Geometric Random Walks: A Survey

SANTOSH VEMPALA

ABSTRACT. The developing theory of geometric random walks is outlined here. Three aspects — general methods for estimating convergence (the "mixing" rate), isoperimetric inequalities in $\mathbb{R}^n$ and their intimate connection to random walks, and algorithms for fundamental problems (volume computation and convex optimization) that are based on sampling by random walks — are discussed.

1. Introduction

A geometric random walk starts at some point in $\mathbb{R}^n$ and at each step, moves to a "neighboring" point chosen according to some distribution that depends only on the current point, e.g., a uniform random point within a fixed distance δ. The sequence of points visited is a random walk. The distribution of the current point, in particular, its convergence to a steady state (or stationary) distribution, turns out to be a very interesting phenomenon. By choosing the one-step distribution appropriately, one can ensure that the steady state distribution is, for example, the uniform distribution over a convex body, or indeed any reasonable distribution in $\mathbb{R}^n$.

Geometric random walks are Markov chains, and the study of the existence and uniqueness of and the convergence to a steady state distribution is a classical field of mathematics. In the geometric setting, the dependence on the dimension (called n in this survey) is of particular interest. Pólya proved that with probability 1, a random walk on an n-dimensional grid returns to its starting point infinitely often for $n \leq 2$, but only a finite number of times for $n \geq 3$.

Random walks also provide a general approach to sampling a geometric distribution. To sample a given distribution, we set up a random walk whose steady state is the desired distribution. A random (or nearly random) sample is obtained by taking sufficiently many steps of the walk. Basic problems such as optimization and volume computation can be reduced to sampling. This con-

nection, pioneered by the randomized polynomial-time algorithm of Dyer, Frieze and Kannan [1991] for computing the volume of a convex body, has lead to many new developments in recent decades.

In order for sampling by a random walk to be efficient, the distribution of the current point has to converge rapidly to its steady state. The first part of this survey (Section 3) deals with methods to analyze this convergence, and describes the most widely used method, namely, bounding the *conductance*, in detail. The next part of the survey is about applying this to geometric random walks and the issues that arise therein. Notably, there is an intimate connection with geometric isoperimetric inequalities. The classical isoperimetric inequality says that among all measurable sets of fixed volume, a ball of this volume is the one that minimizes the surface area. Here, one is considering all measurable sets. In contrast, we will encounter the following type of question: Given a convex set K, and a number t such that $0 < t < 1$, what subset S of volume $t \cdot \text{vol}(K)$ has the smallest surface area inside K (i.e., not counting the boundary of S that is part of the boundary of K)? The inequalities that arise are interesting in their own right.

The last two sections describe polynomial-time algorithms for minimizing a quasi-convex function over a convex body and for computing the volume of a convex body. The random walk approach can be seen as an alternative to the ellipsoid method. The application to volume computation is rather remarkable in the light of results that no deterministic polynomial-time algorithm can approximate the volume to within an exponential (in n) factor. In Section 9, we briefly discuss the history of the problem and describe the latest algorithm.

Several topics related to this survey have been addressed in detail in the literature. For a general introduction to discrete random walks, the reader is referred to [Lovász 1996] or [Aldous and Fill $\geq$ 2005]. There is a survey by Kannan [1994] on applications of Markov chains in polynomial-time algorithms. For an in-depth account of the volume problem that includes all but the most recent improvements, there is a survey by Simonovits [2003] and an earlier article by Bollobás [1997].

Three walks.   Before we introduce various concepts and tools, let us state precisely three different ways to walk randomly in a convex body K in $\mathbb{R}^n$. It might be useful to keep these examples in mind. Later, we will see generalizations of these walks.

The *Grid Walk* is restricted to a discrete subset of K, namely, all points in K whose coordinates are integer multiples of a fixed parameter δ. These points form a grid, and the neighbors of a grid point are the points reachable by changing any one coordinate by $\pm\delta$. Let $e_1, \ldots, e_n$ denote the coordinate vectors in $\mathbb{R}^n$; then the neighbors of a grid point x are $\{x \pm \delta e_i\}$. The grid walk tries to move to a random neighboring point.

> **Grid Walk (δ)**
>
> - Pick a grid point y uniformly at random from the neighbors of the current point x.
> - If y is in K, go to y; else stay at x.

The *Ball Walk* tries to step to a random point within distance δ of the current point. Its state space is the entire set K.

> **Ball Walk (δ)**
>
> - Pick a uniform random point y from the ball of radius δ centered at the current point x.
> - If y is in K, go to y; else stay at x.

Hit-and-run picks a random point along a random line through the current point. It does not need a "step-size" parameter. The state space is again all of K.

> **Hit-and-run**
>
> - Pick a uniform random line ℓ through the current point x.
> - Go to a uniform random point on the chord $\ell \cap K$.

To implement the first step of hit-and-run, we can generate n independent random numbers, $u_1, \ldots, u_n$ each from the standard Normal distribution, and then the direction of the vector $u = (u_1, \ldots, u_n)$ is uniformly distributed. For the second step, using the membership oracle for K, we find an interval $[a, b]$ that contains the chord through x parallel to u so that $|a - b|$ is at most twice (say) the length of the chord (this can be done by a binary search with a logarithmic overhead). Then pick random points in $[a, b]$ till we find one in K.

For the first step of the ball walk, in addition to a random direction u, we generate a number r in the range $[0, \delta]$ with density $f(x)$ proportional to x^{n-1} and then $z = ru/|u|$ is uniformly distributed in a ball of radius δ.

Do these random walks converge to a steady state distribution? If so, what is it? How quickly do they converge? How does the rate of convergence depend on the convex body K?

These are some of the questions that we will address in analyzing the walks.

2. Basic Definitions

Markov chains. A *Markov chain* is defined using a σ-algebra $(K, \mathtt{A})$, where K is the state space and $\mathtt{A}$ is a set of subsets of K that is closed under complements and countable unions. For each element u of K, we have a probability measure P_u on $(K, \mathtt{A})$, i.e., each set $A \in \mathtt{A}$ has a probability $P_u(A)$. Informally, P_u is the distribution obtained on taking one step from u. The triple $(K, \mathtt{A}, \{P_u : u \in K\})$

along with a starting distribution Q_0 defines a Markov chain, i.e., a sequence of elements of K, $w_0, w_1, \ldots$, where w_0 is chosen from Q_0 and each subsequent w_i is chosen from $P_{w_{i-1}}$. Thus, the choice of w_{i+1} depends only on w_i and is independent of $w_0, \ldots, w_{i-1}$.

A distribution Q on $(K, \mathtt{A})$ is called *stationary* if one step from it gives the same distribution, i.e., for any $A \in \mathtt{A}$,

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

A distribution Q is *atom-free* if there is no $x \in K$ with $Q(x) > 0$.

The *ergodic flow* of subset A with respect to the distribution Q is defined as

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

It is easy to verify that a distribution Q is stationary iff $\Phi(A) = \Phi(K \setminus A)$. The existence and uniqueness of the stationary distribution Q for general Markov chains is a rather technical issue that is not covered in this survey; see [Revuz 1975].[1] In all the chains we study in this survey, the stationary distribution will be given explicitly and can be easily verified. To avoid the issue of uniqueness of the stationary distribution, we only consider *lazy* Markov chains. In a lazy version of a given Markov chain, at each step, with probability $1/2$, we do nothing; with the rest we take a step according to the Markov chain. The next theorem is folklore and will also be implied by convergence theorems that we present later.

THEOREM 2.1. *If Q is stationary with respect to a lazy Markov chain then it is the unique stationary distribution for that Markov chain.*

For some additional properties of lazy Markov chains, see [Lovász and Simonovits 1993, Section 1]. We will hereforth assume that the distribution in the definition of Φ is the unique stationary distribution.

The *conductance* of a subset A is defined as

$$\phi(A) = \frac{\Phi(A)}{\min\{Q(A), Q(K \setminus A)\}}$$

and the conductance of the Markov chain is

$$\phi = \min_A \phi(A) = \min_{0 < Q(A) \leq \frac{1}{2}} \frac{\int_A P_u(K \setminus A) \, dQ(u)}{Q(A)}.$$

The *local* conductance of an element u is $\ell(u) = 1 - P_u(\{u\})$.

The following weaker notion of conductance will also be useful. For any $0 \leq s < \frac{1}{2}$, the *s-conductance* of a Markov chain is defined as

$$\phi_s = \min_{A : s < Q(A) \leq \frac{1}{2}} \frac{\Phi(A)}{Q(A) - s}.$$

---

[1] For Markov chains on discrete state spaces, the characterization is much simpler; see [Norris 1998], for example.

Comparing distributions. We will often have to compare two distributions P and Q (typically, the distribution of the current point and the stationary distribution). There are many reasonable ways to do this. Here are three that will come up.

(i) The *total variation distance* of P and Q is

$$\|P - Q\|_{tv} = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|.$$

(ii) The L_2 *distance* of P with respect to Q is

$$\|P/Q\| = \int_K \frac{dP(u)}{dQ(u)} \, dP(u) = \int_K \left(\frac{dP(u)}{dQ(u)}\right)^2 \, dQ(u).$$

(iii) P is said to be M-*warm* with respect to Q if

$$M = \sup_{A \in \mathcal{A}} \frac{P(A)}{Q(A)}.$$

If Q_0 is $O(1)$-warm with respect to the stationary distribution Q for some Markov chain, we say that Q_0 is a *warm* start for Q.

Convexity. Convexity plays a key role in the convergence of geometric random walks. The following notation and concepts will be used.

The unit ball in R^n is B_n and its volume is $\mathrm{vol}(B_n)$. For two subsets A, B of R^n, their Minkowski sum is

$$A + B = \{x + y : x \in A, \, y \in B\}.$$

The Brunn-Minkowski theorem says that if A, B and $A+B$ are measurable, then

$$\mathrm{vol}(A + B)^{1/n} \geq \mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}. \tag{2–1}$$

Recall that a subset S of R^n is convex if for any two points $x, y \in S$, the interval $[x, y] \subseteq S$. A function $f : \mathsf{R}^n \to \mathsf{R}_+$ is said to be *logconcave* if for any two points $x, y \in \mathsf{R}^n$ and any $\lambda \in [0, 1]$,

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

The product and the minimum of two logconcave functions are both logconcave; the sum is not in general. The following fundamental properties, proved by Dinghas [1957], Leindler [1972] and Prékopa [1973; 1971], are often useful.

THEOREM 2.2. *All marginals as well as the distribution function of a logconcave function are logconcave. The convolution of two logconcave functions is logconcave.*

Logconcave functions have many properties that are reminiscent of convexity. For a logconcave density f, we denote the induced measure by π_f and its centroid by $z_f = \mathsf{E}_f(X)$. The second moment of f refers to $\mathsf{E}_f(|X - z_f|^2)$. The next three lemmas are chosen for illustration from [Lovász and Vempala 2003c]. The first

one was proved earlier by Grünbaum [1960] for the special case of the uniform density over a convex body. We will later see a further refinement of this lemma that is useful for optimization.

LEMMA 2.3. *Let $f : \mathrm{R}^n \to \mathrm{R}_+$ be a logconcave density function, and let H be any halfspace containing its centroid. Then*

$$\int_H f(x)\, dx \geq \frac{1}{e}.$$

LEMMA 2.4. *If X is drawn from a logconcave distribution in R^n, then for any integer $k > 0$,*

$$\mathsf{E}(|X|^k)^{1/k} \leq 2k\mathsf{E}(|X|).$$

Note that this can be viewed as a converse to Hölder's inequality which says that

$$\mathsf{E}(|X|^k)^{1/k} \geq \mathsf{E}(|X|).$$

LEMMA 2.5. *Let $X \in \mathrm{R}^n$ be a random point from a logconcave distribution with second moment R^2. Then $\mathsf{P}(|X| > tR) < e^{-t+1}$.*

A density function $f : \mathrm{R}^n \to \mathrm{R}_+$ is said to be *isotropic*, if its centroid is the origin, and its covariance matrix is the identity matrix. This latter condition can be expressed in terms of the coordinate functions as

$$\int_{\mathrm{R}^n} x_i x_j f(x)\, dx = \delta_{ij}$$

for all $1 \leq i, j \leq n$. This condition is equivalent to saying that for every vector $v \in \mathrm{R}^n$,

$$\int_{\mathrm{R}^n} (v^T x)^2 f(x)\, dx = |v|^2.$$

In terms of the associated random variable X, this means that

$$\mathsf{E}(X) = 0 \quad \text{and} \quad \mathsf{E}(XX^T) = I.$$

We say that f is *near-isotropic up to a factor of C* or *C-isotropic*, if

$$\frac{1}{C} \leq \int_{\mathrm{R}^n} (v^T x)^2\, d\pi_f(x) \leq C$$

for every unit vector v. The notions of "isotropic" and "near-isotropic" extend to non-negative integrable functions f, in which case we mean that the density function $f / \int_{\mathrm{R}^n} f$ is isotropic. For any full-dimensional integrable function f with bounded second moment, there is an affine transformation of the space bringing it to isotropic position, and this transformation is unique up to an orthogonal transformation of the space. Indeed if f is not isotropic, we can make the centroid be the origin by a translation. Next, compute $A = \mathsf{E}(XX^T)$ for the associated random variable X. Now A must be positive semi-definite (since each XX^T is) and so we can write $A = BB^T$ for some matrix B. Then the transformation B^{-1} makes f isotropic.

