A New Approach to Polynomial-Time Generation of Random Points in Convex Bodies*

Russ Bubley Martin Dyer
School of Computer Studies
University of Leeds, Leeds LS2 9JT, England

Mark Jerrum
Department of Computer Science
University of Edinburgh, Edinburgh EH9 3JZ, Scotland

March 28, 1996

Abstract

In this paper we describe a new method for proving the polynomial-time convergence of an algorithm for sampling (almost) uniformly at random from a convex body in high dimension. Previous approaches have been based on estimating conductance via isoperimetric inequalities. We show that a conceptually simpler coupling argument can be used to give a similar result.

1 Introduction

Dyer, Frieze and Kannan [7] gave the first fully polynomial randomized approximation scheme for approximating the volume of a convex body in $\mathbb{R}^n$, for large n. The algorithm was based on a reduction to sampling uniformly at random in such a body. The necessary sampling was achieved, to close enough approximation, by simulating a polynomial number of steps of a random walk in the body.

The approach of [7] to proving "rapid mixing" of the random walk was to estimate the *conductance* [22]. (This methodology was previously used in [13] to prove rapid convergence of an edge-matching random walk proposed in [4].) Conductance estimation requires some form of isoperimetric inequality. In [7] the appropriate inequality was geometric in origin.

*Research supported by EXPRIT Working Group RAND

The result of [7] has been dramatically improved in a sequence of papers. For example, see [2, 6, 18, 8, 19, 16, 15] and [14] for a recent survey. These papers have introduced ideas which are important in their own right, for example new isoperimetric inequalities. We exploit some of these ideas here, specifically the idea of using a surrogate log-concave distribution, and the use of a "Metropolised" rotationally symmetric random walk. These techniques were introduced in [2, 19]. However, while these advances have been made, all approaches to the problem have essentially similarities to [7].

Here we explore a different approach, *coupling*, which previously seemed difficult to apply to this problem. We show that, by viewing the problem from the correct perspective, it is in fact rather easy to show polynomial convergence to the uniform distribution of a certain random walk. To achieve this, we first transform the uniform sampling problem to that of generating a random point from the log-concave distribution of [19]. Then we apply our coupling arguments to a suitably chosen random walk for the transformed problem.

The most important application of the sampling algorithm is to the above-mentioned volume computation problem. We do not consider this here, but refer the interested reader to the papers already cited. The best time bound currently known is given in [15].

Our analysis contains many trade-offs between constants. We do not attempt to make these choices "optimally", and no doubt our estimates can be improved upon. Rather, for ease of exposition, we employ a single constant c which must be chosen large enough for various inequalities to hold.

Similarly, we do not address issues of precision in computation. While there is no essential difficulty here, to do so introduces a level of tedious detail which we prefer to avoid. We generally assume that computations can be carried out to the required precision, and measure the complexity of the algorithm simply by the number of steps of the random walk.

The plan of the paper is as follows. In section 2 we give the necessary backgound, and some initial estimates. In section 3 we analyse the random walk using a coupling argument. In section 4 we describe some improvements, and finally in section 5 we discuss the relationship of our work to the literature cited above.

2 Background and preliminaries

2.1 Convex bodies and gauge functions

Let $K \subseteq \mathbb{R}^n$ be a convex body with $0 \in \operatorname{int} K$ and let ∂K be its boundary. We will assume that K is *well-rounded*, i.e. for some "small" R, $B \subseteq K \subseteq RB$ where B is the unit ball in $\mathbb{R}^n$. A polynomial-time computation will yield $R = O(n^{3/2})$ [10] or, for almost-uniform generation purposes $R = O(n)$ [19]. By a recent result of Kannan, Lovász and Simonovits [16], we may even achieve $R = O(n^{1/2})$. However, the algorithm producing $R = O(n^{1/2})$ requires generating almost uniform points in K, so here we assume only $R = O(n)$.

We define the *gauge function* of K by $f_K(x) = \inf_{t \geq 0}\{x \in tK\}$, and an associated

density function[1]

$$F_K(x) = e^{-f_K(x)} \qquad (x \in \mathbb{R}^n).$$

It is well known that f_K is convex and hence F_K is log-concave. Note that $K \subseteq L$ implies $F_K \leq F_L$ pointwise. Also $f_B(x) = \|x\|$, so if K is well rounded, we have

$$\frac{\|x\|}{R} \leq f_K(x) \leq \|x\|. \tag{1}$$

Note that, since f_K is convex, it possesses at least one *subgradient* $\nabla(x)$ at each point x. Then, for all $y \in \mathbb{R}^n$ and any subgradient ∇ at x,

$$f_K(y) \geq f_K(x) + \nabla.(y - x). \tag{2}$$

It is also easy to show that, for any $v \in \partial B$, the *directional derivative* $\delta_v(x)$, of f_K at x in direction v, satisfies

$$|\delta_v(x)| \leq 1. \tag{3}$$

2.2 Sampling equivalence

We will show that sampling uniformly from K is equivalent to sampling with density F_K. Specifically, we show that there is a measure-preserving bijection between $\mathbb{R}^n$ and $\operatorname{int} K$.

Let Γ_n the distribution with probability density $e^{-t}t^{n-1}/(n-1)!$. We will assume this to be defined for non-integer as well as integer n by using the obvious integral for $(n-1)!$ in such cases. Let $\gamma_n(t) = \int_0^t e^{-u}u^{n-1}\,du/(n-1)!$ be its distribution function. Let $\mathcal{P}_n$ be the distribution on $[0,1]$ with probability density nt^{n-1} and distribution function t^n.

Let us write $w(x) = \|x\|$ and $b(x) = x/\|x\|$. A point $x \in \mathbb{R}^n \setminus \{0\}$ can be represented uniquely by the polar coordinates (w, b), where $b \in \partial B$ and $w \in (0, \infty)$. Let ξ be the uniform probability measure[2] on ∂B. Then the volume element in $\mathbb{R}^n$ is proportional to $w^{n-1}dwd\xi$.

Let $h(b) = 1/f_K(b)$ be the distance from the origin to ∂K in the direction b. Note that $f_K(x) = w/h$. Then the uniform probability measure μ_1 on K has element

$$d\mu_1 = C_1 w^{n-1}dwd\xi \qquad (b \in \partial B, w \in (0, h]),$$

where clearly $C_1 = 1/\operatorname{Vol}(K)$. Similarly, the measure μ_2 determined by F_K has element

$$d\mu_2 = C_2 e^{-w/h}w^{n-1}dwd\xi \qquad (b \in \partial B, w \in (0, \infty)).$$

Integrating with respect to w, the marginal measures ξ_1, ξ_2 on ∂B induced by μ_1 and μ_2 have elements $(C_1/n)h^n d\xi$ and $C_2(n-1)!h^n d\xi$. Comparing, we see that $\xi_1 = \xi_2 = \xi'$ (say), and hence in particular $C_2 = 1/n!\operatorname{Vol}(K)$ (c.f. [19]).

[1] We will use the term *density* when the integral is not necessarily 1. If we wish to insist on integral 1, we will write *probability* density.

[2] Measure will always mean probability measure.

It follows further that, given a point from μ_2 we can generate a point from μ_1. (The reverse is also true, but we will not elaborate the details.) Note, from the expressions for $d\mu_1$ and $d\mu_2$, that the measures ν_1, ν_2 of $f(x) = w/h$ are $\mathcal{P}_n$ and Γ_n respectively. If z is a random point from ν_2, then $\gamma_n(z)^{1/n}$ is a point from ν_1.

Thus, if we have a point x_2 from μ_2, we can construct a point x_1 from μ_1 as follows. If $x_2 = 0$ then $x_1 = 0$. Otherwise, let $b = x_2/\|x_2\| \in \partial B$ be drawn from ξ'. Also, in obvious notation, $w_1 = h\gamma_n(w_2/h)^{1/n}$. Then $x_1 = w_1 b$. Let us write $x_1 = \chi(x_2)$. Then clearly χ is a bijection between $\mathbb{R}^n$ and $\operatorname{int} K$.

