The Annals of Applied Probability
1999, Vol. 9, No. 1, 14–26

LOG-SOBOLEV INEQUALITIES AND SAMPLING FROM LOG-CONCAVE DISTRIBUTIONS

BY ALAN FRIEZE[1] AND RAVI KANNAN[2]

Carnegie Mellon University

We consider the problem of sampling according to a distribution with log-concave density F over a convex body $K \subseteq \mathbf{R}^n$. The sampling is done using a biased random walk and we give improved polynomial upper bounds on the time to get a sample point with distribution close to F.

1. Introduction. This paper is concerned with the efficient sampling of random points from $\mathbf{R}^n$, where the underlying density F is log-concave (i.e., log F is concave). This is a natural restriction which is satisfied by many common distributions, for example, the multivariate normal. The algorithm we use generates a sample path from a Markov chain whose stationary distribution is (close to) F. The algorithm falls into the class of Metropolis algorithms. Using recent developments in the theory of rapidly mixing Markov chains, in particular, the notion of conductance [8, 5], Applegate and Kannan [1] proved a bound on the convergence rate of the chain considered in this paper. In a recent paper, Frieze, Kannan and Polson [3] proved tighter bounds using an approach related to the classical Poincaré inequalities instead of conductance. In this paper, we improve these bounds still further by using logarithmic Sobolev inequalities; see Diaconis and Saloff-Coste [2] for an expository article.

Instead of sampling from the continuum of points in $\mathbf{R}^n$, we discretize the problem by assuming that $\mathbf{R}^n$ is divided into a set of hypercubes $\mathscr{C}_{\mathbf{R}}$ of side δ (δ is a given small positive real number) and the problem is to choose one of these cubes each with probability proportional to the integral of F over the cube. (If necessary, a sample from the continuum can then be picked by standard rejection sampling techniques from the cube chosen; we omit details of this.) Second, we assume that we have a compact convex set of diameter d and we wish to choose points only from K (not all of $\mathbf{R}^n$). This is justified because clearly for any positive real number ε, we can find a compact convex set (e.g., a ball or hypercube) such that the integral of F over the set is at least $(1 - \varepsilon)$ times the integral over $\mathbf{R}^n$.

Let $\mathscr{C}$ denote the set of cubes which intersect K. Let C denote the set of centres of these cubes. For $x \in \mathbf{R}^n$, we denote the cube of side δ and centre x by $C(x)$. [Thus $C(x) \in \mathscr{C}$ if and only if $x \in C$.] We choose our sample point X

Received July 1997; revised February 1998.
[1]Supported in part by NSF Grants CR-92-25008 and CCR-95-30974.
[2]Supported in part by NSF Grant DMS-95-28973.
AMS 1991 *subject classifications*. 68Q20, 60J15
Key words and phrases. Log-Sobolev inequalities, Markov chains, random walks, log-concave

by performing a random walk over C. The walk is biased so that its steady state is (close to) what we want, and we run the walk until it is close enough to the steady state. The result of this paper concerns the rate of convergence of the walk to its steady state.

We may not be able to compute F exactly and so we assume we have good approximations $\overline{F}(x)$, $x \in C$. Further, we assume that $\overline{F}(x)$ is strictly positive for all $x \in C$. We extend $\overline{F}$ to the whole of K by putting $\overline{F}(y) = \overline{F}(x)$ for $y \in C(x)$, $x \in C$.

We can only take advantage of the log-concavity of F if our grid is sufficiently fine and our approximations $\overline{F}(x)$ are sufficiently good. In this context, we will assume that for some small $\alpha > 0$,

$$(1) \qquad (1+\alpha)^{-1}\overline{F}(x) \le F(y) \le (1+\alpha)\overline{F}(x) \qquad \forall y \in C(x).$$

When we have $\overline{F} = F$, it is easy to check that this condition is satisfied if we choose α to be $e^{M\delta} - 1$, where M is the Lipschitz constant of $\ln F$ with respect to the infinity norm [i.e., M satisfies $|\ln F(x) - \ln F(y)| \le M|x - y|_\infty$, $\forall x, y \in K$]. However, an α smaller than $e^{M\delta} - 1$ may satisfy (1); this is, in fact, the case for important functions like $F(x) = e^{-c|x|}$ and $F(x) = \exp(-c|x|^2)$ as tedious, but simple calculations show. As we will see, the rate of convergence to the steady state depends upon $(1 + \alpha)$. In typical applications, one would make $1 + \alpha$ a constant. [For example, this can be ensured by choosing $\delta = O(1/M)$.]

