Convergence of Gibbs Sampling: Coordinate Hit-And-Run Mixes Fast

Aditi Laddha
Georgia Institute of Technology, Atlanta, GA, USA

Santosh S. Vempala
Georgia Institute of Technology, Atlanta, GA, USA

Abstract

The Gibbs Sampler is a general method for sampling high-dimensional distributions, dating back to 1971. In each step of the Gibbs Sampler, we pick a random coordinate and re-sample that coordinate from the distribution induced by fixing all the other coordinates. While it has become widely used over the past half-century, guarantees of efficient convergence have been elusive. We show that for a convex body K in $\mathbb{R}^n$ with diameter D, the mixing time of the Coordinate Hit-and-Run (CHAR) algorithm on K is polynomial in n and D. We also give a lower bound on the mixing rate of CHAR, showing that it is strictly worse than hit-and-run and the ball walk in the worst case.

2012 ACM Subject Classification Mathematics of computing → Markov processes

Keywords and phrases Gibbs Sampler, Coordinate Hit and run, Mixing time of Markov Chain

Digital Object Identifier 10.4230/LIPIcs.SoCG.2021.51

Funding *Aditi Laddha*: Supported in part by NSF awards 1839323 and 2007443.
Santosh S. Vempala: Supported in part by NSF awards 1839323, 1909756 and 2007443.

1 Introduction

Sampling a high-dimensional distribution is a fundamental problem and a basic ingredient of algorithms for optimization, integration, statistical inference, and other applications. Progress on sampling algorithms has led to many useful tools, both theoretical and practical. In the most general setting, given access to a function $f : \mathbb{R}^n \to \mathbb{R}_+$, the goal is to generate a point x whose density is proportional to $f(x)$. Two special cases of particular interest are when f is uniform over a convex body and when f is a Gaussian restricted to a convex set.

The generic approach to sampling is by a Markov chain whose state space is the convex body. The chain is designed so that it is ergodic, time-reversible, and has the desired density as its stationary distribution. The key question is to bound the rate of convergence of the Markov chain. The Ball walk [14, 12, 18] and Hit-and-Run [2, 22, 17] are two Markov chains that work in full generality, and have been shown to mix rapidly (i.e, the convergence rate is polynomial) for arbitrary log-concave densities. Over three decades of improvements, the complexity of this problem has been reduced to a small polynomial in the dimension for the total number of function evaluations with a factor of n^2 per function call for the total number of arithmetic operations. For a log-concave density with support of diameter D, the mixing time is $O^*(n^2D^2)$, taking the same number of function evaluations, with arithmetic complexity of $O^*(n^4D^2)$[12, 15, 17].

A simple and widely-used algorithm that pre-dates these developments considerably is the Gibbs Sampler, proposed by Turchin in 1971 [23]. It is inspired by statistical physics and is commonly used for sampling distributions [5, 6] and Bayesian inference [8, 9, 10]. To sample a multivariate density, at each step, the sampler picks a coordinate (either at random or in order, cycling through the coordinates), fixes all other coordinates, and re-samples this coordinate from the induced distribution. This is very similar to Hit-and-Run, except that instead of picking a direction uniformly at random from the unit sphere, it is picked only from one of the n basis vectors (see [1] for a historical account and more background). It was reported to be significantly faster than Hit-and-Run in state-of-the-art software for volume computation and integration [3, 7, 4]. Gibbs sampling, also called Coordinate Hit-and-Run, has a computational benefit: updating the current point takes $O(n)$ time rather than $O(n^2)$ even for polyhedra since the update is along only one coordinate direction. Thus the overhead per step is reduced from $O(n^2)$ as in all previous algorithms to $O(n)$. However, despite half a century of intense study, the convergence rate of Gibbs sampling has remained an open problem.

In this paper, we show that the Gibbs sampler mixes rapidly for any convex body K. Before stating our main theorem formally, we define the Gibbs sampler.

Coordinate Hit-and-Run

Algorithm 1 describes the Coordinate Hit-and-Run walk for sampling uniformly from a convex body $K \in \mathbb{R}^n$. Let $\{\mathbf{e}_i : 1 \le i \le n\}$ be the standard basis for $\mathbb{R}^n$. The starting point is in the interior of K and is given as an input to the algorithm.

Algorithm 1 Coordinate Hit-and-Run (CHAR).

Input: a point $x^{(0)} \in K$, integer T.
for $i = 1, 2, \cdots, T$ **do**
  Pick a uniformly random axis direction e_j
  Set x^i to be a random point along the line $\ell = \left\{x^{(i-1)} + te_j \;:\; t \in \mathbb{R}\right\}$ chosen uniformly from $\ell \cap K$.
end
Output: x^T.

The stationary distribution of the Coordinate hit-and-run walk is the uniform distribution π_K over K. To sample from a general log-concave density $f : \mathbb{R}^n \to \mathbb{R}_+$ the only change is in Step 2, where the next point y is chosen according to $f(y)$ restricted to ℓ. In both cases, the process is symmetric and ergodic and so the stationary distribution of the Markov chain is the desired distribution.

We can now state our main theorem (see Sec. 1.2 for the definition of a warm start).

