Random Structures and Alg. **4**, 359–412.

RANDOM WALKS IN A CONVEX BODY AND AN IMPROVED VOLUME ALGORITHM

L. Lovász
Department of Computer Science, Eötvös University,
Budapest, Hungary H-1088,
and
Department of Computer Science, Princeton University,
Princeton, NJ 08544

M. Simonovits
Mathematics Research Institute, Hungarian Academy of Sciences
Budapest, Hungary H-1053

Abstract. We give a randomized algorithm using $O(n^7 \log^2 n)$ separation calls to approximate the volume of a convex body with a fixed relative error. The bound is $O(n^6 \log^4 n)$ for centrally symmetric bodies and for polytopes with a polynomial number of facets, and $O(n^5 \log^4 n)$ for centrally symmetric polytopes with a polynomial number of facets. We also give an $O(n^6 \log n)$ algorithm to sample a point from the uniform distribution over a convex body.

Several tools are developed that may be interesting on their own. We extend results of Sinclair–Jerrum (1988) and the authors (1990) on the mixing rate of Markov chains from finite to arbitrary Markov chains. We also analyze the mixing rate of various random walks on convex bodies, in particular the random walk with steps from the uniform distribution over a unit ball.

0. Introduction.

a. Survey of results.

Computing the volume of a (high-dimensional) convex body is an ancient, basic, but extremely difficult task. In fact, there are negative results in this direction: Bárány–Füredi (1986), improving a result of Elekes (1986) proved that if the convex body is given by a separation oracle (a natural framework which allows the polynomial-time solution of many algorithmic problems), then any algorithm that approximates the volume within a factor of $n^{o(n)}$ necessarily takes exponential time. Dyer–Frieze (1988) and Khachiyan (1988, 1989) showed that the problem of computing the volume exactly (deterministically) is # P-hard, even for explicitely described polytopes. (A strong improvement of this is

due to Brightwell and Winkler (1990), who proved that the exact determination of the number of linear extensions of a poset – which is the volume of a related polytope – is also # P-hard.)

A breakthrough in the opposite direction is due to Dyer, Frieze and Kannan (1989) [DFK], who designed a polynomial time *randomized* algorithm to approximate the volume of a convex body K in $\mathbb{R}^n$. Their algorithm has two input parameters $\varepsilon, \delta > 0$, and computes a random variable ζ such that with probability at least $1-\delta$, the volume of K is between $(1-\varepsilon)\zeta$ and $(1+\varepsilon)\zeta$. Several improvements of the original algorithm have been given. The following list describes these improvements. Our main interest is the dependence of the running time on the dimension n, and we include a column to indicate this. (The * after the O means that we suppress factors of $\log n$, as well as factors depending on the error bounds ε, δ and the "badness" of the original data.)

Dyer, Frieze, Kannan	(1989)	$O^*(n^{23})$	$O(n^{23}\log^5 n\varepsilon^{-2}\log(1/\varepsilon)\log(1/\delta))$
Lovász and Simonovits	(1990)	$O^*(n^{16})$	$O(n^{16}\log^5 n\varepsilon^{-4}\log(n/\varepsilon)\log(n/\delta))$
Applegate and Kannan	(1990)	$O^*(n^{10})$	$O(n^{10}\log^2 n\varepsilon^{-2}\log^2(1/\varepsilon)\log(1/\delta)\log\log(1/\delta))$
Lovász	(1991)	$O^*(n^{10})$	
Dyer and Frieze	(1991)	$O^*(n^8)$	$O(n^8\varepsilon^{-2}\log(n/\varepsilon)\log(n/\delta))$
Lovász and Simonovits (this paper)	(1991)	$O^*(n^7)$	$O(n^7\log^2 n\varepsilon^{-2}\log^3(1/\varepsilon)\log(1/\delta))$

It is interesting to give a short survey of those ideas whose various combinations lead to these improvements. The Dyer–Frieze–Kannan algorithm consists of two main phases: the first, preliminary phase makes the body "reasonable round" by applying an affine transformation to the given body K so that the image contains the unit ball B and is contained in the concentrical ball with radius $n^{3/2}$; this is achieved by a known application of the ellipsoid method (see Grötschel, Lovász and Schrijver (1988) [GLS]). Then a sequence of bodies $K_0 = B \subseteq K_1 \subseteq \ldots \subseteq K_k = K$ is constructed so that the ratio of the volumes of consecutive bodies is at most 2; these ratios are then estimated by a Monte-Carlo method, sampling from K_{i+1} and counting how often K_i is hit.

The most important ingredient is an algorithm to generate a random point from the uniform distribution over a convex body. This is achieved by taking a random walk on the lattice points inside the body, and stopping after an appropriately large (but polynomial) number of steps. The analysis of the algorithm depends on two factors: a theorem of Sinclair and Jerrum (1988) [SJ] on the mixing rate of time-reversible Markov chains and on an isoperimetric inequality for subsets of a convex body.

The arguments are complicated by singularities on the surface of the body in two ways: first, they mean "corners " that the random walk will reach too slowly and second, they weaken the isoperimetric inequality, whose proof by Dyer, Frieze and Kannan depends on methods from differential geometry. Dyer, Frieze and Kannan get around this difficulty by approximating the body by another convex body with a smooth surface, at the cost of a substantial increase in the running time, which amounts to the solution of $O^*(n^{23})$ convex programs, or (using ellipsoid-type methods for the solution of these programs), to $O^*(n^{27})$ membership tests in K.

Khachiyan–Karzanov [KK] and Lovász–Simonovits [LS] proved the isoperimetric inequality in a best possible form (up to a constant). [LS] gives a new, more elementary proof

method, which facilitates further generalizations. [LS] also contains a generalization of the Sinclair–Jerrum result, by allowing small exceptional sets, and by using information about the initial distribution in estimating the convergence speed to the uniform distribution. These ideas improve the running time to $O^*(n^{16})$ membership tests.

Applegate and Kannan [AK] suggest another way to handle some of the difficulties caused by non-smoothness: consider volume computation as the integration of the characteristic function of the body, and approximate this characteristic function by a smooth function. This requires the extension of the method of sampling from a uniform distribution over a convex body to sampling from a distribution with a log-concave density function. They show that the random walk technique, combined with the so-called "Metropolis rule", gives a fast sampling procedure. The proof involves an extension of the isoperimetric inequality from the usual volume (Lebesgue measure) to measures with a log-concave density function.

A further improvement comes from the simple but elegant observation that by "sandwiching" the body between two concentrical cubes instead of two concentrical balls, a ratio of $O(n)$ can be achieved instead of the ratio of $O(n^{3/2})$. (The running time includes the square of this ratio.) They give an $O^*(n^{10})$ implemetation of these ideas.

Dyer and Frieze [DF] show that the Applegate–Kannan method can be further improved by two further ideas: using the improvement of the Sinclair–Jerrum estimate as mentioned above, and by showing that the errors being almost independent, they accumulate in a slower rate. They improve the running time to $O^*(n^8)$ membership tests. They also obtain the best possible constant in the isoperimetric inequality.

All the previously mentioned methods are based on random walks on lattice points. Lovász [L] sketches the analysis of a random walk where each step is chosen from the uniform distribution over the unit ball centered at the current position. Somewhat surprisingly, the analysis depends on the same isoperimetric inequality. Replacing the random walk on lattice points by this, but leaving the other ingredients of the algorithm in [LS] essentially intact, improves the running time to $O(n^{10})$.

In this paper we describe a randomized volume algorithm which requires $O^*(n^7)$ membership tests. Our algorithm builds on virtually all of the ideas mentioned above; yet we feel that the combination is very natural and leads to an algorithm which is conceptually quite simple.

A first observation is that once we replace the volume computation by integration, the integrand need not approximate the characteristic function of the body: only its integral has to approximate the volume. This enables us to use much smoother functions, and a large ball as the domain of integration, which eliminates most of the errors due to boundary effects.

Another, unexpected but important benefit is that in the first phase, only "approximate sandwiching" is needed: it suffices to achieve by an affine transformation that 2/3 of the volume of the unit ball B is contained in K and 2/3 of the volume of K is contained in the "almost circumscribed" ball mB. We show that this can be achieved in randomized polynomial time with $m=n$, and in special cases even better (e.g. for centrally symmetric polyhedra, $m=O(\log n)$). Since m^2 is a factor in the running time, this leads to substantial improvements: $O^*(n^6)$ for centrally symmetric bodies or for polyhedra with polynomially

many facets, and $O^*(n^5)$ for centrally symmetric polyhedra with polynomially many facets. We suspect that the ratio $m = n$ is not optimal even for general convex bodies; as far as we can tell, even $m = O(\log n)$ is possible. This could lead to substantial improvement in the running time.

While our volume algorithm is basically simple, we have to do some more extensive preparations. One of these consists of extending the Sinclair–Jerrum result and its generalizations to Markov chains whose state space is an arbitrary measure space (Section 1). These extensions are basically rather straightforward from the discrete case, especially having found the right formulation (see Feller (1968) for treatment of general Markov chains).

Section 2 introduces log-concave functions and gives the proof of an even more general version of the isoperimetric inequality. While the Applegate-Kannan form would suffice to prove the error estimates for our main algorithm (with a little worse bound), the more general form is more convenient to apply and has some corollaries that are interesting also from a geometric point of view.

Section 3 discusses the issue of generating a random point from a distribution with a log-concave density function. While the volume algorithm involves very "nice" functions only, we derive general bounds on the mixing rates of various random walks on general convex bodies. These bounds depend on isoperimetric inequalities; while these inequalities use the basic isoperimetric inequality mentioned above, they may be of interest on their own right. Our sampling algorithm is "truely" polynomial, i.e., polynomial in both n and in $\log(1/\varepsilon)$, where ε is the error bound.

Section 4 contains the volume algorithm and its analysis.

To conclude this introduction we have to remark that we do not take the shortest route to obtain an $O^*(n^7)$ volume algorithm. Substantial parts of our preparations aim at gaining factors of $\log n$ or $\log(1/\varepsilon)$, or simply mathematical beauty. Thus, we could restrict ourselves to finite Markov chains; s-conductance could be left out; the central limit theorem and its treatment in the Hilbert space could be replaced by the results of 1.c, at the cost of a factor of $\log n$; it would suffice to use the isoperimetric inequality of Applegate and Kannan instead of the results in 2.b, at the cost of a factor of $1/\varepsilon$; and we could restrict ourselves to stepping in a unit ball in Section 3.

b. Computational model.

There is no standard finite encoding of a general convex body; various subclasses may be encoded as solution sets of systems of linear inequalities (polyhedra), convex hulls of lists of vectors (polytopes), level sets of concave functions (e.g., unit balls of norms) etc. A general approach is to consider an *oracle*, i.e., a black box that answers certain queries about the body. The most natural oracle is a *membership oracle*, which returns, for every query $x \in \mathbb{R}^n$, the answer "YES" or "NO" to the question "Is $x \in K$?". This is, however, sometimes too strong, sometimes too weak. For algorithmic purposes, the most usual way to describe a convex body K as an input is a *well-guaranteed weak separation oracle*. We refer to Grötschel, Lovász and Schrijver (1988) for a discussion of the relative strengths of these oracles; under rather reasonable technical assumptions, these are equivalent.

Definition (Weak separation oracle). For any $y \in \mathbb{Q}^n$ and error tolerance $\delta > 0$, the oracle returns either "YES" or "NO". The "YES" answer means that the distance of y

from K is at most δ; the "NO" answer means that the distance of y from $\mathbb{R}^n \setminus K$ is less than δ. In this case, it also returns a "proof" of this fact, in the form a hyperplane $c^T x \le \gamma$ through y which almost separates y from K in the sense that

$$\max\{c^T x : x \in K\} \le \gamma + \delta|c|.$$

(So if y is near the boundary of K, then either answer is legal.)

In addition, we assume that we know the radius r of some ball contained in K (but we do not necessarily know the center of the ball) and the radius R of another ball, with center 0, containing K. (In [GLS] this assumption is phrased as "the oracle is well guaranteed".) The number of bits needed to describe both of these balls is part of the input size. Without loss of generality, we may assume that this contributes $|\log R| + |\log r|$ bits to the input size.

We remark that for the main part of our algorithm, a *weak membership oracle* (defined by relaxing the weak separation oracle in the obvious way) suffices.

Notation and preliminaries.

B denotes the euclidean unit ball, $\mathrm{vol}(K)$ is the volume (Lebesgue measure) of the set K, and $\mathrm{conv}(X)$ denotes the convex hull of the set $X \in \mathbb{R}^n$.

We shall repeatedly use the following fact.

0.1. Lemma. *Let H be a halfspace in $\mathbb{R}^n$ and B, a unit ball whose center is at a distance t from H. (So we speak of the halfspace not containing the center.) Then*

(a) *if $t \le 1/\sqrt{n}$, then*

$$\mathrm{vol}(H \cap B) > \left(\frac{1}{2} - \frac{t\sqrt{n}}{2}\right)\mathrm{vol}(B);$$

(b) *if $t > 1/\sqrt{n}$ then*

$$\frac{1}{10t\sqrt{n}}(1-t^2)^{(n+1)/2}\mathrm{vol}(B) \;<\; \mathrm{vol}(H \cap B) \;<\; \frac{1}{t\sqrt{n}}(1-t^2)^{(n+1)/2}\mathrm{vol}(B).$$

∎

1. Conductance of Markov chains.

a. Markov schemes and Markov chains.

We extend the theory of conductance and rapid mixing from the finite case to arbitrary Markov chains. See Halmos (1974) for fundamentals of measure theory, and [LS] for the discrete versions of some of these results.

Let $(\Omega, \mathcal{A})$ be a σ-algebra. For every $u \in \Omega$, let P_u be a probability measure on Ω, and assume that for every $A \in \mathcal{A}$, the value $P_u(A)$ is measurable as a function of u. We call the triple $\mathcal{M} = (\Omega, \mathcal{A}, \{P_u : u \in \Omega\})$ a *Markov scheme*. A Markov scheme, together with an

initial distribution Q_0 on Ω, defines a *Markov chain*, i.e. a sequence of random variables $w_0, w_1, w_2, \ldots$ with values from Ω such that w_0 is chosen from distribution Q_0 and w_{i+1} is chosen from distribution P_{w_i} (independently of the value of $w_0, \ldots, w_{i-1}$). So we have

$$\text{Prob}(w_{i+1} \in A \mid w_1 = u_1, \ldots, w_i = u_i) = \text{Prob}(w_{i+1} \in A \mid w_i = u_i) = P_{u_i}(A)$$

for every $u_1, \ldots, u_i \in \Omega$ and $A \in \mathcal{A}$.

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f : \Omega \times \Omega \to \mathbb{R}$ be an integrable function (with respect to the product measure $\mu \times \mu$) such that $\int_\Omega f(u,v) d\mu(v) = 1$ for all $u \in \Omega$. Then f defines a Markov scheme by

$$P_u(A) = \int_A f(u,v) d\mu(v).$$

If such a function f exists, then we call it the *transition function* of the Markov scheme (with respect to the measure μ). The transition function is *symmetric* if $f(x,y) = f(y,x)$. We could describe (somewhat artificially) our volume algorithm using Markov chains having transition functions, but some related Markov chains – which could conceivably replace them – do not have. We give the proofs for the general case, but the reader may find them easier to follow in terms of transition functions.

A probability measure Q on $(\Omega, \mathcal{A})$ is a *stationary distribution* for the Markov scheme if choosing w_0 from this distribution, w_1 will have the same distribution (then, of course, so does every w_i). This is equivalent to saying that for all $A \in \mathcal{A}$,

$$\int_\Omega P_u(A)\, dQ(u) = Q(A).$$

From now on we shall fix one stationary distribution Q. (We are not concerned here with the existence of such a distribution; this will exist and in fact be explicitly given in the applications we consider. In all cases relevant for us, the uniqueness of the stationary distribution will be implied e.g. by Theorem 1.4 below.)

In the theory of finite Markov chains, matrices and their eigenvalues play a central role. To extend some results from finite to general Markov chains, we consider the Hilbert space $L^2 = L^2(\Omega, \mathcal{A}, Q)$ with the inner product

$$\langle f, g \rangle = \int_\Omega f g\, dQ.$$

Every Markov scheme defines a positive linear operator $M : L^2 \to L^2$ by

$$(Mf)(u) = \int_\Omega f(v)\, dP_u(v).$$

So $(Mf)(u)$ is the expected value of $f(w_{i+1})$, given that $w_i = u$. More generally, $(M^k f)(u)$ is the expected value of $f(w_{i+k})$, given that $w_i = u$.

Consider a Markov chain where the first element (and hence the other elements as well) are from the stationary distribution. In terms of the Hilbert space, we have for any function $f \in L^2$,

$$\mathbf{E}(f(w_i)) = \mathbf{E}(f(w_0)) = \langle f, 1\rangle,$$
$$\mathbf{E}(f(w_i)^2) = \mathbf{E}(f(w_0)^2) = \langle f, f\rangle,$$
$$\mathbf{E}(f(w_i)f(w_{i+k})) = \mathbf{E}(f(w_0)f(w_k)) = \langle f, M^k f\rangle.$$

A Markov scheme is *time-reversible* if (roughly speaking) for any two sets $A, B \in \mathcal{A}$, it steps from A to B as often as from B to A. Formally, this means that

$$\int_B P_u(A)\, dQ(u) = \int_A P_u(B)\, dQ(u). \tag{1.1}$$

It is easy to see that it suffices to require this relation for disjoint sets A and B. Condition (1.1) can be written as

$$\int_A \int_B 1\, dP_u(v)\, dQ(u) = \int_B \int_A 1\, dP_u(v)\, dQ(u),$$

implying that for any function $F : \Omega \times \Omega \to \mathbb{R}$ for which the integrals exist, we have

$$\int_\Omega \int_\Omega F(u,v)\, dP_u(v)\, dQ(u) = \int_\Omega \int_\Omega F(v,u)\, dP_u(v)\, dQ(u). \tag{1.2}$$

Another equivalent formulation is that the operator M is self-adjoint. If the Markov scheme can be described by a transition function f with respect to Q, then time-reversibility is equivalent to the symmetry of the transition function.

