from [3] which says :

$$\sum_{y\in X}|P^t(x,y)-\pi(y)| \leq \sqrt{\frac{1-\pi(x)}{\pi(x)}}(\chi^*)^t$$

.

where, $\chi^* = \max(\chi_1, |\chi_m|)$, where χ_1 is the second largest eigenvalue and χ_m is the smallest eigenvalue of P, the matrix of transition probabilities.

We find an upper bound on χ_1 by appealing to a result of Sinclair and Jerrum's [12], which is quoted also as Proposition 6 of [3], namely, $\chi_1 \leq 1 - \frac{\hat{\phi}^2}{2}$, where $\hat{\phi}$ is a lower bound on the "conductance" (defined in section 6.1) of the Markov chain. We find a lower bound on the conductance in section 6.1 below. A lower bound on χ_m is obtained by appealing to Proposition 2 of [3], described in detail in section 6.2.

We also prove that (see sections 6.1-6.2) $|\chi_m| \leq 1 - \frac{\hat{\phi}^2}{2}$. Thus, we have:

$$\sum_{y \in X} |P^t(x,y) - \pi(y)| \leq \frac{1}{\sqrt{\pi(x)}}(1 - \frac{\hat{\phi}^2}{2})^t.$$

What we wish to do is find t so that $\sum_{y \in X} |P^t(x,y) - \pi(y)|$ is under 1/12. Recall that by Theorem 2, the states of our Markov chain corresponding to cubes that cover the "tip" have mass at least 1/6. Thus, we will have a chance of at least $1/6 - 1/12 = 1/12$ that a random walk of t steps will end in one of the states corresponding to a cube covering the tip (i.e. close to the optimum). For this, it suffices to have

$$t \geq \ln\left(\frac{\sqrt{\pi(x)}}{12}\right) / \ln\left(1 - \frac{\hat{\phi}^2}{2}\right) \geq \ln\left(\frac{12}{\sqrt{\pi(x)}}\right) / \left(\frac{\hat{\phi}^2}{2}\right) = \frac{2\ln(12) - \ln(\pi(x))}{\hat{\phi}^2}. \tag{15}$$

We now wish to prove a lower bound on $\pi(x)$, for feasible x, so that we can apply the above inequality.

We assume that the walk is started deterministicly at x^f. Since we know that at least half of the mass of $B_{(L,T)}$ is in the body and the highest possible stationary probability of a cube intersecting K_L is at most $e^{\beta(c\cdot(x^f+2\delta\vec{1})-L)}\pi(x^f)$ and there are at most $\prod_{i=1}^n \left\lceil \frac{x_i^u - x_i^l}{2\delta} + 1 \right\rceil$ states in K_L (or even U), we know

$$\frac{\rho 10/11}{\rho^{-1}12/11}\frac{1}{2} \leq \pi(x^f) e^{\beta(c\cdot(x^f+2\delta\vec{1})-L)} \prod_{i=1}^n \left\lceil \frac{x_i^u - x_i^l}{2\delta} + 1 \right\rceil$$

which yields

$$\pi(x^f) \geq \frac{\rho^2 5 e^{\beta(L - c\cdot(x^f+2\delta\vec{1}))}}{12\prod_{i=1}^n \left\lceil \frac{x_i^u - x_i^l}{2\delta} + 1 \right\rceil}$$

but we will just use the easier form:

$$\frac{5\rho^2 e^{\beta(L - c\cdot(x^f+2\delta\vec{1}))}}{12\left\lceil \frac{\left\|x^u - x^l\right\|_\infty}{2\delta} + 1 \right\rceil^n}. \tag{16}$$

Plugging in, we get

Theorem 3 *If the conditions of Theorem 2 are met then running the above Markov chain for at least*

$$\frac{2\ln(12)+\ln\left(\frac{12}{5\rho^2}\right)+\beta(c\cdot(x^f+2\delta\vec{1})-L)+n\ln\left(\left\lceil\frac{\|x^u-x^l\|_\infty}{2\delta}+1\right\rceil\right)}{\hat{\phi}^2} \tag{17}$$

is sufficient to ensure with probability at least $\frac{1}{12}$, the chain will stop at a state x such that $x+\delta\vec{1}$ is feasible and within cost $T+2\delta\|c\|_1$ of the optimal point.

Although at the end of Section 4, we had shown that the steady state probability of being in the tip is at least $\frac{1}{6}$, now we only guarantee the probability that the sample is in the tip at the end of t steps is (at least) $\frac{1}{12}$. Thus, the three sources of errors–discretization, approximation of the gauge function, walking for t steps–reduce the tip probability from $\frac{1}{4}$ to $\frac{1}{12}$.

