Sampling from convex sets with a cold start using multiscale decompositions

Hariharan Narayanan[*] Amit Rajaraman[†] Piyush Srivastava[‡]

Abstract

A standard approach for sampling approximately uniformly from a convex body $K \subseteq \mathbb{R}^n$ is to run a random walk within K. The requirement is that starting from a suitable initial distribution, the random walk should "mix rapidly", i.e., after a number of steps that is polynomial in n and the aspect ratio R/r (here, K is assumed to contain a ball of radius r and to be contained within a ball of radius R), the distribution of the random walk should come close to the uniform distribution π_K on K. Different random walks differ in aspects such as the ease of implementation of each step, or suitability for a specific class of convex bodies. Therefore, the rapid mixing of a wide variety of random walks on convex bodies has been studied.

Many proofs of rapid mixing of such random walks however require that the initial distribution of the random walk is not too different from the target distribution π_K. In particular, they require that the probability density function of the initial distribution with respect to the uniform distribution π_K on K must be bounded above by $\mathrm{poly}(n)$: this is called a *warm start*. Achieving such a warm start often requires a non-trivial pre-processing step before the random walk can be started. This motivates the problem of proving rapid mixing from "cold starts", i.e., when the density of the initial distribution with respect to π_K can be as high as $\exp(\mathrm{poly}(n))$. In contrast to warm starts, a cold start is usually trivial to achieve. However, rapid mixing from a cold start may not hold for every random walk, e.g., the well-known "ball walk" does not have rapid mixing from an arbitrary cold start. On the other hand, for the "hit-and-run" random walk, Lovász and Vempala proved rapid mixing from a cold start. For the related *coordinate* hit-and-run (CHR) random walk, which has been found to be promising in computational experiments, a rapid mixing result starting from a warm start was proven only recently, while the question of whether CHR mixes rapidly from a cold start remained open.

In this paper, we construct a family of Markov chains inspired by classical multiscale decompositions of subsets of $\mathbb{R}^n$ into countably many axis-aligned cubes. We show that even with a cold start, the mixing times of these chains are bounded by a polynomial in n and the aspect ratio of the body. Our main technical ingredient is an isoperimetric inequality for K for a metric that magnifies distances between points that are close to the boundary of K. As a byproduct of the analysis of this new family of chains, we show that the coordinate hit-and-run (CHR) random walk also mixes rapidly from a cold start, and also from any point that is not too close to the boundary of the body.

[*]Tata Institute of Fundamental Research, Mumbai. Email: hariharan.narayanan@tifr.res.in.

[†]Indian Institute of Technology Bombay. Email: amit.rajaraman@gmail.com.

[‡]Tata Institute of Fundamental Research, Mumbai. Email: piyush.srivastava@tifr.res.in.

HN and PS acknowledge support from the Department of Atomic Energy, Government of India, under project no. RTI4001, 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.

Contents

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 [DFK91]: 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 [AK91, DF91, LS90]: we refer to [LS93] for a more complete discussion of the history. Soon after, Lovász [Lov90] and Lovász and Simonovits [LS93] considered a more geometric random walk not supported on a discrete lattice: the so-called *ball walk*. Here, one fixes a radius parameter δ, and given a current point $x \in K$, proposes a next point y from the Euclidean ball of radius δ centered at x, and moves to y if $y \in K$. They prove (see [LS93, Remark on p. 398]) that when δ is chosen appropriately, the lazy[1] version of the ball walk *mixes rapidly*, i.e., it reaches a distribution that is ϵ-close in total variation distance to the uniform distribution π_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 μ supported on K is said to be M-*warm* if the density function of μ with respect to π_K is bounded above by M.) Another natural geometric random walk is the *hit-and-run* walk (see [Smi84], 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 $\hat{u}$ from the unit sphere $\mathbb{S}^{n-1}$, and then picking y uniformly at random from the chord of K in direction $\hat{u}$ passing through x. Lovász [Lov99] proved that the lazy hit-and-run walk also mixes in time polynomial in n, $1/\epsilon$ and R/r, again assuming that the initial point is sampled from a poly (n)-warm start.

While a poly (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 [LV06].

An interesting variant of the hit-and-run walk is the *coordinate* hit-and-run (CHR) walk, where the direction $\hat{u}$ 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 [Smi84, 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 [HCT+17, FSA20]. However, few theoretical guarantees were known for the CHR walk, and it was only recently that Laddha and Vempala [LV21] and Narayanan and Srivastava [NS22] proved that with a poly (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]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.

1.1 Contributions

We construct a new family of Markov chains inspired by classical multiscale decompositions of bounded sets of $\mathbb{R}^n$ into axis-aligned dyadic (i.e., of sidelength equal to a integral power of two) cubes. Our chains $\mathcal{M}_p$ are parameterized by the ℓ_p norms on $\mathbb{R}^n$, $1 \le p \le \infty$. 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 ϵ total variation distance of the uniform distribution π_K, even when started with an $\exp(\mathrm{poly}\,(n))$-warm start. However, before describing the $\mathcal{M}_p$ 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 ℓ_∞ unit ball B_∞ 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 $r \cdot B_\infty \subseteq K \subseteq R \cdot B_\infty$. Then starting from an M-warm start, the lazy coordinate hit-and-run walk comes within total variation distance at most ϵ of the uniform distribution π_K on K after $O\left(n^9 (R/r)^2 \log(M/\epsilon)\right)$ 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., $\exp(\mathrm{poly}\,(n))$-warm starts. As described above, polynomial time mixing for the CHR walk had only been proved so far starting from a $\mathrm{poly}\,(n)$-warm start [NS22,LV21]: the dependence on M in the mixing time bounds obtained in [NS22,LV21] are proportional to $\mathrm{poly}\,(M)$, as compared to the $\log M$ dependence in our Theorem 1.1. The above result can also be extended to show that CHR mixes after a polynomial number of steps even when the starting distribution is concentrated on a single point in K, provided that the point is not too close to the boundary of K – its ℓ_∞-distance from the boundary of K should be least $R \exp(-\,\mathrm{poly}\,(n))$. See Section 6.3 for further discussion of this extension.

We now proceed to describe our main technical result: the construction of the $\mathcal{M}_p$ random walks and their rapid mixing from a cold start. The random walks $\mathcal{M}_p$ are inspired from the 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 [Whi34] (see, e.g., [Fef05,FK09] for more recent examples of their use). We now informally describe the decomposition of K that we use for the $\mathcal{M}_p$ chain. For simplicity, we assume that K is contained in the interior of the ℓ_∞ 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 ℓ_p-distance of the center of Q from $\mathbb{R}^n \setminus K$ is at least *twice* the ℓ_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° 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 $\mathcal{M}_p$ chains to the coordinate hit-and-run random walk.

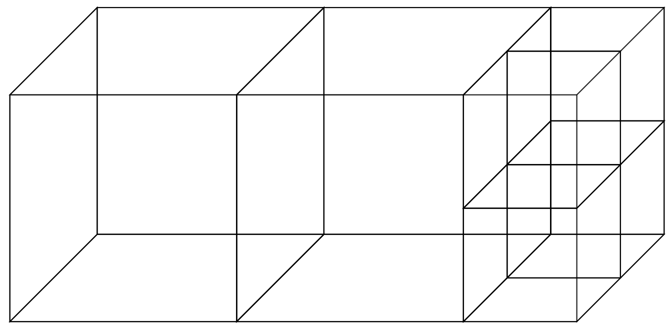


Figure 1: Local geometry of Whitney decompositions: the sidelengths of adjacent cubes are within a factor of two of each other.

The chain $\mathcal{M}_p$ can be seen both as a random walk on K and also as a random walk on the countably infinite set of cubes in the Whitney decomposition $\mathcal{F} = \mathcal{F}^{(p)}$ of K described above, but the latter view is easier to describe first. The stationary distribution π of $\mathcal{M}_p$ is given by $\pi(Q) = \mathrm{vol}\,(Q)\,/\mathrm{vol}\,(K)$ for each cube in $\mathcal{F}^{(p)}$. Given the current cube Q, the walk chooses to stay at Q with probability $1/2$. With the remaining probability, it performs the following step. Pick a point x uniformly at random from the boundary ∂Q 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 π. The Metropolis filter ensures that the walk is in fact *reversible* with respect to π, i.e., $\pi(Q)P_{\mathcal{M}_p}(Q, Q') = \pi(Q')P_{\mathcal{M}_p}(Q', Q)$ where $P_{\mathcal{M}_p}(Q, Q')$ is the probability of transitioning to cube Q' in one step, when starting from cube Q. This implies that π is a stationary distribution of $\mathcal{M}_p$ (see Section 4 for details).

As stated above, $\mathcal{M}_p$ can equivalently be seen as a Markov chain on K itself. To see this, note that corresponding to any probability distribution ν on $\mathcal{F}^{(p)}$, there is a probability distribution ν_K on K obtained by sampling a cube Q from $\mathcal{F}^{(p)}$ according to ν, and then a point x uniformly at random from Q. It is easy to see that the uniform distribution π_K on K can be generated from the distribution π above in this fashion. Further, one can also show that the total variation distance between the distributions ν_K and π_K on K is at most the total variation distance between the distributions ν and π on $\mathcal{F}^{(p)}$ (this follows directly from the definition, and the easy details are given in the proof of Theorem 1.2 on page 27). Similarly, given a probability distribution ν_K on K that is M-warm with respect to π_K, one can obtain a distribution ν on $\mathcal{F}^{(p)}$ that is M-warm with respect to π. This is done as follows. Sample a point x according to ν_K. With probability 1, x lies in the interior of some cube $Q \in \mathcal{F}^{(p)}$ (this follows because ν_K is M-warm with respect to π_K and because the probability measure under π_K of the union of the boundaries of the countably many cubes in $\mathcal{F}^{(p)}$ is zero). ν is then defined to be the probability distribution of this random cube Q. Then $\nu(Q) = \nu_K(Q) \leq M\pi_K(Q) = M\mathrm{vol}\,(Q)\,/\mathrm{vol}\,(K) = M\pi(Q)$.

Our main theorem for $\mathcal{M}_p$ chains is the following. Here, $B_p := \left\{x \in \mathbb{R}^n \,:\, \|x\|_p < 1\right\}$ is the unit ℓ_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 $\mathcal{M}_p$ random walk on K comes within total variation distance at most ϵ of the uniform distribution π_K on K after $O\left(n^{4+\frac{2}{p}} \cdot (R/r)^2 \cdot \log(M/\epsilon)\right)$ 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 $\mathcal{M}_p$ 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 $\frac{1}{\mathrm{poly}(n)}$ 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 $\mathcal{M}_p$ We note that the tools we develop for the analysis of our multi-scale chains $\mathcal{M}_p$ play a crucial rule in our result for the CHR walk (Theorem 1.1). In addition, $\mathcal{M}_p$ chains are of algorithmic interest in their own right. However, it may not be immediately clear how to algorithmically implement each step of the $\mathcal{M}_p$ 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 $\mathcal{M}_1$ 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 $\mathcal{M}_p$ chain requires access to an oracle for the ℓ_p-distance of a point $x \in K$ to the boundary ∂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 π_K on K, the conductance $\Phi_{\mathcal{W}}(S)$ 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 [LS93] that if $\Phi_{\mathcal{W}}(S) \geq 1/\mathrm{poly}\,(n, R/r)$ for every measurable $S \subseteq Q$, then the random walk $\mathcal{W}$ mixes rapidly from a $\exp(\mathrm{poly}\,(n))$-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* [LS93, 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 $\mathrm{poly}\,(n)$-warm start (see, e.g., [LS93, 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 [Lov99] 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 [Lov99, Theorem 3], and therefore required a $\mathrm{poly}\,(n)$-warm start. In contrast, the later proof by Lovász and Vempala [LV06] 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) lower bound of the chain using an *isoperimetric inequality* for an appropriate metric (roughly speaking, an isoperimetric inequality puts a lower bound on $\mathrm{vol}\,(K \setminus (S_1 \cup S_2))$ proportional to the product of volumes of S_1 and S_2 and the distance δ between S_1 and S_2, at least when δ 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 [Lov99], has been to prove such an isoperimetric inequality when the underlying metric is non-Euclidean. For example, the underlying metric in [Lov99] 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

[2] A notable exception to this general strategy is the work of Bubley, Dyer and Jerrum [BDJ98], discussed in more detail later in the introduction.

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 [LV06] 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 [Lov99]: 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 [KN12] of another random walk, called the *Dikin walk*, on polytopes. The Dikin walk was generalized by Narayanan [Nar16] 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 [CDWY18] and by Laddha, Lee, and Vempala [LLV20]. The geodesic walk of Lee and Vempala [LV17] 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 [Lov99] 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 [LV17] and used Jacobi fields among other tools. For a more detailed discussion of these and related developments, we refer to the recent survey [LV22] 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 [LV21] 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 ℓ_0-metric and that are not too small in volume (the ℓ_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 ℓ_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 $\mathcal{M}_p$ chains on Whitney decompositions of K that we introduced above. The conductance analysis of the $\mathcal{M}_p$ chains, in turn, proceeds by introducing a kind of degenerate Finsler metric on K (see Section 5), which is a scaled version of ℓ_∞ that magnifies distances in the vicinity of a point x in K by a factor of $1/\mathrm{dist}_{\ell_p}(x, \mathbb{R}^n \setminus K)$. 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 [KLM06], 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 [LS93]. In the more refined analysis (Theorem 5.6) of the $\mathcal{M}_p$ chain from a fixed state that we alluded to in the remark following Theorem 1.2, we also use results of Lovász and Kannan [LK99] relating rapid mixing to *average conductance* rather than worst-case conductance, thereby saving ourselves a factor of $\tilde{O}(n)$ 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.

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 ∂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 [Lov99]. 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 [LV21, NS22]: in both these proofs, an isoperimetric inequality could only be applied after excluding a part of K close to ∂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 $\mathcal{M}_p$ 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 ℓ_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 results for ℓ_p-norms alone (e.g., those in [LS93, KLM06]) 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 ∂K of K. More concretely, to analyze the chain $\mathcal{M}_p$, we consider the metric g_p which magnifies ℓ_∞-distances in the vicinity of a point $x \in K$ by a factor of $1/\mathrm{dist}_{\ell_p}(x, \partial K)$ (see Section 5 for the formal definition of the metric g_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 g_p metric and the uniform (rescaled Lebesgue) probability measure. We show that $\mathrm{vol}\,(K \setminus (S_1 \cup S_2))$ is significant in proportion to $\min\{\mathrm{vol}\,(S_1), \mathrm{vol}\,(S_2)\}$ 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 a isoperimetric inequality of Kannan, Lovász and Montenegro [KLM06] for the standard ℓ_p norms: this is Part 1 (page 17) of the proof of Theorem 5.1.

The case that requires more work is Part 2 (page 18), which is when a large constant fraction (about 0.95 in our proof) of the volume of S_1 lies within ℓ_p-distance C_1/n of the boundary ∂K of K for some parameter C_1. Our proof of this part is inspired by the localization idea of [LS93], 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 (see Definition 5.2), 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 be too close to the boundary ∂K. This is because the g_p metric magnifies distances close to ∂K, so that a segment that is close to ∂K and is of length δ 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_3-segment in a needle be too *far* from the boundary ∂K. This is because, by definition, a needle 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_3-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 ∂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 ∂K up to a C_2/n distance along the needle, where C_2 is an appropriate factor that depends upon the needle (see eq. (61) and Definition 5.2 for the formal definition). For an appropriate choice of C_1 and C_2, we can show that for at least a constant fraction of needles (see the definition of good needles in eq. (53)), 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.

For a formal description, see eqs. (72) and (74) in conjunction with fig. 2. Together, these facts can be used to show that the inner part of the stub contains a large segment of S_3 (see page 21). 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 ∂K by definition).

Mixing time for the $\mathcal{M}_p$ chains We then show in Section 5.2 that the isoperimetric inequality above implies a conductance lower bound for the $\mathcal{M}_p$ 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 $\mathcal{M}_p$ chain can be used to improve the mixing time from a fixed state by a factor of $\tilde{O}(n)$ 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 $\mathcal{M}_p$ chain is tight up to a logarithmic factor in the dimension.

Rapid mixing from cold start for coordinate hit-and-run Finally, we prove rapid mixing from 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 [LV21, NS22], 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 ∂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 $\mathcal{M}_\infty$ (Theorem 6.3). As discussed above, the conductance of the latter can be bounded from below using the isoperimetry result for the g_∞ 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 [LV21], who had proved an "ℓ_0-isoperimetry" result for such sets. However, as discussed above, their isoperimetry result gives non-trivial conductance 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 ℓ_0-isoperimetry result for axis-disjoint sets (Theorem 6.2) that applies to sets of all sizes, and that involves the conductance of the multiscale $\mathcal{M}_\infty$ chain described above.

Finally, we show in Corollary 6.5 that the mixing result for CHR from a cold start can be extended to show that CHR mixes in polynomially many steps even when started from a point that is not too close to the boundary. Roughly speaking, the mixing time scales with $\log(R/\delta)$ where δ is the ℓ_∞-distance of the starting point to the boundary of the body K and where $K \subseteq R \cdot B_\infty$. This extension formalizes the intuition that after about $O(n \log n)$ steps of CHR (i.e., when the chain has had the opportunity, with high probability, to have made a step in each of the n coordinate directions), the resulting distribution is close to a "cold start" in the sense of the mixing result in Theorem 1.1.

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 $\mathcal{M}_\infty$ walk to the analysis of the coordinate hit-and-run walk is whether the $\mathcal{M}_p$ 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 inequalities for analyzing mixing times for random walks on convex sets is suggested by an interesting paper of Bubley, Dyer and Jerrum [BDJ98], where a certain gauge transformation is used to push forward the uniform measure on a convex set on to a log concave measure supported on $\mathbb{R}^n$, whereafter 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, under 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 isoperimetric 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 $\mathcal{M}_p$ 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 [CE22].

Another direction for investigation would be to make the implementation of each step of $\mathcal{M}_p$, especially in the case $p > 1$, more efficient. In the current naive implementation of a step of the $\mathcal{M}_p$ chain (when $p > 1$) on polytopes that is described in Section 4.1, the distances of a given point x to all the facets of the polytope are computed in order to find the Whitney cube which contains x. In principle, it may be possible to ignore far away facets as has been done by Mangoubi and Vishnoi [MV19] in the context of the ball walk, leading to savings in the implementation time.

2 Preliminaries

2.1 Markov chains

We follow mostly the Markov chain notation used by Lovász and Simonovits [LS93], which we reproduce here for reference. For the following definitions, let $\mathcal{M}$ be a Markov chain on a state space Ω, and let $P(\cdot, \cdot) = P_{\mathcal{M}}(\cdot, \cdot)$ denote the transition kernel of the Markov chain. Let π be the stationary distribution of the chain: this means that for any measurable subset $A \subseteq \Omega$,

$$\int_{x\in\Omega} \pi(dx)P(x,A) = \pi(A). \tag{1}$$

A Markov chain is said to be *lazy* if for every $x \in \Omega$, $P(x,x) \geq 1/2$. All Markov chains we consider in this paper will be lazy.

Definition 2.1 (Ergodic flow and reversible chains)**.** Given measurable subsets A and B of Ω, the *ergodic flow* $\Psi_{\mathcal{M}}(A,B)$ is defined as

$$\Psi_{\mathcal{M}}(A,B) := \int_{x\in A} \pi(dx)P(x,B). \tag{2}$$

Informally, the ergodic flow from A to B is the probability of landing in B after the following process: first sample a point from A with "weight" proportional to π, and then take one step of the chain.

We also denote $\Psi_{\mathcal{M}}(A, \Omega \setminus A)$ as $\Psi_{\mathcal{M}}(A)$. Note that for *any* Markov chain on Ω with stationary distribution π, $\Psi_{\mathcal{M}}(A) = \Psi_{\mathcal{M}}(\Omega \setminus A)$ (see, e.g., [LS93, Section 1.C]). $\mathcal{M}$ is said to be *reversible* with respect to π if $\Psi_{\mathcal{M}}(A,B) = \Psi_{\mathcal{M}}(B,A)$ for all measurable $A, B \subseteq \Omega$.

Definition 2.2 (Conductance)**.** Given a measurable subset A of Ω, the *conductance* $\Phi_{\mathcal{M}}(A)$ of A is defined as $\Psi_{\mathcal{M}}(A)/\pi(A)$. The *conductance* $\Phi_{\mathcal{M}}$ of $\mathcal{M}$ is defined as the infimum of $\Phi_{\mathcal{M}}(S)$ over all measurable $S \subseteq \Omega$ such that $\pi(S) \leq \frac{1}{2}$:

$$\Phi_{\mathcal{M}} := \inf_{S\,:\,\pi(S)\leq 1/2} \Phi_{\mathcal{M}}(S). \tag{3}$$

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

When the underlying chain $\mathcal{M}$ is clear from the context, we will drop the subscript $\mathcal{M}$ from the quantities in the above definitions.