The walk we consider fits into the scheme of Metropolis algorithms introduced in Metropolis, Rosenberg, Rosenbluth, Teller and Teller [7]. It was used by Applegate and Kannan [1] in their paper on volume computation and was further studied by Frieze, Kannan and Polson [3].

In the following text, for any natural number m, $[m] = \{1, 2, \ldots, m\}$ and $e_1, e_2, \ldots, e_m$ are the standard basis vectors of $\mathbf{R}^m$. For convenience, we consider the continuous time version, where the time between transitions is an independent negative exponential with mean 1.

The random walk. This generates a random trajectory $X(t)$, $t \ge 0$, where $X(0)$ is picked according to some initial distribution $p_0(x)$. Fix a time t, where a transition has just taken place. At a random time $t' = t + \tau$, where τ is a negative exponential with mean 1, we do the following:

STEP 1. Choose j randomly from $[n]$. Choose σ randomly from $\{\pm 1\}$.

STEP 2. Let $y = X(t) + \delta e_j$.

STEP 3. If $y \notin C$, then $X(t') = X(t)$; replace t by t' and return to Step 1. Otherwise, put $X(t') = y$ with probability $\theta = \text{Min}\{1, \overline{F}(y)/\overline{F}(X_t)\}$ and $X(t') = X(t)$ with probability $1 - \theta$.

Let $\Gamma(x) = \{y \neq x: P(x, y) > 0\}$. Then the transition probabilities $P(x, y) = \Pr(X(t') = y \mid X(t) = x)$ are formally given by

$$P(x, y) = \frac{1}{2n}\text{Min}\left\{1, \frac{\bar{F}(y)}{\bar{F}(x)}\right\} \quad \text{for } y \in \Gamma(x)$$

and

$$P(x, x) = 1 - \sum_{y \neq x} P(x, y).$$

We refer to this as "the random walk" in this paper.

The process has a steady state with probabilities $\pi(x)$ with $\lim_{t \to \infty} \Pr(X(t) = x) = \pi(x)$ for all x independent of the distribution of $X(0)$. It is easy to verify that

$$\pi(x) = \bar{F}(x)/\Delta,$$

where $\Delta = \sum_{x \in C} \bar{F}(x)$. We assume that the $\bar{F}(x)$ are sufficiently good approximations so that sampling according to π can be considered to be our objective.

Note that this process is *time-reversible*, that is,

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{for } x, y \in C.$$

Let $p_t(x) = \Pr(X_t = x)$ be the distribution at time t and let $f_t(x) = p_t(x)/\pi(x)$. Then let $M = \max_{x \in C} f_0(x)\log f_0(x)$. For $0 \leq \theta \leq 1$, let $C_\theta = \{x \in C: \text{vol}(C(x) \cap K) \geq \theta\delta^n\}$ and $\pi_\theta = \sum_{x \notin C_\theta} \pi(x)$.

The *variational distance* between p_t and π is given by

$$\|p_t - \pi\|_{\text{TV}} = \sum_{x \in C} |p_t(x) - \pi(x)|.$$

Our main result is the following theorem.

THEOREM 1. *Assume $d \geq \delta n^{1/2}$. There is an absolute constant $\gamma > 0$ such that*

$$\|p_t - \pi\|_{\text{TV}} \leq \left(\frac{1}{2}\left(\exp\left\{-\frac{\gamma\theta t\delta^2}{nd^2}\right\}\log \pi_*^{-1} + \frac{M\pi_\theta nd^2}{\gamma\delta^2}\right)\right)^{1/2},$$

where $\pi_* = \min \pi(x)$.

Generally speaking, it is not difficult to choose $p_0(x)$ so that for $\theta = 1/10$ say, $M\pi_\theta$ is exponentially small. Usually one has p_0 concentrated on a small set S_0 and then f_0 is zero outside this set. One can then blow up K so that π_θ is sufficiently small, while only marginally changing $p_0(x)$ for $x \in S_0$. This improves results of [3], essentially by replacing π_*^{-1} by $\log \pi_*^{-1}$.

2. Back to a continuous problem. The *entropy* $\text{Ent}_\pi(\mu)$ of measure μ is given by

$$\text{Ent}_\pi(\mu) = \sum_{x \in C} \mu(x)\log \frac{\mu(x)}{\pi(x)}.$$

It follows from the convexity of x^x that

$$\mathrm{Ent}_\pi(\mu) \le \log \pi_*^{-1}. \tag{2}$$

Inequality (2.8) of [2] shows that for any measure μ, we have

$$2\|\mu - \pi\|_{\mathrm{TV}}^2 \le \mathrm{Ent}_\pi(\mu). \tag{3}$$

(For a proof, see the Appendix.)