If the Markov scheme is time-reversible, then for any function $f \in L^2$ we have (using (1.2))

$$\langle f, f\rangle - \langle f, Mf\rangle = \frac{1}{2}\int_\Omega \int_\Omega (f(x) - f(y))^2\, dP_y(x)\, dQ(y) \geq 0, \tag{1.3}$$

Equality holds here for a constant function. Thus the spectral radius of M is exactly 1.

b. Laziness of Markov chains.

We call a Markov scheme *lazy* if $P_u(u) \geq 1/2$ at each node. This condition is technical; its main adventage is that (in the time-reversible case) it implies that the operator M associated with the Markov scheme is positive semidefinite. In the discrete case this follows from the fact that if the diagonal elements of a symmetric matrix majorize the sum of absolute values of the rest of their row, then the matrix is positive semidefinite. To see the positive definiteness in general, observe that $2M - I$ is also a self-adjoint operator associated with a Markov scheme, and hence for any function $f \in L^2$, we have

$$\langle f, Mf\rangle = \frac{1}{2}\langle f, f\rangle - \frac{1}{2}\langle f, (2M - I)f\rangle \geq 0.$$

Every Markov scheme can be made lazy by simply tossing a coin at each step and making a move only if it is tails. So (at the cost of a little slow-down) we can assume that M is positive semidefinite, which will be very convenient. One nice consequence of laziness is that if the chain is time-reversible and we generate a Markov chain from the stationary distribution, then the elements of the chain are "positively correlated" in the following sense:

1.1. Lemma. *Let $w_1, w_2, \ldots$ be a time-reversible Markov chain generated by a lazy Markov scheme $\mathcal{M}$ with w_0 from the stationary distribution Q. Then for any function $f \in L^2$,*

$$\mathbf{E}(f(w_i)f(w_j)) \geq \mathbf{E}(f(w_i))\mathbf{E}(f(w_j)) = \mathbf{E}(f(w_0))^2.$$

Proof. Let, say $j > i$. Since all the w_i have the same distribution Q by the definition of stationary distribution, we have here

$$\mathbf{E}(f(w_i)f(w_j)) = \mathbf{E}(f(w_0)f(w_{j-i})) = \langle f, M^{j-i}f\rangle \geq 0,$$

since M is positive semidefinite. Applying this inequality to the function $f - \mathbf{E}(f(w_0))$, we obtain the lemma. ∎

c. Conductance and rapid mixing.

We define the *ergodic flow* $\Phi : \mathcal{A} \to [0,1]$ of the Markov scheme by

$$\Phi(A) = \int_A P_u(\Omega \setminus A)\, dQ(u).$$

This value is the probability of the event that choosing w_0 from the stationary distribution, we have $w_0 \in A$ but $w_1 \notin A$. From the assumption that Q is stationary, we get

$$\begin{aligned}
\Phi(A) - \Phi(\Omega \setminus A) &= \int_A P_u(\Omega \setminus A)\, dQ(u) - \int_{\Omega\setminus A} P_u(A)\, dQ(u) \\
&= \int_A (1 - P_u(A))\, dQ(u) - \int_{\Omega\setminus A} P_u(A)\, dQ(u) \\
&= Q(A) - \int_A P_u(A)\, dQ(u) - \int_{\Omega\setminus A} P_u(A)\, dQ(u) \\
&= Q(A) - \int_\Omega P_u(A)\, dQ(u) = 0\,.
\end{aligned}$$

Note that this computation also works backward: if Q' is any probability distribution on Ω such that the set-function

$$\Phi'(A) = \int_A P_u(\Omega \setminus A)\, dQ'(u)$$

is invariant under complementation, then Q' is stationary. This observation provides a sometimes convenient way to verify that a given distribution is stationary.

The *conductance* of the Markov scheme is

$$\Phi = \inf_{0<Q(A)<1/2} \frac{\Phi(A)}{Q(A)};$$

for every $0 \leq s \leq 1$, the s-conductance is defined by

$$\Phi_s = \inf_{s<Q(A)\leq 1/2} \frac{\Phi(A)}{Q(A) - s}.$$

We call the value $1-P_u(u)$ the *local conductance* of the Markov scheme at element u. If u is an atom (i.e., $Q(u)>0$), then this is just $\Phi(u)/Q(u)$. So in this case the local conductance is an upper bound on the conductance. More generally, let

$$H_t = \{u \in \Omega :\ P_u(u) > 1-t\},$$

and let $s = Q(H_t)$. Then

$$\Phi(H_t) = \int_{H_t} P_u(\Omega \setminus H_t)\, dQ(u) < tQ(H_t).$$

As a consequence, the $(s/2)$-conductance of the scheme is at most $2t$.

Let Q_k denote the distribution of w_k, i.e., let

$$Q_k(A) = \mathrm{Prob}(w_k \in A)$$

for $A \in \mathcal{A}$. By definition, we have the recurrence

$$Q_k(A) = \int_\Omega P_u(A)\, dQ_{k-1}(u).$$

It is well-known (and will also follow from our results below) that if $\Phi(A)>0$ for all $A \in \mathcal{A}$ with $Q(A)>0$, then $Q_k \to Q$ in the ℓ_1 distance. Our main point will be to give a bound on the rate of convergence.

To measure the speed of convergence, we consider all measurable functions

$$g :\ \Omega \to [0,1] \quad \text{with} \quad \int_\Omega g dQ(u) = x,$$

and define the *distance function* of Q and Q_k by

$$h_k(x) = \sup_g \int_\Omega g\, dQ_k - x = \int_\Omega g[dQ_k - dQ],$$

where the supremum is extended over all these functions. (We shall see below that this supremum is always attained.) In the case of a finite Markov scheme with N atoms and with uniform stationary distribution, $h_k(j/N)$ can be obtained by adding up the j largest "errors" $Q_k(\omega) - (1/N)$.

It follows easily from the definition that $h_k(x)$ is a concave function of x. Since $0 \le h_k(0) \le 1$ and $h_k(1) = 0$, we have $0 \le h_k(x) \le 1-x$.

This somewhat artificial definition will be clearer from the following lemma.

1.2. Lemma. *(i) For every set $A \in \mathcal{A}$ with $Q(A) = x$, we have*

$$-h_k(1-x) \le Q_k(A) - Q(A) \le h_k(x).$$

(ii) If Q is atom-free, then

$$h_k(x) = \sup_{\substack{A\in\mathcal{A}\\ Q(A)=x}} \big[Q_k(A)-Q(A)\big].$$

(iii) For every $0<x<1$, there exists a function G that is 0-1 valued except possibly on a Q-atom, attaining the supremum in the definition of $h_k(x)$.

Proof. The upper bound in (i) is immediate if we notice that the incidence function of A is one of the functions g considered in the definition of $h_k(Q(A))$. The lower bound follows by complementation. Assertion (ii) follows from assertion (iii) immediately, so it suffices to prove (iii).

Let $g\colon \Omega\to[0,1]$ be any measurable function with $\int_\Omega g dQ = x$. Let U be a measurable subset of Ω with $Q(U)=0$ and $Q_k(U)$ maximum (such a subset clearly exists). Let Q' and Q'_k be the restriction of Q and Q_k to $\Omega\setminus U$. Then Q'_k is absolutely continuous with respect to Q', so the Radon-Nikodym derivative $\phi = dQ'_k/dQ'$ of Q'_k with respect to Q' exists.

Let, for $t\ge 0$, $A_t = U\cup\{u\in\Omega\setminus U : \phi(u)\ge t\}$ and $s=\inf\{t\colon\ Q(A_t)\le x\}$. Note that $A_s = \cap\{A_t\colon\ t<s\}$. Let $A' = \cup\{A_t\colon\ t>s\}$, then $A'\subseteq A_s$ and $\phi(u)=s$ for every $u\in A_s\setminus A'$. Moreover, $Q(A')\le x\le Q(A_s)$. Choose a measurable set B with $A'\subseteq B\subseteq A_s$, $Q(B)\le x$, and $Q(B)$ maximum (such a set clearly exists).

Assume first that $Q(B)=x$. Then

$$\begin{aligned}
\int_\Omega g dQ_k &= \int_U g dQ_k + \int_{\Omega\setminus U} g\phi dQ\\
&= \int_U g dQ_k + \int_{B\setminus U}\phi dQ + \int_{B\setminus U}(g-1)\phi dQ + \int_{\Omega\setminus B} g\phi dQ\\
&\le Q_k(B) + s\int_{B\setminus U}(g-1)dQ + s\int_{\Omega\setminus B} g\,dQ = Q_k(B) - sQ(B) + s\int_\Omega g dQ\\
&= Q_k(B) - sx + sx = Q_k(B).
\end{aligned}$$

So the incidence function of B achieves the supremum in the definition of $h_k(x)$.

Second, assume that $Q(B)<x$. Then for every $W\subseteq A_s\setminus B$, we have either $Q(W)=0$ or $Q(W)>x-Q(B)$. Hence the measure Q, restricted to $A_s\setminus B$, is concentrated on atoms; let V be any of these. So $Q(V)>x-Q(B)$ and for every subset of $V'\subseteq V$, we have either $Q(V')=Q(V)$ or $Q(V')=0$. Let $B'=B\cup V$.

Now we have, similarly as above,

$$\begin{aligned}
\int_\Omega g\,dQ_k &= \int_U g\,dQ_k + \int_{B\setminus U} g\,dQ_k + \int_V g\,dQ_k + \int_{\Omega\setminus B'} g\,dQ_k\\
&= \int_U g dQ_k + \int_{B\setminus U}\phi dQ + \int_{B\setminus U}(g-1)\phi dQ + s\int_V g\,dQ + \int_{\Omega\setminus B'} g\phi dQ\\
&\le Q_k(U) + Q_k(B\setminus U) + s\int_{B\setminus U}(g-1)dQ + s\int_V g dQ + s\int_{\Omega\setminus B'} g\,dQ\\
&= Q_k(B) - sQ(B) + s\int_\Omega g\,dQ = Q_k(B) + s[x-Q(B)].
\end{aligned}$$

Let f be the incidence function of B and f', the incidence function of B'. Let $\lambda = (x - Q(B))/Q(V)$. Then

$$\int_\Omega [(1-\lambda)f + \lambda f']\, dQ = Q(B) + \lambda Q(V) = x,$$

and

$$\int_\Omega [(1-\lambda)f + \lambda f']\, dQ_k = Q_k(B) + \lambda Q_k(V) = Q_k(B) + s[x - Q(B)]x,$$

So the function $(1-\lambda)f + \lambda f'$ achieves the supremum in the definition of $h_k(x)$. ∎

Below we shall call x an *extreme point* of a convex [concave] function, if it is not on (in case of a discontinous function above [below]) a chord of the function.

It follows from this lemma that the extreme points of the concave function $h_k(x)$ are those where the supremum is attained by the incidence function of a measurable set. It also follows that

$$\sup_x h_k(x) = \sup_A |Q_k(A) - Q(A)|$$

is half of the ℓ_1-distance of Q and Q_k. The Markov chain is called *rapidly mixing* if for some $\theta < 1$, this distance is $O(\theta^k)$.

The main lemma we need is a natural extension of Lemma 1.4 of Lovász and Simonovits (1990):

1.3. Lemma. *Let $k \geq 1$. If $s \leq x \leq 1/2$, then*

$$h_k(x) \leq \frac{1}{2}\Big(h_{k-1}\big(x - 2\Phi_s(x-s)\big) + h_{k-1}\big(x + 2\Phi_s(x-s)\big)\Big).$$

If $1/2 \leq x \leq 1-s$, then

$$h_k(x) \leq \frac{1}{2}\Big(h_{k-1}\big(x - 2\Phi_s(1-x-s)\big) + h_{k-1}\big(x + 2\Phi_s(1-x-s)\big)\Big).$$

Proof. We prove the first inequality; the second is analogous. Since both sides of the inequality are concave functions of x, it suffices to verify the inequality for the extreme points of the function on the left hand side. So by Lemma 1.2, we may assume that there exists a set A with $Q(A) = x$ and $h_k(x) = Q_k(A) - Q(A)$. Define, for $u \in \Omega$,

$$g_1(u) = \begin{cases} 2P_u(A) - 1, & \text{if } u \in A, \\ 0, & \text{if } u \notin A, \end{cases}$$

$$g_2(u) = \begin{cases} 1, & \text{if } u \in A, \\ 2P_u(A), & \text{if } u \notin A. \end{cases}$$

Also set $x_i = \int_\Omega g_i\, dQ$. Then – by the laziness of the walk – $0 \leq g_i \leq 1$ and $0 \leq x_i \leq 1$. Moreover, since Q is stationary,

$$x_1 + x_2 = \int_\Omega (g_1 + g_2)\, dQ = 2\int_\Omega P_u(A)\, dQ = 2x.$$

Now we have

$$Q_k(A) = \int_\Omega P_u(A)\, dQ_{k-1} = \frac{1}{2}\int_\Omega g_1 dQ_{k-1} + \frac{1}{2}\int_\Omega g_2\, dQ_{k-1}.$$

So

$$\begin{aligned} h_k(x) &= Q_k(A) - x = \frac{1}{2}\left(\int_\Omega g_1 dQ_{k-1} - x_1\right) + \frac{1}{2}\left(\int_\Omega g_2\, dP_{k-1} - x_2\right) \\ &\le \frac{1}{2}h_{k-1}(x_1) + \frac{1}{2}h_{k-1}(x_2). \end{aligned}$$

Moreover, we have

$$x_2 - x = 2\int_{\Omega\setminus A} P_u(A)\, dQ + Q(A) - x = 2\int_{\Omega\setminus A} P_u(A)\, dQ(u) \ge 2\Phi_s(x-s),$$

and so

$$x_2 - x = x - x_1 \ge 2\Phi_s(x-s).$$

Hence, by the concavity of h_{k-1},

$$h_k(x) \le \frac{1}{2}h_{k-1}(x_1) + \frac{1}{2}h_{k-1}(x_2) \le \frac{1}{2}h_{k-1}(x - 2\Phi_s(x-s)) + \frac{1}{2}h_{k-1}(x + 2\Phi_s(x-s)).$$

∎

The following theorem is a slight extension of Theorem 1.2 of [LS].

1.4. Theorem. *Let $0 \le s \le 1/2$, and assume that c_1, c_2 are chosen so that for every $s \le x \le 1-s$ we have*

$$h_0(x) \le c_1 + c_2 \min\{\sqrt{x-s}, \sqrt{1-s-x}\}.$$

Then for every $k \ge 0$ and $s \le x \le 1-s$ we have

$$h_k(x) \le c_1 + c_2 \min\{\sqrt{x-s}, \sqrt{1-s-x}\}\left(1 - \frac{1}{2}\Phi_s^2\right)^k.$$

Proof. By induction on k. If $k=0$, then the assertion is obvious. Assume that $k \ge 1$, let $s \le x \le 1/2$, and apply Lemma 1.3:

$$\begin{aligned} h_k(x) &\le \frac{1}{2}\Big(h_{k-1}\big(x - 2\Phi_s(x-s)\big) + h_{k-1}\big(x + 2\Phi_s(x-s)\big)\Big) \\ &\le c_1 + \frac{c_2}{2}\left(\sqrt{x - 2\Phi_s(x-s) - s}\left(1 - \frac{1}{2}\Phi_s^2\right)^{k-1} + \sqrt{x + 2\Phi_s(x-s) - s}\left(1 - \frac{1}{2}\Phi_s^2\right)^{k-1}\right) \\ &\le c_1 + \frac{c_2}{2}\sqrt{x-s}\left(\sqrt{1-2\Phi_s} + \sqrt{1+2\Phi_s}\right)\left(1 - \frac{1}{2}\Phi_s^2\right)^{k-1} \\ &\le c_1 + c_2\sqrt{x-s}\left(1 - \frac{1}{2}\Phi_s^2\right)^k. \end{aligned}$$

For $1/2 \le x \le 1-s$ the bound follows similarly. ∎

1.5. Corollary. (a) *Let* $M = \sup_A Q_0(A)/Q(A)$. *Then for every* $A \subseteq \Omega$,

$$|Q_k(A) - Q(A)| \le \sqrt{M}\left(1 - \frac{1}{2}\Phi^2\right)^k.$$

(b) *Let* $0 < s \le 1/2$ *and* $H_s = \sup\{|Q_0(A) - Q(A)| \colon Q(A) \le s\}$. *Then for every* $A \subseteq \Omega$,

$$|Q_k(A) - Q(A)| \le H_s + \frac{H_s}{s}\left(1 - \frac{1}{2}\Phi_s^2\right)^k.$$

Proof. (a) The definition of M implies that for all $0 \le x \le 1$, we have $h_0(x) \le Mx$. Also trivially $h_0(x) \le 1 - x$. Hence $h_0(x) \le \sqrt{M}\min\{\sqrt{x}, \sqrt{1-x}\}$. Thus by Theorem 1.4,

$$h_k(x) \le \sqrt{M}\min\{\sqrt{x}, \sqrt{1-x}\}\left(1 - \frac{1}{2}\Phi^2\right)^k < \sqrt{M}\left(1 - \frac{1}{2}\Phi^2\right)^k,$$

from where the assertion follows.

