A survey on Hit-and-run algorithm

Siyu Liu

January 2026

1 Introduction

Problem: sample uniformly from a convex body K in $\mathbb{R}^n$

Algorithm (Hit-and-run):
It's a Markov Chain with transition:

- From a starting point X_t,
- sample uniformly a direction $\vec{u}$ from $\mathbb{S}^{n-1}$ (sphere in $\mathbb{R}^n$)
- sample X_{t+1} uniformly from $K \cap (X_t + \theta\vec{u}), \quad \theta \in \mathbb{R}$
- move to X_{t+1}.

Sometimes it's not that simple to sample from an arbitrary line.
So we have

Algorithm (coordinate Hit-and-run):
First fix a set of orthonormal basis $\{e_i\}_{i=1}^n$.
One step transition:

- sample uniform $i \leftarrow [n]$.
- sample X_{t+1} uniformly from $K \cap (X_t + \theta e_i), \quad \theta \in \mathbb{R}$
- move to X_{t+1}.

2 Previous Result:

The fundamental work mainly due to great mathematician Lászlo Lovász.

2.1 Upper bound

Theorem 1 (Lovász)**.** *For a convex body $K \subseteq \mathbb{R}^n$, $B_2 \subseteq K \subseteq R \cdot B_2$, where B_2 is ℓ_2-norm ball. Hit-and-run algorithm mixes in $\tilde{O}(n^2R^2)$ from a "warm-start" or from arbitrary interior point.*

To be specific, call distribution ν is M-warm w.r.t. π if the Radon-Nikodym derivative $\frac{d\nu}{d\pi}$ is pointwise bounded $\leq M$. If x_0 is sampled from a M-warm starting distribution w.r.t. uniform, then Hit-and-run gets ε-approximate to uniform in Total Variance after $\tilde{O}\left(n^2R^2 \log \frac{M}{\varepsilon}\right)$ steps.

Furthermore, if x_0 is any interior point at distance d from the boundary, then Hit-and-run gets ε-approximate to uniform in total variance after $\tilde{O}\left(n^3R^2 \log \frac{R}{d\varepsilon}\right)$ steps.

Later, Lovász generalized it to sampling any log-concave distribution.

First we shall generalize the algorithm.
Now we aim to sample from given distribution π on $\mathbb{R}^n$.

Hit-and-run(Gibbs sampling version)
one step transition:

- sample a direction $\vec{u}$ uniformly from S^{n-1}
- sample X_{t+1} with density $\pi_{\vec{u}}(x) \propto \pi(x \mid x \in \{X_t + \theta\vec{u}\})$
- move to X_{t+1}

Theorem 2 (Lovász)**.** *Let f be a c-isotropic log-concave density function in $\mathbb{R}^n$. Hit-and-run gets ε-approximate to f's distribution in total variance after $O\left(\frac{c^4M^4n^3}{\varepsilon^4}\log^3\frac{2M}{\varepsilon}\right)$ steps from a M-warm start,*

where c-isotropic refers to $\frac{1}{c} \leq \int_{\mathbb{R}^n}(v^Tx)^2 f(x)dx \leq c$ for every unit vector v and $\int_{\mathbb{R}^n} xf(x)dx = 0$.

2.2 Lower bound

Recall

Spectral gap γ is defined as $(1 - \text{second largest eigenvalue } \lambda_2)$ for reversible Markov chain P. Formally,

$$\lambda_2 := \sup_{\substack{\|g\|_\pi=1 \\ E_\pi[g]=0}} \langle Pg, g\rangle_\pi$$

Also, we have

$$\gamma = \inf_g \frac{\mathcal{E}(g,g)}{\mathrm{Var}_\pi(g)}$$

where

$$\begin{aligned}\mathcal{E}(f,g) &= \langle f, Lg\rangle_\pi \quad (L = I - P \text{ is Laplacian.})\\ &= \frac{1}{2}\iint (f(x)-f(y))(g(x)-g(y))P_{x\to dy}\pi(dx)\end{aligned}$$

is the Dirichlet form.