6.1 Conductance.

For any $V \subseteq U$ and $\bar{V} = U \setminus V$, and

$$\begin{aligned} V_\delta &= \bigcup_{x\in V} C_\delta(x) \\ \bar{V}_\delta &= \bigcup_{x\in \bar{V}} C_\delta(x) \end{aligned}$$

We define the conductance of V by

$$\phi_V = \frac{\sum_{x\in V, y\in \bar{V}\cap N(x)} \pi(x)P(x,y)}{\min(\pi(V),\pi(\bar{V}))} = \frac{\sum_{x\in V, y\in \bar{V}\cap N(x)} \min(\hat{F}(x),\hat{F}(y))}{2n\min(\hat{F}(V),\hat{F}(\bar{V}))}. \tag{18}$$

The conductance of the chain defined by

$$\phi = \min_{V\subseteq U} \phi_V. \tag{19}$$

We use an "isoperimetric inequality" to find a lower bound for ϕ. An isoperimetric inequality was proved in [5]; a simpler proof of a stronger inequality was given in [11]. The inequality was generalized to the case of log-concave functions in [1]. We use here a version of this from [4]. If $\mathrm{dist}(x,y) = ||x-y||$ where $||\cdot||$ is an arbitrary norm on $\mathcal{R}^n$ and $\mathrm{diam}(K) = \max_{x,y\in K}\mathrm{dist}(x,y)$ we have the following theorem from [4]:

Theorem 4 *Let $J \subseteq \mathcal{R}^n$ be a convex body and F a log-concave function defined on* $\operatorname{int} J$ *and μ the induced measure. Let $S_1, S_2 \subseteq J$, and $t \leq \operatorname{dist}(S_1, S_2)$ and $d \geq \operatorname{diam}(J)$. If $B = J \setminus (S_1 \cup S_2)$, then*

$$\min(\mu(S_1), \mu(S_2)) \leq \frac{1}{2}(d/t)\mu(B). \tag{20}$$

We will use $\operatorname{dist}(x, y) = \|x - y\|_\infty$.

Lemma 4 *If δ is fine then $\phi \geq \frac{\delta}{3n\|x^u - x^l\|_\infty} = \hat{\phi}$ (say).*

Proof Let η be a small positive real that will tend to zero. Let B_δ be the $\eta/2$ neighborhood of $V_\delta \cap \bar{V}_\delta$ and $B_\delta(x, y)$ be the $\eta/2$ neighborhood of $C_\delta(x) \cap C_\delta(y)$. Let S_1, S_2 and B be $V_\delta \setminus B_\delta, \bar{V}_\delta \setminus B_\delta$ and $B_\delta \cap J$ respectively.

From inequalities (7), (8), (9), (10) and (12), it is clear that

$$\begin{aligned}
\sum_{x \in V, y \in \bar{V} \cap N(x)} \min(\hat{F}(x), \hat{F}(y)) &\geq \frac{10}{11} \sum_{x \in V, y \in \bar{V} \cap N(x)} \min(F(x), F(y)) \\
&\geq \frac{10}{11}\sigma \sum_{x \in V, y \in \bar{V} \cap N(x)} F\left(\frac{x+y}{2}\right) \\
&\geq \frac{10}{11}\sigma \sum_{x \in V, y \in \bar{V} \cap N(x)} \rho \frac{1}{\eta(2\delta + \eta)^{n-1}} \int_{B_\delta(x,y)} F \\
&\geq \frac{10}{11} \frac{\sigma\rho}{\eta(2\delta + \eta)^{n-1}} \int_{B_\delta} F \\
&\geq \frac{10}{11} \frac{\sigma\rho}{\eta(2\delta + \eta)^{n-1}} \int_{B} F \\
\text{similarly} \quad \min(\hat{F}(V), \hat{F}(\bar{V})) &\leq \frac{12}{11} \frac{1}{\sigma\rho(2\delta)^n} \min\left(\int_{V_\delta} F, \int_{\bar{V}_\delta} F\right) \\
&\leq \frac{12}{11} \frac{1}{\sigma\rho(2\delta - \eta)^n} \min\left(\int_{S_1} F, \int_{S_2} F\right)
\end{aligned}$$

From the isoperimetric inequality,

$$\frac{\int_B F}{\min(\int_{S_1} F, \int_{S_2} F)} \geq \frac{2\eta}{\|x^u - x^l\|_\infty + 4\delta}.$$

Combining this with the inequalities above and taking the limit $\eta \to 0$ we get:

$$\phi \geq \frac{2\frac{10}{11}\sigma^2\rho^2\delta}{\frac{12}{11}n(\|x^u - x^l\|_\infty + 4\delta)}. \tag{21}$$

Since δ is fine (from inequalities (10),(9))

$$\phi \geq \frac{\delta}{3n\|x^u - x^l\|_\infty} = \hat{\phi}. \tag{22}$$

$\square$

6.2 Comparison of χ_1 and $|\chi_m|$.

Here we find a lower bound for χ_m, by the *canonical odd path* argument outlined in Proposition 2 of [3].