(b) First we show that for every $0 \le x \le 1$,

$$h_0(x) \le H_s + \frac{H_s}{s}\sqrt{x - s}. \tag{1.4}$$

It is enough to prove this for the case when x is an extreme point of h_0. For $0 \le x \le s$ we have

$$h_0(x) = \sup_{\substack{A \in \mathcal{A} \\ Q(A) = x}} (Q_0(A) - Q(A)) \le H_s.$$

Similarly, for $1 - s \le x \le 1$ we have

$$\begin{aligned} h_0(x) &= \sup_{\substack{A \in \mathcal{A} \\ Q(A) = x}} (Q_0(A) - Q(A)) = \sup_{\substack{B \in \mathcal{A} \\ Q(B) = 1 - x}} \big(Q_0(\Omega \setminus B) - Q(\Omega \setminus B)\big) \\ &= \sup_{\substack{B \in \mathcal{A} \\ Q(B) = 1 - x}} (Q(B) - Q_0(B)) \le H_s. \end{aligned}$$

By the concavity of $h_0(x)$, we have for every $s \le x \le 1$ that

$$h_0(x) \le \frac{h_0(s)}{s}x \le \frac{H_s}{s}x \le H_s + \frac{H_s}{s}(x - s) \le H_s + \frac{H_s}{s}\sqrt{x - s}.$$

This last upper bound also holds for $x < s$ by the definition of H_s, and hence (1.4) follows. Similarly, we have for $0 \le x \le 1$

$$h_0(x) \le H_s + \frac{H_s}{s}\sqrt{1 - s - x}.$$

So by Theorem 1.4,

$$h_k(x) \le H_s + \frac{H_s}{s}\left(1 - \frac{1}{2}\Phi_s^2\right)^k \min\{\sqrt{x - s}, \sqrt{1 - s - x}\} < H_s + \frac{H_s}{s}\left(1 - \frac{1}{2}\Phi_s^2\right)^k.$$

∎

d. A Central Limit Theorem.

We prove a simple "Central Limit Theorem" for time-reversible Markov chains. It is quite possible that this can be extended to all Markov chains by methods similar to those in the previous chapter; but currently our proof makes use of time reversibility, mainly through Lemma 1.1.

We need a lemma which may be viewed as a "spectral" result about the operator M. For finite Markov chains, several of the first results establishing the rapid mixing property used the spectrum of the matrix M. Some of these techniques extend to the infinite case without any difficulty.

1.6. Lemma. *Let $\mathcal{M}$ be a time-reversible Markov scheme with conductance Φ. Then for every function $f \in L^2$ with $\mathbf{E}(f) = 0$, we have*

$$\langle f, Mf \rangle \le \left(1 - \frac{\Phi^2}{2}\right) \|f\|^2.$$

Proof. We adapt the proof in [SJ]. We use the identity (1.3). Choose a real number r such that $Q(\{x\colon\ f(x) > r\}) \le 1/2$ and $Q(\{x\colon\ f(x) < r\}) \le 1/2$. Let $g(x) = \max\{f(x) - r, 0\}$. Replacing f by $-f$ if necessary, we may assume that

$$\int_\Omega g^2(x)\, dQ(x) \ge \frac{1}{2} \int_\Omega (f(x) - r)^2\, dQ(x) = \frac{1}{2}\|f\|^2 + \frac{1}{2} r^2 \ge \frac{1}{2}\|f\|^2.$$

Let $A(t) = \{x \in \Omega\colon\ g^2(x) > t\}$. Then we have

$$\begin{aligned}
\int_\Omega \int_\Omega |g^2(x) - g^2(y)|\, dP_y(x)\, dQ(y) &= 2 \int_\Omega \int_{A(g^2(y))} (g^2(x) - g^2(y))\, dP_y(x)\, dQ(y) \\
&= 2 \int_\Omega \int_{g^2(y)}^\infty P_y(A(t))\, dt\, dQ(y) = 2 \int_0^\infty \int_{\Omega \setminus A(t)} P_y(A(t))\, dQ(y)\, dt \\
&\ge 2\Phi \int_0^\infty Q(A(t))\, dt = 2\Phi \int_\Omega g^2(x)\, dx = 2\Phi \|g\|^2,
\end{aligned}$$

since by the choice of r, $Q(A(t)) \le 1/2$ if $t \ge 0$. On the other hand, we have by Cauchy-Schwarz,

$$\begin{aligned}
&\int_\Omega \int_\Omega |g^2(x) - g^2(y)|\, dP_y(x)\, dQ(y) \\
&\qquad \le \left(\int_\Omega \int_\Omega (g(x) - g(y))^2\, dP_y(x)\, dQ(y) \right)^{1/2} \left(\int_\Omega \int_\Omega (g(x) + g(y))^2\, dP_y(x)\, dQ(y) \right)^{1/2}.
\end{aligned}$$

Here the second factor is easily estimated:

$$\begin{aligned}
&\left(\int_\Omega \int_\Omega (g(x) + g(y))^2\, dP_y(x)\, dQ(y) \right)^{1/2} \le \left(\int_\Omega \int_\Omega 2(g^2(x) + g^2(y))\, dP_y(x)\, dQ(y) \right)^{1/2} \\
&= 2 \left(\int_\Omega g^2(x)\, dx \right)^{1/2} = 2\|g\|.
\end{aligned}$$

So

$$\begin{aligned}\int_\Omega\int_\Omega (g(x)-g(y))^2\,dP_y(x)\,dQ(y)&\\ \ge \left(\int_\Omega\int_\Omega |g^2(x)-g^2(y)|\,dP_y(x)\,dQ(y)\right)^2 &\Big/ \int_\Omega\int_\Omega (g(x)+g(y))^2\,dP_y(x)\,dQ(y)\\ \ge 4\Phi^2\|g\|^4/2\|g\|^2 = 2\Phi^2\|g\|^2 \ge \Phi^2\|f\|^2.&\end{aligned}$$

Hence

$$\int_\Omega\int_\Omega (f(x)-f(y))^2\,dP_y(x)\,dQ(y) \ge \int_\Omega\int_\Omega (g(x)-g(y))^2\,dP_y(x)\,dQ(y) \ge \Phi^2\|f\|^2,$$

which proves the lemma by (1.3). ∎

1.7. Corollary. *Let $\mathcal{M}$ be a time-reversible Markov scheme with conductance Φ. Then for every function $f\in L^2$ with $\mathbf{E}(f)=0$, we have*

$$\langle f, M^k f\rangle \le \left(1-\frac{\Phi^2}{2}\right)^k \|f\|^2.$$

Proof. Let $\hat{M}$ denote the restriction of M to the invariant subspace $\mathbf{E}(f)=0$. From Lemma 1.6 we get (see Riesz–Sz.-Nagy 19xx, VI, §1) that $\|\hat{M}\|\le 1-\Phi^2/2$, and hence

$$\|\hat{M}^k\| \le \|\hat{M}\|^k \le \left(1-\frac{\Phi^2}{2}\right)^k.$$

∎

1.8. Theorem. *Let $\mathcal{M}$ be a time-reversible Markov scheme with stationary distribution Q, let $w_1, w_2, \dots$ be a Markov chain generated by $\mathcal{M}$ with initial distribution Q. Let $F\in L^2$ and $\xi=\sum_{i=0}^{T-1} F(w_i)$. Then*

$$\mathbf{D}^2(\xi) \le \frac{4T}{\Phi^2}\|F\|^2.$$

Proof. We may assume that $\mathbf{E}(\xi)=\int_\Omega F\,dQ=0$. Then we have, by Lemmas 1.1 and 1.6,

$$\begin{aligned}\mathbf{D}^2(\xi)=\mathbf{E}(\xi^2) &= \sum_{0\le i,j\le T-1}\mathbf{E}(F(w_i)F(w_j)) = \sum_{0\le i,j\le T-1}\mathbf{E}(F(w_0)F(w_{|j-i|}))\\ &= T\langle F,F\rangle + \sum_{k=1}^{T-1} 2(T-k)\langle F, M^k F\rangle < 2T\sum_{k=0}^{T-1}\langle F, M^k F\rangle\\ &\le 2T\sum_{k=0}^{\infty}\langle F, M^k F\rangle \le 2T\sum_{k=0}^{\infty}\left(1-\frac{\Phi^2}{2}\right)^k\|F\|^2 = \frac{4T}{\Phi^2}\|F\|^2.\end{aligned}$$

∎

e. The Metropolis Filter.

We conclude this section with describing a version of the Metropolis algorithm (1953), which can be used to modify a given time-reversible Markov scheme so as to achieve a given stationary distribution. The use of this method in volume computations was initiated by Applegate and Kannan.

Consider a time-reversible Markov scheme $\mathcal{M}$ on the σ-algebra $(\Omega, \mathcal{A})$, and let $F\colon \Omega \to \mathbb{R}$ be a non-negative measurable function. Assume that the integral

$$\overline{F} = \int_\Omega F\, dQ$$

is finite. Recall that we denote by μ_F the measure with density function F, i.e.,

$$\mu_F(S) = \int_S F(x) dQ(x)$$

for every measurable set S. Clearly $Q_F = (1/\overline{F})\mu_F$ is a probability distribution on $(\Omega, \mathcal{A})$.

We generate a Markov chain $(w_0, w_1, \ldots)$ of elements of Ω as follows. Given w_k, we generate a random element u from the distribution P_{w_k}. If $F(u) \geq F(w_k)$, we let $w_{k+1} = u$. If $F(u) < F(w_k)$, then we "flip a biased coin" with heads probability $F(u)/F(w_k)$. If we see head, we move to $w_{k+1} = u$. Else, we let $w_{k+1} = w_k$. Note the simple but very important property of this method that we do not use the density function $F/\overline{F}$ of the desired limit distribution, but only the ratios $F(u)/F(v)$.

The transition probabilities of this new Markov scheme are given by

$$P_u^F(A) = \begin{cases} \displaystyle\int_A \min\left\{1, \frac{F(v)}{F(u)}\right\} dP_u(v), & \text{if } u \notin A, \\ \displaystyle\int_A \min\left\{1, \frac{F(v)}{F(u)}\right\} dP_u(v) + \ell(u), & \text{if } u \in A, \end{cases}$$

where

$$\ell(u) = \int_\Omega \max\left\{0, 1 - \frac{F(v)}{F(u)}\right\} dP_u(u)$$

is the probability that we had to stay because the coin tossing came out wrong. We call this new Markov scheme $\mathcal{M}/F$ the *filtering* of $\mathcal{M}$ by F.

1.9. Lemma. *If $\mathcal{M}$ is time-reversible, then the Markov scheme $\mathcal{M}/F$ is also time-reversible with stationary distribution Q_F.*

Proof. By our remark after the definition of ergodic flow, it suffices to verify that for any two disjoint measurable sets A and B, we have

$$\int_A P_u^F(B)\, dQ_F(u) = \int_B P_u^F(A)\, dQ_F(u).$$

Substituting for P_u^F and Q_F, this is equivalent to

$$\int_A \int_B \min\left\{1, \frac{F(v)}{F(u)}\right\} dP_u(v) \frac{F(u)}{\overline{F}}\, dQ(u) = \int_B \int_A \min\left\{1, \frac{F(v)}{F(u)}\right\} dP_u(v) \frac{F(u)}{\overline{F}}\, dQ(u),$$

or

$$\frac{1}{\overline{F}} \int_A \int_B \min\{F(u),\ F(v)\}\, dP_u(v)\, dQ(u) = \frac{1}{\overline{F}} \int_B \int_A \min\{F(u), F(v)\}\, dP_u(v)\, dQ(u).$$

This equality follows from the assumption that $\mathcal{M}$ is time-reversible. ∎

Remark. For us, the important property of $\mathcal{M}/F$ will be that Q_F is stationary. This is only true if the original Markov scheme $\mathcal{M}$ is time-reversible (the time-reversibility of $\mathcal{M}/F$ is a convenient, but not crucial side benefit). There are many other ways to generate a Markov chain with given stationary distribution Q_F (from the known Markov chain $\mathcal{M}$), but the Metropolis filter is in a sense optimal.

2. Log-concave functions and isoperimetric inequalities

a. Log-concave functions.

We collect some known facts about convex bodies and log-concave functions.

Definition. (Log-concave function.) A function $f\colon \mathbb{R}^n \to \mathbb{R}_+$ is *log-concave* if it satisfies, for all $x, y \in \mathbb{R}^n$ and $0 < \lambda < 1$,

$$f(\lambda x + (1-\lambda)y) \le f(x)^\lambda f(y)^{(1-\lambda)}.$$

Equivalently, a (non-negative) function is log-concave if its support $K = \{x \in \mathbb{R}^n\colon\ f(x) > 0\}$ is convex, and $\log f$ is a concave function on K.

Every non-negative function that is concave over a convex domain is log-concave (we define its value to be 0 outside the original domain). In particular, the characteristic function of a convex body is log-concave. Log-concave functions include many density functions important in statistics, e.g. e^{-x^2} and $e^{-|x|}$.

It is obvious that the product and the minimum of two log-concave functions is log-concave (not their sum!). The following fact (Dinghas 1957, Prékopa 1971, 1973) is much less obvious:

2.1. Lemma. *The convolution $h(x) = \int_{\mathbb{R}^n} g(u) f(x-u)\, du$ of two log-concave functions (assuming that it is well-defined) is log-concave.* ∎

If F is a non-negative integrable function on $\mathbb{R}^n$, then we denote by μ_F the measure with density function F, i.e.,

$$\mu_F(A) = \int_A F(x)\, dx.$$

Applying Lemma 2.1 with g chosen to be the characteristic function of the convex body $-K$, we obtain the following Brunn-Minkowski type theorem:

2.2. Corollary. *Let $K \subseteq \mathbb{R}^n$ be a convex body and $F\colon \mathbb{R}^n \to \mathbb{R}_+$ a log-concave function. Then $\mu_F(x+K)$ is a log-concave function of x.* ∎

Specializing further, we obtain

2.3. Corollary. *Let K and K' be two convex bodies and $t>0$. If the set $\{x \in \mathbb{R}^n : \mathrm{vol}((K'+x)\cap K) > t\}$ has an interior point, then it is a convex body. In particular, the set $K_s = \{x \in K : \mathrm{vol}((x+K)\cap K) \geq (1-s)\mathrm{vol}(K)\}$ is a convex body for any $0<s<1$.* ∎

Another consequence of Lemma 2.1 we shall need is the following, which we obtain by choosing K in Corollary 2.2 a rectangle aligned with the axes having edges of length ε in k directions and $1/\varepsilon$ in the remaining directions, and then letting $\varepsilon \to 0$.

2.4. Corollary. *If $F\colon \mathbb{R}^n \to \mathbb{R}_+$ is a log-concave function with finite integral, then for any subset $\{x_1,\dots,x_k\}$ of variables, $\int_{\mathbb{R}} \dots \int_{\mathbb{R}} F\,dx_1 \dots dx_k$ is a log-concave function of the remaining variables.* ∎

b. Isoperimetric inequalities for log-concave functions.

The following general lemma is a refinement of the bisection method introduced in [LS].

2.5. Lemma. [Localization Lemma.] *Let g and h be upper semi-continuous Lebesgue integrable functions on $\mathbb{R}^n$ such that*

$$\int_{\mathbb{R}^n} g(x)\,dx > 0 \quad \text{and} \quad \int_{\mathbb{R}^n} h(x)\,dx > 0.$$

Then there exist two points $a,b \in \mathbb{R}^n$ and a linear function $\ell\colon [0,1] \to \mathbb{R}_+$ such that

$$\int_0^1 \ell(t)^{n-1} g((1-t)a+tb)\,dt > 0 \quad \text{and} \quad \int_0^1 \ell(t)^{n-1} g((1-t)a+tb)\,dt > 0.$$

(We may formulate the conclusion informally, in a more transparent way, as follows: if both g and h have positive integrals, then there exists an infinitesimally narrow truncated cone such that the restrictions of g and h to this "body" have positive integrals. It is easy to show that we could not further restrict the family of these "test bodies": cones, or cylinders would not be enough.)

Proof. We may assume that g and h are continuous: in fact, they arise as limits of monotone (strictly) increasing sequences of integrable continuous functions $g_k \to g$ and $h_k \to h$, and we have

$$\lim_{k\to\infty} \int_{\mathbb{R}^n} g_k = \int_{\mathbb{R}^n} g > 0,$$

so for k large enough, we have $\int_{\mathbb{R}^n} g_k > 0$, and similarly $\int_{\mathbb{R}^n} h_k > 0$. If we know the validity of the lemma for continuous functions, then it follows that there exist two points $a, b \in \mathbb{R}^n$ and a non-negative linear function ℓ such that

$$\int_0^1 \ell(t)^{n-1} g_k((1-t)a+tb)\,dt > 0 \quad \text{and} \quad \int_0^1 \ell(t)^{n-1} h_k((1-t)a+tb)\,dt > 0,$$

and then

$$\int_0^1 \ell(t)^{n-1} g((1-t)a+tb)\,dt > 0 \quad \text{and} \quad \int_0^1 \ell(t)^{n-1} h((1-t)a+tb)\,dt > 0.$$

Note also that it is enough to know that the conclusion holds with ≥ 0 instead of > 0. Indeed, if we know the result with ≥ 0, then we can apply it to $g-a$ and $h-a$ in place of g and h, where a is an everywhere positive continuous function with a small integral, which implies strict inequality for the original functions.

Claim 1. *There exists a sequence of convex bodies $K_0 \subseteq K_1 \subseteq K_2, \ldots$ such that*

$$\int_{K_i} g(x)\,dx > 0 \quad \text{and} \quad \int_{K_i} h(x)\,dx > 0, \tag{2.1}$$

and $K = \cap_i K_i$ is a point or a segment.