The conductance $\Phi(P)$ is defined by $\inf_{\pi(S)\le\frac{1}{2}} \frac{\int_S P_{x\to S^c}\pi(dx)}{\pi(S)}$

Lemma 1 (Spectral Gap theorem)**.** *Mixing time of reversible Markov chain P has lower bound*

$$\tau(\varepsilon) \ge \frac{1-2\varepsilon}{2}\frac{1}{\Phi(P)}$$

Recall Cheeger's inequality: $\frac{\gamma}{2} \le \Phi(P) \le \sqrt{2\gamma}$.
So $\tau(\varepsilon) \ge \frac{1-2\varepsilon}{2}\cdot\frac{1}{\sqrt{2\gamma}}$.

Observation 1. *Consider a cylinder $B_2^{n-1}\times[-D,D]$. Hit-and-run mixing time is lower bounded by $\Omega(n^2D^2)$. The intuition is that each step projected to the axis of the cylinder is $\Theta(\frac{1}{n})$. So it takes $\Omega(n^2D^2)$ for this random walk to cover the whole axis.*

This result means Lovász's analysis is optimal in general setting.

2.3 Further work

There're several directions for improvements.

1. The cylinder counterexample is trivially bad, namely it's not isotropic at all. Then uniformly choose a direction in hit-and-run algorithm will be inappropriate. Can we give more tight analysis for isotropic distribution? (see "Hit-and-run mixing via localization schemes")

2. A higher point of view for hit-and-run is Gibbs sampling, roughly each time choose a subspace and update this subspace according to its marginal probability while maintaining the others. Can we generalize it to Gibbs sampling? (see "Entropy contraction of Gibbs sampler under log-concavity")

3. Coordinate hit-and-run is practically faster than hit-and-run in industry. However, neither upper nor lower bound was established before. Can we use the similar method to analyze Coordinate hit-and-run? (see "On the mixing time of coordinate Hit-and-Run")

We dive into each in order.

3 "Hit-and-run mixing via localization schemes"

3.1 Main result:

- Cheeger's isoperimetric constant of a measure ν:

$$\psi_\nu := \inf_{A \subseteq \mathbb{R}^n} \frac{\int_{\partial A} \nu}{\min\{\nu(A), 1-\nu(A)\}}$$

- A measure ν is **isotropic** if $\mathbb{E}_{X\sim\nu}[X] = 0$ and $\mathrm{Var}_{X\sim\nu}[X] = I_n$.
  A convex body is **isotropic** if the uniform distribution on it is isotropic.

Theorem 3. *Let μ be uniform distribution on isotropic convex body $K \subset \mathbb{R}^n$. Then there exists constant $C, c > 0$ s.t. for $n \geq c \log \frac{M}{\varepsilon}$, from a M-warm start initial distribution μ_{init}, the ε-mixing time of hit-and-run is:*

$$\leq C \cdot \frac{n^2}{\psi_n^2} \left(\frac{M}{\varepsilon}\right)^{11} \log^5 \frac{M}{\varepsilon} = \tilde{O}\left(\frac{n^2}{\psi_n^2}\right)$$

where $\psi_n = \inf_{\substack{\nu \text{ isotropic, log-concave} \\ \text{on } \mathbb{R}^n}} \psi_\nu$.

Remark 1. For isotropic convex K, "diameter" R is "typically" $\tilde{O}(\sqrt{n})$.

By "typically", we mean all but a fraction of ε of K is contained in a ball of $2\sqrt{2n \log \frac{1}{\varepsilon}}$. The outside part is negligible within ε total variance from uniform.

In fact, we remark by the way here that for general convex body K, we can also find an affine transformation converting it to a near isotropic one, giving R an upper bound $\tilde{O}(\sqrt{n})$ in consequent hit-and-run.

We give a heuristic statement of it here.

Start from $i = 0$. $K_0 = B_2 \subseteq K$. $K_i = K \cap 2^{\frac{i}{n}} B_2$.

K_0 is isotropic, so we can sample efficiently.

Provided that K_i is 2-isotropic, then K_{i+1} is 6-isotropic.

Sample $\tilde{O}(n)$ points uniformly from K_{i+1} to approximate the Variance of K_{i+1}.