Let

$$\Delta = \frac{\delta}{8n \left\|x^u - x^l\right\|_\infty}. \tag{23}$$

For each state x let ω_x be the smallest non-negative integer such that $P(x+2\omega_x\delta E^1, x+2\omega_x\delta E^1) \geq \Delta$ (this is always possible since $P(a,a) \geq \frac{1}{2n} \geq \Delta$ on the border of our bounding region). Let σ_x be the $2\omega_x + 1$ step path of from x to x given by:

$$\underline{x} \mapsto \underline{x+2\delta E^1} \mapsto \underline{x+4\delta E^1} \mapsto \cdots \overbrace{\underline{x+2\omega_x\delta E^1} \mapsto \underline{x+2\omega_x\delta E^1}}^{\text{“self loop”}} \mapsto \cdots \underline{x+4\delta E^1} \mapsto \underline{x+2\delta E^1} \mapsto \underline{x}$$

We will call σ_x “the canonical odd path for x”.

Proposition 6 of [3] states for any selection of canonical odd paths we have $\chi_m \geq -1 + \frac{2}{\iota}$, where

$$\iota \stackrel{\text{def}}{=} \max_{(a,b)} \sum_{\sigma_x \ni (a,b)} \|\sigma_x\|_P \, \pi(x), \tag{24}$$

$$\text{where} \quad \|\sigma_x\|_P \stackrel{\text{def}}{=} \sum_{(a,b)\in\sigma_x} \frac{1}{\pi(a)P(a,b)}. \tag{25}$$

To prove the bound for χ^*, we will show that $|\chi_m|$ is less than the upper bound for χ_1. From the discussion above, it is sufficient to show the following.

Lemma 5 $\iota \leq \frac{4}{\hat{\phi}^2}$

Proof By our choice of paths, we have for any $i \leq \omega_x - 1$:

$$\begin{array}{rcccl} \frac{1}{2n} - \Delta & \leq & P(x+i2\delta E^1, x+(i+1)2\delta E^1) & \leq & \frac{1}{2n} \\ & & P(x+(i+1)2\delta E^1, x+i2\delta E^1) & \leq & \frac{1}{2n} \end{array}$$

and by time reversibility

$$\pi(x+(i+1)2\delta E^1) = \frac{P(x+i2\delta E^1, x+(i+1)2\delta E^1)}{P(x+(i+1)2\delta E^1, x+i2\delta E^1)}\pi(x+i2\delta E^1) \geq \frac{\frac{1}{2n}-\Delta}{\frac{1}{2n}}\pi(x+i2\delta E^1).$$

For any x we have,

$$\begin{aligned}
\|\sigma_x\|_P &\leq 2\sum_{i=1}^{\omega_x} \frac{1}{\pi(x)(1-2\Delta n)^i(\frac{1}{2n}-\Delta)} + \frac{1}{\pi(x)(1-2\Delta n)^{\omega_x}\Delta} \\
&\leq 2 \sum_{i=1}^{\left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil} \frac{1}{\pi(x)(1-2\Delta n)^i(\frac{1}{2n}-\Delta)} + \frac{1}{\pi(x)\Delta(1-2\Delta n)^{\left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil}} \\
&\leq 2\left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil \frac{2n}{\pi(x)\left(1-\left\lceil \frac{x_1^u - x_1^l}{2\delta}+2 \right\rceil 2\Delta n\right)} + \frac{1}{\pi(x)\Delta\left(1-\left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil 2\Delta n\right)}.
\end{aligned}$$

Using Proposition 1 it is easy to show that

$$\left\lceil \frac{x_1^u - x_1^l}{2\delta}+2 \right\rceil 2\Delta n \leq \frac{1}{3}.$$

Continuing,

$$\|\sigma_x\|_P \leq \left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil \frac{6n}{\pi(x)} + \frac{3}{2\pi(x)\Delta}.$$

Because each edge can be used by at most $\left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil$ canonical odd paths, we have

$$\iota \leq \left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil \left(\left\lceil \frac{x_1^u - x_1^l}{2\delta}+1 \right\rceil 6n + \frac{3}{2\Delta}\right). \tag{26}$$

The result now follows from Lemma 4, inequality (26) and Proposition 1.