For K_0 we may choose any sufficiently large ball. Given K_i, we choose a halfspace H such that

$$\int_{K_i \cap H} g(x)\,dx = \frac{1}{2} \int_{K_i} g(x)\,dx.$$

Let us call the boundary hyperplane of such a halfspace *bisecting*. Replacing H by the complementary halfspace if necessary, we may assume that

$$\int_{K_i \cap H} h(x)\,dx > 0.$$

Thus $K_{i+1} = K_i \cap H$ will satisfy (2.1). What we have to show is that we can choose H so that the K_i shrink to a 0-dimensional or 1-dimensional body. To this end, let us remark that given any $(n-2)$-dimensional affine subspace A, there is at least one bisecting hyperplane containing it. This follows by a trivial continuity argument, obtained by rotating the hyperplane about A.

Let $A_0, A_1, \ldots$ be all the $(n-2)$-dimensional affine subspaces with rational coordinates, ordered in a single sequence. Let K_{i+1} be obtained from K_i by cutting it into two by a bisecting hyperplane P_i through A_i, and choosing the appropriate half.

We show that $K = \cap_i K_i$ is at most 1-dimensional. Assume that it is 2-dimensional, then its projection onto one of the planes spanned by two coordinate axes (say, axes 1 and 2) is still 2-dimensional, and so has a rational interior point (r_1, r_2). The affine subspace defined by $x_1 = r_1,\ x_2 = r_2$ is one of the A_i; but then P_i properly bisects K, and so also K_{i+1}, which contradicts the construction.

If K is a single point a, then it follows from the continuity of g and h that $g(a) \geq 0$ and $h(a) \geq 0$, and so the conclusion holds for $b = a$ and $\ell \equiv 1$. So assume that K is a proper segment, and let a and b denote the endpoints of K.

Claim 2. *There exists a concave function $\psi : [0,1] \to \mathbb{R}_+$, not identically zero, such that*

$$\begin{aligned} \int_0^1 \psi(t)^{n-1} g((1-t)a + tb)\, dt \geq 0, \\ \int_0^1 \psi(t)^{n-1} h((1-t)a + tb)\, dt \geq 0. \end{aligned} \tag{2.2}$$

Without loss of generality, $a = 0$, $b = e_1$. Consider an $i \geq i_0$. Set

$$Z_t = \{x \in \mathbb{R}^n :\ x_1 = t\}$$

and

$$\psi_i(t) = \left(\frac{\mathrm{vol}_{n-1}(K_i \cap Z_t)}{\mathrm{vol}(K)} \right)^{1/(n-1)}.$$

Let α_i and β_i denote the minimum and maximum of x_1 over K_i. Clearly $\alpha_i \leq 0$ and $\beta_i \geq 1$; moreover, $\alpha_i \to 0$ and $\beta_i \to 1$ as $i \to \infty$. The function ψ_i is concave on the interval $[\alpha_i, \beta_i]$ by the Brunn-Minkowski Theorem.

We can select a subsequence of the indices i for which $\psi_i(t)$ converges to a limit $\psi(t)$ for every $0 \leq t \leq 1$. (This is a version of the Blaschke Principle; alternatively, we can notice that the functions ψ_i are uniformly equicontinuous in every interval $[s,t]$ with $0 < s < t < 1$, and hence the Arzela–Ascoli Lemma implies that the sequence has a pointwise convergent subsequence.) Obviously, the limit function ψ is also non-negative, concave, and

$$\int_0^1 \psi(t)^{n-1}\, dt = 1.$$

Now we have (with $x = (t, y)$, $t \in \mathbb{R}$, $y \in \mathbb{R}^{n-1}$)

$$\int_{K_i} g(x)\, dx = \int_{\alpha_i}^{\beta_i} \int_{K_i \cap Z_t} g(t,y)\, dy\, dt = \int_{\alpha_i}^{\beta_i} \left[\frac{1}{\mathrm{vol}_{n-1}(K_i \cap Z_t)} \int_{K_i \cap Z_t} g(t,y)\, dy \right] \mathrm{vol}(K_i) \psi_i(t)^{n-1}\, dt,$$

and hence

$$\frac{1}{\mathrm{vol}(K_i)}\int\limits_{K_i} g(x)\,dx = \int\limits_{\alpha_i}^{\beta_i}\left[\frac{1}{\mathrm{vol}_{n-1}(K_i\cap Z_t)}\int\limits_{K_i\cap Z_t} g(t,y)\,dy\right]\psi_i(t)^{n-1}\,dt.$$

The left hand side is non-negative, while the right hand side tends to $\int_0^1 g(t,0)\psi(t)^{n-1}\,dt$. So this integral is non-negative. The same argument applies to h. This proves the claim.

To conclude the proof, we choose $a,b\in K$ so that for an appropriate not identically zero concave function $\psi\geq 0$, (2.2) is satisfied, and

(i) $|a-b|$ is minimal.

We may assume again that $a=0$ and $b=e_1$. We may also assume that ψ is continuous (it could only have discontinuities at 0 or 1). If ψ is linear, we are done, so suppose that ψ is not linear. Let $0\leq\alpha\leq\beta\leq 1$ be chosen so that

(ii) ψ is linear on the interval $[\alpha,\beta]$ and $\beta-\alpha$ is maximum (subject to (i)).

(It follows by a standard compactness argument that such a pair of points and function ψ exist.) Define

$$\hat{g}(x)=g(p(x)) \quad \text{and} \quad \hat{h}(x)=h(p(x)),$$

where $p(x)$ is the first coordinate of x. For every convex body L, let

$$\psi_L(t)=\mathrm{vol}(L\cap Z_t)^{1/(n-1)}.$$

Also consider the convex body K' defined by

$$0\leq x_1\leq 1, \qquad x_2,\ldots,x_n\geq 0,$$

$$x_2+\cdots+x_n\leq\psi(x_1).$$

(In other words, we consider for every $0\leq t\leq 1$ the $(n-1)$-dimensional simplex S_t spanned by the points $te_1, te_1+\psi(t)e_2,\ldots,te_1+\psi(t)e_n$. The union of these simplices is a convex body K'.) Then

$$\int\limits_{K'}\hat{g}(x)\,dx=\frac{1}{n!}\int\limits_0^1\psi(t)^{n-1}g(te_1)\,dt\geq 0,$$

and similarly

$$\int\limits_{K'}\hat{h}(x)\,dx\geq 0.$$

By (i), one of these integrals is 0; we may assume without loss of generality that

$$\int\limits_{K'}\hat{g}(x)\,dx=0.$$

Consider two real numbers $0<\sigma<1$ and $\tau>0$, and the $(n-2)$-dimensional affine subspace A defined by the equations $x_1=\sigma$ and $x_2+\cdots+x_n=\tau$. Assume that A intersects

the interior of K'. Then there is a hyperplane H through A that splits K' into two convex bodies L_A and L'_A such that

$$\int_{L_A} \hat{g}(x)\,dx = \int_{L'_A} \hat{g}(x)\,dx = 0.$$

We may assume that the notation is chosen so that

$$\int_{L_A} \hat{h}(x)\,dx \geq 0.$$

Then $(*)$ L_A must intersect both Z_0 and Z_1; else, ψ_{L_A} satisfies (2.2) and violates (i). There will be two different possibilities: either L_A "faces down" (it contains the point $(1/2)e_1$) or it "faces up". Condition $(*)$ implies that H cannot be orthogonal to the x_1-axis, and hence (using that it contains an $(n-2)$-dimensional subspace A of a special form) it can be defined by an equation $x_2+\cdots+x_n=\ell(x_1)$ where ℓ is a linear function. If L_A "faces down", then $(*)$ implies that $\ell(0),\ell(1)\geq 0$ and

$$\psi_{L_A}(t) = \min\{\psi(t),\ell(t)\}. \tag{2.3}$$

We are going to show that chosing A appropriately, we get a contradiction at either (i) or (ii).

Case 1. $\psi(0)=\psi(1)=0$. Consider the affine subspace A defined by $x_1=1/2$ and $x_2+\cdots+x_n=\tau$. Then L_A cannot "face up" by $(*)$. It follows from $\ell(0),\ell(1)\geq 0$ and $\ell(1/2)=\tau$ that $\ell(t)$ tends to 0 uniformly on $[0,1]$ if $\tau\to 0$, and so (2.3) implies that ψ_{L_A} is linear on an interval whose length tends to 1, contradicting (ii).

Case 2. $\psi(0)=0$ and $\psi(1)>0$ (say). Consider the affine subspace A defined by $x_1=\sigma$ and $x_2+\cdots+x_n=\tau$, where $\tau=\psi(1)\sigma$. If L_A "faces up", then $(*)$ implies that $\ell(t)=\psi(1)t$ and so considering ψ_{L_A} we get a contradiction by case 1. So assume that L_A "faces down". Then $(*)$ implies that $0\leq\ell(1)\leq\psi(1)$ and so (2.3) implies that ψ_0 is linear in the interval $[\sigma,1]$, which is longer than $\beta-\alpha$ if σ is small enough.

Case 3. $\psi(0),\psi(1)>0$. Consider all continuous convex (!) functions $\eta:[0,1]\to\mathbb{R}_+$ such that $0\leq\eta(0)\leq\psi(0)$ and $0\leq\eta(1)\leq\psi(1)$, and the convex body K_η defined by

$$0\leq x_1\leq 1, \qquad x_2,\ldots,x_n\geq 0,$$

$$\eta\leq x_2+\cdots+x_n\leq\psi(x_1)$$

satisfies

$$\int_{K_\eta}\hat{g}(x)\,dx=0, \quad \text{and} \quad \int_{K_\eta}\hat{h}(x)\,dx\geq 0.$$

Choose an η for which $\int_0^1\eta(t)\,dt$ is maximal. If $\eta(0)=\psi(0)$ or $\eta(1)=\psi(1)$, then considering

$$\psi_{K_\eta}=\psi-\eta$$

we get a contradiction by either Case 1 or Case 2. So assume that $\eta(0) < \psi(0)$ and $\eta(1) < \psi(1)$.

Let (σ, τ) denote the intersection point of the segments connecting $(0, \eta(0))$ to $(1, \psi(0))$ and $(0, \psi(0))$ to $(1, \eta(0))$, and consider a hyperplane H through the affine subspace defined by $x_1 = \sigma$, $x_2 + \cdots + x_n = \tau$, which cuts K_η into two parts M and M' for which

$$\int\limits_M \hat{g}(x)\,dx = \int\limits_{M'} \hat{g}(x)\,dx = 0.$$

Note that by the choice of (σ, τ), either both M and M' intersect Z_0 and Z_1 or none of them does; but the latter case is ruled out by $(*)$. Let M "face up" and M' "face down". Then $M = K_\ell$ and so by the maximality of η, we must have

$$\int\limits_M \hat{h}(x)\,dx < 0.$$

But the $K' \setminus M$ is a truncated cone and we have

$$\int\limits_{K' \setminus M} g(x)\,dx = \int\limits_{K'} g(x)\,dx - \int\limits_M g(x)\,dx = 0,$$

and

$$\int\limits_{K' \setminus M} h(x)\,dx = \int\limits_{K'} h(x)\,dx - \int\limits_M h(x)\,dx > 0,$$

which contradicts (ii). ∎

As an application, we derive the following isoperimetric inequality, which is a slight extension of the inequality by Dyer and Frieze [DF] (which, in turn, is a slight extension of the inequality by Applegate and Kannan [AK]).

2.6. Theorem. *Let $K \subseteq \mathbb{R}^n$ be a convex body, $0 < t < 1$, and $K_1 \cup K_2 \cup K_3$ be a partition of K into three measurable sets such that for any two points $a, b \in K$, the distance of $K_1 \cap [a,b]$ and $K_2 \cap [a,b]$ is at least $t|a-b|$. Then for any log-concave function with support K,*

$$\mu_F(K_3) \geq \frac{2t}{1-t} \min\{\mu_F(K_1),\ \mu_F(K_2)\}.$$

(This theorem is tight when $F \equiv 1$ and K is a cylinder.)

Proof. Assume, by way of contradiction, that

$$\mu_F(K_3) < \frac{2t}{1-t} \min\{\mu_F(K_1),\ \mu_F(K_2)\}.$$

We may assume that K_1 and K_2 are open (we can delete the boundary of K and enlarge K_1 and K_2 slightly to open sets so that both the assumptions and the indirect hypothesis remain true).

Let

$$g(x)=\begin{cases} F(x), & \text{if } x\in K_1,\\ -\dfrac{1-t}{2t}F(x), & \text{if } x\in K_3,\\ 0, & \text{otherwise,}\end{cases}$$

and

$$h(x)=\begin{cases} F(x), & \text{if } x\in K_2,\\ -\dfrac{1-t}{2t}F(x), & \text{if } x\in K_3,\\ 0, & \text{otherwise,}\end{cases}$$

Then, by the indirect hypothesis,

$$\int_{\mathbb{R}^n} g(x)\,dx>0, \quad \text{and} \quad \int_{\mathbb{R}^n} h(x)\,dx>0.$$

Thus by Lemma 2.5, there exist two points $a,b\in\mathbb{R}^n$ and a linear function $\ell\colon\ [0,1]\to\mathbb{R}^n_+$ such that

$$\int_0^1 \ell(u)^{n-1}g((1-u)a+ub)\,du>0 \quad \text{and} \quad \int_0^1 \ell(u)^{n-1}g((1-u)a+ub)\,du>0. \tag{2.4}$$

Let

$$H_i=\{u\colon\ (1-u)a+ub\in K_i\},$$

and

$$G(u)=\ell(u)^{n-1}F((1-u)a+ub).$$

Then (2.4) can be written as

$$\int_{H_3} G(u)\,du<\frac{2t}{1-t}\min_{i=1,2}\int_{H_i} G(u)\,du.$$

Let μ_G be the measure on $[0,1]$ with density function G.

Claim. *For $0\le s<s+t\le 1$ we have*

$$\mu_G([s,s+t])\ge\frac{2t}{1-t}\min\{\mu_G([0,s]),\ \mu_G([s+t,1]).\}$$

(Note that if we were content with the coefficient t instead of $2t/(1-t)$, then the claim would be trivial, and in fact at this point we would only have to use that the function G is unimodal.)

We may assume that $G(u)=e^{cu}$ for some constant $c>0$. To see this, note that there are constants $c_0>0$ and c such that $G(s)=c_0e^{cs}$ and $G(s+t)=c_0e^{c(s+t)}$. By the log-concavity of G, we have $G(u)\ge c_0e^{cu}$ for $s<u<s+t$ and $G(u)\le c_0e^{cu}$ else. Thus it suffices to prove the claim for $G(u)=c_0e^{cu}$. Obviously, we may assume that $c_0=1$. If $c=0$ then

the assertion is obvious, so we may assume that $c \neq 0$ and (without loss of generality) that $c > 0$. Thus we want to show that

$$e^{c(s+t)} - e^{cs} \geq \frac{2t}{1-t} \min\left\{e^{cs} - 1,\ e^c - e^{c(s+t)}\right\}.$$

One can easily see that the worst case is when $e^{cs} - 1 = e^c - e^{c(s+t)}$, i.e., when

$$e^{cs} = \frac{1+e^c}{1+e^{ct}},$$

in which case we want to show that

$$(e^{ct} - 1)\frac{1+e^c}{1+e^{ct}} \geq \frac{2t}{1-t}\left(\frac{1+e^c}{1+e^{ct}} - 1\right).$$

Introducing $x = e^{ct}$ and $\lambda = 1/t$ we get

$$(\lambda - 1)x^{\lambda+1} - (\lambda+1)x^\lambda + (\lambda+1)x - (\lambda-1) \geq 0.$$

We have $x > 1$ and $\lambda > 1$. Moreover, the function $f(x)$ on the right hand side satisfies $f(1) = f'(1) = f''(1) = 0$ and

$$f''(x) = (\lambda+1)\lambda(\lambda-1)(x^{\lambda-1} - x^{\lambda-2}) \geq 0$$

for $x \geq 1$, and therefore $f(x) \geq 0$ for all $x \geq 1$. This proves the claim.

The claim proves the assertion for the (intuitively worst) case when H_3 consists of a single interval. In the general case, one could easily simplify H_3 until it consists of one or two intervals, and treat this case directly. To give an argument that is easier to describe, for each maximal interval $I = (a, b)$ contained in H_3 and of length at least t we color the interval $[0, a]$ red if $\mu_G([0, a]) < \mu_G([b, 1])$; else, color the interval $[b, 1]$ red. By the claim, each interval I introduces a red set with μ_G-measure at most $(1-t)/(2t)\mu_G([I])$. So the whole red set will have measure at most $(1-t)/(2t)\mu_G(H_3)$.

It suffices to show that either H_1 or H_2 is totally red. Suppose not, then the uncolored set U intersects both H_1 and H_2. By our construction, U is an open interval. By the assumption that the distance of H_1 and H_2 is at least t, $U \cap H_3$ contains an interval of length at least t. But then there is an adjacent red interval, a contradiction. ∎

2.7. Corollary. *Let $K \subseteq \mathbb{R}^n$ be a convex set, $0 < t < 1$, and $K_1 \cup K_2 \cup K_3$, a partition of K into three measurable sets such that for any two points $a, b \in K$, the distance of $K_1 \cap [a, b]$ and $K_2 \cap [a, b]$ is at least $t|a - b|$. Then*

$$\mathrm{vol}(K_3) \geq \frac{2t}{1-t} \min\{\mathrm{vol}(K_1), \mathrm{vol}(K_2)\}.$$

In particular, if (in any norm) K has diameter d and the distance of K_1 and K_2 is at least 1, then

$$\mathrm{vol}(K_3) \geq \frac{2}{d-1} \min\{\mathrm{vol}(K_1), \mathrm{vol}(K_2)\}.$$

The Localization Lemma has several applications in convex geometry; for example, it gives a simple proof of Lemma 2.1, and implies Brunn-Minkowski type results for balls and other bodies. These will be treated in detail elsewhere; here we formulate one consequence, which will be handy in our volume algorithm.