It follows that for an isotropic distribution in $\mathbb{R}^n$, the second moment is

$$\mathsf{E}(|X|^2) = \sum_i \mathsf{E}(X_i^2) = n.$$

Further, Lemma 2.5 implies that for an isotropic logconcave distribution f,

$$\mathsf{P}(X > t\sqrt{n}) < e^{-t},$$

which means that most of f is contained in a ball of radius $O(\sqrt{n})$, and this is sometimes called its effective diameter.

Computational model. If the input to an algorithm is a convex body K in $\mathbb{R}^n$, we assume that it is given by a *membership oracle* which on input $x \in \mathbb{R}^n$ returns Yes if $x \in K$ and No otherwise. In addition we will have some bounds on K—typically, $B_n \subseteq K \subseteq RB_n$, i.e., K contains a unit ball around the origin and is contained in a ball of given radius. It is enough to have any point x in K and the guarantee that a ball of radius r around x is contained in K and one of radius R contains K (by translation and scaling this is equivalent to the previous condition). Sometimes, we will need a separation oracle for K, i.e., a procedure which either verifies that a given point x is in K or returns a hyperplane that separates x from K. The complexity of the algorithm will be measured mainly by the number of oracle queries, but we will also keep track of the number of arithmetic operations.

In the case of a logconcave density f, we have an oracle for f, i.e., for any point x it returns $Cf(x)$ where C is an unknown parameter independent of x. This is useful when we know a function proportional to the desired density, but not its integral, e.g., in the case of the uniform density over a bounded set, all we need is the indicator function of the support. In addition, we have a guarantee that the centroid of f satisfies $|z_f|^2 < Z$ and the eigenvalues of the covariance matrix of f are bounded from above and below by two given numbers. Again, the complexity is measured by the number of oracle calls. We will say that an algorithm is *efficient* if its complexity is polynomial in the relevant parameters.

To emphasize the essential dependence on the dimension we will sometimes use the $O^*(.)$ notation which suppresses logarithmic factors and also the dependence on error parameters. For example, $n \log n / \varepsilon = O^*(n)$.

Examples. For the ball walk in a convex body, the state space K is the convex body, and $\mathcal{A}$ is the set of all measurable subsets of K. Further,

$$P_u(\{u\}) = 1 - \frac{\text{vol}\,(K \cap (u + \delta B_n))}{\text{vol}(\delta B_n)}$$

and for any measurable subset A, such that $u \notin A$,

$$P_u(A) = \frac{\text{vol}\,(A \cap (u + \delta B_n))}{\text{vol}(\delta B_n)}.$$

If $u \in A$, then

$$P_u(A) = P_u(A \setminus \{u\}) + P_u(\{u\}).$$

It is straightforward to verify that the uniform distribution is stationary, i.e.,

$$Q(A) = \frac{\mathrm{vol}(A)}{\mathrm{vol}(K)}.$$

For hit-and-run, the one-step distribution for a step from $u \in K$ is given as follows. For any measurable subset A of K,

$$P_u(A) = \frac{2}{\mathrm{vol}_{n-1}(\partial B_n)} \int_A \frac{dx}{\ell(u,x)|x-u|^{n-1}} \qquad (2\text{--}2)$$

where $\ell(u,x)$ is the length of the chord in K through u and x. The uniform distribution is once again stationary. One way to see this is to note that the one-step distribution has a density function and the density of stepping from u to v is the same as that for stepping from v to u.

These walks can be modified to sample much more general distributions. Let $f : \mathrm{R}^n \to \mathrm{R}_+$ be a nonnegative integrable function. It defines a measure π_f (on measurable subsets of R^n):

$$\pi_f(A) = \frac{\int_A f(x)\,dx}{\int_{\mathrm{R}^n} f(x)\,dx}.$$

The following extension of the ball walk, usually called the *ball walk with a Metropolis filter* has π_f as its stationary distribution (it is a simple exercise to prove, but quite nice that this works).

Ball walk with Metropolis filter (δ, f)

- Pick a uniformly distributed random point y in the ball of radius δ centered at the current point x.
- Move to y with probability $\min\left\{1, \frac{f(y)}{f(x)}\right\}$; stay at x with the remaining probability.

Hit-and-run can also be extended to sampling from such a general distribution π_f. For any line ℓ in R^n, let $\pi_{\ell,f}$ be the *restriction* of π to ℓ, i.e.,

$$\pi_{\ell,f}(S) = \frac{\int_{p+tu \in S} f(p+tu)\,dt}{\int_\ell f(x)\,dx},$$

where p is any point on ℓ and u is a unit vector parallel to ℓ.

Hit-and-run (f)

- Pick a uniform random line ℓ through the current point x.
- Go to a random point y along ℓ chosen from the distribution $\pi_{\ell,f}$.

Once again, it is easy to verify that π_f is the stationary distribution for this walk. One way to carry out the second step is to use a binary search to find the point p on ℓ where the function is maximal, and the points a and b on both sides of p on ℓ where the value of the function is $\varepsilon f(p)$. We allow a relative error of ε, so the number of oracle calls is only $O(\log(1/\varepsilon))$. Then select a uniformly distributed random point y on the segment $[a, b]$, and independently a uniformly distributed random real number in the interval $[0, 1]$. Accept y if $f(y) > rf(p)$; else, reject y and repeat.

3. Convergence and Conductance

So far we have seen that random walks can be designed to approach any reasonable distribution in R^n. For this to lead to an efficient sampling method, the convergence to the stationary distribution must be fast. This section is devoted to general methods for bounding the rate of convergence.

One way to define the *mixing rate* of a random walk is the number of steps required to reduce some measure of the distance of the current distribution to the stationary distribution by a factor of 2 (e.g., one of the distance measures from page 577). We will typically use the total variation distance. For a discrete random walk (i.e., the state space is a finite set), the mixing rate is characterized by the *eigenvalues gap* of the transition matrix P whose ijth entry is the probability of stepping from i to j, conditioned on currently being at i. Let $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$ be the eigenvalues of P. The top eigenvalue is 1 (by the definition of stationarity) and let $\lambda = \max\{\lambda_2, |\lambda_n|\}$ (in the lazy version of any walk, all the λ_i are nonnegative and $\lambda = \lambda_2$). Then, for a random walk starting at the point x, with Q_t being the distribution after t steps, the following bound on the convergence can be derived (see [Lovász 1996], for example). For any point $y \in K$,

$$|Q_t(y) - Q(y)| \leq \sqrt{\frac{Q(y)}{Q(x)}} \lambda^t. \tag{3-1}$$

Estimating λ can be difficult or impractical even in the discrete setting (if, for example, the state space is too large to write down P explicitly).

Intuitively, a random walk will "mix" slowly if it has a bottleneck, i.e., a partition $S, K \setminus S$ of its state space, such that the probability of stepping from S to $K \setminus S$ (the ergodic flow out of S) is small compared to the measures of S and $K \setminus S$. Note that this ratio is precisely the conductance of S, $\phi(S)$. It takes about $1/\phi(S)$ steps in expectation to even go from one side to the other. As we will see in this section, the mixing rate is bounded from above by $2/\phi^2$. Thus, conductance captures the mixing rate upto a quadratic factor. This was first proved for discrete Markov chains by Jerrum and Sinclair [1989] who showed that conductance can be related to the eigenvalue gap of the transition matrix. A similar relationship for a related quantity called *expansion* was found by Alon

[1986] and by Dodziuk and Kendall [1986]. The inequality below is a discrete analogue of Cheeger's inequality in differential geometry.

THEOREM 3.1.
$$\frac{\phi^2}{2} \leq 1 - \lambda \leq 2\phi.$$

As a consequence of this and (3–1), we get that for a discrete random walk starting at x, and any point $y \in K$,

$$|Q_t(y) - Q(y)| \leq \sqrt{\frac{Q(y)}{Q(x)}} \left(1 - \frac{\phi^2}{2}\right)^t. \qquad (3\text{–}2)$$

For the more general continuous setting, Lovász and Simonovits [1993] proved the connection between conductance and convergence. Their proof does not use eigenvalues. We will sketch it here since it is quite insightful, but does not seem to be well-known. It also applies to situations where the conductance can be bounded only for subsets of bounded size (i.e., the s-conductance, ϕ_s, can be bounded from below for some $s > 0$). We remind the reader that we have assumed that our Markov chains are lazy.

To show convergence, we need to prove that $|Q_t(A) - Q(A)|$ falls with t for every measurable subset A of K. However, this quantity might converge at different rates for different subsets. So we consider

$$\sup_{A:Q(A)=x} Q_t(A) - Q(A)$$

for each $x \in [0, 1]$. A bound for every x would imply what we want. To prove inductively that this quantity decreases with t, Lovász and Simonovits define the following formal upper bound. Let G_x be the set of functions defined as

$$\mathsf{G}_x = \left\{ g : K \to [0, 1] \ : \ \int_{u \in K} g(u) \, dQ(u) = x \right\}.$$

Using this, define

$$h_t(x) = \sup_{g \in \mathsf{G}_x} \int_{u \in K} g(u) \, (dQ_t(u) - dQ(u)) = \sup_{g \in \mathsf{G}_x} \int_{u \in K} g(u) \, dQ_t(u) - x,$$

It is clear that for A with $Q(A) = x$, $h_t(x) \geq Q_t(A) - Q(A)$ since the indicator function of A is in G_x. The function $h_t(x)$ has the following properties.

LEMMA 3.2. *Let t be a positive integer.*

a. *The function h_t is concave.*
b. *If Q is atom-free, then $h_t(x) = \sup_{A:Q(A)=x} Q_t(A) - Q(A)$ and the supremum is achieved by some subset.*
c. *Let Q be atom-free and $t \geq 1$. For any $0 \leq x \leq 1$, let $y = \min\{x, 1 - x\}$. Then,*
$$h_t(x) \leq \tfrac{1}{2} h_{t-1}(x - 2\phi y) + \tfrac{1}{2} h_{t-1}(x + 2\phi y).$$

The first part of the lemma is easily verified. We sketch the second part: to maximize h_t, we should use a function g that puts high weight on points u with $dQ_t(u)/dQ(u)$ as high as possible. Let A be a subset with $Q(A) = x$, so that for any point y not in A, the value of $dQ_t(y)/dQ(y)$ is no more than the value for any point in A (i.e., A consists of the top x fraction of points according to $dQ_t(u)/dQ(u)$). Let g be the corresponding indicator function. These points give the maximum payoff per unit of weight, so it is optimal to put as much weight on them as possible. We mention in passing that the case when Q has atoms is a bit more complicated, namely we might need to include one atom fractionally (so that $Q(A) = x$). In the general case, $h_t(x)$ can be achieved by a function g that is $0 - 1$ valued everywhere except for at most one point.

The third part of the lemma, which is the key to convergence, is proved below.