▶ **Theorem 1.** *Let K be a convex body in $\mathbb{R}^n$ containing a unit ball. Let R^2 be the expected squared distance of a uniform random point in K from the centroid of K. Then the mixing time of Coordinate Hit-and-Run from a warm start in K is $\widetilde{O}\left(n^9 R^2\right)$.*

By applying an affine transformation, R can be made $O(\sqrt{n})$. We note that from a warm start both the Ball Walk and Hit-and-Run have a mixing time of $\widetilde{O}(n^2R^2)$ [12, 17]. While our bound is likely not the best polynomial bound for CHAR, in Section 4, we show that it is necessarily higher than the bound for hit-and-run.

Concurrently and independently, Narayanan and Srivastava [20] also proved a polynomial bound on the mixing rate of Coordinate Hit-and-Run, with a different proof. They showed that CHAR mixes in $\tilde{O}(n^7 R_1^4)$ steps where R_1 is the smallest number s.t., $B_\infty \subseteq K \subseteq R_1 B_\infty$, i.e., the cube sandwiching ratio (R_1 can be larger than R in our theorem by a factor of $\sqrt{n}$). This quantity R_1 can be bounded by $O(n)$ after an affine transformation.

A key ingredient of our proof is a new "ℓ_0"-isoperimetric inequality. We will need the following definition.

▶ **Definition 2** (Axis-disjoint). *Two measurable sets S_1, S_2 are called* axis-disjoint *if* $\forall x \in S_1, \forall y \in S_2, |\{i \in [n] : x_i = y_i\}| \le n - 2$.

In other words, no point from S_1 is on the same axis-parallel line as any point in S_2 (see Fig. 1).

The main component of the proof of Theorem 1 is the following isoperimetric inequality for axis-disjoint subsets of a convex body.

▶ **Theorem 3.** *Let K be a convex body in $\mathbb{R}^n$ containing a unit ball with $R^2 = \mathbb{E}_K(\|x - z_K\|^2)$ where z_K is the centroid of K. Let $S_1,S_2 \subseteq K$ be two measurable subsets of K such that S_1, S_2 are axis-disjoint. Then for any $\epsilon \geq 0$, the set $S_3 = K \setminus \{S_1 \cup S_2\}$ satisfies*

$$\text{vol}(S_3) \geq \frac{\epsilon}{48 \cdot 10^3 \cdot n^{3.5} R \log n} \left(\min\{\text{vol}(S_1), \text{vol}(S_2)\} - \epsilon \text{vol}(K)\right).$$

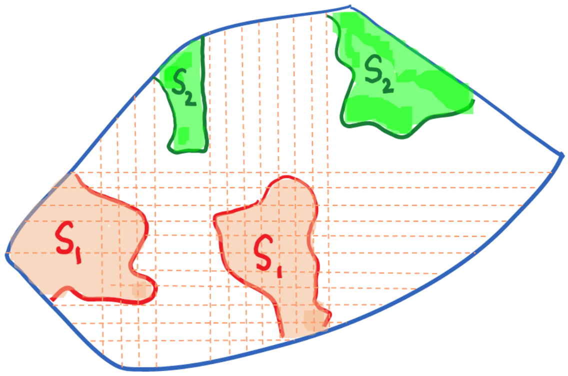

Figure 1 Axis-disjoint subsets S_1 and S_2.

1.1 Approach

At a high level, we follow the proof of rapid mixing based on the conductance of Markov chains [21] in the continuous setting [16]. We give a simple, new one-step coupling lemma which reduces the problem of lower bounding the conductance of the underlying Markov chain to an isoperimetric inequality about axis-disjoint sets in high dimension. Roughly speaking, the inequality says the following: If two subsets of a convex body are axis-disjoint, then the remaining mass of the body is proportional to the smaller of the two subsets. This inequality is our main technical contribution. In comparison, the inequality for Euclidean distance says that for any two subsets of a convex body, the remaining mass is proportional to their (minimum) Euclidean distance times the smaller of the two subset volumes.

Standard approaches to proving such inequalities, notably localization [11, 13], which reduce the desired high-dimensional inequality to a one-dimensional inequality, do not seem to be directly applicable to proving this "ℓ_0-type" inequality. So we develop a first-principles approach where we first prove the inequality for cubes, taking advantage of their product structure, and then for general bodies using a tiling of space with cubes. In the course of the latter part, we will use several known properties of convex bodies, including Euclidean isoperimetry.

1.2 Preliminaries

- **Markov chain**: Let $\mathcal{M}$ be a Markov chain with state space K and stationary distribution Q. For any measurable subset $S \subseteq K$ and $x \in K$, let $P_x(S)$ be the probability that one step of $\mathcal{M}$ from x goes to a point in S. Q being stationary implies that for any measurable subset $S \subseteq K$

  $$\int_K P_x(S) dQ(x) = Q(S).$$

  The Markov chain is **time-reversible** if for any two subsets $A, B \subseteq K$

  $$\int_A P_x(B) dQ(x) = \int_B P_x(A) dQ(u).$$

  For a measurable subset S of the state space of a Markov chain with stationary distribution Q, the ergodic flow of S, denoted by $p(S)$ is defined as

  $$p(S) = \int_S P_x(K \setminus S)\, dQ(x).$$

- **Conductance**: The conductance of a subset $S \subseteq K$, denoted by $\phi(S)$, is defined as

  $$\phi(S) = \frac{\int_S P_x(K \setminus S)\, dQ(x)}{\min\{Q(S), Q(K \backslash S)\}},$$

  and the conductance of $\mathcal{M}$ is defined as

  $$\phi = \inf_{0 < Q(S) \leq 1/2} \phi(S).$$

  For any $s \in [0, 1/2]$ the s-conductance of the Markov chain is:

  $$\phi_s = \inf_{S: s < Q(S) \leq \frac{1}{2}} \frac{p(S)}{Q(S) - s}.$$

- **Warm Start**: Given distributions P and Q on the same state space $\mathcal{A}$, P is said to be M-warm with respect to Q if