2.8. Theorem. *Let F be a log-concave function on $\mathbb{R}^n$, and let*

$$\int_{\mathbb{R}^n\setminus B} F(x)\,dx = \theta \int_{\mathbb{R}^n} F(x)\,dx.$$

Then for every $u \geq 1$,

$$\int_{\mathbb{R}^n\setminus uB} F(x)\,dx \leq \theta^{(u+1)/2} \int_{\mathbb{R}^n} F(x)\,dx.$$

Proof. Assume that

$$\int_{\mathbb{R}^n\setminus uB} F(x)\,dx > \theta^{(u+1)/2} \int_{\mathbb{R}^n} F(x)\,dx.$$

Similarly as in the proof of Theorem 2.6, it follows from Lemma 2.5 that there exist two points $a, b \in \mathbb{R}^n$ and a linear function $\ell : [0,1] \to \mathbb{R}_+$ such that, setting

$$G(t) = \ell(t)^{n-1} F((1-t)a + tb),$$

$$H_1 = \{t :\ 0 \leq t \leq 1,\ (1-t)a + tb \in B\}, \quad H_2 = \{t :\ 0 \leq t \leq 1,\ (1-t)a + tb \in uB \setminus B\},$$
$$H_3 = [0,1] \setminus H_1 \setminus H_2,$$

we have

$$\int_{H_2 \cup H_3} G(t)\,dt \leq \theta \int_0^1 G(t)\,dt, \tag{2.5}$$

and

$$\int_{H_3} G(t)\,dt > \theta^{(u+1)/2} \int_0^1 G(t)\,dt. \tag{2.6}$$

Clearly G is log-concave, H_1 is an interval, and H_2 and H_3 consist of one or two intervals.

We may assume that $0 \in H_1$ (i.e., $a \in B$); else, we may choose a point $s \in H_1$ such that

$$\frac{\int_{[0,s]\cap H_1} G(t)\,dt}{\int_{[s,1]\cap H_1} G(t)\,dt} = \frac{\int_0^s G(t)\,dt}{\int_s^1 G(t)\,dt}.$$

Then we can replace $[a,b]$ by either $[a,(1-s)a+sb]$ or $[(1-s)a+sb,b]$. So let $H_1 = [0,\alpha]$, $H_2 = [\alpha,\beta]$ and $H_3 = [\beta,1]$. It follows easily that $\beta \geq \frac{u+1}{2}\alpha$.

We can choose $c_0, c > 0$ so that

$$\int_0^\alpha G(t)\,dt = \int_0^\alpha c_0 e^{-ct}\,dt, \tag{2.7}$$

and

$$\int_\beta^1 G(t)\,dt = \int_\beta^\infty c_0 e^{-ct}\,dt. \tag{2.8}$$

Then we have $G(t) \geq c_0 e^{-ct}$ for all $\alpha \leq t \leq \beta$; else, it would follow from the log-concavity of G that $G(t) < c_0 e^{-ct}$ either for all $0 \leq t \leq \alpha$ or for all $\beta \leq t \leq 1$, contradicting the choice of c_0 and c. So, by (2.7) and (2.8),

$$\int_0^1 G(t)\,dt \geq \int_0^\infty c_0 e^{-ct}\,dt. \tag{2.9}$$

But then by (2.5) we have

$$e^{-c\alpha} = \frac{\int_\alpha^\infty c_0 e^{-ct}\,dt}{\int_0^\infty c_0 e^{-ct}\,dt} \leq \frac{\int_\alpha^1 G(t)\,dt}{\int_0^1 G(t)\,dt} \leq \theta,$$

and so, by (2.9),

$$\frac{\int_\beta^1 G(t)\,dt}{\int_0^1 G(t)\,dt} \leq \frac{\int_\beta^\infty c_0 e^{-ct}\,dt}{\int_0^\infty c_0 e^{-ct}\,dt} = e^{-c\beta} \leq e^{-c\alpha(u+1)/2} \leq \theta^{(u+1)/2},$$

contradicting (2.6). ∎

As a special case we obtain

2.9. Corollary. *Let K be a convex body in $\mathbb{R}^n$ and let $\theta = \mathrm{vol}(K \setminus B)/\mathrm{vol}(K)$. Then $\mathrm{vol}(K \setminus uB) \leq \theta^{(u+1)/2}\mathrm{vol}(K)$.* ∎

For this special case, the bound in Corollary 2.9 could be improved to $\big(1-(1-\theta^{1/n})(u+1)/2\big)^n$ if $(1-\theta^{1/n})(u+1)/2 < 1$ and 0 otherwise. We shall only need a simpler version of the last assertion, to which we give a simple proof.

2.10. Lemma. *Let K be a convex body in $\mathbb{R}^n$ and let $\theta = \mathrm{vol}(K \setminus B)/\mathrm{vol}(K)$. Then $K \subseteq \dfrac{2n}{1-\theta} B$.*

Proof. Let x be a point in K farthest from the origin, and let $R = |x|$. Replacing K by $\mathrm{conv}((K \cap B) \cup \{x\})$ we decrese θ, and therefore decrease $2n/(1-\theta)$. So we may assume that $K = \mathrm{conv}((K \cap B) \cup \{x\})$. Blowing up $K \setminus B$ from x by a factor of $(R+1)/(R-1)$, we get a set containing K. Thus

$$\mathrm{vol}(K) \leq \left(\frac{R+1}{R-1}\right)^n \mathrm{vol}(K \setminus B) \leq \theta \left(\frac{R+1}{R-1}\right)^n \mathrm{vol}(K).$$

Hence

$$R \leq \frac{2}{1-\theta^{1/n}} - 1 < \frac{2n}{1-\theta}.$$

∎

We conclude this section with one more lemma of similar type.

2.11. Lemma. *Let $0 \le t \le 1$. If $\mathrm{vol}(K \setminus (x+K)) \le (1/2)\mathrm{vol}(K)$, then $\mathrm{vol}(K \setminus (tx+K)) \le (2t/3)\mathrm{vol}(K)$.*

Proof. Consider the function defined for $u \in [0,1]$ by

$$\psi(u) = \frac{\mathrm{vol}(K \cap (ux+K))}{\mathrm{vol}(K)}.$$

By Corollary 2.2, ψ is log-concave. Moreover, $\psi(0) = 1$ and $\psi(1) \ge 1/2$. Hence we have

$$\psi(t) \ge 2^{-t},$$

and so

$$\mathrm{vol}(K \setminus (tx+K)) \le \left(1 - 2^{-t}\right) \mathrm{vol}(K) < \frac{2t}{3}\mathrm{vol}(K).$$

∎

3. Random walks and sampling in convex bodies

a. Random walks with G-steps.

In Section 1 we analyzed general Markov chains, i.e., random walks in abstract measure spaces. Now we turn to the special case when the underlying domain is $\mathbb{R}^n$ (or a convex body), and the step is chosen uniformly from the unit ball about the current point. For the time being, we allow an arbitrary norm, but for the algorithmic applications we only need the case of a euclidean ball.

Let G be a centrally symmetric convex body in $\mathbb{R}^n$, with its center at the origin, and K, any convex body. We generate a Markov chain of points in K as follows. Let v_0 be drawn from some initial distribution on K. Given v_k, we flip a coin and if it is heads, we let $v_{k+1} = v_k$. Else, we generate a vector u from the uniform distribution on G, and consider $v_k + u$. If $v_k + u \in K$, we let $v_{k+1} = v_k + u$. Else, we let $v_{k+1} = v_k$. We call this the *lazy random walk in K with G-steps*. It is straightforward to see that this is a time-reversible Markov scheme $\mathcal{M}(K,G)$ and the uniform distribution on K is stationary.

We shall analyze, more generally, a filtered version of this random walk. Let F be a non-negative log-concave function on K, and let $\overline{F} = \int_K F(x)dx$. Let μ_F denote the measure with density function F and $Q_F = (1/\overline{F})\mu_F$. Specializing the definition of the Metropolis F-filtering for this case, we have, for any measurable set A with $x \notin A$,

$$\begin{aligned} P_x(A) &= \frac{1}{2\mathrm{vol}(G)} \int\limits_{(x+G)\cap A} \min\left\{1, \frac{F(y)}{F(x)}\right\} dy \\ &= \frac{1}{2\mathrm{vol}(G)} \int\limits_{(x+G)\cap A} \min\left\{\frac{1}{F(y)}, \frac{1}{F(x)}\right\} d\mu_F(y). \end{aligned} \tag{3.1}$$

and

$$P_x(x) = \frac{1}{2} + \frac{1}{2\mathrm{vol}(G)} \int\limits_{x+G} \max\left\{0, 1 - \frac{F(y)}{F(x)}\right\} dy. \tag{3.2}$$

By Lemma 1.9, the measure Q_F is stationary for our Markov scheme. Also note that this Markov scheme is lazy.

Our main goal is to estimate the mixing rate of the F-filtered lazy random walk with G-steps. By the results of Section 1, it suffices to estimate its conductance. Our main result in this section asserts that the conductance of this Markov scheme is determined in a sense by its local conductance.

Recall that we denote by H_t the set of points $x \in \mathbb{R}^n$ such that we have a chance of less than t of making a step from x (where $0 \le t \le 1/2$), i.e.,

$$\int\limits_{x+G} \min\{F(x), F(y)\}\, dy < tF(x)\mathrm{vol}(G).$$

We have noticed that setting $s = Q_F(H_t)/2$, the s-conductance is at most $2t$.

The main theorem of this section asserts that if the local conductance is large and the "diameter" of K is small, then the (global) conductance is large. Here the "diameter" of K is measured by $1/\theta$, where θ is the largest number such that for all $x, y \in K$,

$$\mathrm{vol}(G \cap (\theta(x-y)+G)) \ge \frac{1}{2}\mathrm{vol}(G).$$

Then $1/\theta$ is indeed the diameter in the norm whose unit ball is $\{x:\ \mathrm{vol}(G \cap (x+G)) \ge \mathrm{vol}(G)/2\}$. If $G = B$ is the euclidean ball, then Lemma 0.1 implies that θ is asymptotically $c/(\sqrt{n}d)$, where d is the euclidean diameter of K and $c = 1.4825\ldots$.

3.2. Theorem. *Let $0 \le t \le 1/2$, $0 < \theta < 1$ and $s = Q_F(H_t)$. Assume that for all $x, y \in K$,*

$$\mathrm{vol}(G \cap (\theta(x-y)+G)) \ge \frac{1}{2}\mathrm{vol}(G).$$

Then the $(7s/t)$-conductance of the Markov scheme $\mathcal{M}(K,G)/F$ is at least $(1/6)t^2\theta$.

Proof. Translating into more elementary terms, we want to prove the following. Consider a splitting of K into two measurable sets S_1 and S_2. Then

$$\int\limits_{S_1} \int\limits_{S_2 \cap (x+G)} \min\{F(x), F(y)\}\, dy\, dx \ge \frac{t^2\theta\mathrm{vol}(G)}{6} \min\{Q_F(S_1) - \frac{7s}{t},\ Q_F(S_2) - \frac{7s}{t}\}.$$

Of course, we may assume that $\mu_F(S_i) > 7s/t$ for $i = 1, 2$, else the assertion is obvious. Note that $\mathrm{vol}(G)\min\{\mu_F(S_1),\ \mu_F(S_2)\}$ is a trivial upper bound on left hand side.

Let, for $i = 1, 2$,

$$S_i' = \left\{x \in S_i:\ \mu_F((x+G) \cap S_{3-i}) \ge \frac{t}{3}\mathrm{vol}(G)F(x)\right\},$$

$$S_i'' = S_i \setminus S_i' \setminus H_t,$$

and

$$S_3 = S_1' \cup S_2' \cup H_t.$$

The key observation in the proof is the following.

Claim. *If $x_1 \in S_1''$ and $x_2 \in S_2''$ then* $\mathrm{vol}((x_1+G)\setminus(x_2+G)) > \dfrac{t}{3}\mathrm{vol}(G)$.

Assume (by way of contradiction) that

$$\mathrm{vol}((x_1+G)\setminus(x_2+G)) < \frac{t}{3}\mathrm{vol}(G). \tag{3.3}$$

Note that (3.3) is symmetric in x_1 and x_2. So we may assume that $F(x_1) \le F(x_2)$. Let $G_i = x_i + G$. Since $x_2 \notin H_t$, we have

$$\int\limits_{G_2} \min\{F(x_2), F(y)\}\, dy \ge tF(x_2)\mathrm{vol}(G). \tag{3.4}$$

Moreover, (3.3) implies that

$$\int\limits_{G_2\setminus G_1} \min\{F(x_2), F(y)\}\, dy < \frac{t}{3}F(x_2)\mathrm{vol}(G). \tag{3.5}$$

Subtracting (3.5) from (3.4), we obtain

$$\int\limits_{G_1\cap G_2} \min\{F(x_2), F(y)\}\, dy > \frac{2t}{3}F(x_2)\mathrm{vol}(G). \tag{3.6}$$

Since $x_2 \notin S_2'$, we have

$$\int\limits_{G_2\cap S_1} \min\{F(x_2), F(y)\}\, dy \le \frac{t}{3}F(x_2)\mathrm{vol}(G),$$

and hence

$$\int\limits_{G_1\cap G_2\cap S_1} \min\{F(x_2), F(y)\}\, dy \le \frac{t}{3}F(x_2)\mathrm{vol}(G). \tag{3.7}$$

Subtracting (3.7) from (3.6), we obtain

$$\int\limits_{G_1\cap G_2\cap S_2} \min\{F(x_2), F(y)\}\, dy > \frac{t}{3}F(x_2)\mathrm{vol}(G).$$

Using the trivial inequality

$$\min\{F(x_2), F(y)\} \le \frac{F(x_2)}{F(x_1)}\min\{F(x_1), F(y)\},$$

we derive that

$$\int\limits_{G_1\cap G_2\cap S_2} \min\{F(x_1), F(y)\}\, dy > \frac{t}{3}F(x_1)\mathrm{vol}(G),$$

and hence

$$\int\limits_{G_1\cap S_2} \min\{F(x_1),F(y)\}\,dy > \frac{t}{3}F(x_1)\mathrm{vol}(G).$$

But this inequality says that $x_1\in S_1'$, a contradiction. This proves the claim.

We can rephrase the claim as follows: If $x_1\in S_1''$ and $x_2\in S_2''$ then $\mathrm{vol}(G\setminus((x_2-x_1)+G))>\frac{t}{3}\mathrm{vol}(G)$. Hence Lemma 2.11 implies that for all $x_1\in S_1''$ and $x_2\in S_2''$,

$$\mathrm{vol}(G\setminus(\frac{2}{t}(x_2-x_1)+G))>\frac{1}{2}\mathrm{vol}(G).$$

Consider any $a,b\in K$ such that $S_1''\cap[a,b]$ and $S_2''\cap[a,b]$ are non-empty, and let $\rho|b-a|$ be the distance of these sets. Then by the above, $\mathrm{vol}(G\setminus((2/t)\rho(b-a)+G)>(1/2)\mathrm{vol}(G)$. On the other hand, by the definition of θ, $\mathrm{vol}(G\setminus(\theta(b-a)+G))\le(1/2)\mathrm{vol}(G)$. Hence $\rho>t\theta/2$.

So we can apply Theorem 2.6 to get that

$$\mu_F(S_3)>\frac{1}{2}t\theta\min\{\mu_F(S_1''),\mu_F(S_2'')\}.$$

Here

$$\mu_F(S_i'')\ge\mu_F(S_i)-\mu_F(S_i')-\mu_F(H_t)\ge\mu_F(S_i)-\mu_F(S_3)-s,$$

and so

$$\left(1+\frac{1}{2}t\theta\right)\mu_F(S_3)\ge\frac{1}{2}t\theta\min\{\mu_F(S_1)-s,\ \mu_F(S_2)-s\}.$$

Since $t\theta<1$, this implies that

$$\mu_F(S_3)\ge\frac{1}{3}t\theta\min\{\mu_F(S_1)-s,\ \mu_F(S_2)-s\}.$$

To complete the proof, we estimate the conductance as follows.

$$\begin{aligned}\int\limits_{S_1}\int\limits_{\substack{S_2\\ y\in x+G}}\min\{F(x),F(y)\}\,dy\,dx &\ge \int\limits_{S_1'}\int\limits_{\substack{S_2\\ y\in x+G}}\min\{F(x),F(y)\}\,dy\,dx\\ &\ge\int\limits_{S_1'}\frac{t}{3}F(x)\mathrm{vol}(G)\,dx=\frac{t}{3}\mathrm{vol}(G)\mu_F(S_1').\end{aligned}$$

Similarly,

$$\int\limits_{S_1}\int\limits_{\substack{S_2\\ y\in x+G}}\min\{F(x),F(y)\}\,dy\,dx\ge\frac{t}{3}\mathrm{vol}(G)\mu_F(S_2'),$$

and hence

$$\int_{S_1}\int_{\substack{S_2\\ y\in x+G}} \min\{F(x),F(y)\}\,dy\,dx \ge \frac{t}{6}\mathrm{vol}(G)\mu_F(S_1'\cup S_2')$$

$$= \frac{t}{6}\mathrm{vol}(G)\mu_F(S_3\setminus H_t)$$

$$\ge \frac{t}{6}\mathrm{vol}(G)(\mu_F(S_3)-s) \ge \frac{t^2\theta}{18}\mathrm{vol}(G)\min\{\mu_F(S_1)-s,\ \mu_F(S_2)-s\}-\frac{t}{6}s$$

$$\ge \frac{t^2\theta}{18}\mathrm{vol}(G)\min\{\mu_F(S_1)-\frac{7s}{t},\ \mu_F(S_2)-\frac{7s}{t}\}.$$

3.3. Corollary. *If the local conductance of $\mathcal{M}(K,G)/F$ is at least t at each point, then its conductance is at least $t^2\theta/18$.*

The following "Lipschitz type" condition gives an easy way to handle the local conductance of the filtered Markov scheme:

3.4. Corollary. *If $F(u)\ge t_1F(v)$ whenever $u-v\in G$, and the local conductance of $\mathcal{M}(K,G)$ is at least t_2 at each point, then the conductance of $\mathcal{M}(K,G)/F$ is at least $(t_1t_2)^2\theta/18$.*

b. Sampling from a convex body.