Given variance Σ of K_{i+1}, $x \mapsto \Sigma^{-\frac{1}{2}} x$ (affine transformation) brings K_{i+1} to be isotropic. An estimate of Σ can give an affine transformation bringing K_{i+1} to be 2-isotropic. Continue the iteration until $i = O(n \log R)$. $K = K \cap RB = K_{n \cdot \log R}$ is 2-isotropic after affine transformation.

Remark 2. $\psi_n \gtrsim \frac{1}{\log^5(n)}$. So the result improved $\tilde{O}(n^2 R^2)$ to $\tilde{O}(n^2)$.

3.2 Proof technique: (Localization scheme)

Remark. It's a very brief sketch of what localization scheme does.

We first introduce a powerful conclusion in terms of bounding mixing time, which we'll use many times.

Theorem 4 (Lovász, Simonovits). *Let $S \subseteq \mathbb{R}^n$ with finite measure. P be a Markov chain on S, μ_0 is initial distribution, π is stationary. Assume π is atom-free, i.e. $\pi(\{u\}) = 0 \quad \forall u \in S$.*
Fix $0 < s < \frac{1}{2}$, let $H_s := \sup_{A \subseteq S : \pi(A) \leq s} (\mu_0(A) - \pi(A))$.
The s-conductance $\bar{\Phi}_s := \inf_{A : s < \pi(A) \leq \frac{1}{2}} \frac{\int_A P_{x \to A^c} \pi(dx)}{\pi(A) - s}$.
μ_k is the distribution after k step Markov transition.
Then $\forall k$,

$$d_{TV}(\mu_k, \pi) \leq \left(1 + \frac{(1 - \Phi_s^2/2)^k}{s}\right) H_s$$

In particular, for a M-warm start μ_0, $H_s \leq s \cdot M$, and

$$d_{TV}(\mu_k, \pi) \leq M \cdot s + M \left(1 - \frac{\Phi_s^2}{2}\right)^k$$

We lower bound Φ_s by localization scheme.
Basic idea: Hit-and-run finds it difficult to get rid of the boundary; When trapped in boundary, movement has low probability. Uniform distribution has rather large probability in boundary. We consider turn to a more interiorly "localized" distribution.

Stochastic localization process (SL): defined as

$$\mu_t(x) \propto \exp\left(c_t^T x - \frac{1}{2} x^T B_t x\right) \mu_0(x)$$

where

$$\begin{aligned} dc_t &= C_t^{1/2} dW_t + C_t b(\mu_t) dt \\ dB_t &= C_t dt \\ b(\mu_t) &= \int x \mu_t(x) dx \end{aligned}$$

C_t, called control matrix, is positive semi-definite.
Here we take $G_t = I_n$ for all t.
It turns out μ_t is a truncated Gaussian.
A **truncated Gaussian on** K is with density

$$\nu_{\beta,m}(x) \propto e^{-\frac{m}{2}\|x - \beta\|^2} \mathbf{1}_K.$$

We have

$$\mu_t = \nu_{\frac{c_t}{t}, t}$$

It turns out $\frac{c_t}{t}$ is an interior point with large probability, resulting in convenient analysis for s-conductance on μ_T.

It remains to be shown s-conductance on $\mu_T \leq s$-conductance on μ_0. This is shown by proving Dirichlet form

$$D_t := \mathcal{E}_{P_{\bullet\to\bullet}(\mu_t)}(f, f)$$

is a super-martingale.

Note $\int_S P_{x\to S^c}\mu(dx) = \mathcal{E}(\mathbf{1}_S, \mathbf{1}_S)$.

So $\int_S P_{x\to S^c}\mu_0(dx) \geq \mathbb{E}\left[\int_S P^{(\mu_T)}_{x\to S^c}\mu_T(dx)\right]$ via super-martingale.

We still need so called "approximate conservation of variance", stating that for E with $\mu_0(E)$ away from 0 and 1, so does $\mu_t(E)$.

A combination of these ingredients give lower bound of s-conductance.

4 "Entropy contraction of Gibbs sampler under log-concavity"

Remark. I didn't get much intuition from the proof technique other than the beautiful deduction, so I omit it here.

