

# SAMPLING AND INTEGRATION OF NEAR LOG-CONCAVE FUNCTIONS

(preliminary version)

David Applegate

Ravi Kannan

**Abstract** An important class of functions that arise in statistics and other areas are the log-concave functions. We give an algorithm for sampling according to log-concave functions. We prove fast convergence of our algorithm using the theory of rapidly mixing Markov chains. We use sampling to develop an algorithm for integrating such functions. Our algorithms work for all functions that assume only positive real values; the running time depends on a natural parameter that measures how close they are to being log-concave. As one application, we are able to develop an algorithm for approximating the volume of convex bodies given by an oracle; we do so by enclosing the given body in a cube, defining a log-concave function that is 1 on the body and exponentially falls off outside and integrating this function. The running time of our algorithm is better than previous algorithms.

## 1 Introduction

Suppose  $F(\cdot)$  is a real positive valued function defined on the cube  $A = \{x \in \mathbf{R}^n : 0 \leq x_i \leq d\}$  in Euclidean  $n$ -space (where  $d$  is some positive integer.) Let  $\gamma$  be a (small) positive real and  $L$  the grid of points

$$L = \{x : x_i = \text{an integer multiple of } \gamma \text{ for } i = 1, 2, \dots, n\}.$$

We consider the following three fundamental problems :

1. **Sampling according to  $F$**  Pick a (random) point of  $L \cap A$  such that the probability  $p(x)$  of picking point  $x$  is proportional to  $F(x)$ . The problem we solve here is the following approximate version : given a positive constant  $\epsilon$ ,  $p(x)$  must satisfy

$$|p(x) - cF(x)| \leq \epsilon \quad \forall x \in L \cap A.$$

for some constant  $c$  independent of  $x$ .

2. **Integration** Find the integral of  $F(x)$  over  $A$ . In the approximate version we solve, we are given an  $\epsilon > 0$  which is an upper bound on the relative error allowed. Since we will give a probabilistic algorithm, we are also given an upper bound  $\epsilon'$  on the probability with which our algorithm is allowed to fail.
3. **Volume of convex bodies** Given a convex body in Euclidean  $n$ -space by a membership oracle, find its volume with relative error at most  $\epsilon$ ; the algorithm should have failure probability at most  $\epsilon'$ .

Our algorithm for the third problem will be based on our solution of the first two problems, so we discuss them first. It is known that for functions  $F$  satisfying certain special conditions, these problems can be solved efficiently, i.e., with running times polynomially

bounded in certain parameters of the problem. In particular, for **concave** functions  $F$ , this is true by the results of Dyer, Frieze and Kannan (1989); they gave the first polynomial time algorithm for solving problem 3) and as well problems 1) and 2) in the case of concave functions. A very important class of functions that occurs in practice are **log-concave** (which is a weaker condition than concavity - it is easy to see that a positive concave function is also log-concave) : i.e., their logarithms are concave, but not necessarily the function itself. Most of the important distributions in statistics have this property : for example, the **normal, exponential, gamma** and many others. In fact, the statistical applications provided our initial motivation. The class of log-concave functions has another important property : it is closed under multiplications.

In this paper, we describe a general random walk and use it to solve problems 1) and 2) efficiently for “smooth” log-concave functions. Our algorithms work in fact for any general function  $F$ , but the running times depend upon two parameters  $\alpha$  and  $\beta$ , the first of which measures how smooth  $F$  is and the second how close it is to being log-concave.

Here are the definitions of  $\alpha$  and  $\beta$ . First we let

$$f(x) = \ln F(x).$$

Since  $F$  is strictly positive on  $A$ ,  $f$  is defined throughout the cube  $A$ . Then  $\alpha$  and  $\beta$  are any real numbers satisfying :

$$|f(x) - f(y)| \leq \alpha \left( \max_i^n |x_i - y_i| \right) \quad \forall x, y \in A.$$

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y) - \beta \quad \forall x, y \in A \text{ and } \forall \lambda \in [0, 1].$$

Note that  $\alpha$  may be taken to be the Lipschitz constant of  $f$  with respect to the infinity norm. Since  $\lambda$  can be zero,  $\beta$  must be nonnegative. If  $f$  is concave,  $\beta$  can be taken to be zero. If  $f$  is sufficiently differentiable, we can also take  $\alpha$  to be  $\sqrt{n}$  times any upper bound on the length of the gradient vector of  $f$  at all points of  $A$ . Similarly  $\beta$  can be defined in terms of the the maximum value of the second derivative of  $f$  over all directions and all points of  $A$ . These considerations are not used in the rest of the paper.