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

  If the initial distribution Q_0 is $O(1)$-warm with respect to the stationary distribution Q for some Markov chain, we say that Q_0 is a warm start for Q.
- **Lazy chain**: A lazy version of a Markov chain with transition probability P is one where we use the transition probability $P_x(\{y\}) = P_x(\{y\})/2 + \mathbf{1}(x = y)/2$, so that with probability $1/2$, the chain feels lazy and stays in the same state.
- For a body $K \subseteq \mathbb{R}^n$, let π_K denote the uniform distribution on K and $\mathbb{E}_K(X)$ denote the expected value of X with respect to π_K.

The following theorem shows that the s-conductance of a Markov chain bounds its rate of convergence from a warm start.

▶ **Theorem 4** ([16]). *Suppose that a lazy, time-reversible Markov chain with stationary distribution Q has s-conductance at least ϕ_s. Then with initial distribution Q_0, and*

$$H_s = \sup \{|Q(A) - Q_0(A)| \,:\, A \subset K, Q(A) \leq s\},$$

the distribution Q_t after t steps satisfies

$$d_{TV}(Q_t, Q) \leq H_s + \frac{H_s}{s}\left(1 - \frac{\phi_s^2}{2}\right)^t.$$

2 The isoperimetric inequality

Before proving Theorem 3, we need a few definitions.

▶ **Definition 5** (Axis-aligned Line). *A line ℓ in $\mathbb{R}^n$ is called* axis-aligned *if $\ell = \{x \in \mathbb{R}^n : x = c + te_i, t \in \mathbb{R}\}$ for an $i \in \{1, \dots, n\}$ and a point $c \in \mathbb{R}^n$.*

▶ **Definition 6** (Axis-aligned Cube). *An axis-aligned cube C is defined as*

$$C = \{x \in \mathbb{R}^n : \|x - y\|_\infty \leq c\},$$

where $y \in \mathbb{R}^n$ is a fixed point and c is a positive constant.

▶ **Definition 7** (Axis-parallel Extension). *For a body K in $\mathbb{R}^n$ and a measurable subset $S \subseteq K$, the axis-parallel extension of S in K, denoted by $ext_K(S)$ is defined as*

$$\text{ext}_K(S) = \{x \in K \backslash S : \exists y \in S \text{ such that } |\{i \in [n] : x_i = y_i\}| = n - 1\}.$$

In other words, $ext_K(S)$ is the set of points in $K \backslash S$ obtained by changing exactly 1 coordinate from a point in S.

We will bound the conductance of CHAR in an axis aligned cube by its mixing time.

▷ Claim 8. The mixing time of the Coordinate Hit-and-Run chain in an axis-aligned cube $C \in \mathbb{R}^n$ is $O(n \log n)$.

Proof. WLOG, assume that C is a cube with side length 1. In a step of CHAR, if axis e_i is selected, then the i-th coordinate of the current point is re-sampled uniformly at random from $[0, 1]$. So starting from any point in C, after every axis direction has been picked at least once, the distribution induced by CHAR will be uniform on C. Using the coupon collector bound, the number of steps of CHAR needed to ensure that every coordinate has been re-sampled at least once is at most $4n \log n = O(n \log n)$ in expectation. ◁

▶ **Lemma 9** (Cube isoperimetry). *For an axis-aligned cube $C \in \mathbb{R}^n$, and any two axis-disjoint subsets $S_1, S_2 \subseteq C$, with $S_3 = C \setminus \{S_1 \cup S_2\}$, the following holds:*

$$\text{vol}(S_3) \geq \frac{1}{4n \log n} \cdot \min \{\text{vol}(S_1), \text{vol}(S_2)\}.$$

▶ Remark 10. We believe that the bound above is not optimal, and even an absolute constant factor might be possible. In the appendix, we give a different proof achieving a weaker bound.

Proof. Let $\text{vol}(S_1) \leq \text{vol}(S_2)$. For a Markov chain with mixing time t_{mix}, the conductance is at least $1/t_{mix}$ [19]. From Claim 8, we get

$$\phi(S_1) = \frac{\int_{S_1} P_x(C \setminus S_1)\, dx}{\text{vol}(S_1)} \geq \frac{1}{4n \log n}$$

$$\int_{S_1} P_x(C \setminus S_1)\, dx \geq \frac{\text{vol}(S_1)}{4n \log n} \tag{1}$$

Since S_1 and S_2 are axis-disjoint, the probability of moving from a point in S_1 to a point in S_2 in one step is 0 and hence

$$\int_{S_1} P_x(C \setminus S_1)\, dx \leq \int_{S_1} P_x(S_3)\, dx = \int_{S_3} P_x(S_1)\, dx \leq \text{vol}(S_3). \tag{2}$$

Combining equation (1) and equation (2), we get

$$\text{vol}(S_3) \geq \frac{1}{4n \log n} \cdot \text{vol}(S_1).$$ ◀

The next lemma is an isoperimetric inequality from [11].