For $\phi \in \mathbf{R}^C$, we introduce the quantities

$$\mathscr{E}(\phi, \phi) = \tfrac{1}{2} \sum_{x \in C} \sum_{y \in \Gamma(x)} (\phi(x) - \phi(y))^2 \pi(x) P(x, y) \tag{4}$$

and

$$\begin{aligned} \mathscr{L}(\phi) &= \sum_{x \in C} \phi(x)^2 \log\left(\frac{\phi(x)^2}{\|\phi\|_2^2}\right) \pi(x) \\ &= \|\phi\|_2^2 \mathrm{Ent}_\pi(\mu), \end{aligned}$$

where

$$\|\phi\|_2^2 = \sum_{x \in C} \phi(x)^2 \pi(x) \quad \text{and} \quad \mu(x) = \pi(x)\phi(x)^2/\|\phi\|_2^2 \quad \text{for } x \in C. \tag{5}$$

We now define the *log-Sobolev* constant α by

$$\begin{aligned} \alpha &= \inf\left\{ \frac{\mathscr{E}(\phi, \phi)}{\mathscr{L}(\phi)} : \mathscr{L}(\phi) \ne 0 \right\} \\ &= \inf\left\{ \frac{\mathscr{E}\left(\sqrt{\mu/\pi}, \sqrt{\mu/\pi}\right)}{\mathrm{Ent}_\pi(\mu)} : \mu \text{ is a measure on } C \right\}. \end{aligned}$$

Putting $\phi_t(x) = \sqrt{f_t(x)}$ for $x \in C$, it is shown in [2] (see Appendix for a proof) that

$$\begin{aligned} \frac{d}{dt}\mathrm{Ent}_\pi(p_t) &\le -4\mathscr{E}(\phi_t, \phi_t) \\ &\le -4\alpha\, \mathrm{Ent}_\pi(p_t). \end{aligned} \tag{6}$$

It follows that

$$\mathrm{Ent}_\pi(p_t) \le e^{-4\alpha t}\mathrm{Ent}_\pi(p_0).$$

We need to modify this in order to account for the *border* cubes $C(x)$, $x \notin C_\theta$.

2.1. *Two integrals.* Given $\phi \in \mathbf{R}^N$ and a small $\varepsilon > 0$, we define Φ_ε: $K \to \mathbf{R}$ as follows: suppose $z \in C(x)$ for some $x \in C$. Let $C(x, \varepsilon)$ denote the cube centred at x with side $\delta - 2\varepsilon$. If $z \in C(x, \varepsilon)$, we let $\Phi_\varepsilon(z) = \phi(x)$. If $z \notin C(x, \varepsilon)$, let D be a face of $C(x)$ which is closest to z. (If there is a tie for D, the value of Φ_ε does not matter, as we will see.) Suppose first that $D = C(x) \cap C(y)$ for some $y \in C$ and that $\mathrm{dist}(z, D) = \eta\varepsilon$, where $0 \le \eta < 1$.

In this case, we let $\Phi_\varepsilon(z) = ((1+\eta)\phi(x) + (1-\eta)\phi(y))/2$. In this way, if we start at a point on a face of $C(x,\varepsilon)$ parallel to D and move toward D, then Φ_ε changes linearly from $\phi(x)$ to $\phi(y)$ over a distance 2ε. Finally, if the hypercube on the other side of D to $C(x)$ is not in $\mathscr{C}$, then we keep $\Phi_\varepsilon(z) = \phi(x)$.

We consider the two integrals

$$I_\varepsilon = \int_K |\nabla\Phi_\varepsilon(z)|^2 F(z) g(z)\, dz$$

and

$$J_\varepsilon = \int_K \Phi_\varepsilon(z)^2 \log\left(\frac{\Phi_\varepsilon(z)^2 \int_K F(\zeta) g(\zeta)\, d\zeta}{\int_K \Phi_\varepsilon(\zeta)^2 F(\zeta) g(\zeta)\, d\zeta}\right) F(z) g(z)\, dz.$$

