Sampling from convex bodies in poly-time: random walks in $\mathbb{R}^n$

Stephen Tu

Abstract

This note is meant to highlight some of the key ideas in establishing convergence results for random walks on convex bodies in $\mathbb{R}^n$. We describe two simple geometric random walk algorithms: ball walk, and hit and run. For simplicity, we focus the convergence analysis on the ball walk algorithm. This note contains a small subset of the results in the survey paper [Vem05].

Notation and setup. In what follows, let μ denote the Lebesgue measure (dimension will be clear from context) and let $\|\cdot\|$ denote the standard l_2 norm. Let $B_n(x, \delta)$ denote the n-dimensional Euclidean ball with radius δ: $B_n(x, \delta) = \{z \in \mathbb{R}^n : \|z - x\| \le \delta\}$. Furthermore, let S_n denote the boundary of $B_n(0, 1)$, e.g. $S_n = \{z \in \mathbb{R}^n : \|z\| = 1\}$. Finally, given two point $u, v \in \mathbb{R}^n$ and a subset $K \subset \mathbb{R}^n$, let $l_K(u, v)$ denote $l_K(u, v) = \{\lambda \in \mathbb{R} : u + \lambda v \in K\}$.

Goal. We assume we are given a full convex body $K \subset \mathbb{R}^n$, and a membership oracle which we can query with x to learn if $x \in K$. Let π_K denote the uniform measure on K: our goal is to produce random samples from π_K.

1 Random walk algorithms

For the following algorithms, let x_0 denote some starting point, and define the algorithm on time $t \ge 1$.

Ball walk. Fix a $\delta > 0$. At each time step t do:

(i) Pick a point y uniformly from $B_n(x_t, \delta)$.

(ii) If $y \in K$, set $x_{t+1} = y$.

Hit and run. At each step t do:

(i) Pick a point v uniformly from S_n.

(ii) Pick a uniformly random point λ from $l_K(x_t, v)$.

(iii) Set $x_{t+1} = x_t + \lambda v$.

2 Preliminaries

Unfortunately, we need a bit of notation and formalism to talk about continuous state space Markov chains. A discrete-time Markov chain $(K, \mathcal{B}, \{P_u : u \in K\})$ consists of a state space $K \subset \mathbb{R}^n$, a σ-field $\mathcal{B}$ of subsets of K, and a family of probability measures (often called a transition or stochastic kernel) $\{P_u : u \in K\}$ on $(K, \mathcal{B})$.

Definition 1. *(Stationary distribution of a Markov chain) A distribution Q is called* stationary *for a Markov chain if the following holds for all $A \in \mathcal{B}$.*

$$Q(A) = \int_K P_u(A)Q(du) \tag{1}$$

For the purposes of our discussion, we will assume the stationary distribution Q is unique for a Markov chain, and that our Markov chains converge to the stationary distribution in the limit. The following definition helps to establish a stationary distribution.

Definition 2. *(Time reversible) A Markov chain is called* time reversible *w.r.t. the distribution Q, if the following holds for all $A, B \in \mathcal{B}$*

$$\int_A P_u(B)Q(du) = \int_B P_u(A)Q(du)$$

This definition is useful because of the following fact.

Proposition 1. *If a chain is time reversible w.r.t. to a distribution Q, then Q is stationary for the chain.*

Proof.

$$\int_K P_u(A)Q(du) = \int_A P_u(K)Q(du) = \int_A Q(du) = Q(A)$$

□

For technical reasons, it is useful to consider a *lazy* Markov chain.

Definition 3. *(Lazy Markov chain) A Markov chain is called* lazy *if $P_u(\{u\}) \geq \frac{1}{2}$ for every $u \in K$.*

Given a chain M, we can make a lazy variant M' by simply flipping a fair coin and following M if it comes up heads, doing nothing otherwise. This construction preserves the stationary distribution of M.

Proposition 2. *Given a Markov chain M with stationary distribution Q, the lazy chain M' also has stationary distribution Q.*

Proof. Let P'_u denote the measure at u of the lazy M'. Clearly for all $A \in \mathcal{B}$,

$$P'_u(A) = \frac{1}{2}\mathbf{1}_A(u) + \frac{1}{2}P_u(A)$$

The claim follows from

$$\int_K P'_u(A)Q(du) = \frac{1}{2}\int_K \mathbf{1}_A(u)Q(du) + \frac{1}{2}\int_K P_u(A)Q(du) = \frac{1}{2}Q(A) + \frac{1}{2}Q(A) = Q(A)$$

□

We now describe some definitions which attempt to characterize some geometric notations about the Markov chain.