□

7 DESCRIPTION AND ANALYSIS OF THE ALGORITHM.

Suppose we are given a feasible point x^f, a relative accuracy goal (ϵ) and a desired upper bound on the probability that the algorithm fails (κ). Let ν be the ratio of $c \cdot x^f$ to $c \cdot x^{opt}$. We assume that without loss of generality, the problem has been rescaled so $c_i = 1$ for all i ($x_i \to c_i x_i$, $c_i \to \frac{c_i}{c_i} = 1$).

7.1 In the Worst Case.

Here, we assume that $n \geq 2$ and $\epsilon \leq 1$. We present the following algorithm:

Input: x^f, ϵ, κ, c and a membership oracle for K.

rescale problem so $c = \vec{1}$

$x^l \leftarrow 0$ (pointwise lower bound)

$L \leftarrow 0$ (probalistic cost lower bound)

$x^{\text{best}} \leftarrow x^f$ (the best feasible solution observed)

while $\sum_i x_i^f - L > \epsilon L$

$x_i^u \leftarrow 2 \sum_j x_j^f$, $i = 1 \cdots n$ (pointwise upper bound)

$T \leftarrow \frac{\sum_j x_j^f - L}{3}$ ($\frac{1}{3}$ of the current gap)

$\beta \leftarrow \frac{n}{T}$ (objective function "bias")

$\alpha \leftarrow \frac{7n^2 \sum_i x_i^f}{T}$ (gauge function gain)

$\delta \leftarrow \frac{T}{49n^2}$ (step size)

repeat $\left\lceil \log_{\frac{12}{11}} \left(\frac{\lceil \log_2(\frac{7}{\epsilon}) \rceil + 1}{\kappa} \right) \right\rceil$ times or until $\sum_i x_i^{\text{best}} \leq \frac{\sum_j x_j^f + L}{2}$

run the random walk of section 5 with the above parameters for the number of steps prescribed in Theorem 3, let x be the stopping point of the walk.

if $x + \delta\vec{1}$ feasible and $\sum_i (x_i + \delta) < \sum_i x_i^{\text{best}}$

then $x^{\text{best}} \leftarrow x + \delta\vec{1}$

endrepeat

if $\sum_i x_i^{\text{best}} > \frac{\sum_j x_j^f + L}{2}$

then $L \leftarrow \sum_i x_i^{\text{best}} - T - 2n\delta$

$x^f \leftarrow x^{\text{best}}$

endwhile

return x^f

Figure 1: **AlgorithmA.**

The analysis of **AlgorithmA** is fairly straight forward.

- It is easy to see that x^u such that $x^u_i = \sum_j x^f_j$ must dominate any optimal point.
- Each time we draw a sample according to Theorem (3) we have at least a chance of $\frac{1}{12}$ that it is feasible and within a cost of $T + 2n\delta$ of the optimum. We have picked T and δ such that $T + 2n\delta < \frac{\sum_j x^f_j - L}{2}$.
- When the **repeat** loop terminates either
  - $\sum_i x^{\text{best}}_i \leq \frac{\sum_j x^f_j + L}{2}$, or
  - enough samples have been drawn such that with confidence at least $(1-\kappa)^{\frac{1}{\left\lceil \log_2\left(\frac{7}{\epsilon}\right)\right\rceil + 1}}$ one of them was feasible and within distance $T + 2n\delta$ of the optimum.

  Either way, the distance from the new x^f to the new L is no more than half of the old distance.
- Thus, the outer **while** loop will run no more than $\lceil \log_2\left(\frac{\nu}{\epsilon}\right)\rceil$ times.
- Furthermore, we see the first time the algorithm alters the lower bound (it must establish a lower bound to halt), we have $L \geq \frac{\sum_i x^f_i}{2} - T - 2n\delta \geq \frac{1}{7}\sum_i x^f_i$.
  - Therefore, the outer while loop will cycle no more than $\left\lceil \log_2\left(\frac{7}{\epsilon}\right)\right\rceil$ more times.
  - Thus the lower bound can be altered at most $\left\lceil \log_2\left(\frac{7}{\epsilon}\right)\right\rceil + 1$ times.
  - Since each lower bound alteration is correct with chance at least $(1-\kappa)^{\frac{1}{\left\lceil \log_2\left(\frac{7}{\epsilon}\right)\right\rceil + 1}}$ we see that they all are correct with odds at least $1-\kappa$.