Our algorithm solves problem 1) (namely, draws one sample) in time

$$O(n^2 d^2 \alpha^2 e^{2\beta} (\log \frac{1}{\epsilon} + \log \frac{1}{q(x_0)})).$$

We use our sampling algorithm in a straightforward way to solve the problem of integration in time

$$O\left(\frac{n^5}{\epsilon^2} d^2 \alpha^2 e^{2\beta} \log \frac{n}{\epsilon'} (n \log \frac{d\alpha}{\epsilon} + \log \frac{1}{q(x_0)})\right).$$

We use the algorithms for the first two problems to solve problem 3). This is done by enclosing the given convex body  $K$  in a cube  $A$ . Then we define a log-concave function  $F$  which is 1 everywhere on  $K$  and falls off exponentially outside  $K$ . This enables us to argue that the integral of  $F$  over  $A$  is at least the volume of  $K$  and at most a constant times the volume of  $K$ . We compute this integral using our solution to problem 2) and then sample

using our solution to problem 1) to find the ratio of the volume of  $K$  to the integral. Our algorithm has running time :

$$O\left(\frac{n^{10}}{\epsilon^2}(\log n)^2\left(\log \frac{1}{\epsilon}\right)^2\left(\log \frac{1}{\epsilon'}\right)\left(\log \log \frac{1}{\epsilon'}\right)\right).$$

The paper of Dyer, Frieze, and Kannan was the first to give a polynomial time algorithm. Their running time was  $O(n^{23}(\log n)^5\epsilon^{-2}\log(1/\epsilon)\log(1/\epsilon'))$ . The running time's dependence on  $n$  was improved by Lovász and Simonovits (1990) to  $O(n^{16}(\log n)^6\log(n/\epsilon)\log(n/\epsilon')\epsilon^{-4})$ . The proof of convergence of all our algorithms uses a generalization of a lemma in the Lovász and Simonovits paper. This lemma asserted a lower bound on the surface area of a film inside a convex set in terms of the minimum of the volumes on the two sides of the film in the convex set. Our lemma considers the weighted surface area of the film versus the weighted volume of the two parts - the weight used is the function  $F$ .

## 2 The basic Random Walk

The sampling algorithm is based on the following simple random walk.

We divide the cube  $A$  into small cubes of side  $\delta$ , where  $\delta$  will be  $1/\lceil\alpha\rceil$ . The states of our random walk will be the set of these small cubes. We call this set  $V$ . We call the cube of side  $\delta$  with  $x$  as center  $C_x$ . We interchangeably say the random walk is in  $x$  or  $C_x$ . The random walk will start in some point in  $V$  to be specified later. At a general time, the random walk is in some  $x \in V$ . With probability  $1/2$ , it decides to stay where it is; otherwise, with probability  $1/(2n)$  each it picks one of the  $2n$  coordinate directions. If the adjacent cube in that direction is not in  $A$ , the walk stays where it is. Otherwise, it moves to that adjacent cube with probability

$$\min\left\{1, \frac{F(y)}{F(x)}\right\}$$

where  $y$  is the center of the adjacent cube. More precisely, for any centers  $x, y$  of two different cubes in  $A$ ,  $p_{xy}$  the probability of going from  $x$  to  $y$  is defined by :

- $p_{xy} = 0$  if they do not share a facet
- $p_{xy} = \left(\frac{1}{4n}\right)\left(\min\left\{1, \frac{F(y)}{F(x)}\right\}\right)$  if they do .

Further, the probability  $p_{xx}$  of staying in  $x$  is  $1 - \sum_{y \neq x} p_{xy}$ .

**Proposition** The Markov Chain is connected and aperiodic, i.e., it is ergodic.

**Proof** Follows since  $F$  is positive on  $A$ .

**Lemma** The stationary distribution of the Markov chain  $\{q(x) : x \in V\}$  satisfies :

$$q(x) = F(x) / \sum_{y \in V} F(y).$$

**Proof** It suffices to check that for each  $x$ , we have

$$F(x) = p_{xx}F(x) + \sum_{y \neq x} F(y)p_{yx}.$$

We have  $\sum_{y \neq x} F(y)p_{yx} = (1/(4n)) \sum_y \min\{F(y), F(x)\}$ , where the last sum is over  $y$  such that  $C_y$  is adjacent to  $C_x$ . Now  $p_{xx} = 1 - (1/(4n)) \sum_y \min\{1, F(y)/F(x)\}$ . So the lemma follows. ■

### 3 The basic random walk mixes rapidly

The proof of rapid mixing of the basic random walk is based on the notion of conductance and its relation to the rate of convergence proved by Sinclair and Jerrum (1988). Informally, their notion of conductance of a subset of states measures the “escape probability” from the subset. Their theorem stated below says that if this escape probability is high for each subset of states, then the Markov chain approaches the steady state fast. Here are the precise definitions.

A Markov chain is said to be “time-reversible” if for every pair of states  $x, y$  of the chain, we have  $q(x)p_{xy} = q(y)p_{yx}$ . This is clearly true for our Markov chain, where both the above quantities are either zero or

$$(1/(4n)) \min\{F(x), F(y)\} / \sum_z F(z).$$

Henceforth, when we say Markov chain, we will mean a time-reversible Markov chain.