PROOF OF LEMMA 3.2C. Assume that $0 \le x \le \frac{1}{2}$. The other range is proved in a similar way. We will construct two functions, g_1 and g_2, and use these to bound $h_t(x)$. Let A be a subset to be chosen later with $Q(A) = x$. Let

$$g_1(u) = \begin{cases} 2P_u(A) - 1 & \text{if } u \in A, \\ 0 & \text{if } u \notin A, \end{cases} \quad \text{and} \quad g_2(u) = \begin{cases} 1 & \text{if } u \in A, \\ 2P_u(A) & \text{if } u \notin A. \end{cases}$$

First, note that $\frac{1}{2}(g_1 + g_2)(u) = P_u(A)$ for all $u \in K$, which means that

$$\frac{1}{2}\int_{u \in K} g_1(u) \, dQ_{t-1}(u) + \frac{1}{2}\int_{u \in K} g_2(u) \, dQ_{t-1}(u) = \int_{u \in K} P_u(A) \, dQ_{t-1}(u)$$
$$= Q_t(A) \qquad\qquad (3\text{--}3)$$

By the laziness of the walk ($P_u(A) \ge \frac{1}{2}$ iff $u \in A$), the range of the functions g_1 and g_2 is between 0 and 1 and if we let

$$x_1 = \int_{u \in K} g_1(u) \, dQ(u) \quad \text{and} \quad x_2 = \int_{u \in K} g_2(u) \, dQ(u),$$

then $g_1 \in \mathsf{G}_{x_1}$ and $g_2 \in \mathsf{G}_{x_2}$. Further,

$$\frac{1}{2}(x_1 + x_2) = \frac{1}{2}\int_{u \in K} g_1(u) \, dQ(u) + \frac{1}{2}\int_{u \in K} g_2(u) \, dQ(u)$$
$$= \int_{u \in K} P_u(A) \, dQ(u) = Q(A) = x$$

since Q is stationary. Next, since Q is atom-free, there is a subset $A \subseteq K$ that achieves $h_t(x)$. Using this and (3–3),

$$h_t(x) = Q_t(A) - Q(A)$$
$$= \frac{1}{2}\int_{u \in K} g_1(u) \, dQ_{t-1}(u) + \frac{1}{2}\int_{u \in K} g_2(u) \, dQ_{t-1}(u) - Q(A)$$
$$= \frac{1}{2}\int_{u \in K} g_1(u) \, (dQ_{t-1}(u) - dQ(u)) + \frac{1}{2}\int_{u \in K} g_2(u) \, (dQ_{t-1}(u) - dQ(u))$$
$$\le \frac{1}{2}h_{t-1}(x_1) + \frac{1}{2}h_{t-1}(x_2).$$

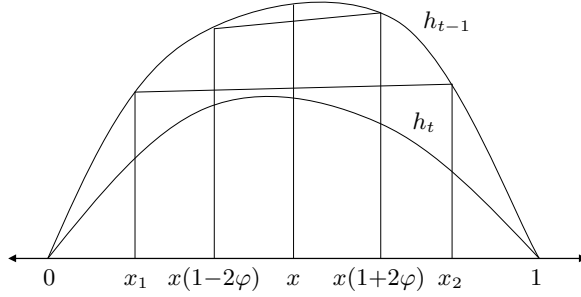**Figure 1.** Bounding h_t.

We already know that $x_1 + x_2 = 2x$. In fact, x_1 and x_2 are separated from x.

$$x_1 = \int_{u \in K} g_1(u) \, dQ(u)$$

$$= 2 \int_{u \in A} P_u(A) \, dQ(u) - \int_{u \in A} dQ(u)$$

$$= 2 \int_{u \in A} (1 - P_u(K \setminus A)) \, dQ(u) - x$$

$$= x - 2 \int_{u \in A} P_u(K \setminus A) \, dQ(u)$$

$$= x - 2\Phi(A)$$

$$\leq x - 2\phi x$$

$$= x(1 - 2\phi).$$

(In the penultimate step, we used the fact that $x \leq \frac{1}{2}$.) Thus we have,

$$x_1 \leq x(1 - 2\phi) \leq x \leq x(1 + 2\phi) \leq x_2.$$

Since h_{t-1} is concave, the chord from x_1 to x_2 on h_{t-1} lies below the chord from $x(1 - 2\phi)$ to $x(1 + 2\phi)$ (see Figure 1). Therefore,

$$h_t(x) \leq \tfrac{1}{2} h_{t-1}(x(1 - 2\phi)) + \tfrac{1}{2} h_{t-1}(x(1 + 2\phi))$$

which is the conclusion of the lemma.

In fact, a proof along the same lines implies the following generalization of part (c).

LEMMA 3.3. *Let Q be atom-free and $0 \leq s \leq 1$. For any $s \leq x \leq 1 - s$, let $y = \min\{x - s, 1 - x - s\}$. Then for any integer $t > 0$,*

$$h_t(x) \leq \frac{1}{2} h_{t-1}(x - 2\phi_s y) + \frac{1}{2} h_{t-1}(x + 2\phi_s y).$$

Given some information about Q_0, we can now bound the rate of convergence to the stationary distribution. We assume that Q is atom-free in the next theorem

and its corollary. These results can be extended to the case when Q has atoms with slightly weaker bounds [Lovász and Simonovits 1993].

THEOREM 3.4. *Let $0 \leq s \leq 1$ and C_0 and C_1 be such that*

$$h_0(x) \leq C_0 + C_1 \min\{\sqrt{x-s}, \sqrt{1-x-s}\}.$$

Then

$$h_t(x) \leq C_0 + C_1 \min\{\sqrt{x-s}, \sqrt{1-x-s}\} \left(1 - \frac{\phi_s^2}{2}\right)^t.$$

PROOF. By induction on t. The inequality is true for $t = 0$ by the hypothesis. Now, suppose it holds for all values less than t. Assume $s = 0$ (for convenience) and without loss of generality that $x \leq 1/2$. From Lemma 3.3, we know that

$$h_t(x) \leq \frac{1}{2}h_{t-1}(x(1-2\phi)) + \frac{1}{2}h_{t-1}(x(1+2\phi))$$

$$\leq C_0 + \frac{1}{2}C_1 \left(\sqrt{x(1-2\phi)} + \sqrt{x(1+2\phi)}\right) \left(1 - \frac{\phi^2}{2}\right)^{t-1}$$

$$= C_0 + \frac{1}{2}C_1\sqrt{x} \left(\sqrt{1-2\phi} + \sqrt{1+2\phi}\right) \left(1 - \frac{\phi^2}{2}\right)^{t-1}$$

$$\leq C_0 + C_1\sqrt{x} \left(1 - \frac{\phi^2}{2}\right)^t.$$

Here we have used $\sqrt{1-2\phi} + \sqrt{1+2\phi} \leq 2(1 - \frac{\phi^2}{2})$.

The following corollary, about convergence from various types of "good" starting distributions, gives concrete implications of the theorem.

COROLLARY 3.5. a. *Let $M = \sup_A Q_0(A)/Q(A)$. Then,*

$$\|Q_t - Q\|_{tv} \leq \sqrt{M} \left(1 - \frac{\phi^2}{2}\right)^t.$$

b. *Let $0 < s \leq \frac{1}{2}$ and $H_s = \sup\{|Q_0(A) - Q(A)| : Q(A) \leq s\}$. Then,*

$$\|Q_t - Q\|_{tv} \leq H_s + \frac{H_s}{s} \left(1 - \frac{\phi_s^2}{2}\right)^t.$$

c. *Let $M = \|Q_0/Q\|$. Then for any $\varepsilon > 0$,*

$$\|Q_t - Q\|_{tv} \leq \varepsilon + \sqrt{\frac{M}{\varepsilon}} \left(1 - \frac{\phi^2}{2}\right)^t.$$

PROOF. The first two parts are straightforward. For the third, observe that the L_2 norm,

$$\|Q_0/Q\| = \mathsf{E}_{Q_0} \left(\frac{dQ_0(x)}{dQ(x)}\right).$$

So, for $1-\varepsilon$ of Q_0, $dQ_0(x)/dQ(x) \leq M/\varepsilon$. We can view the starting distribution as being generated as follows: with probability $1 - \varepsilon$ it is a distribution with

$\sup Q_0(A)/Q(A) \leq M/\varepsilon$; with probability ε it is some other distribution. Now using part (a) implies part (c).

Conductance and s-conductance are not the only known ways to bound the rate of convergence. Lovász and Kannan [1999] have extended conductance to the notion of *blocking conductance* which is a certain type of average of the conductance over various subset sizes (see also [Kannan et al. 2004]). In some cases, it provides a sharper bound than conductance. Let $\phi(x)$ denote the minimum conductance over all subsets of measure x. Then one version of their main theorem is the following.

THEOREM 3.6. *Let π_0 be the measure of the starting point. Then, after*

$$t > C \ln \left(\frac{1}{\varepsilon} \right) \int_{\pi_0}^{\frac{1}{2}} \frac{dx}{x \phi(x)^2}$$

steps, where C is an absolute constant, we have $\|Q_t - Q\|_{tv} \leq \varepsilon$.

The theorem can be extended to continuous Markov chains. Another way to bound convergence which we do not describe here is via the *log-Sobolev* inequalities [Diaconis and Saloff-Coste 1996].

4. Isoperimetry

How to bound the conductance of a geometric random walk? To show that the conductance is large, we have to prove that for any subset $A \subset K$, the probability that a step goes out of A is large compared to $Q(A)$ and $Q(K \setminus A)$. To be concrete, consider the ball walk. For any particular subset S, the points that are likely to "cross over" to $K \setminus S$ are those that are "near" the boundary of S inside K. So, showing that $\phi(S)$ is large seems to be closely related to showing that there is a large volume of points near the boundary of S inside K. This section is devoted to inequalities which will have such implications and will play a crucial role in bounding the conductance.

To formulate an isoperimetric inequality for convex bodies, we consider a partition of a convex body K into three sets S_1, S_2, S_3 such that S_1 and S_2 are "far" from each other, and the inequality bounds the minimum possible volume of S_3 relative to the volumes of S_1 and S_2. We will consider different notions of distance between subsets. Perhaps the most basic is the Euclidean distance:

$$d(S_1, S_2) = \min\{|u - v| : u \in S_1, v \in S_2\}.$$

Suppose $d(S_1, S_2)$ is large. Does this imply that the volume of $S_3 = K \setminus (S_1 \cup S_2)$ is large? The classic counterexample to such a theorem is a dumbbell — two large subsets separated by very little. Of course, this is not a convex set!

The next theorem, proved in [Dyer and Frieze 1991] (improving on a theorem in [Lovász and Simonovits 1992] by a factor of 2; see also [Lovász and Simonovits 1993]) asserts that the answer is yes.

THEOREM 4.1. *Let S_1, S_2, S_3 be a partition into measurable sets of a convex body K of diameter D. Then,*

$$\text{vol}(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\text{vol}(S_1), \text{vol}(S_2)\}.$$

A limiting version of this inequality is the following: For any subset S of a convex body of diameter D,

$$\text{vol}_{n-1}(\partial S \cap K) \geq \frac{2}{D} \min\{\text{vol}(S), \text{vol}(K \setminus S)\}$$

which says that the surface area of S inside K is large compared to the volumes of S and $K \setminus S$. This is in direct analogy with the classical isoperimetric inequality, which says that the surface area to volume ratio of any measurable set is at least the ratio for a ball.

How does one prove such an inequality? In what generality does it hold? (i.e., for what measures besides the uniform measure on a convex set?) We will address these questions in this section. We first give an overview of known inequalities and then outline the proof technique.

Theorem 4.1 can be generalized to arbitrary logconcave measures. Its proof is very similar to that of 4.1 and we will presently give an outline.

THEOREM 4.2. *Let f be a logconcave function whose support has diameter D and let π_f be the induced measure. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\pi_f(S_1), \pi_f(S_2)\}.$$

In terms of diameter, this inequality is the best possible, as shown by a cylinder. A more refined inequality is obtained in [Kannan et al. 1995; Lovász and Vempala 2003c] using the average distance of a point to the center of gravity (in place of diameter). It is possible for a convex body to have much larger diameter than average distance to its centroid (e.g., a cone). In such cases, the next theorem provides a better bound.

THEOREM 4.3. *Let f be a logconcave density in $\mathbb{R}^n$ and π_f be the corresponding measure. Let z_f be the centroid of f and define $M(f) = \mathsf{E}_f(|x - z_f|)$. Then, for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{\ln 2}{M(f)} d(S_1, S_2) \pi_f(S_1) \pi_f(S_2).$$

For an isotropic density, $M(f)^2 \leq \mathsf{E}_f(|x - z_f|^2) = n$ and so $M(f) \leq \sqrt{n}$. The diameter could be unbounded (e.g., an isotropic Gaussian).

A further refinement, based on isotropic position, has been conjectured in [Kannan et al. 1995]. Let λ be the largest eigenvalue of the inertia matrix of f,

i.e.,

$$\lambda = \max_{v:|v|=1} \int_{\mathbb{R}^n} f(x)(v^T x)^2 \, dx. \tag{4-1}$$

Then the conjecture says that there is an absolute constant c such that

$$\pi_f(S_3) \geq \frac{c}{\sqrt{\lambda}} d(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

Euclidean distance and isoperimetric inequalities based on it are relevant for the analysis of "local" walks such as the ball walk. Hit-and-run, with its nonlocal moves, is connected with a different notion of distance.

The *cross-ratio distance* between two points u, v in a convex body K is computed as follows: Let p, q be the endpoints of the chord in K through u and v such that the points occur in the order p, u, v, q. Then

$$d_K(u, v) = \frac{|u - v||p - q|}{|p - u||v - q|} = (p : v : u : q).$$

where $(p : v : u : q)$ denotes the classical cross-ratio. We can now define the cross-ratio distance between two sets S_1, S_2 as

$$d_K(S_1, S_2) = \min\{d_K(u, v) : u \in S_1, v \in S_2\}.$$

The next theorem was proved in [Lovász 1999] for convex bodies and extended to logconcave densities in [Lovász and Vempala 2003d].

THEOREM 4.4. *Let f be a logconcave density in $\mathbb{R}^n$ whose support is a convex body K and let π_f be the induced measure. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq d_K(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

All the inequalities so far are based on defining the distance between S_1 and S_2 by the *minimum* over pairs of some notion of pairwise distance. It is reasonable to think that perhaps a much sharper inequality can be obtained by using some *average* distance between S_1 and S_2. Such an inequality was proved in [Lovász and Vempala 2004]. As we will see in Section 6, it leads to a radical improvement in the analysis of random walks.