In practice, we can only sample approximately from F_K. We will take *variation distance* as the measure of distributional approximation. Let μ_1, μ_2 be two measures defined on the Borel sets in $\mathbb{R}^n$. (All measures we employ satisfy this assumption, and henceforward we presume the necessary measurability qualification on sets.) Then their variation distance is defined by

$$\mathrm{d_{TV}}(\mu_1, \mu_2) = \max_A \{\mu_1(A) - \mu_2(A)\},$$

where $A \subseteq \mathbb{R}^n$ is a Borel set. See, for example, Appendix A.1 of [3] for more information.

Variation distance is a metric on measures, but we will extend the notation to random variables. If X, Y are random variables, we will write $\mathrm{d_{TV}}(X, Y)$ for the variation distance between the measures $\mathbf{P}(X \in \cdot)$, $\mathbf{P}(Y \in \cdot)$ associated with X and Y,

Thus, let us denote the approximating measure to μ_2 by $\hat{\mu}_2$, and let $\hat{\mu}_1$ be the resulting approximation to μ_1. Then, since χ is a bijection, we have from the definition of total variation distance that

$$\mathrm{d_{TV}}(\mu_1, \hat{\mu}_1) = \mathrm{d_{TV}}(\mu_2, \hat{\mu}_2).$$

Thus the transformation preserves the quality of the approximation.

Note that, in view of this transformation, we can regard a random walk on $\mathbb{R}^n$ converging to the F_K distribution as a random walk on K converging to the uniform distribution. Thus our development below could be described equivalently in terms of K rather than F_K. For example, we define below a distance metric on $\mathbb{R}^n$ to gauge the convergence of the random walk for F_K. The bijection χ then induces a metric on K. We could equally well work with this as a measure for convergence of the random walk in K.

In practice, it is not necessary to determine $w_1 = h\gamma_n(w_2/h)^{1/n}$. We can merely generate an independent uniform $[0, 1]$ variable U and let $w_1 = U^{1/n}$. The additional generator $\hat{\nu}_1$, approximating ν_1, introduces further error. We have to check that the effect of this on the overall approximation is small. Letting $\tilde{\mu}_1$ be the resulting approximation to μ_1, this follows from

Lemma 1

$$\mathrm{d_{TV}}(\mu_1, \tilde{\mu}_1) \leq \mathrm{d_{TV}}(\mu_2, \hat{\mu}_2) + \mathrm{d_{TV}}(\nu_1, \hat{\nu}_1).$$

Proof: Let $\hat{\xi}'$ be the measure resulting from projecting $\hat{\mu}_2$ onto ∂B. Then it is clear from the definition of variation distance that

$$\mathrm{d_{TV}}(\xi', \hat{\xi}') \leq \mathrm{d_{TV}}(\mu_2, \hat{\mu}_2).$$

Let $\eta = \mathrm{d}_{\mathrm{TV}}(\nu_1, \hat{\nu}_1)$. Now, for any $A \subseteq \mathbb{R}^n$ and given $b \in \partial B$, let $A_b = A \cap (0,h)b$. We have $\mu_1(A) = \int_{\partial B} \nu_1(A_b) d\xi'$ and

$$\hat{\mu}_1(A) = \int_{\partial B} \hat{\nu}_1(A_b) d\hat{\xi}' \leq \int_{\partial B} (\nu_1(A_b) + \eta) d\hat{\xi}'.$$

Hence

$$\hat{\mu}_1(A) - \mu_1(A) \leq \int_{\partial B} \nu_1(A_b)(d\hat{\xi}' - d\xi') + \eta \leq \int_{B'} (d\hat{\xi}' - d\xi') + \eta,$$

where $B' = \{b \in \partial B : d\hat{\xi}' \geq d\xi'\}$. But

$$\int_{B'} (d\hat{\xi}' - d\xi') = \hat{\xi}'(B') - \xi'(B') \leq \mathrm{d}_{\mathrm{TV}}(\xi', \hat{\xi}'),$$

from which the Lemma follows. □

Thus, if we can generate approximately from F_K, we can sample almost uniformly from K.

2.3 Modifying F_K

We need to be able to sample from the log-concave distribution defined by F_K. It will be easier to show that we can do this for a slightly modified function, which differs from F_K only at "large" distances from the origin. We will first show that effect on the sampling distribution will be negligible.

Let $\rho = cnR$, for some constant $c \geq 2$. For convenience, we will assume without loss that $R \geq c^4$. Then, if X has our target distribution, we have using (1),

$$\mathbf{P}(\|X\| \geq \rho) = \mathbf{P}(\|X\|/R \geq cn) \leq \mathbf{P}(f_K(X) \geq cn)$$

But we know from section 2.2 that $f_K(X)$ has the Γ_n density. Therefore, using Stirling's approximation and Lemma 2 of section 2.7 below, we have

$$\mathbf{P}(\|X\| \geq \rho) \leq 2e^{-cn}(cn)^{n-1}/(n-1)! < (c/e^{c-1})^n. \tag{4}$$

Thus, by suitable choice of c, $\|X\| < \rho$ with very high probability.

In view of this, we may replace $f_K(x)$ outside ρB by a larger function without significantly altering the distribution determined by f_K for any polynomial-time sampling scheme. We will choose

$$f'_K(x) = \max\{f_K(x), 2(\|x\| - \rho)\}.$$

Note that this is convex. If $\|x\| \leq \rho$, then $f'_K(x) = f_K(x)$, and if $\|x\| \geq 2\rho$, we have $f'_K(x) = 2(\|x\| - \rho)$ using (1) . For our purposes below, the advantage of this function is that it behaves "essentially" like $\|x\|$ at large distances from the origin. Note that f'_K has (to within the required approximations) the same properties as f_K, except that (3) must be modified to $|\delta_v(x)| \leq 2$.

Henceforward we will simply write f for f'_K and $F = e^{-f}$, assuming only a condition of the form

$$|\delta_v(x)| \leq \lambda, \tag{5}$$

for some given constant $\lambda \geq 1$.

2.4 Metropolis random walks

The following is sufficient for our purpose, but can be generalised. (See, for example, [14, 19].) Consider a *lazy* (discrete time, continuous space) Markov random walk X_t on $\mathbb{R}^n$. Denote the (nonnegative) transition density by $p(x, x')$ where, for each $x \in \mathbb{R}^n$, we suppose $p(x, x') > 0$ in some neighbourhood of x and $\int_{\mathbb{R}^n} p(x, x')\, dx' \leq 1$. Then we assume $\mathbf{P}(X_{t+1} \in x' + dx' \mid X_t = x) = p(x, x')dx'$ for $x \neq x'$, and $\mathbf{P}(X_{t+1} = x \mid X_t = x) = 1 - \int_{\mathbb{R}^n} p(x, x')\, dx'$. (Here dx' is an infinitesimal neighbourhood of x', and we abuse notation by employing the same symbol for its volume.)

The random walk is *time-reversible* if there exists a probability density $\pi(x)$ on $\mathbb{R}^n$ such that the conditions

$$\pi(x)p(x, x') = \pi(x')p(x', x) \qquad (\text{for all } x, x' \in \mathbb{R}^n \text{ with } x \neq x'),$$

are satisfied. Then it follows that $\pi(x)$ is the asymptotic density of X_t. These conditions are called *detailed balance*.