Here $g\colon K \to \mathbf{R}$ is defined by

$$g(y) = \frac{\overline{F}(z)}{F(y)} \quad \text{for } y \in C(x),\ x \in C.$$

Thus, in particular,

$$\int_{C(x)} F(y) g(y)\, dy = \delta^n \overline{F}(x), \qquad x \in C,$$

and

$$(1+\alpha)^{-1} \le g(y) \le 1+\alpha. \tag{7}$$

On a set Z of measure zero (consisting of points for which there is a tie for D), Φ_ε is not differentiable. We can, however, easily "smooth out" Φ_ε close to Z so that (4) and (1) imply

$$\begin{aligned} I_\varepsilon &= \sum_{x\in C} \int_{K\cap C(x)} |\nabla\Phi_\varepsilon(z)|^2 \overline{F}(z)\, dz \\ &\le \sum_{x\in C} \int_{C(x)} |\nabla\Phi_\varepsilon(z)|^2 \overline{F}(z)\, dz \\ &\le \varepsilon^{-1}\delta^{n-1} \sum_{x\in C} \sum_{y\in\Gamma(x)} \left(\frac{\phi(x)-\phi(y)}{2}\right)^2 \overline{F}(x) + O(1) \qquad [O(1) \text{ as } \varepsilon\to 0] \\ &\le \left(\frac{1+\alpha}{4}\right)\varepsilon^{-1}\delta^{n-1} \sum_{x\in C} \sum_{y\in\Gamma(x)} (\phi(x)-\phi(y))^2 \min\{\overline{F}(x), \overline{F}(y)\} \\ &\quad + O(1) \\ &= (1+\alpha)\varepsilon^{-1}\Delta n \delta^{n-1} \mathscr{E}(\phi,\phi) + O(1), \end{aligned} \tag{8}$$

where the term $O(1)$ may depend on n, F and ϕ. So

$$\mathscr{E}(\phi,\phi) \ge \frac{\varepsilon I_\varepsilon}{(1+\alpha)^2(1+f(\varepsilon))n\delta^{n-1}\Delta} - O(\varepsilon). \tag{9}$$

We decompose

$$\mathrm{Ent}_\pi(p_t) = \mathrm{En}_I(p_t) + \mathrm{En}_B(p_t), \tag{10}$$

where $\mathrm{En}_I(p_t) = \sum_{x \in C_\theta} p_t(x) \log f_t(x)$ and so forth.

Arguing as we did for $\mathscr{E}(\phi, \phi)$, we see that

$$\begin{aligned} J_\varepsilon &\geq \sum_{x \in C_\theta} \int_{K \cap C(x)} \Phi_\varepsilon(z)^2 \log\left(\frac{\Phi_\varepsilon(z)^2 \int_K F(\zeta) g(\zeta)\, d\zeta}{\int_K \Phi_\varepsilon(\zeta)^2 F(\zeta) g(\zeta)\, d\zeta} \right) F(z) g(z)\, dz \\ &\geq \theta \delta^n \Delta \sum_{x \in C_\theta} \phi(x)^2 \log\left(\frac{\phi(x)^2}{\|\phi\|_2^2} \right) \pi(x) + O(\varepsilon) \\ &= \theta \delta^n \Delta\, \mathrm{En}_I(\mu) + O(\varepsilon), \end{aligned} \tag{11}$$

where μ is as in (5).

Applying (6) and (9) (with $\phi = \phi_t$) and (11) (with $\mu = p_t$), we find that for some absolute constant $A_1 > 0$,

$$\begin{aligned} \frac{d}{dt}\mathrm{Ent}_\pi(p_t) &\leq -A_1 \frac{\varepsilon I_\varepsilon}{n \delta^{n-1} \Delta} + O(\varepsilon) \\ &\leq -A_1 \frac{\theta \delta \varepsilon I_\varepsilon}{n J_\varepsilon} \mathrm{En}_I(p_t) + O(\varepsilon). \end{aligned} \tag{12}$$

In the next section we prove the following theorem.

THEOREM 2. *Suppose K is a convex set in $\mathbf{R}^n$ of diameter d, F is a (positive real valued) log-concave function on K and g is any sufficiently smooth real valued function on K satisfying* (7). *Then with $f = \Phi_\varepsilon$ and $d \geq \delta n^{1/2}$, we have*

$$\begin{aligned} &\int_K f(x)^2 \log\left(\frac{f(z)^2 \int_K Fg\, d\zeta}{\int_K f^2 Fg\, d\zeta} \right) F(z) g(z)\, dz \\ &\qquad \leq A_2 d^2 \varepsilon \delta^{-1} \int_K |\nabla f(z)|^2 F(z) g(z)\, dx \end{aligned} \tag{13}$$

for some absolute constant $A_2 > 0$.

