Sampling from Convex Sets with a Cold Start using Multiscale Decompositions

Hariharan Narayanan  
Tata Institute of Fundamental Research  
Mumbai, Maharashtra, India  
hariharan.narayanan@tifr.res.in

Amit Rajaraman  
Indian Institute of Technology Bombay  
Mumbai, Maharashtra, India  
amit.rajaraman@gmail.com

Piyush Srivastava  
Tata Institute of Fundamental Research  
Mumbai, Maharashtra, India  
piyush.srivastava@tifr.res.in

ABSTRACT

Running a random walk in a convex body \(K \subseteq \mathbb{R}^n\) is a standard approach to sample approximately uniformly from the body. The requirement is that from a suitable initial distribution, the distribution of the walk comes close to the uniform distribution \(\pi_K\) on \(K\) after a number of steps polynomial in \(n\) and the aspect ratio \(R/r\) (i.e., when \(rB_2 \subseteq K \subseteq RB_2\)). Proofs of rapid mixing of such walks often require the probability density \(\eta_0\) of the initial distribution with respect to \(\pi_K\) to be at most \(\text{poly}(n)\): this is called a “warm start”. Achieving a warm start often requires non-trivial pre-processing before starting the random walk. This motivates proving rapid mixing from a “cold start”, wherein \(\eta_0\) can be as high as \(\exp(\text{poly}(n))\). Unlike warm starts, a cold start is usually trivial to achieve. However, a random walk need not mix rapidly from a cold start: an example being the well-known “ball walk”. On the other hand, Lovász and Vempala proved that the “hit-and-run” random walk mixes rapidly from a cold start. For the related coordinate hit-and-run (CHR) walk, which has been found to be promising in computational experiments, rapid mixing from a warm start was proved only recently but the question of rapid mixing from a cold start remained open.

We construct a family of random walks inspired by classical decompositions of subsets of \(\mathbb{R}^n\) into countably many axis-aligned dyadic cubes. We show that even with a cold start, the mixing times of these walks are bounded by a polynomial in \(n\) and the aspect ratio. Our main technical ingredient is an isoperimetric inequality for \(K\) for a metric that magnifies distances between points close to the boundary of \(K\). As a corollary, we show that the CHR walk also mixes rapidly from a cold start.

1HN and PS acknowledge support from the Department of Atomic Energy, Government of India, under project no. RTH001, the Ramanujan Fellowship of SERB, and the Infosys-Chandrasekharan virtual center for Random Geometry. HN acknowledges support from a Swarna Jayanti fellowship. PS acknowledges support from Adobe Systems Incorporated via a gift to TIFR. The contents of this paper do not necessarily reflect the views of the funding agencies listed above. A full version of this extended abstract is available at [29].

CCS CONCEPTS

- Theory of computation → Random walks and Markov chains; Computational geometry.

KEYWORDS

Markov chains, coordinate hit-and-run, isoperimetry inequalities

ACM Reference Format:

Hariharan Narayanan, Amit Rajaraman, and Piyush Srivastava. 2023. Sampling from Convex Sets with a Cold Start using Multiscale Decompositions. In Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC ’23), June 20–23, 2023, Orlando, FL, USA. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3564246.3585172

1 INTRODUCTION

The problem of generating a point distributed (approximately) uniformly over a convex set \(K \subseteq \mathbb{R}^n\) is an important algorithmic primitive. It is usual to assume that the body \(K\) is presented by means of a “well guaranteed membership oracle”, i.e., a membership oracle for \(K\), along with values \(R > r > 0\) such that the body is contained in the radius \(R\) Euclidean ball and also contains the radius \(r\) Euclidean ball. The ratio \(R/r\) is then referred to as the aspect ratio of the body.

The first provably polynomial time algorithm for this problem was given by Dyer, Frieze and Kannan [8]: their algorithm used a random walk on a uniformly-spaced lattice of points in a suitable “smoothed” version of the original body \(K\). More refined analyses of such lattice walks were given in subsequent works [1, 7, 24]. We refer to [25] for a more complete discussion of the history. Soon after, Lovász [22] and Lovász and Simonovits [25] considered a more geometric random walk not supported on a discrete lattice: the so-called ball walk. Here, one fixes a radius parameter \(\delta\), and given a current point \(x \in K\), proposes a next point \(y\) from the Euclidean ball of radius \(\delta\) centered at \(x\), and moves to \(y\) if \(y \in K\). They prove (see [25, Remark on p. 398]) that when \(\delta\) is chosen appropriately, the lazy\(^1\) version of the ball walk mixes rapidly, i.e., it reaches a distribution that is \(\epsilon\)-close in total variation distance to the uniform distribution \(\pi_K\) on \(K\), after a number of steps which is polynomial in \(n, 1/\epsilon\) and \(R/r\), provided that the initial point of the random walk is chosen according to a poly \((n)\)-warm start. (A distribution \(\mu\) supported on \(K\) is said to be \(M\)-warm if the density function of \(\mu\) with respect to \(\pi_K\) is bounded above by \(M\)). Another

\(^1\)A random walk is called lazy if the probability that it stays at its current state after one step is at least \(1/2\). A lazy version of any random walk \(W\) can be obtained by considering the random walk in which at each step, the walk simply stays at the current state with probability \(1/2\), and takes a step according to \(W\) with probability \(1/2\). Considering only lazy versions of walks is a standard device for avoiding pathological periodicity issues, and therefore we will always work with lazy walks in this paper.
natural geometric random walk is the hit-and-run walk (see [31], where it is attributed to earlier work of Boneh and Golan, and of Smith). Here, if the current state is \( x \in K \), then the next point \( y \) is sampled by first choosing a uniformly random direction \( \tilde{u} \) from the unit sphere \( S^{n-1} \), and then picking \( y \) uniformly at random from the chord of \( K \) in direction \( \tilde{u} \) passing through \( x \). Lovász [23] proved that the lazy hit-and-run walk also mixes in polynomial time in \( n \), \( 1/e \) and \( R/r \), again assuming that the initial point is sampled from a \( (n) \)-warm start.

While a \( (n) \)-warm start can be achieved in polynomial time, it requires sophisticated pre-processing. In contrast, a "cold start", i.e., an \( M \)-warm start where \( M \) can be as large as \( \exp(\text{poly}(n)) \), is very easy to generate when \( R/r \) is at most \( \exp(\text{poly}(n)) \): one can simply sample the initial point uniformly at random from the radius \( r \) Euclidean ball. The first polynomial time mixing time result for the hit-and-run walk from such a cold start, without the need for any further pre-processing, was proved by Lovász and Vempala [26].

An interesting variant of the hit-and-run walk is the coordinate hit-and-run (CHR) walk, where the direction \( \tilde{u} \) is chosen uniformly at random from one of the coordinate directions. The CHR walk is attractive in part because the implementation of each step of the chain can potentially be quite efficient: Smith [31, pp. 1302-1303] already mentioned some preliminary computational experiments of Telgen supporting such an expectation in the important special case when \( K \) is a polytope described by a small number of sparse inequalities. More recent computational work has also explored the CHR walk in various application areas [10, 15]. However, few theoretical guarantees were known for the coordinate direction walk, and it was only recently that Laddha and Vempala [18] and Narayanan and Srivastava [30] proved that with a \( (n) \)-warm start, the lazy CHR walk mixes in polynomial time. The question of its mixing time from a "cold start", i.e., from a \( \exp(\text{poly}(n)) \)-warm start, however has remained open.

1.1 Contributions

We construct a new family of Markov chains inspired by classical multiscale decompositions of bounded subsets of \( \mathbb{R}^n \) into axis-aligned dyadic (i.e., of sidelength equal to a integral power of two) cubes. Our chains \( M_p \) are parameterized by the \( \ell_p \) norms on \( \mathbb{R}^n \), \( 1 \leq p \leq \infty \). Our first contribution is to show that all of these chains require only a polynomial in \( n \) and the aspect ratio \( R/r \), as before number of steps to come within \( \varepsilon \) total variation distance of the uniform distribution \( \pi_K \), even when started with an \( \exp(\text{poly}(n)) \)-warm start. However, before describing the \( M_p \) chains and our mixing result for them in detail, we state the following result that we obtain as a byproduct of their analysis. (Given the special status of the coordinate directions in the coordinate hit-and-run walk, we parametrize the aspect ratio in terms of the \( t_{\infty} \) unit ball \( B_{\infty} \) rather than in terms of the Euclidean unit ball \( B_2 \).)

**Theorem 1.1** (see Corollary 6.4). Let \( K \subseteq \mathbb{R}^n \) be a convex body such that \( r \cdot B_{\infty} \subseteq K \subseteq r \cdot B_{2} \). Then starting from an \( M \)-warm start, the lazy coordinate hit-and-run walk comes within total variation distance at most \( \varepsilon \) of the uniform distribution \( \pi_K \) on \( K \) after \( O\left(n^2(R/r)^2\log(M/\varepsilon)\right) \) steps.

The above result shows that the coordinate hit-and-run (CHR) random walk also mixes in polynomial (in \( n \) and the aspect ratio \( R/r \) time even from "cold", i.e., \( \exp(\text{poly}(n)) \)-warm starts. As described above, polynomial time mixing for the CHR walk had only been proved so far starting from a \( (n) \)-warm start [18, 30]: the dependence on \( M \) in the mixing time bounds obtained in [18, 30] are proportional to \( \text{poly}(M) \), as compared to the \( \log M \) dependence in our Theorem 1.1.

We now proceed to describe our main technical result: the construction of the \( M_p \) random walks and their rapid mixing from a cold start. The random walks \( M_p \) are inspired by a classical decomposition of bounded subsets of \( \mathbb{R}^n \) into axis-aligned cubes with disjoint interiors. Such decompositions have been used since the work of Whitney [33] (see, e.g., [11, 12] for more recent examples of their use). We now informally describe the decomposition of \( K \) that we use for the \( M_p \) chain. For simplicity, we assume that \( K \) is contained in the interior of the \( t_{\infty} \) ball of radius \( 1 \). We start with the standard tiling of \( \mathbb{R}^n \) by unit cubes with vertices in \( \mathbb{Z}^n \), and also consider all scalings of this tiling by factors of the form \( 2^{-k} \), where \( k \) is a positive integer. Our decomposition \( \mathcal{F} = \mathcal{F}_{(p)} \) of \( K \) into cubes with disjoint interiors is then obtained by considering these cubes in decreasing order of sidelength and including those cubes \( Q \) for which

1. \( Q \) is contained within \( K \), and in fact, relative to its own diameter, \( Q \) is "far away" from the exterior \( \mathbb{R}^n \setminus K \) of \( K \): the \( \ell_p \)-distance of the center of \( Q \) from \( \mathbb{R}^n \setminus K \) is at least twice the \( \ell_p \)-diameter of \( Q \), and
2. no "ancestor" cube of \( Q \), i.e., a cube containing \( Q \) is part of the decomposition \( \mathcal{F} \).

A formal description of the construction of \( \mathcal{F} \) is given in Section 3, where it is also shown that such a decomposition fully covers the interior \( K_{\infty} \) of \( K \), and also that if two cubes in \( \mathcal{F} \) abut along an \( (n-1) \)-dimensional facet, then their sidelengths must be within a factor of two of each other. We note that this "bounded geometry": namely that the ratio of the side lengths of abutting cubes are within a factor of two of each other (see fig. 1), is a very useful feature of this construction for our purposes. In particular, this feature plays an important role in relating the properties of the \( M_p \) chains to the coordinate hit-and-run random walk.