▶ **Lemma 11** (Euclidean isoperimetry). *[11] Let $K \subset \mathbb{R}^n$ be a convex body containing a unit ball and $R^2 = \mathbb{E}_K(\|x - z_k\|^2)$ where z_K is the centroid of K. For a subset $S \subseteq K$, let $\partial_K(S)$ denote the boundary of S, relative to K. Then for any $S \subseteq K$ of volume at most* $\mathrm{vol}(K)/2$, *we have*

$$\mathrm{vol}_{n-1}(\partial_K S) \geq \frac{\ln 2}{R} \mathrm{vol}(S).$$

We can now prove the new isoperimetric inequality, restated below for convenience.

▶ **Theorem 3.** *Let K be a convex body in $\mathbb{R}^n$ containing a unit ball with $R^2 = \mathbb{E}_K(\|x - z_K\|^2)$ where z_K is the centroid of K. Let $S_1, S_2 \subseteq K$ be two measurable subsets of K such that S_1, S_2 are axis-disjoint. Then for any $\epsilon \geq 0$, the set $S_3 = K \backslash \{S_1 \cup S_2\}$ satisfies*

$$\mathrm{vol}(S_3) \geq \frac{\epsilon}{48 \cdot 10^3 \cdot n^{3.5} R \log n} \left(\min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\} - \epsilon \mathrm{vol}(K) \right).$$

Proof of Theorem 3. Let $K' = (1 - \alpha)K$ for $\alpha = \frac{\epsilon}{20n}$, and $S'_i = S_i \cap K'$ for $i \in \{1, 2\}$. Assume $\mathrm{vol}(S'_1) \leq \mathrm{vol}(S'_2)$. For any subset $X \subseteq K$, we have

$$\mathrm{vol}(X \cap K') \geq \mathrm{vol}(X) - (1 - (1 - \alpha)^n)\mathrm{vol}(K) \geq \mathrm{vol}(X) - \frac{\epsilon}{2} \cdot \mathrm{vol}(K).$$

Next consider a standard lattice of width δ, with each lattice point inducing a cube of side length δ. We choose $\delta = \frac{\alpha}{4\sqrt{n}}$ to ensure that the cubes which intersect K' are fully contained in K. Let $\mathcal{C}$ be the set of hypercubes in this lattice that intersect S_1. We partition the set of cubes in $\mathcal{C}$ as

- $\mathcal{C}_1$: the set of cubes in $\mathcal{C}$ where S_1 takes up less than $\frac{2}{3}$ of the volume of the cube, and
- $\mathcal{C}_2$: the set of cubes in $\mathcal{C}$ where S_1 takes up at least $\frac{2}{3}$ of each cube.

If $\mathrm{vol}(\mathcal{C}_1 \cap S_1) \geq \mathrm{vol}(S_1)/2$, i.e., at least $\frac{1}{2}$ of $\mathrm{vol}(S_1)$ resides in cubes in $\mathcal{C}_1$, then we can apply Lemma 9 to every cube in $\mathcal{C}_1$ individually to get a bound on $\mathrm{vol}(S_3)$. However, the cubes in $\mathcal{C}_1$ might not be fully contained in K and so before using the cube isoperimetry, we need to move to K' and consider the subset of cubes of $\mathcal{C}_1$ that intersect K'. Let $\mathcal{C}'_1 = \{c \in \mathcal{C}_1 : c \cap K' \neq \emptyset\}$. By our choice of α, we have $\mathcal{C}'_1 \subseteq K$ and

$$\mathrm{vol}(\mathcal{C}'_1 \cap S_1) \geq \mathrm{vol}(\mathcal{C}_1 \cap S_1 \cap K') \geq \mathrm{vol}(\mathcal{C}_1 \cap S_1) - \frac{\epsilon}{2} \cdot \mathrm{vol}(K).$$

Now consider a cube c in $\mathcal{C}'_1$. Let $x = \frac{\mathrm{vol}(c \cap S_1)}{\mathrm{vol}(c)}$ and $y = \frac{\mathrm{vol}(c \cap S_2)}{\mathrm{vol}(c)}$ where $0 < x \leq 2/3$ and $x + y \leq 1$. Then, applying Lemma 9for any feasible values of x and y, we have

$$\frac{\mathrm{vol}(S_3 \cap c)}{\mathrm{vol}(c)} \geq \max\left\{1 - x - y, \frac{1}{4n \log n} \min\{x, y\}\right\} \geq \frac{1}{4n \log n} \cdot \frac{x}{4}$$

Applying this argument to each cube in $\mathcal{C}'_1$, we get

$$\begin{aligned}
\mathrm{vol}(S_3) &\geq \frac{1}{16n \log n} \cdot \sum_{c \in \mathcal{C}'_1} \mathrm{vol}(c \cap S_1) = \frac{1}{16n \log n} \cdot \mathrm{vol}(\mathcal{C}'_1 \cap S_1) \\
&\geq \frac{1}{16n \log n} \left(\mathrm{vol}(\mathcal{C}_1 \cap S_1) - \frac{\epsilon}{2} \cdot \mathrm{vol}(K) \right) \\
&\geq \frac{1}{16n \log n} \left(\frac{1}{2} \cdot \mathrm{vol}(S_1) - \frac{\epsilon}{2} \cdot \mathrm{vol}(K) \right) \\
&\geq \frac{1}{32n \log n} \left(\mathrm{vol}(S_1) - \epsilon \cdot \mathrm{vol}(K) \right).
\end{aligned}$$