**Definition** Suppose a Markov chain has a finite set of states  $V$  and  $S$  is subset of  $V$ . The conductance of  $S$  is defined to be

$$\frac{\sum_{x \in S, y \in V \setminus S} q(x)p_{xy}}{\min\left\{\sum_{x \in S} q(x), \sum_{y \in V \setminus S} q(y)\right\}}.$$

The conductance of the Markov chain, denoted  $\phi$  henceforth is defined to be the minimum conductance of a subset  $S$  of states.

**Theorem:** (Sinclair and Jerrum) If the basic random walk is started in state  $x_o$ , then after  $t$  steps, the probability distribution  $p_t(\cdot)$  satisfies

$$|p_t(x) - q(x)| \leq \frac{\sqrt{q(x)}}{\sqrt{q(x_o)}} \left(1 - \frac{\phi^2}{2}\right)^t \quad \forall x \in V.$$

We will also use the following corollary :

**Corollary :** If we start the random walk in a state  $x_o$  with  $\nu q(x_o) \geq q(x) \forall x$ , then we have

$$|p_t(x) - q(x)| \leq \sqrt{\nu} \left(1 - \frac{\phi^2}{2}\right)^t.$$

In order to prove that the conductance of the basic random walk is not too small, we need following generalization of a theorem from Lovász and Simonovits:

**Theorem:** (weighted isoperimetry) Let  $K$  be a convex set in Euclidean  $n$ -space with  $l_\infty$  diameter (the maximum  $l_\infty$  distance between any pair of points)  $d$ . Let  $K_1 \cup B \cup K_2$  be a decomposition of  $K$  into three closed parts such that the  $l_\infty$  distance between any point of  $K_1$  and any point of  $K_2$  is at least  $t$ . Then

$$\min\{\mu(K_1), \mu(K_2)\} \leq \frac{d}{t} e^\beta \mu(B)$$

where  $\mu$  is the measure with weight function  $F$ . (i.e.,  $\mu(T) = \int_T F$  for any measurable subset  $T$  of  $K$ .)

**Proof:** Let  $\tau$  be an arbitrarily small positive real. Let  $E$  be the Löwner-John ellipsoid of  $K$ . The reduction to the “needle-like” case, where all but one of the axes of  $E$  are shorter than  $\tau$  follows that of Lovász and Simonovits, except that one uses the weighted Ham-Sandwich theorem to bisect the weighted  $K_1$  and  $K_2$ , and uses  $F(x) > 0$  for all  $x \in K$ , a bounded closed set, to get a non-zero lower bound on  $F(x)$  to guarantee that the repeated bisection of the largest two axis of the Löwner-John ellipsoid results in a “needle-like” body. The details are simple and are deferred to the final version.

Now, for the “needle-like” case: Let  $a_1$  and  $a_2$  be the endpoints of the projection of  $K$  onto the longest axis of  $E$ . Let  $r = |a_2 - a_1|_\infty$ ,  $e = (a_2 - a_1)/r$ , and let  $H_u$  be the hyperplane through  $a_1 + ue$  orthogonal to this axis.

Assume that  $a_1 \in K_1$ . Assume that (moving from  $a_1$  toward  $a_2$ )  $u_1$  is the last  $u$  for which  $H_u$  contains a point in  $K_1$  and  $u_2$  is the first one for which  $H_u$  contains a point in  $K_2$ . Then  $|u_2 - u_1|_\infty \geq t - \tau\sqrt{n}$ .

Let  $s(u)$  be the  $(n - 1)$ -dimensional volume of  $H_u \cap K$ . We know from the Brunn-Minkowski theorem (see e.g. Bonnesen and Fenchel (1934)) that  $s^{1/(n-1)}$  is a concave function, from this it follows that  $\ln s$  is a concave function. Let  $G(u) = s(u)F(a_1 + ue)$ . Let  $G(z) = \min_{x \in [u_1, u_2]} G(x)$ , and  $G(y) = \max_{x \in [a_1, a_2]} G(x)$ . By symmetry we may assume that  $y \geq z$ .

Let  $x \in [a_1, u_1]$ . We show that  $G(x) \leq G(z)e^\beta$ :

Case 1:  $G(y) \leq G(z)e^\beta$ . Since  $G(y) = \max_{x \in [a_1, a_2]} G(x)$ ,  $G(x) \leq G(y) \leq G(z)e^\beta$ .

Case 2:  $G(y) > G(z)e^\beta$ . With suitable  $0 \leq \lambda \leq 1$  we have:

$$f(z) \geq \lambda f(x) + (1 - \lambda)f(y) - \beta$$

$$\ln s(z) \geq \lambda \ln s(x) + (1 - \lambda) \ln s(y)$$

Adding and noting that  $\ln G(y) > \ln G(z) + \beta$  we must have  $\ln G(x) \leq \ln G(z) + \beta$ . Therefore, since every point  $x$  on  $H_u$  has  $e^{-\alpha\tau\sqrt{n}} \leq \frac{F(x)}{F(a_1+ue)} \leq e^{\alpha\tau\sqrt{n}}$  we have that

$$\mu(K_1) \leq G(z)re^\beta e^{\alpha\tau\sqrt{n}} \quad \text{and} \quad \mu(B) \geq G(z)(t - \tau\sqrt{n})e^{-\alpha\tau\sqrt{n}}.$$

$$\text{Thus, } \mu(K_1) \leq \frac{r}{(t - \tau\sqrt{n})} e^{2\alpha\tau\sqrt{n}} e^\beta \mu(B).$$