![Figure 1: Local geometry of Whitney decompositions: the sidelengths of adjacent cubes are within a factor of two of each other.](image)
With the remaining probability, it performs the following step. Pick a point $x$ uniformly at random from the boundary $\partial K$ of the cube $Q$. With probability 1, there is a unique cube $Q' \neq Q$ in $\mathcal{F}(p)$ to which $x$ belongs. The walk proposes a move to this cube $Q'$, and then accepts it based on a standard Metropolis filter with respect to $\pi$. The Metropolis filter ensures that the walk is in fact reversible with respect to $\pi$, i.e., $\pi(Q) P_{M}(Q, Q') = \pi(Q') P_{M}(Q', Q)$ where $P_{M}(Q, Q')$ is the probability of transitioning to cube $Q'$ in one step, when starting from cube $Q$. This implies that $\pi$ is a stationary distribution of $M_p$ (see Section 4 for details).

As stated above, $M_p$ can equivalently be seen as a Markov chain on $K$ itself. To see this, note that corresponding to any probability distribution $\nu$ on $\mathcal{F}(p)$, there is a probability distribution $\nu_K$ on $K$ obtained by sampling a cube $Q$ from $\mathcal{F}(p)$ according to $\nu$, and then a point $x$ uniformly at random from $Q$. It is easy to see that the uniform distribution $\nu_K$ on $K$ can be generated from the distribution $\nu$ above in this fashion. Further, one can also show that the total variation distance between the distributions $\nu_K$ and $\pi_K$ on $K$ is at most the total variation distance between the distributions $\nu$ and $\pi$ on $\mathcal{F}(p)$ (this follows directly from the definition). Similarly, given a probability distribution $\nu_K$ on $K$ that is $M$-warm with respect to $\pi_K$, one can obtain a distribution $\nu$ on $\mathcal{F}(p)$ that is $M$-warm with respect to $\pi$. This is done as follows. Sample a point $x$ according to $\nu_K$. With probability 1, $x$ lies in the interior of some cube $Q \in \mathcal{F}(p)$ (this follows because $\nu_K$ is $M$-warm with respect to $\pi_K$ and because the probability measure under $\pi_K$ of the union of the boundaries of the countably many cubes in $\mathcal{F}(p)$ is zero). $v$ is then defined to be the probability distribution of this random walk $Q$. Then $v(Q) = v_K(Q) \leq M \nu_K(Q) = M \text{vol} (Q) / \text{vol} (K) = M \pi(Q)$.

Our main theorem for $M_p$ chains is the following. Here, $B_p := \{x \in \mathbb{R}^n : \|x\|_p < 1\}$ is the unit $p$-ball in $\mathbb{R}^n$; note that the requirement $r \cdot B_p \subseteq K$ is weaker than the requirement $r \cdot B_\infty \subseteq K$ when $p < \infty$.

**Theorem 1.2 (see Corollary 5.5).** Fix $1 \leq p \leq \infty$. Let $K \subseteq \mathbb{R}^n$ be a convex body such that $r \cdot B_p \subseteq K \subseteq R \cdot B_\infty$. Then, starting from an $M$-warm start, the $M_p$ random walk on $K$ comes within total variation distance at most $\epsilon$ of the uniform distribution $\pi_K$ on $K$ after $\tilde{O}(n^{4p} / (R/r)^2 \cdot \log (M/\epsilon))$ steps.

**Remark 1.3.** In Theorem 5.6, we use a more refined analysis to establish essentially the same dependence of the mixing time on $n$ and $R/r$ even when $M_p$ is started from a starting distribution supported on a single cube $Q \in \mathcal{F}(p)$ (or equivalently, in light of the discussion in the previous paragraph, a point $x \in Q$) that is at least $\frac{1}{\text{poly}(n)}$ away from the boundary of $K$ (a direct application of Theorem 1.2 would lose an extra factor of $\tilde{O}(n)$ in this setting).

**Algorithmic implementation of $M_p$.** We note that the tools we develop for the analysis of our multiscale chains $M_p$ play a crucial role in our result for the CHR walk (Theorem 1.1). In addition, $M_p$ chains are of algorithmic interest in their own right. However, it may not be immediately clear how to algorithmically implement each step of the $M_p$ chain from the above description of the chain and its state space $\mathcal{F}(p)$ of Whitney cubes. We show in Section 4.1 that each step of the $M_1$ chain can be algorithmically implemented in $O(n)$ time using only a membership oracle for $K$. When $p > 1$, algorithmically implementing one step of the $M_p$ chain requires access to an oracle for the $\ell_p$-distance of a point $x \in K$ to the boundary $\partial K$ of $K$; such oracles can be implemented efficiently for polytopes. We describe this construction as well in Section 4.1.

We now proceed to discuss the context for our results in the light of existing literature. Following this, we give a overview of our results and proof techniques in Section 1.3.

### 1.2 Discussion

The notion of conductance has played a central role in most rapid mixing results for random walks on convex bodies. For the discussion below, we fix a convex body $K \subseteq \mathbb{R}^n$ such that $rB_2 \subseteq K \subseteq RB_2$.

Given a random walk $\mathcal{W}$ with stationary distribution as the uniform distribution $\pi_K$ on $K$, the conductance $\Phi_w(S)$ of a subset $S \subseteq K$ is defined as the probability of the following randomly chosen point lying in $K \setminus S$: choose a point uniformly at random from $S$, and then take a step according to $\mathcal{W}$. It follows from standard results in the theory of Markov chains [25] that if $\Phi_w(S) \geq 1/\text{poly}(n, R/r)$ for every measurable $S \subseteq Q$, then the random walk $\mathcal{W}$ mixes rapidly from a $\exp(\text{poly}(n))$-warm start. However, in several cases, one only gets the weaker result that only large enough subsets have good conductance: a formalization of this is through the notion of $s$-conductance [25, p. 367], which can capture the phenomenon that, roughly speaking, the lower bound obtained on the conductance of $S$ degrades as the volume $s$ of the set $S$ becomes smaller. Under such a bound, one usually only gets rapid mixing from a $\exp(\text{poly}(n))$-warm start (see, e.g., [25, Corollary 1.6]). The reason that one can only get a lower bound on the conductance of large sets may have to do with the properties of the walk $\mathcal{W}$ itself (which is the case with the ball walk). However, it may also be an artefact of the proof method rather than a property of the walk itself. For example, the original proof of Lovász [23] for the rapid mixing of the hit-and-run walk was built upon an $s$-conductance lower bound that approached zero as the size parameter $s$ approached zero [23, Theorem 3], and therefore required a $\exp(\text{poly}(n))$-warm start. In contrast, the later proof by Lovász and Vempala [26] established a conductance bound for the same chain and thereby achieved rapid mixing from a cold start.

Rapid mixing proofs of random walks on convex sets often follow the plan of establishing a conductance (or $s$-conductance) bound for the chain using an isoperimetric inequality for an appropriate metric (roughly speaking, an isoperimetric inequality puts a lower bound on $\text{vol} (K \setminus (S_1 \cup S_2))$ proportional to the product of volumes of $S_1$ and $S_2$ and the distance $\delta$ between $S_1$ and $S_2$, at least when $\delta$ is a sufficiently small positive number). $^2$ A unifying theme in the analysis of many random walks for sampling from convex sets, starting from the work of Lovász [23], has been to prove such an isoperimetric inequality when the underlying metric is non-Euclidean. For example, the underlying metric in [23] is the Hilbert metric defined using the logarithm of certain cross-ratios. This isoperimetric inequality was then used to give an inverse polynomial lower bound for the $s$-conductance of the hit-and-run walk that degraded gracefully to zero as the size parameter $s$ approached zero, thereby leading to a rapid mixing result for the hit-and-run walk under a warm start. In later work, Lovász and Vempala [26] $^2$ A notable exception to this general strategy is the work of Bubley, Dyer and Jerrum [2], discussed in more detail later in the introduction.
obtained an inverse-polynomial lower bound on the conductance of the hit-and-run walk by refining the isoperimetric inequality for the Hilbert metric proved in [23]: this improvement in the isoperimetric inequality thus led to a rapid mixing result for the hit-and-run walk without the need of a warm start.

The Hilbert metric also appears in the analysis by Kannan and Narayanan [16] of another random walk, called the Dikin walk, on polytopes. The Dikin walk was generalized by Narayanan [28] to more general convex sets equipped with a weighted combination of logarithmic, hyperbolic and self-concordant barriers, and was analysed using a different Riemannian metric whose metric tensor is derived from the Hessian of the combined barrier. The isoperimetric properties of this Riemannian metric were established by comparison to the Hilbert metric. Improvements on this walk with better mixing times have been obtained by Chen, Dwivedi, Wainwright and Yu [5] and by Laddha, Lee, and Vempala [17]. The geodesic walk of Lee and Vempala [19] uses geodesics of the Riemannian metric associated with the logarithmic barrier to define a walk on polytopes, whose properties again hinge on the isoperimetric properties of the convex set equipped with the Hilbert metric and the uniform measure.

Beyond proving the isoperimetric inequality, there is also the need to relate these Markov chains to the reference metric introduced. This was done for hit-and-run in [23] using in part the well-known theorem of Menelaus in Euclidean geometry. This step for the Dikin walk used facts from interior point methods developed by Nesterov and Nemirovski. The analogous analysis was particularly involved in [19] and used Jacobi fields among other tools. For a more detailed discussion of these and related developments, we refer to the recent survey [20] by Lee and Vempala.

Unfortunately, it has not been possible to exploit the Hilbert metric to analyze the coordinate hit-and-run (CHR) walk. However, in recent work, Laddha and Vempala [18] showed how to implement the program of proving an s-conductance bound for the CHR walk using an isoperimetric inequality for an appropriate metric: they proved rapid mixing for the CHR walk from a warm start via an isoperimetric inequality for subsets of $K$ that are far in the $f_0$-metric and that are not too small in volume (the $f_0$-distance between two points in $\mathbb{R}^n$ is the number of coordinates on which they differ).

Our result for the CHR walk (Theorem 1.1) also hinges on a similar $f_0$-isoperimetric inequality, Theorem 6.2, which however extends to sets of all volumes (including arbitrarily small volumes). This is the main technical ingredient that allows us to remove the requirement of a warm start in Theorem 1.1.

The proof of Theorem 6.2 itself goes via the proof of a conductance lower bound for the $M_p$ chains on Whitney decompositions of $K$ that we introduced above. The conductance analysis of the $M_p$ chains, in turn, proceeds by introducing a kind of degenerate Finsler metric on $K$ (see Section 5), which is a scaled version of $f_d$ that magnifies distances in the vicinity of a point $x$ in $K$ by a factor of $1 / \text{dist}_d(x, K)$. Our main technical ingredient is a new isoperimetric inequality (Theorem 5.1) for any convex body $K$ under such a metric. Part of the proof of this inequality requires an existing isoperimetric inequality for convex sets in normed spaces proved by Kannan, Lovász and Montenegro [15], but the bulk of the proof is handled by a detailed analysis of “needles” analogous to those in the celebrated localization lemma of Lovász and Simonovits [25]. In the more refined analysis (Theorem 5.6) of the $M_p$ chain from a fixed state that we alluded to in the remark following Theorem 1.2, we also use results of Lovász and Kannan [21] relating rapid mixing to average conductance rather than worst-case conductance, thereby saving ourselves a factor of $O(n)$ in the mixing time. This in turn is made possible by the fact that for the degenerate Finsler metric we introduce, the lower bounds we can prove on the isoperimetric profile of small sets are actually stronger than those we can prove for large sets.

