Advanced Topics SP25 Final Project: The Hit-and-Run Sampler and a Coordinate-Free Randomized Kaczmarz Algorithm

Kaiwen Zhang

April 2025

1 Introduction

When sampling a multi-dimensional distribution, Gibbs sampling relies on a chosen coordinate system. If the coordinate system is not well aligned with the geometry of the target, the samples may take a long time to mix. This issue motivates the Hit-and-Run sampler, a coordinate-free alternative. The idea is that given the previous sample, one selects a direction on the unit sphere, and let the chain "run" in this chosen direction. The coordinate-free nature allows Hit-and-Run to make "global moves". In our paper [1], the authors rigorously quantify this advantage by proving an estimate on the contraction rate of Hit-and-Run, in the case where the target density is a centered Gaussian. The paper also discusses, as an extension, a coordinate-free version of the randomized Kaczmarz algorithm.

2 Formulation of the Hit-and-Run Sampler

2.1 Algorithm

Suppose the target density is $\mu(x) : \mathbb{R}^d \to \mathbb{R}$. Denote $\mathbb{S}^{d-1}$ as the unit sphere in $\mathbb{R}^d$. Given sample X^k:

- Sample a direction $v \sim \text{Unif}(\mathbb{S}^{d-1})$
- Generate a scalar displacement $H \sim \mu_{x,v}$, where $\mu_{x,v}$ is a measure on $\mathbb{R}$ with density $\mu_{x,v}(h) \propto \mu(x+hv)$
- Update $X^{k+1} = X^k + Hv$

In effect, $X^{k+1} \sim \mu(\cdot|L(X^k, v)) = \mu_{X^k,v} \circ \Delta^{-1}_{X^k,v}$. Here, $L(x, v)$ is the line passing x in direction v. $\mu(\cdot|L(x, v))$ is the regular conditional distribution of μ given the line $L(x, v)$, and $\Delta_{x,v} : \mathbb{R} \to \mathbb{R}^d, \delta \mapsto x + \delta v$ maps a scalar displacement along the line to its corresponding position in $\mathbb{R}^d$.

2.2 Transition Kernel

The transition probabilities can be expressed as

$$p(x, A) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \mathbf{1}_A(x + \delta v)\mu_{x,v}(d\delta)\sigma(dv) \tag{1}$$

where σ is the rotation invariant measure on the $d-1$ dimensional sphere. One can generalize the algorithm by replacing σ with a generic measure on the sphere. The transition kernel satisfies detailed balance with respect to μ, i.e.

$$\int_A p(x, B)\mu(dx) = \int_B p(y, A)\mu(dy)$$

and the proof involves a computation with changes of variables. It relies on the expression for the conditional distribution $\mu_{x,v}(d\delta) = \frac{\mu(x+\delta v)}{\int \mu(x+sv)ds}$ and the fact that if $x \in \mathbb{R}^d, x + \delta v \in A$, then $x \in A - \delta v$.

Suppose now the target density μ is that of a centered Gaussian $N(0, C)$ where C is the covariance matrix. Then

$$\mu(x) = \frac{1}{\sqrt{(2\pi)^d|C|}} e^{-\frac{1}{2}\langle x, C^{-1}x\rangle}$$

Conditional to the line passing x in direction v, the scalar displacement H follows a density proportional to $\mu(x+hv)$, which in this case is

$$\begin{aligned}\mu(x+hv) &\propto \exp\left(-\frac{1}{2}\left\langle x+hv, C^{-1}(x+hv)\right\rangle\right) = \exp\left[-\frac{1}{2}\left\langle x, C^{-1}x\right\rangle - \frac{1}{2}h^2\left\langle v, C^{-1}v\right\rangle - h\left\langle x, C^{-1}v\right\rangle\right] \\ &= \exp\left[-\frac{1}{2}\left\langle x, C^{-1}x\right\rangle - \frac{1}{2}\left\langle v, C^{-1}v\right\rangle\left(h^2 + 2h\frac{\left\langle x, C^{-1}v\right\rangle}{\left\langle v, C^{-1}v\right\rangle}\right)\right] \\ &= \exp\left[-\frac{1}{2}\left\langle v, C^{-1}v\right\rangle\left(h + \frac{\left\langle x, C^{-1}v\right\rangle}{\left\langle v, C^{-1}v\right\rangle}\right)^2\right]\cdot \text{const}\end{aligned}$$

where the constant comes from $\left\langle x, C^{-1}x\right\rangle$ and completing the squares, and depends on x, v, C but not h. Thus in fact

$$H \sim N\left(-\frac{\left\langle x, C^{-1}v\right\rangle}{\left\langle v, C^{-1}v\right\rangle}, \left\langle v, C^{-1}v\right\rangle^{-1}\right), \text{ or } H = -\frac{\left\langle x, C^{-1}v\right\rangle}{\left\langle v, C^{-1}v\right\rangle} + \frac{Z}{\sqrt{\left\langle v, C^{-1}v\right\rangle}}, Z \sim N(0,1) \tag{2}$$

We will use this formulation in numerical experiments.