Thus the algorithm fails with chance less than κ.

We must check that α, β and δ were picked correctly.

- β clearly satisfies inequality (2).
- Assuming that $n \geq 2$, we see that α satisfies inequality (3).
- It is easy to see that inequality (6) is satisfied.
- Invoking Lemma (4), $\hat{\phi} \geq \frac{\epsilon}{6174n^3}$.
- So $t = \frac{(6174)^2 n^6 \left(7 + 3.2n + n\ln\left(\frac{1173n^2}{2\epsilon} + 2\right)\right)}{\epsilon^2}$ steps are enough to draw a sample. Thus, each sample can be drawn in $O\left(\frac{n^7 \log\left(\frac{n}{\epsilon}\right)}{\epsilon^2}\right)$ steps.

- Each step requires at most $O\left(log\left(\frac{n}{\epsilon}\right)\right)$ membership queries to compute the gauge function.

So the total number of membership queries is

$$O\left(\frac{n^7\left(\log\left(\frac{n}{\epsilon}\right)\right)^2\log\left(\frac{\log\left(\frac{1}{\epsilon}\right)}{\kappa}\right)\log\left(\frac{\nu}{\epsilon}\right)}{\epsilon^2}\right), \tag{27}$$

which, if we ignore lesser log factors, can be thought of as $O^{\sim}\left(\frac{n^7\log(\nu)\log\left(\frac{1}{\kappa}\right)}{\epsilon^2}\right)$.

Remark: 1 *With a new result of Frieze, Kannan and Polson that the algorithm outlined here can have it's dependence on n brought down to n^6 by performing the sampling walk in a bounding sphere instead of a bounding box and estimating χ^* without introducing the idea of conductance.*

7.2 With Advantageous Bounds.

Here we analyze the situation where good bounds x^l, L and x^u are known and attempt to lower the dependence of the runtime on n. To do this meaningfully, all dot products (and norms other than infinity) must be removed from the expressions as they hide n's. In this subsection, we work out a bound on run time that explicitly shows all of the powers of n.

We still assume that the problem has been rescaled so $c_i = 1$ for all i ($x_i \to c_i x_i$, $c_i \to \frac{c_i}{c_i} = 1$) and that

$$\frac{\min_i(x_i^u - x_i^f)}{\|x^u\|_\infty - \min_i x_i^l} \geq \frac{1}{2} \tag{28}$$

$$\text{and} \quad \min_i(x_i^u) + \min_i(x_i^l) \geq 2\left\|x^f\right\|_\infty. \tag{29}$$

This is easy to ensure by replacing x_i^u with $\left\|x^f\right\|_\infty + \|x^u\|_\infty$ and has the geometric interpretation of making the problem "well rounded".

We notice that if $\epsilon > \frac{\|x^f\|_\infty}{\min_i x_i^l} - 1$ then x^f is already a solution of the desired accuracy. This and inequality (29) imply

$$\frac{\|x^u\|_\infty - \min_i x_i^l}{\epsilon \min_i x_i^l} \geq 2. \tag{30}$$

Rather than the adaptive approach, first consider a sampling algorithm that comes within ϵ of optimal in one long walk:

- We set $L = \left\|x^l\right\|_1$ and $T = \epsilon\left\|x^l\right\|_1$.

- We will chose β to be at least $\frac{11}{10}\frac{n}{T}$ instead of as stated in inequality (2). The $\frac{11}{10}$ is required to guarantee that we get within T of the optimum instead of the $T+2\delta\|c\|_1$ we could expect because of discretization. To guarantee this, we must show that $\frac{\epsilon}{10}\min_i x_i^l \geq 2n\delta$.
  - So we set
$$\beta = \frac{11}{10\epsilon \min_i x_i^l} \tag{31}$$
$$\text{and } \alpha = \frac{5n(\|x^u\|_\infty - \min_i x_i^l)}{\epsilon \min_i x_i^l}. \tag{32}$$
  - By inequality (30), this satisfies inequality (3).
  - Now setting
$$\delta = \frac{\epsilon \min_i x_i^l}{70n} \tag{33}$$
    satisfies inequality (6), by inequality (28).
  - Clearly, we have $2n\delta \leq \frac{1}{10}\epsilon \min_i x_i^l$.

  So the $2\delta\|c\|_1$ factor has been dealt with.