Letting $\varepsilon \to 0$, we see from (12) and the above theorem that

$$\frac{d}{dt}\mathrm{Ent}_\pi(p_t) \leq -\frac{A_1 \theta \delta^2}{A_2 n d^2} \mathrm{En}_I(p_t). \tag{14}$$

In the proof of (6) [see (29)] we find that

$$f_t(x) = e^{-t(I-P)} f_0(x).$$

This implies that

$$\max_x f_t(x) \leq \max_x f_0(x)$$

and, consequently,

$$\mathrm{En}_B(p_t) \le M\pi_\theta. \tag{15}$$

It follows from (14) and (15) that

$$\mathrm{Ent}_\pi(p_t) \le e^{-\beta t}\mathrm{Ent}_\pi(p_0) + \frac{M\pi_\theta}{\beta}, \tag{16}$$

where $\beta = A_1\theta\delta^2/A_2nd^2$. Theorem 1 now follows from (3) and (16).

3. Proof of Theorem 2. As in [3], we use the *localization lemma* of Lovász and Simonovits [6] to reduce the geometry to one dimension.

LEMMA 1. *Let f_1, f_2 be low semicontinuous functions defined on $\mathbf{R}^n$ such that*

$$\int_{\mathbf{R}^n} f_i(z)\,dz > 0, \qquad i = 1, 2.$$

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

$$\int_{t=0}^{1} f_i(ta + (1-t)b)l(t)^{n-1}dt > 0, \qquad i = 1, 2.$$

We use the fact that (13) *fails* to hold if and only if there exists $\lambda > 0$ such that

$$\int_K F(x)g(x)\,dx > \lambda\int_K f(x)^2F(x)g(x)\,dx$$

and

$$\int_K f(x)^2F(x)g(x)\log\bigl(\lambda f(x)^2\bigr)\,dx > Ad^2\varepsilon\delta^{-1}\int_K |\nabla f(x)|^2F(x)g(x)\,dx.$$

So we put

$$f_1 = Fg\bigl(1 - \lambda f^2\bigr)\chi_K$$

and

$$f_2 = Fg\bigl(f^2\log\bigl(\lambda f^2\bigr) - Ad^2\varepsilon\delta^{-1}|\nabla f|^2\bigr)\chi_K,$$

where χ_K is the indicator function of the body K.

Let a, b and l be as in Lemma 1. We observe that we can take $a, b \in K$ because of the factor χ_K. Let $f(t) = \Phi_\varepsilon((1-t)a + tb)$, $g(t) = g((1-t)a + tb)$, $h(t) = F((1-t)a + tb)l(t)^{n-1}$ and $\tilde{g}(t) = |\nabla\Phi_\varepsilon((1-t)a + tb)|$. Note that $h(t)$ is log-concave. We can assume that $|b-a| = d$ as $|b-a| \le d$ and $|b-a|$ can replace d in our proof.

We then see that if (13) fails to hold, then, where $\psi(t) = g(t)h(t)$,

$$\int_{t=0}^{1} f(t)^2 \log\left(\frac{f(t)^2 \int_{t=0}^{1} \psi(t)\, dt}{\int_{t=0}^{1} f(t)^2 \psi(t)\, dt} \right) \psi(t)\, dt > Ad^2 \varepsilon \delta^{-1} \int_{t=0}^{1} \tilde{g}(t)^2 \psi(t)\, dt. \tag{17}$$

Suppose on this line segment, h attains its maximum at ζ. We consider the two parts of the line segment $[0, \zeta]$ and $[\zeta, 1]$ separately. The arguments are symmetric and we give only one part. In fact, we will assume for simplicity that $\zeta = 0$ and h decreases monotonically on $[0, 1]$. We can make the following normalizations (the first by scaling h and the second by then scaling f, neither of which changes the theorem):

$$\begin{aligned} \int_{t=0}^{1} \psi(t)\, dt &= 1, \\ \int_{t=0}^{1} f(t)^2 \psi(t)\, dt &= 1. \end{aligned} \tag{18}$$

Let $X \in [0, 1]$ be a random variable with density function ψ. Now

$$f(t)^2 \le 2(f(t) - f(0))^2 + 2f(0)^2. \tag{19}$$

Also (from the fact that log is a concave function),

$$\begin{aligned} \int_{t=0}^{1} f(0)^2 \log\big(f(t)^2\big) \psi(t)\, dt &\le f(0)^2 \log\left(\int_{t=0}^{1} f(t)^2 \psi(t)\, dt \right) \\ &= 0. \end{aligned} \tag{20}$$