Generating an approximately uniformly distributed point in a convex body K is not only the crucial step in the algorithm of Dyer, Frieze and Kannan, but an important algorithmic question in statistics, optimization, simulation and other fields; cf. Smith (1984), Berbie at al (1987), McGeoch and Hoogs (1990). In this paper we need a modified version, sampling from a non-uniform distribution on a ball. But we shall also make use of the uniform case, to test whether a convex body is (approximately) contained in another. In fact, uniform sampling from a general convex body can be reduced to non-uniform sampling from a ball, as we are going to show now.

We generate this random point by using a random walk with B-steps. One could of course use any other centrally symmetric convex body G, provided one has an easy way of generating a random point from the uniform distribution over G. This would allow cubes, cross-polytopes, and others (and in fact the cube would perhaps be the most attractive from the programming point of view). The only reason why we choose the euclidean ball is that it is this case in which we can guarantee the best value of θ in Theorem 3.2.

A random point in the euclidean ball can be generated e.g. as follows. Let $\xi_1,\dots,\xi_n$ be independent random variables from a standardized normal distribution, and let η be uniformly distributed in $[0,1]$. Let $\mu_i=\eta^{1/n}\xi_i/\sqrt{\xi_1^2+\dots+\xi_n^2}$, then $v_0=(\mu_1,\dots,\mu_n)$ is uniformly distributed over G.

The idea is contained in the following lemmas. Let K be a convex body containing the origin in the interior and $x\in\mathbb{R}^n$. We denote by $\phi(x)=\phi_K(x)$ the least non-negative number t for which $x\in tK$. (If K is centrally symmetric, this is just the norm determined by K.) We set $F(x)=F_K(x)=e^{-\phi(x)}$, then $0<F(x)\le 1$.

3.5. Lemma.

$$\mathrm{vol}(K) = \frac{1}{n!} \int\limits_{\mathbb{R}^n} F(x)dx.$$

Proof. Let $h\colon \mathbb{R}_+ \to \mathbb{R}$ be any function such that $\int_0^\infty h(t)t^{n-1}dt$ exists. To evaluate

$$\int_{\mathbb{R}^n} h(\phi(x))\,dx$$

we partition the space into layers $S_t = \{x \,:\, t \le \phi(x) \le t+dt\}$. The measure of such a layer is

$$(t+dt)^n \mathrm{vol}(K) - t^n \mathrm{vol}(K) = nt^{n-1}\mathrm{vol}(K)\,dt.$$

Hence

$$\int\limits_{\mathbb{R}} h(\phi(x))\,dx = n \cdot \mathrm{vol}(K) \cdot \int\limits_0^\infty h(t)t^{n-1}dt. \tag{3.8}$$

In particular, we have

$$\int\limits_{\mathbb{R}^n} F(x)\,dx = n \cdot \mathrm{vol}(K) \cdot \int\limits_0^\infty t^{n-1}e^{-t}\,dt = n!\,\mathrm{vol}(K).$$

∎

(The fact that we obtain the factor $1/n!$ is nice but irrelevant: the point is that if g is any function of ϕ, then the integral of g over the space is the volume of K times a constant that depends on n only. It would be perfectly sufficient to use a good approximation of this constant.)

Put

$$\lambda(s) = \frac{1}{s}\left(\frac{1}{(n-1)!}\int\limits_0^s e^{-t}t^{n-1}\,dt\right)^{1/n}.$$

3.6. Lemma. *If v is a random vector in $\mathbb{R}^n$ with density function $e^{-\phi(v)}/(n-1)!$, then*

$$H(v) = \lambda(\phi(v))v$$

is uniformly distributed over K.

Proof. Set $h(s) = \frac{1}{(n-1)!}e^{-s}$. The set sK is mapped by H^{-1} onto the set $s\lambda(s)K$. The probability of $s\lambda(s)K$ in the uniform distribution is $(s\lambda(s))^n$; the probability of sK in the distribution Q_h is

$$\int\limits_{sK} h(x)\,dx = \int\limits_0^s h(s)s^{n-1}\,ds.$$

These two numbers are equal (that's how λ was chosen). Similar assertion holds for the intersection of these sets with any fixed cone. This implies that $H(v)$ is uniform.

Alternatively, the lemma can be verified directly by calculating the Jacobian of H. ∎

Let K be a convex body and assume the following:

(A1) More than $2/3$ of the volume of B is contained in K.

(A2) For some parameter $1 \le m \le n^{3/2}$, more than $2/3$ of the volume of K is contained in the convex body mB.

Let an "error bound" $\varepsilon > 0$ be also given.

The following algorithm generates a random point in K.

Sampling Algorithm. Set

$$q = 2mn\log(4/\varepsilon),$$
$$t = 10^{10} n q^2 (n\log n + \log(2/\varepsilon)).$$

Starting from the uniform distribution over B, do a lazy random walk $(v_0, \dots, v_t)$ in qB, with B-steps, filtered by F. Compute

$$w = H(v_t).$$

3.7. Theorem. *For every Lebesgue measurable set $A \subseteq K$, the random point w computed by the Sampling Algorithm satisfies*

$$\left| \mathrm{Prob}(H(v_t) \in A) - \frac{\mathrm{vol}(A)}{\mathrm{vol}(K)} \right| < \varepsilon.$$

Moreover, it needs $O(n^3 m^2 \log^2(1/\varepsilon)(n\log n + \log(1/\varepsilon)))$ membership tests and $O(n^4 m^2 \log^2(1/\varepsilon)(n\log n + \log(1/\varepsilon)))$ arithmetic operations using numbers of $O(\log n)$ bits.

We need some lemmas. Note that it follows from assumption (A1) and Lemma 0.1 that the ball $\frac{1}{3} n^{-1/2} B$ is contained in K, and from assumption (A2) and Lemma 2.10 that the ball $3nmB$ contains all of K. Hence

$$|x|/(3nm) \le \phi(x) \le 3\sqrt{n}|x|. \tag{3.9}$$

The first lemma enables us to restrict our attention to qB.

3.8. Lemma.

$$\left(1 - \frac{\varepsilon}{4}\right) \mathrm{vol}(K) < \frac{1}{n!} \int_{qB} F(x)dx.$$

Proof. Let $\overline{K} = K \cap 2mB$, $\overline{\phi}(x) = \phi_{\overline{K}}(x)$ and $\overline{F}(x) = e^{-\overline{\phi}(x)}$. Then clearly $\overline{F}(x) < F(x)$. Applying (3.8) with

$$h(t) = \begin{cases} e^{-t}, & \text{if } 0 \le t \le 2n, \\ 0, & \text{otherwise,} \end{cases}$$

we obtain

$$\int_{2nmB} F(x)\,dx \geq \int_{2n\overline{K}} F(x)\,dx \geq \int_{2n\overline{K}} \overline{F}(x)\,dx = n\mathrm{vol}(\overline{K}) \int_0^{2n} e^{-t}t^{n-1}\,dt > \frac{39}{40}n!\,\mathrm{vol}(\overline{K})$$
$$\geq \frac{13}{20}n!\,\mathrm{vol}(K).$$

So

$$\int_{\mathbb{R}^n \setminus 2nmB} F(x)\,dx \leq \frac{7}{20}n!\,\mathrm{vol}(K),$$

and hence by Theorem 2.8,

$$\int_{\mathbb{R}^n - 4\log(1/\varepsilon)nmB} F(x)\,dx < \left(\frac{7}{20}\right)^{\log(4/\varepsilon)} n!\,\mathrm{vol}(K) < \frac{\varepsilon}{4}n!\,\mathrm{vol}(K).$$

∎

We shall also need an estimate on the conductance of the random walk used in the algorithm.

3.9. Lemma. *If K satisfies (A1) and $p \geq 10\sqrt{n}$, then the conductance of $\mathcal{M}(pB, B)/F_K$ is at least $1/(20000p\sqrt{n})$.*

Proof. First we show that the local conductance of $\mathcal{M}(pB, B)/F_K$ is at least $1/25$. Let $x \in pB$. If y is a point chosen randomly from the uniform distribution over $x + B$, then the probability that $y \notin pB$ is less than $5/9$, by the assumption that $p \geq 10\sqrt{n}$. By (A1), the probability that $y \notin x + K$ is at most $1/3$; thus with probability at least $1/9$, we have $y \in pB \cap (x + K)$. Now for each such y, we have

$$y \in x + K \subseteq \phi(x)K + K = (\phi(x) + 1)K,$$

and hence $\phi(y) \leq \phi(x) + 1$ and thus $F_K(y) \geq (1/e)F_K(x)$. So the local conductance is at least $1/(9e) > 1/25$.

Second, Lemma 0.1 shows that

$$\theta > \frac{2}{3pn^{1/2}}.$$

Thus the lemma follows by Corollary 3.3. ∎

Proof of Theorem 3.7. Let, for $A \subseteq \mathbb{R}^n$,

$$Q_F(A) = \int_A F(u)\,du \Big/ \int_{\mathbb{R}^n} F(u)\,du\,,$$

$$Q'_F(A) = \int_{A\cap qB} F(u)\,du \Big/ \int_{qB} F(u)\,du\,,$$

and

$$Q_t(A) = \mathrm{Prob}(v_t \in A).$$

Then for every $A \subseteq K$,

$$\mathrm{Prob}(H(v_t) \in A) = \mathrm{Prob}(v_t \in H^{-1}(A)) = Q_t(H^{-1}(A))$$

and by Lemmas 3.5 and 3.6,

$$\frac{\mathrm{vol}(A)}{\mathrm{vol}(K)} = Q_F(H^{-1}(A)).$$

Thus it suffices to show that for every Lebesque measurable $U \subseteq \mathbb{R}^n$, we have

$$|Q_t(U) - Q_F(U)| \le \varepsilon.$$

By Lemma 3.8, we have

$$\begin{aligned} Q_F(U) - Q'_F(U) &= \frac{\mu_F(U)}{\mu_F(\mathbb{R}^n)} - \frac{\mu_F(U\cap qB)}{\mu_F(qB)} \le \frac{\mu_F(U)}{\mu_F(\mathbb{R}^n)} - \frac{\mu_F(U\cap qB)}{\mu_F(\mathbb{R}^n)} \\ &= \frac{\mu_F(U\setminus qB)}{\mu_F(\mathbb{R}^n)} \le \frac{\varepsilon}{4}, \end{aligned}$$

and similarly

$$\begin{aligned} Q'_F(U) - Q_F(U) &= \frac{\mu_F(U\cap qB)}{\mu_F(qB)} - \frac{\mu_F(U)}{\mu_F(\mathbb{R}^n)} \le \frac{\mu_F(U\cap qB)}{\mu_F(qB)} - \frac{\mu_F(U\cap qB)}{\mu_F(\mathbb{R}^n)} \\ &= \frac{\mu_F(U\cap qB)\mu_F(\mathbb{R}^n\setminus qB)}{\mu_F(qB)\mu_F(\mathbb{R}^n)} \le \frac{\varepsilon}{4}. \end{aligned}$$

On the other hand, we can estimate $Q'_F - Q_t$ by Corollary 1.5:

$$|Q_t(U) - Q'_F(U)| \le \sqrt{M}\left(1 - \frac{1}{2}\Phi^2\right)^t,$$

where, by (3.9),

$$M = \sup_A \frac{Q_0(A)}{Q'_F(A)} = \frac{Q_0(B)}{Q'_F(B)} = \frac{\mu_F(qB)}{\mu_F(B)} \le \frac{n!\,\mathrm{vol}(K)}{e^{-3\sqrt{n}}\mathrm{vol}(B)} \le n^{5n},$$

and by Lemma 3.9,

$$\Phi \ge \frac{1}{20000q\sqrt{n}}.$$

Hence

$$|Q_t(U) - Q'_F(U)| \le n^{(5/2)n}\left(1-\frac{1}{2}\Phi^2\right)^t < \frac{\varepsilon}{2}.$$

The running time estimate is straightforward, except perhaps for the precision, whose discussion we omit here since it will be discussed in Section 3 in connection with the volume algorithm. ∎

Remark. The following algorithm to generate a random point in K is perhaps more natural: Let v_0 be uniformly distributed over B. We do a lazy random walk $(v_0, v_1, \ldots)$ in K with G-steps, where $G = \frac{\varepsilon}{100n}B$ (no filtering!), and stop after $t = O^*(n^4m^2)$ steps. This algorithm achieves the same result, but the analysis is more complicated since we have to use s-conductance instead of conductance.

4. A volume algorithm

a. Outline.

Let K be a convex body in $\mathbb{R}^n$, given by a separation oracle, together with two numbers $r, R > 0$ such that K is contained in the ball RB and it contains some ball with radius r. In addition to K, the input to the algorithm consists of two numbers $0 < \delta, \varepsilon < 1$, bounds on the error probability and relative error.

Our algorithm computes a (random) real number ζ such that with probability at least $1-\delta$,

$$(1-\varepsilon)\mathrm{vol}(K) \le \zeta \le (1+\varepsilon)\mathrm{vol}(K).$$

We may assume that $\varepsilon,\ \delta < 1$. The algorithm will be polynomial in $\log(R/r)$, n, $1/\varepsilon$, and $\log(1/\delta)$.

The algorithm consists of a preliminary part and a main part. The *preliminary part* transforms K by an affine transformation to a convex body $A(K)$ so that the following two conditions (also formulated in Section 2) are met:

(A1) More than $2/3$ of the volume of B is contained in $A(K)$.

(A2) For some parameter $m \in [1, n^{3/2}]$, more than $2/3$ of the volume of $A(K)$ is contained in the convex body mB.

The smaller m we can achieve, the faster the main part of our algorithm. We can always achieve $m = n^{3/2}$ deterministically, and $m = n$ randomized, but for special cases like centrally symmetric bodies, or polyhedra with a polynomial number of facets, smaller values of m can be chosen, thereby reducing the running time of the main part.

The second, *main part* computes the volume, based on the identity in Lemma 3.5. In this part we only need a membership oracle for the body.

We describe the algorithm in exact real arithmetic; we then show that rounding does not introduce substantial errors.

b. A statistical trick.

It will be convenient not to have to worry about the error probability δ. The following well-known trick [see e.g. Jerrum, Valiant and Vazirani (1986)] shows that it suffices to solve the problem with $\delta=1/3$. Assume that we have an algorithm that computes a random variable ζ that falls in the interval $I=[(1-\varepsilon)\mathrm{vol}(K),\ (1+\varepsilon)\mathrm{vol}(K)]$, with probability at least $2/3$. Let $s=10\log(1/\delta)$, and let us compute $2s+1$ independent values of ζ, say $\zeta_1,\dots,\zeta_{2s+1}$. Let ζ_{i_0} be the median of them, i.e. the $(s+1)$-st of them when ordered increasingly. We claim that

$$\mathrm{Prob}(\zeta_{i_0}\in I)\geq 1-\delta.$$

In fact, if the median is not in the interval I, then at least $s+1$ of these $2s+1$ values are outside I. But by Chernoff's inequality (1952) [see also e.g. Bollobás (1987)], the probability that out of $2s+1$ independent events, all having probability at most $1/3$, more than half occur is at most $e^{-(s/10)}\leq\delta$.

c. Preliminary part.

We describe a deterministic and a randomized algorithm to achieve (A1) and (A2) in the general case, and randomized algorithms for special classes of convex bodies.

(1) A standard application of the ellipsoid method (see [GLS]) gives that we can find an affine transformation A such that $B\subseteq A(K)\subseteq n^{3/2}B$; such an affine tranformation A can be computed using $O\big(n^4\log(R/r)\big)$ operations with numbers with $O\big(n^2(|\log R|+|\log r|)\big)$ digits. This way we have achieved $m=n^{3/2}$. This "rounding" was used in [DFK] and [LS]. For polyhedra given by explicit linear inequalities, $m=O(n)$ can be achieved by the same method.

To improve this result, we have to examine why we loose a factor of $n^{1/2}$ for general convex bodies. Recall that the ellipsoid method maintains an ellipsoid E that contains the body K. An *ellipsoid step* replaces E by another ellipsoid containing K with smaller volume. For this, we use a cutting plane: a halfspace H also containing K. If H does not contain the ellipsoid $E/(2n)$ obtained from E by shrinking it from its center by a factor of $2n$, then one computes an ellipsoid $E'\supseteq E\cap H\supseteq K$ with smaller volume than E.

The crucial step in implementing this scheme is testing whether K contains $E/(2n)$ or, equivalently, whether an appropriate affine image $A(K)$ contains the unit ball B. Unfortunately, there is no polynomial-time test available for this (in fact, the problem is exponentially hard, by the results of [BF]). In [GLS] the following approximation method was used: we test if the vectors $\pm e_i$ $(i=1,\dots,n)$ belong to $A(K)$. If the answer is no, we know $A(K)$ does not contain B, and have a vector in $B\setminus A(K)$. If the answer is yes, we know that the ball $n^{-1/2}B$ belongs to $A(K)$ (loosing the factor of $n^{1/2}$).