Now, by Lemma (4) and Theorem (3), we have

$$\hat{\phi} = \frac{\epsilon \min_i x_i^l}{210n^2\|x^u - x^l\|_\infty} \tag{34}$$
$$t = \left\lceil \left(7 + \frac{11n\left\|x^f - x^l\right\|_\infty}{10\epsilon \min_i x_i^l} + n\ln\left(\left\lceil \frac{70n\left\|x^u - x^l\right\|_\infty}{2\epsilon \min_i x_i^l} + 1\right\rceil\right)\right)\left(\frac{210n^2\left\|x^u - x^l\right\|_\infty}{\epsilon \min_i x_i^l}\right)^2\right\rceil \tag{35}$$

Which, if we take the middle term of the sum to be dominant, is

$$O\left(n^5\left(\frac{\left\|x^u - x^l\right\|_\infty}{\epsilon \min_i x_i^l}\right)^3\right) \tag{36}$$

steps.

We will call this algorithm **AlgorithmB**. **AlgorithmB** can then be repeated $\left\lceil \frac{\ln(\kappa)}{\ln(11/12)} \right\rceil$ times to amplify the chance of success to at least $1-\kappa$.

The dependence on $\frac{\left\|x^u - x^l\right\|_\infty}{\epsilon \min_i x_i^l}$ can be improved by designing a new algorithm (**AlgorithmC**) that runs **AlgorithmB** in stages like we did for **AlgorithmA**.

The analysis is as before and the run time comes out to

$$O\left(n^5\left(\frac{\left\|x^u-x^l\right\|_\infty}{\epsilon\min_i x_i^l}\right)^2\left(\ln\left(\frac{n\left\|x^u-x^l\right\|_\infty}{\epsilon\min_i x_i^l}\right)\right)^2\ln(\frac{1}{\kappa})\right) \tag{37}$$

membership queries.

Appendices

A Proof of NP Completeness

Theorem 5 *Given m points $y^1, y^2, \ldots y^m$ in $\mathcal{R}^n$, a fraction γ, a positive vector c and a real number ζ, deciding if there exist an $x \in \mathcal{R}^n$ such that $x \geq y^i$ is satisfied for at least γm different i 's and $c \cdot x \leq \zeta$ is NP complete.*

Proof sketch: Let $G = (V, E)$ be an undirected graph with vertices $V = \{1, 2 \cdots n\}$ and edges E. Given an integer k, the clique problem is: "does G have a complete induced subgraph on some k vertices?" For each $e \in E$ let $y^e \in \mathcal{R}^n$ be the vector such that

$$y_i^e = \begin{cases} 1 & \text{vertex } i \text{ is an endpoint of edge } e \\ 0 & \text{otherwise} \end{cases}$$

Let $c = \vec{1}$, $\gamma = (k-1)k/2$ and $\beta = k$.

Let x be a solution to the above problem. WLOG assume x is a 0-1 vector and define $G_x = (V_x, E_x)$ to be the induced subgraph of G where $V_x = \{i \in V \mid x_i \geq 1\}$. We then have $|G_x| = c \cdot x$ and $|E_x| = |\operatorname{dom}(x)|$. So we see that x such that $c \cdot x \leq k$ and $|\operatorname{dom}(x)| \geq k(k-1)/2$ correspond precisely to induced subgraphs of G with $\leq k$ vertices and $\geq k(k-1)/2$ edges: k-cliques. □

B Errors Due to Discretization

For every $p \in J$ we need to determine a lower bound on ρ such that for some continuous monotone decreasing function ξ such that $\xi(0) = 0$ and all $\eta \geq 0$ in some open neighborhood of 0: if C is any rectangular region with center p contained in $C_{\delta+\eta}(p) \cap J$ then

$$\rho(1-\xi(\eta))\left(\frac{1}{Vol(C)}\int_C F\right) \leq F(p) \leq (\rho(1-\xi(\eta)))^{-1}\left(\frac{1}{Vol(C)}\int_C F\right). \tag{38}$$

We will also need a lower bound on σ such that

$$\sigma F(x) \leq F(x + \lambda E^i) \tag{39}$$

for all $x \in K$, $i \in \{1 \cdots n\}$ and all $|\lambda| \leq 2\delta$.

For vectors a and b let ab be the vector $(ab)_i = a_i b_i$. To facilitate the analysis we break F into its two constituent parts f and g where $f(x) = e^{-\alpha(\max(\psi_{(K_L,x^u)}(x)-1,0))}$, $g(x) = e^{-\beta c \cdot x}$, and $F(x) = f(x)g(x)$.