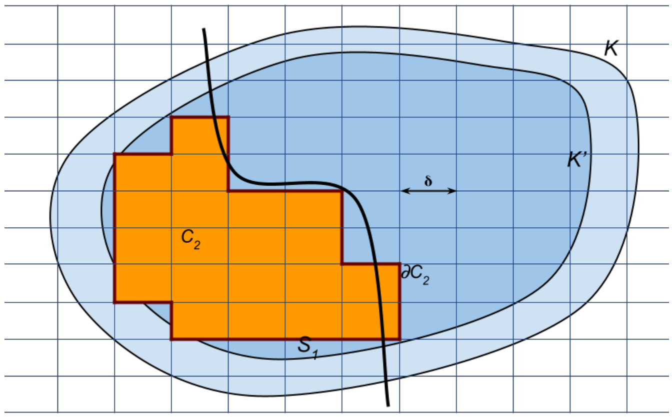

Figure 2 Illustration of the isoperimetry proof.

So assume that $\mathrm{vol}(\mathcal{C}_1 \cap S_1) \leq \frac{1}{2} \cdot \mathrm{vol}(S_1)$ and consequently $\mathrm{vol}(\mathcal{C}_2 \cap S_1) \geq \frac{1}{2} \cdot \mathrm{vol}(S_1)$. Let $\partial\mathcal{C}_2$ be the internal boundary of $\mathcal{C}_2$ in K. Because $\mathcal{C}_2$ is comprised of hypercubes, $\partial\mathcal{C}_2$ consists of facets of axis-aligned n-dimensional cubes. Consider a facet f on this boundary with normal axis e_f, let the cube adjacent to this facet in $\mathcal{C}_2$ be f_2 and the cube adjacent to the facet not in $\mathcal{C}_2$ be f_1. So, $\mathrm{vol}_{n-1}(f) \cdot \delta = \mathrm{vol}(f_1)$ and f_1 cannot be in $\mathcal{C}_2$ (it can belong to $\mathcal{C}_1$ but we account for that in the next step). Since at least 2/3 of the volume f_2 is in S_1, the support of marginal of S_1 along any axis direction will be at least 2/3 of the support of marginal of f_2 along that axis. Therefore, at least 2/3 of the mass of f and by extension f_1 is reachable from a point in S_1 along e_f and therefore cannot be in S_2. Now if $f_1 \in \mathcal{C}_1$, there are 2 possibilities:

1. $\mathrm{vol}(f_1 \cap S_1) \leq \mathrm{vol}(f_1)/3$, we can simply subtract this volume from the volume of S_3 gained from f_1 and get
$$\mathrm{vol}(f_1 \cap S_3) \geq \frac{2}{3} \cdot \delta\mathrm{vol}_{n-1}(f) - \frac{1}{3} \cdot \mathrm{vol}(f_1) = \frac{2}{3} \cdot \mathrm{vol}(f_2) - \frac{1}{3} \cdot \mathrm{vol}(f_1) = \frac{1}{3} \cdot \mathrm{vol}(f_1).$$
2. Otherwise, $\frac{1}{3} \cdot \mathrm{vol}(f_1) \leq \mathrm{vol}(f_1 \cap S_1) \leq \frac{2}{3} \cdot \mathrm{vol}(f_1)$ and using Lemma 9, we get
$$\mathrm{vol}(f_1 \cap S_3) \geq \frac{1}{4n \log n} \cdot \frac{\mathrm{vol}(f_1 \cap S_1)}{4} = \frac{1}{48n \log n} \cdot \mathrm{vol}(f_1).$$

So, for every facet $f \in \partial_K(\mathcal{C}_2)$, we get
$$\min\left\{\frac{1}{48n \log n}, \frac{1}{4}\right\} \cdot \delta \cdot \mathrm{vol}_{n-1}(f) = \frac{1}{48n \log n} \cdot \delta \cdot \mathrm{vol}_{n-1}(f). \tag{3}$$
Since every such neighboring cube can be counted at most $2n$ times using this argument, we get at most $1/2n$ of the above volume in S_3. But f_2 might not be (fully) contained in K. So, we need to move to K'. Let $\mathcal{C}'_2 = \{c \in \mathcal{C}_2 : c \cap K' \neq \emptyset\}$. Our choice of α ensures that the cubes in $\mathcal{C}'_2$ and all their neighboring cubes are fully contained in K. Let $\partial_{K'}(\mathcal{C}'_2)$ be the boundary of $\mathcal{C}'_2$ relative to K'. Then $\partial_{K'}(\mathcal{C}'_2) \subseteq \partial_K(\mathcal{C}_2)$ as $\partial_{K'}(\mathcal{C}'_2)$ only consists of the boundary of $\mathcal{C}_2 \cap K'$ internal to K'. So, every facet $f \in \partial_{K'}(\mathcal{C}'_2)$ contributes at least
$$\frac{1}{2n} \cdot \frac{1}{48n \log n} \cdot \delta \cdot \mathrm{vol}_{n-1}(f)$$
to $\mathrm{vol}(S_3)$. Because S_1 occupies at least 2/3 of every cube in $\mathcal{C}'_2$,
$$\mathrm{vol}(\mathcal{C}'_2 \cap S_1) \leq \mathrm{vol}(\mathcal{C}'_2) \leq \frac{3}{2}\mathrm{vol}(S_1 \cap K') \leq \frac{3}{4}\mathrm{vol}(K') \tag{4}$$