One can study the scheme under the coordinate change $x \mapsto C^{-1/2}x$, where $(C^{-1/2})^T C^{-1/2} = C^{-1}$. Given previous sample X^k, knowing that $X^{k+1} = X^k + Hv$ and using formula 2, we have

$$C^{-1/2}X^{k+1} = C^{-1/2}(X^k + Hv) = \left(I - \frac{(C^{-1/2}v)(C^{-1/2}v)^T}{|C^{-1/2}v|^2}\right)C^{-1/2}X^k + \frac{(C^{-1/2}v)(C^{-1/2}v)^T}{|C^{-1/2}v|^2}C^{-1/2}X^k \tag{3}$$

$$+\left(-\frac{\left\langle X^k, C^{-1}v\right\rangle}{\left\langle v, C^{-1}v\right\rangle} + \frac{Z}{\sqrt{\left\langle v, C^{-1}v\right\rangle}}\right)C^{-1/2}v \tag{4}$$

$$= \left(I - \frac{(C^{-1/2}v)(C^{-1/2}v)^T}{|C^{-1/2}v|^2}\right)C^{-1/2}X^k + Z\frac{C^{-1/2}v}{|C^{-1/2}v|} \tag{5}$$

Hence we see that after selecting a direction v, the algorithm "resamples" the component of vector $C^{-1/2}X^k$ in the direction $C^{-1/2}v$, with a standard Gaussian determining the updated magnitude and sign.

2.3 Contraction Properties

We focus on the case of centered Gaussian densities. In proving contraction, the authors use an observation on random projections:

Lemma 1. *η is a probability measure on $\mathbb{R}^d$. For $w \in \mathbb{R}^d$ define the projection onto the orthogonal complement of w as*

$$\Pi_w = I - \frac{ww^T}{|w|^2}$$

Then, for any $z \in \mathbb{R}^d$,

$$\mathbb{E}_{w\sim\eta}|\Pi_w z|^2 \le \left[1 - \inf_{|\zeta|=1}\mathbb{E}_{w\sim\eta}\left(\left\langle \zeta, \frac{w}{|w|}\right\rangle^2\right)\right]|z|^2 \tag{6}$$

Proof: consider first $w \in \mathbb{R}^d, |w| = 1$. Because $\Pi_w z$ is orthogonal to span of w,

$$|\Pi_w z|^2 = |z - \langle z, w\rangle w|^2 = |z|^2 - \langle z, w\rangle^2 = \left(1 - \left\langle \frac{z}{|z|}, w\right\rangle^2\right)|z|^2$$

z is deterministic, so taking expectation yields, for $|w| = 1$,

$$\mathbb{E}_{w\sim\eta}|\Pi_w z|^2 \le \left(1 - \mathbb{E}_{w\sim\eta}\left\langle \frac{z}{|z|}, w\right\rangle\right)|z|^2 \le \left[1 - \inf_{|\zeta|=1}\mathbb{E}_{w\sim\eta}\left(\langle\zeta, w\rangle^2\right)\right]|z|^2$$

Then for general w, noticing $\Pi_w = \Pi_{\frac{w}{|w|}}$ and plugging in the expression above finish the proof.

Using this lemma, the authors show the following contraction estimate concerning the Wasserstein distance between the transition probability distributions starting from two different points in $\mathbb{R}^d$. In this lemma, we will use the Wasserstein distance with respect to the metric on $\mathbb{R}^d$ induced by $C^{-1/2}$, i.e. $|x|_{C^{-1/2}} = |C^{-1/2}x|$.

Lemma 2. *For $x, y \in \mathbb{R}^d$, denote $p(x,\cdot), p(y,\cdot)$ as transition probabilities starting from x and y respectively. τ being the distribution from which we sample direction v,*

$$W^2_{C^{-1/2}}(p(x,\cdot),p(y,\cdot)) \leq (1-\rho)|x-y|_{C^{-1/2}}, \text{ where } \rho = \frac{1}{2}\inf_{|\zeta|=1}\mathbb{E}_{v\sim\tau}\left[\left\langle \zeta, \frac{C^{-1/2}v}{|C^{-1/2}v|}\right\rangle^2\right] \tag{7}$$