We now proceed to give a more detailed overview of our techniques. Most proofs and technical details are omitted from this extended abstract, and can be found in the full version of the paper [29].

1.3 Technical Overview

Our result follows the general schema of establishing a conductance lower bound for the chain using an isoperimetric inequality for an appropriate metric. As discussed above, the requirement of a warm start in rapid mixing proofs is often a consequence of the fact that non-trivial bounds for the conductance of the chain are available only for sets of somewhat large volumes. This in turn is often due to having to “throw away” a part of the volume of $K$ that is close to the boundary $\partial K$ of $K$ before applying the isoperimetric inequality: this is the case, for example, with the original warm start rapid mixing proof of the hit-and-run walk [23]. The same issue also arose in two different proofs of rapid mixing for the coordinate hit-and-run (CHR) walk starting with a warm start [18, 30]: in both these proofs, an isoperimetric inequality could only be applied after excluding a part of $K$ close to $\partial K$.

Our motivation for considering a multiscale walk comes partly from the desire to avoid this exclusion of the part of $K$ close to its boundary. Notice that as our multiscale chain $M_p$ approaches the boundary of $K$, the underlying cubes also become proportionately smaller, and the chain can still make progress to neighboring cubes at a rate that is not much worse than what it would be from larger cubes in the deep interior of the body. Note, however, that this progress cannot be captured in terms of usual $f_p$-norms: while the chain does move to adjacent cubes, the distances between the centers of these adjacent cubes shrink as the chain comes closer to the boundary of $K$. Thus, it seems unlikely that isoperimetric properties for $f_p$-norms alone (e.g., those in [15, 25]) would be able to properly account for the progress the multiscale chain makes when it is close to the boundary of $K$.

A metric and an isoperimetry result. In order to properly account for this progress, we introduce metrics that magnify distances close to the boundary $\partial K$ of $K$. More concretely, to analyze the chain $M_p$, we consider the metric $\eta_p$ that magnifies $\text{dist}_d$-distances in the vicinity of a point $x \in K$ by a factor of $1 / \text{dist}_d(x, K)$ (see Section 5 for the formal definition of the metric $\eta_p$). Because of this scaling, this metric captures the intuition that the chain’s progress close to the boundary is not much worse than what it is in the deep interior of $K$. Our main technical result is an isoperimetry result for $K$ endowed with the $\eta_p$ metric and the uniform (rescaled Lebesgue) probability measure. We show that $\text{vol}(K \setminus (S_1 \cup S_2))$ is significant in proportion to $\text{min} \{\text{vol}(S_1), \text{vol}(S_2)\}$ whenever $S_1$ and $S_2$ are
subsets of $K$ that are far in the $g_p$ distance: see Theorem 5.1 for the
detailed statement.

Our proof of Theorem 5.1 is divided into two cases depending
upon whether $S_1$, the smaller of the sets $S_1$ and $S_2$, has a significant
mass close to the boundary of $K$ or not. The easy case is when $S_1$
does not have much mass close to the boundary, and in this case we
are able to appeal to an isoperimetric inequality of Kannan, Lovász
and Montenegro [15] for the standard $\ell_p$ norms.

The case that requires more work is when a large constant frac-
tion (about 0.95 in our proof) of the volume of $S_1$ lies within $\ell_p$-
distance $C_1/n$ of the boundary $\partial K$ of $K$ for some parameter $C_1$. Our
proof of this part is inspired by the localization idea of [25], but we
are unable to directly apply their localization lemma in a black box
manner. Instead, we proceed by radially fibering the body $K$ into
one-dimensional needles, where the needles correspond to radial
line segments in a spherical polar coordinate system centered at a
point $x_0$ in the deep interior of $K$. The intuition is that since $S_1$ and
$S_2$ are at distance at least $\delta > 0$ in the $g_p$ metric, a large fraction of
these needles contain a large segment intersecting $S_3 = K \setminus (S_1 \cup S_2)$. This
intuition however runs into two competing requirements.

(1) First, the $S_3$-segment in a needle cannot be too close to
the boundary $\partial K$. This is because the $g_p$ metric magnifies
distances close to $\partial K$, so that a segment that is close to $\partial K$
is of length $\delta$ in the $g_p$ metric may have a much smaller
length in the usual Euclidean norm. The contribution to the
volume of $S_3$ of such a segment would therefore also be
small.

(2) Second, neither can the $S_2$-segment in a needle be too far
from the boundary $\partial K$. This is because, by definition, a nee-
dle $N$ is a radial line in a polar coordinate system centered
at a point $x_0$ deep inside $K$, so that the measure induced
on $N$ by the standard Lebesgue measure is proportional to
$t^{n-1}$, where $t$ is the Euclidean distance from $x_0$. Thus,
the measure of an $S_2$-segment that lies close to the center $x_0$ of
the polar coordinate system may be attenuated by a large
factor compared to the measure of a segment of the same
Euclidean length that lies closer to $\partial K$.

For dealing with these two requirements together, we consider the outer “stub” of each needle, which is the part of the needle starting
from $\partial K$ up to a $C_2/n$ distance along the needle, where $C_2$ is an
appropriate factor that depends upon the needle. For an appropri-
ate choice of $C_1$ and $C_2$, we can show that for at least a significant
fraction of needles, the following conditions are simultaneously satisfied:

(1) The stub of the needle contains a non-zero volume of $S_1$.
(2) A large fraction of the inner part of the stub (i.e., the part
farthest from the boundary) is not in $S_1$.

Together, these facts can be used to show that the inner part of
the stub contains a large segment of $S_3$. This achieves both the
requirements above: the segment of $S_3$ found does not lie too close
to the boundary (because it is in the inner part of the stub), but is
not too far from the boundary either (because the stub as a whole
is quite close to $\partial K$ by definition).

**Mixing time for the $M_p$ chains.** We then show in Section 5.2 that
the isoperimetric inequality above implies a conductance lower
bound for the $M_p$ chain, in accordance with the intuition outlined
for the definition of the $g_p$ metric. Rapid mixing from a cold start
(Corollary 5.5) then follows immediately from standard theory. In
Section 5.5, we show that the fine-grained information that one
 obtains about the conductance profile of the $M_p$ chain can be
used to improve the mixing time from a fixed state by a factor of $O(n)$ over what the vanilla mixing time result from a cold start
(Corollary 5.5) would imply. We also show in Proposition 5.4 that
the conductance lower bound we obtain for the $M_p$ chain is tight
up to a logarithmic factor in the dimension.

**Rapid mixing from cold start for coordinate hit-and-run.** Finally,
we show rapid mixing from a cold start for the coordinate hit-and-
run (CHR) walk in Section 6. As described above, two different
proofs were recently given for the rapid mixing for this chain from
a warm start [18, 30], and in both of them, the bottleneck that led
to the requirement of a warm start was a part of the argument that
had to “throw away” a portion of $K$ close to $\partial K$. In Section 6, we
show that the conductance (even that of arbitrarily small sets) of the
CHR walk can be bounded from below in terms of the conductance
of the multiscale chain $M_{\infty}$ (Theorem 6.3). As discussed above,
the conductance of the latter can be bounded from below using the
isoperimetry result for the $g_\infty$ metric. Together, this gives a rapid
mixing result for the CHR walk from a cold start (Corollary 6.4
and Theorem 1.1). To prove Theorem 6.3, we build upon the notion of
axis-disjoint sets introduced by Laddha and Vempala [18], who
had proven an “$g_0$-isoperimetry” result for such sets. However, as
discussed above, their isoperimetry result gives non-trivial conduc-
tance lower bounds only for sets of somewhat large volume. This
was in part due their result being based on a (partial) tiling of $K$
by cubes of fixed sidelength, thereby necessitating the exclusion of
a part of the volume of the body close to the boundary. The main
technical ingredient underlying our result for the CHR chain is a
new $g_\infty$-isoperimetry result for axis-disjoint sets (Theorem 6.2) that
applies to sets of all sizes, and that involves the conductance of the
multiscale $M_{\infty}$ chain described above.

**1.4 Open Problems**

We conclude the introduction with a discussion of some directions
for future work suggested by this work. The natural question raised
by the application of the $M_{\infty}$ walk to the analysis of the coordinate
hit-and-run walk is whether the $M_p$ chains, or the notion of
multiscale decompositions in general, can be used to analyze the
rapid mixing properties of other random walks on convex sets.

An alternative to the approach of using isoperimetric inequali-
ties for analyzing mixing times for random walks on convex sets is
suggested by an interesting paper of Bubley, Dyer and Jerrum [2],
in which a certain gauge transformation is used to push forward the
uniform measure on a convex set on to a log concave measure sup-
ported on $\mathbb{R}^n$, whereafater a Metropolis-filtered walk is performed
using Gaussian steps. The analysis of this walk (which mixes in
polynomial time from a cold start, or even from the image, un-
der the gauge transformation, of a fixed point not too close to the
boundary) proceeds via a coupling argument, and does not use
the program of relating the conductance of the chain to an isoperi-
metric inequality. Such coupling arguments have also been very
successful in the analysis of a variety of Markov chains on finite
state spaces. It would be interesting to explore if a coupling based analysis can be performed for the $M_p$ random walks or for the CHR random walk. Another possible approach to attacking these questions on rapid mixing could be the recent localization scheme framework of Chen and Eldan [6].

Another direction for investigation would be to make the implementation of each step of $M_p$, especially in the case $p > 1$, more efficient. In the current naive implementation of a step of the framework of Chen and Eldan [6].

Questions on rapid mixing could be the recent localization scheme 

\[ \Phi \subseteq \mathcal{M} \]

Definition 2.3 (Conductance profile). For $\alpha \in (0,1/2]$, we define the value $\Phi_{\alpha,M}$ of the conductance profile of $M$ at $\alpha$ as the infimum of $\Phi_M(S)$ over all measurable $S \subseteq \Omega$ such that $\pi(S) \leq \frac{\alpha}{2}$.

\[ \Phi_M := \inf_{S: \pi(S) \leq \frac{\alpha}{2}} \Phi_M(S). \]
Lemma 2.1 ([25, Corollary 1.8]). Suppose that the lazy Markov chain $M$ on $\Omega$ is reversible with respect to a probability distribution $\pi$ on $\Omega$. Let $v_0$ have density $\eta_0$ with respect to $\pi$, and define $\eta_t$ to be the density of the distribution $\nu_t = v_0 \pi^t$ obtained after $t$ steps of the Markov chain starting from the initial distribution $v_0$. Then
\[
\|\eta_t - \pi\|_{L^1(\pi)} \leq \left(1 - \frac{\Phi}{2}\right)^t \|\eta_0 - \pi\|_{L^1(\pi)}
\]
where $\Phi$ is the conductance of $M$.