Now, putting $u = (b - a)/d$, we get

$$\begin{aligned} (f(t) - f(0))^2 &= \left(\int_{s=0}^{t} f'(s)\, ds \right)^2 \\ &= d^2 \left(\int_{s=0}^{t} \sum_{j=1}^{n} u_j \chi_j(s) \frac{\partial \Phi_\varepsilon}{\partial x_j}((1 - s)a + sb)\, ds \right)^2, \end{aligned}$$

where χ_j is defined by

$$\chi_j(s) = \begin{cases} 1, & \text{if } \dfrac{\partial \Phi_\varepsilon}{\partial x_j}((1 - s)a + sb) \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

By the Cauchy–Schwarz inequality, we have

$$\sum_{j=1}^{n} \left| u_j \chi_j(s) \frac{\partial \Phi_\varepsilon}{\partial x_j}((1 - s)a + sb) \right| \le \left(\sum_{j=1}^{n} u_j^2 \chi_j(s) \right)^{1/2} \tilde{g}(s).$$

So we get, with another application of Cauchy–Schwarz,

$$\begin{aligned}(f(t)-f(0))^2 &\le d^2\left(\int_{s=0}^{t}\left(\sum_{j=1}^{n}u_j^2\chi_j(s)\right)^{1/2}\tilde{g}(s)\,ds\right)^2\\ &\le d^2\left(\int_{s=0}^{t}\sum_{j=1}^{n}u_j^2\chi_j(s)\,ds\right)\left(\int_{s=0}^{t}\tilde{g}(s)^2\,ds\right).\end{aligned}$$

Now each time the line from 0 to t crosses a hyperplane of the form $x_j = m\delta$, m an integer, we get a contribution of $2\varepsilon/du_j$ to $\int \chi_j(s)\,ds$. Furthermore, the number of such crossings is at most

$$\frac{du_j\,t}{\delta}+1. \tag{21}$$

So we get (using the facts that $\sum_{j=1}^{n}u_j^2 = 1$ and $\sum|u_j| \le \sqrt{n}$)

$$\begin{aligned}(f(t)-f(0))^2 &\le 2d^2\varepsilon\left(\frac{t}{\delta}+\frac{n^{1/2}}{d}\right)\int_{s=0}^{t}\tilde{g}(s)^2\,ds\\ &\le B\int_{s=0}^{t}\tilde{g}(s)^2\,ds,\end{aligned}$$

where

$$B = 2d^2\varepsilon\left(\frac{1}{\delta}+\frac{n^{1/2}}{d}\right).$$

Now let $T = \{t\colon f(t)^2 \ge 1\}$ and $T_s = T \cap [s,1]$ for $0 \le s \le 1$. Then

$$\begin{aligned}&\int_{t=0}^{1}(f(t)-f(0))^2\log\bigl(f(t)^2\bigr)\psi(t)\,dt\\ &\quad\le B\int_{t\in T}\int_{s=0}^{t}\tilde{g}(s)^2\,ds\log\bigl(f(t)^2\bigr)\psi(t)\,dt\\ &\quad= B\int_{s=0}^{1}\tilde{g}(s)^2\Pr(X\in T_s)\int_{t\in T_s}\log\bigl(f(t)^2\bigr)\frac{\psi(t)}{\Pr(X\in T_s)}\,dt\,ds\\ &\quad\le B\int_{s=0}^{1}\tilde{g}^2\Pr(X\in T_s)\log\left(\int_{t\in T_s}\frac{f(t)^2\psi(t)}{\Pr(X\in T_s)}\,dt\right)ds\\ &\quad\le B\int_{s=0}^{1}\tilde{g}^2\Pr(X\in T_s)\log\left(\frac{1}{\Pr(X\in T_s)}\right)ds.\end{aligned}$$

We will prove that

$$\begin{aligned}\Pr(X\in T_s)\log\left(\frac{1}{\Pr(X\in T_s)}\right) &\le (1+\alpha)(1+e^{-1})h(s)\\ &\le (1+\alpha)^2(1+e^{-1})\psi(s)\quad\text{by (7)}.\end{aligned} \tag{22}$$