(2) We could stop here having achieved (A1) and (A2) with $m=n^{3/2}$; but this would cost a factor of n in the running time of the main part. Instead, we go on using a randomized test for $B\subseteq A(K)$ by selecting $T=\lceil 3\log n\rceil$ independent random points in B, and testing if they belong to $A(K)$. If one of them does not, we have a point in $B\setminus A(K)$ needed to carry out the ellipsoid step. If all of them belong to $A(K)$, we conclude that at least $2/3$ of the volume of B must be in $A(K)$. If less than $2/3$ of the volume of B is contained in $A(K)$, then the probability that we do not find a point in the difference is less than

$(2/3)^T \le 1/(100n^2 \log n)$. The procedure lasts at most $25n^2 \log n$ ellipsoid steps, so the probability that it halts with $A(K)$ containing less than 2/3 of B is less than 1/4.

This gives a randomized algorithm that achieves (A1) and (A2), with $m = O(n)$, with error probability less than 1/4. Note that (A2) is achieved in a much stronger sense: the ball mB in fact contains $A(K)$. (Independently, a similar randomized rounding algorithm was worked out by U. Faigle)

(3) In the case of centrally symmetric bodies, exactly the same improvement over the ratio n in [GLS] works and one achieves (A1) and (strong) (A2) with $m = O(n^{1/2})$.

(4) We can also turn things around and find an affine transformation that achieves (A1) in the strong sense that B is contained in $A(K)$, while achieving (A2) in the original sense. For this, we can apply a "polar" form of the ellipsoid method, used by Lenstra (1983). In this, we maintain that the body $A(K)$ contains the unit ball, and apply affine transformations to reduce the radius of the circumscribed ball. If $A(K)$ contains a vector x with $|x| > 6n$, then we can apply the following affine transformation: we expand the body K by a factor of $1 + (1/n)$ in all directions orthogonal to x, shrink it by a factor of 3 in the direction of x, and translate it by $-(2/3)x/|x|$. The resulting body still contains B, and this operation reduces the volume of $A(K)$ by a factor of $(1+1/n)^{n-1}/3 < e/3$. If we start with the result of (1), then after at most $O(n \log n)$ applications of this step, the circumscribed ball can be reduced to $6nB$.

Of course, we need to test whether $A(K) \subseteq 6nB$ and if not then find a point in $6nB \setminus A(K)$. We can do this by generating $T = \lceil \log_{3/2}(10n \log n) \rceil$ independent random points in $A(K)$ and testing if they belong to $6nB$. If one of them does not, we have a point in $A(K) \setminus 6nB$ needed to carry out the ellipsoid step. If all of them belong to $6nB$, we conclude that at least 2/3 of the volume of $A(K)$ must be in $6nB$. If less than 2/3 of the volume of B is contained in $A(K)$, then the probability that we do not find a point in the difference is less than $(2/3)^T \le 1/(10n \log n)$.

The cost of generating random points in $A(K)$ is substantial: if we use the Sampling Algorithm from the previous section (with ε a small constant), we need $O(n^4 m^2 \log n)$ membership tests and $O(n^5 m^2 \log n)$ arithmetic operations. With a little trick, we can take $m = O(n)$ here: we can apply the Sampling Algorithm to $K' = A(K) \cap (18n)B$, since it follows from Theorem 2.7(b) that if $\mathrm{vol}(A(K) \setminus (6nB)) = \theta \mathrm{vol}(A(K))$ where $\theta > 1/3$, then $\mathrm{vol}(A(K) \setminus (18nB)) \le \theta^2 \mathrm{vol}(A(K))$ and hence $\mathrm{vol}(K' \setminus (6n)B) \ge \theta/(1+\theta)\mathrm{vol}(K') > (1/4)\mathrm{vol}(K')$.

We need to generate $O(n \log^2 n)$ random points altogether, which makes the cost of this phase $O(n^7 \log^3 n)$, at most a log factor more than the cost of the main part.

At this point, it is not clear that we have won anything since the ratio $m = O(n)$ can be achieved in a much simpler fashion following (2). But for the special case (which is perhaps the most important) when $A(K)$ is a polytope with a polynomial number of facets, we do gain more. If $A(K)$ has n^c facets, then we can shrink $A(K)$ by a factor of $\Omega(\sqrt{n}/\log n)$ and still have 2/3 of the volume of B in $A(K)$. In fact, each facet of $A(K)$ cuts off a fraction of less that $1/(3n^c)$ of the volume of B by Lemma 2.1. So for polytopes with a polynomial number of facets, (A1) and (A2) can be achieved with $m = O(\sqrt{n} \log n)$.

(5) Similarly, if K is a centrally symmetric polytope with a polynomial number of

facets, then $m = O(\log n)$ can be achieved in time not exceeding the time of the main part.

d. The main volume algorithm.

Let us assume that K is a convex body such that $\mathrm{vol}(K \cap B) \geq (2/3)\mathrm{vol}(B)$ and $\mathrm{vol}(K \cap (mB)) \geq (2/3)\mathrm{vol}(K)$. Let ϕ and F be defined as in Section 2. For $i = 0, 1, \ldots$, we define the function

$$F_i(x) = \min\{F(x), \exp(-|x| n^{1/2} 2^{-i/n})\}.$$

Instead of computing the volume, we shall integrate the function F, using the formula of Lemma 3.5:

$$\mathrm{vol}(K) = \frac{1}{n!} \int_{\mathbb{R}^n} F(x) dx,$$

and the approximation provided by Lemma 3.8. Our algorithm follows the same general pattern as the algorithms in [DFK], [LS], [AK] and [DF], but details are different (and in fact somewhat simpler).

Volume Algorithm. (a) Set

$$q = 4mn \log(1/\varepsilon),$$
$$k = 4n \log n,$$
$$t = 10^{11} n k q^2 \varepsilon^{-2}.$$

(b) For $i = 1, \ldots, k$, we do a lazy random walk $(v_0^i, \ldots, v_{3t}^i)$ in qB, with B-steps, filtered by F_i. We use the notation of Section 2, but with qB playing the role of K. Thus

$$Q_{F_i}(A) = \int_A F_i(u)\, du \Big/ \int_{qB} F_i(u)\, du.$$

The first walk starts at a point generated from the distribution Q_{F_0}. The i-th walk starts where the $(i-1)$-st ended. We compute

$$\alpha_i = \frac{1}{t} \sum_{j=t+1}^{2t} \frac{F_{i-1}(v_j^i)}{F_i(v_j^i)}$$

(The first t steps of each phase serve to get close to the stationary distribution of the given phase. So the points v_j^i $(j = t+1, \ldots, 2t)$ will have a distribution over qB that is very close to Q_{F_i}. Also note that $F_i = F_{K_i}$, where $K_i = K \cap 2^{i/n} B$. Therefore, α_i will be a good estimate on

$$\int_{qB} \frac{F_{i-1}}{F_i}\, dQ_{F_i} = \int_{qB} F_{i-1}(x)\, dx \Big/ \int_{qB} F_i(x)\, dx \approx \frac{\mathrm{vol}(K_{i-1})}{\mathrm{vol}(K_i)}.$$

The last t steps ensure independence from the next phase. The first t steps of each phase could be replaced by a shorter walk and the last t steps could be omitted, at the cost of a more complicated analysis.)

(c) Estimate $\mathrm{vol}(K)$ by

$$\zeta = (\alpha_1 \cdots \alpha_k)^{-1} \mathrm{vol}(B).$$

e. Analysis of the algorithm.

4.1. Theorem. *If K is a convex body such that* $\mathrm{vol}(K\cap B)\geq(2/3)\mathrm{vol}(B)$ *and* $\mathrm{vol}(K\cap(mB))\geq(2/3)\mathrm{vol}(K)$*, and ζ is the estimate on* $\mathrm{vol}(K)$ *given by the Volume Algorithm, then*

$$\mathrm{Prob}((1-\varepsilon)\zeta<\mathrm{vol}(K)<(1+\varepsilon)\zeta)\geq 3/4.$$

Moreover, the algorithm uses

$$kt=O(n^3m^2\varepsilon^{-2}(\log m+\log\log(1/\varepsilon))^2(n+\log(1/\varepsilon))^2\log(1/\varepsilon)^2)$$

membership oracle calls and

$$ktO(n)=O(n^4m^2\varepsilon^{-2}(\log m+\log\log(1/\varepsilon))^2(n+\log(1/\varepsilon))^2\log(1/\varepsilon)^2)$$

arithmetic operations on numbers with $O(\log n+\log(1/\varepsilon))$ bits.

(Ordinarily, a membership test takes at least n arithmetic operations (it has to look at the coordinates), so the running time will be dominated by the membership tests.)

Combining this theorem with our discussions in the Preliminary part, we obtain several corollaries. For sake of simplicity, we assume here that $\varepsilon>2^{-n}$, and $R/r<n^n$. We only give the number of oracle calls; each oracle call goes with $O(n)$ arithmetic operations on numbers with $O(\log n)$ bits.

4.2. Corollary. (a) *If K is a convex body given by a well-guaranteed separation oracle, then we can compute a random value ζ using $O(n^7\varepsilon^{-2}\log^2 n\log^2(1/\varepsilon)\log(1/\delta))$ oracle calls, so that*

$$\mathrm{Prob}((1-\varepsilon)\zeta\leq\mathrm{vol}(K)\leq(1+\varepsilon)\mathrm{vol}(K))\geq 1-\delta.$$

(b) *If K is centrally symmetric, or K is a polytope with a polynomial (in n) number of facets, then we need only $O(n^6\varepsilon^{-2}\log^4 n\log^2(1/\varepsilon)\log(1/\delta))$ membership tests.*

(c) *If K is a centrally symmetric polytope with a polynomial (in n) number of facets, then we need only $O(n^5\varepsilon^{-2}\log^4 n\log^2(1/\varepsilon)\log(1/\delta))$ oracle calls.* ∎

Proof of Theorem 4.1. We need some preliminary observations. Since $F_i=F_{K_i}$, the functions F_i are log-concave. Note that each K_i also satisfies conditions (A1) and (A2), and we have

$$F_0=\exp(-|x|n^{1/2})\leq F_1\leq\ldots\leq F_k=F.$$

Let $W_i=\int_{qB}F_i\,dy$. Then Lemma 3.8 implies that

$$\left(1-\frac{\varepsilon}{4}\right)n!\,\mathrm{vol}(K_i)<W_i<n!\,\mathrm{vol}(K_i).$$

For the case $i=0$, the value

$$\int_{\mathbb{R}^n}\exp(-\sqrt{n}|x|)\,dx=n^{-n/2}n!\,\mathrm{vol}(B)$$

is easily computed, and W_0 is within a factor of $(1-\frac{\varepsilon}{4})$ of this value.

Claim 1.

$$\frac{1}{2} \le \frac{W_{i-1}}{W_i} \le 1. \tag{4.1}$$

Up to an error of ε, this follows immediately from the observation that it is essentially the ratio of the volumes of K_{i-1} and K_i. Precisely, we have

$$\begin{aligned} W_{i-1} &= \int_{qB} F_{i-1}(y)dy \ge \int_{2^{-1/n}qB} F_{i-1}(y)\,dy \\ &= \frac{1}{2}\int_{qB} F_{i-1}(2^{-1/n}x)\,dx \ge \frac{1}{2}\int_{qB} F_i(x)\,dx = \frac{1}{2}W_i. \end{aligned}$$

∎

Let P_p^i denote the distribution of the p-th point in the i-th random walk, and consider, as in Section 1, the distance function h_p^i of P_j^i and Q_{F_i}. Since the latter measure is atom-free, we have by Lemma 1.2(ii),

$$h_p^i(x) = \sup P_p^i(A) - x,$$

$(i=1,\dots,k)$, where A ranges over all measurable subsets of qB with $Q_{F_i}(A)=x$. For $i=0$ we define $h_p^0(x)=0$. Let $\Phi=\min_i \Phi_i$, and $\eta=\left(1-\Phi^2/2\right)^t$. By Lemma 3.9,

$$\Phi \ge \frac{1}{20000q\sqrt{n}}.$$

Hence $\eta < e^{-t\Phi^2/2} < m^{-128n/\varepsilon^2}$.

Claim 2. *For every $0\le x\le 1$, $i\ge 1$ and $t\le p\le 3t$, we have*

$$h_p^i(x) \le \begin{cases} 2\eta\min\{\sqrt{x},\sqrt{1-x}\}, & \text{if } t\le p\le 3t, \\ 4\min\{\sqrt{x},\sqrt{1-x}\}, & \text{otherwise.} \end{cases}$$

Proof. By induction on i. For $i=0$ and $p=3t$ the assertion may be considered true. Let $i>0$. Note that we have, for a suitable measurable set $A\subseteq qB$ with $Q_{F_i}(A)=x$,

$$h_0^i(x) = P_0^i(A) - x = P_{3t}^{i-1}(A) - x.$$

Let $y=Q_{F_{i-1}}(A)$. Then

$$\begin{aligned} y &= \frac{1}{W_{i-1}}\int_A F_{i-1}(u)\,du \le \frac{1}{W_{i-1}}\int_A F_i(u)\,du \\ &\le \frac{2}{W_i}\int_A F_i(u)\,du = 2Q_{F_i}(A) = 2x. \end{aligned}$$

Hence $y-x\le x$. Similarly $1-y\le 2(1-x)$. Therefore, by the induction hypothesis,

$$\begin{aligned} h_0^i(x) &= P_{3t}^{i-1}(A) - x = P_{3t}^{i-1}(A) - y + (y-x) \le h_{3t}^{i-1}(y) + \min\{x,1-x\} \\ &\le 2\eta\min\{\sqrt{y},\sqrt{1-y}\} + \min\{\sqrt{x},\sqrt{1-x}\} \\ &< (1+4\eta)\min\{\sqrt{x},\sqrt{1-x}\}. \end{aligned}$$

Hence by Theorem 1.4, we have for $p \geq t$,

$$h_p^i(x) \leq (1+4\eta)\min\{\sqrt{x}, \sqrt{1-x}\}\left(1-\frac{1}{2}\Phi_i^2\right)^p$$
$$\leq (1+4\eta)\min\{\sqrt{x}, \sqrt{1-x}\}\eta < 2\eta\min\{\sqrt{x}, \sqrt{1-x}\}.$$

The case $p < t$ follows trivially. ∎

Remark. This claim remains valid if we start our random walk from any other distribution P satisfying

$$|P(A) - Q_{F_1}(A)| < 4\min\{\sqrt{Q_{F_1}(A)},\ \sqrt{1-Q_{F_1}(A)}\},$$

and it is enough to assume that for each $j < i$, we walk at least t steps filtered by F_j.

Consider two vertices $u = v_a^i$ and $w = v_b^j$ where $i < j$ and $t < a, b \leq 2t$. The next lemma estimates the correlation between $f(u) = F_{i-1}(u)/F_i(u)$ and $g(w) = F_{j-1}(w)/F_j(w)$.

Claim 3. $|\mathbf{E}(f(u)g(w)) - \mathbf{E}(f(u))\mathbf{E}(g(w))| \leq 6\eta$.

Proof. Let us start another (inhomogeneous) random walk $\overline{v}_a^i, \overline{v}_{a+1}^i, \ldots, \overline{v}_b^j$, where the transition probabilities from a $\overline{v}$ to the next are the same as from the v with the same indices, but the distribution of the starting point $\overline{v}_a^i$ is "filtered" by the function f in the following sense. Put

$$c_f = \frac{1}{\int_{qB} f\, dP_a^i} = \frac{W_i}{W_{i-1}},$$

then $1 \leq c_f \leq 2$. Further, put

$$\overline{P}_a^i(A) = \mathrm{Prob}(v_a^i \in A) = \frac{\int_A f\, dP_a^i}{\int_{qB} f\, dP_a^i} = c_f \cdot \int_A f\, dP_a^i.$$

Let $\overline{P}_r^k$ denote the distribution of $\overline{v}_r^k$ ($i \leq k \leq j$, $0 \leq r \leq 2t$), and let

$$\overline{h}_r^k(x) = \sup_{Q_{F_k}(A)=x} \overline{P}_r^k(A) - Q_{F_k}(A).$$

Then we have

$$\mathbf{E}(f(u)g(w)) = \int_{qB} \mathbf{E}(f(x)g(w) \mid u = x)\, dP_a^i(x) = \int_{qB} \mathbf{E}(g(w) \mid u = x) f(x)\, dP_a^i(x)$$
$$= \int_{qB} \mathbf{E}(g(w) \mid u = x)\, d\overline{P}_a^i(x) \int_{qB} f\, dP_a^i = \mathbf{E}(g(\overline{v}_b^j))\mathbf{E}(f(u)).$$

Hence, using that $0 \leq f, g \leq 1$,

$$|\mathbf{E}(f(u)g(w)) - \mathbf{E}(f(u))\mathbf{E}(g(w))| = \left|\mathbf{E}(f(u))\left(\mathbf{E}(g(\overline{v}_b^j)) - \mathbf{E}(g(v_b^j))\right)\right|$$
$$= \mathbf{E}(f(u))\left|\int_{qB} g(y)\,(d\overline{P}_b^j(y) - dP_b^j(y))\right| \leq |\overline{P}_b^j - P_b^j|$$
$$\leq |\overline{P}_b^j - Q_{F_i}| + |P_b^j - Q_{F_i}|.$$

Here the second term is at most 2η by Claim 2. To estimate the first, we use that

$$\overline{h}_a^i(z) = \sup_{Q_{F_i}(A)=z} \left[\overline{P}_a^i(A) - Q_{F_i}(A)\right].$$

Let $A \subseteq qB$, $Q_{F_i}(A) = z$. Then

$$\begin{aligned}\overline{P}_a^i(A) - Q_{F_i}(A) &= \int_A c_f f(x)\, dP_a^i(x) - z \\ &\le \int_A c_f f\,[dP_a^i - dQ_{F_i}] + \int_A c_f f(x)\, dQ_{F_i}(x) - z \le c_f h_a^i(y) + c_f y - z,\end{aligned}$$

where $y = \int_A f(x)\, dQ_{F_i}(x) \le z$. Hence by Claim 2,

$$\overline{h}_a^i(z) \le 5\sqrt{z}.$$

On the other hand, trivially

$$\overline{h}_a^i(z) \le 1 - z \le 5\sqrt{1-z}.$$

Thus we may apply the argument of Claim 2 to the other Markov chain to get

$$\overline{h}_b^j(z) \le 2\eta,$$

and so

$$|P_b^j - Q_{F_i}| \le 2\eta.$$

∎

Now we can turn to estimating the error of the algorithm. The error of our estimate ζ comes from three main sources:

I. $W_k \ne \mathrm{vol}(K)$, $W_0 \ne \mathrm{vol}(B)$. This is taken care of by Lemma 3.8.

II. The **probabilistic error** of our estimate ζ has itself three possible sources:

(E1) The distribution of the points v_{t+1}^i generated in step (c) is not exactly $(1/W_i)Q_{F_i}$.