Definition 2.4 (Density and warmth)**.** Given probability distributions π and ν on Ω, we say that ν has *density* f with respect to π if there is a measurable function $f : \Omega \to [0, \infty)$ such that for every measurable subset A of Ω,

$$\nu(A) = \int_{x\in A} f(x)\pi(dx). \tag{4}$$

We will also use the notation $f\pi$ to denote the probability distribution that has density f with respect to π. Note that this implicitly requires that $\mathbb{E}_{X\sim\pi}[f(X)] = \nu(\Omega) = 1$.

A probability distribution ν is said to be *M-warm* with respect to π if it has a density f with respect to π such that $f(x) \leq M$ for all $x \in \Omega$. Note also that π has as density the constant function $\mathbf{1}$ with respect to itself.

Given a probability distribution π on Ω, one can define the norms $L^p(\pi)$, $1 \leq p \leq \infty$ on the set of bounded measurable real valued functions on Ω as follows:

$$\|f\|_{L^p(\pi)} := \left(\int_{x\in\Omega} |f(x)|^p\,\pi(dx)\right)^{1/p} = \mathbb{E}_{X\sim\pi}\left[|f(X)|^p\right]^{1/p}. \tag{5}$$

We will need only the norms $L^1(\pi)$ and $L^2(\pi)$ in this paper. If ν has density f with respect to π, then the *total variation distance* $d_{TV}(\nu, \pi)$ between ν and π can be written as

$$d_{TV}(\nu, \pi) = \sup_{A \subseteq Q} |\pi(A) - \nu(A)| = \frac{1}{2}\|f - \mathbf{1}\|_{L^1(\pi)}. \tag{6}$$

From Jensen's inequality we also have that for every bounded measurable f,

$$\|f\|_{L^1(\pi)} \leq \|f\|_{L^2(\pi)}. \tag{7}$$

Corresponding to $L^2(\pi)$, we also have the inner product

$$\langle f, g\rangle_{L^2(\pi)} := \int_{x\in\Omega} f(x)g(x)\pi(dx) = \mathbb{E}_{X\sim\pi}[f(X)g(X)], \tag{8}$$

so that $\langle f, f\rangle_{L^2(\pi)} = \|f\|^2_{L^2(\pi)}$. Note that any Markov chain $\mathcal{M}$ can be seen as a linear operator acting on probability measures ν on Ω as

$$(\nu\mathcal{M})(A) := \int_{x\in\Omega} P_{\mathcal{M}}(x, A)\nu(dx), \quad \text{for every measurable } A \subseteq \Omega, \tag{9}$$

and also on real valued function f on Ω as

$$(\mathcal{M}f)(x) := \int_{y\in\Omega} f(y)P_{\mathcal{M}}(x, dy), \quad \text{for every } x \in \Omega. \tag{10}$$

When $\mathcal{M}$ is *reversible* with respect to π, we have (see, e.g., [LS93, eq. (1.2)])

$$\langle \mathcal{M}f, g\rangle_{L^2(\pi)} = \langle f, \mathcal{M}g\rangle_{L^2(\pi)}, \tag{11}$$

and also that the probability distribution $\nu\mathcal{M}$ has density $\mathcal{M}f$ with respect to π when the probability distribution ν has density f with respect to π. We will need the following result of Lovász and Simonovits [LS93] connecting the mixing properties of reversible chains to their conductance (the result builds upon previous work of Jerrum and Sinclair for finite-state Markov chains [JS88]).

Lemma 2.1 ([LS93, Corollary 1.8]). *Suppose that the lazy Markov chain $\mathcal{M}$ on Ω is reversible with respect to a probability distribution π on Ω. Let ν_0 have density η_0 with respect to π, and define η_t to be the density of the distribution $\nu_t = \nu_0\mathcal{M}^t$ obtained after t steps of the Markov chain starting from the initial distribution ν_0. Then*

$$\|\eta_t - \mathbf{1}\|^2_{L^2(\pi)} \leq \left(1 - \frac{\Phi^2}{2}\right)^{2t} \|\eta_0 - \mathbf{1}\|^2_{L^2(\pi)},$$

where Φ is the conductance of $\mathcal{M}$.

Proof. Note that since $\mathcal{M}$ is reversible with respect to π, the density η_t of $\nu_t = \nu_0\mathcal{M}^t$ with respect to π is $\mathcal{M}^t\eta_0$. We now apply Corollary 1.8 of [LS93] with f in the statement of that corollary set to $\eta_0 - \mathbf{1}$, and T set to $2t$. This ensures that $\mathbb{E}_{X\sim\pi}[f(X)] = 0$. The corollary then gives

$$\left\langle \eta_0 - \mathbf{1}, \mathcal{M}^{2t}(\eta_0 - \mathbf{1})\right\rangle \leq \left(1 - \frac{\Phi^2}{2}\right)^{2t} \|\nu_0 - \mathbf{1}\|^2_{L^2(\pi)}. \tag{12}$$

Now, by reversibility of $\mathcal{M}$, we get (from eq. (11)) that

$$\left\langle \eta_0 - \mathbf{1}, \mathcal{M}^{2t}(\eta_0 - \mathbf{1})\right\rangle = \left\langle \mathcal{M}^t(\eta_0 - 1), \mathcal{M}^t(\eta_0 - 1)\right\rangle. \tag{13}$$

The claim now follows since, as observed above, the reversibility of $\mathcal{M}$ implies that $\eta_t = \mathcal{M}^t\eta_0$. ◻

2.2 Geometric facts

Notation For any subset $S \subseteq \mathbb{R}^n$ we will denote by S° its open interior (i.e., the union of all open sets contained in S), and by ∂K the *boundary* of S, defined as $\bar{S} \setminus S^\circ$, where $\bar{S}$ 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 $f(x) = \mathrm{dist}_{\ell_p}(x, \partial K)$ is concave.*

Proof. Consider $x, y \in K$ such that $f(x) = \mathrm{dist}_{\ell_p}(x, \partial K) = a$ and $f(y) = \mathrm{dist}_{\ell_p}(y, \partial K) = b$. If both a and b are zero then there is nothing to prove. Otherwise, for any $\lambda \in (0, 1)$, let $z = \lambda x + (1 - \lambda)y$. Now, for any vector v such that $\|v\|_p \leq \lambda a + (1 - \lambda)b$, we have $z + v = \lambda x' + (1 - \lambda)y'$ where $x' := x + \frac{a}{\lambda a + (1-\lambda)b} \cdot v$ and $y' := y + \frac{b}{\lambda a + (1-\lambda)b} \cdot v$. By construction, $\mathrm{dist}_{\ell_p}(x, x') \leq a$ and $\mathrm{dist}_{\ell_p}(y, y') \leq b$ so that x', y', and therefore $z + v$ are all elements of K. Since v was an arbitrary vector with $\|v\|_p \leq \lambda a + (1 - \lambda)b$, this shows that $f(z) = \mathrm{dist}_{\ell_p}(z, \partial K) \geq \lambda a + (1 - \lambda)b = \lambda f(a) + (1 - \lambda)f(b)$. □

In the proof of Theorem 5.6, we will need the following well-known direct consequence of Cauchy's surface area formula (see, e.g., [TV17]).

Proposition 2.5. *Let K and L be convex bodies in $\mathbb{R}^n$ such that $K \subseteq L$. Then $\mathrm{vol}_{n-1}(\partial K) \leq \mathrm{vol}_{n-1}(\partial L)$.*

3 Whitney decompositions

Hassler Whitney introduced a decomposition of an open set in a Euclidean space into cubes in a seminal paper [Whi34]. The goal of this work was to investigate certain problems involving interpolation. Such decompositions were further developed by Calderón and Zygmund [CZ52]. For more recent uses of decompositions of this type, see Fefferman [Fef05] and Fefferman and Klartag [FK09]. We begin with the procedure for constructing a Whitney decomposition of K°, i. e. the interior of K, for the ℓ_p-norm, along the lines of Theorem 1, page 167 of [Ste70].

As in the statement of Theorem 1.2 we assume that $K \subseteq \{x : \|x\|_\infty < R_\infty\}$ 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 $\mathcal{Q}_0$, which is a collection of cubes: namely all cubes of unit side length, whose vertices are points of the above lattice. The mesh $\mathcal{Q}_0$ leads to an infinite chain of such meshes $\{\mathcal{Q}_k\}_0^\infty$, with $\mathcal{Q}_k = 2^{-k}\mathcal{Q}_0$. Thus, each cube in the mesh $\mathcal{Q}_k$ gives rise to 2^n cubes in $\mathcal{Q}_{k+1}$ which are termed its *children* and are obtained by bisecting its sides. The cubes in the mesh $\mathcal{Q}_k$ each have sides of length 2^{-k} and are thus of ℓ_p-diameter $n^{\frac{1}{p}} 2^{-k}$.

We now inductively define sets $\mathcal{F}_i = \mathcal{F}_i^{(p)}$, $i \geq 0$ as follows. Let $\mathcal{F}_0$ consist of those cubes $Q \in \mathcal{Q}_0$ for which $\mathrm{dist}_{\ell_p}(\mathrm{center}(Q), K) \leq \frac{n^{1/p}}{2}$. Fix $\lambda = 1/2$. A cube Q in $\mathcal{F}_k$ is subdivided into its children in $\mathcal{Q}_{k+1}$ if

$$\lambda \, \mathrm{dist}_{\ell_p}(\mathrm{center}(Q), \partial K) < \mathrm{diam}_{\ell_p}(Q), \tag{14}$$

which are then declared to belong to $\mathcal{F}_{k+1}$. Otherwise Q is not divided and its children are not in $\mathcal{F}_{k+1}$.

Let $\mathcal{F} = \mathcal{F}^{(p)} = \{Q_1, Q_2, \ldots, Q_k, \ldots\}$ denote the set of all cubes Q such that

1. There exists a k for which $Q \in \mathcal{F}_k = \mathcal{F}_k^{(p)}$ but the children of Q do not belong to $\mathcal{F}_{k+1} = \mathcal{F}_{k+1}^{(p)}$.

2. $\mathrm{center}(Q) \in K^\circ$.

We will refer to $\mathcal{F}^{(p)}$ as a *Whitney decomposition* of K, and the cubes included in $\mathcal{F}^{(p)}$ as *Whitney cubes*. In our notation, we will often suppress the dependence of $\mathcal{F}^{(p)}$ on the underlying ℓ_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 $1 \le p \le \infty$. 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 $\mathcal{F} = \mathcal{F}^{(p)}$ of K.*

1. *$\bigcup_{Q \in \mathcal{F}} Q = K^\circ$. Further, if $Q \in \mathcal{F}$, then $Q \notin \mathcal{Q}_0$.*

2. *The interiors Q_k° are mutually disjoint.*

3. *For any Whitney cube $Q \in \mathcal{F}$,*

$$2\,\mathrm{diam}_{\ell_p}(Q) \le \mathrm{dist}_{\ell_p}(\mathrm{center}(Q), \mathbb{R}^n \setminus (K^\circ)) \le \frac{9}{2}\,\mathrm{diam}_{\ell_p}(Q).$$

4. *For any Whitney cube $Q \in \mathcal{F}$ and $y \in Q$,*

$$\frac{3}{2}\,\mathrm{diam}_{\ell_p}(Q) \le \mathrm{dist}_{\ell_p}(y, \mathbb{R}^n \setminus (K^\circ)) \le 5\,\mathrm{diam}_{\ell_p}(Q).$$

   In particular, this is true when $\mathrm{dist}_{\ell_p}(y, \mathbb{R}^n \setminus (K^\circ)) = \mathrm{dist}_{\ell_p}(Q, \mathbb{R}^n \setminus (K^\circ))$.

5. *The ratio of sidelengths of any two abutting cubes lies in $\{1/2, 1, 2\}$.*

The proof of this theorem can be found in Appendix A.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_∞. For any p such that $1 \le p \le \infty$, denote by $\mathcal{F}^{(p)}(K/2^a)$ the Whitney decomposition of the scaled body $K/2^a$ (which can be constructed as above since $R_\infty/2^k < 1$). Now scale each cube in the decomposition $\mathcal{F}^{(p)}(K/2^a)$ up by a factor of 2^a, and declare this to be the Whitney decomposition $\mathcal{F}^{(p)}$ of K. Since only linear scalings are performed, all properties guaranteed by Theorem 3.1 for $\mathcal{F}^{(p)}(K/2^a)$ remain true for $\mathcal{F}^{(p)}$, except possibly for the property that unit cubes $Q \in \mathcal{Q}_0$ do not belong to $\mathcal{F}^{(p)}$. Henceforth, we will therefore drop the requirement that K has to be strictly contained in B_∞ for it to have a Whitney decomposition $\mathcal{F}^{(p)}$.

4 Markov chains on Whitney decompositions

Fix a convex body K as in the statement of Theorem 3.1, and a p such that $1 \le p \le \infty$. We now proceed to define the Markov chain $\mathcal{M}_p$.

The state space and the stationary distribution The state space of the chain $\mathcal{M}_p$ is the set $\mathcal{F} = \mathcal{F}^{(p)}$ as in the statement of Theorem 3.1. The stationary distribution π is defined as

$$\pi(Q) := \frac{\mathrm{vol}\,(Q)}{\mathrm{vol}\,(K)} \text{ for every } Q \in \mathcal{F}. \tag{15}$$

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 $\frac{\text{sidelength}(Q')}{\text{sidelength}(Q)}$ 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 π described in eq. (15). Let $P(Q, Q') = P_{\mathcal{M}_p}(Q, Q')$ denote the probability of transitioning to $Q' \in \mathcal{F}$ in one step, starting from $Q \in \mathcal{F}$. We then have

$$P(Q, Q') = \frac{1}{2} \cdot \frac{\text{vol}_{n-1}(\partial Q \cap \partial Q')}{\text{vol}_{n-1}(\partial Q)} \cdot \min\left\{1, \frac{\text{sidelength}(Q')}{\text{sidelength}(Q)}\right\}. \tag{16}$$

We thus have (since $\text{sidelength}(Q) \cdot \text{vol}_{n-1}(Q) = 2n \cdot \text{vol}_n(Q)$)

$$\begin{aligned}
\pi(Q)P(Q, Q') &= \frac{\text{vol}(Q)}{\text{vol}(K)} \cdot \frac{1}{2} \cdot \frac{\text{vol}_{n-1}(\partial Q \cap \partial Q')}{\text{vol}_{n-1}(\partial Q)} \cdot \min\left\{1, \frac{\text{sidelength}(Q')}{\text{sidelength}(Q)}\right\} && (17)\\
&= \frac{1}{4n} \cdot \frac{\text{vol}_{n-1}(\partial Q \cap \partial Q')}{\text{vol}(K)} \cdot \min\{\text{sidelength}(Q'), \text{sidelength}(Q)\} && (18)\\
&= \pi(Q')P(Q', Q), \text{ by its symmetry in } Q \text{ and } Q'. && (19)
\end{aligned}$$

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 $z \in K^\circ$ 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 $\text{sidelength}(q) = 2^{-b}$, 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 $2^{-b}(v + (1/2)\mathbf{1})$. It thus remains to find b.

Assume now that we have access to an "ℓ_p-distance inequality oracle" for K, which, on input a point $x \in K$ and an algebraic number γ answers "YES" if

$$\text{dist}_{\ell_p}(x, \mathbb{R}^n \setminus (K^\circ)) > \gamma$$

and "NO" otherwise, along with an "approximate ℓ_p-distance oracle", which outputs an $2^{\pm 0.01}$-factor multiplicative approximation $\tilde{d}$ of $d(x) := \text{dist}_{\ell_p}(x, \mathbb{R}^n \setminus (K^\circ))$ 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 ℓ_p-distance oracles for K.

Now, since $\tilde{d}$ is a $2^{\pm 0.01}$-factor multiplicative approximation of $d(x)$, item 4 of Theorem 3.1 implies that

$$2^{-0.01} \cdot \frac{1}{5} \cdot \frac{\tilde{d}}{n^{1/p}} \leq \text{sidelength}\,(q) \leq \frac{2}{3} \cdot \frac{\tilde{d}}{n^{1/p}} \cdot 2^{0.01}.$$

Since sidelength $(q) = 2^{-b}$, this gives

$$\left\lceil \log_2\left(\frac{3n^{1/p}}{2d}\right) - 0.01 \right\rceil \leq b \leq \left\lfloor \log_2\left(\frac{3n^{1/p}}{2d}\right) + \log_2\left(\frac{10}{3}\right) + 0.01 \right\rfloor. \tag{20}$$

Let $b_{\min}$ and $b_{\max}$ be the lower and upper bounds in eq. (20). 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. (14). 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 ℓ_p-distance inequality oracle for K.

4.1.1 ℓ_p-distance oracles for K

When $p = 1$, the distance oracle can be implemented to $O(2^{-L})$ precision as follows. Given a point $x \in K^\circ$, consider for each canonical basis vector e_i and each sign $\sigma = \pm 1$, the ray $\{y : (y - x) \in \mathbb{R}_+ \cdot \sigma e_i\}$. The intersection $y_{i,\sigma}$ 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 ℓ_1 distance to the complement of K from x equals $\min(\|y_{i,\sigma} - x\|_1 : i \in [n], \sigma \in \{-1, 1\})$, provided all the $\|y_{i,\sigma} - x\|_1$ are finite and the point x is not in K otherwise. The implementation of the ℓ_1-distance inequality oracle also follows from the same consideration: for $x \in K$, $\mathrm{dist}_{\ell_1}(x, \partial K) > \gamma$ if and only if all of the points $\{x + \sigma\gamma e_i : 1 \leq i \leq n, \sigma \in \{-1, 1\}\}$ are in K°.

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

We compute the ℓ_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 ℓ_p-distance of x to $\mathbb{R}^n \setminus H_i$ is given by

$$\inf_{y \in \mathbb{R}^n \setminus H_i} \|y - x\|_p = \|a_i\|_q^{-1}, \tag{21}$$

for $1/p + 1/q = 1$. To see (21), note that for any a_i, equality in $\|y - x\|_p \cdot \|a_i\|_q \geq 1$, can be achieved by some y in $\overline{\mathbb{R}^n \setminus H_i}$ 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 abutting cubes lie in $\{\frac{1}{2}, 1, 2\}$ (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 $n, R/r$ and 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 ℓ_1 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 ℓ_p-multiscale chain $\mathcal{M}_p$, 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 ℓ_∞, by a factor of $\mathrm{dist}_{\ell_p}(x, \partial K)^{-1}$ so that the distance to the boundary of K in the local norm is always greater than $\Omega\left(n^{-\frac{1}{p}}\right)$. In order to prove the results we need on the isoperimetric profile, we need to lower bound the volume of a tube of thickness δ 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 [KLM06]. 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 (namely those given by eq. (53)) 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 $F_p : K^\circ \times \mathbb{R}^n \to \mathbb{R}_+$, $p \geq 1$, defined by

$$F_p(x, v) := (\mathrm{dist}_{\ell_p}(x, \partial K))^{-1} \|v\|_\infty \tag{22}$$

for each $x \in K^\circ$ 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 $x \in K^\circ, v \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Given this, the length $\mathrm{length}_{g_p}(\gamma)$ (for each $p \geq 1$) of any piecewise differentiable curve $\gamma : [0, 1] \to K^\circ$, is defined as

$$\mathrm{length}_{g_p}(\gamma) := \int_0^1 F_p(\gamma(t), \gamma'(t))\, dt. \tag{23}$$

(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° as usual: for $x, y \in K^\circ$,

$$\mathrm{dist}_{g_p}(x, y) := \inf_\gamma \mathrm{length}_{g_p}(\gamma), \tag{24}$$

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

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 $r_p B_p \subseteq K \subseteq R_\infty B_\infty$. Let S_1, S_2, S_3 be a partition of K into three parts such that $\mathrm{dist}_g(S_1, S_2) > \delta$, and $\mathrm{vol}\,(S_1) \leq \frac{1}{2}\mathrm{vol}\,(K)$. Define $\rho_p := r_p/R_\infty \leq 1$. Then, for $\delta \leq 1$, we have the following: if $\mathrm{vol}\,(S_1) > \exp(-C_0 n) \cdot \mathrm{vol}\,(K)$ then*

$$\mathrm{vol}\,(S_3) \geq C_1 \cdot \frac{\rho_p}{n} \cdot \delta \cdot \mathrm{vol}\,(S_1) \cdot \log\left(1 + 0.9\frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(S_1)}\right). \tag{25}$$

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

$$\mathrm{vol}\,(S_3) \geq C_2 \cdot \rho_p \cdot \delta \cdot \mathrm{vol}\,(S_1)\,. \tag{26}$$

In the proof of Theorem 5.1, we will need to consider needles analogous to those that appear in the localization lemma of [LS93]. 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 some $x_0 \in K$ such that $\mathrm{dist}_{\ell_p}(x_0, \partial K) = \max_{x \in K} \mathrm{dist}_{\ell_p}(x, \partial K) \geq r_p$. By a *needle*, we mean a set

$$N_u := K \cap \{x_0 + tu \, : \, t \geq 0\}$$

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

Note that by the choice of x_0, $\ell_p(N) \geq r_p$ for any p and any needle N. Similarly, we also have $\ell_\infty(N) \leq 2R_\infty$.