The random walks in which we are interested arise in the following way. We have a function $q(x, x') : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, symmetric in x and x'. We further assume $q(x, \cdot)$ is a probability density for each x, positive in some neighbourhood of x. Suppose $F : \mathbb{R}^n \to \mathbb{R}$ is a log-concave function, integrable on $\mathbb{R}^n$, such that the derivative of $f = -\ln F$ is bounded in absolute value by λ in all directions at every point. Let

$$M(x) = e^{-\lambda\|x\|}, \text{ and } A(x, x') = \sqrt{M(x' - x)F(x')/F(x)}.$$

Clearly $0 \leq A(x, x') \leq 1$ everywhere, since $|f(x') - f(x)| \leq \lambda\|x' - x\|$ is a consequence of the derivative condition. Thus $A(x, x')$ is a probability. Also $A(x, x')/A(x', x) = F(x')/F(x)$, since $M(x' - x) = M(x - x')$. Now define a lazy random walk as above by $p(x, x') = A(x, x')q(x, x')$. Clearly this satisfies

$$F(x)p(x, x') = F(x')p(x', x) \qquad (x, x' \in \mathbb{R}^n),$$

and hence the asymptotic density of the random walk is $\pi(x) = F(x)/\int F(x)$. This is a modification of the usual Metropolis random walk (see [14]), but is convenient for our proofs. This random walk is easily implemented as follows. If $X_t = x$, we choose a trial point x' with density $q(x, \cdot)$. We accept this point with probability $A(x, x')$. If we accept the point, $X_{t+1} = x'$, otherwise $X_{t+1} = x$. We will let $\mathcal{A}(x, x')$ be the 0-1 indicator variable of the acceptance event, so $\mathbf{P}(\mathcal{A}(x, x') = 1) = A(x, x')$. We will denote the complementary quantities $(1 - A)$, $(1 - \mathcal{A})$ by $\bar{A}$, $\bar{\mathcal{A}}$ respectively.

2.5 The random walk

Given a log-concave function of the type discussed at the end of section 2.4, we define a random walk as in section 2.4. We choose $q(x, x')$ to be Gaussian. Let ϕ be the unit Normal density on $\mathbb{R}$, and G_n be the standard Normal density on $\mathbb{R}^n$, i.e.

$$\phi(x) = e^{-x^2/2}/\sqrt{2\pi} \text{ and } G_n(x) = \prod_{i=1}^{n} \phi(x_i).$$

Let U be a random variable with density G_n. We take $q(x, x')$ so that $x' = x + \sigma U$, for some constant σ, to be determined. Then the trial step is to x', and it is accepted with probability $A(x, x')$.

In fact, for convenience, we will modify this random walk slightly. Let $r = \sigma\sqrt{n}$. Then, for large n, r is the approximate step size for X_t. Now, with c as in (4), let us redefine the acceptance multiplier $M(x)$ to be

$$M(x) = \exp\left(-\lambda \max\left\{cr, \|x\|\right\}\right).$$

Now $\mathbf{P}(\|x' - x\| > cr) = \mathbf{P}(\|U\|^2 > c^2 n)$. But it is well known that $\frac{1}{2}\|U\|^2$ has the $\Gamma_{n/2}$ density. Hence, using Lemma 2 and Stirling's approximation,

$$\mathbf{P}(\|x' - x\| > cr) \leq (c^2/e^{c^2-1})^{n/2} \leq (c/e^{c-1})^n.$$

Thus, with a suitable choice of c, we may assume that with very high probability there is no step of size greater than cr during any polynomial number of steps of the random walk. Thus $M(x) = \exp(-\lambda cr)$, where henceforth we will assume without loss that $\lambda r \leq c^{-4}$. In section 4.2 below, we will describe a further modification to $M(x)$ which asymptotically improves the number of steps required, but for ease of exposition here we use the expression above.

2.6 Coupling

The proof technique we use below is *coupling*, and has been employed in similar contexts by Jerrum [12] and Matthews [20]. An overview of the use of couplings for Markov chains can be found, for example, in [1] and [5]. Details of the extension of the coupling method to more general processes can be found in [9], and [23].

To analyse the convergence rate of our random walk X_t to its stationary distribution, we consider a second random walk, Y_t. The "Coupling Lemma" (which is very easy to prove) then states that

$$\mathrm{d}_{\mathrm{TV}}(X_t, Y_t) \leq \mathbf{P}(X_t \neq Y_t).$$

This Lemma is usually attributed to Aldous and holds for any (not necessarily Markovian) processes X_t, Y_t. In order to apply it to convergence of Markov processes, we will require that marginally, X_t and Y_t are both faithful copies of the process and that Y_0 is chosen randomly from the stationary distribution (and hence Y_t has this property throughout).

The Coupling Lemma will give us an upper bound on the variation distance of the random walk from its stationary distribution. To get good bounds, however, we have to construct the joint process (X_t, Y_t) so that we quickly get $X_t = Y_t$. In order to achieve this here, we use two different couplings depending on $\|X_t - Y_t\|$. The first encourages the two random walks to come closer to each other when they are distant, and the second encourages the random walks to meet if they are sufficiently close.

We wish to have $\mathrm{d}_{\mathrm{TV}}(X_t, Y_t) \leq \varepsilon$ for arbitrary ε. However, it is sufficient to have τ, of size polynomial in n, such that $\mathbf{P}(X_\tau \neq Y_\tau) \leq \frac{1}{2}$. Then we may regard t as divided into

blocks of length τ, each constituting a coupling "trial" with "success" probability at least $\frac{1}{2}$. The blocks are independent. and hence if $k = \lceil \lg(1/\varepsilon) \rceil$,

$$\mathrm{d_{TV}}(X_{k\tau}, Y_{k\tau}) \leq \mathbf{P}(X_{k\tau} \neq Y_{k\tau}) \leq (\tfrac{1}{2})^k \leq \varepsilon.$$

Thus we can guarantee variation distance ε by increasing the execution time only by a factor $O(\log(1/\varepsilon))$. Therefore, we aim below simply to attain a probability $\frac{1}{2}$ of coupling.

2.7 Technical Results

We collect here some simple lemmas which are used in the analysis of section 3. All but the most tenacious reader may wish to skip this material, at least on first reading. We first give the proof of a result used in section 2.3.

Lemma 2 *Let Z have the Γ_{k+1} density, and let $\mu \geq 2k$, then*

$$e^{-\mu}\mu^k/k! \leq \mathbf{P}(Z \geq \mu) \leq 2e^{-\mu}\mu^k/k!$$

Proof: The left hand inequality is easily proved by proved by parts integration. Also, since $\mu \geq 2k$,

$$\begin{aligned} \mathbf{P}(Z \geq \mu) &= \frac{1}{k!}\int_\mu^\infty e^{-t}t^k dt \\ &\leq \frac{1}{k!}e^{-\mu/2}\mu^k \int_\mu^\infty e^{-t/2}dt \\ &= 2e^{-\mu}\mu^k/k! \end{aligned}$$

□

We now state a very simple bound on $\ln n$, which we use repeatedly without comment. It may be proved by elementary calculus.

Lemma 3 *If $\alpha > 0$, then $\ln n/n^\alpha \leq 1/(e\alpha)$.*

We next prove two simple approximation results.

Lemma 4 *Let $a \geq \frac{1}{4}$, and $b > 0$. Then $\sqrt{a+b} \leq \sqrt{a} + b$.*

Proof: This follows from $a + b \leq a + 2\sqrt{a}\,b \leq (\sqrt{a} + b)^2$, on taking square roots. We require only $2\sqrt{a} \geq 1$, i.e. $a \geq \frac{1}{4}$. □

Corollary 1 *Let $a \geq \frac{1}{4}$, and $b > 0$. Then $\sqrt[4]{a+b} \leq \sqrt[4]{a} + b$.*

Proof: Applying Lemma 4 twice

$$\sqrt[4]{a+b} \leq \sqrt{\sqrt{a} + b} \leq \sqrt[4]{a} + b,$$

where the first application is valid for $a \geq \frac{1}{4}$, and the second for $a \geq \frac{1}{16}$. □

Lemma 5 *If $|z| \leq \frac{1}{2}$, then $\sqrt{1-z} \leq 1 - \frac{1}{2}z - \frac{1}{10}z^2$.*

Proof: Since $1 - \frac{1}{2}z - \frac{1}{10}z^2$ is positive for $|z| \leq 1$, we may obtain an equivalent inequality by squaring both sides. Simplifying this gives $z^2 + 10z + 5 \geq 0$. This is satisfied for all $|z| \leq \frac{1}{2}$. $\square$

We now prove a simple bound on the tail of the Normal distribution.