Proof: given $x, y \in \mathbb{R}^d$, one way to couple the distributions $p(x,\cdot), p(y,\cdot)$ is to have them use the same random variables $v \sim \text{Unif}(\mathbb{S}^{d-1}), Z \sim N(0,1)$. Then, using computations in 5, one sees that, in the notation of lemma 1 if $X \sim p(x,\cdot), Y \sim p(y,\cdot)$,

$$C^{-1/2}(X-Y) = \Pi_{C^{-1/2}v}C^{-1/2}(x-y)$$

Denote $z = C^{-1/2}(x-y)$ as in lemma 1. The Wasserstein distance $W^2_{C^{-1/2}}(p(x,\cdot),p(y,\cdot))$ concerns optimal coupling between X and Y, so we can bound this optimal coupling distance by our particular choice of coupling. Taking expectation yields:

$$W^2_{C^{-1/2}}(p(x,\cdot),p(y,\cdot)) \leq \sqrt{\mathbb{E}[|C^{-1/2}(X-Y)|^2]} = \sqrt{\mathbb{E}[|\Pi_{C^{-1/2}v}z|^2]}$$
$$\leq \sqrt{\left[1 - \inf_{|\zeta|=1}\mathbb{E}_{v\sim\tau}\left(\left\langle \zeta, \frac{C^{-1/2}v}{|C^{-1/2}v|}\right\rangle^2\right)\right]|z|^2} \leq \left[1 - \frac{1}{2}\inf_{|\zeta|=1}\mathbb{E}_{v\sim\tau}\left(\left\langle \zeta, \frac{C^{-1/2}v}{|C^{-1/2}v|}\right\rangle^2\right)\right]|z|$$

We used $\sqrt{1-x} \leq 1 - \frac{1}{2}x$ in the last estimate. Recalling $|x|_{C^{-1/2}} = |C^{-1/2}x|$ and $z = C^{-1/2}(x-y)$ finishes the proof. Moreover, the authors showed via computations that in fact

Lemma 3. *Any other constant ρ' satisfying estimate 7 must also satisfy $\rho' \leq 3\rho$.*

The authors also showed a total variation distance bound between $p(x,\cdot), p(y,\cdot)$ through highly technical computations.

3 Numerical Experiments re. the Hit-and-Run Sampler

3.1 2D Correlated Gaussian

Consider a bivariate normal distribution centered at the origin, where both coordinates have variance 1, with correlation coefficient ρ. To use expression 2 in implementing Hit-and-Run, we compute that the covariance matrix satisfies

$$C = \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}, C^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}, C^{-1/2} = \frac{1}{\sqrt{1-\rho^2}}\begin{bmatrix}\sqrt{\frac{1-\rho}{2}} & \sqrt{\frac{1-\rho}{2}}\\ \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1+\rho}{2}}\end{bmatrix}$$

In this case, a Gibbs sampler that scans each coordinate in order generates a Markov chain such that given $X^k = (X_1^k, X_2^k)$,

$$X_1^{k+1} \sim N(\rho X_2^k, 1-\rho^2), X_2^{k+1} \sim N(\rho X_1^{k+1}, 1-\rho^2)$$

We select $\rho = 0.75$, and generate a chain of length 1,000,000 using a Gibbs sampler and a Hit-and-Run sampler, respectively. To visualize the "moves" that the chains take, we purposefully start the chain at $(-10,-10)$ and record the first 10 updates of the chains. Scatterplots of the samples and the trajectories are shown in figures 1 and 2. For the Gibbs sampler, we interpret each "update" as a scan of two coordinates. Hence, for each "update", the Gibbs chain "moves" twice, once in x and once in y.

From the figures, we see that the Gibbs chain moves in a zigzag path. In comparison, because of its coordinate-free nature, the Hit-and-Run chain is sometimes able to make more ambitious global moves in the correct direction. We also note that for both samplers, more often than not the chains make local moves, especially when they reach regions where the target density is high.

Next we test the efficiency of these samplers by looking at IATs. We compute the IAT for two observables: given sample X, we study:

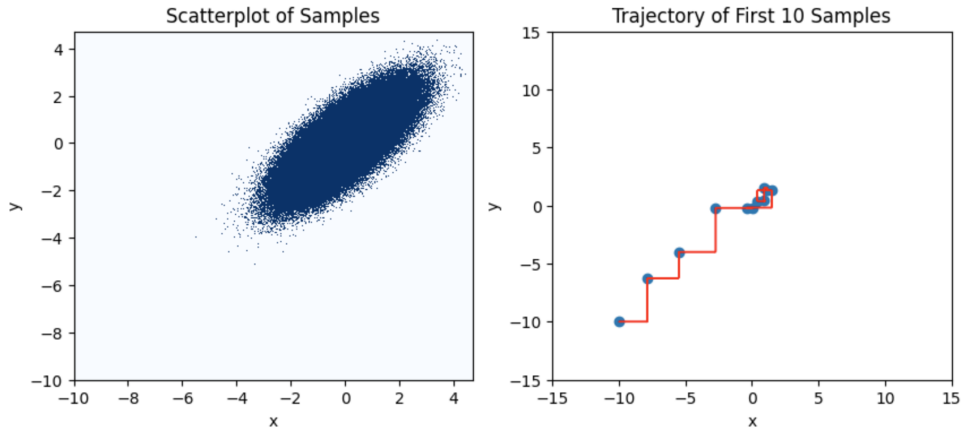


Figure 1: 2D Correlated Gaussian, Scatterplot of Samples and Trajectory - Gibbs Sampler

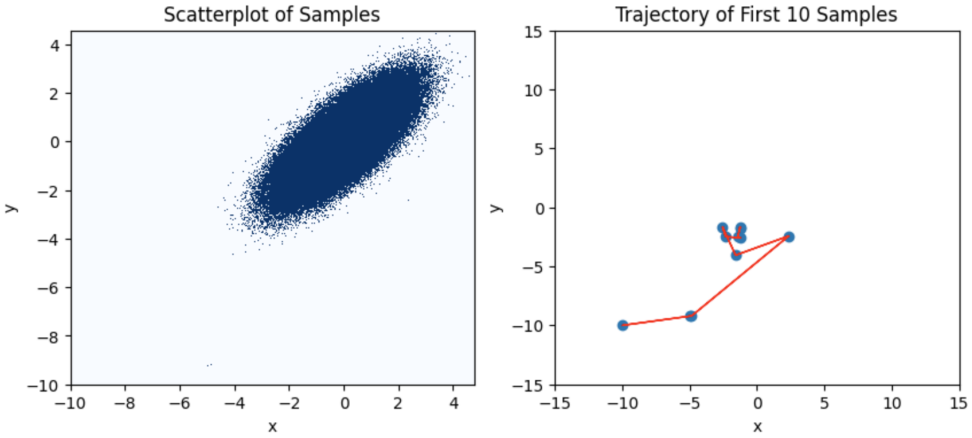


Figure 2: 2D Correlated Gaussian, Scatterplot of Samples and Trajectory - Hit-and-Run Sampler

- $R = \sqrt{\langle X, C^{-1}X \rangle}$ - we call this the radial IAT, used to test the radial exploration of the algorithms;
- $P = X_1 + X_2$ - we call this the angular IAT, as a projection to $(1, 1)$, the direction to which the Gaussian is rotated.

The plots of the autocorrelation functions for the Hit-and-Run sampler are shown in figure 3, and the IAT estimates are reported in table 1. The IAT estimates are shown in number of "updates". For the Gibbs sampler, one update means two "moves" in the chain, so one might multiply the IAT for the Gibbs chain by 2 and then compare. In number of "moves", the IAT for R is not significantly different for the two chains; Hit-and-Run seems to yield a slightly lower IAT for P, suggesting that it is better at capturing the angular aspects of the target distribution.

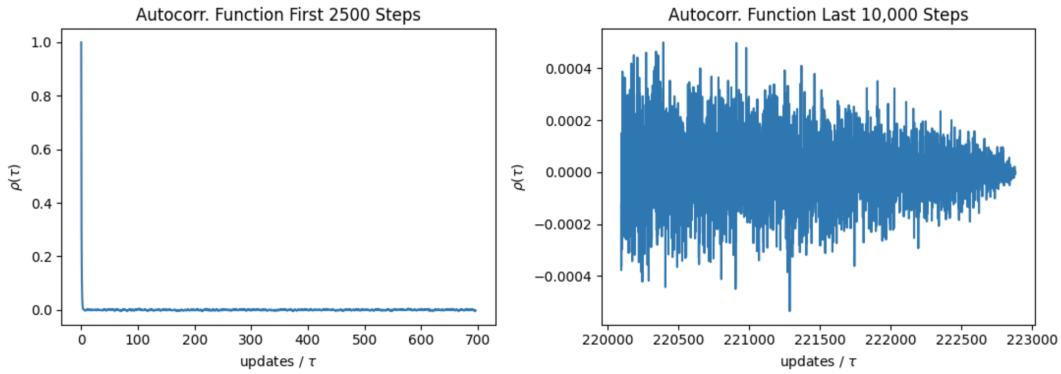
	$R = \sqrt{\langle X, C^{-1}X \rangle}$		$P = X_1 + X_2$	
	Gibbs	Hit-and-Run	Gibbs	Hit-and-Run
IAT (updates)	1.78	3.59	4.03	6.35