4.1 Algorithm (Random Scan Gibbs Sampler)

Goal: sample from π on $\mathbb{R}^d = \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_M$.
Transition:

- sample $m \in [M]$ uniformly.
- sample $X^{(t+1)}$'s m-th dimension $X_m^{(t+1)}$ from its marginal $\pi(\cdot \mid X_{-m}^{(t)})$, where $X_{-m}^{(t)} = (X_1^{(t)}, \ldots, X_{m-1}^{(t)}, X_{m+1}^{(t)}, \ldots, X_M^{(t)})$.
- update $X^{(t)}$'s m-th dimension with $X_m^{(t+1)}$, remain the other the same.

4.2 Requirement for π:

Requirement A: π has density $\exp(-U)$ and $U : \mathbb{R}^d \to \mathbb{R}$ is

(i) L-smooth: $\nabla U : \mathbb{R}^d \to \mathbb{R}^d$ is L-Lipschitz.

(ii) λ-convex: $U(x) - \frac{\lambda}{2}\|x\|^2$ is convex.

$\kappa = L/\lambda$ be the condition number.

Requirement B: π has density $\exp(-U)$ and U decomposes as

$$U(x) = U_0(x) + \sum_{m=1}^{M} U_m(x_m)$$

where $U_m : \mathcal{X}_m \to \mathbb{R}$ is convex for $m = 1, \ldots, M$. $U_0 : \mathbb{R}^d \to \mathbb{R}$ satisfies

(i) $\mathcal{X}_m \mapsto U_0(\dots, x_m, \dots)$ is L_m-smooth for every $m = 1, \dots, M$ and every assignment to "..." .

(ii) $U_0(x) - \frac{\lambda^*}{2} \sum_{m=1}^{M} L_m \|x_m\|^2$ is convex.
$\kappa^* = \frac{1}{\lambda^*}$ be coordinate-wise condition number.

Lemma 2. *A implies B with* $1 \leq \kappa^* \leq \kappa$*.*

4.3 Main Result:

Theorem 5. *Let* π *satisfy Assumption B. Then*

$$KL(\mu P^{GS} \| \pi) \leq \left(1 - \frac{1}{\kappa^* M}\right) KL(\mu \| \pi)$$

for every μ*, where* P^{GS} *is Gibbs sample Markov transition kernel.*

Remark 3. This implies a mixing time of $\tilde{O}(\kappa^* M \log \frac{1}{\varepsilon})$.

Remark 4. The theorem is tight, by π taken as Gaussian and analyzing spectral gap.

Theorem 6 (non-strongly convex case)**.** *Let* π *satisfy Assumption B but* $\lambda^* = 0$*. Let* x^* *be minimum point of* U*.*

$$R^2 = 2 \sup_{n \geq 0} \int_{\mathbb{R}^d} \left(\sum_{m=1}^{M} L_m \|x_m - x_m^*\|^2 \right) \mu^{(n)}(dx)$$

Then

$$KL(\mu^{(n)} \| \pi) \leq \frac{2M \max\{KL(\mu^{(0)} \| \pi), R^2\}}{n + 2M}$$

Remark 5. R is analogous to "diameter" in convex body sampling. It implies mixing time of $\tilde{O}(\frac{M}{\varepsilon})$.

Corollary 1. *Let* π *satisfy Assumption A. Then*

$$KL(\mu P^{HR} \| \pi) \leq \left(1 - \frac{1}{\kappa d}\right) KL(\mu \| \pi)$$

where P^{HR} *is transition kernel of Hit-and-run on* $\mathbb{R}^d$*.*

5 "On the mixing time of coordinate Hit-and-Run"

Remark. I'd like to dive into proof details a little bit here.

5.1 Main result:

Theorem 7. *Let $K \subset \mathbb{R}^n$ be a closed convex body s.t.*

$$B_\infty \subseteq K \subseteq R \cdot B_\infty, \quad B_\infty \text{ is } L_\infty\text{-norm ball.}$$

Starting from M-warm distribution, coordinate Hit-and-Run gets ε-approximation from uniform in total variance after

$$k = \tilde{O}\left(\frac{n^7 R^4 M^4}{\varepsilon^4}\right) \text{ steps.}$$

5.2 Proof technique:

Basic idea is a mimick to Lovász's proof of Hit-and-run. To lower bound s-conductance $\bar{\Phi}_s$, Lovász introduce an "overlap property" of Hit-and-run, roughly stating that when $u, v \in K$ are "close", the probability distributions generated by one-step transition from u and from v are not too far.