Let $h(t) = e^{-\eta(t)}$, where $\eta(t)$ is convex. Let $\lambda = \eta'(s)$. Then

$$h(t) \le h(s)e^{-\lambda(t-s)}. \tag{23}$$

So,

$$\begin{aligned}\Pr(X \in T_s) \le \Pr(X \ge s) &= \int_{t=s}^{1} \psi(t)\,dt \\ &\le (1+\alpha)\int_{t=s}^{\infty} h(s)e^{-\lambda(t-s)}\,dt \qquad (24)\\ &= \frac{(1+\alpha)h(s)}{\lambda}.\end{aligned}$$

We also have

$$(1+\alpha)^{-1} \le h(0) \le h(s)e^{\lambda}. \tag{25}$$

The first inequality follows from the monotonicity of h and (18). The second follows from (23). Thus

$$h(s) \ge (1+\alpha)^{-1}e^{-\lambda}. \tag{26}$$

The function $\xi \log(1/\xi)$ increases monotonically from 0 at $\xi = 0$ to e^{-1} at $\xi = e^{-1}$ and decreases monotonically from then on. So if $h(s) \ge 1/((1+\alpha)e)$, then

$$\Pr(X \in T_s)\log\bigl(1/\Pr(X \in T_s)\bigr) \le e^{-1} \le h(s)(1+\alpha), \tag{27}$$

which implies (22) in this case.

So assume that $h(s) < e^{-1}/(1+\alpha)$. Now $\Pr(X \in T_s) \le \Pr(X \ge s) = \int_{t=s}^{1} \psi(t)\,dt \le (1+\alpha)h(s) \le e^{-1}$. So,

$$\begin{aligned}\Pr(X \in T_s)\log\left(\frac{1}{\Pr(X \in T_s)}\right) &\le (1+\alpha)h(s)\log\left(\frac{1}{(1+\alpha)h(s)}\right)\\ &\le (1+\alpha)h(s)\lambda \quad \text{from (26)}.\end{aligned}$$

So $\lambda \le 1$ will imply that (27) holds again. We can therefore assume that $\lambda > 1$, but then $h(s)/\lambda < h(s) < e^{-1}/(1+\alpha)$ and so we obtain from (24) that

$$\begin{aligned}\Pr(X \in T_s)\log\left(\frac{1}{\Pr(X \in T_s)}\right) &\le \frac{(1+\alpha)h(s)}{\lambda}\log\left(\frac{\lambda}{(1+\alpha)h(s)}\right)\\ &\le \frac{(1+\alpha)h(s)}{\lambda}\log(\lambda e^{\lambda}) \quad \text{from (26)} \qquad (28)\\ &= (1+\alpha)h(s)\left(\frac{\log\lambda}{\lambda}+1\right)\\ &\le (1+\alpha)(e^{-1}+1)h(s)\end{aligned}$$

as claimed. So if

$$C = 2B(1+\alpha)^2(1+e^{-1}),$$

then

$$\int_{t=0}^{1} (f(t) - f(0))^2 \log\bigl(f(t)^2\bigr)\psi(t)\,dt \le C\int_{t=0}^{1} \tilde{g}^2\psi(t)\,dt.$$

Theorem 2 follows. □

There is an alternative random walk, the *ball walk*, which has been applied in this area [4–6]. It is, in some sense, preferable to the one we have studied, since better upper bounds are known on its "mixing time."

Ball walk. Replace Steps 1 and 2 of the previously discussed walk by the following steps.

STEP 1′. Choose u uniformly at random from the set $\{u \in \mathbf{R}^n: |u| \le \delta\}$.

STEP 2′. Let $y = X(t) + u$.

PROBLEM. Estimate the log-Sobolev constant for the ball walk.

APPENDIX

PROOF OF (3). Observe first that for $u > 0$, we have

$$3(u-1)^2 \le (4 + 2u)(u\log u - u + 1).$$

Applying this with $u = \mu(x)/\pi(x)$ and then multiplying by $\pi(x)$, we obtain

$$\sqrt{3}\,|\mu(x) - \pi(x)| \le (4\pi(x) + 2\mu(x))^{1/2}(\mu(x)\log(\mu(x)/\pi(x)) - \mu(x) + \pi(x))^{1/2}.$$

Applying the Cauchy–Schwarz inequality,

$$\begin{aligned}\bigl(\sqrt{3}\,\|\mu - \pi\|_{\mathrm{TV}}\bigr)^2 &\le \sum_x (4\pi(x) + 2\mu(x)) \sum_x (\mu(x)\log(\mu(x)/\pi(x)) - \mu(x) + \pi(x)) \\ &= 6\,\mathrm{Ent}_\pi(\mu).\end{aligned}$$