Lemma 6 *Let $v \sim N(0, \sigma^2)$ and $\ell = c\sigma \ln n$, then $\mathbf{P}(|v| > \ell) < n^{-c^2 \ln n/2}$.*

Proof:

$$\begin{aligned} \mathbf{P}(|v| > \ell) &= \frac{2}{\sigma}\int_{\ell}^{\infty} \phi(t/\sigma)\, dt \leq \frac{2}{\sigma\ell}\int_{\ell}^{\infty} t\phi(t/\sigma)\, dt \\ &= \frac{2\sigma}{\ell}\phi(\ell/\sigma) < n^{-c^2 \ln n/2}. \end{aligned}$$

$\square$

Next we prove two easy results to deal with mild conditioning in the analysis.

Lemma 7 *Let Z be a real valued random variable, symmetrically distributed about the origin. For $t \geq 0$, let E_t be the event $|Z| \leq t$. Then if $f(z)$ is an*

(i) odd function, $\mathbf{E}\,(f(Z)|E_t) = 0$;

(ii) even convex function, $\mathbf{E}\,(f(Z)|E_t) \leq \mathbf{E}\,(f(Z))$.

Proof: The first statement follows by symmetry, the second by noting that $f(z_1) \leq f(z_2)$ for $|z_1| \leq |z_2|$. $\square$

Lemma 8 *Let $Z \sim N(0, \sigma^2)$, and for $t \geq 3\sigma$ let E_t be the event $|Z| \leq t$, then $\mathbf{E}\,(Z^2|E_t) \geq 0.97\sigma^2$.*

Proof: It clearly suffices to prove the Lemma for $\sigma = 1$. Then $\mathbf{E}\,(Z^2|E_t) = 1 - t\phi(t)/(\Phi(t) - \frac{1}{2})$. This expression is increasing for all $t > 0$, and putting $t = 3$ gives $(1 - 0.0133/0.4987) > 0.97$. $\square$

Finally we give a bound on a simple function.

Lemma 9 *For $z < \frac{1}{10}$, ze^{-z} is increasing and bounded above by $z - \frac{19}{20}z^2$.*

Proof: The first claim may be proved by elementary calculus. If $z \leq 0$, the second follows from $e^{|z|} \geq 1 + |z|$. If $z > 0$, note that the series for e^{-z} is alternating with deceasing terms for $z < 1$, and hence

$$e^{-z} < 1 - z + \tfrac{1}{2}z^2 = 1 - z(1 - \tfrac{1}{2}z) \leq 1 - \tfrac{19}{20}z,$$

from which the result follows. $\square$

3 Analysis of the random walk

We assume $n \geq 2$. We present the coupling analysis in three sections. In 3.1 we prove boundedness, in 3.2 we show approximate coupling, and in 3.3 we show that exact coupling will eventually occur.

3.1 Boundedness of the walk

It will be necessary for us to be sure that with high probability the random walks do not stray too far from the origin. For Y_t (started in the stationary distribution) this can be established easily by a calculation like that leading to (4). However for X_t (started at an arbitrary point) it is not so straightforward. To bound this we consider the induced random walk $\Xi_t = \|X_t\|$. We assume only that $\Xi_0 < 2\rho$, so in fact the analysis applies to Y_t also.

Let us call any set of consecutive values of t for which $\Xi_t > 2\rho$ an *excursion*. We wish to show that any excursion does not include values of $\Xi_t > 3\rho$. Suppose $X_t = x$ with $\Xi_t = \xi \geq 2\rho$, and let $s = \Xi_{t+1} - \Xi_t$. Then we know that $f(x) = 2(\xi - \rho)$, and hence we will assume $\lambda \geq 2$ for the remainder of this section. (The argument may be modified, with a weaker conclusion, for any positive λ.) Also, if $x' = x + \sigma U$ is the trial point as in section 2.5, then letting $\xi' = \|x'\|$ and $s' = \xi' - \xi$, we have $F(x')/F(x) = e^{-2s'}$.

Now write $u = x/\xi$, and let $v = \sigma U.u$ and $w = \sqrt{\sigma^2\|U\|^2 - v^2}$. Then

$$\xi' = \sqrt{(\xi + v)^2 + w^2} = \xi\sqrt{(1 + v/\xi)^2 + (w/\xi)^2}.$$

From Lemma 6 we may assume that $|v| \leq \ell = c\sigma \ln n \leq \frac{1}{21}$. Denote the event $|v| \leq \ell$ by $\mathcal{E}_1$, and the event $w < 2cr/3$ by $\mathcal{E}_2$. Let $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$. Then with high probability, $\mathcal{E}$ occurs at every step for polynomially many steps, provided c is large enough. Note that $\mathcal{E}$ implies $\sigma\|U\| \leq cr\sqrt{4/9 + (\ln n)^2/n} < cr$. Thus, if $\mathcal{E}$ occurs, the acceptance indicator $\mathcal{A}$ has

$$\mathbf{P}(\mathcal{A} = 1) = A = \sqrt{e^{-\lambda cr}e^{-2s'}} = e^{-(c'+s')},$$

where $c' = \lambda cr/2 \leq \frac{1}{16}$.

Assuming the same event, $v/\xi \geq -\ell/2cnR = -r\ln n/2Rn^{3/2} \geq -1/4000$. Hence we may apply Lemma 4 to give

$$\xi' \leq \xi(1 + v/\xi + (w/\xi)^2) = \xi + v + w^2/\xi \leq \xi + v + \epsilon,$$

for some constant ϵ, where $0 \leq \epsilon < 4(cr)^2/18cnR \leq \sigma^2/144$. Also it follows that $\xi' \geq \xi + v$, and hence

$$v \leq s' \leq v + \epsilon.$$

Since $\epsilon < \ell/1000$, we have $|s| \leq |s'| \leq \ell + \epsilon < 1.001\ell \leq \frac{1}{20}$. Now, using Lemmas 7, 8, and 9,

$$\mathbf{E}\,(s|\mathcal{E}) \;=\; \mathbf{E}\left(s'e^{-(c'+s')}|\mathcal{E}\right)$$

$$
\begin{aligned}
&\leq e^{-c'}\mathbf{E}\left((v+\epsilon)e^{-(v+\epsilon)}|\mathcal{E}_1\right)\\
&\leq e^{-c'}\mathbf{E}\left((v+\epsilon)-0.95(v+\epsilon)^2|\mathcal{E}_1\right)\\
&= e^{-c'}\mathbf{E}\left(\epsilon(1-0.95\epsilon)+(1-1.9\epsilon)v-0.95v^2|\mathcal{E}_1\right)\\
&\leq e^{-c'}(\epsilon-0.92\sigma^2)\\
&\leq -0.85\sigma^2,
\end{aligned}
$$

using $\epsilon \leq \sigma^2/144$ and $c' \leq 1/16$. Hence $|\Xi_{t+1}-\Xi_t| < 1.001\, c\sigma \ln n$ and $\mathbf{E}\,(\Xi_{t+1}-\Xi_t) \leq -0.85\sigma^2$. Thus, suppose the excursion starts at $t = a+1$, so $\Xi_a \leq 2\rho$. Let $k_0 = \lfloor c^4(\ln n)^4/2.1\sigma^2 \rfloor$. Then, since the process Ξ_t is Markovian, Hoeffding's inequality for Martingales [11] (see also [21]) gives,

$$
\mathbf{P}(\exists k, 1 \leq k \leq k_0 : \Xi_{a+k} - \Xi_a + 0.85\sigma^2 k > 0.4c^4(\ln n)^4) \leq n^{-c^2 \ln n/6}. \tag{6}
$$