Since this is true for each positive  $\tau$ , we have the theorem. ■

**Theorem 3:** The conductance  $\phi$  of the random walk satisfies

$$\phi \geq \frac{1}{4e^2nd\lceil\alpha\rceil e^\beta}$$

**Proof:** Let  $S$  be a subset states. The conductance  $\phi_S$  of  $S$  is

$$\phi_S = \frac{\sum_{x \in S, y \notin S} q(x)p_{xy}}{\min\{\sum_{x \in S} q(x), \sum_{x \notin S} q(x)\}} = \frac{\sum_{x \in S, y \notin S} \min\{F(x), F(y)\}/4n}{\min\{\sum_{x \in S} F(x), \sum_{x \notin S} F(x)\}}.$$

The sum in the last numerator is only over adjacent  $x, y$ . Using  $\alpha$  to bound how much  $F$  varies within a small cube gives:

$$\mu(S) \geq \sum_{x \in S} F(x)e^{-\alpha\delta} \delta^n \quad \text{and} \quad \mu(A \setminus S) \geq \sum_{x \notin S} F(x)e^{-\alpha\delta} \delta^n.$$

Let  $B$  be the  $t/2$ -neighborhood of  $S \cap (A \setminus S)$ . Then  $B$  is made up of facets shared by cubes in  $S$  and  $A \setminus S$ , and

$$\mu(B) \leq \sum_{x \in S, y \notin S} \min\{F(x), F(y)\} e^{\alpha\delta} \delta^{n-1} t$$

Finally, weighted isoperimetry with  $K_1 = S \setminus B$  and  $K_2 = A \setminus (S \cup B)$  yields

$$\frac{\mu(B)}{\min\{\mu(S), \mu(A \setminus S)\}} \geq \frac{t}{de^\beta}$$

$$\text{and thus, } \phi_S \geq \frac{\mu(B)}{4n \min\{\mu(S), \mu(A \setminus S)\}} \frac{\delta}{te^{2\alpha\delta}} \geq \frac{\delta}{4nde^\beta e^{2\alpha\delta}} \geq \frac{1}{4e^2nd\lceil\alpha\rceil e^\beta}.$$

(The last inequality uses the fact that  $\delta = 1/\lceil\alpha\rceil$ .)

## 4 Sampling according to $F$

To sample according to  $F$ , we start at some point  $x_0$ , perform the basic random walk until it has “mixed” to get a random point  $x$ , and then use rejection sampling to pick a point  $z$  from within  $L \cap C_x$ . More precisely, the sampling algorithm to generate a random sample  $z$  from  $A \cap L$  according to  $F$  with error  $\epsilon$  is:

1. Run the basic random walk for  $16e^4n^2d^2\alpha^2e^{2\beta}(\log \frac{e^2}{\epsilon} + \log \frac{1}{q(x_0)})$  steps. Let  $x$  be final state of the walk.
2. Pick a point  $z$  uniformly from  $L \cap C_x$ .
3. With probability  $F(z)/(eF(x))$  output  $z$ . Otherwise, start the random walk over again.

**Proposition** : The sampling algorithm terminates in time

$$O(n^2 d^2 \alpha^2 e^{2\beta} (\log \frac{1}{\epsilon} + \log \frac{1}{q(x_0)}) + n \log \frac{d}{\gamma})$$

**Proof** : The first term is the time needed to do the basic random walk. The second term is the time to generate the  $n \log(d/\gamma)$  bits that describe a point of  $A \cap L$ . The probability of accepting is  $\geq 1/e^2$  since  $\delta = 1/\lceil \alpha \rceil$ , so we expect to reject  $O(1)$  times before we accept a sample. ■

**Proposition** : There exists a  $c$  such that for any  $z \in L \cap A$  the probability  $\Pr[z]$  that the sampling algorithm picks  $z$  satisfies  $|\Pr[z] - cF(z)| \leq \epsilon$ .

**Proof** : The proposition follows directly from the conductance of the basic random walk and the choice of  $\delta$ . Details are contained in the Appendix.

**Corollary** : Let  $S \subseteq L \cap A$ . Then  $|\sum_{x \in S} \Pr[x] - \sum_{x \in S} q(x)| \leq \epsilon |L \cap A|$ .

**Remark** : The algorithm can actually sample according to  $F$  over any rectangle, in which case  $d$  is the length of the longest side.

**Remark** : The algorithm as stated is expected to stop after a constant number of repetitions, but has no worst case bound. To make the algorithm fully polynomially time-bounded, if the algorithm has not accepted after  $\log 1/\epsilon$  repetitions, terminate and output  $x$ . The probability of this happening is  $< \epsilon$ .