Table 1: 2D Correlated Gaussian, IAT Estimates

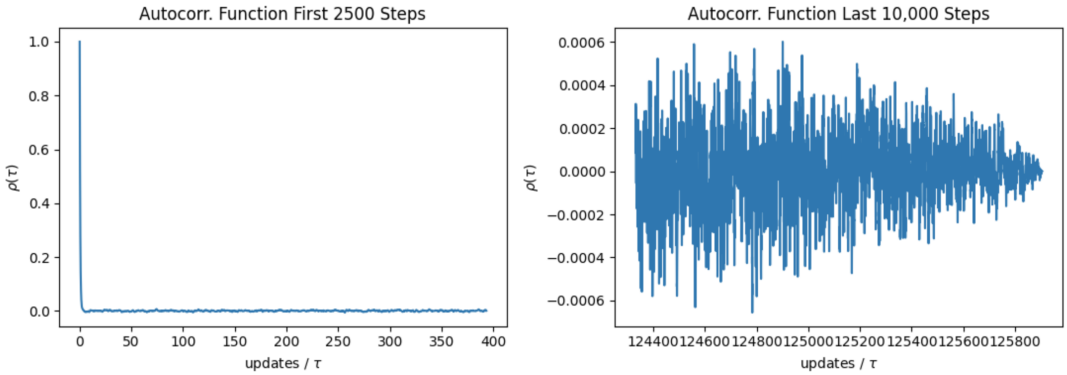
3.2 2D Independent Gaussian with Difference Variances

Consider now a centered bivariate normal whose covariance matrix is diagonal: given $\kappa \geq 1$,

$$C = \begin{bmatrix} \kappa & 0 \\ 0 & 1 \end{bmatrix}, C^{-1} = \begin{bmatrix} \kappa^{-1} & 0 \\ 0 & 1 \end{bmatrix}, C^{-1/2} = \begin{bmatrix} \kappa^{-1/2} & 0 \\ 0 & 1 \end{bmatrix}$$



(a) Radial IAT



(b) Angular IAT

Figure 3: 2D Correlated Gaussians, Autocorrelation Functions for Hit-and-Run Sampler

In this case the two coordinates are actually independent Gaussians, and the Gibbs sampler will sample from one dimensional Gaussians. To implement the Hit-and-Run sampler in this case, we simply plug in the above information on C to expression 2.

In the paper, the authors compute the Wasserstein contraction rate for this case to be $\sim \kappa^{-1/2}$. The computation relies on an observation about the law of $\frac{C^{-1/2}v}{|C^{-1/2}v|}$ where $v \sim \text{Unif}(\mathbb{S}^1)$. Figure 4 is a figure in the original paper and illustrates the idea. Recall that from the interpretation in expression 5, the Hit-and-Run scheme projects off the direction $\frac{C^{-1/2}v}{|C^{-1/2}v|}$ from the previous sample (under coordinate change) and resamples this direction with a standard Gaussian.

Suppose now one wants to move the chain in a direction that explores the density globally. In our context, this would mean, for instance, moving with an angle less than ϕ in figure 4, right panel. Indeed, our covariance matrix is diagonal, with $\kappa \geq 1$, so moving $C^{-1/2}X$ to the angle ϕ or less means moving X to the far right, which is the direction of high variance (large spread). Then one asks what direction v has to be in order for this to happen. In our case, applying $C^{-1/2}$ to v compresses the unit circle in the x direction (since $\frac{1}{\sqrt{\kappa}} \leq 1$), and normalizing pushes the vectors back out to the unit circle. Hence the angle α in the distribution of v that results in angle ϕ in the distribution of $C^{-1/2}v$ are related by

$$\tan\alpha = \frac{y}{x}, \tan\phi = \frac{y}{\frac{1}{\sqrt{\kappa}}x}, \text{ so } \tan\phi = \sqrt{\kappa}\tan\alpha \tag{8}$$

In other words, the amount of room in $\text{Unif}(\mathbb{S}^1)$ that results in a "global" move is on order of $\kappa^{-1/2}$ times the regime of "global" move in the distribution of $\frac{C^{-1/2}v}{|C^{-1/2}v|}$. The computation of contraction rate follows by plugging in relation 8 to expression 7, and realizing that the ζ achieving the infimum in question is e_1.