However, (6) implies

$$
\mathbf{P}(\Xi_{a+k_0} - \Xi_a > 0) \leq n^{-c^2 \ln n/6}.
$$

Thus, with high probability, the excursion must have already ended at step $(a+k_0)$. For $k < k_0$, (6) implies

$$
\mathbf{P}(\exists k : \Xi_{a+k} > 2\rho + 0.4c^4(\ln n)^4) \leq n^{-c^2 \ln n/6}.
$$

But $0.4c^4(\ln n)^4 < 2c^4 n \leq 2nR \leq \rho$. Thus each excursion lies entirely in the ball $3\rho B$, except with probability $n^{-c^2 \ln n/3}$. Since the algorithm can only perform a polynomial number of excursions, we can choose c so that $3\rho B$ includes the whole walk with high probability.

3.2 Bringing the random walks close

We have two random walks X_t, Y_t as in section 2.5, with Y_0 chosen randomly from the equilibrium density π. We will couple them in the following way. Suppose σU is the trial step. Let $u_1, u_2, \ldots, u_n$ be an orthonormal basis for $\mathbb{R}^n$ such that $u_1 = (Y_t - X_t)/\|Y_t - X_t\|$. Then $U = \sum_{i=1}^n U_i u_i$, where U_i has the density G_1. Let us write $U' = \sum_{i=2}^n U_i u_i$, so U' has the density G_{n-1}. Then the trial points are generated by

$$
X_t' = X_t + U_1 u_1 + U', \qquad Y_t' = Y_t - U_1 u_1 + U'.
$$

We accept the trial points *independently* with probabilities $A(X_t, X_t')$, $A(Y_t, Y_t')$ respectively. If we accept X_t', then $X_{t+1} = X_t'$, otherwise $X_{t+1} = X_t$ and similarly for Y_t'. Thus the two random walks are coupled so that, if the two steps are accepted, their steps will be reflections of each other in the hyperplane of points equidistant from X_t, and Y_t. This has similarities to the coupling given in [17].

We will show that, with this coupling, X_t, Y_t have a tendency to converge. Specifically, we show that, for some metric d on $\mathbb{R}^n$, $\mathbf{E}\,(d(X_t, Y_t))$ decreases. The metric we choose is the *square-root* of Euclidean distance, $d(X_t, Y_t) = \sqrt{\|X_t - Y_t\|}$. The rationale for this

metric is to downweight larger values of the "neutral" Euclidean distance. We will write $d_t = d(X_t, Y_t)$, and let D denote $d_t^2 = \|Y_t - X_t\|$.

In notation similar to that in 3.1, let us write $v = U_1$, and $w = \|U'\|$. Also, abbreviating the indicator variables for acceptance to $\mathcal{A}_X$, $\mathcal{A}_Y$, let

$$\begin{aligned} I &= \mathcal{A}_X + \mathcal{A}_Y = 1 + \mathcal{A}_X\mathcal{A}_Y - \bar{\mathcal{A}}_X\bar{\mathcal{A}}_Y \\ J &= |\mathcal{A}_X - \mathcal{A}_Y| = |\bar{\mathcal{A}}_X - \bar{\mathcal{A}}_Y| \end{aligned}$$

where $I \in \{0, 1, 2\}$ and $J \in \{0, 1\}$. Then it follows that

$$d_{t+1} = \sqrt[4]{(d_t^2 - vI)^2 + w^2 J}.$$

We will bound $\mathbf{E}\left(d_{t+1}|d_t\right)$ as a function of d_t.

As in section 3.1, denote the event $|v| \le c\sigma \ln n$ by $\mathcal{E}_1$, the event $w < 2cr/3$ by $\mathcal{E}_2$, and $\mathcal{E}_1 \cap \mathcal{E}_2$ by $\mathcal{E}$. Then, with high probability, $\mathcal{E}$ occurs at every step for sufficiently many steps, and no step size is larger than cr. Now $\mathcal{E}$ implies

$$D - vI \ge D - 2c\sigma \ln n \ge \tfrac{1}{2}D,$$

provided we do not have the event $d_t < 2\sqrt{c\sigma \ln n}$. Let us call this event $\mathcal{C}$. We will say the walks are "close" when $\mathcal{C}$ occurs. Now if $\mathcal{C}$ does not occur, using Corollary 1 and Lemma 5,

$$\begin{aligned} d_{t+1}/d_t &= \sqrt[4]{(1 - vI/D)^2 + w^2 J/D^2}. \\ &\le \sqrt{1 - vI/D} + w^2 J/D^2 \\ &\le 1 - \tfrac{1}{2}vI/D - v^2 I^2/10D^2 + w^2 J/D^2 \end{aligned} \tag{7}$$

We have to bound the terms on the right side of (7) conditioned by $\mathcal{E}$.

Note that if $\mathcal{E}$ occurs then, for A_X, A_Y,

$$e^{-\lambda cr} \le A \le 1, \qquad 0 \le \bar{A} \le 1 - e^{-c\lambda r} \le \lambda cr.$$

Thus, using Lemmas 7 and 8,

$$\begin{aligned} \mathbf{E}\left(w^2 J|\mathcal{E}\right) &\le \mathbf{E}\left(w^2(\bar{A}_X + \bar{A}_Y)|\mathcal{E}\right) \\ &\le 2\lambda cr\mathbf{E}\left(w^2|\mathcal{E}_2\right) \le 2\lambda cr\mathbf{E}\left(w^2\right) < 2\lambda cr^3, \end{aligned} \tag{8}$$

and, providing $c \ln n \ge 3$,

$$\begin{aligned} \mathbf{E}\left(v^2 I^2|\mathcal{E}\right) &= \mathbf{E}\left(v^2(A_X + A_Y + 2A_X A_Y)|\mathcal{E}\right) \\ &\ge 2e^{-c\lambda r}(1 + e^{-c\lambda r})\mathbf{E}\left(v^2|\mathcal{E}_1\right) > 3r^2/n, \end{aligned} \tag{9}$$

since $c\lambda r < \frac{1}{8}$ and $\mathbf{E}\left(v^2|\mathcal{E}_1\right) > 0.97\sigma^2$.

It remains to consider $\mathbf{E}\left(vI\right)$. Now, since $\mathbf{E}\left(v|\mathcal{E}\right) = \mathbf{E}\left(v|\mathcal{E}_1\right)$, using Lemma 7 we have

$$\mathbf{E}\left(vI|\mathcal{E}\right) = \mathbf{E}\left(v + vA_X A_Y - v\bar{A}_X\bar{A}_Y|\mathcal{E}\right) = \mathbf{E}\left(vA_X A_Y|\mathcal{E}\right) - \mathbf{E}\left(v\bar{A}_X\bar{A}_Y|\mathcal{E}\right),$$

We will bound these two contributions separately. We do this by arguing conditionally on both $\mathcal{E}_1$ and the value of w (given that this satisfies $\mathcal{E}_2$). Since the bounds we obtain are independent of w, we may infer the same bound for conditioning on $\mathcal{E}$. In this respect, for either A_X or A_Y, the notation $A(v)$ will mean "A considered as a function of v for a given fixed w".