THEOREM 4.5. *Let K be a convex body in $\mathbb{R}^n$. Let $f : K \to \mathbb{R}_+$ be a logconcave density with corresponding measure π_f and $h : K \to \mathbb{R}_+$, an arbitrary function. Let S_1, S_2, S_3 be any partition of K into measurable sets. Suppose that for any pair of points $u \in S_1$ and $v \in S_2$ and any point x on the chord of K through u and v,*

$$h(x) \leq \frac{1}{3}\min(1, d_K(u, v)).$$

Then

$$\pi_f(S_3) \geq \mathsf{E}_f(h(x)) \min\{\pi_f(S_1), \pi_f(S_2)\}.$$

The coefficient on the RHS has changed from a "minimum" to an "average". The weight $h(x)$ at a point x is restricted only by the minimum cross-ratio distance between pairs u, v from S_1, S_2 respectively, such that x lies on the line between them (previously it was the overall minimum). In general, it can be much higher than the minimum cross-ratio distance between S_1 and S_2.

The localization lemma. The proofs of these inequalities are based on an elegant idea: integral inequalities in R^n can be reduced to one-dimensional inequalities! Checking the latter can be tedious but is relatively easy. We illustrate the main idea by sketching the proof of Theorem 4.2.

For a proof of Theorem 4.2 by contradiction, let us assume the converse of its conclusion, i.e., for some partition S_1, S_2, S_3 of R^n and logconcave density f,

$$\int_{S_3} f(x)\, dx < C \int_{S_1} f(x)\, dx \quad \text{and} \quad \int_{S_3} f(x)\, dx < C \int_{S_2} f(x)\, dx$$

where $C = 2d(S_1, S_2)/D$. This can be reformulated as

$$\int_{\mathrm{R}^n} g(x)\, dx > 0 \quad \text{and} \quad \int_{\mathrm{R}^n} h(x)\, dx > 0 \tag{4–2}$$

where

$$g(x) = \begin{cases} Cf(x) & \text{if } x \in S_1, \\ 0 & \text{if } x \in S_2, \\ -f(x) & \text{if } x \in S_3. \end{cases} \quad \text{and} \quad h(x) = \begin{cases} 0 & \text{if } x \in S_1, \\ Cf(x) & \text{if } x \in S_2, \\ -f(x) & \text{if } x \in S_3. \end{cases}$$

These inequalities are for functions in R^n. The main tool to deal with them is the *localization lemma* [Lovász and Simonovits 1993] (see also [Kannan et al. 1995] for extensions and applications).

LEMMA 4.6. *Let $g, h : \mathrm{R}^n \to \mathrm{R}$ be lower semi-continuous integrable functions such that*

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

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

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

The points a, b represent an interval A and one may think of $l(t)^{n-1} dA$ as the cross-sectional area of an infinitesimal cone with base area dA. The lemma says that over this cone truncated at a and b, the integrals of g and h are positive. Also, without loss of generality, we can assume that a, b are in the union of the supports of g and h.

The main idea behind the lemma is the following. Let H be any halfspace such that

$$\int_H g(x)\, dx = \frac{1}{2} \int_{\mathrm{R}^n} g(x)\, dx.$$

Let us call this a bisecting halfspace. Now either

$$\int_H h(x)\, dx > 0 \quad \text{or} \quad \int_{R^n \setminus H} h(x)\, dx > 0.$$

Thus, either H or its complementary halfspace will have positive integrals for both g and h. Thus we have reduced the domains of the integrals from R^n to a halfspace. If we could repeat this, we might hope to reduce the dimensionality of the domain. But do there even exist bisecting halfspaces? In fact, they are aplenty: for any $(n-2)$-dimensional affine subspace, there is a bisecting halfspace with A contained in its bounding hyperplane. To see this, let H be halfspace containing A in its boundary. Rotating H about A we get a family of halfspaces with the same property. This family includes H', the complementary halfspace of H. Now the function $\int_H g - \int_{R^n \setminus H} g$ switches sign from H to H'. Since this is a continuous family, there must be a halfspace for which the function is zero, which is exactly what we want (this is sometimes called the "ham sandwich" theorem).

If we now take all $(n-2)$-dimensional affine subspaces given by some $x_i = r_1$, $x_j = r_2$ where r_1, r_2 are rational, then the intersection of all the corresponding bisecting halfspaces is a line (by choosing only rational values for x_i, we are considering a countable intersection). As long as we are left with a two or higher dimensional set, there is a point in its interior with at least two coordinates that are rational, say $x_1 = r_1$ and $x_2 = r_2$. But then there is a bisecting halfspace H that contains the affine subspace given by $x_1 = r_1, x_2 = r_2$ in its boundary, and so it properly partitions the current set. With some additional work, this leads to the existence of a concave function on an interval (in place of the linear function ℓ in the theorem) with positive integrals. Simplifying further from concave to linear takes quite a bit of work.

Going back to the proof sketch of Theorem 4.2, we can apply the lemma to get an interval $[a, b]$ and a linear function ℓ such that

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

$$(4\text{--}3)$$

(The functions g, h as we have defined them are not lower semi-continuous. However, this can be easily achieved by expanding S_1 and S_2 slightly so as to make them open sets, and making the support of f an open set. Since we are proving strict inequalities, we do not lose anything by these modifications).

Let us partition $[0, 1]$ into Z_1, Z_2, Z_3.

$$Z_i = \{t \in [0, 1] : (1-t)a + tb \in S_i\}.$$

Note that for any pair of points $u \in Z_1, v \in Z_2$, $|u - v| \geq d(S_1, S_2)/D$. We can rewrite (4–3) as

$$\int_{Z_3} \ell(t)^{n-1} f((1-t)a + tb)\, dt < C \int_{Z_1} \ell(t)^{n-1} f((1-t)a + tb)\, dt$$

and

$$\int_{Z_3} \ell(t)^{n-1} f((1-t)a + tb)\, dt < C \int_{Z_2} \ell(t)^{n-1} f((1-t)a + tb)\, dt.$$

The functions f and $\ell(.)^{n-1}$ are both logconcave, so $F(t) = \ell(t)^{n-1} f((1-t)a + tb)$ is also logconcave. We get,

$$\int_{Z_3} F(t)\, dt < C \min \left\{ \int_{Z_1} F(t)\, dt, \int_{Z_2} F(t)\, dt \right\}. \tag{4–4}$$

Now consider what Theorem 4.2 asserts for the function $F(t)$ over the interval $[0, 1]$ and the partition Z_1, Z_2, Z_3:

$$\int_{Z_3} F(t)\, dt \geq 2 d(Z_1, Z_2) \min \left\{ \int_{Z_1} F(t)\, dt, \int_{Z_2} F(t)\, dt \right\}. \tag{4–5}$$

We have substituted 1 for the diameter of the interval $[0, 1]$. Also, $d(Z_1, Z_2) \geq d(S_1, S_2)/D = C/2$. Thus, Theorem 4.2 applied to the function $F(t)$ contradicts (4–4) and to prove the theorem in general, it suffices to prove it in the one-dimensional case.

In fact, it will be enough to prove (4–5) for the case when each Z_i is a single interval. Suppose we can do this. Then, for each maximal interval (c, d) contained in Z_3, the integral of F over Z_3 is at least C times the smaller of the integrals to its left $[0, c]$ and to its right $[d, 1]$ and so one of these intervals is "accounted" for. If all of Z_1 or all of Z_2 is accounted for, then we are done. If not, there is an unaccounted subset U that intersects both Z_1 and Z_2. But then, since Z_1 and Z_2 are separated by at least $d(S_1, S_2)/D$, there is an interval of Z_3 of length at least $d(S_1, S_2)/D$ between $U \cap Z_1$ and $U \cap Z_2$ which can account for more.

We are left with proving (4–5) when each Z_i is an interval. Without the factor of two, this is trivial by the logconcavity of F. To get C as claimed, one can reduce this to the case when $F(t) = e^{ct}$ and verify it for this function [Lovász and Simonovits 1993]. The main step is to show that there is a choice of c so that when the current $F(t)$ is replaced by e^{ct}, the LHS of (4–5) does not increase and the RHS does not decrease.

5. Mixing of the Ball Walk

With the isoperimetric inequalities at hand, we are now ready to prove bounds on the conductance and hence on the mixing time. In this section, we focus on the ball walk in a convex body K. Assume that K contains the unit ball.

A geometric random walk is said to be *rapidly mixing* if its conductance is bounded from below by an inverse polynomial in the dimension. By Corollary 3.5, this implies that the number of steps to halve the variation distance to stationary is a polynomial in the dimension. The conductance of the ball walk in a convex body K can be exponentially small. Consider, for example, starting at point x near the apex of a rotational cone in $\mathbb{R}^n$. Most points in a ball of radius δ around x will lie outside the cone (if x is sufficiently close to the apex) and so the local conductance is arbitrarily small. So, strictly speaking, the ball walk is not rapidly mixing.

There are two ways to get around this. For the purpose of sampling uniformly from K, one can expand K a little bit by considering $K' = K + \alpha B_n$, i.e., adding a ball of radius α around every point in K. Then for $\alpha > 2\delta\sqrt{n}$, it is not hard to see that $\ell(u)$ is at least $1/8$ for every point $u \in K'$. We can now consider the ball walk in K'. This fix comes at a price. First, we need a membership oracle for K'. This can be constructed as follows: given a point $x \in \mathbb{R}^n$, we find a point $y \in K$ such that $|x - y|$ is minimum. This is a convex program and can be solved using the ellipsoid algorithm [Grötschel et al. 1988] and the membership oracle for K, Second, we need to ensure that $\mathrm{vol}(K')$ is comparable to $\mathrm{vol}(K)$. Since K contains a unit ball, $K' \subseteq (1 + \alpha)K$ and so with $\alpha < 1/n$, we get that $\mathrm{vol}(K') < e\,\mathrm{vol}(K)$. Thus, we would need $\delta < 1/2n\sqrt{n}$.

Does large local conductance imply that the conductance is also large? We will prove that the answer is yes. The next lemma about one-step distributions of nearby points will be useful.

LEMMA 5.1. *Let u, v be such that $|u - v| \leq \frac{t\delta}{\sqrt{n}}$ and $\ell(u), \ell(v) \geq \ell$. Then,*

$$\|P_u - P_v\|_{tv} \leq 1 + t - \ell.$$

Roughly speaking, the lemma says that if two points with high local conductance are close in Euclidean distance, then their one-step distributions have a large overlap. Its proof follows from a computation of the overlap volume of the balls of radius δ around u and v.

We can now state and prove a bound on the conductance of the ball walk.

THEOREM 5.2. *Let K be a convex body of diameter D so that for every point u in K, the local conductance of the ball walk with δ steps is at least ℓ. Then,*

$$\phi \geq \frac{\ell^2 \delta}{16\sqrt{n}D}.$$

The structure of most proofs of conductance is similar and we will illustrate it by proving this theorem.

PROOF. Let $K = S_1 \cup S_2$ be a partition into measurable sets. We will prove that

$$\int_{S_1} P_x(S_2)\,dx \geq \frac{\ell^2 \delta}{16\sqrt{n}D}\,\min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\} \qquad (5\text{--}1)$$

Note that since the uniform distribution is stationary,

$$\int_{S_1} P_x(S_2)\,dx = \int_{S_2} P_x(S_1)\,dx.$$

Consider the points that are "deep" inside these sets, i.e. unlikely to jump out of the set (see Figure 2):

$$S_1' = \left\{ x \in S_1 : P_x(S_2) < \frac{\ell}{4} \right\}$$

and

$$S_2' = \left\{ x \in S_2 : P_x(S_1) < \frac{\ell}{4} \right\}.$$

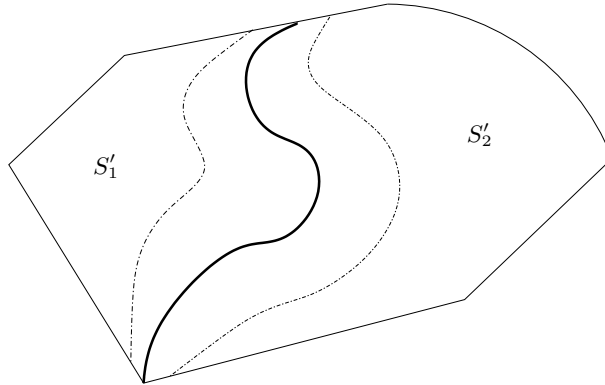Let S_3' be the rest i.e., $S_3' = K \setminus S_1' \setminus S_2'$.

Figure 2. The conductance proof. The dark line is the boundary between S_1 and S_2.

Suppose $\text{vol}(S_1') < \text{vol}(S_1)/2$. Then

$$\int_{S_1} P_x(S_2)\,dx \geq \frac{\ell}{4}\text{vol}(S_1 \setminus S_1') \geq \frac{\ell}{8}\text{vol}(S_1)$$

which proves (5–1).