□

PROOF OF (6). The following sequence of derivations relies heavily on the time-reversibility of our process. We indicate these uses by an asterisk (*) at the end of the line. Observe first that if $f_t(x) = p_t(x)/\pi(x)$ and $H_t = e^{-t}e^{tP}$, then

$$\begin{aligned} f_t(x) &= \sum_{n=0}^{\infty} \frac{t^n e^{-t}}{n!} \sum_{y\in C} \frac{p_0(y)P^n(y,x)}{\pi(x)} \\ &= \sum_{n=0}^{\infty} \frac{t^n e^{-t}}{n!} \sum_{y\in C} \frac{p_0(y)P^n(x,y)}{\pi(y)} \quad * \\ &= H_t f_0(x). \end{aligned} \tag{29}$$

We apply this to compute the rate of change of $\mathrm{Ent}_\pi(p_t)$ with respect to t:

$$\begin{aligned}
&\frac{d}{dt}\mathrm{Ent}_\pi(p_t)\\
&\quad=\sum_{x\in C}\pi(x)\frac{d}{dt}\big(f_t(x)\log f_t(x)\big)\\
&\quad=\sum_{x\in C}\pi(x)\big(1+\log f_t(x)\big)\frac{d}{dt}f_t(x)\\
&\quad=\sum_{x\in C}\pi(x)\big(1+\log f_t(x)\big)\frac{d}{dt}H_tf_0(x)\\
&\quad=\sum_{x\in C}\pi(x)\big(1+\log f_t(x)\big)(P-I)f_t(x)\\
&\quad=\sum_{x\in C}\pi(x)\big(1+\log f_t(x)\big)\sum_{y\in C}P(x,y)\big(f_t(y)-f_t(x)\big)\\
&\quad=\sum_{x\in C}\pi(x)\log f_t(x)\sum_{y\in C}P(x,y)\big(f_t(y)-f_t(x)\big) \qquad *\\
&\quad=-\frac{1}{2}\sum_{x,\,y\in C}\pi(x)P(x,y)\big(\log f_t(x)-\log f_t(y)\big)\big(f_t(x)-f_t(y)\big). \qquad *
\end{aligned}$$

Now (2.7) of [2] shows that if $a\ge b\ge 0$, then

$$(\log a-\log b)(a-b)\ge 4\big(a^{1/2}-b^{1/2}\big)^2.$$

Hence,

$$\begin{aligned}
\frac{d}{dt}\mathrm{Ent}_\pi(p_t)&\le -2\sum_{x,\,y\in C}\pi(x)P(x,y)\big(f_t(x)^{1/2}-f_t(y)^{1/2}\big)^2\\
&=-4\mathscr{E}\big(f_t^{1/2},f_t^{1/2}\big)\\
&\le -4\alpha\,\mathrm{Ent}_\pi(p_t),
\end{aligned}$$

which proves (6). □

Acknowledgment. We thank Laci Lovász for pointing out an error in an earlier version of the paper.

REFERENCES

[1] APPLEGATE, D. and KANNAN, R. (1991). Sampling and integration of log-concave functions. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing* 156–163. ACM, New York.

[2] DIACONIS, P. and SALOFF-COSTE, L. (1998). Logarithmic Sobolev inequalities for finite Markov chains. Unpublished manuscript.

[3] FRIEZE, A. M., KANNAN, R. and POLSON, N. (1994). Sampling from log-concave distributions. *Ann. Appl. Probab.* **4** 812–837.

[4] KANNAN, R., LOVÁSZ, L. and SIMONOVITS, M. (1998). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms*. To appear.

[5] LOVÁSZ, L. and SIMONOVITS, M. (1990). Mixing rate of Markov chains, an isoperimetric inequality and computing the volume. In *Proceeding of the 31st Annual IEEE Symposium on Foundations of Computer Science* 346–355. IEEE, New York.
[6] LOVÁSZ, L. and SIMONOVITS, M. (1993). Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms* **4** 359–412.
[7] METROPOLIS, N., ROSENBERG, ROSENBLUTH, TELLER and TELLER (1953). Equation of state calculation by fast computing machines. *J. Chemical Physics* **21** 1087–1092.
[8] SINCLAIR, A. J. and JERRUM, M. R. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82** 93–133.

DEPARTMENT OF MATHEMATICAL SCIENCES
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213

DEPARTMENT OF COMPUTER SCIENCE
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213