## 5 Integration of $F$

To integrate  $F$ , we divide the cube in half. Samples according to  $F$  will then lie on one half with probability equal to that half's integral divided by the total integral. We sample to determine that ratio, and recursively compute the integral of the larger half, to determine the whole integral. We end the recursion when the cube has become small enough that the value at its center point is close to its value over the entire cube. In this situation, uniform sampling is sufficient to approximate the volume.

In more detail, the algorithm to integrate  $F$  over a cube  $A$  with relative error  $\epsilon$  and failure probability  $\epsilon'$  is:

1. Let  $\gamma = \frac{\epsilon}{24\alpha n}$ , and let  $L$  be the corresponding grid of points. Let  $N = |L \cap A| = \left(\frac{d}{\gamma}\right)^n$ .
2. If  $d \leq 1/\lceil \alpha \rceil$ , let  $X$  be a set of  $\frac{\epsilon^2}{\epsilon'} \log \frac{2}{\epsilon'}$  points picked uniformly at random from  $A \cap L$ . Return

$$\frac{\sum_{x \in X} F(x)}{|X|} \text{Vol}(A)$$

3. Otherwise, compute a cube  $A' \subset A$  of side  $d/2$ , for which we know  $R = \frac{\int_{A'} F}{\int_A F}$  to within a relative error of  $\epsilon/4$ . To do this:

- (a) Start with the rectangle  $A^0 = A$  and the ratio  $R^0 = 1$ .
  - (b) For  $i = 1, \dots, n$  split  $A^{i-1}$  into  $A_l^i$  and  $A_r^i$  by splitting  $A^{i-1}$  in half along the  $i$ -th dimension.
  - (c) Let  $X^{i-1}$  be a set of  $(\frac{24n}{\epsilon})^2 \log \frac{2n}{\epsilon'}$  points picked according to  $F$  from  $A^{i-1} \cap L$  by the sampling algorithm with error  $\frac{\epsilon}{24Nn}$ .
  - (d) Either  $|A_l^i \cap X^{i-1}|$  or  $|A_r^i \cap X^{i-1}|$  will be at least  $|X^{i-1}|/2$ . Let  $|A^i|$  be that half of  $A^{i-1}$  and let  $R^i = R^{i-1}|X^{i-1}|/|A^i \cap X^{i-1}|$ .
  - (e) Let  $A' = A^n$  and  $R = R^n$ .
4. Recursively compute the integral of  $F$  over  $A'$  to within a relative error of  $\epsilon/\sqrt{2}$  and failure probability  $\frac{\epsilon'}{2}$ , and return the product  $R \int_{A'} F$ .

**Proposition** : The integration algorithm takes time

$$O(n^5 \frac{1}{\epsilon^2} d^2 \alpha^2 e^{2\beta} \log \frac{n}{\epsilon'} (n \log \frac{d\alpha}{\epsilon} + \log \frac{1}{q(x_0)}))$$

**Proof** : Step 2) takes time  $O(\frac{1}{\epsilon^2} \log(\frac{1}{\epsilon'}) n \log \frac{d}{\gamma})$  and only gets executed once. Step 3c) takes time  $O(\frac{n^2}{\epsilon^2} \log(\frac{n}{\epsilon'})(n^2 d^2 \alpha^2 e^{2\beta} (\log \frac{d}{\gamma} + \log \frac{1}{q(x_0)}) + n \log \frac{d}{\gamma})) = O(n^4 \frac{1}{\epsilon^2} d^2 \alpha^2 e^{2\beta} \log \frac{n}{\epsilon'} (n \log \frac{dn\alpha}{\epsilon} + \log \frac{1}{q(x_0)}))$  and gets executed  $n$  times. Thus,  $T(n, d, \epsilon, \epsilon') = O(n^5 \frac{1}{\epsilon^2} d^2 \alpha^2 e^{2\beta} (n \log \frac{d\alpha}{\epsilon} + \log \frac{1}{q(x_0)}) + T(n, \frac{d}{2}, \frac{\epsilon}{\sqrt{2}}, \frac{\epsilon'}{2}))$  and the result follows. ■

**Proposition** : With probability  $\geq 1 - \epsilon'$  the integration algorithm's result  $V$  satisfies:

$$(1 - \epsilon) \int_A F \leq V \leq (1 + \epsilon) \int_A F$$

**Proof** : This follows from the following theorem from statistics on large deviations (see for example Theorem 1 of Hoeffding (1963)). Details of the proof are contained in the Appendix.