So we can assume that $\text{vol}(S_1') \geq \text{vol}(S_1)/2$ and similarly $\text{vol}(S_2') \geq \text{vol}(S_2)/2$. Now, for any $u \in S_1'$ and $v \in S_2'$,

$$\|P_u - P_v\|_{tv} \geq 1 - P_u(S_2) - P_v(S_1) > 1 - \frac{\ell}{2}.$$

Applying Lemma 5.1 with $t = \ell/2$, we get that

$$|u - v| \geq \frac{\ell\delta}{2\sqrt{n}}.$$

Thus $d(S_1, S_2) \geq \ell\delta/2\sqrt{n}$. Applying Theorem 4.1 to the partition S_1', S_2', S_3', we have

$$\text{vol}(S_3') \geq \frac{\ell\delta}{\sqrt{n}D} \min\{\text{vol}(S_1'), \text{vol}(S_2')\}$$

$$\geq \frac{\ell\delta}{2\sqrt{n}D} \min\{\text{vol}(S_1), \text{vol}(S_2)\}$$

We can now prove (5–1) as follows:

$$\int_{S_1} P_x(S_2)\, dx = \frac{1}{2}\int_{S_1} P_x(S_2)\, dx + \frac{1}{2}\int_{S_2} P_x(S_1)\, dx \geq \tfrac{1}{2}\text{vol}(S_3')\frac{\ell}{4}$$

$$\geq \frac{\ell^2\delta}{16\sqrt{n}D} \min\{\text{vol}(S_1), \text{vol}(S_2)\}.$$

As observed earlier, by going to $K' = K + (1/n)B_n$ and using $\delta = 1/2n\sqrt{n}$, we have $\ell \geq 1/8$. Thus, for the ball walk in K', $\phi = \Omega(1/n^2 D)$. Using Corollary 3.5, the mixing rate is $O(n^4 D^2)$.

We mentioned earlier that there are two ways to get around the fact that the ball walk can have very small local conductance. The second, which we describe next, is perhaps a bit cleaner and also achieves a better bound on the mixing rate. It is based on the observation that only a small fraction of points can have small local conductance. Define the points of high local conductance as

$$K_\delta = \left\{u \in K : \ell(u) \geq \frac{3}{4}\right\}$$

LEMMA 5.3. *Suppose that K is a convex body containing a unit ball in R^n.*

a. *K_δ is a convex set.*
b. *$\text{vol}(K_\delta) \geq (1 - 2\delta\sqrt{n})\text{vol}(K)$.*

The first part follows from the Brunn-Minkowski inequality (2–1). The second is proved by estimating the average local conductance [Kannan et al. 1997] and has the following implication: if we set $\delta \leq \varepsilon/2\sqrt{n}$, we get that at least $(1 - \varepsilon)$ fraction of points in K have large local conductance. Using this, we can prove the following theorem.

THEOREM 5.4. *For any $0 \leq s \leq 1$, we can choose the step-size δ for the ball walk in a convex body K of diameter D so that*

$$\phi_s \geq \frac{s}{200nD}.$$

PROOF. The proof is quite similar to that of Theorem 5.2. Let S_1, S_2 be a partition of K. First, since we are proving a bound on the s-conductance, we can assume that $\text{vol}(S_1), \text{vol}(S_2) \geq s\text{vol}(K)$. Next, we choose $\delta = s/4\sqrt{n}$ so that by Lemma 5.3,

$$\text{vol}(K_\delta) \geq (1 - \tfrac{1}{2}s)\text{vol}(K).$$

So only an $s/2$ fraction of K has small local conductance and we will be able to ignore it. Define

$$S_1' = \left\{ x \in S_1 \cap K_\delta : P_x(S_2) < \tfrac{3}{16} \right\},$$
$$S_2' = \left\{ x \in S_2 \cap K_\delta : P_x(S_1) < \tfrac{3}{16} \right\}.$$

As in the proof of Theorem 5.2, these points are "deep" in S_1 and S_2 respectively and they are also restricted to be in K_δ. Recall that the local conductance of every point in K_δ is at least $3/4$. We can assume that $\mathrm{vol}(S_1') \geq \mathrm{vol}(S_1)/3$. Otherwise,

$$\int_{S_1} P_x(S_2)\, dx \geq \left(\tfrac{2}{3} \mathrm{vol}(S_1) - \tfrac{1}{2} s \mathrm{vol}(K) \right) \tfrac{3}{16} \geq \tfrac{1}{32} \mathrm{vol}(S_1).$$

which implies the theorem. Similarly, we can assume that $\mathrm{vol}(S_2') \geq \mathrm{vol}(S_2)/3$.

For any $u \in S_1'$ and $v \in S_2'$,

$$\|P_u - P_v\|_{tv} \geq 1 - P_u(S_2) - P_v(S_1) > 1 - \tfrac{3}{8}.$$

Applying Lemma 5.1 with $t = 3/8$, we get that

$$|u - v| \geq \frac{3\delta}{8\sqrt{n}}.$$

Thus $d(S_1, S_2) \geq 3\delta/8\sqrt{n}$. Let $S_3' = K_\delta \setminus (S_1' \cup S_2')$ Applying Theorem 4.1 to the partition S_1', S_2', S_3' of K_δ, we have

$$\mathrm{vol}(S_3') \geq \frac{3\delta}{4\sqrt{n}D} \min\{\mathrm{vol}(S_1'), \mathrm{vol}(S_2')\}$$
$$\geq \frac{s}{16nD} \min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}$$

The theorem follows:

$$\int_{S_1} P_x(S_2)\, dx \geq \tfrac{1}{2} \mathrm{vol}(S_3') \tfrac{3}{16} > \frac{s}{200nD} \min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}.$$

Using Corollary 3.5(b), this implies that from an M-warm start, the variation distance of Q_t and Q is smaller than ε after

$$t \geq C \frac{M^2}{\varepsilon^2} n^2 D^2 \ln\left(\frac{2M}{\varepsilon} \right) \tag{5-2}$$

steps, for some absolute constant C.

There is another way to use Lemma 5.3. In [Kannan et al. 1997], the following modification of the ball walk, called the *speedy* walk, is described. At a point x, the speedy walk picks a point uniformly from $K \cap x + \delta B_n$. Thus, the local conductance of every point is 1. However, there are two complications with this. First, the stationary distribution is not uniform, but proportional to $\ell(u)$. Second, each step seems unreasonable — we could make $\delta > D$ and then we would only need one step to get a random point in K. We can take care of the first problem with a rejection step at the end (and using Lemma 5.3). The root of the second problem is the question: how do we implement one step of

the speedy walk? The only general way is to get random points from the ball
around the current point till one of them is also in K. This process is the ball
walk and it requires $1/\ell(u)$ attempts in expectation at a point u. However, if we
count only the proper steps, i.e., ones that move away from the current point,
then it is possible to show that the mixing rate of the walk is in fact $O(n^2 D^2)$
from *any* starting point [Lovász and Kannan 1999]. Again, the proof is based
on an isoperimetric inequality which is slightly sharper than Theorem 4.2. For
this bound to be useful, we also need to bound the total number of improper or
"wasted" steps. If we start at a random point, then this is the number of proper
steps times $\mathsf{E}(1/\ell(u))$, which can be unbounded. But, if we allow a small overall
probability of failure, then with the remaining probability, the expected number
of wasted steps is bounded by $O(n^2 D^2)$ as well.

The bound of $O(n^2 D^2)$ on the mixing rate is the best possible in terms of the
diameter, as shown by a cylinder. However, if the convex body is isotropic, then
the isoperimetry conjecture (4–1) implies a mixing rate of $O(n^2)$.

For the rest of this section, we will discuss how these methods can be extended
to sampling more general distributions. We saw already that the ball walk can
be used along with a Metropolis filter to sample arbitrary density functions.
When is this method efficient? In [Applegate and Kannan 1990] and [Frieze
et al. 1994] respectively, it is shown that the ball walk and the lattice walk are
rapidly mixing from a warm start, provided that the density is logconcave and
it does not vary much locally, i.e., its Lipschitz constant is bounded. In [Lovász
and Vempala 2003c], the assumptions on smoothness are eliminated, and it is
shown that the ball walk is rapidly mixing from a warm start for any logconcave
function in R^n. Moreover, the mixing rate is $O(n^2 D^2)$ (ignoring the dependence
on the start), which matches the case of the uniform density on a convex body.
Various properties of logconcave functions are developed in [Lovász and Vempala
2003c] with an eye to the proof. In particular, a smoother version of any given
logconcave density is defined and used to prove an analogue of Lemma 5.3. For
a logconcave density f in R^n, the smoother version is defined as

$$\hat{f}(x) = \min_C \frac{1}{\operatorname{vol}(C)} \int_C f(x + u)\, du,$$

where C ranges over all convex subsets of the ball $x + rB_n$ with $\operatorname{vol}(C) =
\operatorname{vol}(B_n)/16$. This function is logconcave and bounded from above by f every-
where (using Theorem 2.2). Moreover, for δ small enough, its integral is close to
the integral of f. We get a lemma very similar to Lemma 5.3. The function $\hat{f}$
can be thought of as a generalization of K_δ.

LEMMA 5.5. *Let f be any logconcave density in R^n. Then*

(i) *The function $\hat{f}$ is logconcave.*
(ii) *If f is isotropic, then $\int_{\mathsf{R}^n} \hat{f}(x)\, dx \geq 1 - 64\delta^{1/2} n^{1/4}$.*

Using this along with some technical tools, it can be shown that ϕ_s is large. Perhaps the main contribution of [Lovász and Vempala 2003c] is to move the smoothness conditions from requirements on the input (i.e., the algorithm) to tools for the proof.

In summary, analyzing the ball walk has led to many interesting developments: isoperimetric inequalities, more general methods of proving convergence (ϕ_s) and many tricks for sampling to get around the fact that it is not rapidly mixing from general starting points (or distributions). The analysis of the speedy walk shows that most points are good starting points. However, it is an open question as to whether the ball walk is rapidly mixing from a pre-determined starting point, e.g., the centroid.

6. Mixing of Hit-and-Run

Hit-and-run, introduced in [Smith 1984], offers the attractive possibility of long steps. There is some evidence that it is fast in practice [Berbee et al. 1987; Zabinsky et al. 1993].

Warm start. Lovász [1999] showed that hit-and-run mixes rapidly from a warm start in a convex body K. If we start with an M-warm distribution, then in

$$O\left(\frac{M^2}{\varepsilon^2}n^2 D^2 \ln\left(\frac{M}{\varepsilon}\right)\right)$$

steps, the distance between the current distribution and the stationary is at most ε. This is essentially the same bound as for the ball walk, and so hit-and-run is no worse. The proof is based on cross-ratio isoperimetry (Theorem 4.5) for convex bodies and a new lemma about the overlap of one-step distributions. For $x \in K$, let y be a random step from x. Then the step-size $F(x)$ at x is defined as

$$\mathsf{P}\left(|x - y| \leq F(x)\right) = \tfrac{1}{8}.$$

The lemma below asserts that if u, v are close in Euclidean distance and cross-ratio distance then their one-step distributions overlap substantially. This is analogous to Lemma 5.1 for the ball walk.

LEMMA 6.1. *Let $u, v \in K$. Suppose that*

$$d_K(u, v) < \frac{1}{8} \ and \ |u - v| < \frac{2}{\sqrt{n}} \max\{F(u), F(v)\}.$$

Then

$$\|P_u - P_v\|_{tv} < 1 - \frac{1}{500}.$$

Hit-and-run generalizes naturally to sampling arbitrary functions. The isoperimetry, the one-step lemma and the bound on ϕ_s were all extended to arbitrary logconcave densities in [Lovász and Vempala 2003d]. Thus, hit-and-run is rapidly mixing for any logconcave density from a warm start. While the analysis

is along the lines of that in [Lovász 1999] and uses the tools developed in [Lovász and Vempala 2003c], it has to overcome substantial additional difficulties.

So hit-and-run is at least as fast as the ball walk. But is it faster? Can it get stuck in corners (points of small local conductance) like the ball walk?

Any start. Let us revisit the bad example for the ball walk: starting near the apex of a rotational cone. If we start hit-and-run at any interior point, then it exhibits a small, but inverse polynomial, drift towards the base of the cone. Thus, although the initial steps are tiny, they rapidly get larger and the current point moves away from the apex. This example shows two things. First, the "step-size" of hit-and-run can be arbitrarily small (near the apex), but hit-and-run manages to escape from such regions. This phenomenon is in fact general as shown by the following theorem, proved recently in [Lovász and Vempala 2004].

THEOREM 6.2. *The conductance of hit-and-run in a convex body of diameter D is $\Omega(1/nD)$.*

Unlike the ball walk, we can bound the conductance of hit-and-run (for arbitrarily small subsets). From this we get a bound on mixing time.