To see how the IAT grows as κ increases, we select 20 choices of κ between 9 and 100, generate a chain of length 2,500,000 using Hit-and-Run for each choice, and plot the relationship between κ and the radial and angular IAT estimates, shown in figure 5. As κ increases, the density becomes more ill-conditioned, so the IAT increases. The log-log plot indicates approximate power law growth of radial and angular IAT with

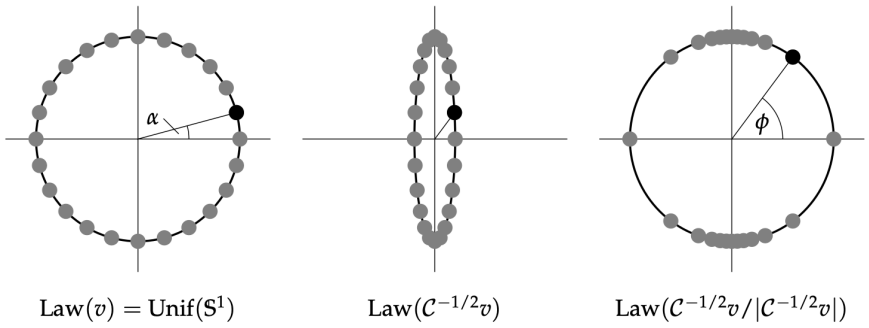


Figure 4: Illustration of Contraction Rate with 2D Gaussian with Different Variances (from the paper)

respect to κ, and the angular IAT grows faster (in fact, at a rate around $\sqrt{\kappa}$). The dotted lines are results of linear least-squares fitting from data of $\log \kappa$ and $\log IAT$.

However, the residuals in the log-log plot seems to show a subtle convex pattern, which leads to worries that the growth could be more than power law. Thus we plot a semilogy plot to check if the growth is exponential. The dotted lines there are linear least-squares results from the data of κ and $\log IAT$. The semilogy plot clearly shows the IAT grows sub-exponentially, so our power law observation should not be too far from the truth.

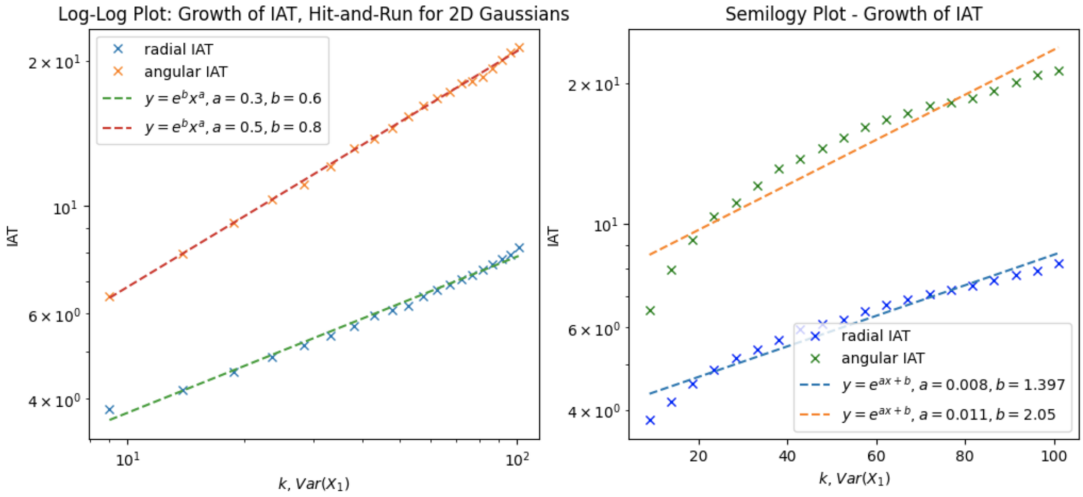


Figure 5: 2D Independent Gaussian with Large and Small Variances, Growth of IAT using Hit-and-Run.

There is a word to be said on the implementation of Hit-and-Run. Specifically, when sampling $v \sim \text{Unif}(\mathbb{S}^1)$, one way is sampling a 2D standard Gaussian and projecting it to the unit sphere, and another way is sampling an angle $\phi \in [0, 2\pi]$ and having $v = (\cos\phi, \sin\phi)$. The Gaussian approach is valid because the standard multivariate Gaussian distribution is invariant under rotation. Its advantage is that it generalizes easily to higher dimensions, since programming spherical coordinates in dimension more than 3 seems to be more work. Its disadvantage is that, in dimension d, to sample v it generates a size d Gaussian variable, so combined with the $H \sim N(0, 1)$ generated afterwards, one update of Hit-and-Run needs $d + 1$ Gaussian random variables. The approach with spherical coordinates, however, generates v with only $d - 1$ uniform random variables, so combined with $H \sim N(0, 1)$ it needs only d random variables per update. This difference does not lead to anything significant in terms of running time, however, as $d/(d + 1) \to 1$.