**Proposition** : Let  $S_n = X_1 + \dots + X_n$ , where the  $X_i$  are independent, identically distributed simple random variables with mean  $m$  and standard deviation  $\sigma$ . Then

$$\Pr \left[ \left| \frac{S_n}{n} - m \right| \geq \epsilon m \right] \leq 2e^{-\frac{\epsilon^2 m^2 n}{2\sigma^2}}$$

**Corollary** : Let  $S_n = X_1 + \dots + X_n$ , where the  $X_i$  are simple Bernoulli trials with probability  $p$ . Then

$$\Pr \left[ \left| \frac{S_n}{n} - p \right| \geq \epsilon p \right] \leq 2e^{-\frac{\epsilon^2 pn}{2}}$$

## 6 Computing the volume

For computing the volume of a convex body  $K$ , given by a “weak-membership oracle”, we proceed as follows : Just as in Dyer, Frieze and Kannan, we also first make the body well-rounded. We, however, make one observation that saves us some running time : since we are finally going to subdivide into cubes, it will be better to get a pair of “concentric” cubes one inside the body and one outside after applying a linear transformation to it, rather than get a pair of such spheres. More precisely, we do the following :

- Find  $n + 1$  points  $v_0, v_1, \dots, v_n$  in  $K$  such that the volume of the simplex with these  $n$  points as vertices is at least  $1/2$  of the maximum volume of any simplex with all vertices points of  $K$ . Lenstra (1983) gives an algorithm to accomplish this using the ellipsoid algorithm.
- Apply an affine transformation to space that makes  $v_0$  into the origin and the  $n$  vectors  $v_1, v_2, \dots$  into the  $n$  unit vectors  $e_1, e_2, \dots, e_n$  respectively; this changes the volume of  $K$  by a factor equal to the determinant of the transformation which we remember.

We see that now  $K$  must be contained in the cube

$$A' = \{x : -2 \leq x_i \leq 2 \ \forall i\}$$

for otherwise, we may more than double the volume of the simplex by taking a point that violates one of these inequalities. Also, since the origin and the unit vectors are in  $K$ , we see that the cube  $B$  of side  $1/(2n)$  with center  $p = (1/(2n), 1/(2n), \dots, 1/(2n))$  is contained in  $K$ . For convenience, we apply an affine transformation that sends  $p$  to the origin and  $B$  to a cube of side 2 (so, now  $B$  is the unit ball of the  $l_\infty$  metric). Thus we have that  $K$  contains the unit ball of the  $l_\infty$  metric and is contained in a ball of ( $l_\infty$ ) radius  $4n + 1$  which we call  $A$ .

We now define the following real valued functions on  $A \setminus \{0\}$  :

$$t'(x) = \inf\{r > 0 : x/r \in K\} \quad t(x) = \max\{0, t'(x) - 1\} \quad F(x) = e^{-2nt(x)}.$$

Note that  $F$  is 1 on  $K$  and is positive everywhere on  $A$ .

**Proposition** :  $F$  is log-concave and has  $\alpha = 2n$ .

**Proof** The proof is straightforward and deferred to the final paper.

**Lemma** The integral of  $F$  over  $A$  is at least the volume of  $K$  and at most twice the volume of  $K$ .

**Proof** The first part is obvious. For the upper bound, we have  $\int_A F - \text{Vol}(K)$  is at most

$$\int_0^\infty [\text{Vol}((1+t+dt)K) - \text{Vol}((1+t)K)] e^{-2nt} \leq n \text{Vol}(K) \int_0^\infty (1+t)^{n-1} e^{-2nt} dt.$$

Noting that  $1+t \leq e^t$ , we have the required bound. ■

We estimate the integral of  $F$  over  $A$  using the algorithm of the last section. We start at a state that has  $q(x_o)$  at least  $1/2$  the maximum value of  $q(x)$ , this is possible, since there is a polynomial time algorithm to (nearly) maximize a concave function over a convex set. The details are given in Grötschel, Lovász and Schrijver (1988). For the basic random walk, we would take  $\delta = 1/(2n)$ . In doing the random walk, we must evaluate  $F(x)/F(y)$  for pairs of points  $x, y$  that are  $\delta$  apart. We can only do this with some relative error. We argue in the Appendix that it suffices to make the relative error  $O(\epsilon^{O(1)}/(n^{O(1)}(\log 1/\epsilon')))$ . To evaluate  $F(x)/F(y)$  with relative error  $\xi$  we proceed as follows: we first find (by a simple binary search procedure using queries to the oracle for  $K$ )  $t(x)$  and  $t(y)$  each with absolute error at most  $O(\xi)$ . This takes time  $O(\log(n/\xi))$ . Thus we will then know their difference with absolute error at most  $O(\xi)$ . We find the exponential of this by power series, noting that since the difference is not more than  $O(1)$  in absolute value, we need only  $O(\log(1/\xi))$  terms of the power series to approximate  $F(x)/F(y)$  to relative error at most  $O(\xi)$ . So, each step of the random walk takes times  $O(\log n + \log(1/\xi))$ ; the same argument gives a similar bound on the time required to do the rejection sampling to pick a point of the finer grid of side  $\gamma$ . Using the proposition of the last section that gives the time required for integration, and substituting  $d = O(n)$ ,  $\alpha = O(n)$ ,  $1/q(x_o) = O((d/\delta)^n) = O(n^{2n})$ , we have that the time required to find the integral of  $F$  over  $A$  with relative error  $O(\epsilon)$  is

$$O\left(\left(\frac{n^{10}}{\epsilon^2}\right)(\log n)^2\left(\log \frac{1}{\epsilon}\right)^2\left(\log \frac{1}{\epsilon'}\right)\left(\log \log \frac{1}{\epsilon'}\right)\right).$$