THEOREM 6.3. *Let K be a convex body that contains a unit ball and has centroid z_K. Suppose that $\mathsf{E}_K(|x - z_K|^2) \le R^2$ and $\|Q_0/Q\| \le M$. Then after*

$$t \ge Cn^2R^2 \ln^3 \frac{M}{\varepsilon},$$

steps, where C is an absolute constant, we have $\|Q_t - Q\|_{tv} \le \varepsilon$.

The theorem improves on the bound for the ball walk (5–2) by reducing the dependence on M and ε from polynomial (which is unavoidable for the ball walk) to logarithmic, while maintaining the (optimal) dependence on R and n. For a body in near-isotropic position, $R = O(\sqrt{n})$ and so the mixing time is $O^*(n^3)$. One also gets a polynomial bound starting from *any single* interior point. If x is at distance d from the boundary, then the distribution obtained after one step from x has $\|Q_1/Q\| \le (n/d)^n$ and so applying the above theorem, the mixing time is $O(n^4 \ln^3(n/d\varepsilon))$.

The main tool in the proof is a new isoperimetric inequality based on "average" distance (Theorem 4.5). The proof of conductance is on the same lines as shown for Theorem 5.2 in the previous section. It uses Lemma 6.1 for comparing one-step distributions.

Theorems 6.2 and 6.3 have been extended in [Lovász and Vempala 2004] to the case of sampling an exponential density function over a convex body, i.e., $f(x)$ is restricted to a convex body and is proportional to $e^{a^T x}$ for some fixed vector a. It remains open to determine if hit-and-run mixes rapidly from any starting point for arbitrary logconcave functions.

As in the ball walk analysis, it is not known (even in the convex body case) if starting at the centroid is as good as a warm start. Also, while the theorem is

the best possible in terms of R, it is conceivable that for an isotropic body the mixing rate is $O(n^2)$.

7. Efficient Sampling

Let f be a density in R^n with corresponding measure π_f. Sampling f, i.e., generating independent random points distributed according to π_f is a basic algorithmic problem with many applications. We have seen in previous sections that if f is logconcave there are natural random walks in R^n that will converge to π_f. Does this yield an efficient sampling algorithm?

Rounding. Take the case when f is uniform over a convex body K. The convergence depends on the diameter D of K (or the second moment). So the resulting algorithm to get a random point would take $\mathrm{poly}(n, D)$ steps. However, the input to the algorithm is only D and an oracle. So we would like an algorithm whose dependence on D is only logarithmic. How can this be done? The ellipsoid algorithm can be used to find a transformation that achieves $D = O(n^{1.5})$ in $\mathrm{poly}(n, \log D)$ steps.

Isotropic position provides a better solution. For a convex body in isotropic position $D \leq n$. For an isotropic logconcave distribution, $(1 - \varepsilon)$ of its measure lies in a ball of radius $\sqrt{n} \ln(1/\varepsilon)$. But how to make f isotropic? One way is by sampling. We get m random points from f and compute an affine tranformation that makes this set of points isotropic. We then apply this transformation to f. It is shown in [Rudelson 1999], that the resulting distribution is near-isotropic with $m = O(n \log^2 n)$ points for convex bodies and and $m = O(n \log^3 n)$ for logconcave densities [Lovász and Vempala 2003c] with large probability.

Although this sounds cyclic (we need samples to make the sampling efficient) one can bootstrap and make larger and larger subsets of f isotropic. For a convex body K such an algorithm was given in [Kannan et al. 1997]. Its complexity is $O^*(n^5)$. This has been improved to $O^*(n^4)$ in [Lovász and Vempala 2003b]. The basic approach in [Kannan et al. 1997] is to define a series of bodies, $K_i = K \cap 2^{i/n} B_n$. Then $K_0 = B_n$ is isotropic upto a radial scaling. Given that K_i is 2-isotropic, K_{i+1} will be 6-isotropic and so we can sample efficiently from it. We use these samples to compute a transformation that makes K_{i+1} 2-isotropic and continue. The number of samples required in each phase is $O^*(n)$ and the total number of phases is $O(n \log D)$. Since each sample is drawn from a near-isotropic convex body, the sample complexity is $O^*(n^3)$ on average ($O^*(n^4)$ for the first point and $O^*(n^3)$ for subsequent points since we have a warm start). This gives an overall complexity of $O^*(n^5)$. The improvement to $O^*(n^4)$ uses ideas from the latest volume algorithm [Lovász and Vempala 2003b], including sampling from an exponential density and the pencil construction (see Section 9).

A similar method also works for making a logconcave density f isotropic [Lovász and Vempala 2003c]. We consider a series of level sets

$$L_i = \left\{ x \in \mathsf{R}^n : f(x) \geq \frac{M_f}{2^{(1+1/n)^i}} \right\}$$

where M_f is the maximum value of f. In phase i, we make the restriction of f to L_i isotropic. The complexity of this algorithm is $O^*(n^5)$. It is an open question to reduce this to $O^*(n^4)$.

Independence. The second important issue to be addressed is that of independence. If we examine the current point every m steps for some m, then are these points independent? Unfortunately, they might not be independent even if m is as large as the mixing time. Another problem is that the distribution might not be exactly π_f. The latter problem is easier to deal with. Suppose that x is from some distribution π so that $\|\pi - \pi_f\|_{tv} \leq \varepsilon$. Typically this affects the algorithm using the samples by some small function of ε. There is a general way to handle this (sometimes called *divine intervention*). We can pretend that x is drawn from π_f with probability $1 - \varepsilon$ and from some other distribution with probability at most ε. If we draw k samples, then the probability of success (i.e., each sample is drawn from the desired distribution) is at least $1 - k\varepsilon$.

Although points spaced apart by m steps might not be independent, they are "nearly" independent in the following sense. Two random variables X, Y will be called μ-*independent* $(0 < \mu < 1)$ if for any two sets A, B in their ranges,

$$\left| \mathsf{P}(X \in A, \ Y \in B) - \mathsf{P}(X \in A)\mathsf{P}(Y \in B) \right| \leq \mu.$$

The next lemma summarizes some properties of μ-independence.

LEMMA 7.1. (i) *Let X and Y be μ-independent, and f, g be two measurable functions. Then $f(X)$ and $g(Y)$ are also μ-independent.*
(ii) *Let X, Y be μ-independent random variables such that $0 \leq X \leq a$ and $0 \leq Y \leq b$. Then*

$$\left| \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y) \right| \leq \mu ab.$$

(iii) *Let $X_0, X_1, \ldots,$ be a Markov chain, and assume that for some $i > 0$, X_{i+1} is μ-independent from X_i. Then X_{i+1} is μ-independent from $(X_0, \ldots, X_i)$.*

The guarantee that π is close to π_f will imply the following.

LEMMA 7.2. *Let Q be the stationary distribution of a Markov chain and t be large enough so that for any starting distribution Q_0 with $\|Q_0/Q\| \leq 4M$ we have $\|Q_t - Q\|_{tv} \leq \varepsilon$. Let X be a random point from a starting distribution Q_0 such that $\|Q_0/Q\| \leq M$. Then the point Y obtained by taking t steps of the chain starting at X is 2ε-independent from X.*

PROOF. Let $A, B \subseteq \mathsf{R}^n$; we claim that

$$\left| \mathsf{P}(X \in A, Y \in B) - \mathsf{P}(X \in A)\mathsf{P}(Y \in B) \right|$$
$$= \mathsf{P}(X \in A)\left| \mathsf{P}(Y \in B \mid X \in A) - \mathsf{P}(Y \in B) \right|$$
$$\leq 2\varepsilon.$$

Since

$$\left| \mathsf{P}(X \in A, Y \in B) - \mathsf{P}(X \in A)\mathsf{P}(Y \in B) \right|$$
$$= \left| \mathsf{P}(X \in \bar{A}, Y \in B) - \mathsf{P}(X \in \bar{A})\mathsf{P}(Y \in B) \right|$$

we may assume that $Q_0(A) \geq 1/2$. Let Q_0' be the restriction of Q_0 to A, scaled to a probability measure. Then $Q_0' \leq 2Q_0$ and so $\|Q_0'/Q\| \leq 4\|Q_0/Q\| \leq 4M$. Imagine running the Markov chain with starting distribution Q_0'. Then, by the assumption on t,

$$\left| \mathsf{P}(Y \in B \mid X \in A) - \mathsf{P}(Y \in B) \right| = \|Q_t'(B) - Q_t(B)\|_{tv}$$
$$\leq \|Q_t'(B) - Q(B)\|_{tv} + \|Q_t(B) - Q(B)\|_{tv}$$
$$\leq 2\varepsilon,$$

and so the claim holds.

Various versions of this lemma, adapted to the mixing guarantee at hand, have been used in the literature. See [Lovász and Simonovits 1993; Kannan et al. 1997; Lovász and Vempala 2003b] for developments along this line.

We conclude this section with an effective theorem from [Lovász and Vempala 2003a] for sampling from an arbitrary logconcave density.

THEOREM 7.3. *If f is a near-isotropic logconcave density function, then it can be approximately sampled in time $O^*(n^4)$ and in amortized time $O^*(n^3)$ if n or more nearly independent sample points are needed; any logconcave function can be brought into near-isotropic position in time $O^*(n^5)$.*

8. Application I: Convex Optimization

Let $S \subset \mathsf{R}^n$, and $f : S \rightarrow \mathsf{R}$ be a real-valued function. Optimization is the following basic problem: min $f(x)$ s.t. $x \in S$, that is, find a point $x \in S$ which minimizes by f. We denote by x^* a solution for the problem. When the set S and the function f are convex[2], we obtain a class of problems which are solvable in poly$(n, \log(1/\varepsilon))$ time where ε defines an optimality criterion. If x is the point found, then $|x - x^*| \leq \varepsilon$.

The problem of minimizing a convex function over a convex set in R^n is a common generalization of well-known geometric optimization problems such as linear

---

[2]In fact, it is enough for f to be quasi-convex.

programming as well as a variety of combinatorial optimization problems including matchings, flows and matroid intersection, all of which have polynomial-time algorithms [Grötschel et al. 1988]. It is shown in [Grötschel et al. 1988] that the ellipsoid method [Judin and Nemirovskiĭ 1976; Hačijan 1979] solves this problem in polynomial time when K is given by a separation oracle. A different, more efficient algorithm is given in [Vaidya 1996]. Here, we discuss the recent algorithm of [Bertsimas and Vempala 2004] which is based on random sampling.

Note that minimizing a quasi-convex function is easily reduced to the feasibility problem: to minimize a quasi-convex function $f(x)$, we simply add the constraint $f(x) \le t$ and search (in a binary fashion) for the optimal t.

In the description below, we assume that the convex set K is contained in the axis-aligned cube of width R centered at the origin; further if K is non-empty then it contains a cube of width r. It is easy to show that any algorithm with this input specification needs to make at least $n \log(R/r)$ oracle queries. The parameter L is equal to $\log \frac{R}{r}$.

Algorithm.
Input: A separation oracle for a convex set K and L.
Output: A point in K or a guarantee that K is empty.

1. Let P be the axis-aligned cube of side length R and center $z = 0$.
2. Check if z is in K. If so, report z and stop. If not, set

$$H = \{x \mid a^T x \le a^T z\}.$$

   where $a^T x \le b$ is the halfspace containing K reported by the oracle.
3. Set $P = P \cap H$. Pick m random points $y^1, y^2, \ldots, y^m$ from P. Set z to be their average.
4. Repeat steps 2 and 3 at most $2nL$ times. Report K is empty.

The number of samples required in each iteration, m, is $O(n)$. Roughly speaking, the algorithm is computing an approximate centroid in each iteration. The idea of an algorithm based on computing the *exact* centroid was suggested in [Levin 1965]. Indeed, if we could compute the centroid in each iteration, then by Lemma 2.3, the volume of P falls by a constant factor $(1 - 1/e)$ in each iteration. But, finding the centroid, is #P-hard, i.e., computationally intractable.

The idea behind the algorithm is that an approximate centroid can be computed using $O(n)$ random points and the volume of P is likely to drop by a constant factor in each iteration with this choice of z. This is formalized in the next lemma. Although we need it only for convex bodies, it holds for arbitrary logconcave densities.
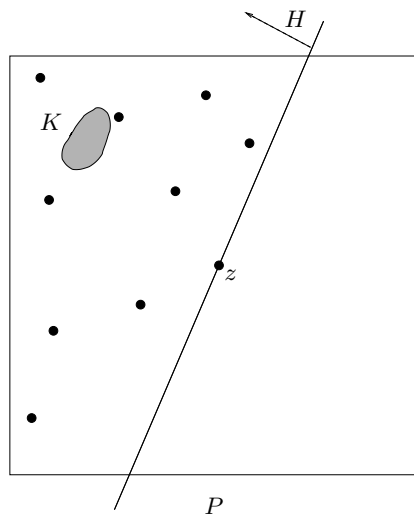
Figure 3. An illustration of the algorithm.