2.2 Geometric Facts

Note. For any subset $S \subseteq \mathbb{R}^n$ we will denote by $S^o$ its open interior (i.e., the union of all open sets contained in $S$), and by $\partial S$ the boundary of $S$, defined as $S \setminus S^o$, where $S$ is the closure of $S$. Note that $\partial S \subseteq S$ if and only if $S$ is closed. A convex body in $\mathbb{R}^n$ is a closed and bounded convex subset of $\mathbb{R}^n$ that is not contained in any proper affine subspace of $\mathbb{R}^n$. We will need the following standard fact.

Lemma 2.2. Fix $p \geq 1$ and let $K$ be any convex body. Then, the function $f : K \to \mathbb{R}$ defined by $f(x) = \text{dist}_p(x, \partial K)$ is concave.

3 WHITNEY DECOMPOSITIONS

Hassler Whitney introduced a decomposition of an open set in a Euclidean space into cubes in a seminal paper [33]. The goal of this work was to investigate certain problems involving interpolation. Such decompositions were further developed by Calderón and Zygmund [4]. For more recent uses of decompositions of this type, see Fefferman [11] and Fefferman and Klartag [12]. We begin with the procedure for constructing a Whitney decomposition of $K^o$, i.e. the interior of $K$, for the $p$-norm, along the lines of Theorem 1, page 167 of [32].

As in the statement of Theorem 1.2 we assume that $K \subseteq \{x : \|x\|_p < R_\infty\}$ where $R_\infty < 1$ is a positive real, and that $K \supseteq \{x : \|x\|_p > r_p\}$ for some positive real $r_p$. The assumption $R_\infty < 1$ is made for notational convenience and can be easily enforced by scaling the body if necessary. We discuss in a remark following Theorem 3.1 below as to how to remove this assumption.

Consider the lattice of points in $\mathbb{R}^n$ whose coordinates are integral. This lattice determines a mesh $Q_0$, which is a collection of cubes: namely all cubes of unit side length, whose vertices are points of the above lattice. The mesh $Q_0$ leads to an infinite chain of such meshes $\{Q_k\}_0^\infty$, with $Q_k = 2^{-k}Q_0$. Thus, each cube in the mesh $Q_k$ gives rise to $2^k$ cubes in $Q_{k+1}$ which are termed its children and are obtained by bisecting its sides. The cubes in the mesh $Q_k$ each have sides of length $2^{-k}$ and are thus of $p$-diameter $2^{-k}$. We now inductively define sets $F_i = F_i^{(p)}$, $i \geq 0$, as follows. Let $F_0$ consist of those cubes $Q \in Q_0$ for which $\text{dist}_p(\text{center}(Q), K) \leq \frac{\lambda}{2}$. Fix $\lambda = 1/2$. A cube $Q$ in $F_k$ is subdivided into its children in $Q_{k+1}$ if
\[
\lambda \text{ dist}_p(\text{center}(Q), \partial K) < \text{diam}_p(Q).
\]
which are then declared to belong to $F_{k+1}$. Otherwise $Q$ is not divided and its children are not in $F_{k+1}$.

Let $F = F^{(p)} = \{Q_1, Q_2, \ldots, Q_k, \ldots\}$ denote the set of all cubes $Q$ such that
\[
\text{(1) There exists a } k \text{ for which } Q \in F_k = F_k^{(p)} \text{ but the children of } Q \text{ do not belong to } F_{k+1} = F_{k+1}^{(p)}.
\]
\[
\text{(2) center}(Q) \in K^o.
\]

We will refer to $F^{(p)}$ as a Whitney decomposition of $K$, and the cubes included in $F^{(p)}$ as Whitney cubes. In our notation, we will often suppress the dependence of $F^{(p)}$ on the underlying $\ell_p$ norm when the value of $p$ being used is clear from the context. The following theorem describes the important features of this construction.

Theorem 3.1. Fix $p$ such that $1 \leq p \leq \infty$. Let $R_\infty < 1$ and let $K \subseteq R_\infty \cdot B_\infty$ be a convex body. Then, the following statements hold true for the Whitney decomposition $F = F^{(p)}$ of $K$.

\[
\text{(1) } \bigcup_{Q \in F} Q = K^o. \text{ Further, if } Q \in F, \text{ then } Q \not\in Q_0.
\]
\[
\text{(2) The interiors } Q^o \text{ are mutually disjoint.}
\]
\[
\text{(3) For any Whitney cube } Q \in F,
\]
\[
2 \text{ diam}_p(Q) \leq \text{dist}_p(\text{center}(Q), \mathbb{R}^n \setminus (K^o)) \leq \frac{9}{2} \text{ diam}_p(Q).
\]
\[
\text{(4) For any Whitney cube } Q \in F \text{ and } y \in Q,
\]
\[
\frac{3}{2} \text{ diam}_p(Q) \leq \text{dist}_p(y, \mathbb{R}^n \setminus (K^o)) \leq 5 \text{ diam}_p(Q).
\]

In particular, this is true when dist$_p(y, \mathbb{R}^n \setminus (K^o)) = \text{dist}_p(Q, \mathbb{R}^n \setminus (K^o))$.
\[
\text{(5) The ratio of sidelengths of any two abutting cubes lies in } \{1/2, 1, 2\}.
\]

Remark 3.2. For notational simplicity, we described the construction of Whitney cubes above under the assumption that $K \subseteq R_\infty \cdot B_\infty$ with $R_\infty < 1$. However, it is easy to see that this assumption can be done away with using a simple scaling operation. If $R_\infty > 1$, let $2^a$ be the smallest integral power of two that is larger than $R_\infty$. For any $p$ such that $1 \leq p \leq \infty$, denote by $F^{(p)}(K/2^a)$ the Whitney decomposition of the scaled body $K/2^a$ (which can be constructed as above since $R_\infty /2^a < 1$). Now scale each cube in the decomposition $F^{(p)}(K/2^a)$ up by a factor of $2^a$, and denote this to be the Whitney decomposition $F^{(p)}$ of $K$. Since only linear scalings are performed, all properties guaranteed by Theorem 3.1 for $F^{(p)}(K/2^a)$ remain true for $F^{(p)}$, except possibly for the property that unit cubes $Q \in Q_0$ do not belong to $F^{(p)}$. Henceforth, we will therefore drop the requirement that $K$ has to be strictly contained in $B_\infty$ for it to have a Whitney decomposition $F^{(p)}$.

4 MARKOV CHAINS ON WHITNEY DECOMPOSITIONS

Fix a convex body $K$ as in the statement of Theorem 3.1, and a $p$ such that $1 \leq p \leq \infty$. We now proceed to define the Markov chain $M_p$.

The state space and the stationary distribution. The state space of the chain $M_p$ is the set $F = F^{(p)}$ as in the statement of Theorem 3.1. The stationary distribution $\pi$ is defined as
\[
\pi(Q) := \frac{\text{vol}(Q)}{\text{vol}(K)} \text{ for every } Q \in F.
\]
Transition probabilities. In describing the transition rule below, we will assume that given a point \( x \) which lies in the interior of an unknown cube \( Q \) in \( \mathcal{F}(P) \), we can determine \( Q \). The details of how to algorithmically perform this operation are discussed in the next section.

The transition rule from a cube \( Q \in \mathcal{F} \) is a lazy Metropolis filter, described as follows. With probability \( 1/2 \) remain at \( Q \). Else, pick a uniformly random point \( x \) on the boundary of \( Q \). Item 4 of Theorem 3.1 implies that \( x \) is in the interior \( K^o \) of \( K \). Additionally, pick a point \( x' \) such that \( \|x' - x\|_2 = \frac{\text{sidelength}(Q)}{4} \) and \( x' - x \) is parallel to the unique outward normal of the face that \( x \) belongs to. With probability \( 1 \), there is a unique abutting cube \( Q' \in \mathcal{F} \) which also contains \( x \). By item 5 of Theorem 3.1, \( Q' \) is also characterised by being the unique cube in \( \mathcal{F} \) that contains \( x' \) in its interior. If this abutting cube \( Q' \) has side length greater or equal to \( Q \), then transition to \( Q' \). Otherwise, do the following: with probability \( \frac{\text{side length}(Q')}{\text{side length}(Q)} \) accept the transition to \( Q' \) and with the remaining probability remain at \( Q \).

We now verify that this chain is reversible with respect to the stationary distribution \( \pi \) described in eq. (13). Let \( P(Q; Q') = P_{M_p}(Q; Q') \) denote the probability of transitioning to \( Q' \in \mathcal{F} \) in one step, starting from \( Q \in \mathcal{F} \). We then have
\[
P(Q; Q') = \frac{1}{2} \frac{\text{vol}_{n-1}((Q \cap Q'))}{\text{vol}_{n-1}(Q)} \cdot \min \left( 1, \frac{\text{side length}(Q')}{\text{side length}(Q)} \right).
\]
We thus have (since \( \text{side length}(Q) \cdot \text{vol}_{n-1}(Q) = 2n \cdot \text{vol}(Q) \))
\[
\pi(Q)P(Q; Q') = \frac{\text{vol}(Q)}{\text{vol}(K)} \cdot \frac{\text{vol}_{n-1}((Q \cap Q'))}{\text{vol}_{n-1}(Q)} \cdot \min \left( 1, \frac{\text{side length}(Q')}{\text{side length}(Q)} \right)
\]
\[
= \frac{1}{4n} \cdot \frac{\text{vol}_{n-1}((Q \cap Q'))}{\text{vol}(K)} \cdot \min \left( 1, \frac{\text{side length}(Q')}{\text{side length}(Q)} \right)
\]
\[
= \pi(Q')P(Q'; Q), \quad \text{by its symmetry in } Q \text{ and } Q'.
\]

4.1 Finding the Whitney Cube Containing a Given Point

The above description of our Markov chain assumed that we can determine the Whitney cube \( q \in \mathcal{F} \) that a point \( z \in K^o \) is contained in. We only needed to do this for points \( z \) that are not on the boundary of such cubes, so we assume that \( z \) is contained in the interior of \( q \). In particular, this implies that \( q \) is uniquely determined by \( z \) (by items 1 and 2 of Theorem 3.1).

Suppose that \( \text{side length}(q) = 2^{-b} \), where \( b \) is a currently unknown non-negative integer. Note that since \( z \) lies in the interior of \( q \), the construction of Whitney cubes implies that given \( b \) and \( z \), \( q \) can be uniquely determined as follows: round each coordinate of the point \( 2^b z \) down to its integer floor to get a vertex \( v \in \mathbb{Z}^n \), and then take \( q \) to be the unique axis-aligned cube of side length \( 2^{-b} \) with center at \( 2^{-b} (v + (1/2)/n) \). It thus remains to find \( b \).