Definition 4. *(Ergodic flow) Let Q be the stationary distribution for a Markov chain. For $A \in \mathcal{B}$, define the* ergodic flow *of A as*

$$\Phi(A) \stackrel{\text{def}}{=} \int_A P_u(K \setminus A)Q(du)$$

From ergodic flow, we have the following fact

Proposition 3. $\Phi(A) = \Phi(K \setminus A)$ *for all* $A \in \mathcal{B}$.

Proof. This follows from definitions

$$\begin{aligned}\Phi(A) - \Phi(K \setminus A) &= \int_A P_u(K \setminus A) Q(du) - \int_{K \setminus A} P_u(A) Q(du) \\ &= Q(A) - \int_A P_u(A) Q(du) - \int_{K \setminus A} P_u(A) Q(du) \\ &= 0\end{aligned}$$

□

Fixing a set A, the ergodic flow can be interpreted as follows: use Q to induce a distribution on A. Now, starting from this distribution and taking a step according to the transition kernel of the chain, the ergodic flow is the probabilty that we leave the set A. Using this, we can define the conductance of a set as

Definition 5. *(Conductance of a set) Let* $A \in \mathcal{B}$. *Define its* conductance *(w.r.t. the stationary distribution* Q *of a Markov chain) as*

$$\phi(A) \stackrel{\text{def}}{=} \frac{\Phi(A)}{\min\{Q(A), 1 - Q(A)\}}$$

Let's step back for a second and parse this. What is the problem we are trying to characterize? Intuitively, we know that Markov chains mix poorly when there are region which don't contain the entire state space that are very hard to leave. Ergodic flow (roughly) tells us the chance that, given we start in a set A, we leave the set. But what if the set is very big? Then the chance we leave it may be small, but we don't care so much about this case.

Conductance of a set tries to account for this via the denominator: likely sets will have small ergodic flow, but large conductance due to the denominator. Similarly, very unlikely sets will have high conductance as long as we can get out. It's precisely the sets which have non-negligible probabilty, which are hard to escape, which have low conductance!

Of course, this is all hand-wavy, so to formalize this, we define the conductance of a Markov chain as follows.

Definition 6. *(Conductance of a chain) Given a Markov chain, define its* conductance ϕ *as*

$$\phi \stackrel{\text{def}}{=} \inf_{A \in \mathcal{B}} \phi(A) = \inf_{A: 0 < Q(A) \leq \frac{1}{2}} \frac{\Phi(A)}{Q(A)}$$

To conclude this section, we end with a way of measuring the distance between two distributions on the same measure space.

Definition 7. *(Total variation distance) Let* P, Q *be distributions on the same measure space. Then define the* total variation *between* P *and* Q *as*

$$\|P - Q\|_{\text{TV}} \stackrel{\text{def}}{=} \sup_{A \in \mathcal{B}} |P(A) - Q(A)|$$

3 Stationary distributions for random walk algorithms

We start with a warm-up.

Proposition 4. π_K *is stationary for ball walk.*

Proof. We verify (1). Fix an $A \in \mathcal{B}$. Note that $\pi_K(A) = \frac{\text{vol}(A)}{\text{vol}(K)}$. By assuming K is a full body in $\mathbb{R}^n$, we have that $\pi_K \ll \mu$, and hence the Radon-Nikodym density of π_K w.r.t. μ exists and is unique a.e. Hence, we can check that $f_K(x) = \frac{d\pi_K}{d\mu}(x) = \frac{1}{\text{vol}(K)}$ is the density a.e., and therefore

$$\int_K P_u(A)\pi_K(du) = \frac{1}{\text{vol}(K)} \int_K P_u(A)\mu(du)$$

It is easy to verify that

$$P_u(\{u\}) = 1 - \frac{\text{vol}(K \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))}$$

$$P_u(A) = \begin{cases} \frac{\text{vol}(A \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))} & \text{if } u \notin A \\ P_u(A \setminus \{u\}) + P_u(\{u\}) & \text{if } u \in A \end{cases}$$

or equivalently

$$P_u(A) = \frac{\text{vol}(A \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))} + \mathbf{1}_A(u) P_u(\{u\})$$