σ is easy to deal with when we apply the well known fact that a gauge function based on K can fall no faster than one based on a convex subset of K (containing x^u). To be precise we apply Corollary 52 from [1] to get:

$$|\psi_{(K_L,x^u)}(x) - \psi_{(K_L,x^u)}(y)| \leq \frac{\|x-y\|_\infty}{\min_i(x_i^u - x_i^f)}$$

which implies

$$e^{-\alpha(\delta+\eta)/(\min_i(x_i^u - x_i^f))} f(x) \leq f(z) \quad \forall z \in C \tag{40}$$

and inequality 39 is satisfied with

$$\sigma = e^{-2\alpha\delta/(\min_i(x_i^u - x_i^f))} e^{-2\beta\delta\|c\|_\infty}. \tag{41}$$

To get a lower bound on ρ we use the simple rule that for $f, g \geq 0$

$$\min_{x \in C}(f(x)) \int_C g \leq \int_C fg \leq \max_{x \in C}(f) \int_C g$$

and use inequality 40 to get the pointwise bounds on f. All that remains is to derive a lower bound ρ' such that

$$\rho' \left(\frac{1}{Vol(C)} \int_C g \right) \leq g(p) \leq {\rho'}^{-1} \left(\frac{1}{Vol(C)} \int_C g \right). \tag{42}$$

We note that g is of the form $g(x) = h(c \cdot x)$ for some non-negative convex function h in the region we are interested in and since C is symmetric about p we know that (from Jensen's inequality)

$$g(p) \leq \frac{1}{Vol(C)} \int_C g,$$

and any ${\rho'}^{-1} \leq 1$ satisfies the right side of inequality 42.

To get the left side of inequality 42 we define

$$\begin{aligned} \Xi_i &\stackrel{\text{def}}{=} \max\{\lambda \in \mathcal{R} \mid p + \lambda E^i \in C\} \\ D &\stackrel{\text{def}}{=} \{x \in \mathcal{R}^n \mid |x_i - c_i p_i| \leq c_i \Xi_i \; \forall i\} \end{aligned}$$

and change variables to get:

$$\frac{1}{\prod_{i=1}^{n} 2c_i\Xi_i}\int_D g.$$

Let $\mu(\lambda)$ be the measure of the set

$$\left\{x \in D \;\middle|\; x\cdot\vec{1} - p\cdot c \geq \lambda\right\}$$

It is easy to see μ is differentiable and $-\mu'(\lambda) \geq 0$ for $\lambda \in [0, c\cdot\Xi]$. We return to our integral (using the estimate $\mu(\lambda) - \mu(\lambda + d\lambda) = -\mu'(\lambda)d\lambda$):

$$\begin{aligned}
&= \frac{1}{\prod_{i=1}^{n} 2c_i\Xi_i}\int_0^{c\cdot\Xi} -\mu'(\lambda)\left(e^{-\beta(c\cdot p+\lambda)} + e^{-\beta(c\cdot p-\lambda)}\right)\mathrm{d}\,\lambda \\
&= g(p)\frac{-2}{\prod_{i=1}^{n} 2c_i\Xi_i}\int_0^{c\cdot\Xi}\cosh(\beta\lambda)\mu'(\lambda)\,\mathrm{d}\,\lambda \\
&= g(p)\frac{-2}{\prod_{i=1}^{n} 2c_i\Xi_i}\left(\cosh(\beta\lambda)\mu(\lambda)|_{\lambda=0}^{\lambda=c\cdot\Xi} - \beta\int_0^{c\cdot\Xi}\sinh(\beta\lambda)\mu(\lambda)\,\mathrm{d}\,\lambda\right) \\
&= g(p)\frac{-2}{\prod_{i=1}^{n} 2c_i\Xi_i}\left(0 - \frac{\prod_{i=1}^{n} 2c_i\Xi_i}{2} - \beta\int_0^{c\cdot\Xi}\sinh(\beta\lambda)\mu(\lambda)\,\mathrm{d}\,\lambda\right) \\
&= g(p)\left(1 + \frac{2\beta}{\prod_{i=1}^{n} 2c_i\Xi_i}\int_0^{c\cdot\Xi}\sinh(\lambda)\mu(\lambda)\,\mathrm{d}\,\lambda\right)
\end{aligned}$$

By Theorem 2 of [8], we have $\mu(\lambda) \leq \left(\prod_{i=1}^{n} 2c_i\Xi_i\right) e^{-\lambda^2/(2\|c\Xi\|_2^2)}$. So continuing we have:

$$\frac{1}{Vol(C)}\int_C g \leq g(p)\left(1 + 2\beta\int_0^{c\cdot\Xi}\sinh(\lambda)e^{-\lambda^2/(2\|c\Xi\|_2^2)}\,\mathrm{d}\,\lambda\right)$$