and by Lemma 11,

$$\begin{aligned}
\mathrm{vol}\left(\partial_{K'}\left(\mathcal{C}_2'\right)\right) &\geq \frac{\ln 2}{R}\cdot\min\left\{\mathrm{vol}\left(\mathcal{C}_2'\right),\mathrm{vol}\left(K'\backslash\mathcal{C}_2'\right)\right\}\\
&\geq \frac{\ln 2}{R}\cdot\min\left\{\mathrm{vol}\left(\mathcal{C}_2\cap K'\right),\frac{1}{4}\mathrm{vol}\left(K'\right)\right\} && \text{(from eq. (4))}\\
&\geq \frac{\ln 2}{R}\cdot\min\left\{\mathrm{vol}\left(\mathcal{C}_2\cap K'\right),\frac{1}{3}\mathrm{vol}\left(\mathcal{C}_2\cap K'\right)\right\}\\
&\geq \frac{\ln 2}{3R}\cdot\mathrm{vol}\left(\mathcal{C}_2\cap K'\right). && (5)
\end{aligned}$$

Combining equation (3) and equation (5), we get

$$\begin{aligned}
\mathrm{vol}(S_3) &\geq \frac{1}{2n}\cdot\sum_{f\in\partial_{K'}(\mathcal{C}_2')}\frac{\delta}{48n\log n}\mathrm{vol}_{n-1}(f)\\
&= \frac{\delta}{96n^2\log n}\cdot\mathrm{vol}(\partial_{K'}(\mathcal{C}_2'))\geq\frac{\delta\ln 2}{300Rn^2\log n}\cdot\mathrm{vol}\left(\mathcal{C}_2\cap K'\right)\\
&\geq \frac{\delta\ln 2}{300Rn^2\log n}\cdot\mathrm{vol}\left(\mathcal{C}_2\cap K'\cap S_1\right)\\
&\geq \frac{\delta\ln 2}{300Rn^2\log n}\cdot\left(\mathrm{vol}\left(\mathcal{C}_2\cap S_1\right)-(1-(1-\alpha)^n)\mathrm{vol}(K)\right)\\
&\geq \frac{\delta\ln 2}{300Rn^2\log n}\cdot\left(\frac{1}{2}\mathrm{vol}(S_1)-\frac{\epsilon}{2}\mathrm{vol}(K)\right)\\
&\geq \frac{\delta\ln 2}{600Rn^2\log n}\cdot\left(\mathrm{vol}(S_1)-\epsilon\mathrm{vol}(K)\right)
\end{aligned}$$

Using $\delta=\frac{\alpha}{4\sqrt{n}}=\frac{\epsilon}{80n\sqrt{n}}$, we have

$$\mathrm{vol}(S_3)\geq\frac{\epsilon\ln 2}{48\cdot 10^3\cdot Rn^{3.5}\log n}\cdot\left(\mathrm{vol}(S_1)-\epsilon\mathrm{vol}(K)\right).$$ ◀

3 Conductance

Here we bound the s-conductance of CHAR. The following simple lemma lets us reduce the s-conductance of K to isoperimetry of axis-disjoint subsets in K.

▶ **Lemma 12.** *Let $S_1\subseteq K$ be a measurable subset of K and $S_2=K\backslash S_1$. Let $S_1'=\{x\in S_1 : P_x(S_2)<\frac{1}{2n}\}$ and $S_2'=\{x\in S_2 : P_x(S_1)<\frac{1}{2n}\}$. Then S_1' and S_2' are axis disjoint.*

Proof. Assume S_1' and S_2' are not axis-disjoint, then let ℓ be an axis-parallel line passing through both S_1' and S_2'. Let $x\in S_1'\cap\ell$ and $y\in S_2'\cap\ell$. Then

$$P_x(S_2)\geq\frac{1}{n}\frac{\mathrm{len}(\ell\cap S_2)}{\mathrm{len}(\ell\cap K)}\Rightarrow\mathrm{len}(\ell\cap S_2)<\frac{\mathrm{len}(\ell\cap K)}{2}$$

and

$$P_y(S_1)\geq\frac{1}{n}\frac{\mathrm{len}(\ell\cap S_1)}{\mathrm{len}(\ell\cap K)}\Rightarrow\mathrm{len}(\ell\cap S_1)<\frac{\mathrm{len}(\ell\cap K)}{2}.$$

This is a contradiction as $\mathrm{len}(\ell\cap K)=\mathrm{len}(\ell\cap S_1)+\mathrm{len}(\ell\cap S_2)$. ◀

▶ **Theorem 13.** *Let K be a convex body in $\mathbb{R}^n$ containing a unit ball with $R^2 = \mathbb{E}_K(\|x - z_K\|^2)$ where z_K is the centroid of K. Then the s-conductance of coordinate hit-and-run in K is at least* $\frac{s}{8 \cdot 10^5 \cdot Rn^{4.5} \log n}$.