First, we note that

$$\begin{aligned}
\int_K \frac{\text{vol}(A \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))}\mu(du) &= \frac{1}{\text{vol}(B_n(0,\delta))} \int_K \text{vol}(A \cap B_n(u,\delta))\mu(du) \\
&= \frac{1}{\text{vol}(B_n(0,\delta))} \int_K \left(\int_K \mathbf{1}_A(v) \mathbf{1}_{B_n(u,\delta)}(v)\mu(dv) \right) \mu(du) \\
&\overset{(a)}{=} \frac{1}{\text{vol}(B_n(0,\delta))} \int_K \left(\int_K \mathbf{1}_A(v) \mathbf{1}_{B_n(u,\delta)}(v)\mu(du) \right) \mu(dv) \\
&\overset{(b)}{=} \frac{1}{\text{vol}(B_n(0,\delta))} \int_K \mathbf{1}_A(v) \left(\int_K \mathbf{1}_{B_n(v,\delta)}(u)\mu(du) \right) \mu(dv) \\
&= \frac{1}{\text{vol}(B_n(0,\delta))} \int_A \left(\int_K \mathbf{1}_K(u) \mathbf{1}_{B_n(v,\delta)}(u)\mu(du) \right) \mu(dv) \\
&= \int_A \frac{\text{vol}(K \cap B_n(v,\delta))}{\text{vol}(B_n(0,\delta))}\mu(dv)
\end{aligned}$$

where in (a) we used Fubini's theorem and in (b) we used the fact that $\mathbf{1}_{B_n(u,\delta)}(v) = \mathbf{1}_{B_n(v,\delta)}(u)$ (when viewed as functions of u and v). Now, note that

$$\begin{aligned}
\int_K \mathbf{1}_A(u) P_u(\{u\})\mu(du) &= \int_K \mathbf{1}_A(u) \left(1 - \frac{\text{vol}(K \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))} \right) \mu(du) \\
&= \int_K \mathbf{1}_A(u)\mu(du) - \int_K \mathbf{1}_A(u) \frac{\text{vol}(K \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))}\mu(du) \\
&= \text{vol}(A) - \int_A \frac{\text{vol}(K \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))}\mu(du)
\end{aligned}$$

and therefore

$$\begin{aligned}
\frac{1}{\text{vol}(K)} \int_K P_u(A)\mu(du) &= \frac{1}{\text{vol}(K)} \left[\int_K \frac{\text{vol}(A \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))}\mu(du) + \int_K \mathbf{1}_A(u) P_u(\{u\})\mu(du) \right] \\
&= \frac{1}{\text{vol}(K)} \left[\int_K \frac{\text{vol}(A \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))}\mu(du) + \text{vol}(A) - \int_A \frac{\text{vol}(K \cap B_n(u,\delta))}{\text{vol}(B_n(0,\delta))}\mu(du) \right] \\
&= \frac{\text{vol}(A)}{\text{vol}(K)}
\end{aligned}$$

which verifies (1). $\square$

Now we show the same result for hit and run. This proof is shorter than the one for ball walk, because we can take advantage of the reversibility of the Markov chain (which we could not for ball walk).

Proposition 5. *π_K is stationary for hit and run.*

Proof. Let $\eta(\cdot)$ denote the uniform measure on S_n. It is not hard to see that

$$P_u(A) = \int_{S_n} \int_{\mathbb{R}} \frac{\mathbf{1}_A(u+\lambda v)}{\mu(l_K(u,v))} \mu(d\lambda)\eta(dv)$$

To show reversibility, simply note that by the symmetry of the sampling,

$$\begin{aligned}
\int_B P_u(A)\mu(dx) &= \int_{S_n} \int_{\mathbb{R}} \frac{\mathbf{1}_A(u+\lambda v)\mathbf{1}_B(u)}{\mu(l_K(u,v))} \mu(d\lambda)\eta(dv) \\
&= \int_{S_n} \int_{\mathbb{R}} \frac{\mathbf{1}_A(u)\mathbf{1}_B(u+\lambda v)}{\mu(l_K(u,v))} \mu(d\lambda)\eta(dv) \\
&= \int_A P_u(B)\mu(dx)
\end{aligned}$$