(E2) The points v_{t+1}^i are not independent.

(E3) The sum used to estimate $\frac{W_{i-1}}{W_i}$ has a standard deviation.

Claims 2 and 3 will be used to handle (E1) and (E2), and Theorem 1.8, to handle (E3).

We want to show that

$$\mathrm{Prob}\left(1 - \frac{\varepsilon}{2} \le \prod_{i=1}^k \alpha_i \Big/ \prod_{i=1}^k E(\alpha_i) \le 1 + \frac{\varepsilon}{2}\right) > 3/4. \tag{4.2}$$

Set $\alpha_i / E(\alpha_i) = 1 + \beta_i$, then (4.2) becomes

$$\mathrm{Prob}\left(1 - \frac{\varepsilon}{2} \le \prod_{i=1}^k (1+\beta_i) \le 1 + \frac{\varepsilon}{2}\right) > 3/4.$$

We split those outcomes when the event fails into three cases, each having probability less that 1/8.

Case 1. There exists an i such that $|\beta_i| > 1/2$. For a given i, we can estimate the probability of this by Chebyshev's inequality. Since $\mathbf{E}(\beta_i) = 0$, and $1/2 \le \mathbf{E}(\alpha_i) \le 1$, we have

$$\text{Prob}\left(|\beta_i| > \frac{1}{2}\right) \le 4D^2(\beta_i).$$

By Theorem 1.8 applied with $F = F_{i-1}/F_i$ (which satisfies $0 \le F \le 1$), we get

$$D^2(t\alpha_i) \le \frac{4t}{\Phi_i^2}\|F\|^2 \le \frac{4t}{\Phi^2},$$

and hence

$$D^2(\alpha_i) \le \frac{4}{t\Phi^2} \le \frac{\varepsilon^2}{32k}.$$

Thus

$$\mathbf{E}(\beta_i^2) = D^2(\beta_i) = \frac{D^2(\alpha_i)}{E^2(\alpha_i)} \le \frac{\varepsilon^2}{8k}. \tag{4.3}$$

Hence the probability that there exists an i with $|\beta_i| > 1/2$ is at most 1/8.

Case 2. A similar argument settles the case when $\sum_{i=1}^k \beta_i^2 > \varepsilon/8$. In fact, by Markov's inequality and (4.3),

$$\text{Prob}\left(\sum_{i=1}^k \beta_i^2 > \frac{\varepsilon}{8}\right) \le \mathbf{E}\left(\sum_{i=1}^k \beta_i^2\right) \Big/ \frac{\varepsilon}{8} \le \frac{1}{8}.$$

Case 3. So assume that $|\beta_i| \le 1/2$ for all i and $\sum_{i=1}^k \beta_i^2 \le \varepsilon/4$. Then

$$|\log(1+\beta_i) - \beta_i| < \beta_i^2.$$

If

$$\left|\prod_{i=1}^k (1+\beta_i) - 1\right| > \frac{\varepsilon}{2}$$

then

$$\left|\sum_{i=1}^k \log(1+\beta_i)\right| > \frac{\varepsilon}{4},$$

and hence

$$\left|\sum_{i=1}^k \beta_i\right| > \frac{\varepsilon}{4} - \sum_{i=1}^k \beta_i^2 > \frac{\varepsilon}{8}.$$

Since $\mathbf{E}(\beta_i)=0$, the probability of this can be estimated by Chebyshev's inequality:

$$\mathrm{Prob}\left(\left|\sum_{i=1}^{k}\beta_i\right|>\frac{\varepsilon}{8}\right)\le\frac{64D^2(\sum_i\beta_i)}{\varepsilon^2}. \tag{4.4}$$

Now here

$$D^2(\sum_i\beta_i)=\mathbf{E}\left(\left(\sum_{i=1}^{k}\beta_i\right)^2\right)=\sum_{i=1}^{k}\mathbf{E}(\beta_i^2)+2\sum_{1\le i<j\le k}\mathbf{E}(\beta_i\beta_j).$$

The first term is at most $\varepsilon^2/8$ by (4.3). To estimate the second, write

$$\mathbf{E}(\beta_i\beta_j)=\frac{1}{\mathbf{E}(\alpha_i)\mathbf{E}(\alpha_j)}\big[\mathbf{E}(\alpha_i\alpha_j)-\mathbf{E}(\alpha_i)\mathbf{E}(\alpha_j)\big].$$

Here, with the notation of Claim 3,

$$\alpha_i\alpha_j=\frac{1}{t^2}\sum_{p=t+1}^{2t}\sum_{q=t+1}^{2t}f(v_a^i)g(v_b^j),$$

and hence by Claim 3,

$$\begin{aligned}\mathbf{E}(\alpha_i\alpha_j)&=\frac{1}{t^2}\sum_{p=t+1}^{2t}\sum_{q=t+1}^{2t}\mathbf{E}(f(v_a^i)g(v_b^j))\le\frac{1}{t^2}\sum_{p=t+1}^{2t}\sum_{q=t+1}^{2t}\mathbf{E}(f(v_a^i))\mathbf{E}(g(v_b^j))+4\eta\\&=\mathbf{E}(\alpha_i)\mathbf{E}(\alpha_j)+4\eta.\end{aligned}$$

Thus

$$\mathbf{E}(\beta_i\beta_j)\le 16\eta<2^{-10}\frac{\varepsilon^2}{k^2}$$

and so

$$D^2(\sum_i\beta_i)\le\frac{\varepsilon^2}{4},$$

whence by (4.4),

$$\mathrm{Prob}\left(\left|\sum_{i=1}^{k}\beta_i\right|>\frac{\varepsilon}{8}\right)<\frac{1}{8}.$$

III. **Numerical errors** come from two sources: first, we can compute $\phi(x)$ only approximately and second, we have desribed the algorithm assuming that arithmetic operations with real numbers can be carried out exactly. We compute $\phi(x)$ with absolute error at most ε/n^2; this means a relative error of ε/n^2 in $F_i(x)$, and hence a relative error of ε/n^2 in α_i. Thus this adds up to a relative error of $k\varepsilon/n^2<\varepsilon/10$ in ζ.

Next we estimate the error coming from the fact that we calculate with finite precision, in fact to $100\log n$ bits after the point. The run of the algorithm is determined

by three random sequences: $b_1^1, b_2^1, \ldots, b_p^i, \ldots$, independent random bits (to secure the laziness), $u_1^1, u_2^1, \ldots, u_p^i, \ldots$ independent uniformly distributed vectors in B (the jumps), and $\tau_1^1, \tau_2^1, \ldots, \tau_p^i, \ldots$ independent uniformly distributed number in $[0,1]$ (the Metropolis filtering). We have

$$v_{p+1}^i = \begin{cases} v_p^i + u_p^i, & \text{if } b_p^i = 1,\ v_p^i + u_p^i \in qB \text{ and } \tau_p^i \le F_i(v_p^i + u_p^i)/F_i(v_p^i), \\ v_p^i, & \text{otherwise.} \end{cases} \tag{4.5}$$

In the actual computation, we get the vectors u_p^i the numbers τ_p^i and the values of F_i with some error: let $\hat{u}_p^i$, $\hat{\tau}_p^i$ and $\hat{F}_i$ be these rounded values. Then we determine the process $\hat{v}_1^1, \hat{v}_2^1, \ldots, \hat{v}_p^i \ldots$ by

$$\hat{v}_{p+1}^i = \begin{cases} \hat{v}_p^i + \hat{u}_p^i, & \text{if } b_p^i = 1,\ \hat{v}_p^i + \hat{u}_p^i \in qB \text{ and } \hat{\tau}_p^i \le \hat{F}_i(v_p^i + u_p^i)/\hat{F}_i(v_p^i), \\ \hat{v}_p^i, & \text{otherwise.} \end{cases} \tag{4.6}$$

We claim that with large probability, the same alternative is chosen in (4.5) and (4.6) at each step; this implies then that $|v_p^i - \hat{v}_p^i| < n^{-10}$ for all i and p, and hence the ζ we compute is within a factor of $1+\varepsilon/2$ to the true value.

Assume that not always the same choice is made in (4.5) and (4.6), and consider the first occurance of this. Then either (a) $v_p^i + u_p^i \in qB$ but $\hat{v}_p^i + \hat{u}_p^i \notin qB$ (or vice versa), or (b) $\tau_p^i \le F_i(v_p^i + u_p^i)/F_i(v_p^i)$ but $\hat{\tau}_p^i > \hat{F}_i(v_p^i + u_p^i)/\hat{F}_i(v_p^i)$ (or vice versa). If (b) occurs, then $|\tau_p^i - F_i(v_p^i + u_p^i)/F_i(v_p^i)| < n^{-100}$, and the probability that this occurs is less than $2/n^{100}$. Assume that (a) occurs. Since we consider the first occurance, we have $|v_p^i - \hat{v}_p^i| < n^{-90}$ and hence the distance of v_p^i from the boundary of qB must be less than n^{-89}. Let S denote that n^{-89}-neighborhood of the boundary of qB. It is easy to compute that $Q_{F_i}(S) < n^{-88}$, and therefore by Claim 2,

$$\text{Prob}(v_p^i \in S) < n^{-44}.$$

So the probability that either (a) or (b) ever occurs is less than $2kt/n^{44} < 1/50$.

The **running time** of the Volume Algorithm is clearly dominated by the time needed to carry out the random walk. This takes kt moves, where each move takes the updating of n coordinates, one test of membership in qB, the evaluation of $\phi(x)$, and a constant number of arithmetic operations (where the evaluation of e^x is considered as a single arithmetic operation; else, we need $O(\log n)$ arithmetic operations). To evaluate $\phi(x)$, we find the largest t with $tx \in K$ by binary search with error ε/n^2; this takes $O(\log n + \log(1/\varepsilon))$ membership tests. Thus the algorithm takes $O(t(\log n + \log(1/\varepsilon)))$ membership tests and $O(tn(\log n + \log(1/\varepsilon)))$ arithmetic operations. Ordinarily, a membership test takes at least n arithmetic operations (it has to look at the coordinates), so the running time will be dominated by the membership tests. ∎

If we combine the preliminary part and the main part to a single algorithm, we face an additional difficulty: we have assumed that membership test can be carried out not only for K but also in its affine image $A(K)$ produced by the preliminary part in a single step. This is not justified if we only have an oracle for the original K, since then we have

to apply each time the inverse of the affine transformation to get the point to be queried from the oracle. If the membership test takes n^2 or more operations, this is majorized by that. However, for most sensible encodings of K (say, by linear or algebraic inequalities), we can compute a new encoding for the affine image and work with this.

We may also leave K invariant and generate the random steps from an appropriate ellipsoid instead of the unit ball. This would solve the problem with the membership test but would increase the cost of generating the random step to $O(n^2)$, and hence would lead to essentially the same analysis.

References

N. Alon (1986): Eigenvalues and expanders, *Combinatorica* **6**, pp. 83–96.

N. Alon and V. D. Milman (1984): Eigenvalues, expanders, and superconcentrators, *Proc. 25th Ann. Symp. on Found. of Comp. Sci.*, pp. 320–322.

N. Alon and V. D. Milman (1985): λ_1, isoperimetric inequalities for graphs and superconcentrators, *J. Combinatorial Theory B* **38** pp. 73–88.

D. Applegate and R. Kannan (1990): Sampling and integration of near log-concave functions (preprint).

I. Bárány and Z. Füredi (1986): Computing the volume is difficult,

H. C. P. Berbie, C. G. E. Boender, A. H. G. R. Kan, C. L. Scheffer, R. L. Smith, and J. Telgen (1987) Hit-and-run algorithms for the identification of nonredundant linear inequalities, *Mat. Programming* **37**, 184–207.

B. Bollobás: Random Graphs, Academic Press, (1985)

T. Bonnesen and W. Fenchel (1934): *Theorie der konvexen Körper*, Springer, Berlin.

G. Brightwell and P. Winkler (1990): Counting linear extensions is #P-complete (Bellcore preprint).

A. Broder (1986): How hard is it to marry at random? (On the approximation of the permanent), *Proc. 18th Annual ACM Symposium on Theory of Computing*, 50–58.

H. Chernoff: A measure of asymptotic efficiency for tests of a Hypothesis based on the sum of observations, Annals of Math. Stat. 23 (1952) 493-509.

L. Danzer, D. Langwitz, H. Lenz (1957), Über das Löwnersche Ellipsoid und sein Analogon unter den einem Eikörper einbeschriebenen Ellipsoiden. *Arch. Math.* **8**, 214–219.

A. Dinghas (1957), Über eine Klasse superadditiver Mengenfunktionale von Brunn–Minkowski–Lusternik-schem Typus, *Math. Zeitschr.* **68**, 111–125.

J. Dodziuk and W. S. Kendall (1986): Combinatorial Laplacians and isoperimetric inequality, in: *From Local Times to Global Geometry, Control and Physics*, (ed. K. D. Ellworthy), Pitman Res. Notes in Math. Series **150**, 68–74.

M. Dyer and A. Frieze (1988): On the complexity of computing the volume of a polytope. *SIAM J. Comp.* **17**, 967–974.

M. Dyer and A. Frieze (1991): Computing the volume of convex bodies: a case where randomness provably helps (preprint).

M. Dyer, A. Frieze and R. Kannan (1989): A Random Polynomial Time Algorithm for Approximating the Volume of Convex Bodies, *Proc. of the 21st Annual ACM Symposium on Theory of Computing*, 375-381.

G. Elekes (1986): A geometric inequality and the complexity of computing volume, *Discrete and Computational Geometry* **1**, 289-292.

W. Feller (1968): *Introduction to the theory of probability and its applications*, I–II, Wiley, New York.

M. R. Garey and D. S. Johnson (1979): *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, San Francisco.

C.C. McGeoch and A. J. Hoogs (1990): Extensions of the hit-and-run algorithm for random point generation (extended abstract)

M. Grötschel, L. Lovász and A. Schrijver (1988): *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag.

P. R. Halmos (1974): *Measure Theory,* Springer-Verlag.

M. R. Jerrum, L. G. Valiant and V. V. Vazirani (1986): Random generation of combinatorial structures from a uniform distribution, *Theoretical Computer Science* **43**, 169–188.

L. G. Khachiyan (1988): On the complexity of computing the volume of a polytope. *Izvestia Akad. Nauk SSSR, Engineering Cybernetics* **3**, 216–217.

L. G. Khachiyan (1989): The problem of computing the volume of polytopes is NP-hard. *Uspekhi Mat. Nauk 44*, 199-200.

L. G. Khachiyan and A. Karzanov (1990): On the conductance of order Markov chains. Rutgers Univ. Tech. Report DCS TR 268.

L. G. Khachiyan and M. J. Todd (1990): On the complexity of estimating the maximal inscribed ellipsoid for a polytope. Cornell Univ. SORIE Tech. Report 893.

J. Lawrence (1989): Polytope volume computation, Preprint NISTIR 89-4123, U.S.Dept. of Commerce, Natl. Inst. of Stand. Tech., Galthersburg.

H. W. Lenstra, Jr. (1983), Integer programming with a fixed number of variables, *Oper. Res.* **8**, 538-548.

L. Lovász (1990): Geometric algorithms and algorithmic geometry, Proc. ICM'90 (Kyoto) (to appear)

L. Lovász (1991): How to compute the volume? (to appear)

L. Lovász and M. Simonovits (1990): Mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. *Proc. 31st Annual Symp. on Found. of Computer Science,* IEEE Computer Soc., 346–355.

L. Lovász and M. Simonovits (1991): (in preparation)

N. Metropolis et al: Equation of state calculation by fast computing machines, *J. Chem. Physics* **21** (1953) 1087-1092.

A. Nesterov and A. S. Nemirovskii (1989), *Self-concordant Functions and Polynomial-Time Methods in Convex Programming* (Russian), Centr. Econ. and Math. Inst., Acad. Nauk SSSR, Chapter 7.

A. Prékopa (1971), Logarithmic concave measures with applications to stochastic programming, *Acta Sci. Math. Szeged.* **32** 301–316.

A. Prékopa (1973), On logarithmic concave measures and functions, *Acta Sci. Math. Szeged.* **34**, 335-343.

R. L. Smith (1984), Efficient Monte-Carlo procedures for generating points uniformly distributed over bounded regions, *Operations Res.* **32**, 1296–1308.

A. Sinclair and M. Jerrum (1988): Conductance and the rapid mixing property for Markov chains: the approximation of the permanent resolved, *Proc. 20th ACM STOC*, pp. 235–244.

B. D. Yudin and A. S. Nemirovskii (1976): Informational complexity and efficient methods for the solution of convex extremal problems. *Ekonomika i Mat. Metody* **12**, 357–369 (Russian) [English transl.: *Matekon* **13**, 25–45].