Assume now that we have access to an "\( \ell_p \)-distance inequality oracle" for \( K \), which, on input a point \( x \in K \) and an algebraic number \( y \) answers "YES" if
\[
\text{dist}_{\ell_p}(x, \mathbb{R}^n \setminus (K^o)) > y
\]
and "NO" otherwise, along with an "approximate \( \ell_p \)-distance oracle", which outputs an \( 2^{\log 0.01} \)-factor multiplicative approximation \( \tilde{d} \) of \( \text{dist}_{\ell_p}(x, \mathbb{R}^n \setminus (K^o)) \) for any input \( x \in K^o \). When \( p = 1 \), such oracles can be efficiently implemented for any convex body \( K \) with a well-guaranteed membership oracle. However, they may be hard to implement for other \( p \) unless \( K \) has special properties. We discuss this issue in more detail in Section 4.1.1 below: here we assume that we have access to such \( \ell_p \)-distance oracles for \( K \).

Now, since \( \tilde{d} \) is a \( 2^{\log 0.01} \)-factor multiplicative approximation of \( \text{dist}(x) \), item 4 of Theorem 3.1 implies that
\[
2^{-0.01} \cdot \frac{1}{5} \cdot \frac{\tilde{d}}{n^{1/p}} \leq \text{side length}(q) \leq \frac{2}{3} \cdot \frac{\tilde{d}}{n^{1/p}} \cdot 2^{0.01}.
\]

Since \( \text{side length}(q) = 2^{-b} \), this gives
\[
\log_2 \left( \frac{3n^{1/p}}{2d} \right) - 0.01 \leq b \leq \log_2 \left( \frac{3n^{1/p}}{2d} \right) + \log_2 \left( \frac{10}{3} \right) + 0.01.
\]

Let \( b_{\min} \) and \( b_{\max} \) be the lower and upper bounds in eq. (14). Note that the range \( [b_{\min}, b_{\max}] \) has at most two integers. We try both these possibilities for \( b \) in decreasing order, and check for each possibility whether the corresponding candidate \( q \) obtained as in the previous paragraph is subdivided in accordance with eq. (12).

By the construction of Whitney cubes, the first candidate \( q \) that is not subdivided is the correct \( q \) (and least one of the candidate cubes is guaranteed to pass this check). Note that this check requires one call to the \( \ell_p \)-distance inequality oracle for \( K \).

4.1.1 \( \ell_p \)-Distance Oracles for \( K \). When \( p = 1 \), the distance oracle can be implemented to \( O(2^{-L}) \) precision as follows. Given a point \( x \in K^o \), consider for each canonical basis vector \( e_i \) and each sign \( \sigma = \pm 1 \), the ray \( \{ y : (y - x) \in \mathbb{R}^+ \cdot e_i \sigma \} \). The intersection \( H_i \sigma \) of this ray with the boundary of the convex set can be computed to a precision of \( O(2^{-L}) \) using binary search and \( L = O(1) \) calls to the the membership oracle. The \( \ell_1 \) distance to the complement of \( K \) from \( x \) equals \( \min \{ \|H_i \sigma - x\|_1 \ : \ i \in [n], \sigma \in \{-1, 1\} \} \), provided all the \( \|H_i \sigma - x\|_1 \) are finite and the point \( x \) is not in \( K \) otherwise. The implementation of the \( \ell_1 \)-distance inequality oracle also follows from the same consideration: for \( x \in K \), \( \text{dist}_{\ell_1}(x, \partial K) > y \) if and only if all of the points \( \{ x + \sigma e_i : 1 \leq i \leq n, \sigma \in \{-1, 1\} \} \) are in \( K^o \).

When \( p > 1 \), and \( K \) is an arbitrary convex body, there is a non-convex optimization involved in computing the \( \ell_p \)-distance. However, for polytopes with \( m \) faces with explicitly given constraints, the following procedure may be used.

We compute the \( \ell_p \)-distance to each face and then take the minimum. These distances have a closed form expression given as follows. Let \( K \) be the intersection of the halfspaces \( H_i \), where \( H_i \) is given by \( \{ y : a_i \cdot (y - x) \leq 1 \} \). The \( \ell_p \)-distance of \( x \) to \( \mathbb{R}^n \setminus H_i \) is given by
\[
\inf_{y \in \mathbb{R}^n \setminus H_i} \|y - x\|_p = \|a_i\|_{q^{-1}}^{-1},
\]
for \( 1/p + 1/q = 1 \). To see (15), note that for any \( a_i \), equality in \( \|y - x\|_p \cdot \|a_i\|_q \geq 1 \) can be achieved by some \( y \in \mathbb{R}^n \setminus H_i \) by the fact that equality in Hölder’s inequality is achievable for any fixed vector \( a_i \).
A note on numerical precision. Since we are only concerned with walks that run for polynomially many steps, it follows as a consequence of the fact that the ratios of the side lengths of aborting cubes lie in \( \{ \frac{1}{4}, 1, 2 \} \) (item 5 of Theorem 3.1) that the distance to the boundary cannot change in the course of the run of the walk by a multiplicative factor that is outside a range of the form \( \exp(n^{-C}), \exp(n^C) \), where \( C \) is a constant. Due to this, the number of bits needed to represent the side lengths of the cubes used is never more than a polynomial in the parameters \( n, R/r, M \) in Theorem 1.2, and thus \( L \) in the description above can also be chosen to be poly \((n, R/r, \log(M)) \) in order to achieve an “approximate \( t \)-distance oracle” of the form considered in Section 4.1.1.

5 ANALYSIS OF MARKOV CHAINS ON WHITNEY DECOMPOSITIONS

5.1 An Isoperimetric Inequality

In this subsection, we take the first step in our strategy for proving a lower bound on the conductance of the \( \ell_p \)-multiscale chain \( \mathcal{N}_p \), which is to equip \( K \) with a suitable metric and prove an isoperimetric inequality for the corresponding metric-measure space coming from the uniform measure on \( K \). We then relate (in Section 5.2) the conductance of the chain to the isoperimetric profile of the metric-measure space.

The metric we introduce is a kind of degenerate Finsler metric, in which the norms on the tangent spaces are rescaled versions of \( \ell_\infty \), by a factor of \( \text{dist}_{\ell_p}(x, \partial K)^{-1} \) so that the distance to the boundary of \( K \) in the local norm is always greater than \( \Omega \left( n^{-\frac{3}{2}} \right) \). In order to prove the results we need on the isoperimetric profile, we need to lower bound the volume of a tube of thickness \( \delta \) around a subset \( S_1 \) of \( K \) whose measure is less than \( 1/2 \). This is done by considering two cases. First, if \( S_1 \) has a strong presence in the deep interior of \( K \), we look at the intersection of \( S_1 \) with an inner parallel body, and get the necessary results by appealing to existing results of Kannan, Lovász, and Montenegro [15]. The case when \( S_1 \) does not penetrate much into the deep interior of \( K \) constitutes the bulk of the technical challenge in proving this isoperimetric inequality. We handle this case by using a radial needle decomposition to fiber \( S_1 \), and then proving on a significant fraction of these needles an appropriate isoperimetric inequality from which the desired result follows. We now proceed with the technical details.

Equip \( K \) with a family of Minkowski functionals \( F_p : K^\times \times \mathbb{R}^n \to \mathbb{R}_+ \), \( p \geq 1 \), defined by

\[
F_p(x, v) := (\text{dist}_{\ell_p}(x, \partial K))^{-1} ||v||_\infty
\]

for each \( x \in K^\times \) and \( v \in \mathbb{R}^n \). Note that each \( F_p \) is a continuous map that satisfies \( F_p(x, \alpha v) = ||\alpha|| F_p(x, v) \), for each \( x \in K^\times \), \( v \in \mathbb{R}^n \) and \( \alpha \in \mathbb{R} \). Given this, the length \( \text{length}_{\ell_p}(y) \) (for each \( p \geq 1 \)) of any piecewise differentiable curve \( y : [0, 1] \to K^\times \), is defined as

\[
\text{length}_{\ell_p}(y) := \int_0^1 F_p(y(t), y'(t)) \, dt.
\]

(Note that the length of a curve defined as above does not change if the curve is re-parameterized.) This defines a metric on \( K^\times \) as usual: for \( x, y \in K^\times \),

\[
\text{dist}_{\ell_p}(x, y) := \inf_y \int F_p(y(t), y'(t)) \, dt.
\]

where the infimum is taken over all piecewise differentiable curves \( y : [0, 1] \to K^\times \) satisfying \( y(0) = x \) and \( y(1) = y \).

We are now ready to state our isoperimetric inequality.

**Theorem 5.1.** There exist absolute positive constants \( C_0, C_1 \) and \( C_2 \) such that the following is true. Let \( K \) be a convex body such that \( r_p B_p \subseteq K \subseteq R_e B_{\infty} \). Let \( S_1, S_2, S_3 \) be a partition of \( K \) into three parts such that \( \text{dist}_{\ell_p}(S_1, S_2) > \delta \), and \( \text{vol}(S_1) \leq \frac{1}{2} \text{vol}(K) \). Define \( \rho_p := r_p/R_{\infty} \leq 1 \). Then, for \( \delta \leq 1 \), we have the following: if \( \text{vol}(S_1) > \exp(-C_0 n) \cdot \text{vol}(K) \) then

\[
\text{vol}(S_1) \geq C_1 \cdot \left( \frac{\rho_p}{n} \cdot \delta \cdot \text{vol}(S_1) \cdot \log \left( 1 + 0.9 \frac{\text{vol}(K)}{\text{vol}(S_1)} \right) \right),
\]

and if \( \text{vol}(S_1) \leq \exp(-C_0 n) \cdot \text{vol}(K) \) then

\[
\text{vol}(S_1) \geq C_2 \cdot \rho_p \cdot \delta \cdot \text{vol}(S_1).
\]

In the proof of Theorem 5.1, we will need to consider needles analogous to those that appear in the localization lemma of [25]. As described above, however, we will have to analyze such needles in detail in part 2 of the proof. We therefore proceed to list some of their properties that will be needed in the proof.

**Definition 5.2 (Needle).** Fix \( x_0 \in K \) such that \( \text{dist}_{\ell_p}(x_0, \partial K) = \max_{x \in K} \text{dist}_{\ell_p}(x, \partial K) \geq r_p \). By a needle, we mean a set

\[
N_u := K \cap \{ x_0 + tu : t \geq 0 \}
\]

where \( u \in S^{n-1} \) is a unit vector. Also define \( \ell_2(N_u) := \sup \{ t : x_0 + tu \in K \} \) and in general, \( f_p(N_u) := \ell_2(N_u) \cdot ||u||_p \).

Note that by the choice of \( x_0 \), \( \ell_2(N_u) \geq r_p \) for any \( p \) and any needle \( N_u \). Similarly, we have \( \ell_\infty(N_u) \leq 2R_{\infty} \).