Thus let $p(v)$ be the conditional density of v given $\mathcal{E}$, and first consider the second term. Then

$$\begin{aligned}
\mathbf{E}\left(v\bar{A}_X\bar{A}_Y|\mathcal{E}_1, w\right) &= \int_{-\infty}^{\infty} v\bar{A}_X(v)\bar{A}_Y(v)p(v)dv \\
&= \int_0^{\infty} v(\bar{A}_X(v)\bar{A}_Y(v) - \bar{A}_X(-v)\bar{A}_Y(-v))p(v)dv \\
&= 2\int_0^{\infty} \frac{\bar{A}_X(v)\bar{A}_Y(v) - \bar{A}_X(-v)\bar{A}_Y(-v)}{2v}\, v^2 p(v)dv \\
&= 2\int_0^{\infty} \left[\frac{d(\bar{A}_X(v)\bar{A}_Y(v))}{dv}\right]_\theta v^2 p(v)dv,
\end{aligned}$$

where $-v \leq \theta \leq v$. Note that if $\bar{A}_X(v)\bar{A}_Y(v)$ is not differentiable everywhere we may consider a suitably close approximant. Now

$$\frac{d(\bar{A}_X(v)\bar{A}_Y(v))}{dv} = -\left(\bar{A}_X(v)A_Y(v)\frac{d(\ln A_Y(v))}{dv} + \bar{A}_Y(v)A_X(v)\frac{d(\ln A_X(v))}{dv}\right),$$

and hence, using the bound on the directional derivative of f,

$$\left|\frac{d(\bar{A}_X(v)\bar{A}_Y(v))}{dv}\right| \leq 2\lambda^2 cr.$$

Thus, using Lemma 7,

$$\begin{aligned}
\mathbf{E}\left(v\bar{A}_X\bar{A}_Y|\mathcal{E}_1, w\right) &\leq 4\lambda^2 cr\int_0^{\infty} v^2 p(v)dv \\
&= 2\lambda^2 cr\mathbf{E}\left(v^2|\mathcal{E}_1\right) \\
&\leq 2\lambda^2 cr\sigma^2 = 2\lambda^2 cr^3/n. \qquad (10)
\end{aligned}$$

Now let $g(v) = \sqrt{e^{-\lambda cr}F(X_t + U' + vu_1)}$. Then

$$g(D-v) = \sqrt{e^{-\lambda cr}F(X_t + U' + Du_1 - vu_1)} = \sqrt{e^{-\lambda cr}F(Y_t + U' - vu_1)}.$$

Note that $g(v)$ is a log-concave function (of one argument), and

$$A_X(v) = \frac{g(v)}{\sqrt{F(X_t)}}, \qquad A_Y(v) = \frac{g(D-v)}{\sqrt{F(Y_t)}}.$$

Therefore, writing $C = (F(X_t)F(Y_t))^{-1/2}$, we have

$$\begin{aligned}
\mathbf{E}\left(vA_XA_Y|\mathcal{E}_1, w\right) &= \int_{-\infty}^{\infty} vA_X(v)A_Y(v)p(v)dv \\
&= C\int_{-\infty}^{\infty} vg(v)g(D-v)p(v)dv \\
&= C\int_{0}^{\infty} (g(v)g(D-v) - g(-v)g(D+v))vp(v)dv, \\
&\geq 0, \qquad (11)
\end{aligned}$$

since the integrand is non-negative by the log-concavity of $g(v)$. To show this, let $t(v) = -\ln g(v)$, so $t(v)$ is convex. We have now to show that $t(v) - t(-v) \leq t(D+v) - t(D-v)$. But this is clear since

$$t(v) - t(-v) = \int_{-v}^{v} t'(v)dv, \qquad t(D+v) - t(D-v) = \int_{D-v}^{D+v} t'(v)dv,$$

for some nondecreasing function $t'(v)$.

We note in passing that it is only to establish (11) that we require log-concavity of F. All our other estimates simply require suitable "smoothness" of F. On the other hand, for something like (11) to be true, it appears we need a property close to log-concavity. It is also the proof of (11) which necessitates our departure from the "usual" Metropolis process, since we have to factor out the different contributions from the points X_t, Y_t.

From (10) and (11) we now have

$$\mathbf{E}\left(vI|\mathcal{E}\right) \geq -2\lambda^2 cr^3/n. \qquad (12)$$

Therefore, putting (8), (9) and (12) into (7), we have

$$\begin{aligned}
\mathbf{E}\left(d_{t+1}/d_t|d_t\right) &\leq 1 + \frac{\lambda^2 cr^3}{nD} - \frac{3r^2}{10nD^2} + \frac{2\lambda cr^3}{D^2} \\
&= 1 + \frac{r^2}{10nD^2}\left(3 - 10\lambda^2 crD - 5\lambda crn\right) \\
&\leq 1 - \frac{r^2}{10nD^2},
\end{aligned}$$

provided $10\lambda^2 crD + 5\lambda crn \leq 2$. This will be true if $20.1\lambda^2 cr\rho \leq 1$, using $R \geq c^4$ and the fact that, with high probablity, $D < 4\rho$ from sections 2.3 and 3.1. Thus let us take

$$r = \frac{1}{20.1c\lambda^2\rho} = \frac{1}{20.1c^2\lambda^2 nR}. \qquad (13)$$

Thus if $R = O(n)$, $r = \Omega(1/n^2)$. Now, using $D < 4\rho$,

$$\mathbf{E}\left(d_{t+1}\right) \leq \mathbf{E}\left(d_t\right)\left(1 - \frac{1}{65000nc^2\lambda^4\rho^4}\right) = \mathbf{E}\left(d_t\right)\left(1 - \frac{1}{65000n^5c^6\lambda^4R^4}\right). \qquad (14)$$

The reduction factor is $(1-\Omega(1/n^9))$, assuming $R=O(n)$. Therefore, since $d_0=O(n)$ and $\sigma=\Omega(n^{-5/2})$, in $O(n^9\log n)$ steps we must either have encountered $\mathcal{C}$ or $\mathbf{E}(d_t)\le \frac{1}{10}\sqrt{c\sigma\ln n}$. In the latter case, by Markov's inequality, $\mathbf{P}(\bar{\mathcal{C}})\le\frac{1}{20}$. Hence, in a sequence of $O(n^9\log n)$ steps, there is probability at least $\frac{19}{20}$ that $\mathcal{C}$ holds at some step.

3.3 Making the random walks meet

When $\mathcal{C}$ occurs, we switch to a different coupling. Note that X_t, Y_t are still relatively far apart compared to the expected movement in any direction in a single step. Therefore we choose instead to correlate a *sequence* of steps in order to have sufficient probability that $X_t=Y_t$ at the end of the sequence.

Thus suppose $\mathcal{C}$ occurs at step τ, but $\bar{\mathcal{C}}$ held at $(\tau-1)$. Let $s_{X,t}=X_t'-X_t$, $s_{Y,t}=Y_t'-Y_t$ denote the trial steps in the two walks. Let $k=\lceil(4c\ln n)^2\rceil$, and

$$S_X=\sum_{t=\tau}^{\tau+k-1}s_{X,t},\qquad Q_X=X_\tau+S_X, \tag{15}$$

and similarly S_Y, Q_Y. Note that $X_{\tau+k}=Q_X$ and $Y_{\tau+k}=Q_Y$ if all trial steps are accepted for $t=\tau,\tau+1,\ldots,\tau+k-1$.

Now S_X,S_Y have density $G_n(x/\sigma\sqrt{k})$. We will couple S_X,S_Y. Then $s_{X,t},s_{Y,t}$ are generated "independently" with density $\phi(x/\sigma)$, subject to the sum condition (15).