LEMMA 8.1. *Let g be a logconcave density in R^n and z be the average of m random points from π_g. If H is a halfspace containing z,*

$$\mathsf{E}\left(\pi_g(H)\right) \geq \left(\frac{1}{e} - \sqrt{\frac{n}{m}}\right).$$

PROOF. First observe that we can assume g is in isotropic position. This is because a linear transformation A affects the volume of a set S as $\mathrm{vol}(AS) = |\det(A)|\mathrm{vol}(S)$ and so the ratio of the volumes of two subsets is unchanged by the transformation. Applying this to all the level sets of g, we get that the ratio of the measures of two subsets is unchanged.

Since $z = \frac{1}{m}\sum_{i=1}^{m} y^i$,

$$\mathsf{E}\left(|z|^2\right) = \frac{1}{m^2}\sum_{i=1}^{m}\mathsf{E}\left(|y^i|^2\right) = \frac{1}{m}\mathsf{E}\left(|y^i|^2\right) = \frac{1}{m}\sum_{j=1}^{n}\mathsf{E}\left((y_j^i)^2\right) = \frac{n}{m},$$

where the first equality follows from the independence between y^i's, and equalities of the second line follow from the isotropic position. Let h be a unit vector normal to H. We can assume that $h = e_1 = (1,0,\ldots,0)$.

Next, let f be the marginal of g along h, i.e.,

$$f(y) = \int_{x:x_1=y} g(x)\,dx. \tag{8-1}$$

It is easy to see that f is isotropic. The next lemma (from [Lovász and Vempala 2003c]; see [Bertsimas and Vempala 2004] for the case of f arising from convex bodies) states that its maximum must be bounded.

LEMMA 8.2. *Let $f : \mathsf{R} \to R_+$ be an isotropic logconcave density function. Then,*

$$\max_y f(y) < 1.$$

Using Lemma 2.3,

$$\begin{aligned}
\int_{z_1}^{\infty} f(y)\, dy &= \int_0^{\infty} f(y)\, dy - \int_0^{z_1} f(y)\, dy \\
&\geq \frac{1}{e} - |z_1| \max_y f(y) \\
&\geq \frac{1}{e} - |z|.
\end{aligned}$$

The lemma follows from the bound on $\mathsf{E}(|z|)$.

The guarantee on the algorithm follows immediately. This optimal guarantee is also obtained in [Vaidya 1996]; the ellipsoid algorithm needs $O(n^2 L)$ oracle calls.

THEOREM 8.3. *With high probability, the algorithm works correctly using at most $2nL$ oracle calls (and iterations).*

The algorithm can also be modified for optimization given a membership oracle only and a point in K. It has a similar flavor: get random points from K; restrict K using the function value at the average of the random points; repeat. The oracle complexity turns out to be $O(n^5 L)$ which is an improvement on previous methods. This has been improved for linear objective functions using a variant of simulated annealing [Kalai and Vempala 2005].

9. Application II: Volume Computation

Finally, we come to perhaps the most important application and the principal motivation behind many developments in the theory of random walks: the problem of computing the volume of a convex body.

Let K be a convex body in R^n of diameter D such that $B_n \subset K$. The next theorem from [Bárány and Füredi 1987], improving on [Elekes 1986], essentially says that a deterministic algorithm cannot estimate the volume efficiently.

THEOREM 9.1. *For every deterministic algorithm that runs in time $O(n^a)$ and outputs two numbers A and B such that $A \leq \mathrm{vol}(K) \leq B$ for any convex body K, there is some convex body for which the ratio B/A is at least*

$$\left(\frac{cn}{a \log n} \right)^n$$

where c is an absolute constant.

So, in polynomial time, the best possible approximation is exponential in n and to get a factor 2 approximation (say), one needs exponential time. The basic idea of the proof is simple. Consider an oracle that answers "yes" for any point

Reference	Complexity	New ingredient(s)
[Dyer et al. 1991]	n^{23}	Everything
[Lovász and Simonovits 1990]	n^{16}	Localization lemma
[Applegate and Kannan 1990]	n^{10}	Logconcave sampling
[Lovász 1990]	n^{10}	Ball walk
[Dyer and Frieze 1991]	n^{8}	Better error analysis
[Lovász and Simonovits 1993]	n^{7}	Many improvements
[Kannan et al. 1997]	n^{5}	Isotropy, speedy walk
[Lovász and Vempala 2003b]	n^{4}	Annealing, hit-and-run

Table 1. Complexity comparison

in a unit ball and "no" to any point outside. After m "yes" answers, the convex body K could be anything between the ball and the convex hull of the m query points. The ratio of these volumes is exponential in n.

Given this lower bound, the following result of Dyer, Frieze and Kannan [Dyer et al. 1991] is quite remarkable.

THEOREM 9.2. *For any convex body K and any $0 \le \varepsilon, \delta \le 1$, there is a randomized algorithm which computes an estimate V such that with probability at least $1 - \delta$, we have $(1 - \varepsilon)\mathrm{vol}(K) \le V \le (1 + \varepsilon)\mathrm{vol}(K)$, and the number of oracle calls is $poly(n, 1/\varepsilon, \log(1/\delta))$.*

Using randomness, we can go from an exponential approximation to an arbitrarily small one!

The main tool used in [Dyer et al. 1991] is sampling by a random walk. They actually used the grid walk and showed that by "fixing up" K a bit without changing its volume by much, the grid walk can sample nearly random points in polynomial time. Even though the walk is discrete, its analysis relies on a continuous isoperimetric inequality, quite similar to the one used for the analysis of the ball walk. The original algorithm of Dyer, Frieze and Kannan had complexity $O^*(n^{23})$. In the years since, there have been many interesting improvements. These are summarized in Table 9. In this section, we describe the latest algorithm from [Lovász and Vempala 2003b] whose complexity is $O^*(n^4)$.

Let us first review the common structure of previous volume algorithms. Assume that the diameter D of K is $poly(n)$. All these algorithms reduce volume computation to sampling from a convex body, using the "Multi-Phase Monte-Carlo" technique. They construct a sequence of convex bodies $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = K$, where $K_0 = B_n$ or some body whose volume is easily computed. They estimate the ratios $\mathrm{vol}(K_{i-1})/\mathrm{vol}(K_i)$ by generating sufficiently many independent (nearly) uniformly distributed random points in K_i and counting the fraction that lie in K_{i-1}. The product of these estimates is an estimate of $\mathrm{vol}(K_0)/\mathrm{vol}(K)$.

In order to get a sufficiently good estimate for the ratio $\mathrm{vol}(K_{i-1})/\mathrm{vol}(K_i)$, one needs about $m\,\mathrm{vol}(K_i)/\mathrm{vol}(K_{i-1})$ random points. So we would like to have the ratios $\mathrm{vol}(K_i)/\mathrm{vol}(K_{i-1})$ be small. But, the ratio of $\mathrm{vol}(K)$ and $\mathrm{vol}(K_0)$ could be $n^{\Omega(n)}$ and so m has to be $\Omega(n)$ just to keep the ratios $\mathrm{vol}(K_i)/\mathrm{vol}(K_{i-1})$ polynomial. The best choice is to keep these ratios bounded; this can be achieved, for instance, if $K_0 = B_n$ and $K_i = K \cap (2^{i/n} B_n)$ for $i = 1, 2, \ldots, m = \Theta(n \log n)$. Thus, the total number of random points used is $O(m^2) = O^*(n^2)$. Since $\mathrm{vol}(K_i) \leq 2\mathrm{vol}(K_{i-1})$ for this sequence, a random point in K_{i-1} provides a warm start for sampling from K_i. So each sample takes $O^*(n^3)$ steps to generate, giving an $O^*(n^5)$ algorithm. In [Applegate and Kannan 1990; Lovász and Simonovits 1993], sampling uniformly from K_i was replaced by sampling from a smooth logconcave function to avoid bad local conductance and related issues.

The number of phases, m, enters the running time as its square and one would like to make it as small as possible. But, due to the reasons described above, $m = \Theta(n \log n)$ is optimal for this type of algorithm and reducing m any further (i.e., $o(n)$) seems to be impossible for this type of method.

The algorithm in [Lovász and Vempala 2003b] can be viewed as a variation of *simulated annealing*. Introduced in [Kirkpatrick et al. 1983], simulated annealing is a general-purpose randomized search method for optimization. It walks randomly in the space of possible solutions, gradually adjusting a parameter called "temperature". At high temperature, the random walk converges to the uniform distribution over the whole space; as the temperature drops, the stationary distribution becomes more and more biased towards the optimal solutions.

Instead of a sequence of bodies, the algorithm in [Lovász and Vempala 2003b] constructs a sequence of functions $f_0 \leq f_1 \leq \ldots \leq f_m$ that "connect" a function f_0 whose integral is easy to find to the characteristic function f_m of K. The ratios $(\int f_{i-1})/(\int f_i)$ can be estimated by sampling from the distribution whose density function is proportional to f_i, and averaging the function f_{i-1}/f_i over the sample points. Previous algorithms can be viewed as the special case where the functions f_i are characteristic functions of the convex bodies K_i. By choosing a different set of f_i, the algorithm uses only $m = O^*(\sqrt{n})$ phases, and $O^*(\sqrt{n})$ sample points in each phase. In fact, it uses exponential functions of the form $f(x) = e^{-a^T x/T}$ restricted to some convex body. The temperature T will start out at a small value and increase gradually. This is the reverse of what happens in simulated annealing.

Besides annealing, the algorithm uses a pre-processing step called the *pencil* construction. We describe it next.

Let K be the given body in $\mathbb{R}^n$ and $\varepsilon > 0$. Let C denote the cone in $\mathbb{R}^{n+1}$ defined by

$$C = \left\{ x \in \mathbb{R}^{n+1} : x_0 \geq 0, \sum_{i=1}^{n} x_i^2 \leq x_0^2 \right\}$$

where $x = (x_0, x_1 \ldots, x_n)^\mathsf{T}$. We define a new convex body $K' \in \mathrm{R}^{n+1}$ as follows:

$$K' = \big([0, 2D] \times K\big) \cap C.$$

In words, K' is a sharpened $(n + 1)$-dimensional "pencil" whose cross-section is K and its tip is at the origin. See Figure 4. The idea of the algorithm is to start
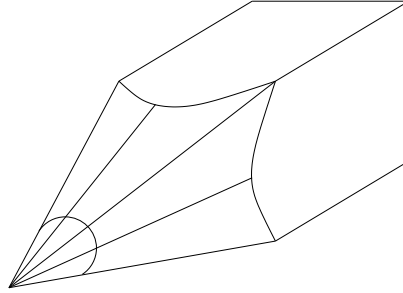

Figure 4. The pencil construction when K is a square.

with a function that is concentrated near the tip of the pencil, and is thus quite close to an exponential over a cone, and gradually move to a nearly constant function over the whole pencil, which would give us the volume of the pencil. The integral of the starting function is easily estimated. The sharpening takes less than half of the volume of the pencil away. Hence, if we know the volume of K', it is easy to estimate the volume of K by generating $O(1/\varepsilon^2)$ sample points from the uniform distribution on $[0, 2D] \times K$ and then counting how many of them fall into K'.

We describe the annealing part of the algorithm in a bit more detail. For each real number $a > 0$, let

$$Z(a) = \int_{K'} e^{-ax_0} \, dx$$

where x_0 is the first coordinate of x. For $a \leq \varepsilon^2/D$, an easy computation shows that

$$(1 - \varepsilon)\mathrm{vol}(K') \leq Z(a) \leq \mathrm{vol}(K).$$

On the other hand, for $a \geq 2n$ the value of $Z(a)$ is essentially the same as the integral over the whole cone which is easy to compute. So, if we select a sequence $a_0 > a_1 > \ldots > a_m$ for which $a_0 \geq 2n$ and $a_m \leq \varepsilon^2/D$, then we can estimate $\mathrm{vol}(K')$ by

$$Z(a_m) = Z(a_0) \prod_{i=0}^{m-1} \frac{Z(a_{i+1})}{Z(a_i)}.$$

The algorithm estimates each ratio $R_i = Z(a_{i+1})/Z(a_i)$ as follows. Let μ_i be the probability distribution over K' with density proportional to $e^{-a_i x_0}$, i.e., for $x \in K'$,

$$\frac{d\mu_i(x)}{dx} = \frac{e^{-a_i x_0}}{Z(a_i)}.$$

To estimate the ratio R_i, the algorithm draws random samples $X^1, \ldots, X^k$ from μ_i, and computes

$$W_i = \frac{1}{m} \sum_{j=1}^{m} e^{(a_i - a_{i+1})(X^j)_0}.$$

It is easy to see that $\mathsf{E}(W_i) = R_i$. The main lemma in the analysis is that the second moment of W_i is small.