The result now follows by Proposition 1. Note that it can be shown [BRS93] that $P_u(A)$ is given by

$$P_u(A) = \frac{1}{\text{surf}(S_n)} \int_A \frac{2}{\mu(l_K(u,x))\|x-u\|^{n-1}} \mu(dx)$$

where $\text{surf}(S_n)$ denotes the surface area of S_n, from which the symmetry is even more apparent. □

4 The role of conductance in convergence

Now that we have established the stationary distribution is of interest, we turn to the question of how fast can we get there? It turns out, conductance plays a key role in helping to answer these questions. In what follows, fix a lazy Markov chain with stationary distribution Q, initial distribution Q_0, and distribution at time t denoted Q_t. We have a very neat theorem

Theorem 1. *(Corollary 1.5, [LS93]) Put $M = \sup_{A\in\mathcal{B}} \frac{Q_0(A)}{Q(A)}$. Then,*

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

Note: here, we see explicitly that smaller conductance leads to slower convergence.

Proof. (Sketch) We won't detail the entire proof, see the references for further derivations. The high level ideas are as follows. For $x \in [0,1]$, define the set $\mathcal{G}_x$ as

$$\mathcal{G}_x \stackrel{\text{def}}{=} \{g : K \longrightarrow [0,1], g \text{ measurable} : \mathbb{E}_Q[g] = x\}$$

Next, for a time t, define the distance function $h_t(x)$ as

$$h_t(x) \stackrel{\text{def}}{=} \sup_{g\in\mathcal{G}_x} \mathbb{E}_{Q_t}[g] - \mathbb{E}_Q[g] = \sup_{g\in\mathcal{G}_x} \mathbb{E}_{Q_t}[g] - x$$

Proposition 6. *$h_t(x)$ is a concave function in x.*

Proof. Fix $x, y, \theta \in [0,1]$. By the linearity of expectation, if $g_1 \in \mathcal{G}_x$ and $g_2 \in \mathcal{G}_y$, then $\theta g_1 + (1-\theta) g_2 \in \mathcal{G}_{\theta x + (1-\theta) y}$. Hence, for any $g_i \in \mathcal{G}_i$, $i = 1, 2$, we have

$$\sup_{g \in \mathcal{G}_{\theta x + (1-\theta)y}} \mathbb{E}_{Q_t}[g] \geq \mathbb{E}_{Q_t}[\theta g_1 + (1-\theta) g_2] = \theta \mathbb{E}_{Q_t}[g_1] + (1-\theta)\mathbb{E}_{Q_t}[g_2]$$

and hence

$$\sup_{g \in \mathcal{G}_{\theta x + (1-\theta)y}} \mathbb{E}_{Q_t}[g] \geq \sup_{g_1 \in \mathcal{G}_x, g_2 \in \mathcal{G}_y} \theta \mathbb{E}_{Q_t}[g_1] + (1-\theta)\mathbb{E}_{Q_t}[g_2] = \theta \sup_{g \in \mathcal{G}_x} \mathbb{E}_{Q_t}[g] + (1-\theta) \sup_{g \in \mathcal{G}_y} \mathbb{E}_{Q_t}[g]$$

The result now clearly follows. □

$h_t(x)$ can be thought of as a proxy for the TV distance. First, for all A with $Q(A) = x$, we have

$$-h_t(1-x) \leq Q_t(A) - Q(A) \leq h_t(x)$$

Second, (we state without proof, see e.g. Lemma 1.2 [LS93]) if Q is atom-free, then

$$h_t(x) = \sup_{A \in \mathcal{B}: Q(A) = x} Q_t(A) - Q(A) \tag{2}$$

and the supremum is attained by some $A \in \mathcal{B}$. Hence if we upper bound $\sup_{x \in [0,1]} h_t(x)$, we also have an upper bound for $\sup_{A \in \mathcal{B}} |Q_t(A) - Q(A)|$. To do this, we develop the following recurrance on h_t.