Finally, we let  $\gamma = O(\epsilon/n)$ . Let  $L$  be as usual the grid of side  $\gamma$ . Let  $N = |L \cap A|$ . We have  $\log N = n(\log n - \log \epsilon)$ . With these parameters, we solve problem 1) of sampling in  $L \cap A$  according to the function  $F$  where we take the error allowed (the  $\epsilon$  in the definition of problem 1) to be  $O(\epsilon/N)$ . For  $z \in L \cap A$ , let  $\Pr[z]$  be the probability of picking  $z$ . Let  $\Pr[K]$  be  $\sum_{z \in L \cap K} \Pr[z]$ . Finally, let  $X = \sum_{z \in L \cap A} F(z)$ . Then by the corollary in section 4, we have

$$|\Pr[K] - |L \cap K|/X| \leq O(\epsilon).$$

So, by doing  $O((1/\epsilon^2)(\log \frac{1}{\epsilon}))$  trials, we may estimate  $|L \cap K|/X$  to relative error  $O(\epsilon)$  with failure probability  $\epsilon'$ . (The proof of this fact uses the same result on large deviations as in section 5 and we defer this proof to the final paper.) To finish the proof, we need two facts:  $\gamma^n X$  is close to the integral of  $F$  over  $A$  which follows easily and secondly we prove below that  $\gamma^n |L \cap K|$  is close to the volume of  $K$ :

If  $K$  intersects any cube of side  $\gamma$ , then  $(1 + \gamma)K$  contains the center and conversely, if  $K$  contains the center, but not the whole cube, then  $(1 - \gamma)K$  misses the cube. So we have

$$(1 - \gamma)^n \text{Vol}(K) \leq \gamma^n |L \cap K| \leq (1 + \gamma)^n \text{Vol}(K).$$

The time for the sampling part does not change the total time since it is easily seen to be at most a constant times the time for the integration.

## 7 References

T.Bonnesen and W.Fenchel, *Theorie der konvexen Körper*, Springer, Berlin.

M.Dyer, A.Frieze, and R.Kannan, *A random polynomial time algorithm for approximating the volume of convex bodies*, Proceedings of the 21st Annual ACM Symposium on Theory of Computing (1989), pp. 375-381.

M.Grötschel, L.Lovász and A.Schrijver, *Geometric algorithms and combinatorial optimization*, Springer-Verlag (1988).

W.Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association 58 (1963) 13-30.

M.R.Jerrum and A.J.Sinclair, *Conductance and the rapid mixing property for Markov chains: the approximation of the permanent resolved*, Proceedings of the Twentieth Annual ACM Symposium on the Theory of Computing (1988) 235-244.

H.W.Lenstra, *Integer programming with a fixed number of variables*, Mathematics of Operations research, Volume 8, Number 4 Nov (1983) pp 538-548

L. Lovász and M. Simonovits, *The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume*, Proceedings of the Thirty-First IEEE Symposium on Foundations of Computer Science (1990) pp. 346-355

## Appendix

**Proposition** : There exists a  $c$  such that for any  $z \in L \cap A$  the probability  $\Pr[z]$  that the sampling algorithm picks  $z$  satisfies  $|\Pr[z] - cF(z)| \leq \epsilon$ .

**Proof** : The probability that a pass picks the point  $z$  is equal to the probability that it picks the square  $C_x$  containing  $z$ , and picks  $z$  from within that square, and accepts  $z$ , or

$$\Pr[z \text{ on this pass}] = p_t(x) \frac{1}{|C_x \cap L|} \frac{F(z)}{eF(x)}$$

From the conductance of the basic random walk, we have

$$|p_t(x) - q(x)| \leq \frac{\sqrt{q(x)}}{\sqrt{q(x_0)}} \left(1 - \frac{\phi^2}{2}\right)^t \leq \frac{1}{\sqrt{q(x_0)}} \left(1 - \frac{\phi^2}{2}\right)^{\frac{1}{\phi^2}(\log \frac{\epsilon^2}{c} + \log \frac{1}{q(x_0)})} \leq \frac{\epsilon}{e^2}$$

$$\text{so } \Pr[z \text{ on this pass}] \leq \left(\frac{F(x)}{\sum_{y \in V} F(y)} + \frac{\epsilon}{e^2}\right) \frac{1}{|C_x \cap L|} \frac{F(z)}{eF(x)} \leq F(z) \frac{1}{|C_x \cap L| e \sum_{y \in V} F(y)} + \frac{\epsilon}{e^2}$$

$$\text{and similarly, } \Pr[z \text{ on this pass}] \geq F(z) \frac{1}{|C_x \cap L| e \sum_{y \in V} F(y)} - \frac{\epsilon}{e^2}$$

$$\text{Since } \Pr[z] = \frac{\Pr[z \text{ on this pass}]}{\Pr[\text{somebody on this pass}]} \quad \text{we have}$$

$$\left| \Pr[z] - F(z) \frac{\Pr[\text{somebody on this pass}]}{|C_x \cap L| e \sum_{y \in V} F(y)} \right| \leq \epsilon$$