However, overlap property does not hold for coordinate Hit-and-run: If $u - v$ has nonzero component along all basis, after one step transition, resulting distributions have disjoint support.

However, an auxiliary "Gaussian walk", behaving not so differently from coordinate Hit-and-run, turns out to have overlap property.

Definition 1 (overlap property)**.** *Let K be a convex body in $\mathbb{R}^n$. A Markov scheme P on K is called to have $(\varepsilon, \delta, \eta)$-overlap property w.r.t. $K' \subseteq K$ if*

(i) $vol(K') \geq vol(K)(1 - \varepsilon)$

(ii) $\forall \|u - v\|_2 \leq \delta,\ u, v \in K'$, we have $d_{TV}(P(u, \cdot), P(v, \cdot)) \leq 1 - \eta$

Definition 2 (Gaussian walk)**.** *Given X_t.*
Sample $i \in [n]$ uniformly, $\kappa \sim N(0, \sigma^2)$
Let $u = X_t + \kappa e_i$. If $u \in K$, $X_{t+1} := u$.
If not, reject the proposal, $X_{t+1} = X_t$.
Denote $G_v^{(\tau)}$ the distribution of X_τ given $X_0 = v$.

Definition 3 (Robust interior)**.** $K_\varepsilon := \{x \in K : \forall v, \|v\|_\infty \leq \varepsilon \Rightarrow x + v \in K\}$

Proposition 1. *For $\varepsilon \in (0, 1)$, K is a closed convex body with $B_\infty \subseteq K$. Then $(1 - \varepsilon)K \subseteq K_\varepsilon$. In particular, $vol(K_\varepsilon) \geq (1 - \varepsilon)^n vol(K)$.*

Lemma 3. *$G^{(t)}$, $t = \lceil 20n \ln n \rceil$ steps regarded as one step transition, has $(\varepsilon, \sigma, \frac{1}{4})$-overlap property w.r.t. $K_{100\sigma \ln n}$.*

It's still not convenient to prove directly.

We approximate $G^{(\tau)}$.

Given a multi-index $\mathbb{I} = (i_1, \ldots, i_n) \in \mathbb{N}^n$, define $G_{v,\mathbb{I}}$ be the Gaussian centered at v and with covariance $\mathrm{diag}(i_j \sigma^2)$.

Let $M_{n,\tau} = \left\{(i_1, \dots, i_n) \mid \sum_{j=1}^n i_j = \tau\right\}$.
$\lambda_{\mathbb{I}} = \binom{\tau}{\mathbb{I}} n^{-\tau}$, where $\binom{\tau}{\mathbb{I}} = \frac{\tau!}{\prod_{j=1}^n i_j!}$ is the multinomial coefficient.

The intuition is to enumerate how many times Gaussian walk choose e_j and temporarily forget to reject the proposal.

Lemma 4. *Fix σ, $\tau = \lceil 20n \ln n \rceil$. Let $v \in K$ s.t. $v \in K_{100\sigma \ln n}$. Then*

$$d_{TV}\left(G_v^{(\tau)}, \sum_{\mathbb{I} \in M_{n,\tau}} \lambda_{\mathbb{I}} G_{v,\mathbb{I}}\right) \leq 2n^{-5}$$

Proof. Denote $\mathcal{H} = \sum_{\mathbb{I} \in M_{n,\tau}} \lambda_{\mathbb{I}} G_{v,\mathbb{I}}$.
Coupling between $G_v^{(\tau)}$ and $\mathcal{H}$: given pair (X_t, X_t'), pick $i \in [n]$ u.a.r. and $\kappa \sim N(0, \sigma^2)$.

- $X_{t+1} = X_t + \kappa e_i$ if the point is in K.
- $X_{t+1} = X_t$ o.w.

$X_{t+1}' = X_t' + \kappa e_i$.