Let $\mathcal{N}$ denote the set of all needles. Clearly, $\mathcal{N}$ is in bijection with $\mathbb{S}^{n-1}$, and we will often identify a needle with the corresponding element of $\mathbb{S}^{n-1}$. Let σ denote the uniform (Haar) probability measure on $\mathbb{S}^{n-1}$, and ω_n the $(n-1)$-dimensional surface area of $\mathbb{S}^{n-1}$. 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., [BC13, Corollary 2.2] and [EG15, Theorem 3.12]) to write

$$\begin{aligned}
\mathrm{vol}\,(S) &= \int\limits_{x \in \mathbb{R}^n \setminus \{x_0\}} I_S(x) dx && (27)\\
&= \omega_n \int\limits_{r=0}^{\infty} \int\limits_{\hat{u} \in \mathbb{S}^{n-1}} I_S(x_0 + r\hat{u}) r^{n-1} dr \, \sigma(d\hat{u}) && (28)\\
&= \frac{\omega_n}{n} \int\limits_{\hat{u} \in \mathbb{S}^{n-1}} \ell_2\,(N_{\hat{u}})^n \cdot \mu_{N_{\hat{u}}}(S \cap N_{\hat{u}}) \sigma(du), && (29)
\end{aligned}$$

where for any needle $N_{\hat{u}}$, the probability measure $\mu_{N_{\hat{u}}}$ on $N_{\hat{u}}$ is defined as

$$\begin{aligned}
\mu_{N_{\hat{u}}}(S \cap N_{\hat{u}}) &:= \frac{n}{\ell_2\,(N_{\hat{u}})^n} \cdot \int\limits_{r=0}^{\infty} I_S(x_0 + r\hat{u}) r^{n-1} dr && (30)\\
&= \frac{n}{\ell_2\,(N_{\hat{u}})^n} \cdot \int\limits_{r=0}^{\ell_2(N_{\hat{u}})} I_S(x_0 + r\hat{u}) r^{n-1} dr \quad \text{for every measurable } S \subseteq K. && (31)
\end{aligned}$$

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

$$T = S \cap \bigcup_{\hat{u} \in A} N_{\hat{u}}. \qquad (32)$$

Then T itself is measurable and

$$\frac{\omega_n}{n} \int_{\hat{u} \in \mathbb{S}^{n-1}} I_A(\hat{u}) \cdot \ell_2 \left(N_{\hat{u}}\right)^n \cdot \mu_{N_{\hat{u}}}(S \cap N_{\hat{u}}) \sigma(d\hat{u}) = \omega_n \int_{r=0}^{\infty} \int_{\hat{u} \in \mathbb{S}^{n-1}} I_A(\hat{u}) I_S(x_0 + r\hat{u}) r^{n-1} dr \sigma(d\hat{u}) \tag{33}$$

$$= \omega_n \int_{r=0}^{\infty} \int_{\hat{u} \in \mathbb{S}^{n-1}} I_T(x_0 + r\hat{u}) r^{n-1} dr \sigma(d\hat{u}) \tag{34}$$

$$= \mathrm{vol}\,(T)\,. \tag{35}$$

The following alternative description of $\mu_{N_{\hat{u}}}$ will be useful. Let us label the points in $N = N_{\hat{u}}$ by their ℓ_p distance, along N, from the boundary ∂K: thus x_0 is labelled $\ell_p(N)$. In what follows, we will often identify, without comment, a point $x \in N$ with its label $\mathrm{label}_N(x) \in [0, \ell_p(N)]$. Note that $\mathrm{label}_N(x_0 + t\hat{u}) = \ell_p(N) \cdot (1 - t/\ell_2(N))$. By a slight abuse of notation, we will denote the inverse of the bijective map label_N by N. Thus, the label of the point $N(x) \in N$, where $x \in [0, \ell_p(N)]$, is x. The pushforward of μ_N under the bijective label map label_N is then a probability measure on $[0, \ell_p(N)]$ with the following density, which we denote again by μ_N by a slight abuse of notation:

$$\mu_N(x) = \frac{n}{\ell_p(N)} \left(1 - \frac{x}{\ell_p(N)}\right)^{n-1}. \tag{36}$$

We are now ready to begin with the proof of Theorem 5.1.

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 $\mathbb{P}_{x \sim S_1}\left[\mathrm{dist}_{\ell_p}(x, \partial K) \geq c_1 r_p/n\right]$.

Part 1: Suppose that $\mathbb{P}_{x \sim S_1}\left[\mathrm{dist}_{\ell_p}(x, \partial K) \geq c_1 r_p/n\right] \geq \beta$.
In this case, let

$$K' := \{x \in K : \mathrm{dist}_{\ell_p}(x, \partial K) \geq c_1 r_p/n\}. \tag{37}$$

Lemma 2.2 then implies that K' is convex, and further that $x_0 + \{(1 - c_1/n)\,(z - x_0) \,:\, z \in K\} \subseteq K'$. This inclusion implies that

$$\mathrm{vol}\,(K') \geq 0.95\,\mathrm{vol}\,(K) \tag{38}$$

provided $c_1 \leq 0.05$. Now, by the assumption in this part,

$$\frac{\mathrm{vol}\,(S_1 \cap K')}{\mathrm{vol}\,(S_1)} = \mathbb{P}_{x \sim S_1}[x \in K'] \geq \beta. \tag{39}$$

If $\mathrm{vol}\,(S_3) \geq \frac{1}{2}\mathrm{vol}\,(S_1)$, then we already have the required lower bound on the volume of S_3. So we assume that $\mathrm{vol}\,(S_3) \leq \frac{1}{2}\mathrm{vol}\,(S_1)$, and get $\mathrm{vol}\,(S_1 \cup S_3) \leq \frac{3}{4}\mathrm{vol}\,(K)$ (where we use the fact that $\mathrm{vol}\,(S_1) \leq \frac{1}{2}\mathrm{vol}\,(K)$). From eq. (38), we therefore get

$$\mathrm{vol}\,(S_2 \cap K') \geq \mathrm{vol}\,(K') - \mathrm{vol}\,(S_1 \cup S_3) \geq \frac{1}{8}\mathrm{vol}\,(K)\,. \tag{40}$$

Note also that since $\mathrm{dist}_{\ell_p}(x, \partial K) \geq c_1 r_p/n$ for every $x \in K'$, we have

$$\mathrm{dist}_{\ell_\infty}(S_1 \cap K', S_2 \cap K') \geq \frac{c_1 r_p}{n} \cdot \mathrm{dist}_{g_p}(S_1 \cap K', S_2 \cap K') \geq \frac{c_1 r_p}{n} \cdot \mathrm{dist}_{g_p}(S_1, S_2) \geq \frac{c_1 r_p \delta}{n}. \tag{41}$$

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 [KLM06]), applied with the underlying norm being the ℓ_∞ norm.[3] Applying this result to K', along with eqs. (39) to (41) we get

$$\begin{aligned}
\mathrm{vol}\,(S_3) &\geq \mathrm{vol}\,(S_3 \cap K') && (42)\\
&\geq \frac{c_1 r_p \delta}{2nR_\infty} \cdot \frac{\mathrm{vol}\,(S_1 \cap K')\,\mathrm{vol}\,(S_2 \cap K')}{\mathrm{vol}\,(K')} \cdot \log\left(1 + \frac{\mathrm{vol}\,(K')^2}{\mathrm{vol}\,(S_1 \cap K')\,\mathrm{vol}\,(S_2 \cap K')}\right) && (43)\\
&\geq \frac{\beta c_1 r_p \delta}{16nR_\infty} \cdot \mathrm{vol}\,(S_1) \cdot \log\left(1 + \frac{\mathrm{vol}\,(K')}{\mathrm{vol}\,(S_1)}\right) && (44)\\
&\geq \frac{\beta c_1 r_p \delta}{16nR_\infty} \cdot \mathrm{vol}\,(S_1) \cdot \log\left(1 + \frac{9}{10} \cdot \frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(S_1)}\right), \text{ using eq. (38).} && (45)
\end{aligned}$$

Part 2: Suppose now that $\mathbb{P}_{x \sim S_1}\left[\mathrm{dist}_{\ell_p}(x, \partial K) \geq c_1 r_p/n\right] < \beta$.
In this case, for any needle N, define the sets

$$\begin{aligned}
N_{in} &:= N \cap \left\{x \,:\, \mathrm{dist}_{\ell_p}(x, \partial K) > c_1 r_p/n\right\}, \text{ and} && (46)\\
N_{out} &:= N \cap \left\{x \,:\, \mathrm{dist}_{\ell_p}(x, \partial K) \leq c_1 r_p/n\right\}, && (47)
\end{aligned}$$

and consider the set B_{α,c_1} of *inside-heavy* needles, defined as a subset of $\mathbb{S}^{n-1}$ as

$$B_{\alpha,c_1} := \left\{\hat{u} \,:\, \mu_{N_{\hat{u}}}\left(S_1 \cap N_{\hat{u},in}\right) \geq \alpha \cdot \mu_{N_{\hat{u}}}(N_{\hat{u}} \cap S_1)\right\}. \quad (48)$$

Note that B_{α,c_1} is a measurable subset of $\mathbb{S}^{n-1}$. Integrating this inequality over all such needles using the formula in eq. (35), we get

$$\begin{aligned}
\alpha \cdot \mathrm{vol}\left(S_1 \cap \bigcup_{\hat{u} \in B_{\alpha,c_1}} N_{\hat{u}}\right) &\leq \mathrm{vol}\left(S_1 \cap \left\{x \,:\, \mathrm{dist}_{\ell_p}(x, \partial K) > c_1 r_p/n\right\} \cap \bigcup_{\hat{u} \in B_{\alpha,c_1}} N_{\hat{u}}\right) && (49)\\
&\leq \mathrm{vol}\left(S_1 \cap \left\{x \,:\, \mathrm{dist}_{\ell_p}(x, \partial K) > c_1 r_p/n\right\}\right). && (50)
\end{aligned}$$

By our assumption for this case, we have

$$\frac{\mathrm{vol}\left(S_1 \cap \left\{x \,:\, \mathrm{dist}_{\ell_p}(x, \partial K) > c_1 r_p/n\right\}\right)}{\mathrm{vol}\,(S_1)} = \mathbb{P}_{x \sim S_1}\left[\mathrm{dist}_{\ell_p}(x, \partial K) > c_1 r_p/n\right] < \beta. \quad (51)$$

Substituting eq. (51) in eq. (50) gives

$$\mathrm{vol}\left(S_1 \cap \bigcup_{\hat{u} \in B_{\alpha,c_1}} N_{\hat{u}}\right) \leq \frac{\beta}{\alpha} \cdot \mathrm{vol}\,(S_1). \quad (52)$$

which shows that when β is small enough compared to α, a point sampled randomly from S_1 is unlikely to land in an inside-heavy needle. Let G_{α,c_1} be the set of *good* needles defined as follows (again, as a subset of $\mathbb{S}^{n-1}$):

$$G_{\alpha,c_1} := \mathbb{S}^{n-1} \setminus B_{\alpha,c_1} = \left\{\hat{u} \in \mathbb{S}^{n-1} \,:\, \mu_{N_{\hat{u}}}\left(N_{\hat{u},in} \cap S_1\right) < \alpha \mu_{N_{\hat{u}}}(N_{\hat{u}} \cap S_1)\right\}. \quad (53)$$

[3] The inequality in [KLM06] is stated for distance and diameters in ℓ_2 only, but exactly the same proof works when the distance and the diameters are both in ℓ_∞ (or in any ℓ_p norm), because the final calculation is on a straight line, just as is the case for [LS93]. See Appendix B.1 for details.

Equation (52) then gives that

$$\mathrm{vol}\left(S_1 \cap \bigcup_{\hat{u} \in G_{\alpha,c_1}} N_{\hat{u}}\right) \geq \left(1 - \frac{\beta}{\alpha}\right) \mathrm{vol}\left(S_1\right). \tag{54}$$

Our goal now is to show that for every $\hat{u} \in G_{\alpha,c_1}$, we have for the needle $N = N_{\hat{u}}$

$$\mu_N(S_3 \cap N) \geq C \cdot \mu_N(S_1 \cap N). \tag{55}$$

for some $C = C(K)$. Indeed, given eq. (55), we get that (identifying each needle with the corresponding element in $\mathbb{S}^{n-1}$ for ease of notation)

$$\begin{aligned}
\mathrm{vol}\left(S_3\right) &\geq \mathrm{vol}\left(S_3 \cap \bigcup_{N \in G_{\alpha,c_1}} N\right) && (56)\\
&\overset{\text{eq. (35)}}{=} \frac{\omega_n}{n} \int_{G_{\alpha,c_1}} \ell_2(N)^n \cdot \mu_N(S_3 \cap N)\sigma(dN) && (57)\\
&\overset{\text{eq. (55)}}{\geq} C \cdot \frac{\omega_n}{n} \int_{G_{\alpha,c_1}} \ell_2(N)^n \mu_N(S_1 \cap N)\sigma(dN) && (58)\\
&\overset{\text{eq. (35)}}{=} C \cdot \mathrm{vol}\left(S_1 \cap \bigcup_{N \in G_{\alpha,c_1}} N\right) && (59)\\
&\overset{\text{eq. (54)}}{\geq} C\left(1 - \frac{\beta}{\alpha}\right) \mathrm{vol}\left(S_1\right). && (60)
\end{aligned}$$

We now proceed to the proof of eq. (55). Given a needle $N \in G_{\alpha,c_1}$ as above and $c_2 \geq c_1$, define $\mathrm{stub}_{c_2}(N)$ as

$$\mathrm{stub}_{c_2}(N) := \left\{N(x) \,:\, x \in [0, c_2 \ell_p(N)/n]\right\}. \tag{61}$$

In particular, for every $\gamma \in [0,1]$ we have

$$\frac{\gamma}{2} \leq \mu_N(\mathrm{stub}_\gamma(N)) = 1 - \left(1 - \frac{\gamma}{n}\right)^n \leq \gamma. \tag{62}$$

Note also that since $N(\ell_p(N)) = x_0$ with $\mathrm{dist}_{\ell_p}(x_0, \partial K) \geq r_p$, and $N(0) \in \partial K$, Lemma 2.2 implies that

$$\mathrm{dist}_{\ell_p}\left(N\left(\frac{\gamma \ell_p(N)}{n}\right), \partial K\right) \geq \frac{\gamma r_p}{n}. \tag{63}$$

More generally, the concavity of $\mathrm{dist}_{\ell_p}(\cdot, \partial K)$ along with the fact that $\mathrm{dist}_{\ell_p}(x_0, \partial K) \geq r_p > c_1 r_p/n$ implies that the labels of the points in the sets N_{in} and N_{out} form a partition of the interval $[0, \ell_p(N)]$ into disjoint intervals, with $x_0 = N(\ell_p(N)) \in N_{in}$ and $N(0) \in N_{out}$. Further, from eq. (63), we get that

$$\begin{aligned}
N_{out} &\subseteq N([0, c_1 \ell_p(N)/n]) \subseteq \mathrm{stub}_{c_2}(N), \text{ and} && (64)\\
N_{in} &\supseteq N((c_1 \ell_p(N)/n, \ell_p(N)]). && (65)
\end{aligned}$$

whenever $c_1 \leq c_2$.

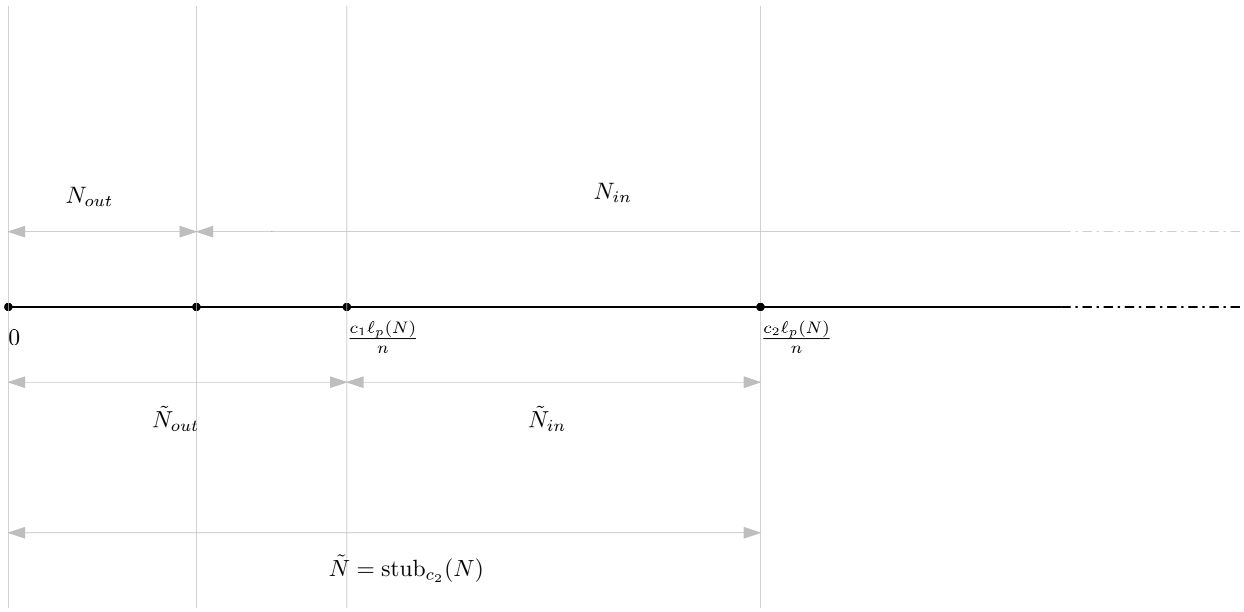


Figure 2: Various parts of a needle

Some estimates We now record some estimates that follow directly from the above computations. Recall that $c_1 \le c_2$. Let $\tilde{N}$ denote $\mathrm{stub}_{c_2}(N)$. Similarly, define (see fig. 2)

$$\tilde{N}_{out} := \mathrm{stub}_{c_1}(N), \qquad \text{and} \qquad \tilde{N}_{in} := \mathrm{stub}_{c_2}(N) \setminus \mathrm{stub}_{c_1}(N). \tag{66}$$

From eqs. (64) and (65), we then get that

$$N_{out} \subseteq \tilde{N}_{out} \qquad \text{and} \qquad \tilde{N}_{in} \subseteq \tilde{N} \cap N_{in}. \tag{67}$$

We then have

$$\mu_N(N_{out}) \overset{\text{eq. (67)}}{\le} \mu_N(\tilde{N}_{out}) \overset{\text{eq. (62)}}{\le} c_1, \tag{68}$$

$$\mu_N(N_{in}) \overset{\text{eq. (68)}}{\ge} 1 - c_1, \text{ since } N_{in} \text{ and } N_{out} \text{ form a partition of } N, \text{ and} \tag{69}$$

$$\mu_N(N_{in} \cap \tilde{N}) \overset{\text{eq. (67)}}{\ge} \mu_N(\tilde{N}_{in}) = \mu_N(\tilde{N}) - \mu_N(\tilde{N}_{out}) \overset{\text{eq. (62)}}{\ge} \frac{c_2}{2} - c_1. \tag{70}$$

Now, when $N \in G_{\alpha,c_1}$, we also have

$$\mu_N(\tilde{N}_{in} \cap S_1) \overset{\text{eq. (67)}}{\le} \mu_N(N_{in} \cap \tilde{N} \cap S_1) \le \alpha, \text{ and} \tag{71}$$

$$\mu_N(N_{in} \cap \tilde{N} \cap (S_2 \cup S_3)) \overset{\text{eq. (67)}}{\ge} \mu_N(\tilde{N}_{in} \cap (S_2 \cup S_3)) \ge \frac{c_2}{2} - c_1 - \alpha. \tag{72}$$

Here, eq. (71) follows from eq. (53) because $N \in G_{\alpha,c_1}$. Equation (72) then follows from eqs. (70) and (71). Along with the fact that N_{in} and N_{out} form a partition of N, eq. (53) for $N \in G_{\alpha,c_1}$ also yields

$$\mu_N(S_1 \cap N) < \frac{\mu_N(S_1 \cap N_{out})}{1-\alpha} \le \frac{\mu_N(N_{out})}{1-\alpha} \overset{\text{eq. (68)}}{\le} \frac{c_1}{1-\alpha}. \tag{73}$$

Proving eq. (55). Since the conclusion of eq. (55) is trivial when $\mu_N(S_1 \cap N) = 0$, we assume that $\mu_N(N \cap S_1) > 0$. As $N \in G_{\alpha,c_1}$, eq. (73) yields

$$\mu_N(\tilde{N}_{out} \cap S_1) \overset{\text{eq. (67)}}{\geq} \mu_N\left(N_{out} \cap S_1\right) \overset{\text{eq. (73)}}{>} (1-\alpha)\mu_N(N \cap S_1) > 0. \tag{74}$$

We now have two cases.