3.3 3D Independent Gaussian with Different Variances

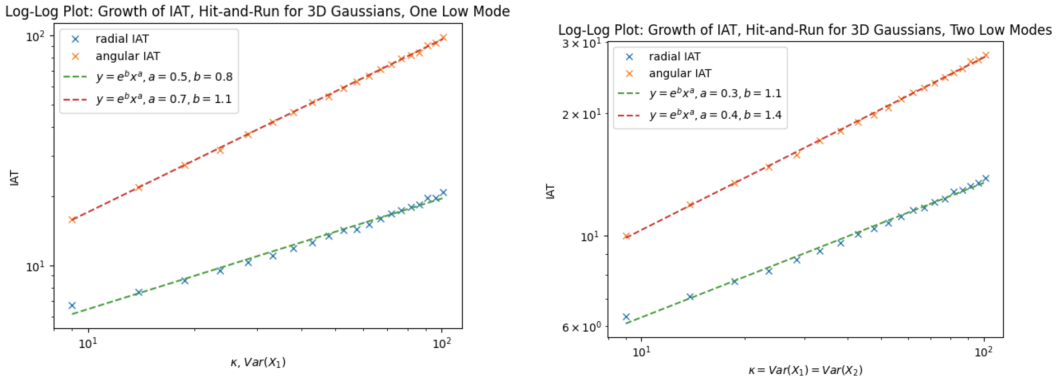
In more than two dimensions, the Hit-and-Run sampler could exhibit more diverse behaviors, depending on how the target density is set up. We illustrate this using two examples of covariance matrices. In the following experiments, our target density will be the joint density of three independent Gaussians. Consider,

for example, for $\kappa \geq 1$,

$$C_1 = \begin{bmatrix} \kappa & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, C_2 = \begin{bmatrix} \kappa & 0 & 0 \\ 0 & \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

C_1 corresponds to a density with *one* high variance direction, and C_2 a density with *two* high variance directions. The density with one high-variance direction is arguably more ill-conditioned, because it is spread out in only one direction and concentrated in the other two. The authors make similar calculations as section 3.2 to determine that in the one high-variance direction case, the contraction rate of Hit-and-Run is a superdiffusive $\sim \kappa^{-1} \log \kappa$, whereas in the other case, the contraction rate is still $\kappa^{-1/2}$.

To study the difference, we choose 20 values of κ between 9 and 100, generate a chain of length 5,000,000 for each κ, and estimate the radial and angular IAT. The log-log plot of IAT as κ grows is shown in figure 6. We see that indeed, for the density with only one high-variance direction, the IATs grow faster as κ increases, indicating slower mixing rate. For the density with two high-variance directions, the growth of IAT is at a speed that's somewhat more similar to the 2D Gaussian case. The dotted lines in the plots are results of linear least-squares from the data of $\log \kappa$ and $\log IAT$. We remark that similar to last section, the residuals may exhibit subtle convexity, but we checked in the code through semilogy plots that the IAT grows sub-exponentially. The authors extend these discussions to higher dimensions in a similar spirit.



(a) One High Variance Direction

(b) Two High Variance Directions

Figure 6: Growth of Radial and Angular IAT for 3D Independent Gaussians

4 A Coordinate-Free Randomized Kaczmarz Algorithm

4.1 Formulation and Convergence

Given $b \in \mathbb{R}^d$ and full rank $A \in \mathbb{R}^{d \times m}$ where $d \geq m$, the Kaczmarz algorithm is an iterative method that solves an overdetermined linear system $Ax = b$ by projecting the previous step iterate to the hyperplane

$$H(e_i) = \{x \in \mathbb{R}^d, \langle e_i, Ax - b \rangle = 0\}$$

In this formulation, the algorithm is bound to a chosen coordinate system. Hence, the coordinate-free nature of the Hit-and-Run sampler inspires a generalized formulation of a randomized Kaczmarz algorithm: given iterate X^k

- randomly select a vector $v \sim \tau = \text{Unif}(\mathbb{S}^{d-1})$
- project X^K such that

$$X^{k+1} \in H(v) := \{x \in \mathbb{R}^d, \langle v, Ax - b \rangle = 0\} \tag{9}$$

By characterizing $H(v)$ as $\langle x, A^T v \rangle = \langle b, v \rangle$, we can write any $x \in H(v)$ as $x = z + \frac{A^T v}{|A^T v|^2} \langle b, v \rangle$ where $z \perp A^T v$. Hence the update can be expressed as

$$X^{k+1} = \Pi_{A^T v} X^k + \frac{A^T v}{|A^T v|^2} \langle b, v \rangle \tag{10}$$

where the projection Π is onto the orthogonal complement of $A^T v$, as defined in lemma 1. Since $\frac{A^T v}{|A^T v|^2}\langle b, v\rangle$ is in $span(A^T v)$, the map $x \mapsto \Pi_{A^T v} x + \frac{A^T v}{|A^T v|^2}\langle b, v\rangle$ is itself also a projection.