This finishes the proof, with  $c = \frac{\Pr[\text{somebody on this pass}]}{|C_x \cap L| e \sum_{y \in V} F(y)}$ . ■

**Proposition** : With probability  $\geq 1 - \epsilon'$  the integration algorithm's result  $V$  satisfies:

$$(1 - \epsilon) \int_A F \leq V \leq (1 + \epsilon) \int_A F$$

**Proof** : If  $d \leq 1/\lceil \alpha \rceil$ ,  $V$  is determined by step 2). In that case, each  $F(x)$  is an independent random variable with mean  $m = \sum_{y \in L \cap A} F(y)/N$  and variance  $\sigma^2 = \sum_{y \in L \cap A} F(y)^2/N - m^2 \leq \sum_{y \in L \cap A} m^2 e^{2d\alpha}/N - m^2 \leq e^2 m^2$ . Thus

$$\Pr \left[ \left| \frac{\sum_{x \in X} F(x)}{|X|} - m \right| \geq \frac{\epsilon}{2} m \right] \leq 2e^{-\frac{\epsilon^2 m^2 |X|}{2\sigma^2}} \leq \epsilon'$$

$$\text{and since } \frac{\sum_{x \in L \cap A} F(x)}{N} e^{-\alpha\gamma} \leq \frac{\int_A F}{\text{Vol}(A)} \leq \frac{\sum_{x \in L \cap A} F(x)}{N} e^{\alpha\gamma} \quad \text{we get}$$

$$\Pr \left[ \left| V - \int_A F \right| \geq \epsilon \int_A F \right] \leq \epsilon'$$

If  $d > 1/\lceil \alpha \rceil$ ,  $V$  is determined by step 3) and the recursive call. We will show that for each  $i$ , with probability at least  $1 - \frac{\epsilon'}{2n}$  the relative difference between  $\frac{|X^{i-1}|}{|A^i \cap X^{i-1}|}$  and  $\frac{\int_{A^{i-1}} F}{\int_{A^i} F}$  is less than  $\frac{\epsilon}{4n}$ . This will mean that with probability at least  $1 - \frac{\epsilon'}{2}$  the relative error in  $R$  will be less than  $\frac{\epsilon}{4}$ , and the result follows.

At each  $i$ , each point  $x \in X^{i-1}$  can be viewed as a Bernoulli trial, where a success is when  $x \in A^i$ . Thus, letting  $p(A^i)$  denote the probability that the sampling algorithm selected a point in  $A^i$ , we have

$$\Pr \left[ \left| \frac{|X^{i-1} \cap A^i|}{|X^{i-1}|} - p(A^i) \right| \geq \frac{\epsilon}{12n} p(A^i) \right] \leq 2e^{-\frac{\epsilon^2 p(A^i) |X^{i-1}|}{12^2 n^2}} \leq \frac{\epsilon'}{2n}$$

since we chose  $A^i$  so that  $p(A^i) \geq 1/2$ .

Now, by the corollary to the sampling theorem, we know that  $p(A^i)$  has relative error  $\leq \frac{\epsilon}{12n}$  from  $\frac{\sum_{x \in A^i} F(x)}{\sum_{x \in A^{i-1}} F(x)}$ , and by our choice of  $\gamma$ , we know that this has relative error  $\leq \frac{\epsilon}{12n}$  from  $\frac{\int_{A^i} F}{\int_{A^{i-1}} F}$ . Combining these all, we find that  $\frac{|X^{i-1} \cap A^i|}{|X^{i-1}|}$  approximates  $\frac{\int_{A^i} F}{\int_{A^{i-1}} F}$  with a relative error  $\leq \frac{\epsilon}{4n}$  with probability of failure  $\leq \frac{\epsilon'}{2n}$ . ■

The following proposition can be directly used to show that it suffices to estimate  $F(x)/F(y)$  for the random walk in section 6 to relative error  $O(\epsilon^{O(1)}/(n^{O(1)}(\log(1/\epsilon'))^{O(1)}))$ . To derive this, we just substitute  $t =$  number of steps for which the random walk is executed  $= O(\epsilon^{O(1)}/(n^{O(1)}(\log(1/\epsilon'))^{O(1)}))$  into the proposition.

**Proposition :** Suppose  $P$  is the transition matrix of a Markov chain (with row sums = 1) with column sums at most  $c$ . Suppose  $P_1$  is another transition matrix of a Markov chain on the same states with  $|P_1 - P|_\infty \leq \epsilon_1$ . Then for any natural number  $t$ ,  $|P_1^t - P^t|_\infty \leq ct\epsilon_1$ .

**Proof :** By induction on  $t$ . ■