Case 1: $\mu_N(\tilde{N}_{in} \cap S_2) = 0$. In this case, eq. (72) implies that

$$\mu_N(S_3 \cap N) \geq \mu_N(S_3 \cap \tilde{N}_{in}) \geq \frac{c_2}{2} - c_1 - \alpha. \tag{75}$$

Combined with eq. (73), this gives

$$\mu_N(S_3 \cap N) \geq (1-\alpha) \cdot \frac{c_2 - 2(c_1+\alpha)}{2c_1} \cdot \mu_N(S_1 \cap N). \tag{76}$$

Case 2: $\mu_N(\tilde{N}_{in} \cap S_2) > 0$. In this case, we define

$$t' := \inf\left\{x \;:\; N(x\ell_p(N)/n) \in \tilde{N}_{in} \cap S_2\right\} = \inf\left\{x \in (c_1, c_2) \;:\; N(x\ell_p(N)/n) \in S_2\right\}. \tag{77}$$

and note that that the assumption for the case means that t' exists and satisfies $t' \geq c_1$. We then define

$$s := \sup\left\{x < t' \;:\; N(x\ell_p(N)/n) \in S_1\right\}. \tag{78}$$

It follows from eq. (74) that s is well defined. Again using eq. (74) followed by the definition of s, we then have

$$(1-\alpha)\mu_N(S_1 \cap N) \leq \mu_N(S_1 \cap \tilde{N}_{out}) \leq \mu_N(\mathrm{stub}_s(N)) \overset{\text{eq. (62)}}{\leq} s. \tag{79}$$

Now, define

$$t := \inf\left\{x > s \;:\; N(x\ell_p(N)/n) \in S_2\right\}. \tag{80}$$

Note that $s \leq t \leq t'$, and the open segment of N between the points $N(s\ell_p(N)/n)$ and $N(t\ell_p(N)/n)$ is contained in S_3. Thus,

$$\mu_N(S_3 \cap N) \geq \mu_N\left(N\left(\left(s\ell_p(N)/n, t\ell_p(N)/n\right)\right)\right). \tag{81}$$

Further, since $\mathrm{dist}_{g_p}(S_1, S_2) \geq \delta$, we get that the g_p length of the segment from $N(s\ell_p(N)/n)$ to $N(t\ell_p(N)/n)$ along N must also be at least δ. From eq. (63), we see that for any point τ on this segment,

$$\mathrm{dist}_{\ell_p}(\tau, \partial K) \geq \frac{sr_p}{n}. \tag{82}$$

Using the definition of the g_p metric (see eqs. (22) to (24)) we therefore get

$$\delta \leq \frac{\mathrm{dist}_{\ell_\infty}\left(N(s\ell_p(N)/n), N(t\ell_p(N)/n)\right)}{\frac{sr_p}{n}} \leq \frac{\ell_\infty(N)}{r_p}\frac{t-s}{s}, \tag{83}$$

where $\ell_\infty(N)$ is the ℓ_∞ length of N (this is because the segment from $N(s\ell_p(N)/n)$ to $N(t\ell_p(N)/n)$ of N constitutes a $(t-s)/n$ fraction of the length of N, in any ℓ_q norm). Rearranging, we get

$$t - s \geq \frac{r_p}{\ell_\infty(N)} \cdot s\delta. \tag{84}$$

Now, a direct calculation using eq. (62) and the convexity of the map $x \mapsto (1 - x/n)^n$ gives

$$\begin{aligned}
\mu_N(S_3 \cap N) &\geq \mu_N\left(N\left(\left(s\ell_p(N)/n, t\ell_p(N)/n\right)\right)\right), \text{ from eq. (81)}, && (85)\\
&= (1 - s/n)^n - (1 - t/n)^n, \text{ from eq. (62)} && (86)\\
&\geq (1 - t/n)^{n-1}(t - s) \geq (1 - c_2)(t - s), \text{ since } t \leq c_2, && (87)\\
&\geq (1 - c_2) \cdot \delta \cdot \frac{r_p}{\ell_\infty(N)} \cdot s, \text{ from eq. (84)} && (88)\\
&\geq (1 - \alpha) \cdot (1 - c_2) \cdot \delta \cdot \frac{r_p}{\ell_\infty(N)} \cdot \mu_N(S_1 \cap N), \text{ from eq. (79)}. && (89)
\end{aligned}$$

Combining eqs. (76) and (89) we therefore get that eq. (55) holds with

$$C = \min\left\{(1 - \alpha) \cdot (1 - c_2) \cdot \delta \cdot \frac{r_p}{\ell_\infty(N)},\ (1 - \alpha) \cdot \frac{c_2 - 2(c_1 + \alpha)}{2c_1}\right\}. \tag{90}$$

We can now choose $c_1 = 0.05, c_2 = 0.5, \alpha = 0.1$ and $\beta = 0.05$. Then, since $\delta \leq 1$, and $\ell_\infty(N) \leq 2R_\infty$, 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 $\rho_p = r_p/R_\infty \leq 1$).

We now combine the results for the two parts (eq. (45), and eqs. (60) and (90) and the discussion in the previous paragraph, respectively) to conclude that there exist positive constants C_1' and C_2' such that

$$\operatorname{vol}(S_3) \geq \min\left\{C_1', \frac{1}{n}\log\left(1 + 0.9 \cdot \frac{\operatorname{vol}(K)}{\operatorname{vol}(S_1)}\right)\right\} \cdot C_2'\rho_p\delta \cdot \operatorname{vol}(S_1). \tag{91}$$

The existence of constants C_0, C_1 and C_2 as in the statement of the theorem follows immediately from eq. (91), 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 prove Theorem 5.3 which gives a lower bound on the conductance of the $\mathcal{M}_p$ random walks on Whitney cubes described earlier. In the proof, we will need the following two geometric lemmas, whose proofs can be found in Appendix A.1. In Section 5.3, we further show that in the worst case, the conductance lower bound we obtain here for the $\mathcal{M}_p$ random walks is tight up to a factor of $O(\log n)$, where n is the dimension.

Lemma 5.1. *Let K be a convex body, $\mathcal{F}$ a Whitney decomposition of it as described in Section 3, and consider any set $S \subseteq \mathcal{F}$. As before, we identify S also with the union of cubes in S. For any cube $Q \in S$, we have*

$$\operatorname{vol}_{n-1}(\partial S \cap \partial Q) = \lim_{\epsilon \downarrow 0} \frac{\operatorname{vol}_n\left((Q + \epsilon B_\infty) \setminus S\right)}{\epsilon}. \tag{92}$$

Recall that the general definition of surface area uses Minkowski sums with scalings of the Euclidean unit ball B_2. The reason we can work instead with the ℓ_∞-unit ball B_∞ in Lemma 5.1 is because all the surfaces involved are unions of finitely many axis-aligned cuboidal surfaces.

The following lemma relates distances in the ℓ_∞-norm to distances in the g_p metric defined before the statement of Theorem 5.1.

Lemma 5.2. *Fix a convex body $K \subseteq \mathbb{R}^n$. Then there exists $\delta = \delta(n, p)$ such that for all $\epsilon \in [0, \delta]$, and all points $x, y \in K^\circ$ with*

$$\|x - y\|_\infty \geq \epsilon \cdot \mathrm{dist}_{\ell_p}(x, \partial K), \tag{93}$$

it holds that

$$\mathrm{dist}_{g_p}(x, y) \geq \frac{\epsilon}{2}. \tag{94}$$

We are now ready to state and prove our conductance lower bound.

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 $\Phi = \Phi_{\mathcal{M}_p}$ of the chain $\mathcal{M}_p$ on the Whitney decomposition $\mathcal{F}^{(p)}$ of K satisfies*

$$\Phi \geq \frac{\rho_p}{O(n^{2+\frac{1}{p}})}. \tag{95}$$

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

$$\Phi_\alpha \geq \frac{\rho_p}{O(n^{2+\frac{1}{p}})} \cdot \log\left(1 + \frac{0.9}{\alpha}\right), \tag{96}$$

and for $\alpha \leq \exp(-C_0 n)$, Φ_α satisfies

$$\Phi_\alpha \geq \frac{\rho_p}{O(n^{1+\frac{1}{p}})}. \tag{97}$$

Proof. Let $S \subseteq \mathcal{F} = \mathcal{F}^{(p)}$ be such that $\pi(S) \leq (1/2)$. We shall often also view S as the subset of K corresponding to the union of the cubes in it. For each $Q \in S$ and $\epsilon > 0$, consider the set Q_ϵ defined by

$$Q_\epsilon = \bigcup_{x \in Q} \left(x + 2\epsilon\, \mathrm{dist}_{\ell_p}(x, \partial K) B_\infty\right). \tag{98}$$

Then, by item 4 of Theorem 3.1, we have that $\mathrm{dist}_{\ell_p}(x, \partial K) \leq 5\, \mathrm{diam}_{\ell_p}(Q)$. Thus, we get that for some absolute constant $C > 0$

$$Q_\epsilon \subseteq Q + C\epsilon\, \mathrm{diam}_{\ell_p}(Q) B_\infty = Q + \epsilon \cdot C n^{1/p} \mathrm{sidelength}\,(Q)\, B_\infty.$$

Consequently, using Lemma 5.1,

$$\begin{aligned}
\lim_{\epsilon \downarrow 0} \frac{\mathrm{vol}\,(Q_\epsilon \setminus S)}{\epsilon} &\leq \lim_{\epsilon \downarrow 0} \frac{\mathrm{vol}\left((Q + C\epsilon\, \mathrm{diam}_{\ell_p}(Q) B_\infty) \setminus S\right)}{\epsilon} && (99)\\
&= C\, \mathrm{diam}_{\ell_p}(Q) \mathrm{vol}_{n-1}\,(\partial Q \cap \partial S)\\
&= C n^{1/p} \mathrm{sidelength}\,(Q)\, \mathrm{vol}_{n-1}\,(\partial Q \cap \partial S)\,. && (100)
\end{aligned}$$

Define S_ϵ similarly by

$$S_\epsilon = \bigcup_{x \in S} \left(x + 2\epsilon\, \mathrm{dist}_{\ell_p}(x, \partial K) B_\infty\right). \tag{101}$$

Clearly, $S_\epsilon = \bigcup_{Q \in S} Q_\epsilon$. Before proceeding, we also note that it follows from Lemma 5.2 that when $\epsilon \in (0, 1)$ is sufficiently small (as a function of n and p) then

$$\mathrm{dist}_{g_p}(S, (K \setminus S_\epsilon)) \geq \epsilon. \tag{102}$$

We now compute the ergodic flow of $\mathcal{M}_p$ out of S.

$$\begin{aligned}
\Psi(S) &= \sum_{Q\in S}\sum_{Q'\notin S}\pi(Q)P(Q,Q') && (103)\\
&= \sum_{Q\in S}\sum_{Q'\notin S}\frac{1}{2}\cdot\frac{\mathrm{vol}\,(Q)}{\mathrm{vol}\,(K)}\cdot\frac{\mathrm{vol}_{n-1}\,(\partial Q\cap\partial Q')}{\mathrm{vol}_{n-1}\,(\partial Q)}\cdot\min\left\{1,\frac{\mathrm{sidelength}\,(Q')}{\mathrm{sidelength}\,(Q)}\right\} \quad \text{(from eq. (16))} && (104)\\
&= \frac{1}{4n\,\mathrm{vol}\,(K)}\sum_{Q\in S}\mathrm{sidelength}\,(Q)\sum_{Q'\notin S}\mathrm{vol}_{n-1}\,(\partial Q\cap\partial Q')\cdot\min\left\{1,\frac{\mathrm{sidelength}\,(Q')}{\mathrm{sidelength}\,(Q)}\right\} && (105)\\
&\geq \frac{1}{8n\mathrm{vol}\,(K)}\sum_{Q\in S}\mathrm{sidelength}\,(Q)\cdot\mathrm{vol}_{n-1}\,(\partial Q\cap\partial S) \quad \text{(from item 5 of Theorem 3.1)} && (106)\\
&\geq \frac{1}{8Cn^{1+1/p}\mathrm{vol}\,(K)}\sum_{Q\in S}\lim_{\epsilon\downarrow 0}\frac{\mathrm{vol}\,(Q_\epsilon\setminus S)}{\epsilon} \quad \text{(from eq. (100))} && (107)\\
&= \frac{1}{8Cn^{1+1/p}\mathrm{vol}\,(K)}\lim_{\epsilon\downarrow 0}\sum_{Q\in S}\frac{\mathrm{vol}\,(Q_\epsilon\setminus S)}{\epsilon} \quad \text{(from the dominated convergence theorem, see below)}\\
&\geq \frac{1}{8Cn^{1+1/p}\mathrm{vol}\,(K)}\limsup_{\epsilon\downarrow 0}\frac{\mathrm{vol}\,(S_\epsilon\setminus S)}{\epsilon} \quad \left(\text{since } S_\epsilon = \bigcup_{Q\in S} Q_\epsilon\right). && (108)
\end{aligned}$$

Here, to interchange the limit and the sum in eq. (107), one can use, e.g., the dominated convergence theorem, after noting that $\sum_{Q\in S}\mathrm{vol}\,(Q) = \mathrm{vol}\,(S) < \infty$, and that for all $0 < \epsilon \leq 1$, and all $Q \in S$, the discussion following the definition of Q_ϵ in eq. (98) implies that

$$\frac{\mathrm{vol}\,(Q_\epsilon\setminus S)}{\epsilon} \leq \frac{\mathrm{vol}\,(Q_\epsilon\setminus Q)}{\epsilon} \leq \mathrm{vol}\,(Q)\cdot\frac{\left(1+2\epsilon\cdot Cn^{1/p}\right)^n-1}{\epsilon} \leq \mathrm{vol}\,(Q)\cdot\xi_{n,p},$$

where $\xi_{n,p} := \left(1+2Cn^{1/p}\right)^n$ is a finite positive number that depends only on n and p, and not on Q or ϵ. Now, as noted above, Lemma 5.2 implies that when $\epsilon \in (0,1)$ is sufficiently small (as a function of n and p) then $\mathrm{dist}_{g_p}(S,(K\setminus S_\epsilon)) \geq \epsilon$. We can therefore apply Theorem 5.1 after setting $\delta = \epsilon$, $S_1 = S$, $S_3 = (S_\epsilon\setminus S)$, and $S_2 = (K\setminus S_\epsilon)$. Let C_0 be as in the statement of Theorem 5.1. Applying Theorem 5.1, we then get that if S is such that $\pi(S) \leq \exp(-C_0 n)$, then

$$\frac{\mathrm{vol}\,(S_\epsilon\setminus S)}{\epsilon} \geq \Theta(1)\cdot\rho_p\cdot\mathrm{vol}\,(S_1)\,. \tag{109}$$

Substituting this in eq. (108), we thus get that for such S,

$$\Psi(S) \geq \frac{\rho_p}{O(n^{1+\frac{1}{p}})}\cdot\pi(S). \tag{110}$$

We thus get that the value of the conductance profile Φ_α at $\alpha \leq \exp(-C_0 n)$ is

$$\Phi_\alpha \geq \frac{\rho_p}{O(n^{1+\frac{1}{p}})}. \tag{111}$$

Similarly, when S is such that $\frac{1}{2} \geq \pi(S) > \exp(-C_0 n)$, Theorem 5.1 gives

$$\frac{\mathrm{vol}\,(S_\epsilon\setminus S)}{\epsilon} \geq \frac{\rho_p}{O(n)}\cdot\mathrm{vol}\,(S_1)\cdot\log\left(1+0.9\frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(S_1)}\right). \tag{112}$$

Substituting this in eq. (108), we thus get that for such S,

$$\Psi(S) \geq \frac{\rho_p}{O(n^{2+\frac{1}{p}})} \cdot \pi(S) \cdot \log\left(1 + \frac{0.9}{\pi(S)}\right). \tag{113}$$

Combining this with eq. (111), we get that the value of the conductance profile Φ_α at $\alpha > \exp(-C_0 n)$ therefore satisfies

$$\Phi_\alpha \geq \frac{\rho_p}{O(n^{2+\frac{1}{p}})} \cdot \log\left(1 + \frac{0.9}{\alpha}\right). \tag{114}$$

Combining eqs. (111) and (114), we also get that the conductance $\Phi = \Phi_{\mathcal{M}_p}$ of the chain $\mathcal{M}_p$ satisfies

$$\Phi \geq \frac{\rho_p}{O(n^{2+\frac{1}{p}})}. \qquad \square$$

5.3 Tightness of the conductance bound

We now show that at least in the worst case, the conductance lower bound proved above for the $\mathcal{M}_p$ 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 $\mathcal{M}_p$ on the Whitney decomposition $\mathcal{F} = \mathcal{F}^{(p)}$ of K. We then have*

$$\Phi_{\mathcal{M}_p} \leq O\left(\frac{\log n}{n^{2+\frac{1}{p}}}\right).$$

Proof. We consider the half-cube $S_1 = K \cap H$, where H is the half-space $\{x : x_1 \leq 0\}$. Note that the construction of $\mathcal{F}$ implies that the boundary of H does not intersect the interior of any Whitney cube in $\mathcal{F}$. Thus, the set $S \subseteq \mathcal{F}$ of Whitney cubes lying inside S_1 in fact covers S_1 fully. In particular, this implies that $\pi(S) = \pi_K(S_1) = 1/2$.

We now proceed to bound the ergodic flow $\Psi_{\mathcal{M}_p}(S)$ from above. Let $A \subseteq S$ be the set of Whitney cubes whose boundary has a non-zero intersection with the boundary of H. We then note that $P_{\mathcal{M}_p}(q, \mathcal{F} \setminus S) = 0$ when $q \in S \setminus A$ (because $\mathcal{M}_p$ only moves to abutting cubes) and also that

$$P_{\mathcal{M}_p}(q, \mathcal{F} \setminus S) \leq \frac{1}{2n} \text{ when } q \in A, \tag{115}$$

since $\mathcal{M}_p$ chooses a point uniformly at random from the boundary ∂q of q to propose the next step, and only one of the $2n$ faces of q can have a non-trivial intersection with the boundary of H. It follows that

$$\Psi_{\mathcal{M}_p}(S) \leq \frac{1}{2n}\pi(A). \tag{116}$$

We will show now that $\pi(A) \leq O\left(\frac{\log n}{n^{1+\frac{1}{p}}}\right)$, which will complete the proof.