Proof. Let $S_1 \subseteq K$ be a measurable subset of K with $s < \pi_K(S_1) \leq 1/2$ and let $S_2 = K \backslash S_1$. Let

$$S_1' = \{x \in S_1 : P_x(S_2) < \frac{1}{2n}\} \quad \text{and} \quad S_2' = \{x \in S_2 : P_x(S_1) < \frac{1}{2n}\}.$$

Let $S_3' = K \backslash \{S_1' \cup S_2'\}$. From Lemma 12, we know that S_1' and S_2' are axis-disjoint. Thus, from Theorem 3 with $\epsilon = s/2$, we get

$$\text{vol}(S_3') \geq \frac{s}{96 \cdot 10^3 \cdot Rn^{3.5} \log n} \left(\min\{\text{vol}(S_1'), \text{vol}(S_2')\} - \frac{s}{2}\text{vol}(K)\right).$$

If $\text{vol}(S_1') < \text{vol}(S_1)/2$, then

$$\begin{aligned}\int_{x \in S_1} P_x(S_2)dx &= \int_{x \in S_1'} P_x(S_2)dx + \int_{x \in S_1 \backslash S_1'} P_x(S_2)dx \\ &\geq \frac{1}{2n}\text{vol}(S_1 \backslash S_1') \geq \frac{1}{4n}\text{vol}(S_1),\end{aligned}$$

and $\phi_s(S_1) \geq \frac{1}{4n}$. If $\text{vol}(S_2') < \text{vol}(S_2)/2$, then

$$\int_{x \in S_1} P_x(S_2)dx = \int_{x \in S_2} P_x(S_1)dx \geq \frac{1}{4n}\text{vol}(S_2).$$

So, assume that $\text{vol}(S_1') \geq \text{vol}(S_1)/2$ and $\text{vol}(S_2') \geq \text{vol}(S_2)/2$. Then,

$$\int_{x \in S_1} P_x(S_2)dx \geq \int_{x \in S_1 \backslash S_1'} P_x(S_2)dx \geq \frac{1}{2n} \cdot \text{vol}(S_1 \backslash S_1'), \tag{6}$$

and

$$\int_{x \in S_1} P_x(S_2)dx \geq \int_{x \in S_1} P_x(S_2 \backslash S_2')dx = \int_{y \in S_2 \backslash S_2'} P_y(S_1)dy \geq \frac{1}{2n}\text{vol}(S_2 \backslash S_2'). \tag{7}$$

Thus, from equations (6) and (7),

$$\begin{aligned}\int_{x \in S_1} P_x(S_2)dx &\geq \frac{1}{2} \cdot \frac{1}{2n} \cdot (\text{vol}(S_1 \backslash S_1') + \text{vol}(S_2 \backslash S_2')) = \frac{1}{4n} \cdot \text{vol}(S_3') \\ &\geq \frac{s}{400 \cdot 10^3 \cdot Rn^{4.5} \log n} \cdot \left(\min\{\text{vol}(S_1'), \text{vol}(S_2')\} - \frac{s}{2} \cdot \text{vol}(K)\right) \\ &\geq \frac{s}{800 \cdot 10^3 \cdot Rn^{4.5} \log n} (\min\{\text{vol}(S_1), \text{vol}(S_2)\} - s \cdot \text{vol}(K)) \\ &\geq \frac{s}{8 \cdot 10^5 \cdot Rn^{4.5} \log n} (\text{vol}(S_1) - s \cdot \text{vol}(K)) \\ &= \frac{\text{vol}(K)s}{8 \cdot 10^5 \cdot Rn^{4.5} \log n} (\pi_K(S_1) - s) \\ \Rightarrow \int_{S_1} P_x(K \backslash S_1)d\pi_K(x) &\geq \frac{s}{8 \cdot 10^5 \cdot Rn^{4.5} \log n} (\pi_K(S_1) - s).\end{aligned}$$

Thus, for any $S_1 \subseteq K$ with $s < \pi_K(S_1) \leq 1/2$, we get

$$\frac{p(S_1)}{\pi_K(S_1) - s} \geq \frac{\int_{x \in S_1} P_x(K \backslash S_1)d\pi_K(x)}{\pi_K(S_1) - s} \geq \frac{s}{8 \cdot 10^5 \cdot Rn^{4.5} \log n},$$

which implies that the s-conductance of the CHAR Markov chain is

$$\phi_s \geq \frac{s}{8 \cdot 10^5 \cdot Rn^{4.5} \log n}.$$ ◀

Proof of Theorem 1. For a convex body K, let π_0 be the starting distribution and π_t be the distribution after t steps of CHAR. Assume that π_0 is M-warm with respect to π_K. From Theorem 4 and selecting $s = \frac{\epsilon}{2M}$, we get

$$d_{TV}(\pi_t, \pi_K) \leq \frac{\epsilon}{2} + M\left(1 - \frac{\phi_s^2}{2}\right)^t.$$

Combining this with Theorem 13,

$$t = \frac{4}{\phi_s^2} \log \frac{2M}{\epsilon} = O\left(\frac{M^2 R^2 n^9 \log^2 n}{\epsilon^2} \log \frac{2M}{\epsilon}\right)$$

steps of the CHAR Markov chain suffice to ensure $d_{TV}(\pi_t, \pi_K) \leq \epsilon$. ◀

4 Lower bound

▶ **Theorem 14.** *For a convex body K in $\mathbb{R}^n$ with diameter D and containing a unit ball, the conductance of Coordinate Hit-and-Run is $O(1/n^2 D)$.*