We need an upper bound for

$$2\beta\int_0^{c\cdot\Xi}\sinh(\lambda)e^{-\lambda^2/(2\|c\Xi\|_2^2)}\,\mathrm{d}\,\lambda$$

which comes out to (by standard integral tables)

$$\begin{aligned}
&\beta\|c\Xi\|_2\sqrt{\frac{\pi}{2}}e^{\beta^2\|c\Xi\|_2^2/2}\left(2\,\mathrm{erf}\left(\frac{\beta\|c\Xi\|_2}{\sqrt{2}}\right) + \mathrm{erf}\left(\frac{c\cdot\Xi - \beta\|c\Xi\|_2^2}{\sqrt{2}\|c\Xi\|_2}\right) - \mathrm{erf}\left(\frac{c\cdot\Xi + \beta\|c\Xi\|_2^2}{\sqrt{2}\|c\Xi\|_2}\right)\right) \\
\leq\;& \sqrt{2\pi}\beta\|c\Xi\|_2\,\mathrm{erf}\left(\frac{\beta\|c\Xi\|_2}{\sqrt{2}}\right)e^{\beta^2\|c\Xi\|_2^2/2} \\
\leq\;& \sqrt{2\pi}\beta(\delta + \sqrt{n}\eta)\|c\|_2\,\mathrm{erf}\left(\frac{\beta(\delta+\sqrt{n}\eta)\|c\|_2}{\sqrt{2}}\right)e^{\beta^2(\delta+\sqrt{n}\eta)^2\|c\|_2^2/2}
\end{aligned}$$

which for sufficiently small $\eta > 0$ (and $\eta = 0$) is

$$\leq \sqrt{2\pi}\beta\delta\|c\|_2\,\mathrm{erf}\left(\frac{\beta\delta\|c\|_2}{\sqrt{2}}\right)e^{\beta^2\delta^2\|c\|_2^2/2}(1 + o(1))$$

combining this with our pointwise bound on f we get

$$\rho \geq \frac{e^{-\alpha\delta/(\min_i(x_i^u - x_i^f))}}{1 + \sqrt{2\pi}\beta\delta\|c\|_2\,\mathrm{erf}\left(\frac{\beta\delta\|c\|_2}{\sqrt{2}}\right)e^{\beta^2\delta^2\|c\|_2^2/2}}$$

References

[1] Applegate, David and Ravi Kannan. *Sampling and integration of near log-concave functions*, 23rd ACM Symposium on the Theory of Computing, 1991.

[2] Boppana, R. B. and M. M. Halldórsson. *Approximating maximum independent sets by excluding subgraphs*, Proc. of 2nd Scand. Workshop on Algorithm Theory, pp. 13-25, Springer-Verlag Lecture Notes in Computer Science #447, July, 1990.

[3] Diaconis, P. and D. Stroock. *Geometric Bounds for Eigenvalues of Markov Chains*, Annals of Applied Probability, 1991, Vol. 1, No. 1, pp. 36-61.

[4] Dyer, M. and A. Frieze. *Computing the volume of convex bodies: a case where randomness provably helps*, Proceedings of Symposia in Applied Mathematics, Vol 44, 1991.

[5] Dyer, M., A. Frieze and R. Kannan. *A Random Polynomial-Time Algorithm for Approximating the Volume of Convex Bodies*, Journal of the Association for Computing Machinery, Vol. 38, No.1 January 1991, pp.1-17.

[6] Garey, Michael and David Johnson. *Computers and Intractability, A Guide to the Theory of NP-Completeness*, W.H. Freeman, New York, 1979.

[7] Grötschel, M., László Lovász and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, New York, 1988.

[8] Hoeffding, W. *Probability Inequalities For Sums of Bounded Random Variables*, American Statistical Association Journal, March 1963, pp. 13-30.

[9] Jayaraman, J., R. Srinivasan, R. Roundy and S. Tayur. *Procurement of Common Components in a Stochastic Environment*, IBM Technical Report, November 1992.

[10] Jerrum, M.R., L. G. Valiant and V.V. Vazirani.*Random Generation of Combinatorial Structures from a Uniform Distribution*, Theoretical Computer Science 43 (1986), 169-188.

[11] Lovász, László and Miklós Simonovits. *The Mixing Rate of Markov Chains, an Isoperimetric Inequality, and Computing the Volume*, 31st IEEE Symposium on the Foundations of Computer Science, 1990, pp. 356-354.

[12] Sinclair, Alistair and Mark Jerrum. *Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains*, Information and Computation 82, pp. 93-133, 1989.

[13] Wets, R. *Stochastic Programming* in Handbooks in Operations Research and Management Science, Vol. 1, North Holland, 1989.