To estimate $\pi(A)$, we recall that sampling a cube from the probability distribution π on $\mathcal{F}$ can also be described as first sampling a point x according to the uniform distribution π_K on K, and then choosing the cube $q \in \mathcal{F}$ containing x (as discussed earlier, with probability 1 over the choice of x, there is a unique q

containing x). Now, let $\tilde{A}$ denote the set of those $x \in S_1$ which are at ℓ_p-distance at most $\frac{10 \log n}{n}$ from the boundary of K, and also at ℓ_∞-distance at least $\frac{100 \log n}{n^{1+\frac{1}{p}}}$ from the boundary ∂H of H. Formally,

$$\tilde{A} := \left\{ x \,:\, \mathrm{dist}_{\ell_p}(x, \partial K) \leq \frac{10 \log n}{n} \right\} \cap \left\{ x \,:\, \mathrm{dist}_{\ell_\infty}(x, \partial H) \geq \frac{100 \log n}{n^{1+\frac{1}{p}}} \right\} \cap S_1 \tag{117}$$

The first condition for x being in $\tilde{A}$ implies, due to item 4 of Theorem 3.1, that the Whitney cube $q \in \mathcal{F}$ containing x has ℓ_p-diameter at most $\frac{10 \log n}{n}$. The sidelength of this q is therefore at most $\frac{10 \log n}{n^{1+\frac{1}{p}}}$. Combined with the second condition in the definition of $\tilde{A}$, this implies that if $x \in \tilde{A}$, then $q \in S$ and $q \notin A$. Since $A \subseteq S$, we can now use the translation described above between the probability distributions π on $\mathcal{F}$ and π_K on K to get

$$\pi_K(\tilde{A}) = \mathbb{P}_{x \sim \pi_K}\left[x \in \tilde{A} \right] \leq \mathbb{P}_{q \sim \pi}\left[q \in S \text{ and } q \notin A \right] = \pi(S \setminus A) = \pi(S) - \pi(A). \tag{118}$$

Now, note that since K is an axis aligned cube, we have $\mathrm{dist}_{\ell_p}(x, \partial K) = \mathrm{dist}_{\ell_\infty}(x, \partial K)$ for any $x \in K$ and any ℓ_p-norm, where $1 \leq p \leq \infty$. Using this, a direct calculation gives

$$\pi_K(\tilde{A}) \geq \frac{1}{2} \cdot \left(1 - \left(1 - \frac{20 \log n}{n} \right)^n \right) - \frac{100 \log n}{n^{1+\frac{1}{p}}} \geq \frac{1}{2}\left(1 - n^{-20} \right) - \frac{100 \log n}{n^{1+\frac{1}{p}}}. \tag{119}$$

Plugging this into eq. (118) and using $\pi(S) = 1/2$ gives $\pi(A) \leq n^{-20}/2 + \frac{100 \log n}{n^{1+\frac{1}{p}}}$. Using eq. (116), this gives $\Psi_{\mathcal{M}_p}(S) \leq O\left(\frac{\log n}{n^{2+\frac{1}{p}}} \right)$. Since $\pi(S) = 1/2$, this yields the claimed upper bound on the conductance $\Phi_{\mathcal{M}_p}(S) = \Psi_{\mathcal{M}_p}(S)/\pi(S)$ of S. □

5.4 Rapid mixing from a cold start: Proof of Theorem 1.2

Given the conductance bound, the proof for rapid mixing from a cold start will from a result of Lovász and Simonovits [LS93] ([LS93, Corollary 1.8], as stated in Lemma 2.1). Recall that we denote the multiscale chain corresponding to the ℓ_p-norm by $\mathcal{M}_p$. We say that a starting density η_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 $\mathcal{M}_p$ to achieve a total variation distance of ϵ, from any M-warm start (in the $L^2(\pi)$ sense), obeys*

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

Proof. If $\frac{R}{r}$ is bounded above by a polynomial in n, so is ρ_p^{-1}. Recall that $f\pi$ denotes the probability distribution with density f with respect to π. As in Lemma 2.1, let η_T denote the density (with respect to π) of the distribution obtained after T steps of chain. Since we have

$$2\, d_{TV}(\eta_T \pi, \pi) = \|\eta_T - 1\|_{L^1(\pi)} \leq \sqrt{\langle \eta_T - 1, \eta_T - 1 \rangle_{L^2(\pi)}},$$

this corollary follows from Lemma 2.1 and the fact that

$$\Phi \geq \frac{\rho_p}{O(n^{2+\frac{1}{p}})},$$

as shown in eq. (95) in Theorem 5.3. □

Proof of Theorem 1.2. The theorem follows immediately from Corollary 5.5 after a few substitutions. Recall that the notation $f\nu$ denotes the probability distribution that has density f with respect to ν. Let η_0' be the density of the initial M-warm start (with respect to π_K) in the statement of Theorem 1.2, and let $\nu_0' = \eta_0' \pi_K$ be the corresponding probability distribution on K. Let η_0 be the density (with respect to the distribution π on $\mathcal{F}^{(p)}$) of the probability distribution ν_0 on $\mathcal{F}^{(p)}$ obtained by first sampling a point x according to ν_0' and then choosing the cube in $\mathcal{F}_q$ in which x lies. As argued in the paragraph preceding Theorem 1.2, the probability distribution ν_0 is also M-warm with respect to π in the sense that $\|\eta_0\|_\infty \leq M$. This implies that $\|\eta_0 - 1\|_{L^2(\pi)} \leq M + 1$. Next, note that $\rho_p = R/r$ by definition, where R and r are as in the statement of Theorem 1.2. Corollary 5.5 then implies that the total variation distance between the probability distributions $\nu_T := \eta_T \pi$ and π on $\mathcal{F}^{(p)}$ is at most ϵ. This implies that for any function $f : \mathcal{F}^{(p)} \to [0, 1]$,

$$\mathbb{E}_{Q\sim\pi}[f(X)] - \mathbb{E}_{Q\sim\nu_T}[f(X)] \leq \epsilon. \tag{120}$$

To prove Theorem 1.2, we now need to show that the probability distribution $\nu_T' := \eta_T' \pi_K$ on K obtained by first sampling a cube Q from $\nu_T = \eta_T \pi$, and then sampling a point uniformly at random from Q (as discussed in the paragraph preceding Theorem 1.2) is within total variation distance at most ϵ from the uniform distance π_K on K. To do this, we only need to show that for any measurable subset S of K, we have

$$\pi_K(S) - \nu_T'(S) \leq \epsilon. \tag{121}$$

Recall that π_W denotes the uniform probability distribution on W, where W is any measurable set. For any measurable subset S of K and a cube $Q \in \mathcal{F}^{(p)}$, we then have $\pi_K(S \cap Q) = \pi_Q(S \cap Q)\pi_K(Q) = \pi_Q(S \cap Q)\pi(Q)$ and $\nu_T'(S \cap Q) = \pi_Q(S \cap Q)\nu_T(Q)$. Now, since the cubes in $\mathcal{F} = \mathcal{F}^{(p)}$ form a countable partition of K° (items 1 and 2 of Theorem 3.1), we get that

$$\pi_K(S) - \nu_T'(S) = \sum_{Q\in\mathcal{F}} \pi_Q(S \cap Q)\pi(Q) - \sum_{Q\in\mathcal{F}} \pi_Q(S \cap Q)\nu_T(Q) \tag{122}$$

$$= \mathbb{E}_{Q\sim\pi}\left[\pi_Q(S \cap Q)\right] - \mathbb{E}_{Q\sim\nu_T}\left[\pi_Q(S \cap Q)\right] \overset{\text{eq. (120)}}{\leq} \epsilon. \tag{123}$$

Here, the inequality in eq. (120) is applied with $f(Q) := \pi_Q(S \cap Q) \in [0, 1]$. Equation (123) thus proves eq. (121) and hence completes the proof of the theorem. □

5.5 An extension: rapid mixing from a given state

In the following theorem, we state an upper bound on the time taken by $\mathcal{M}_p$ to achieve a total variation distance of ϵ from the stationary distribution π on the set of cubes starting from a given state. By using the notion of "average conductance" introduced by Lovász and Kannan in [LK99], we save a multiplicative factor of $\tilde{O}(n)$ (assuming that the starting state is at least $1/\text{poly}(n)$ 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 $\mathcal{M}_p$ defined on the Whitney decomposition $\mathcal{F}^{(p)}$ of K. Let X_t be a Markov chain evolving according to $\mathcal{M}_p$, where $X_0 = Q \in \mathcal{F}$. Suppose that $\text{dist}_{\ell_p}(\text{center}(Q), \partial K) = d$. Given $\epsilon \in (0, 1/2)$, after*

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

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

In the proof of this theorem, we will need an auxiliary finite version of the $\mathcal{M}_p$ chain, that we now proceed to define. For notational convenience, we will also assume in the proof that the body K is scaled so that $R_\infty \leq 1$ (to translate the calculations below back to the general setting, we will need to replace d by d/R_∞).

5.5.1 A family of auxiliary chains $\mathcal{M}_{p,a}$

We consider a finite variant of the multiscale chain $\mathcal{M}_p$, which we term $\mathcal{M}_{p,a}$ where p corresponds to the ℓ_p-norm used and $a \geq 1$ is a natural number. In this chain, all the states Q in $\mathcal{M}_p$ that correspond to cubes of side length less or equal to 2^{-a} are fused into a single state, which we call Q_∞, which we also identify with the union of these cubes. The transition probabilities $P_{p,a}$ of $\mathcal{M}_{p,a}$ to and from Q_∞ are defined as follows, in terms of the transition probabilities P_p of $\mathcal{M}_p$. Suppose (the interior of) Q' is disjoint from Q_∞.

$$P_{p,a}(Q', Q_\infty) := \sum_{\mathcal{F} \ni Q \subseteq Q_\infty} P_p(Q', Q).$$

$$P_{p,a}(Q_\infty, Q') := \frac{\sum\limits_{\mathcal{F} \ni Q \subseteq Q_\infty} \pi(Q) P_p(Q, Q')}{\sum\limits_{\mathcal{F} \ni Q \subseteq Q_\infty} \pi(Q)}.$$

For two cubes Q, Q' that are disjoint from Q_∞,

$$P_{p,a}(Q, Q') := P_p(Q, Q').$$

Finally

$$P_{p,a}(Q_\infty, Q_\infty) := \frac{\sum\limits_{\mathcal{F} \ni Q'' \subseteq Q_\infty} \sum\limits_{\mathcal{F} \ni Q \subseteq Q_\infty} \pi(Q) P_p(Q, Q'')}{\sum\limits_{\mathcal{F} \ni Q \subseteq Q_\infty} \pi(Q)}.$$

We now proceed with the proof of Theorem 5.6.

Proof of Theorem 5.6. To avoid cluttering of notation in this proof, we will adopt the standard convention that different occurrences of the letters C and c can refer to different absolute constants.

By item 3 of Theorem 3.1, for any Whitney cube $Q \in \mathcal{F}$,

$$\mathrm{dist}_{\ell_p}(\mathrm{center}(Q), \mathbb{R}^n \setminus (K^\circ)) \leq \frac{9}{2} \mathrm{diam}_{\ell_p}(Q).$$

Let 2^{-b} denote the side length of Q, where b is a positive natural number. Thus,

$$2^{-b} = n^{-\frac{1}{p}} \mathrm{diam}_{\ell_p}(Q) \geq \frac{2dn^{-\frac{1}{p}}}{9}.$$

With the notation of Lemma 2.1, we have

$$\langle \eta_0 - 1, \eta_0 - 1 \rangle_{L^2(\pi)} = 2^{-nb}(2^{2nb})(\mathrm{vol}\,(K)) - 2(2^{-nb})(2^{nb}) + 1 < 2^{nb}(\mathrm{vol}\,(K)).$$

Thus, using the inequality $1 - x \le \exp(-x)$ for positive x together with Lemma 2.1, we have

$$\langle \eta_T - 1, \eta_T - 1 \rangle_{L^2(\pi)} \le \exp\left(-\Phi^2 T\right) 2^{nb}(\mathrm{vol}\,(K)).$$

Since

$$\|\eta_T - 1\|_{L^1(\pi)} \le \|\eta_T - 1\|_{L^2(\pi)} = \sqrt{\langle \eta_T - 1, \eta_T - 1 \rangle_{L^2(\pi)}},$$

and $R_\infty \le 1$, we have

$$\|\eta_T - 1\|_{L^1(\pi)} \le \exp\left(-\frac{\Phi^2 T}{2}\right) 2^{\frac{n(b+1)}{2}}.$$

Substituting $\Phi \ge \frac{r_p}{O(n^{2+\frac{1}{p}}) \cdot R_\infty}$ from eq. (95) we have

$$\|\eta_T - 1\|_{L^1(\pi)} \le \exp\left(-\left(\frac{r_p}{O(n^{2+\frac{1}{p}}) \cdot R_\infty}\right)^2 T\right) 2^{\frac{n(b+1)}{2}}.$$

In order to have $\|\eta_T - 1\|_{L^1(\pi)} \le 2\epsilon$, it suffices to have

$$T \ge \tilde{T}(\epsilon) := \left(\log\left((2\epsilon)^{-1}\right) + \frac{n(b+1)\ln 2}{2}\right)\left(\frac{O(n^{2+\frac{1}{p}}) \cdot R_\infty}{r_p}\right)^2.$$

Let $f_{p,a}$ be the function from $\mathcal{M}_p$ to $\mathcal{M}_{p,a}$ that maps a cube $Q \in \mathcal{M}_p$ to Q if Q is not contained in Q_∞, and otherwise maps Q to Q_∞. Recall that X_t is a Markov chain evolving according to $\mathcal{M}_p$, where $X_0 = Q \not\subseteq Q_\infty$. So $f_{p,a}(X_t)$ evolves according to $\mathcal{M}_{p,a}$ until the random time $\tau(Q, Q_\infty)$ that it hits Q_∞. Let $\tilde{a}(\epsilon)$ be the minimum a such that Q_∞ satisfies the following property:

$$\mathbb{P}\left[\tau(Q, Q_\infty) \ge \tilde{T}\left(\frac{\epsilon}{2}\right)\right] \ge 1 - \frac{\epsilon}{2}. \tag{124}$$

Let π^a denote the stationary measure of the chain $\mathcal{M}_{p,a}$ for any natural number a. Note that $\pi^a(Q') = \pi(Q')$ for all Q' that are not contained in Q_∞. Therefore, for any $T \le \tilde{T}\left(\frac{\epsilon}{2}\right)$, we have the following upper bound:

$$d_{TV}(X_T, \pi) \le d_{TV}(f_{p,\tilde{a}(\epsilon)}(X_T), \pi^{\tilde{a}(\epsilon)}) + \frac{\epsilon}{2}. \tag{125}$$

We will next obtain an upper bound on the right hand side by finding an upper bound on the mixing time of $\mathcal{M}_{p,a}$. The conductance profile of $\mathcal{M}_{p,a}$ dominates that of $\mathcal{M}_p$ because for any subset S of the states of $\mathcal{M}_{p,a}$, $f_{p,a}^{-1}(S)$ is a subset of the states of $\mathcal{M}_p$ of the same measure, and the transitions of $\mathcal{M}_{p,a}$ correspond to that of a chain obtained from fusing the states in Q_∞ as stated in Section 5.5.1.

In preparation for applying the average conductance result of Lovász and Kannan [LK99] to the chain $\mathcal{M}_{p,\tilde{a}(\epsilon)}$, we denote

$$\mathcal{H} := \frac{1}{\Phi_{1/2}} + \int_{(2^{\tilde{a}(\epsilon)n}\mathrm{vol}\,(K))^{-1}}^{\frac{1}{2}} \frac{d\alpha}{\alpha \Phi_\alpha^2}. \tag{126}$$

Here, $1/(2^{\tilde{a}(\epsilon)n}\mathrm{vol}\,(K))$ is a lower bound on the stationary probability $\pi^{\tilde{a}(\epsilon)}(q)$ of any state q of the finite state Markov chain $\mathcal{M}_{p,\tilde{a}(\epsilon)}$, and $\Phi_\alpha = \Phi_{\mathcal{M}_p,\alpha}$ is the conductance profile of $\mathcal{M}_p$. As argued above, this conductance profile is dominated by that of $\mathcal{M}_{p,\tilde{a}(\epsilon)}$, i.e., $\Phi_{\mathcal{M}_{p,\tilde{a}(\epsilon)},\alpha} \geq \Phi_{\mathcal{M}_{p},\alpha}$ for every $\pi_0 \leq \alpha \leq 1/2$, where π_0 is the minimum stationary probability of any state of $\mathcal{M}_{p,\tilde{a}(\epsilon)}$. Using the discussion in [LK99, p. 283, especially Theorem 2.2] (see also the corrected version [MS01, Theorem 2.2]), we thus get that the mixing time to reach a total variation distance of $\epsilon/2$ from any state of the finite Markov chain $\mathcal{M}_{p,\tilde{a}(\epsilon)}$ is at most (for some absolute constant C)

$$C\mathcal{H} \log\left(\frac{2}{\epsilon}\right). \tag{127}$$

Therefore, by eq. (125), $C\mathcal{H}\log\left(\frac{2}{\epsilon}\right)$ is an upper bound on the the time needed by $\mathcal{M}_p$ starting at the state Q to achieve a total variation distance of ϵ to stationarity. It remains to estimate (127) from above. As a consequence of eqs. (96) and (97) in Theorem 5.3,

$$\Phi_{\exp(-x)} \geq \begin{cases} \frac{\rho_p}{O(n^{2+\frac{1}{p}})} \cdot (x + \log(0.9)), & \text{if } x < C_0 n \\ \frac{\rho_p}{O(n^{1+\frac{1}{p}})}, & \text{if } x \geq C_0 n. \end{cases}$$

Therefore,

$$\begin{aligned} \mathcal{H} &= \frac{1}{\Phi_{1/2}} + \int_{(2^{\tilde{a}(\epsilon)n}\mathrm{vol}(K))^{-1}}^{\frac{1}{2}} \frac{d\alpha}{\alpha\Phi_\alpha^2} && (128) \\ &= \frac{1}{\Phi_{1/2}} + \int_{\ln 2}^{\ln\left(2^{\tilde{a}(\epsilon)n}\mathrm{vol}(K)\right)} \Phi_{\exp(-x)}^{-2}\,dx && (129) \\ &\leq \left(\frac{\rho_p}{O(n^{2+\frac{1}{p}})}\right)^{-2} + \tilde{a}(\epsilon)n\left(\frac{\rho_p}{O(n^{1+\frac{1}{p}})}\right)^{-2}, && (130) \end{aligned}$$

where in the last line we use the assumption $R_\infty \leq 1$ to upper bound $\mathrm{vol}\,(K)$.

Lastly, we will find an upper bound for $\tilde{a}(\epsilon)$. Observe that, by reversibility (see the discussion following eq. (11)), we have that for all $t \geq 0$, $\|\eta_{t+1}\|_\infty = \|\mathcal{M}_{p,\tilde{a}(\epsilon)}\eta_t\|_\infty \leq \|\eta_t\|_\infty$, and consequently, for all positive t,

$$\|\eta_t\|_\infty \leq \|\eta_0\|_\infty \leq 2^{nb}\mathrm{vol}\,(K)\,. \tag{131}$$

We now proceed to upper bound the total volume of all cubes contained in Q_∞, for any value of the parameter a used in its definition. Since any cube in Q_∞ has side length at most 2^{-a}, it follows from item 4 of Theorem 3.1 that any point in such a cube must be at an Euclidean distance of at most $5n^2 2^{-a} < Cn^C 2^{-a}$ from ∂K. Define the inner parallel body K_r to be the set of all points in K at a Euclidean distance at least r from ∂K. This set is convex as can be seen from Lemma 2.2. From the above discussion, we also $\mathrm{vol}\,(Q_\infty) \leq \mathrm{vol}\,(K^\circ \setminus K_{Cn^C 2^{-a}})$.

Now, by the coarea formula for the closed set $\mathbb{R}^n \setminus K^\circ$ (see Lemma 3.2.34 on p. 271 of [Fed96]), letting $u(x)$ be the ℓ_2 distance of x to $\mathbb{R}^n \setminus K^\circ$ (and therefore also to ∂K), we have

$$\mathrm{vol}\,(K^\circ \setminus K_r) = \int_0^r H_{n-1}(u^{-1}(t))\;dt = \int_0^r \mathrm{vol}_{n-1}\,(\partial K_t)\;dt \tag{132}$$

where H_{n-1} is the $(n-1)$-dimensional Hausdorff measure. In particular, this, together with

$$\mathrm{vol}_{n-1}\left(\partial K_t\right) \leq \mathrm{vol}_{n-1}\left(\partial K\right),$$

(which follows from Proposition 2.5) implies

$$\mathrm{vol}\left(K^\circ \setminus K_r\right) = \int_0^r \mathrm{vol}_{n-1}\left(\partial K_t\right) dt \leq r \cdot \mathrm{vol}_{n-1}\left(\partial K\right). \tag{133}$$

Thus (again using Proposition 2.5, along with the assumption $R_\infty \leq 1$, which implies $K \subseteq B_\infty$),

$$\mathrm{vol}\left(Q_\infty\right) < C2^{-a} n^C \mathrm{vol}_{n-1}\left(\partial K\right) < C2^{-a} n^C \mathrm{vol}_{n-1}\left(\partial B_\infty\right) < C2^{-a+n} n^{C+1}.$$

Thus, for all a such that

$$C2^{-a+n} n^{C+1} < \frac{\epsilon}{2^{nb+2}\left(\tilde{T}\left(\frac{\epsilon}{2}\right)\right)},$$

we have (using eq. (131), and recalling that $\eta_t \pi^a$ is the probability distribution with density η_t with respect to π^a)

$$\mathbb{P}\left[\tau(Q, Q_\infty) < \tilde{T}\left(\frac{\epsilon}{2}\right)\right] \leq \sum_{t=0}^{\tilde{T}\left(\frac{\epsilon}{2}\right)} (\eta_t \pi^a)(Q_\infty) \leq 2^{nb} \cdot \left(1 + \tilde{T}\left(\frac{\epsilon}{2}\right)\right) \cdot \mathrm{vol}\left(Q_\infty\right) < \frac{\epsilon}{2}.$$

Thus, from eq. (124), we see that $\tilde{a}(\epsilon)$ only needs to satisfy

$$2^{-\tilde{a}(\epsilon)} < \frac{c\epsilon}{2^{n(b+1)} n^{C+1}\left(\tilde{T}\left(\frac{\epsilon}{2}\right)\right)}.$$

Since

$$\tilde{T}(\epsilon/2) \leq \left(\log \epsilon^{-1} + nb\right)\left(O(n^C \rho_p^{-2})\right),$$

we see that $\tilde{a}(\epsilon)$ can be chosen so that

$$\tilde{a}(\epsilon) \leq \log\left(\frac{2^{n(b+1)} n^{C+1}\left(\log \epsilon^{-1} + nb\right)\left(O(n^C \rho_p^{-2})\right)}{c\epsilon}\right),$$

which simplifies to

$$\tilde{a}(\epsilon) < C\left(nb + \log\left(\frac{n}{\rho_p \epsilon}\right)\right). \tag{134}$$

Finally, putting together (130) and (134), we have for all $\epsilon < \frac{1}{2}$,

$$C\mathcal{H} \log \frac{2}{\epsilon} \leq C \log \epsilon^{-1}\left(\frac{n^{4+\frac{2}{p}}}{\rho_p^2}\right)\left(\log \frac{n}{d} + n^{-1} \log\left(\frac{n}{\rho_p \epsilon}\right)\right).$$

In light of the discussion following eq. (127) and the choice of $\tilde{a}(\epsilon)$, this completes the proof of the claimed mixing time for $\mathcal{M}_p$ as well (recall that we assumed by scaling the body that $R_\infty \leq 1$). $\square$

6 Coordinate hit-and-run

Given a convex body K in $\mathbb{R}^n$ and $x_1 \in K$, the steps $x_1, x_2, \dots,$ 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 ℓ 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 ℓ. In this section, we prove 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 [LV21].