Let \( N \) denote the set of all needles. Clearly, \( N \) is in bijection with \( S^{n-1} \), and we will often identify a needle with the corresponding element of \( S^{n-1} \). Let \( \sigma \) denote the uniform (Haar) probability measure on \( S^{n-1} \), and \( c_n \) the \((n-1)\)-dimensional surface area of \( S^{n-1} \). Then, for any measurable subset \( S \) of \( K \), we can use a standard coordinate transformation to polar coordinates followed by Fubini’s theorem (see, e.g., [3, Corollary 2.2] and [9, Theorem 3.12]) to write

\[
\text{vol}(S) = \frac{c_n}{n} \int_{u \in S^{n-1}} \ell_2(N_u)^n \cdot \mu_{N_u}(S \cap N_u) \sigma(du).
\]

where for any needle \( N_u \), the probability measure \( \mu_{N_u} \) on \( N_u \) is defined as follows for every measurable \( S \subseteq K \):

\[
\mu_{N_u}(S \cap N_u) := \frac{n}{\ell_2(N_u)^n} \int_{r=0}^{\ell_2(N_u)} \ell_2(S_3(u + ru)r^{n-1}) \, dr.
\]

More generally, Fubini’s theorem also yields the following. Let \( A \) be a Haar-measurable subset of \( S^{n-1} \), \( S \) a measurable subset of \( K \), and set

\[
T = S \cap \bigcup_{\hat{u} \in A} N_{\hat{u}}.
\]

Then \( T \) itself is measurable and

\[
\text{vol}(T) = \frac{c_n}{n} \int_{\hat{u} \in S^{n-1}} I_A(\hat{u}) \cdot \ell_2(N_{\hat{u}})^n \cdot \mu_{N_{\hat{u}}}(S \cap N_{\hat{u}}) \sigma(du).
\]
The following alternative description of $\mu_N$ will be useful. Let us label the points in $N = N_0$ by their $t_p$ distance, along $N$, from the boundary $\partial K$: thus $x_0$ is labelled $t_p(N)$. In what follows, we will often identify, without comment, a point $x \in N$ with its label $t_p(x) = [0, t_p(N)]$. Note that $t_p(N) = t_p(N) \cdot (1 - t / t_2(N))$. By a slight abuse of notation, we will denote the inverse of the bijective map $t_p(N)$ by $N$. Thus, the label of the point $N(x) \in N$, where $x \in [0, t_p(N)]$, is $x$. The pushforward of $\mu_N$ under the bijective map $t_p(N)$ is then a probability measure on $[0, t_p(N)]$ with the density
\[
\tilde{\mu}_N(x) = \frac{n}{t_p(N)} \left(1 - \frac{x}{t_p(N)}\right)^{n-1}.
\]
(25)

We are now ready to begin with the proof of Theorem 5.1. Some of the more technical derivations are omitted in the proof given below, and these can be found in the full version [29].

**Proof of Theorem 5.1.** Let $c_1 \leq c_2 < 1$ and $\beta < \alpha \leq 1/2$ be positive constants to be fixed later. The proof is divided into two parts, based on the value of $E_{X-S_1} \left[\text{dist}_{t_p}(x, \partial K) \geq c_1 t_p/n\right]$.

**Part 1:** Suppose that $E_{X-S_1} \left[\text{dist}_{t_p}(x, \partial K) \geq c_1 t_p/n\right] \geq \beta$. In this case, let
\[K' = \{x \in K : \text{dist}_{t_p}(x, \partial K) \geq c_1 t_p/n\}.
\]
(26)

Lemma 2.2 implies that $K'$ is convex, and also that $\text{vol}(K') \geq 0.95 \text{vol}(K)$ provided $c_1 \leq 0.05$. Now, by the assumption in this part,
\[\text{vol}(S_1 \cap K') \geq \text{vol}(S_1) \cdot \text{dist}_{\tilde{t}_p}(S_1 \cap K') \geq \beta.
\]
(27)
If $\text{vol}(S_3) \geq \frac{1}{2} \text{vol}(S_1)$, then we already have the required lower bound on the volume of $S_1$. So we assume that $\text{vol}(S_3) \leq \frac{1}{2} \text{vol}(S_1)$, and get $\text{vol}(S_1 \cup S_3) \leq \frac{1}{8} \text{vol}(K)$. Therefore,
\[\text{vol}(S_2 \cap K') \geq \text{vol}(K') - \text{vol}(S_1 \cup S_3) \geq \frac{1}{8} \text{vol}(K).
\]
(28)
Note also that since $\text{dist}_{t_p}(x, \partial K) \geq c_1 t_p/n$ for every $x \in K'$,
\[\text{dist}_{t_p}(S_1 \cap K', S_2 \cap K') \geq \begin{cases} c_1 t_p/n \cdot \text{dist}_{t_p}(S_1 \cap K', S_2 \cap K') \geq \frac{c_1 t_p}{n}. \end{cases}
\]
(29)
The isoperimetric constant of $K'$ can now be bounded from below using a "multiscale" isoperimetric inequality of Kannan, Lovász and Montenegro (Theorem 4.3 of [15]), applied with the underlying norm being the $\ell_p$ norm. Applying this result to $K'$, along with eqs. (27) to (29), we get
\[\text{vol}(S_3) \geq \beta \frac{c_1 t_p}{16 n R_{e_1}} \cdot \text{vol}(S_1) \cdot \left(1 + \frac{9}{10} \cdot \frac{\text{vol}(K)}{\text{vol}(S_1)}\right).
\]
(30)

**Part 2:** Suppose that $E_{X-S_1} \left[\text{dist}_{t_p}(x, \partial K) \geq c_1 t_p/n\right] < \beta$. In this case, for any needle $N$, define the sets
\[N_{in} := N \cap \left\{x : \text{dist}_{t_p}(x, \partial K) > c_1 t_p/n\right\}, \quad \text{and}
\]
\[N_{out} := N \cap \left\{x : \text{dist}_{t_p}(x, \partial K) \leq c_1 t_p/n\right\}.
\]
(31)
(32)

Let $G_{a,c}$ be the set of good needles defined as follows (as a subset of $\mathbb{R}^{n-1}$):
\[G_{a,c} := \{\hat{u} \in \mathbb{R}^{n-1} : \mu_N(S_{\hat{u},in} \cap S_1) < \alpha \mu_N(S_{\hat{u},in} \cap S_1)\}.
\]
(33)

An application of eq. (24) with our assumption for this case gives that
\[\text{vol}\left[\left\{S_1 \cap \hat{u} \in G_{a,c} \right\}\right] \geq \left(1 - \frac{\beta}{\alpha}\right) \text{vol}(S_1).
\]
(34)
which says that when $\beta/\alpha$ is small, a point sampled from $S_1$ is likely to land in a good needle.

It suffices to show that for every $\hat{u} \in G_{a,c}$, we have for the needle $N = N_{\hat{u}}$
\[\mu_N(S_1 \cap N) \geq C \cdot \mu_N(S_1 \cap N).
\]
(35)
for some $C = C(K)$. Indeed, given this, we can use eqs. (24) and (34) to prove that
\[\text{vol}(S_3) \geq C \left(1 - \frac{\beta}{\alpha}\right) \text{vol}(S_1).
\]
(36)
We now prove eq. (35). Given a needle $N \in G_{a,c}$ as above and $c_2 \geq c_1$, define
\[\text{stub}_{c_1}(N) := \{N(x) : x \in \left[0, c_2 t_p(N)/n\right]\}.
\]
(37)
For every $\gamma \in [0, 1]$ we have
\[\frac{\gamma}{2} \leq \mu_N(\text{stub}_{c_1}(N)) = 1 - \left(1 - \frac{\gamma}{n}\right)^n \leq \gamma.
\]
(38)
Note also that since $N(t_p(N)) = x_0$ with $\text{dist}_{t_p}(x_0, \partial K) \geq r_p$, and $N(0) \in \partial K$, Lemma 2.2 implies that
\[\text{dist}_{t_p}(N \left(\frac{\gamma t_p(N)}{n}\right), \partial K) \geq \frac{\gamma t_p}{n}.
\]
(39)
More generally, the concavity of $t_p$, along with the fact that $\text{dist}_{t_p}(x_0, \partial K) \geq r_p > c_1 t_p/n$ implies that the labels of the points in the sets $N_{in}$ and $N_{out}$ form a partition of the interval $[0, t_p(N)]$ into disjoint intervals, with $x_0 = N(t_p(N)) \in N_{in}$ and $N(0) \in N_{out}$. Further, from eq. (39), we get that
\[N_{out} \subseteq N \cap \left[0, c_1 t_p(N)/n\right] \subseteq \text{stub}_{c_1}(N), \quad \text{and}
\]
\[N_{in} \subseteq N \cap \left[0, c_1 t_p(N)/n, t_p(N)\right].
\]
(40)
(41)
whenever $c_1 \leq c_2$.

Some estimates. We now record some estimates that follow directly from the above computations. Recall that $c_1 \leq c_2$. Let $N$ denote stub$_{c_1}(N)$. Similarly, define (see fig. 2)
\[N_{out} := \text{stub}_{c_1}(N), \quad \text{and} \quad N_{in} := \hat{N} \setminus N_{out}.
\]
(42)
From eqs. (40) and (41), we then get that
\[N_{out} \subseteq \hat{N}_{out} \quad \text{and} \quad \hat{N}_{in} \subseteq \hat{N} \cap N_{in}.
\]
(43)
Consequently, it may be shown that when \( N \in G_{\alpha, c_1} \),
\[
\mu_N(\tilde{N}_{in} \cap S_1) \leq \alpha,  \quad (44)
\]
\[
\mu_N(\tilde{N}_{in} \cap (S_2 \cup S_1)) \geq \frac{c_2}{2} - c_1 - \alpha, \quad (45)
\]
\[
\mu_N(S_1 \cap N) \leq \frac{c_1}{1 - \alpha}.  \quad (46)
\]

Proving eq. (35). Since the conclusion of eq. (35) is trivial when \( \mu_N(N \cap S_1) = 0 \), we assume that \( \mu_N(N \cap S_1) > 0 \). As \( N \in G_{\alpha, c_1} \), eqs. (43) and (46) yield
\[
\mu_N(\tilde{N}_{out} \cap S_1) > (1 - \alpha)\mu_N(N \cap S_1) > 0.  \quad (47)
\]

We now have two cases.