Then

$$\begin{aligned} d_{TV}(G_v^{(\tau)}, \mathcal{H}) &\leq \mathbb{P}[X_\tau \neq X_\tau'] \\ &\leq \mathbb{P}[\exists t \leq \tau, \text{proposal at } t \text{ was rejected}] \end{aligned}$$

Suppose A_i is the number of times e_i was chosen in the above coupling.

$$\mathbb{P}[\exists i. A_i \geq 50 \ln n] \leq n \cdot \exp\left(-\frac{\tau}{3n}\right) \leq n^{-5}$$

By Gaussian tail bounds,

$$\begin{aligned} &\mathbb{P}[\exists t.\text{ proposal at } t \text{ was rejected} \mid A_i \leq 50 \ln n \ \forall i] \\ &\leq \sum_{j=1}^n \sum_{s=1}^\tau \exp\left(-\frac{100^2 \sigma^2 \ln^2 n}{100\sigma^2 \ln n}\right) \leq n^{-90} \end{aligned}$$

So $\mathbb{P}[\exists t.\text{ proposal at } t \text{ was rejected}] \leq 2n^{-5}$.

□

Bridging via $G_{v,\mathbb{I}}$, the overlap-property can be proved.
The calculation is less of my interest.

Next, we use overlap property to deduce s-conductance.

Lemma 5. *Let $\varepsilon, \delta, \eta \in (0, \frac{1}{2})$. A reversible Markov scheme P has $(\varepsilon, \delta, \eta)$-overlap property w.r.t. K'.*
Then $\forall S_1, S_2$ a partition of K,

$$\begin{aligned}\Phi_\pi(S_1) &\geq \frac{\eta\delta}{4(R-\delta)} \min\{\pi(K' \cap S_1), \pi(K' \cap S_2)\} \\ &\geq \frac{\eta\delta}{4(R-\delta)} \min\{\pi(S_1) - \varepsilon, \pi(S_2) - \varepsilon\}\end{aligned}$$

Thus ε-conductance $\bar{\Phi}_s \geq \frac{\eta\delta}{4(R-\delta)}$.

Proof. Since $\pi(K') \geq 1 - \varepsilon$, we have

$$\pi(K' \cap S_1) \geq \pi(S_1) - \varepsilon, \quad \pi(K' \cap S_2) \geq \pi(S_2) - \varepsilon.$$

Consider

$$\begin{aligned}S_1' &:= \{x \in K' \cap S_1 : P(x, S_2) < \frac{\eta}{2}\} \\ S_2' &:= \{x' \in K' \cap S_2 : P(x', S_1) < \frac{\eta}{2}\}\end{aligned}$$

By overlap property, $\mathrm{dist}(S_1', S_2') > \delta$.

$$\begin{aligned}\Phi(S_1, K' \cap S_2) &= \Phi(K' \cap S_2, S_1) \\ &= \int_{K' \cap S_2} P(u, S_1)\pi(du) \\ &\geq \int_{(K' \cap S_2) - S_2'} P(u, S_1)\pi(du) \geq \frac{\eta}{2}\pi((K' \cap S_2) - S_2')\end{aligned}$$

Combining with an isoperimetric inequality stated below, we finished proof, whose detail is less of my interest. □

Theorem 8. *Let $\delta > 0$. Fix any ℓ-norm $\|\cdot\|$. K be a convex body, and $K_1, K_2 \subseteq K$ with dist$(K_1, K_2) > \delta$. Let the diameter of $K \leq D$. Then*

$$vol(K - (K_1 \cup K_2)) \geq \frac{2\delta}{D - \delta} \min\{vol(K_1), vol(K_2)\}$$

Finally,

Lemma 6.

$$\Phi_{\textit{Coordinate Hit-and-run}}(S, K - S) \geq \frac{\sigma\sqrt{\pi}}{R\sqrt{2}} \Phi_{\textit{Gaussian walk}}(S, K - S)$$

Proof. The proof is directly from the Radon-Nikodym derivative between them $\geq \frac{\sqrt{2\pi}\sigma}{2R}$. □

Now we finishing lower bounding s-conductance.

6 Future work

A simple lower bound of coordinate Hit-and-run is also by cylinder, which is $\Omega(n)$. There's a large gap here.

Reusing Gibb's sampler's proof is promising since they are quite similar.

However, there's a tricky difference between them. The setting of "Gibbs sampler" require $\log \pi \in C^1$, which not contain sampling in convex body.

Whether two settings can be unified is also open.