6.1 Isoperimetric inequality for axis-disjoint sets

We need the following definition, due to Laddha and Vempala [LV21].

Definition 6.1 (Axis-disjoint sets [LV21]). 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 ith coordinate direction. In other words, it is not possible to "reach" S_2 from S_1 by moving along a coordinate direction.

The main technical result of this section is the following isoperimetric inequality for axis-disjoint sets.

Theorem 6.2. *Let K be a convex body in $\mathbb{R}^n$. Denote by $\Phi_{\mathcal{M}_\infty}$ the conductance of the Markov chain $\mathcal{M}_\infty$ defined on the Whitney decomposition $\mathcal{F}^{(\infty)}$ 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,*

$$\mathrm{vol}\,(S_3) \geq \Omega\left(\frac{\Phi_{\mathcal{M}_\infty}}{n^{3/2}}\right) \cdot \min\{\mathrm{vol}\,(S_1)\,, \mathrm{vol}\,(S_2)\}.$$

Combined with the results already proved for the multiscale chain $\mathcal{M}_\infty$, 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 [LV21, Theorem 3; Theorem 2 in the arXiv version]. The result of [LV21] 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,

$$\mathrm{vol}\,(S_3) \geq \epsilon \cdot \Omega\left(\frac{r}{n^{3/2} \cdot R}\right) \cdot \left(\min\{\mathrm{vol}\,(S_1)\,, \mathrm{vol}\,(S_2)\} - \epsilon \cdot \mathrm{vol}\,(K)\right), \tag{135}$$

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 $\mathrm{vol}\,(S_3)$ and $\min\{\mathrm{vol}\,(S_1)\,, \mathrm{vol}\,(S_2)\}$ only when the latter is at least $\epsilon \cdot \mathrm{vol}\,(K)$. Further, due to the ϵ pre-factor, the volume guarantee that it gives for $\mathrm{vol}\,(S_3)$ as a multiple of $\min\{\mathrm{vol}\,(S_1)\,, \mathrm{vol}\,(S_2)\}$ 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 [LV21] requiring a warm start. The proof strategy employed by Narayanan and Srivastava [NS22] for the polynomial time mixing of CHR from a warm start was different from that of [LV21], 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: see the proof of Theorem 6.3.

The proof of the inequality in eq. (135) by Laddha and Vempala [LV21] built upon an isoperimetry result for cubes (Lemma 6.1 below). 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 [LS93].

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 ∂K, which led to the troublesome $-\epsilon\mathrm{vol}(K)$ term in eq. (135) 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 $\mathcal{M}_\infty$ chain) our isoperimetric inequality (Theorem 5.1), which is especially oriented for handling sets with a significant amount of mass close to the boundary ∂K.

We now proceed to the proofs of Theorems 6.2 and 6.3. We begin by listing some results and observations of Laddha and Vempala [LV21] about axis-disjoint sets.

Lemma 6.1 (Laddha and Vempala [LV21, Lemma 1 in the arXiv version]). *Let S_1, S_2 be measurable axis-disjoint subsets of an axis-aligned cube Q. Set $S_3 := Q \setminus (S_1 \cup S_2)$. Then,*

$$\pi_Q(S_3) \geq \frac{1}{2n}\min\{\pi_Q(S_1), \pi_Q(S_2)\} \geq \frac{1}{2n}\cdot\pi_Q(S_1)\cdot\pi_Q(S_2).$$

In particular, if $\pi_Q(S_1) \leq (2/3)$, then

$$\pi_Q(S_3) \geq \frac{1}{8n}\pi_Q(S_1).$$

For completeness, we include the proof.

Proof of Lemma 1 of [LV21]. Without loss of generality, we assume that $Q = [0,1]^n$ and $\mathrm{vol}(S_1) \leq \mathrm{vol}(S_2)$, so that $\pi_Q(S) = \mathrm{vol}(S_1) \leq (1/2)$. Denote by $\mathrm{proj}_j(S_1) \subseteq \mathbb{R}^{n-1}$ the $(n-1)$-dimensional projection

$$\mathrm{proj}_j(S_1) = \{(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) : x \in S_1\}$$

of S_1 onto the jth hyperplane. Since S_1 and S_2 are axis-disjoint, we have, for every $1 \leq j \leq n$,

$$S_1 \subseteq \left\{x \in Q : (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in \mathrm{proj}_j(S_1)\right\} \subseteq (S_1 \cup S_3),$$

so that $\mathrm{vol}(S_3) \geq \mathrm{vol}_{n-1}\left(\mathrm{proj}_j(S_1)\right) - \mathrm{vol}(S_1)$. Averaging this over j,

$$\begin{aligned}
\pi_Q(S_3) = \mathrm{vol}(S_3) &\geq \frac{1}{n}\sum_{j=1}^{n}\left(\mathrm{vol}_{n-1}\left(\mathrm{proj}_j(S_1)\right) - \mathrm{vol}(S_1)\right)\\
&\geq \left(\prod_{j=1}^{n}\mathrm{vol}_{n-1}\left(\mathrm{proj}_j(S_1)\right)\right)^{1/n} - \mathrm{vol}(S_1) \qquad \text{(AM-GM inequality)}\\
&\geq \mathrm{vol}(S_1)^{1-(1/n)} - \mathrm{vol}(S_1) \qquad \text{(Loomis-Whitney inequality)}\\
&\geq \mathrm{vol}(S_1)\left(2^{1/n} - 1\right) \qquad \text{(because } \mathrm{vol}(S_1) \leq 1/2\text{)}\\
&\geq \frac{1}{2n}\mathrm{vol}(S_1) = \frac{1}{2n}\min\{\pi_Q(S_1), \pi_Q(S_2)\}.
\end{aligned}$$

The final comment in the statement follows by considering separately the cases $\pi_Q(S_2) \leq 1/4$ and $\pi_Q(S_2) > 1/4$. ∎

We will need the following corollary.

Lemma 6.2. *Let S be a countable union of axis-aligned cubes in $\mathbb{R}^n$ with disjoint interiors. Let $S_1, S_2 \subseteq S$ be axis-disjoint and define $S_3 := S \setminus (S_1 \cup S_2)$. Suppose that $\pi_Q(S_1) \le (2/3)$ for each cube $Q \in S$ and $\pi_S(S_1 \cup S_3) \ge \eta$. Then,*

$$\pi_S(S_3) \ge \frac{\eta}{16n}.$$

Proof. If $\pi_S(S_1) \le (\eta/2)$, we are done. If $(\eta/2) \le \pi_S(S_1)$, the previous lemma gives that

$$\pi_S(S_3) = \sum_{Q \in S} \pi_Q(S_3)\pi_S(Q) \ge \frac{1}{8n} \sum_{Q \in S} \pi_Q(S_1)\pi_S(Q) = \frac{1}{8n}\pi_S(S_1) \ge \frac{\eta}{16n}. \qquad \square$$

We will also need the following simple observation.

Lemma 6.3. *Let q, q' be axis-aligned cuboids with a common facet, and S_1, S_2, S_3 a partition of $(q \cup q')$ such that S_1 and S_2 are axis-disjoint. Then, $\pi_{q'}(S_1 \cup S_3) \ge \pi_q(S_1)$.*

Proof. Assume without loss of generality that the standard unit vector e_1 is normal to the common facet f. By the definition of axis-disjointness, S_2 does not intersect $T := (S_1 \cap q) + e_1\mathbb{R}$. In particular,

$$q' \cap T \subseteq q' \cap (S_1 \cup S_3).$$

To conclude,

$$\pi_{q'}(S_1 \cup S_3) \ge \frac{\mathrm{vol}\,(q' \cap T)}{\mathrm{vol}\,(q')} = \frac{\mathrm{vol}\,(q \cap T)}{\mathrm{vol}\,(q)} \ge \pi_q(S_1). \qquad \square$$

We are now ready to prove Theorem 6.2.

Proof of Theorem 6.2. Assume without loss of generality that $\mathrm{vol}\,(S_1) \le \mathrm{vol}\,(S_2)$. Let $\mathcal{F} = \mathcal{F}^{(\infty)}$ be a multiscale partition of K into axis-aligned cubes as described in Section 3, and split it as

$$\mathcal{F}_1 = \left\{Q \in \mathcal{F} : \pi_Q(S_1) < 2/3\right\} \text{ and}$$
$$\mathcal{F}_2 = \left\{Q \in \mathcal{F} : \pi_Q(S_1) \ge 2/3\right\}.$$

As before, given a collection of cubes $\mathcal{F}' \subseteq \mathcal{F}$, we interchangeably use it to denote $\bigcup_{Q \in \mathcal{F}'} Q$. Note that $\mathrm{vol}\,(S_1 \cap \mathcal{F}_1) + \mathrm{vol}\,(S_1 \cap \mathcal{F}_2) = \mathrm{vol}\,(S_1)$. Now, if $\mathrm{vol}\,(S_1 \cap \mathcal{F}_1) \ge (1/2)\mathrm{vol}\,(S_1)$, then

$$\begin{aligned}
\mathrm{vol}\,(S_3) &\ge \sum_{Q \in \mathcal{F}_1} \mathrm{vol}\,(S_3 \cap Q) \\
&\ge \sum_{Q \in \mathcal{F}_1} \frac{1}{8n}\mathrm{vol}\,(S_1 \cap Q) \qquad \text{(Lemma 6.1)} \\
&= \frac{1}{8n}\mathrm{vol}\,(S_1 \cap \mathcal{F}_1) \ge \frac{1}{16n}\mathrm{vol}\,(S_1)\,,
\end{aligned}$$

and the claimed lower bound on the volume of S_3 follows. Therefore, we can assume that

$$\mathrm{vol}\,(\mathcal{F}_2) \ge \mathrm{vol}\,(S_1 \cap \mathcal{F}_2) \ge (1/2)\mathrm{vol}\,(S_1)\,. \tag{136}$$

For any cube $q \in \mathcal{F}_2$ and facet f of q, denote by η_f the fraction of the facet that is incident on $\mathcal{F}_1$, and set

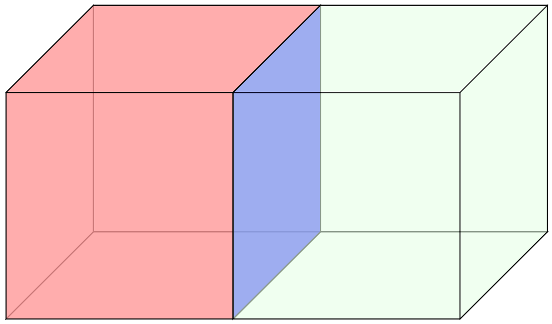
$$\eta_q := \frac{1}{2n} \sum_{\text{facet } f \text{ of } q} \eta_f. \tag{137}$$

We shall show that for some universal constant c,

$$\mathrm{vol}\left(S_3\right) \geq \frac{c}{n^{3/2}} \sum_{q \in \mathcal{F}_2} \eta_q \mathrm{vol}\left(q\right). \tag{138}$$

Let $q_2 \in \mathcal{F}_2$ and f a facet of q_2 such that $\eta_f \neq 0$.

1. Case 1. f is between two cubes of the same sidelength.

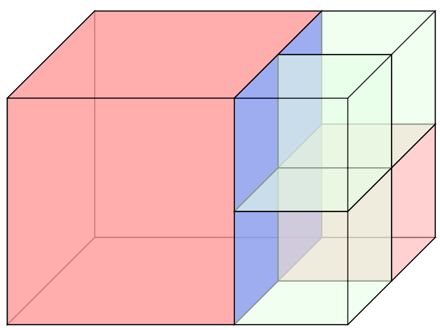


Case 1. The red cube is $q_2 \in \mathcal{F}_2$ and the green is $q \in \mathcal{F}_1$. The blue facet is f.

Let q be the cube other than q_2 that is bordering f. Observe that $\eta_f = 1$.
Using Lemma 6.3 on q_2 and q, we get that $\pi_q(S_1 \cup S_3) \geq \pi_{q_2}(S_1) \geq (2/3)$. By Lemma 6.2, $\pi_q(S_3) \geq 1/(24n)$ so

$$\mathrm{vol}\left(S_3 \cap q\right) \geq \frac{1}{24n} \eta_f \mathrm{vol}\left(q_2\right). \tag{139}$$

2. Case 2. The cubes other than q_2 incident on f are smaller than q_2.



Case 2. The red cubes are in $\mathcal{F}_2$ and the green in $\mathcal{F}_1$. The blue facet is f.

Let T be the set of all cubes of smaller size incident on f, and set $T' := T \cap \mathcal{F}_1$. Recall from item 5 of Theorem 3.1 that all cubes in T must then have sidelength exactly half the sidelength of q_2. We use two consequences of this fact. First, that, $\mathrm{vol}\left(T'\right) = \eta_f \mathrm{vol}\left(T\right) = (1/2)\eta_f \mathrm{vol}\left(q_2\right)$. Second, that the graph with vertex set equal to the set of cubes in T in which two vertices are adjacent if and only if the corresponding cubes are adjacent and have a facet in common is exactly the $(n-1)$-dimensional Boolean hypercube.

(a) Case 2(a). $\eta_f \geq (1/2)$. In this case, $\mathrm{vol}\,(T \setminus T') \leq (1/2)\mathrm{vol}\,(T)$. Using Lemma 6.3 on q_2 and T, $\pi_T(S_1 \cup S_3) \geq \pi_{q_2}(S_1) \geq (2/3)$, so $\mathrm{vol}\,(T \cap (S_1 \cup S_3)) \geq (2/3)\mathrm{vol}\,(T)$. Combining, we get

$$\mathrm{vol}\,(T' \cap (S_1 \cup S_3)) \geq \mathrm{vol}\,(T \cap (S_1 \cup S_3)) - \mathrm{vol}\,(T \setminus T') \geq \frac{1}{6}\mathrm{vol}\,(T) \geq \frac{1}{6}\mathrm{vol}\,(T')\,.$$

As a result, $\pi_{T'}(S_1 \cup S_3) \geq (1/6)$. By Lemma 6.2, $\pi_{T'}(S_3) \geq \frac{1}{96n}$ (since each cube in T' lies in $\mathcal{F}_1$), so that

$$\mathrm{vol}\,(S_3 \cap T) \geq \mathrm{vol}\,(S_3 \cap T') \geq \frac{1}{96n}\mathrm{vol}\,(T') \geq \frac{1}{200n}\eta_f \mathrm{vol}\,(q_2)\,. \tag{140}$$

(b) Case 2(b). $\eta_f < (1/2)$. By Lemma 6.3, for any cube q in T' adjacent to a cube in $(T \setminus T') \subseteq \mathcal{F}_2$, we have $\pi_q(S_1 \cup S_3) \geq (2/3)$. Since $q \in T' \subseteq \mathcal{F}_1$, Lemma 6.2 then implies $\pi_q(S_3) \geq 1/(24n)$, so that for such cubes

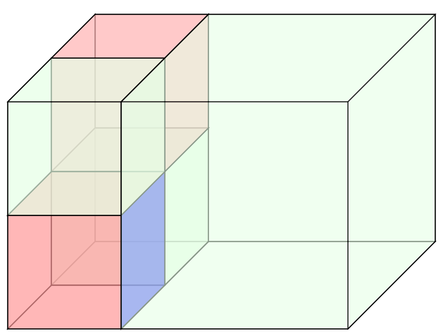
$$\mathrm{vol}\,(q \cap S_3) \geq \frac{1}{24n} \cdot \mathrm{vol}\,(q)\,. \tag{141}$$

We shall show that there are many such cubes.

To do this, consider the $(n-1)$-dimensional hypercube graph with vertex set equal to the set of cubes T and with two vertices being adjacent if and only if the corresponding cubes are adjacent in the sense of sharing a facet. Due to Harper's Theorem [Har66], the vertex expansion of T' in the hypercube graph is $\Omega(n^{-1/2})$ (see Corollary B.3 for more details). In particular, an $\Omega(n^{-1/2})$ fraction of the cubes in T', which therefore constitute at least an $\eta_f\Omega(n^{-1/2})$ fraction of the cubes in T, are adjacent to a cube in $(T \setminus T')$. The total volume of these cubes is thus $\eta_f\Omega(n^{-1/2}) \cdot \mathrm{vol}\,(q_2)\,/2$. Consequently,

$$\mathrm{vol}\,(S_3 \cap T) \geq \mathrm{vol}\,(S_3 \cap T') \overset{\text{eq. (141)}}{\geq} \frac{1}{48n} \cdot \Omega(n^{-1/2})\eta_f \mathrm{vol}\,(q_2) = \Omega(n^{-3/2})\eta_f \mathrm{vol}\,(q_2)\,. \tag{142}$$

3. Case 3. The cube q_1 other than q_2 incident on f is larger than q_2.



Case 3. The red cubes are in $\mathcal{F}_2$ and the green in $\mathcal{F}_1$.
The blue facet is f. f_1 is the larger facet it is part of.

In this case, $\eta_f = 1$ again. Further, from item 5 of Theorem 3.1, the sidelength of q_1 is exactly twice that of q_2. Let f_1 be the facet of q_1 that contains f, and let η_{f_1} be the fraction of f_1 that is incident on $\mathcal{F}_2$. Let T be the set of all cubes of smaller size incident on f_1, and $T' = T \cap \mathcal{F}_2$ (note that T' is now defined in terms of $\mathcal{F}_2$, unlike in Case 2). Clearly, $\mathrm{vol}\,(T') = \eta_{f_1}\mathrm{vol}\,(T) = (\eta_{f_1}/2)\mathrm{vol}\,(q_1)$. For each cube q in T', let f_q be the facet of q contained in f_1, so $\eta_{f_q} = 1$ for all such q.

Noting that $\pi_T(S_1) \geq \pi_{T'}(S_1)\pi_T(T') \geq (2/3)\eta_{f_1}$ and using Lemma 6.3 on T and q_1, we get that $\pi_{q_1}(S_1 \cup S_3) \geq (2/3)\eta_{f_1}$, so that by Lemma 6.2,

$$\pi_{q_1}(S_3 \cap q_1) \geq \frac{\eta_{f_1}}{24n}.$$

So,

$$\mathrm{vol}\,(S_3 \cap q_1) \geq \frac{\eta_{f_1}}{24n}\mathrm{vol}\,(q_1) = \frac{1}{12n}\mathrm{vol}\,(T') = \frac{1}{12n}\sum_{q \in T'} \eta_{f_q}\mathrm{vol}\,(q)\,. \tag{143}$$

This allows us to associate a volume of $S_3 \cap q_1$ of measure $\frac{1}{12n}\eta_{f_q}\mathrm{vol}\,(q)$ to each cube $q \in T'$, such that the volumes associated with distinct cubes in T' are disjoint.

From eqs. (139), (140), (142) and (143), we see that each facet f of a cube q in $\mathcal{F}_2$ which abuts a cube in $\mathcal{F}_1$ can be associated with a volume of S_3 which lies in cubes in $\mathcal{F}_1$ that abut f and which is of measure at least $c \cdot n^{-3/2} \cdot \eta_f \mathrm{vol}\,(q)$, for some universal constant c.

Observe also that for distinct facets f, f' of cubes in $\mathcal{F}_2$ whose normal vectors pointing out of their respective cubes point in the same direction, these S_3 volumes are disjoint. Indeed, if f falls in Cases 1 or 2 above, then the cubes in $\mathcal{F}_1$ which contain the S_3 volume associated with f and f' are disjoint. If f falls in Case 3, f' can share the cube in $\mathcal{F}_1$ which contains the S_3-volume associated with f only if f' abuts the same cube q_1 in $\mathcal{F}_1$ as f does. But in this case, the remark after eq. (143) shows that the S_3 volumes associated with f, f' are still distinct.