Proof. Fix a simplex C in $\mathbb{R}^{n-1}$ with center of gravity at zero containing a unit ball. We construct a convex body K in $\mathbb{R}^n$ so that $K(x_1)$, the slice of K with the first coordinate x_1, is $C + (x_1, 0, \ldots, 0)$ for $x_1 \in [0, D]$ and empty outside this range of x_1. We choose $D \geq 2n$. Let $S \subset K$ be the set of all points in K with $x_1 \leq D/2$. We now observe that the volume of axis-aligned extension of S, $\text{ext}_K(S)$ is bounded by $O(1/nD)$ times the volume of S. To see this, note that the shadow of the cross-section has volume that can be computed as

$$\int_0^h t^{n-2}(h-t) = h^{n-1}\left(\frac{1}{n-1} - \frac{1}{n}\right) = \frac{h^{n-1}}{n(n-1)} = \frac{A}{n}$$

where A is the area of the cross-section. This shows that the isoperimetric ratio is $O(1/nD)$. Next, we note that the extension of S goes beyond S only along e_1, and the probability that CHAR chooses e_1 at any step is only $1/n$. This gives a conductance bound of $O(1/(n^2 D))$. ◀

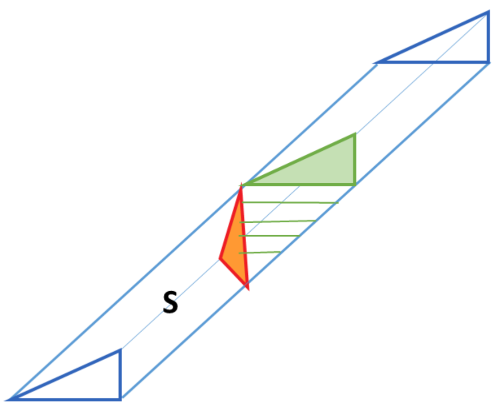

Figure 3 The lower bound construction.

We expect that this translates to a lower bound of $\widetilde{\Omega}(n^3D^2)$ on the mixing rate even from a warm start. Consider two subsets of K at opposite ends: $K \cap \{x : x_1 \leq D/4\}$ and $K \cap \{x : x_1 \geq 3D/4\}$. Suppose we start with a uniformly random point in the first set. Then in order to mix, the current point must reach the latter set. Even though this is worse than the $\widetilde{O}(n^2D^2)$ mixing rate of hit-and-run, it is an interesting open problem to determine the precise mixing rate of CHAR.

References

1. Hans C. Andersen and Persi Diaconis. Hit and run as a unifying device. *Journal de la société française de statistique*, 148(4):5–28, 2007. URL: `http://www.numdam.org/item/JSFS_2007__148_4_5_0`.
2. Arnon Boneh. Preduce – a probabilistic algorithm identifying redundancy by a random feasible point generator (rfpg). In *Redundancy in Mathematical Programming*, pages 108–134, Berlin, Heidelberg, 1983. Springer Berlin Heidelberg.
3. Ben Cousins and Santosh Vempala. Volume computation of convex bodies. *MATLAB File Exchange*, 2013. URL: `http://www.mathworks.com/matlabcentral/fileexchange/43596-volume-computation-of-convex-bodies`.
4. Ben Cousins and Santosh Vempala. A practical volume algorithm. *Mathematical Programming Computation*, 8(2):133–160, 2016.
5. Persi Diaconis, Kshitij Khare, and Laurent Saloff-Coste. Gibbs sampling, conjugate priors and coupling. *Sankhya A*, 72(1):136–169, 2010.
6. Persi Diaconis, Gilles Lebeau, and Laurent Michel. Gibbs/metropolis algorithms on a convex polytope. *Mathematische Zeitschrift*, 272(1-2):109–129, 2012.
7. Ioannis Z Emiris and Vissarion Fisikopoulos. Efficient random-walk methods for approximating polytope volume. In *Proceedings of the thirtieth annual symposium on Computational geometry*, pages 318–327, 2014.
8. Jenny Rose Finkel, Trond Grenager, and Christopher D Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 363–370, 2005.
9. Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741, 1984. `doi:10.1109/TPAMI.1984.4767596`.
10. Edward I George and Robert E McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.
11. R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.
12. R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms*, 11:1–50, 1997.
13. Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *Proc. of IEEE FOCS*, 2017.
14. L. Lovász. How to compute the volume? *Jber. d. Dt. Math.-Verein, Jubiläumstagung 1990*, pages 138–151, 1990.
15. L. Lovász. Hit-and-run mixes fast. *Math. Prog.*, 86:443–461, 1998.
16. L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. In *Random Structures and Alg.*, volume 4, pages 359–412, 1993.
17. L. Lovász and S. Vempala. Hit-and-run from a corner. *SIAM J. Computing*, 35:985–1005, 2006.
18. L. Lovász and S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307–358, 2007. `doi:10.1002/rsa.v30:3`.

19 László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random structures & algorithms*, 4(4):359–412, 1993.

20 Hariharan Narayanan and Piyush Srivastava. On the mixing time of coordinate hit-and-run. *arXiv preprint*, 2020. `arXiv:2009.14004`.

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

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

23 V. Turchin. On the computation of multidimensional integrals by the monte-carlo method. *Theory of Probability & Its Applications*, 16(4):720–724, 1971. `doi:10.1137/1116083`.