**Case 1:** \( \mu_N(\tilde{N}_{in} \cap S_2) = 0 \). In this case, eqs. (45) and (46) imply that
\[
\mu_N(S_3 \cap N) \geq (1 - \alpha) \cdot \frac{c_2 - 2(c_1 + \alpha)}{2c_1} \cdot \mu_N(S_1 \cap N).  \quad (48)
\]

**Case 2:** \( \mu_N(\tilde{N}_{in} \cap S_2) > 0 \). In this case, we define
\[
t' := \inf \left\{ x \in (c_1, c_2) : N(x\ell_p(N)/n) \in S_2 \right\}.  \quad (49)
\]
and note that the assumption for the case means that \( t' \) exists and satisfies \( t' \geq c_1 \). We then define
\[
s := \sup \left\{ x < t' : N(x\ell_p(N)/n) \in S_1 \right\}.  \quad (50)
\]
It follows from eq. (47) that \( s \) is well defined. Using eqs. (38) and (47) with the definition of \( s \), we have
\[
(1 - \alpha)\mu_N(S_1 \cap N) \leq s.  \quad (51)
\]

Now, define
\[
t := \inf \left\{ x > s : N(x\ell_p(N)/n) \in S_2 \right\}.  \quad (52)
\]
Note that \( s \leq t \leq t' \), and the open segment of \( N \) between the points \( N(x\ell_p(N)/n) \) and \( N(t\ell_p(N)/n) \) is contained in \( S_1 \). Thus,
\[
\mu_N(S_3 \cap N) \geq \mu_N \left( \left\{ (x\ell_p(N)/n, t\ell_p(N)/n) \right\} \right).  \quad (53)
\]

Further, since \( \text{dist}_{\ell_p}^2(S_1, S_2) \geq \delta \), we get that the \( \ell_p \) length of the segment from \( N(x\ell_p(N)/n) \) to \( N(t\ell_p(N)/n) \) along \( N \) must also be at least \( \delta \). From eq. (39), we see that for any point \( r \) on this segment,
\[
\text{dist}_{\ell_p}(r, \partial K) \geq \frac{r_p}{\ell_p(N)} \cdot t.  \quad (54)
\]
Using the definition of the \( \ell_p \) metric (see eqs. (16) to (18)) we therefore get
\[
\delta \leq \frac{\text{dist}_{\ell_p}(N(x\ell_p(N)/n), N(t\ell_p(N)/n))}{\frac{r_p}{\ell_p(N)}} \leq \ell_p(N) - t - s.  \quad (55)
\]
Rearranging,
\[
t - s \geq \frac{r_p}{\ell_p(N)} \cdot s.  \quad (56)
\]

Now, a direct calculation using eq. (38) and the convexity of the map \( x \mapsto (1 - x/n)^\alpha \) gives (see [29] for details)
\[
\mu_N(S_3 \cap N) \geq \left(1 - \alpha \right)(1 - c_2) \cdot \frac{r_p\delta}{\ell_p(N)}. \quad (57)
\]
Finally, combining eqs. (48) and (56), eq. (35) holds with
\[
C = \min \left\{ \left(1 - \alpha \right)(1 - c_2) \cdot \frac{r_p\delta}{\ell_p(N)}, \left(1 - \alpha \right) \cdot \frac{c_2 - 2(c_1 + \alpha)}{2c_1} \right\}.  \quad (58)
\]
We now choose \( c_1 = 0.05, c_2 = 0.5, \alpha = 0.1 \) and \( \beta = 0.05 \). Then, since \( \delta \leq 1 \), and \( \ell_p(N) \leq 2\ell_\infty \), the right hand side above is at least \( C' \rho_p \delta \) for some absolute constant \( C' \) (recall from the statement of the theorem that \( \rho_p = r_p/\ell_\infty \leq 1 \)).

We now combine the results for the two parts (eq. (30), and eqs. (36) and (57) and the discussion in the previous paragraph, respectively) to conclude that there exist positive constants \( C'_1 \) and \( C'_2 \) such that
\[
\text{vol}(S_3) \geq \min \left\{ C'_1 \cdot \frac{1}{n} \log \left( 1 + 0.9 \cdot \frac{\text{vol}(K)}{\text{vol}(S_1)} \right), C'_2 \rho_p \delta \cdot \text{vol}(S_1) \right\}.  \quad (59)
\]

The existence of constants \( C_0, C_1 \) and \( C_2 \) as in the statement of the theorem follows immediately from eq. (58), by considering when
each of the two quantities in the minimum above is the smaller one. □

5.2 Bounding the Conductance

In this subsection, we state Theorem 5.3 which gives a lower bound on the conductance of the $M_p$ random walks on Whitney cubes described earlier. In Section 5.3, we further show that in the worst case, the conductance lower bound we obtain here for the $M_p$ random walks is tight up to a factor of $O(\log n)$, where $n$ is the dimension.

**Theorem 5.3.** Fix $p$ such that $1 \leq p \leq \infty$. Let $K$ be a convex body such that $r_p \cdot B_p \subseteq K \subseteq R_\infty \cdot B_\infty$. Define $\rho_p := r_p / R_\infty \leq 1$ as in the statement of Theorem 5.1. The conductance $\Phi = \Phi_{M_p}$ of the chain $M_p$ on the Whitney decomposition $F(p)$ of $K$ satisfies

$$\Phi \geq \frac{\rho_p}{O(n^{1/2} \bar{p})}.$$  \hfill (59)

More precisely, letting $C_0$ be as in the statement of Theorem 5.1, the conductance profile $\Phi_p$ for $x > \exp(-C_0 n)$ satisfies

$$\Phi_p \geq \frac{\rho_p}{O(n^{1/2} \bar{p})} \cdot \log \left(1 + \frac{0.9}{\alpha} \right),$$  \hfill (60)

and for $x \leq \exp(-C_0 n)$, $\Phi_p$ satisfies

$$\Phi_p \geq \frac{\rho_p}{O(n^{1/2} \bar{p})}.$$  \hfill (61)

5.3 Tightness of the Conductance Bound

In the worst case, the conductance lower bound proved above for the $M_p$ chains is tight up to a logarithmic factor in the dimension.

**Proposition 5.4.** Fix $1 \leq p \leq \infty$, and the convex body $K = [-\frac{1}{2}, \frac{1}{2}]^n$, and consider the Markov chain $M_p$ on the Whitney decomposition $F = F(p)$ of $K$. We then have

$$\Phi_{M_p} \leq O\left(\log n / n^{1/2} \bar{p}\right).$$

The above is proved by considering the half-cube $S_1 = K \cap H$, where $H$ is the half-space $\{x : x_1 \leq 0\}$; see [29] for details.

5.4 Rapid Mixing from a Cold Start

Given the conductance bound, rapid mixing from a cold start follows from a result of Lovász and Simonovits [25] ([25, Corollary 1.8], as stated in Lemma 2.1). Recall that we denote the multiscale chain corresponding to the $\ell_2$-norm by $M_p$. We say that a starting density $\eta_0$ is $M$-warm in the $L^2(\pi)$ sense if $\|\eta_0 - 1\|_{L^2(\pi)} \leq M$. Note that if $\|\eta_0\|_\infty \leq M - 1$, then this criterion is satisfied.

**Corollary 5.5.** Let $0 < \epsilon < 1/2$. The mixing time $T$ of $M_p$ to achieve a total variation distance of $\epsilon$, from any $M$-warm start (in the $L^2(\pi)$ sense), obeys

$$T \leq O\left(\frac{n^{4+2/p}}{\rho_p^2} \log \frac{M}{\epsilon}\right).$$

5.5 An Extension: Rapid Mixing from a Given State

In the following theorem, we state an upper bound on the time taken by $M_p$ to achieve a total variation distance of $\epsilon$ from the stationary distribution $\pi$ on the set of cubes starting from a given state. By using the notion of “average conductance” introduced by Lovász and Kannan in [21], we save a multiplicative factor of $\tilde{O}(n)$ (assuming that the starting state is at least 1/poly $(n)$ away from the boundary of the body $K$) from what would be obtained from a direct application of the conductance bound above. This is possible because our lower bound on the value of the conductance profile for small sets is significantly larger than our lower bound on the worst case value of the conductance.

**Theorem 5.6 (Mixing time from a given state).** Fix $p$ such that $1 \leq p \leq \infty$. Let $K$ be a convex body such that $r_p \cdot B_p \subseteq K \subseteq R_\infty \cdot B_\infty$. Define $\rho_p := r_p / R_\infty \leq 1$ as in the statement of Theorem 5.1. Consider the Markov chain $M_p$ defined on the Whitney decomposition $F(p)$ of $K$. Let $X_0$ be a Markov chain evolving according to $M_p$, where $X_0 = Q \in F$. Suppose that $\text{dist}_p(\text{center}(Q), \partial K) = d$. Given $\epsilon \in (0, 1/2)$, after

$$T = C \log \epsilon^{-1} \left(\frac{n^{4+2/p}}{\rho_p^2}\right) \left(\log \frac{n R_\infty}{d} + n^{-1} \log \left(\frac{n}{\rho_p \epsilon}\right)\right)$$

steps, the total variation distance $d_{TV}(X_T, \pi)$ is less than $\epsilon$, for a universal constant $C$.

6 COORDINATE HIT-AND-RUN

Given a convex body $K$ in $\mathbb{R}^n$ and $x_1 \in K$, the steps $x_1, x_2, \ldots$, of the Coordinate Hit-and-Run (CHR) random walk are generated as follows. Given $x_i$, with probability 1/2, we stay at $x_i$. Otherwise, we uniformly randomly draw $j$ from $\{n\}$ and let $\ell$ be the chord $(x_i + e_j \mathbb{R}) \cap K$, and then set $x_{i+1}$ to be a uniformly random point from this segment $\ell$. In this section, we describe the proof of Theorem 1.1, which shows that the CHR random walk on convex bodies mixes rapidly even from a cold start. Our main technical ingredient is an improvement (Theorem 6.2) of an isoperimetric inequality of Laddha and Vempala [18].

6.1 Isoperimetric Inequality for Axis-Disjoint Sets

We need the following definition, due to Laddha and Vempala [18].

**Definition 6.1 (Axis-disjoint sets [18]).** Subsets $S_1, S_2$ of $\mathbb{R}^n$ are said to be axis-disjoint if for all $i \in [n]$, $(S_1 + e_i \mathbb{R}) \cap S_2 = \emptyset$, where $e_i$ is the standard unit vector in the $i$th coordinate direction. In other words, it is not possible to “reach” $S_2$ from $S_1$ by moving along a coordinate direction.

Our main technical result in this direction is the following isoperimetric inequality for axis-disjoint sets.

**Theorem 6.2.** Let $K$ be a convex body in $\mathbb{R}^n$. Denote by $\Phi_{M_{\infty}}$ the conductance of the Markov chain $M_{\infty}$ defined on the Whitney decomposition $F(\infty)$ of $K$. Suppose that $K = S_1 \cup S_2 \cup S_3$ is a partition of $K$ into measurable sets such that $S_1, S_2$ are axis-disjoint. Then,

$$\text{vol}(S_1) \geq \Omega\left(\frac{\Phi_{M_{\infty}}}{n^{3/2}}\right) \cdot \min\{\text{vol}(S_1), \text{vol}(S_2)\}.$$
Combined with the results already proved for the multiscale chain \( M_\infty \), this implies a conductance bound (Theorem 6.3), followed by rapid mixing from a cold start (Theorem 1.1), for the CHR walk. Theorem 6.2 should be compared against the main isoperimetric inequality of Laddha and Vempala [18, Theorem 3; Theorem 2 in the arXiv version]. The result of [18] essentially required the sets \( S_1 \) and \( S_2 \) to be not too small: they proved that for any \( \epsilon > 0 \) and under the same notation as in Theorem 6.2,

\[
\operatorname{vol}(S_1) \geq \Omega \left( \frac{r e}{n^{3/2} \cdot R} \right) \cdot (\min(\operatorname{vol}(S_1), \operatorname{vol}(S_2)) - \epsilon \operatorname{vol}(K)),
\]

(62)

when the body \( K \) satisfies \( rB_2 \subseteq K \subseteq RB_2 \). Such an inequality gives a non-trivial lower bound on the ratio of \( \operatorname{vol}(S_1) \) and \( \min(\operatorname{vol}(S_1), \operatorname{vol}(S_2)) \). Only when the latter is at least \( \epsilon \operatorname{vol}(K) \). Further, due to the \( e \)-pre-factor, the volume guarantee that it gives for \( \operatorname{vol}(S_2) \) as a multiple of \( \min(\operatorname{vol}(S_1), \operatorname{vol}(S_2)) \) degrades as the lower bound imposed on the volumes of the sets \( S_1 \) and \( S_2 \) is lowered. Thus, it cannot lead to a non-trivial lower bound on the conductance of arbitrarily small sets. As discussed in the technical overview, this was the main bottleneck leading to the rapid mixing result of Laddha and Vempala [18] requiring a warm start. The proof strategy employed by Narayanan and Srivastava [30] for the polynomial time mixing of CHR from a warm start was different from that of [18], but still faced a similar bottleneck: non-trivial conductance bounds could be obtained only for sets with volume bounded below. In contrast, Theorem 6.2 allows one to prove a non-trivial conductance bound for sets of arbitrarily small size.