LEMMA 9.3. *Let X be a random sample from $d\mu_i$ and*

$$Y = e^{(a_i - a_{i+1})X_0}.$$

Then,

$$\frac{\mathsf{E}(Y^2)}{\mathsf{E}(Y)^2} \leq \frac{a_{i+1}^2}{a_i(2a_{i+1} - a_i)}.$$

From the lemma, it follows that with

$$a_{i+1} = a_i \left(1 - \frac{1}{\sqrt{n}} \right)$$

we get that $\mathsf{E}(Y^2)/\mathsf{E}(Y)^2$ is bounded by a constant. So with k samples,

$$\frac{\mathsf{E}(W_i^2)}{\mathsf{E}(W_i)^2} \leq \left(1 + \frac{O(1)}{k} \right).$$

Hence, the standard deviation of the estimate $Z = W_1 W_2 \ldots W_m$ is at most ε times $\mathsf{E}(W_1 W_2 \ldots W_m) = \mathrm{vol}(K)$, for $k = O(m/\varepsilon^2)$. Further, the number of phases needed to go from $a_0 = 2n$ to $a_m \leq \varepsilon^2/D$ is only $\sqrt{n} \log(D/\varepsilon^2)$. So the total number of sample points needed is only $O^*(n)$ (it would be interesting to show that this is a lower bound for any algorithm that uses a blackbox sampler).

As mentioned earlier, the samples obtained are not truly independent. This introduces technical complications. In previous algorithms, the random variable estimating the ratio was bounded (between 1 and 2) and so we could directly use Lemma 7.1. For the new algorithm, the individual ratios being estimated could be unbounded. To handle this further properties of μ-independence are developed in [Lovász and Vempala 2003b]. We do not go into the details here.

How fast can we sample from μ_i? Sampling from μ_0 is easy. But each μ_{i-1} no longer provides a warm start for μ_i, i.e., $d\mu_{i-1}(x)/d\mu_i(x)$ could be unbounded. However, as the next lemma asserts, the L_2 distance is bounded.

LEMMA 9.4. $\|\mu_{i-1}/\mu_i\| < 8$.

The proof of this lemma and that of Lemma 9.3 are both based on the following property of logconcave functions proved in [Lovász and Vempala 2003b].

LEMMA 9.5. *For $a > 0$, any convex body K and logconcave function f in R^n, the function*

$$Z(a) = a^n \int_K f(ax) \, dx$$

is logconcave.

Finally, hit-and-run only needs bounded L_2 norm to sample efficiently, and by the version of Theorem 6.3 for the exponential density, we get each sample in $O^*(n^3)$ time. Along with the bound on the number of samples, this gives the complexity bound of $O * (n^4)$ for the volume algorithm.

It is apparent that any improvement in the mixing rate of random walks will directly affect the complexity of volume computation. Such improvements seem to consistently yield interesting new mathematics as well.

References

[Aldous and Fill ≥ 2005] D. Aldous and J. Fill, "Reversible Markov chains and random walks on graphs". Available at http://stat-www.berkeley.edu/users/aldous/RWG/book.html.

[Alon 1986] N. Alon, "Eigenvalues and expanders", *Combinatorica* **6**:2 (1986), 83–96.

[Applegate and Kannan 1990] D. Applegate and R. Kannan, "Sampling and integration of near log-concave functions", pp. 156–163 in *Proc. 23th ACM STOC*, 1990.

[Bárány and Füredi 1987] I. Bárány and Z. Füredi, "Computing the volume is difficult", *Discrete Comput. Geom.* **2**:4 (1987), 319–326.

[Berbee et al. 1987] H. C. P. Berbee, C. G. E. Boender, A. H. G. Rinnooy Kan, C. L. Scheffer, R. L. Smith, and J. Telgen, "Hit-and-run algorithms for the identification of nonredundant linear inequalities", *Math. Programming* **37**:2 (1987), 184–207.

[Bertsimas and Vempala 2004] D. Bertsimas and S. Vempala, "Solving convex programs by random walks", *J. Assoc. Comput. Mach.* **51**:4 (2004), 540–556.

[Bollobás 1997] B. Bollobás, "Volume estimates and rapid mixing", pp. 151–182 in *Flavors of geometry*, edited by S. Levy, Math. Sci. Res. Inst. Publ. **31**, Cambridge Univ. Press, Cambridge, 1997.

[Diaconis and Saloff-Coste 1996] P. Diaconis and L. Saloff-Coste, "Logarithmic Sobolev inequalities for finite Markov chains", *Ann. Appl. Probab.* **6**:3 (1996), 695–750.

[Dinghas 1957] A. Dinghas, "Über eine Klasse superadditiver Mengenfunktionale von Brunn-Minkowski-Lusternikschem Typus", *Math. Z.* **68** (1957), 111–125.

[Dodziuk and Kendall 1986] J. Dodziuk and W. S. Kendall, "Combinatorial Laplacians and isoperimetric inequality", pp. 68–74 in *From local times to global geometry, control and physics* (Coventry, 1984/85), edited by K. D. Ellworthy, Pitman Res. Notes Math. Ser. **150**, Longman Sci. Tech., Harlow, 1986.

[Dyer and Frieze 1991] M. Dyer and A. Frieze, "Computing the volume of convex bodies: a case where randomness provably helps", pp. 123–169 in *Probabilistic combinatorics and its applications* (San Francisco, CA, 1991), edited by B. Bollobás, Proc. Sympos. Appl. Math. **44**, Amer. Math. Soc., Providence, RI, 1991.

[Dyer et al. 1991] M. Dyer, A. Frieze, and R. Kannan, "A random polynomial-time algorithm for approximating the volume of convex bodies", *J. Assoc. Comput. Mach.* **38**:1 (1991), 1–17.

[Elekes 1986] G. Elekes, "A geometric inequality and the complexity of computing volume", *Discrete Comput. Geom.* **1**:4 (1986), 289–292.

[Frieze et al. 1994] A. Frieze, R. Kannan, and N. Polson, "Sampling from log-concave distributions", *Ann. Appl. Probab.* **4**:3 (1994), 812–837. Correction in **4** (1994), 1255.

[Grötschel et al. 1988] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric algorithms and combinatorial optimization*, Algorithms and Combinatorics **2**, Springer, Berlin, 1988.

[Grünbaum 1960] B. Grünbaum, "Partitions of mass-distributions and of convex bodies by hyperplanes.", *Pacific J. Math.* **10** (1960), 1257–1261.

[Hačijan 1979] L. G. a. Hačijan, "A polynomial algorithm in linear programming", *Dokl. Akad. Nauk SSSR* **244**:5 (1979), 1093–1096.

[Jerrum and Sinclair 1989] M. Jerrum and A. Sinclair, "Approximating the permanent", *SIAM J. Comput.* **18**:6 (1989), 1149–1178.

[Judin and Nemirovskiĭ 1976] D. B. Judin and A. S. Nemirovskiĭ, "Estimation of the informational complexity of mathematical programming problems", *Èkonom. i Mat. Metody* **12**:1 (1976), 128–142. Translated in *Matekon* **13**:2 (1976), 3–45.

[Kalai and Vempala 2005] A. Kalai and S. Vempala, "Simulated annealing for convex optimization", Preprint, 2005. Available at http://www-math.mit.edu/~vempala/papers/anneal.ps.

[Kannan 1994] R. Kannan, "Markov chains and polynomial time algorithms", pp. 656–671 in *35th Annual Symposium on Foundations of Computer Science* (Santa Fe, NM, 1994), IEEE Comput. Soc. Press, Los Alamitos, CA, 1994.

[Kannan et al. 1995] R. Kannan, L. Lovász, and M. Simonovits, "Isoperimetric problems for convex bodies and a localization lemma", *Discrete Comput. Geom.* **13**:3-4 (1995), 541–559.

[Kannan et al. 1997] R. Kannan, L. Lovász, and M. Simonovits, "Random walks and an $O^*(n^5)$ volume algorithm for convex bodies", *Random Structures Algorithms* **11**:1 (1997), 1–50.

[Kannan et al. 2004] R. Kannan, L. Lovász, and R. Montenegro, "Blocking conductance and mixing in random walks", 2004.

[Kirkpatrick et al. 1983] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing", *Science* **220**:4598 (1983), 671–680.

[Leindler 1972] L. Leindler, "On a certain converse of Hölder's inequality", pp. 182–184 in *Linear operators and approximation* (Oberwolfach, 1971), Internat. Ser. Numer. Math. **20**, Birkhäuser, Basel, 1972.

[Levin 1965] A. J. Levin, "An algorithm for minimizing convex functions", *Dokl. Akad. Nauk SSSR* **160** (1965), 1244–1247.

[Lovász 1990] L. Lovász, "How to compute the volume?", pp. 138–151 in *Jber. d. Dt. Math.-Verein, Jubiläumstagung 1990*, B. G. Teubner, Stuttgart, 1990.

[Lovász 1996] L. Lovász, "Random walks on graphs: a survey", pp. 353–397 in *Combinatorics: Paul Erdős is eighty* (Keszthely, Hungary, 1993), vol. 2, edited by D. Miklós et al., Bolyai Soc. Math. Stud. **2**, János Bolyai Math. Soc., Budapest, 1996.

[Lovász 1999] L. Lovász, "Hit-and-run mixes fast", *Math. Program.* **86**:3, Ser. A (1999), 443–461.

[Lovász and Kannan 1999] L. Lovász and R. Kannan, "Faster mixing via average conductance", pp. 282–287 in *Annual ACM Symposium on Theory of Computing* (Atlanta, 1999), ACM, New York, 1999.

[Lovász and Simonovits 1990] L. Lovász and M. Simonovits, "The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume", pp. 346–354 in *31st Annual Symposium on Foundations of Computer Science* (St. Louis, MO, 1990), IEEE Comput. Soc. Press, Los Alamitos, CA, 1990.

[Lovász and Simonovits 1992] L. Lovász and M. Simonovits, "On the randomized complexity of volume and diameter", pp. 482–491 in *33st Annual Symposium on Foundations of Computer Science* (Pittsburg, 1992), IEEE Comput. Soc. Press, Los Alamitos, CA, 1992.

[Lovász and Simonovits 1993] L. Lovász and M. Simonovits, "Random walks in a convex body and an improved volume algorithm", *Random Structures Algorithms* **4**:4 (1993), 359–412.

[Lovász and Vempala 2003a] L. Lovász and S. Vempala, "Logconcave functions: Geometry and efficient sampling algorithms", pp. 640–649 in *44th Annual Symposium on Foundations of Computer Science* (Cambridge, MA, 2003), IEEE Comput. Soc. Press, Los Alamitos, CA, 2003. Available at http://www-math.mit.edu/~vempala/papers/logcon.ps.

[Lovász and Vempala 2003b] L. Lovász and S. Vempala, "Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm", pp. 650–659 in *44th Annual Symposium on Foundations of Computer Science* (Cambridge, MA, 2003), IEEE Comput. Soc. Press, Los Alamitos, CA, 2003. Available at http://www-math.mit.edu/~vempala/papers/vol4.ps. Also to appear in *JCSS*.

[Lovász and Vempala 2003c] L. Lovász and S. Vempala, "The geometry of logconcave functions and an $O^*(n^3)$ sampling algorithm", Technical report MSR-TR-2003-04, 2003. Available at http://www-math.mit.edu/~vempala/papers/logcon-ball.ps.

[Lovász and Vempala 2003d] L. Lovász and S. Vempala, "Hit-and-run is fast and fun", Technical report MSR-TR-2003-05, 2003. Available at http://www-math.mit.edu/~vempala/papers/logcon-hitrun.ps.

[Lovász and Vempala 2004] L. Lovász and S. Vempala, "Hit-and-run from a corner", in *36th Annual Symposium on the Theory of Computing* (Chicago, 2004), ACM, New York, 2004. Available at http://www-math.mit.edu/~vempala/papers/start.ps.

[Norris 1998] J. R. Norris, *Markov chains*, Cambridge University Press, Cambridge, 1998.

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

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

[Revuz 1975] D. Revuz, *Markov chains*, North-Holland, Amsterdam, 1975.

[Rudelson 1999] M. Rudelson, "Random vectors in the isotropic position", *J. Funct. Anal.* **164**:1 (1999), 60–72.

[Simonovits 2003] M. Simonovits, "How to compute the volume in high dimension?", *Math. Program.* **97**:1-2, Ser. B (2003), 337–374.

[Smith 1984] R. L. Smith, "Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions", *Oper. Res.* **32**:6 (1984), 1296–1308.

[Vaidya 1996] P. M. Vaidya, "A new algorithm for minimizing convex functions over convex sets", *Math. Programming Ser. A* **73**:3 (1996), 291–341.

[Zabinsky et al. 1993] Z. B. Zabinsky, R. L. Smith, J. F. McDonald, H. E. Romeijn, and D. E. Kaufman, "Improving hit-and-run for global optimization", *J. Global Optim.* **3**:2 (1993), 171–192.

Santosh Vempala
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
   vempala@mit.edu