As in section 3.2, let $u_1,u_2,\ldots,u_n$ be a basis such that $u_1=(Y_\tau-X_\tau)/D$, where $D=\|Y_\tau-X_\tau\|$. Suppose $S_X=(v,v_2,\ldots,v_n)$ in this coordinate system, where $v,v_2,\ldots,v_n$ are independent with probability density $\psi(x)$ proportional to $\phi(x/\sigma\sqrt{k})$. We now take $S_Y=(-v',v_2,\ldots,v_n)$, where v' will be defined in terms of v. Note that we have $Q_X=Q_Y$ if and only if $v'=D-v$. Thus, for given v, define v' by

$$v'=\begin{cases} D-v, & \text{with probability } \min\{1,\psi(D-v)/\psi(v)\},\\ v, & \text{otherwise.}\end{cases}$$

Intuitively, we are reflecting the tail of the density $\psi(v)$ about the point $v=\frac{1}{2}D$. Then v' has probability density function

$$\psi(D-v')\min\{1,\psi(v')/\psi(D-v')\}+\psi(v')(1-\min\{1,\psi(D-v')/\psi(v')\})=\psi(v').$$

Hence, since $\psi(v')=\psi(-v')$, $-v'$ has density ψ, as required. Also

$$\begin{aligned}\mathbf{P}(Q_X=Q_Y)=\mathbf{P}(v'=D-v) &= \int_{-\infty}^{\infty}\min\{\psi(D-v),\psi(v)\}dv\\ &= 2\int_{-\infty}^{-D/2}\psi(v)dv\\ &= 2\Phi\left(\frac{-D}{2\sigma\sqrt{k}}\right)\\ &\ge 2\Phi(-\tfrac{1}{2})>0.616.\end{aligned}$$

We must now consider the probability that all steps accept. To maximise this, we choose to make the random variables $\mathcal{A}_X$, $\mathcal{A}_Y$ dependent at each step. Their values are generated as follows. Let w be a random variable, independent for each step, with distribution uniform on $[0,1]$. Then $\mathcal{A}_X = 1$ if $w \geq A_X$, otherwise $\mathcal{A}_X = 0$. The value of $\mathcal{A}_Y$ is chosen similarly, using the same w. The advantage of this is that now

$$\mathbf{P}(\mathcal{A}_X = \mathcal{A}_Y = 1) = \min\{A_X,\ A_Y\} \geq e^{-c\lambda r}.$$

Now if $\mathcal{G}$ is the event that all trial steps are accepted in both walks for $t = \tau, \tau+1, \ldots, \tau+k-1$, we have

$$\mathbf{P}(\mathcal{G}) \geq (e^{-c\lambda r})^k \geq e^{-1/18} > 0.945,$$

using the value of r given in (13). Hence we have

$$\mathbf{P}(X_{\tau+k} = Y_{\tau+k}) = \mathbf{P}(Q_X = Q_Y)\mathbf{P}(\mathcal{G}) > 0.616 \times 0.945 > 0.582.$$

The number of steps, $O((\log n)^2)$, is asymptotically negligible compared with the number required in section 3.2 to achieve $\mathbf{P}(\mathcal{C}) > \frac{19}{20}$. Hence, after $O(n^5 R^4 \log n)$ steps, the probability that the walks have failed to couple is at most $\frac{1}{20} + \frac{19}{20}(1 - 0.582) < 0.45$.

The other possible sources of failure to couple, for example that some step is too large, can be made very small by appropriate choice of c. We may arrange that their combined probability is less than 0.05 for all n. Hence, the probability of coupling will be at least $\frac{1}{2}$, as required.

This completes the analysis of the algorithm. Putting $R = O(n)$, our "mixing time" is $O^*(n^9)$, where $O^*(\cdot)$ is the notation which hides factors of $\log n$. This is larger than is known to be possible [15]. Note that the mixing time is $O^*(n\rho^2/r^2) = O^*((\rho/\sigma)^2)$, By comparison with one-dimensional random walk, this is of the optimal form. The problem is that the step size, r, is too small. We attempt to offset this difficulty in the next section.

4 Improvements

In this section we describe two refinements which reduce the required number of steps of the random walk from $O^*(n^9)$ to $O^*(n^{6.5})$.

4.1 A faster simulation of the random walk

An idea similar to that used to couple the walks in section 3.3 can be employed to simulate the random walk with fewer actual steps. A difficulty with our walk is that $r = \Theta(1/\rho)$ is smaller than values known to be achievable [19]. However, the fact that our trial steps are so small means that almost all will be accept, and hence we may simulate many steps simultaneously in a larger "step".

Suppose at step t, we attempt to perform k steps, where k is a power of 2. We will call this a *k-step*. Denote the current k-step by $\mathcal{S}$. Let S be the *vector sum* of the k trial steps

in $\mathcal{S}$, as in section 3.3. We generate the components of S as independent $N(0, k\sigma^2)$. If a is the probability that all these trial steps are accepted, then we have

$$a = \sqrt{e^{-k\lambda cr} F(X_t + S)/F(X_t)},$$

since all other terms cancel. Thus we accept $\mathcal{S}$ with probability a. If we accept, then we update t to $t + k$ and consider the next k-step.

If however, with probability $(1 - a) \leq k\lambda cr$, we do not accept $\mathcal{S}$, we know that at least one constituent step was rejected. We will say $\mathcal{S}$ *rejects*. We subdivide $\mathcal{S}$ into two $\frac{1}{2}k$-steps $\mathcal{S}_1$, $\mathcal{S}_2$ (in that order). We generate the corresponding vector sums S_1, S_2 subject to the condition $S_1 + S_2 = S$. Suppose s_1 is the ith component of S_1, and s the ith component of S. Then s_1 has conditional density

$$\frac{(e^{-{s_1}^2/k\sigma^2}/\sqrt{\pi k\sigma^2})(e^{-(s-s_1)^2/k\sigma^2}/\sqrt{\pi k\sigma^2})}{e^{-s^2/2k\sigma^2}/\sqrt{2\pi k\sigma^2}} = \frac{1}{\sqrt{\frac{1}{2}\pi k\sigma^2}} e^{-2(s_1 - s/2)^2/k\sigma^2}.$$

Thus $s_1 \sim N(\frac{1}{2}s, \frac{1}{4}k\sigma^2)$. Hence we can generate S_1, S_2. Now let

$$a_1 = \sqrt{e^{-k\lambda cr/2} F(X_t + S_1)/F(X_t)}$$

be the unconditioned acceptance probability for $\mathcal{S}_1$. We must condition on the event that $\mathcal{S}$ rejects. Let q be the conditional probability that $\mathcal{S}_1$ does not reject. Then clearly $q = (a_1 - a)/(1 - a)$. We use q to decide (by random generation) if $\mathcal{S}_1$ rejects. If not, we can set $X_{t+k/2} = X_t + S_1$ and consider $\mathcal{S}_2$ (with vector sum S_2), conditional that it rejects. The base case of the recursion, $k = 1$, is clearly a (single) step.

If $\mathcal{S}_1$ rejects, then we (recursively) halve it conditional on this event. Eventually, using this recursion, we compute $X_{t+k/2}$. Then we consider $\mathcal{S}_2$, with vector sum S_2, now as an unconditioned $\frac{1}{2}k$-step. Hence we ultimately compute X_{t+k}.