The proof of the inequality in eq. (62) by Laddha and Vempala [18] is built upon an isoperimetry result for cubes. At a high level, they then combined this with a tiling of the body with a lattice of fixed width, to reduce the problem to a classical isoperimetric inequality for the Euclidean metric [25]. In part due the fact that they used a lattice of fixed width, they had to “throw away” some of the mass of the \( K \) lying close to the boundary \( \partial K \), which led to the troublesome \( -\epsilon \operatorname{vol}(K) \) term in eq. (62) above. The inequality in Theorem 6.2 is able to overcome this barrier and provide a non-trivial conductance bound even for small sets based on the following two features of our argument. First, at a superficial level, the multiscale decomposition ensures that we do not have to throw away any mass (on the other hand, not having a tiling of \( K \) by a fixed lattice makes the argument in the proof of Theorem 6.2 more complicated). Second, and more fundamentally, the multiscale decomposition allows us to indirectly use (through the connection to the conductance of the \( M_\infty \) chain) our isoperimetric inequality (Theorem 5.1), which is especially oriented for handling sets with a significant amount of mass close to the boundary \( \partial K \).

6.2 Rapid Mixing of CHR from a Cold Start

Armed with the new isoperimetric inequality for axis-disjoint sets given by Theorem 6.2, we can now replicate the argument of [18] to get a conductance lower-bound bound even for small sets, in place of the \( s \)-conductance lower bound obtained in that paper, which approached zero as the size of the set approached zero.

Theorem 6.3. Let \( K \) be a convex body in \( \mathbb{R}^n \), and let \( \Phi_{M_\infty} \) denote the conductance of the Markov chain \( M_\infty \) on the Whitney decomposition \( \mathcal{F}^{(\infty)} \) of \( K \). Then, the conductance of the coordinate hit-and-run chain on \( K \) is \( \Omega(\Phi_{M_\infty} n^{-5/2}) \).

Corollary 6.4. Let \( K \) be a convex body such that \( \partial^* \cdot B_\infty \subseteq K \subseteq R_\infty \cdot B_\infty \). Let \( \pi \) denote the uniform measure on \( K \). Let \( k \) denote the indicator of \( K \). Let \( 0 < \epsilon < 1/2 \). The number of steps \( T \) needed for \( CHR \) to achieve a density \( \pi_k \) with respect to \( \pi \) such that \( \| \pi_k - \pi \|_2^2 < \epsilon \) is, from a starting density \( \pi_0 \) (with respect to \( \pi \)) that satisfies \( \| \pi_0 - \pi_k \|_2^2 < M \) obeys

\[
T \leq O \left( \frac{n^9 \log M}{\rho_\infty^2} \right).
\]

The proof of Theorem 1.1 follows from Corollary 6.4 via standard arguments. In fact, Theorem 1.1 can be extended further to imply polynomial time mixing for CHR even when the chain is started from a fixed interior point of the body, provided that the point is sufficiently far from the boundary of the body. The details can be found in the full version [29].

REFERENCES

[1] David Applegate and Ravi Kannan. 1991. Sampling and integration of near log-concave functions. In Proc. 23rd ACM Symposium on Theory of Computing (STOC). ACM, 156–163. https://doi.org/10.1145/103418.103439

[2] Russ Bubley, Martin Dyer, and Mark Jerrum. 1998. An elementary analysis of a procedure for sampling points in a convex body. Random Structures & Algorithms 12, 3 (1998), 213–255. https://doi.org/10.1002/(SICI)1098-2418(199805)12:3<213::AID-RSA1>3.0.CO;2-Y

[3] Peter Bürgisser and Felipe Cucker. 2013. Condition. Springer.

[4] A. P. Calderon and A. Zygmund. 1952. On the existence of certain singular integrals. Acta Mathematica 89 (1952), 85 – 139. https://doi.org/10.1007/BF02392130

[5] Yuansi Chen, Raaz Dwivedi, Martin J Wainwright, and Bin Yu. 2018. Fast MCMC Sampling Algorithms on Polytopes. JMLR 19 (2018), 86.

[6] Yuansi Chen and Ronen Eldan. 2022. Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains (extended abstract). In Proc. 63rd IEEE Symposium on Foundations of Computer Science (FOCS). IEEE, 110–122. https://doi.org/10.1109/FOCS54457.2022.00018 arXiv:2203.04163

[7] Martin Dyer and Alan Frieze. 1991. Computing the volume of convex bodies: a case where randomness provably helps. In Proceedings of Symposia in Applied Mathematics, Béla Bollobás (Ed.). Vol. 44. American Mathematical Society, 123–169. https://doi.org/10.1090/psapm/044/1141926

[8] Martin Dyer, Alan Frieze, and Ravi Kannan. 1991. A Random Polynomial-time Algorithm for Approximating the Volume of Convex Bodies. J. ACM 38, 1 (Jan. 1991), 1–17. https://doi.org/10.1145/102782.102783

[9] L.C. Evans and R.F. Gariepy. 2015. Measure Theory and Fine Properties of Functions, Revised Edition. Chapman and Hall/CRC. https://doi.org/10.1201/b21833

[10] Shirin Fallahi, Hans J. Skau, and Guttorm Aalndal. 2020. A comparison of Monte Carlo sampling methods for metabolic network models. FLOS ONE 15, 7 (July 2020), e0235393. https://doi.org/10.1371/journal.pone.0235393

[11] Charles L. Fefferman. 2005. A Sharp Form of Whitney’s Extension Theorem. Annals of Mathematics 161, 1 (2005), 509–577. http://www.jstor.org/stable/3597349

[12] Charles L. Fefferman and Boaz Klartag. 2009. Fitting a C^\infty Smooth Function to Data II. Revista Matemática Iberoamericana 25, 1 (2009), 49 – 273. https://doi.org/10.4171/RMI/569

[13] Hulda S. Haraldsdóttir, Ben Cousins, Ines Thiele, Ronan M. T. Fleming, and Santosh S. Vempala. 2017. CHR: Coordinate hit-and-run with rounding for uniform sampling of constraint-based models. Bioinformatics 33, 11 (June 2017), 1741–1743. https://doi.org/10.1093/bioinformatics/btx652

[14] Mark Jerrum and Alistair Sinclair. 1988. Conductance and the rapid mixing property for Markov chains: the approximation of permanent resolved. In Proc. 20th ACM Symposium on Theory of Computing (STOC). ACM, 235–244. https://doi.org/10.1145/622162.622348

[15] Ravi Kannan, László Lovász, and Ravi Montenegro. 2006. Blocking Conduction and Mixing in Random Walks. Combinatorics, Probability and Computing 15, 4 (July 2006), 541–570. https://doi.org/10.1017/S0963548306007504

[16] Ravindran Kannan and Harishar Narayanan. 2012. Random Walks on Polytopes and an Affine Interior Point Method for Linear Programming. Mathematics of Operations Research 37, 1 (2012), 1–20. www.jstor.org/stable/4141239
Aditi Laddha, Yin Tat Lee, and Santosh S. Vempala. 2020. Strong self-concordance and sampling. In Proc. 52nd ACM SIGACT Symposium on Theory of Computing (STOC). ACM, 1212–1222. https://doi.org/10.1145/3357713.3384272 arXiv:1911.05656.

Aditi Laddha and Santosh S Vempala. 2021. Convergence of Gibbs Sampling: Coordinate Hit-And-Run Mixes Fast. In Proc. 37th International Symposium on Computational Geometry (SOCG). 51:1–51:12. https://drops.dagstuhl.de/volltexte/2021/13850/pdf/LIPIcs-SoCG-2021-51.pdf arXiv:2009.11338.

Yin Tat Lee and Santosh S. Vempala. 2017. Geodesic walks in polytopes. In Proc. 49th ACM Symposium on Theory of Computing (STOC). ACM, 927–940. https://doi.org/10.1145/3055399.3055416

Yin Tat Lee and Santosh S. Vempala. 2022. The Manifold Joys of Sampling. In Proc. 49th International Colloquium on Automata, Languages, and Programming (ICALP) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 229). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 4:1–4:20. https://doi.org/10.4230/LIPIcs.ICALP.2022.4 Invited talk.

László Lovász and Ravi Kannan. 1999. Faster Mixing via Average Conductance. In Proc. 31st Annual ACM Symposium on Theory of Computing (STOC). ACM, 282–287. https://doi.org/10.1145/301250.301317

L. Lovász. 1990. Geometric Algorithms and Algorithmic Geometry. In Proceedings of the International Congress of Mathematicians (ICM), Vol. 1. 139–154.

László Lovász. 1999. Hit-and-run mixes fast. Math. Program. 86, 3 (Dec. 1999), 443–461. https://doi.org/10.1007/s101079900099

László Lovász and Miklós Simonovits. 1999. Random walks in a convex body and an improved volume algorithm. Random Structures & Algorithms 4, 4 (1993), 359–412. https://doi.org/10.1002/rsa.3240040402

László Lovász and Santosh S. Vempala. 2006. Hit-and-Run from a Corner. SIAM J. Comput. 35, 4 (Jan. 2006), 985–1005. https://doi.org/10.1137/S009753970544727X

Oren Mangoubi and Nisheeth K. Vishnoi. 2019. Faster Polytope Rounding, Sampling, and Volume Computation via a Sub-Linear Ball Walk. In Proc. 60th IEEE Symposium on Foundations of Computer Science (FOCS). IEEE, 1338–1357. https://doi.org/10.1109/FOCS.2019.00082

Hariharan Narayanan. 2016. Randomized interior point methods for sampling and optimization. The Annals of Applied Probability 26, 1 (Feb. 2016), 597–641. https://doi.org/10.1214/15-AAP1184

Hariharan Narayanan, Amit Rajaraman, and Piyush Srivastava. 2022. Sampling from convex sets with a cold start using multiscale decompositions. https://doi.org/10.48550/arXiv.2211.04439 Full version of this extended abstract.

Hariharan Narayanan and Piyush Srivastava. 2022. On the mixing time of coordinate Hit-and-Run. Combinatorics, Probability and Computing 31, 2 (2022), 329–332. https://doi.org/10.1017/S0963548321000328 arXiv:2009.14004.

Robert L. Smith. 1984. Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed Over Bounded Regions. Operations Research 32, 6 (1984), 1296–1308. https://www.jstor.org/stable/170949

Elias M. Stein. 1970. Singular Integrals and Differentiability Properties of Functions (PMS-30). Princeton University Press. http://www.jstor.org/stable/j.ctt1bpmb07

Hassler Whitney. 1934. Analytic Extensions of Differentiable Functions Defined in Closed Sets. Trans. Amer. Math. Soc. 36, 1 (1934), 63–89. http://www.jstor.org/stable/1989708

Received 2022-11-07; accepted 2023-02-06