Since there are only $2n$ distinct directions for the outward normal of a facet (two each in each of the n dimensions) we therefore get

$$2n \cdot \mathrm{vol}\,(S_3) \geq \frac{c}{n^{3/2}} \sum_{q \in \mathcal{F}_2} \sum_{\text{facet } f \text{ of } q} \eta_f \mathrm{vol}\,(q) \tag{144}$$

$$= 2n \cdot \frac{c}{2n^{3/2}} \sum_{q \in \mathcal{F}_2} \eta_q \mathrm{vol}\,(q)\,, \text{using the definition of } \eta_q \text{ in eq. (137)}, \tag{145}$$

proving eq. (138). Now, note that $\pi(\mathcal{F}_2) \leq (3/2)\pi(S_1) \leq (3/4)$. In the multiscale walk discussed in earlier sections, the ergodic flow out of $\mathcal{F}_2 \subseteq \mathcal{F}$ was (noting that $\mathcal{F}_1 = \mathcal{F} \setminus \mathcal{F}_2$)

$$\begin{aligned}\Psi_{\mathcal{M}_\infty}(\mathcal{F}_2, \mathcal{F} \setminus \mathcal{F}_2) &= \sum_{q \in \mathcal{F}_2} \pi(q) \cdot \frac{1}{2} \sum_{\substack{q' \in \mathcal{F}_1 \\ q' \text{ abuts } q}} \frac{\mathrm{vol}_{n-1}\,(\partial q' \cap \partial q)}{\mathrm{vol}_{n-1}\,(\partial q)} \min\left\{1, \frac{\mathrm{sidelength}\,(q')}{\mathrm{sidelength}\,(q)}\right\} \\ &\leq \sum_{q \in \mathcal{F}_2} \eta_q \pi(q).\end{aligned}$$

In particular, because $\pi(\mathcal{F}_2) \leq (3/2)\pi(S_1) \leq (3/4)$, we have that

$$\sum_{q \in \mathcal{F}_2} \eta_q \pi(q) \geq \Psi_{\mathcal{M}_\infty}(\mathcal{F}_2, \mathcal{F} \setminus \mathcal{F}_2) \geq \frac{1}{4}\Phi_{\mathcal{M}_\infty} \cdot \pi(\mathcal{F}_2) \geq \frac{1}{8}\Phi_{\mathcal{M}_\infty} \cdot \pi(S_1).$$

Here, the second inequality uses $\min\{\pi(\mathcal{F}_2), \pi(\mathcal{F} \setminus \mathcal{F}_2)\} \geq \pi(\mathcal{F}_2)\pi(\mathcal{F} \setminus \mathcal{F}_2) \geq \frac{1}{4}\pi(\mathcal{F}_2)$ (which follows since $\pi(\mathcal{F}_2) \leq 3/4$), while the third inequality uses eq. (136) (where it was argued that it is enough to consider the case where $\pi(\mathcal{F}_2 \cap S_1) \geq \pi(S_1)/2$). Plugging this back into eq. (145), we get

$$\mathrm{vol}\,(S_3) \geq \Omega\left(\frac{\Phi_{\mathcal{M}_\infty}}{n^{3/2}}\right)\mathrm{vol}\,(S_1)\,. \qquad \square$$

6.2 Rapid mixing of CHR from a cold start: Proof of Theorem 1.1

Armed with the new isoperimetric inequality for axis-disjoint sets given by Theorem 6.2, we can now replicate the argument of [LV21] to get a conductance lower-bound bound even for small sets, in place of the s-conductance 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 $\Phi_{\mathcal{M}_\infty}$ denote the conductance of the Markov chain $\mathcal{M}_\infty$ on the Whitney decomposition $\mathcal{F}^{(\infty)}$ of K. Then, the conductance of the coordinate hit-and-run chain on K is $\Omega(\Phi_{\mathcal{M}_\infty} n^{-5/2})$.*

Proof. As stated above, the strategy of the proof is essentially identical to that of [LV21], with the new ingredient being the isoperimetry for axis-disjoint sets given by Theorem 6.2. Let $K = S_1 \cup S_2$ be a partition of K into two parts with $\pi(S_1) \le \pi(S_2)$. For $i = 1, 2$, let

$$S_i' = \left\{ x \in S_i \;:\; P_{\mathrm{CHR}}(x, S_{3-i}) < \frac{1}{4n} \right\}.$$

We claim that S_1' and S_2' are axis-disjoint. Suppose instead that they are not, and there is an axis parallel line ℓ intersecting both of them, with $x_i \in \ell \cap S_i'$ for $i = 1, 2$, say. Then,

$$\frac{1}{4n} > P_{\mathrm{CHR}}(x_i, S_{3-i}) \ge \frac{1}{2n} \cdot \frac{\mathrm{vol}_1\left(\ell \cap S_{3-i}\right)}{\mathrm{vol}_1\left(\ell \cap K\right)},$$

so $\mathrm{vol}_1\left(\ell \cap K\right) > 2\mathrm{vol}_1\left(\ell \cap S_i\right)$ for $i = 1, 2$. However, this is clearly impossible as $\mathrm{vol}_1\left(\ell \cap S_1\right) + \mathrm{vol}_1\left(\ell \cap S_2\right) = \mathrm{vol}_1\left(\ell \cap K\right)$.

Now, if $\mathrm{vol}\left(S_1'\right) \le (1/2)\mathrm{vol}\left(S_1\right)$ (or similarly $\mathrm{vol}\left(S_2'\right) \le (1/2)\mathrm{vol}\left(S_2\right)$), then (here $\Psi_{\mathrm{CHR}}(\cdot,\cdot)$ denotes the ergodic flow for the coordinate hit-and-run chain),

$$\Psi_{\mathrm{CHR}}(S_1, S_2) \ge \Psi_{\mathrm{CHR}}((S_1 \setminus S_1'), S_2) \ge \frac{1}{4n}\pi(S_1 \setminus S_1') \ge \frac{1}{8n}\pi(S_1)$$

and we are done. So, assume that $\mathrm{vol}\left(S_i'\right) \ge (1/2)\mathrm{vol}\left(S_i\right)$ for $i = 1, 2$. In this case

$$\begin{aligned}
\Psi_{\mathrm{CHR}}(S_1, S_2) &\ge \frac{1}{2}\left(\Psi_{\mathrm{CHR}}(S_1 \setminus S_1', S_2) + \Psi_{\mathrm{CHR}}(S_1, S_2 \setminus S_2')\right) \text{ by reversibility,}\\
&\ge \frac{1}{8n}\left(\pi(S_1 \setminus S_1') + \pi(S_2 \setminus S_2')\right)\\
&= \frac{1}{8n}\pi(K \setminus (S_1' \cup S_2')), \text{ since } K = S_1 \sqcup S_2,\\
&\ge \frac{c\Phi_{\mathcal{M}_\infty}}{n^{5/2}}\pi(S_1) \text{ (from Theorem 6.2, since } \mathrm{vol}\left(S_i'\right) \ge (1/2)\mathrm{vol}\left(S_i\right))
\end{aligned}$$

for some universal constant c, completing the proof. $\square$

Corollary 6.4. *Let K be a convex body such that $r_\infty \cdot B_\infty \subseteq K \subseteq R_\infty \cdot B_\infty$. Let $\rho_\infty := r_\infty / R_\infty$. Let π denote the uniform measure on K. Let 1_K denote the indicator of K. Let $0 < \epsilon < 1/2$. The number of steps T needed for CHR to achieve a density η_T with respect to π such that $\|\eta_T - 1_K\|_{L^2(\pi)} < \epsilon$, from a starting density η_0 (with respect to π) that satisfies $\|\eta_0 - 1_K\|_{L^2(\pi)} < M$ obeys*

$$T \le O\left(\frac{n^9}{\rho_\infty^2} \log \frac{M}{\epsilon}\right).$$

Proof. Lemma 2.1 applied to CHR gives the following. Let the CHR walk be started from a density $\eta_0 \in L^2(\pi)$ with respect to the uniform measure on K. Then, after T steps, the density η_T, of the measure at time T with respect to π, satisfies

$$\langle \eta_T - 1_K, \eta_T - 1_K \rangle_{L^2(\pi)} \leq \left(1 - \frac{\Phi_{\mathrm{CHR}}^2}{2}\right)^{2T} \langle \eta_0 - 1_K, \eta_0 - 1_K \rangle_{L^2(\pi)}.$$

The corollary now follows by applying Theorem 6.3 and the fact that

$$\Phi_{\mathcal{M}_\infty} \geq \frac{\rho_\infty}{O(n^2)},$$

as shown in eq. (95) in Theorem 5.3. □

Proof of Theorem 1.1. The theorem follows immediately from Corollary 6.4 after a few substitutions. Let η_0 be the density of the initial M-warm start ν_0, with respect to π. This means $\|\eta_0\|_\infty \leq M$, which implies that $\|\eta_0 - 1\|_{L^2(\pi)} \leq M + 1$. Next, note that $\rho_\infty = R/r$ by definition, where R and r are as in the statement of Theorem 1.1. The theorem now follows from Corollary 6.4 by noting that that the density η_T (with respect to π) of the measure ν_T obtained after T steps of the CHR walk satisfies

$$\|\eta_T - 1_K\|_{L^1(\pi)} \leq \|\eta_T - 1_K\|_{L^2(\pi)},$$

so that Corollary 6.4 implies $d_{TV}(\nu_T, \pi) = \frac{1}{2}\|\eta_T - 1_K\|_{L^1(\pi)} \leq \epsilon/2$ for the same T. □

6.3 Mixing of CHR from a point

As discussed in the introduction, a cold start is often trivial to achieve. For example, in the context of Theorem 1.1 for the coordinate hit-and-run (CHR) walk, where the body K satisfies $r \cdot B_\infty \subseteq K \subseteq R \cdot B_\infty$, an $\exp(\mathrm{poly}\,(n))$-warm start can be generated simply by sampling the initial point uniformly at random from $r \cdot B_\infty$, provided that the mild condition that $R/r \leq \exp(\mathrm{poly}\,(n))$ is satisfied. However, for aesthetic reasons, one may want to prove that the chain mixes rapidly even when started from a given point (i.e., when the initial distribution is concentrated on a point). To avoid pathological issues that may arise at a "corner" of the body, it is usual to assume that this initial point is somewhat far from the boundary of the body: say at a distance of at least $R \exp(-\mathrm{poly}\,(n))$.

In this section, we prove the following corollary, which shows that the CHR walk mixes in a polynomial number of steps even when started from a distribution concentrated on a single point of K, provided that that point is not too close to the boundary of K.

Corollary 6.5 (Mixing time of CHR from a point)**.** *There is a universal constant C such that the following is true. Let K be a convex body such that $r \cdot B_\infty \subseteq K \subseteq R \cdot B_\infty$. Consider the coordinate hit-and-run chain on K started from a point X_0 satisfying $\mathrm{dist}_{\ell_\infty}(X_0, \partial K) \geq \delta$, and let X_T denote the random state of the chain after T steps. Let $\epsilon \in (0, 1/2)$ be given, and set $\tau = \lceil 2n \log(6n/\epsilon) \rceil$. Then, provided that*

$$T \geq C \cdot \frac{n^{10} R^2}{r^2} \cdot \left(\log \frac{R}{2\delta} + \tau \cdot \log \frac{6\tau}{\epsilon}\right), \tag{146}$$

the total variation distance $d_{TV}(X_T, \pi_K)$ is less than ϵ.

Lovász and Vempala [LV06, Corollary 1.2] pointed out that their mixing time result for the usual hit-and-run walk from a cold start [LV06, Theorem 1.1] immediately implies a corresponding mixing time result for hit-and-run starting from a point at distance $\delta > 0$ from the boundary as well, by considering the "warmth" of the distribution obtained by taking a single step of the hit-and-run walk from such a point. The intuition for the above result for the CHR walk is similar, with the only apparent difficulty being the fact that after any constant number of steps from a single point, the distribution generated by the CHR walk is a finite mixture of distributions on lower dimensional subsets of $\mathbb{R}^n$, and hence cannot be M-warm with respect to π_K for any finite M (since all such lower dimensional sets have zero probability mass under π_K). The obvious fix is to consider the distribution after about $2n \log(n/\epsilon)$ steps: by this time, with probability at least $1 - \epsilon$, the chain would have taken a step in each of the coordinate directions. All we need to check is that in this time, the chain does not come too close to the boundary either. The proof of Corollary 6.5 formalizes this intuition.

In the proof of Corollary 6.5, we will need the following simple observation.

Observation 6.6. *Let K be a convex body, and let x be a point in K such that $\operatorname{dist}_{\ell_\infty}(x, \partial K) \geq \delta$. Consider a chord PQ of K that passes through x (so that $P, Q \in \partial K$). For $\alpha \in [0, 1]$, define $P_\alpha := \alpha x + (1 - \alpha)P$ and $Q_\alpha := \alpha x + (1 - \alpha)Q$. Then the segment $P_\alpha Q_\alpha$ covers a $(1 - \alpha)$-fraction of the length of PQ, and for any y on the segment $P_\alpha Q_\alpha$,*

$$\operatorname{dist}_{\ell_\infty}(y, \partial K) \geq \alpha\delta. \tag{147}$$

Proof. The claim about the ratios of the length of $P_\alpha Q_\alpha$ and PQ follows immediately from the definition of P_α and Q_α. For the second claim, assume without loss of generality that the point y lies between x and P_α (the argument when it lies between x and Q_α is identical). Then, there exists β satisfying $1 \geq \beta \geq \alpha$ such that $y = \beta x + (1 - \beta)P$. We now use the concavity of $\operatorname{dist}_{\ell_\infty}(x, \partial K)$ (Lemma 2.2) to get

$$\operatorname{dist}_{\ell_\infty}(y, \partial K) = \operatorname{dist}_{\ell_\infty}(\beta x + (1 - \beta)P, \partial K) \geq \beta \operatorname{dist}_{\ell_\infty}(x, \partial K) + (1 - \beta) \operatorname{dist}_{\ell_\infty}(P, \partial K) \geq \alpha\delta. \qquad \square$$

We are now ready to prove Corollary 6.5.

Proof of Corollary 6.5. For convenience of notation, we assume, after scaling the body if necessary, that $R = 1$ (to translate the calculations below back to the general setting, we will need to replace δ by δ/R). Let $\tau = \lceil 2n \log(6n/\epsilon) \rceil$ be as in the statement of the corollary. We will show that after τ steps, the probability distribution ν of the state X_τ of the chain can be written as a convex combination of two probability measures ν_{cold} and ν_{rest}:

$$\nu = p\nu_{\mathrm{cold}} + (1 - p)\nu_{\mathrm{rest}}, \tag{148}$$

where $1 \geq p \geq 1 - \epsilon/3$, ν_{rest} is an arbitrary probability distribution on K, and ν_{cold} is a probability distribution on K which is M-warm with respect to π_K, for an M satisfying

$$\log M \leq 4n \cdot \left(\log \frac{1}{2\delta} + \tau \cdot \log \frac{6\tau}{\epsilon}\right), . \tag{149}$$

Given this, the claimed mixing time result will follow from a direct application of the mixing time bound for CHR from a cold start (Theorem 1.1). We now proceed to prove the decomposition claimed in eq. (148). The intuition is that such a decomposition with a reasonable upper bound on M should follow whenever during the first τ steps, the chain, with high probability, (i) takes a step at least once in each of the n coordinate directions, and (ii) does not come too close to the boundary of K. For the above choice of τ, both of these conditions are seen to be true with high probability. We now proceed to formalize this intuition.

To do this, we consider the following alternative description of the run of the chain till time τ. Let D_i, for $1 \le i \le \tau$, be i.i.d. random variables taking values in $\{0, 1, 2, \ldots, n\}$ such that for each i, $\mathbb{P}[D_i = 0] = 1/2$, and $\mathbb{P}[D_i = j] = 1/(2n)$ for $1 \le j \le n$ (the random variables D_i capture the direction the CHR walk chooses when generating the ith point, with $D_i = 0$ corresponding to the lazy choice of not making a step). Fix $\kappa := \epsilon/(6\tau)$, and let G_i, for $1 \le i \le \tau$, be i.i.d. Bernoulli random variables with a parameter of $(1 - \kappa)$.

Now, the distribution of the random variables $X_1, X_2, \ldots, X_\tau$ representing the first τ states of the CHR walk started at X_0 can be described as follows. Given X_{i-1}, where $1 \le i \le \tau$, X_i is generated as follows: if $D_i = 0$, set $X_i = X_{i-1}$. Otherwise, let $P^{(X_{i-1})}Q^{(X_{i-1})}$ be the chord of K (with $P^{(X_{i-1})}, Q^{(X_{i-1})} \in \partial K$) through X_{i-1}, parallel to the standard basis vector e_{D_i} of $\mathbb{R}^n$. Define $P_\kappa^{(X_{i-1})}, Q_\kappa^{(X_{i-1})}$ as in Observation 6.6. Let A_i be a point chosen uniformly at random from the segment $P_\kappa^{(X_{i-1})}Q_\kappa^{(X_{i-1})}$, and let B_i be a point chosen uniformly at random from the remaining length $P^{(X_{i-1})}Q^{(X_{i-1})} \setminus P_\kappa^{(X_{i-1})}Q_\kappa^{(X_{i-1})}$ of $P^{(X_{i-1})}Q^{(X_{i-1})}$. Set

$$X_i = \begin{cases} A_i & \text{if } G_i = 1, \text{ and} \\ B_i & \text{if } G_i = 0. \end{cases} \tag{150}$$

The independence of the D_i and the G_i, taken together with the fact that $P_\kappa^{(X_{i-1})}Q_\kappa^{(X_{i-1})}$ constitutes a $(1 - \kappa)$ fraction of the chord $P^{(X_{i-1})}Q^{(X_{i-1})}$, then implies that the distribution of $(X_0, X_1, X_2, \ldots, X_\tau)$ generated in the above way is identical to the distribution of the first τ steps of the CHR walk started from X_0. Note also that for any realizations $g \in \{0, 1\}^\tau$ of the G_i and $d \in \{0, 1, 2, \ldots, n\}^\tau$ of the D_i, the above process induces a probability distribution $\nu_{d,g}$ on K, such that the probability distribution ν of X_τ can be written as the convex combination

$$\nu = \sum_{\substack{d \in \{0,1,2,\ldots,n\}^\tau \\ g \in \{0,1\}^\tau}} \mathbb{P}[\boldsymbol{D} = d \text{ and } \boldsymbol{G} = g]\, \nu_{d,g}. \tag{151}$$

If $G_i = 1$, then Observation 6.6 implies that $\mathrm{dist}_{\ell_\infty}(X_i, \partial K) \ge \kappa \cdot \mathrm{dist}_{\ell_\infty}(X_{i-1}, \partial K)$. Thus, if $G_i = 1$ for all i, $1 \le i \le \tau$, then

$$\mathrm{dist}_{\ell_\infty}(X_i, \partial K) \ge \kappa^i \delta \ge \kappa^\tau \delta \text{ for all } 1 \le i \le \tau. \tag{152}$$

Call a vector $d \in \{0, 1, 2, \ldots, n\}^\tau$ *complete* if it includes each of the numbers $\{1, 2, \ldots, n\}$ at least once. Let p be the probability that $G_i = 1$ for all $1 \le i \le \tau$, and that $(D_1, D_2, \ldots, D_\tau)$ is complete. Then, breaking apart the sum in eq. (151) into those d and g that satisfy this requirement and those that do not, we get

$$\nu = p\nu_{\text{good}} + (1 - p)\nu_{\text{rest}}, \tag{153}$$

where ν_{good} is a convex combination of those $\nu_{d,g}$ for which d is complete and $g = \mathbf{1}$, while ν_{rest} is a convex combination of the remaining $\nu_{d,g}$. From a simple union bound argument, we also get

$$1 - p \le \tau\kappa + n\left(1 - \frac{1}{2n}\right)^\tau \le \epsilon/3. \tag{154}$$

Thus, to establish the decomposition in eq. (148), we only need to show that when d is complete and $g = \mathbf{1}$, $\nu_{d,g}$ is M-warm with respect to π_K for an M satisfying eq. (149) (since a convex combination of M-warm distributions is also M-warm).

We now prove that for any realization d of the D_i and g of the G_i, such that d is complete and $g = \mathbf{1}$, $\nu_{d,g}$ satisfies the required warmth condition. For $0 \le i \le \tau$, let $\nu_{d,g,i}$ be the distribution of X_i under this realization. Let $S(i)$ denote the set of coordinate directions corresponding to the distinct non-zero values among $d_1, d_2, \ldots, d_i$. For convenience of notation, we relabel the coordinate directions in the order of their first appearance in d, so that $S(i) = \{e_j : 1 \le j \le |S(i)|\}$. Since d is complete, $|S(\tau)| = n$. Define

$$M_i := (1-\kappa)^{-i}(2\kappa^\tau \delta)^{-|S(i)|}. \tag{155}$$

Note that the probability distribution $\nu_{d,g,i}$ is supported on the $|S(i)|$-dimensional set $(X_0 + \mathbb{R}^{|S(i)|} \times \{0\}^{n-|S(i)|}) \cap K$. We will show by induction that $\nu_{d,g,i}$ is M_i-warm with respect to the standard $|S(i)|$-dimensional Lebesgue measure on $(X_0 + \mathbb{R}^{|S(i)|} \times \{0\}^{n-|S(i)|})$. Let ρ_i denote the density of $v_{d,g,i}$ with respect to the $|S(i)|$ dimensional Lebesgue measure on $(X_0 + \mathbb{R}^{|S(i)|} \times \{0\}^{n-|S(i)|})$ (implicit in the inductive proof of warmth below will be an inductive proof of the existence of these densities). Since $g = \mathbf{1}$, we have from eq. (152) that for all $0 \le j \le \tau$,

$$\rho_j(y) > 0 \text{ only if } \operatorname{dist}_{\ell_\infty}(y, \partial K) \ge \kappa^j \delta \ge \kappa^\tau \delta. \tag{156}$$

In particular, if $\rho_j(y) > 0$, then the length of any chord of K through y in any coordinate direction is at least $2\kappa^j\delta \ge 2\kappa^\tau\delta$.