Lemma 1. *(Lemma 1.3, [LS93]) Fix an $x \in [0,1]$ and put $y = \min(x, 1-x)$. If Q is atom-free, then*

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

Proof. First, suppose $x \leq \frac{1}{2}$ (the $x > \frac{1}{2}$ case is similar). By (2), let A be such that $h_t(x) = Q_t(A) - Q(A)$ and $Q(A) = x$. Now, cleverly define two functions

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

It is trivial to check that $\frac{1}{2}(g_1 + g_2)(u) = P_u(A)$ for all u, therefore

$$\mathbb{E}_{Q_{t-1}}\left[\frac{1}{2}(g_1 + g_2)\right] = \int_K P_u(A) Q_{t-1}(du) = Q_t(A)$$

Now define $x_i = \mathbb{E}_Q[g_i]$ for $i = 1, 2$. By the lazy Markov chain assumption, $0 \leq g_i \leq 1$, and hence $g_i \in \mathcal{G}_{x_i}$. It is easy to check that

$$\frac{1}{2}(x_1 + x_2) = x$$

Therefore,

$$\begin{aligned} h_t(x) &= Q_t(A) - Q(A) \\ &= \mathbb{E}_{Q_{t-1}}\left[\frac{1}{2}(g_1 + g_2)\right] - x \\ &= \frac{1}{2}(\mathbb{E}_{Q_{t-1}}[g_1] - x + \mathbb{E}_{Q_{t-1}}[g_2] - x) \\ &\leq \frac{1}{2}(h_{t-1}(x_1) + h_{t-1}(x_2)) \end{aligned}$$

Now here's where the magic happens to relate x_1, x_2 to conductance. Observe that

$$\begin{aligned} x_2 &= \int_K g_2(u)Q(du) \\ &= \int_K \mathbf{1}_A(u)Q(du) + 2\int_{K\setminus A} P_u(A)Q(du) \\ &= Q(A) + 2\Phi(A) \end{aligned}$$

and hence

$$x_2 - x = x - x_1 = 2\Phi(A) \geq 2\phi x$$

from which we deduce

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

hence by the concavity of h_{t-1}, we have

$$h_t(x) \leq \frac{1}{2}(h_{t-1}(x_1) + h_{t-1}(x_2)) \leq \frac{1}{2}(h_{t-1}(x(1-2\phi)) + h_{t-1}(x(1+2\phi)))$$

□

Now, by unrolling this recursion to $t = 0$, we get the following lemma

Lemma 2. *(Theorem 1.4, [LS93]) Let c_1, c_2 be constants such that for all $x \in [0,1]$ we have*

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

Then we have for $t \geq 0$ and $x \in [0,1]$,

$$h_t(x) \leq c_1 + c_2 \min(\sqrt{x}, \sqrt{1-x}) \left(1 - \frac{\phi^2}{2}\right)^t \tag{3}$$

The theorem now follows. By definition of M, for $x \in [0,1]$ we have

$$h_0(x) \leq \min(Mx, 1-x) \leq \sqrt{M} \min(\sqrt{x}, \sqrt{1-x})$$

so we can plug into (3) to recover the result. □

5 Conductance lower bounds

Establishing convergence rates via conductance, as was done in the previous section, is the easy part. The hard part is getting a lower bound on the conductance of a Markov chain. We simply state some results without proof.

Theorem 2. *(Theorem 5.2, [Vem05]) Let K be a convex body with diameter D, and suppose for every $u \in K$ we have $l(u) \stackrel{\text{def}}{=} 1 - P_u(\{u\}) \geq l$ where l is some constant. Then ball walk with step size δ satisfies*

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

How do we ensure such a lower bound l during ball walk? One way is to relax the convex body slightly. Given a body K, define K' as

$$K' \stackrel{\text{def}}{=} \bigcup_{u\in K} B_n(u, 1/n)$$

Proposition 7. *Ball walk on K' with $\delta = 1/2n\sqrt{n}$ satisfies $l \geq 1/8$.*

Hence, we have a convergence rate bound

Theorem 3. *(Lazy) ball walk on K' with $\delta = 1/2n\sqrt{n}$ satisfies after t steps*

$$\|Q_t - \pi_{K'}\|_{\mathrm{TV}} \leq \left(1 - \Theta\left(\frac{n^2}{D^2}\right)\right)^t$$

References

[BRS93] Claude J. P. Bélisle, H. Edwin Romeijn, and Robert L. Smith. Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res.*, 18(2), 1993.

[LS93] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures and Algorithms*, 4(4), 1993.

[Vem05] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and Computational Geometry*, 2005.