To analyse this procedure, let $t_u(k)$ denote the expected total number of k'-step s which are accepted within $\mathcal{S}$, for all $k' \leq k$. This effectively determines the execution time for $\mathcal{S}$. Let $t_c(k)$ be this number conditional that $\mathcal{S}$ rejects. For notational convenience, let $\beta = \lambda cr$. Then, from the above,

$$\begin{aligned} t_u(k) &\leq a + (1 - a)t_c(k) \leq 1 + k\beta t_c(k) \\ t_c(k) &\leq q(1 + t_c(\tfrac{1}{2}k)) + (1 - q)(t_c(\tfrac{1}{2}k) + t_u(\tfrac{1}{2}k)) \leq 1 + t_c(\tfrac{1}{2}k) + t_u(\tfrac{1}{2}k). \end{aligned}$$

Hence we have

$$t_c(k) \leq 2 + (1 + k\beta)t_c(\tfrac{1}{2}k) \leq 2 + e^{k\beta} t_c(\tfrac{1}{2}k). \tag{16}$$

Now (16) has solution $t_c(k) \leq 2e^{2k\beta} \lg k$, as is easily shown by induction. (Note $t_c(1) = 0$.) Hence

$$t_u(k) \leq 1 + 2e^{2k\beta} k\beta \lg k.$$

Let $k = 2^{\lfloor \lg(1/\beta) \rfloor}$. Note that $k \leq 1/\beta$ and $k = \Omega(1/r)$. Then

$$t_u(k) \leq 1 + 2e^2 \lg(1/\beta) = O(\log(1/r)).$$

Thus we trade a factor $1/r$ in the number of steps for $\log(1/r)$. This reduces the number from $O^*(n^9)$ to $O^*(n^7)$.

4.2 An even faster simulation

We use the factor $M = e^{-\lambda cr}$ in the calculation of A to ensure that the result is at most 1. However this factor is smaller than necessary. From the subgradient inequality (2), we have

$$f(X_t') \geq f(X_t) + \nabla.(X_t' - X_t),$$

where ∇ is a subgradient at X_t. Now $\nabla.(X_t' - X_t) \sim N(0, \sigma^2\|\nabla\|^2)$ and $\|\nabla\| \leq \lambda$. Thus, using Lemma 6,

$$\mathbf{P}(\nabla.(X_t' - X_t) > c\lambda r \ln n/\sqrt{n}) < n^{-c^2 \ln n/2}.$$

Hence $F(X_t')/F(X_t) \leq e^{-c\lambda r \ln n/\sqrt{n}} = M'$, say, with high probability. Let us write $A' = \sqrt{M'F(X_t')/F(X_t)}$, $A_0 = \sqrt{M/M'}$, and note that $A_0 > \frac{3}{4}$ is a constant. Then the acceptance probability $A = A_0 A'$. We write $\mathcal{A} = \mathcal{A}_0\mathcal{A}'$ for the corresponding indicator variables. Hence we may simulate a step of the original algorithm by first choosing $\mathcal{A}_0$ and generating $\mathcal{A}'$ only if $\mathcal{A}_0 = 1$.

Now consider the algorithm which uses only $\mathcal{A}'$. To each step of the modified algorithm there will be a geometrically distributed number of steps of the original algorithm with expectation $1/A_0$. However, observe that it does not matter if we stop at a random time in the original algorithm provided that it is independent of the random walk and it exceeds the bound on the coupling time (at least with high probability).

Thus we need not generate the sum of geometrics if the number of steps of the modified algorithm exceeds the coupling time. Note that the saving in the number steps from generating this sum (as a negative binomial) would be only a small constant factor, and we will not consider it here.

Now consider a k-step of the modified algorithm. We use the same notation as in section 4.1. We have, with high probability, $\|S\| \leq cr\sqrt{k}$, and hence $f(X_t + S) - f(X_t) \leq \lambda cr\sqrt{k}$. Thus

$$a \geq \exp(-\tfrac{1}{2}\lambda cr(k \ln n/\sqrt{n} + \sqrt{k})) = \exp(-\tfrac{1}{2}\beta(k \ln n/\sqrt{n} + \sqrt{k})),$$

and hence $(1 - a) \leq \beta(k \ln n/\sqrt{n} + \sqrt{k})$. With this modification (16) becomes

$$t_c(k) \leq 2 + e^{\beta(k \ln n/\sqrt{n}+\sqrt{k})/2} t_c(\tfrac{1}{2}k),$$

from which it follows that

$$t_u(k) \leq 1 + e^{2\beta(k \ln n/\sqrt{n}+2\sqrt{k})}\beta(k \ln n/\sqrt{n} + \sqrt{k}) \lg k.$$

Hence we may choose $k = 2^{\lfloor \lg(\sqrt{n}/\beta \ln n)\rfloor}$, and simulate $\Omega(\sqrt{n}/r)$ steps of the modified algorithm in $O(\log(\sqrt{n}/r))$ k-steps. This reduces the required number of "steps" to $O^*(n^{6.5})$.

5 Conclusions

We have given a coupling proof for the polynomial-time convergence of a random walk for generating an approximately uniformly distributed point in a convex body. The argument

is completely elementary, and does not require any auxiliary proof of an isoperimetric inequality, as with the usual approach. On the other hand, the running time of the algorithm is inferior to the best conductance methods. We obtain mixing time $O^*(n^{6.5})$, whereas Kannan, Lovász and Simonovits [15] obtain $O^*(n^5)$ (and an amortized time for multiple points which is even smaller). However, the present conductance techniques are the product of years of refinement. We have drawn on some of these developments, but our proof is a first attempt using a different approach. It would be surprising if it cannot be improved further. The present difficulty is the small step size r. We have described modifications to mitigate this problem, but genuine progress would result from an improvement in the arguments to allow r to be increased.

References

[1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilitès XVII 1981/1982*, volume 986 of *Springer-Verlag Lecture Notes in Mathematics*, pages 243–297, 1981.

[2] D. Applegate and R. Kannan. Sampling and integration of near log-concave functions. In *Proceedings of the Twenty Third Annual ACM Symposium on Theory of Computing*, pages 156–163, New Orleans, Louisiana, 1991.

[3] A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford Science, 1992.

[4] A. Z. Broder. How hard is it to marry at random ? In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 50–58, Berkeley, California, 1986.

[5] P. Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *Lecture Notes-Monograph Series*, chapter 4e, pages 84–87. Institute of Mathematical Statistics, Hayward, California, 1988.

[6] M. E. Dyer and A. M. Frieze. Computing the volume of convex bodies: A case where randomness provably helps. In *Symposia in Applied Mathematics*, volume 44, pages 123–169. American Mathematical Society, 1991.

[7] M. E. Dyer, A. M. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38:1–17, 1991.

[8] A. M. Frieze, R. Kannan, and N. Polson. Sampling from log-concave distributions. *Annals of Applied Probability*, to appear.

[9] D. Griffeath. Coupling methods for Markov processes. In Gian-Carlo Rota, editor, *Studies in Probability and Ergodic Theory*, volume 2 of *Advances in Mathematics Supplementary Studies*, pages 1–43. Academic Press, 1978.

[10] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, New York, 1988.

[11] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[12] M. R. Jerrum. A very simple algorithm for estimating the number of k-colourings of a low-degree graph. *Random Structures and Algorithms*, 7:157–165, 1995.

[13] M. R. Jerrum and A. J. Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18:1149–1178, 1989.

[14] R. Kannan. Markov chains and polynomial time algorithms. In *35th Annual IEEE Symposium on Foundations of Computer Science*, pages 656–671, Santa Fe, New Mexico, 1994. IEEE.

[15] R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. Preprint, January 1996.

[16] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric inequalities for convex sets. *Discrete and Computational Geometry*, to appear.

[17] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Annals of Probablilty*, 14:860–872, 1986.

[18] L. Lovász and M. Simonovits. Mixing rate of markov chains, an isoperimetric inequality and computing the volume. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 346–355, St. Louis, Missouri, 1990.

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

[20] P. Matthews. Generating a random linear extension of a partial order. *Annals of Probability*, 19:1367–1392, 1991.

[21] C. McDiarmid. On the method of bounded differences. In J. Siemons, editor, *Surveys in Combinatorics*, LMS Lecture Note Series 141, pages 148–188. Cambridge University Press, 1989.

[22] A. J. Sinclair and M. R. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82:93–133, 1989.

[23] H. Thorisson. Coupling methods in probability theory. *Scandinavian Journal of Statistics*, 22:159–182, 1995.