Regarding convergence, we can obtain an error bound using lemma 1, if the system $Ax = b$ is solvable:

Lemma 4. *Suppose x^* satisfies $Ax^* = b$ and we start iterating from x_0. Then the random variable iterates X^k converge in L^2 to x^*:*

$$\sqrt{\mathbb{E}|X^k - x^*|^2} \leq (1-\rho)^k |x_0 - x^*|, \text{ where } \rho = \frac{1}{2} \inf_{|\zeta|=1} \mathbb{E}_{\zeta \sim \tau} \left(\left\langle \zeta, \frac{A^T v}{|A^T v|} \right\rangle^2 \right) \tag{11}$$

Proof: Notice that since $Ax^* = b$ exactly, $v \perp (Ax^* - b)$ for all v, hence $x^* \in H(v)$ as defined in 9. Hence, projecting x^* onto $H(v)$ gives back x^* itself, so plugging in 10,

$$x^* = \Pi_{A^T v} x^* + \frac{A^T v}{|A^T v|^2}\langle b, v\rangle$$

Subtracting the updating scheme 10 we discover in fact

$$X^{k+1} - x^* = \Pi_{A^T v}(X^k - x^*)$$

and hence by lemma 1,

$$\mathbb{E}|X^{k+1} - x^*|^2 \leq \left[1 - \inf_{|\zeta|=1} \mathbb{E}_{v \sim \tau} \left(\left\langle \zeta, \frac{A^T v}{|A^T v|} \right\rangle^2 \right) \right] |X^k - x^*|^2$$

Iterating this bound and taking square roots complete the proof. Note that $\sqrt{1-x} \leq (1 - \frac{1}{2}x)$.

In cases where $Ax = b$ has no solutions, we can get a similar bound by modifying the scheme 10 to the solvable system $A^T Ax = A^T b$ and applying lemma 4.

4.2 Numerical Experiment

We replicate the experiments in the paper. Set

$$A = \begin{bmatrix} 0 & 1 \\ a & 1 \end{bmatrix}, b = 0, a \in (0, 1)$$

The true solution (since $a \neq 0$) to $Ax = b$ is zero. We study cases of $a = 0.1$ and $a = 0.01$ with an initial guess of $(-10, 0)$. In these cases, the authors compute the convergence rate ρ to be at least $\frac{a}{4}$. The numerical results are shown in figure 7. Consistent with the paper's findings, the iterates converge in root mean squared error faster than an exponential rate of $(1 - \frac{a}{4})^k$.

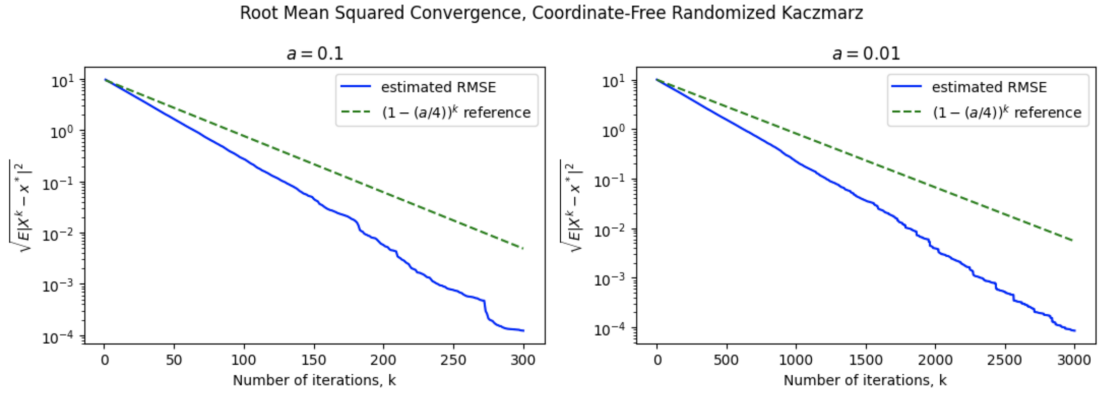


Figure 7: L^2 Convergence of Iterate Random Variables, Coordinate-Free Randomized Kaczmarz

References

[1] Nawaf Bou-Rabee, Andreas Eberle, and Stefan Oberdörster. Ballistic convergence in hit-and-run monte carlo and a coordinate-free randomized kaczmarz algorithm, 2024.