In the base case $i = 0$, we have $S(i) = 0$ and $M_i = 1$, and $\nu_{d,g,0}$ trivially has density $\rho_0 \equiv 1$ (supported on the single point X_0). In the inductive case $i \ge 1$, the density does not change when from step $i-1$ to i if $d_i = 0$, so that we get using the induction hypothesis and the form of the M_i that at each point z, $\rho_i(z) = \rho_{i-1}(z) \le M_{i-1} \le M_i$. We are thus left with the case $d_i \ge 1$. We break this case into further cases.

Case 1: $S(i) = S(i-1)$**.** Note that $X_i = z$ is possible only if $X_{i-1} = y$ lies on the chord $\tilde{P}\tilde{Q}$ of K (where $\tilde{P}, \tilde{Q} \in \partial K$) through z in the direction e_{d_i}. For such a y, we also have $\tilde{P}\tilde{Q} = P^{(y)}Q^{(y)}$ (where, as defined in the description before eq. (150), $P^{(y)}Q^{(y)}$ is the chord through y in the direction e_{d_i}). Further, since $g_i = 1$, we see from eq. (150) that the density at z is given by:

$$\rho_i(z) = \int\limits_{y \in \tilde{P}\tilde{Q}} \rho_{i-1}(y) \cdot \frac{I[z \in P_\kappa^{(y)} Q_\kappa^{(y)}]}{\operatorname{vol}_1\left(P_\kappa^{(y)} Q_\kappa^{(y)}\right)} dy. \tag{157}$$

Using $\tilde{P}\tilde{Q} = P^{(y)}Q^{(y)}$, we get from Observation 6.6 that

$$\operatorname{vol}_1\left(P_\kappa^{(y)} Q_\kappa^{(y)}\right) = (1-\kappa)\operatorname{vol}_1\left(P^{(y)} Q^{(y)}\right) = (1-\kappa)\operatorname{vol}_1\left(\tilde{P}\tilde{Q}\right). \tag{158}$$

Substituting this along with the induction hypothesis in eq. (157), and using $S(i) = S(i-1)$, we get that for every z,

$$\rho_i(z) \le M_{i-1} \cdot \frac{\operatorname{vol}_1\left(\tilde{P}\tilde{Q}\right)}{(1-\kappa)\operatorname{vol}_1\left(\tilde{P}\tilde{Q}\right)} = \frac{M_{i-1}}{1-\kappa} = (1-\kappa)^{-i}(2\kappa^\tau\delta)^{-|S(i-1)|} = M_i, \tag{159}$$

which completes the induction in this case.

Case 2: $|S(i)| = |S(i-1)| + 1$**.** In this case, the ith step is the first occurrence of the direction e_{d_i} in the run of the chain. Again, $X_i = z$ is possible only if $X_{i-1} = y$ lies on the chord $\tilde{P}\tilde{Q}$ through z in the direction e_{d_i}. But in this case, for each $z \in (X_0 + \mathbb{R}^{|S(i)|} \times \{0\}^{n-|S(i)|}) \cap K$, there is at most one $y = y(z)$, obtained by setting the d_ith coordinate of z equal to the d_ith coordinate of x_0, for which $\rho_{i-1}(y) > 0$ and which lies on a chord through z in the direction e_{d_i}. Since we also have $g_i = 1$, we therefore get from eq. (150) that the density ρ_i (which is now with respect to an $(|S_{i-1}| + 1)$-dimensional Lebesgue measure), is given by

$$\rho_i(z) = \rho_{i-1}(y) \cdot \frac{I[z \in P_\kappa^{(y)} Q_\kappa^{(y)}]}{\mathrm{vol}_1\left(P_\kappa^{(y)} Q_\kappa^{(y)}\right)}. \tag{160}$$

Again, since we have $g = \mathbf{1}$, we see from eq. (156) (applied with $j = i - 1$) that if $\rho_{i-1}(y) > 0$ then $\mathrm{vol}_1\left(P_\kappa^{(y)} Q_\kappa^{(y)}\right) = (1-\kappa)\mathrm{vol}_1\left(P^{(y)} Q^{(y)}\right) \geq 2(1-\kappa)(\kappa^\tau\delta)$. Substituting this along with the induction hypothesis in eq. (160), we thus get (using $|S(i)| = |S(i-1)| + 1$)

$$\rho_i(z) \leq \frac{M_{i-1}}{2(1-\kappa)\kappa^\tau\delta} \leq (1-\kappa)^{-i}(2\kappa^\tau\delta)^{-(|S(i-1)|+1)} = M_i, \tag{161}$$

which completes the induction in this case as well.

We thus get that when d is complete and $g = \mathbf{1}$, then $\nu_{d,g} = \nu_{d,g,\tau}$ is M_τ-warm with respect to the n-dimensional Lebesgue measure on $\mathbb{R}^n$ (and is, of course, supported only on K). Since π_K has density $I[\cdot \in K]/\mathrm{vol}_n(K)$ with respect to the Lebesgue measure, it follows that $\nu_{d,g}$ (for such d and g) is $M_\tau \mathrm{vol}(K)$-warm with respect to π_K. Now, since we assumed $K \subseteq RB_\infty$ with $R \leq 1$, we have $\mathrm{vol}(K) \leq 2^n$. Thus, we see that when d is complete and $g = \mathbf{1}$, $\nu_{d,g}$ is $((1-\kappa)^{-\tau} \cdot 2^n \cdot (2\kappa^\tau\delta)^{-n})$-warm with respect to π_K. Given the discussion around eqs. (152) and (153), this completes the proof of the decomposition in eq. (148) (after recalling that $0 < \epsilon < 1/2, \tau = \lceil 2n\log(6n/\epsilon)\rceil$, and $\kappa = \epsilon/(6\tau)$).

Given the decomposition in eq. (148), the claimed bound on the mixing time now follows using an application of Theorem 1.1 for M-warm starts. From that theorem, we see that that there is an absolute constant C' such that when $t \geq T' := C'\frac{n^9R^2}{r^2}\log(6M/\epsilon)$ (where M is as in eq. (149)) then

$$d_{TV}(\nu_{\mathrm{cold}}\mathcal{M}_{CHR}^t, \pi_K) \leq \frac{\epsilon}{6}. \tag{162}$$

Using the decomposition of ν in eq. (148) and the triangle inequality, this implies that

$$d_{TV}(\nu\mathcal{M}_{CHR}^t, \pi_K) \leq p d_{TV}(\nu_{\mathrm{cold}}\mathcal{M}_{CHR}^t, \pi_K) + (1-p) d_{TV}(\nu_{\mathrm{rest}}\mathcal{M}_{CHR}^t, \pi_K) \tag{163}$$

$$\overset{\text{eq. (162)}}{\leq} p\frac{\epsilon}{6} + (1-p) \overset{\text{eq. (154)}}{\leq} \frac{\epsilon}{2}. \tag{164}$$

Since $\nu = \Delta_{X_0}\mathcal{M}_{CHR}^\tau$, where Δ_{X_0} is the probability distribution on K which puts probability mass 1 on X_0, we therefore get that the mixing time from X_0 up to a total variation distance of ϵ to π_K is at most $T' + \tau$, and the claimed bound in the statement of the theorem follows from the expressions given above for M, T' and τ. ◻

Appendix

A Proofs omitted from main text

A.1 Geometry

Here, we provide the omitted proofs of Lemmas 5.1 and 5.2, both of which we restate here for convenience.

Lemma 5.1. *Let K be a convex body, $\mathcal{F}$ a Whitney decomposition of it as described in Section 3, and consider any set $S \subseteq \mathcal{F}$. As before, we identify S also with the union of cubes in S. For any cube $Q \in S$, we have*

$$\mathrm{vol}_{n-1}(\partial S \cap \partial Q) = \lim_{\epsilon \downarrow 0} \frac{\mathrm{vol}_n\left((Q + \epsilon B_\infty) \setminus S\right)}{\epsilon}. \tag{92}$$

Proof. From item 5 of Theorem 3.1, we know that each axis-aligned dyadic cube in S that abuts the axis-aligned dyadic cube Q must have sidelength whose ratio with the sidelength of Q lies in $\left\{2^{-1}, 1, 2\right\}$. Thus, the area $\partial Q \cap \partial S$ can be partitioned into a finite set T of disjoint $(n-1)$-dimensional axis-aligned cuboids lying on the surface of Q, each having a non-zero surface area. Let $a > 0$ be the minimum sidelength over all the sidelengths of cuboids in T, and let $A := \mathrm{vol}_{n-1}(\partial Q \cap \partial S)$ denote the total surface area of cuboids in T. Let ϵ be small enough: e.g. $\epsilon \leq a/100$ suffices. By considering the cuboids obtained by expanding each side of each cuboid in T by ϵ in both directions, we can then sandwich the n-dimensional volume of the set $(Q + \epsilon B_\infty) \setminus S$ as follows:

$$A\epsilon \leq \mathrm{vol}_n\left((Q + \epsilon B_\infty) \setminus S\right) \leq A\epsilon \cdot \left(1 + \frac{\epsilon}{a}\right)^{n-1}. \tag{165}$$

The lemma now follows by dividing by ϵ throughout in eq. (165) and then taking the limit as $\epsilon \downarrow 0$. □

Lemma 5.2. *Fix a convex body $K \subseteq \mathbb{R}^n$. Then there exists $\delta = \delta(n, p)$ such that for all $\epsilon \in [0, \delta]$, and all points $x, y \in K^\circ$ with*

$$\|x - y\|_\infty \geq \epsilon \cdot \mathrm{dist}_{\ell_p}(x, \partial K), \tag{93}$$

it holds that

$$\mathrm{dist}_{g_p}(x, y) \geq \frac{\epsilon}{2}. \tag{94}$$

Proof. Pick $\delta := 1/(2n^{1/p})$. Consider the cube $L := x + 0.9\epsilon\, \mathrm{dist}_{\ell_p}(x, \partial K) B_\infty$. We then have $x \in L$ and $y \notin L$. Further, for any point $z \in L$, we have

$$\mathrm{dist}_{\ell_p}(z, \partial K) \leq \mathrm{dist}_{\ell_p}(x, \partial K) + 0.9\epsilon\, \mathrm{dist}_{\ell_p}(x, \partial K) \cdot n^{1/p} \leq 1.5\, \mathrm{dist}_{\ell_p}(x, \partial K), \tag{166}$$

since $\epsilon \leq \delta = 1/(2n^{1/p})$. Now, let $\gamma : [0, 1] \to K^\circ$ be any piecewise differentiable curve with $\gamma(0) = x$ and $\gamma(1) = y$. To prove the lemma, we only need to show that the length of any such curve (in the g_p metric) is at least $\epsilon/2$. We now proceed to do so, by considering the part of the curve that lies within L. Let $t_0 := \inf\{t : \gamma(t) \notin L\}$. Note that since $\gamma(0) = x \in L$ and $\gamma(1) = y \notin L$, $t_0 \in (0, 1)$ and $\gamma(t_0) \in \partial L$. We then have

$$
\begin{aligned}
\text{length}_{g_p}(\gamma) &\geq \int_0^{t_0} \frac{\|\gamma'(t)\|_\infty}{\text{dist}_{\ell_p}(\gamma(t), \partial K)} dt \\
&\geq \frac{2}{3\,\text{dist}_{\ell_p}(x, \partial K)} \int_0^{t_0} \|\gamma'(t)\|_\infty dt, \text{ applying eq. (166) to } \gamma(t) \in L, \\
&\geq \frac{2}{3\,\text{dist}_{\ell_p}(x, \partial K)} \cdot \|x - \gamma(t_0)\|_\infty \geq \frac{\epsilon}{2},
\end{aligned}
$$

where the last inequality uses that $\gamma(t_0) \in \partial L$ so that $\|x - \gamma(t_0)\|_\infty = 0.9\epsilon\,\text{dist}_{\ell_p}(x, \partial K)$. Since the curve γ is an arbitrary piecewise differentiable curve connecting x to y, this completes the proof. $\square$

A.2 Properties of Whitney cubes

Here, we provide a proof of Theorem 3.1, which we restate here for convenience.

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 $\mathcal{F} = \mathcal{F}^{(p)}$ of K.*

1. *$\bigcup_{Q \in \mathcal{F}} Q = K^\circ$. Further, if $Q \in \mathcal{F}$, then $Q \notin \mathcal{Q}_0$.*

2. *The interiors Q_k° are mutually disjoint.*

3. *For any Whitney cube $Q \in \mathcal{F}$,*
$$2\,\text{diam}_{\ell_p}(Q) \leq \text{dist}_{\ell_p}(\text{center}(Q), \mathbb{R}^n \setminus (K^\circ)) \leq \frac{9}{2}\,\text{diam}_{\ell_p}(Q).$$

4. *For any Whitney cube $Q \in \mathcal{F}$ and $y \in Q$,*
$$\frac{3}{2}\,\text{diam}_{\ell_p}(Q) \leq \text{dist}_{\ell_p}(y, \mathbb{R}^n \setminus (K^\circ)) \leq 5\,\text{diam}_{\ell_p}(Q).$$
In particular, this is true when $\text{dist}_{\ell_p}(y, \mathbb{R}^n \setminus (K^\circ)) = \text{dist}_{\ell_p}(Q, \mathbb{R}^n \setminus (K^\circ))$.

5. *The ratio of sidelengths of any two abutting cubes lies in $\{1/2, 1, 2\}$.*

Proof of Theorem 3.1. Throughout the proof, we use $\lambda = 1/2$. At several places in the proof, we will also use the following simple fact without comment: If Q is an axis-aligned cube and $x \in Q$, then for any $1 \leq p \leq \infty$, $\text{dist}_{\ell_p}(\text{center}(Q), x) \leq (1/2)\,\text{diam}_{\ell_p}(Q)$.

1. Let $x \in K^\circ$. Choose a positive integer k such that $n^{1/p}/2^k < \lambda\,\text{dist}_{\ell_p}(x, \partial K)/4$. Let $Q'_k \in \mathcal{Q}_k$ such that $x \in Q'_k$, and let $(Q'_i)_{i=0}^{k-1}$ be cubes such that $Q'_i \in \mathcal{Q}_i$ is the cube whose child is $Q'_{i+1} \in \mathcal{Q}_{i+1}$. Note that each Q'_i also contains x. Suppose, if possible, that $x \notin \bigcup_{Q \in \mathcal{F}} Q$. Now consider $Q'_0 \in \mathcal{Q}_0$ as above, which, by its definition, contains x. Observe that
$$\text{dist}_{\ell_p}(\text{center}(Q'_0), K) \leq \text{dist}_{\ell_p}(\text{center}(Q'_0), x) \leq (1/2)\,\text{diam}_{\ell_p}(Q'_0) = n^{1/p}/2,$$

so $Q'_0 \in \mathcal{F}_0$. On the other hand, since k was chosen so that $(1/2) \cdot \operatorname{diam}_{\ell_p}(Q'_{k-1}) = \operatorname{diam}_{\ell_p}(Q'_k) = n^{1/p}/2^k < \lambda \operatorname{dist}_{\ell_p}(x, \partial K)/4$, we have

$$\begin{aligned}\lambda \operatorname{dist}_{\ell_p}(\operatorname{center}(Q'_{k-1}), \partial K) &\geq \lambda \operatorname{dist}_{\ell_p}(x, \partial K) - \lambda \operatorname{dist}_{\ell_p}(x, \operatorname{center}(Q'_{k-1})) \\ &\geq \lambda \operatorname{dist}_{\ell_p}(x, \partial K) - (\lambda/2) \operatorname{diam}_{\ell_p}(Q'_{k-1}) \\ &> (2 - \lambda/2) \operatorname{diam}_{\ell_p}(Q'_{k-1}) > \operatorname{diam}_{\ell_p}(Q'_{k-1}),\end{aligned}$$

so that even if Q'_{k-1} were present in $\mathcal{F}_{k-1}$, it would not be subdivided into its children: this implies that $Q'_k \notin \mathcal{F}_k$. Since $Q'_0 \in \mathcal{F}_0$, it follows that there exists a j satisfying $0 \leq j < k$ such that $Q'_j \in \mathcal{F}_j$ but $Q'_{j+1} \notin \mathcal{F}_{j+1}$. We shall show that $Q'_j \in \mathcal{F}$.

To do this, it suffices to show that $\operatorname{center}(Q'_j) \in K^\circ$. By the definition of j, Q'_j is not sub-divided, so that we have $\lambda \operatorname{dist}_{\ell_p}(\operatorname{center}(Q'_j), \partial K) \geq \operatorname{diam}_{\ell_p}(Q'_j)$. Suppose, if possible, that $\operatorname{center}(Q'_j) \in \mathbb{R}^n \setminus (K^\circ)$. Then, $\operatorname{dist}_{\ell_p}(\operatorname{center}(Q'_j), \partial K) = \operatorname{dist}_{\ell_p}(\operatorname{center}(Q'_j), K)$, so that

$$(1/2) \operatorname{diam}_{\ell_p}(Q'_j) \geq \operatorname{dist}_{\ell_p}(\operatorname{center}(Q'_j), x) \geq \operatorname{dist}_{\ell_p}(\operatorname{center}(Q'_j), K) \geq (1/\lambda) \operatorname{diam}_{\ell_p}(Q'_j),$$

which is a contradiction since $\lambda = 1/2$. We have thus shown that $K^\circ \subseteq \bigcup_{Q \in \mathcal{F}} Q$.

The reverse containment follows because for any cube $Q \in \mathcal{F}$,

$$\begin{aligned}\operatorname{dist}_{\ell_p}(\operatorname{center}(Q), \mathbb{R}^n \setminus (K^\circ)) &= \operatorname{dist}_{\ell_p}(\operatorname{center}(Q), \partial K) && \text{(because } \operatorname{center}(Q) \in K^\circ\text{)} \\ &\geq (1/\lambda) \operatorname{diam}_{\ell_p}(Q) && (Q \text{ is not further subdivided}) \\ &> \operatorname{diam}_{\ell_p}(Q), && \text{(since } \lambda = 1/2\text{)}\end{aligned}$$

and as a result, $Q \cap (\mathbb{R}^n \setminus (K^\circ)) = \emptyset$. This proves the first part of item 1. The second part then follows since $K \subseteq R_\infty \cdot B_\infty$ with $R_\infty < 1$ implies that K cannot contain any cube in $\mathcal{Q}_0$.

2. If possible, let $Q \in \mathcal{F}_i \subseteq \mathcal{Q}_i$ and $Q' \in \mathcal{F}_j \subseteq \mathcal{Q}_j$ be distinct Whitney cubes with $i \leq j$, such that $Q^\circ \cap (Q')^\circ \neq \emptyset$. Note that the interiors of any two distinct cubes in $\mathcal{Q}_i$ do not intersect, so it must be the case that $i < j$. Then, since cubes in $\mathcal{Q}_j$ are obtained by a sequence of subdivisions of cubes in $\mathcal{Q}_i$, it must be the case that Q' is a descendant of Q, i.e., obtained by a sequence of subdivisions of Q. However, since $Q \in \mathcal{F}_i$, the construction of $\mathcal{F}_i$ implies that no child of Q can be present in $\mathcal{F}_{i+1}$. This implies that no cube obtained by subdivisions of Q is present in any $\mathcal{F}_j$ for $j > i$, and thus leads to a contradiction since $Q' \in \mathcal{F}_j$ was required to be a descendant of Q.

3. The first inequality is direct since if it did not hold, we would have further subdivided Q, so that Q would not be in $\mathcal{F}$. From item 1, we know that $Q \notin \mathcal{Q}_0 \supseteq \mathcal{F}_0$. So, let $k \geq 0$ be such that $Q \in \mathcal{F}_{k+1}$, and let $Q' \in \mathcal{F}_k$ be its parent whose subdivision led to the inclusion of $Q \in \mathcal{F}_{k+1}$. We then have

$$\begin{aligned}2 \operatorname{diam}_{\ell_p}(Q) &= \operatorname{diam}_{\ell_p}(Q') \\ &> \lambda \operatorname{dist}_{\ell_p}(\operatorname{center}(Q'), \mathbb{R}^n \setminus (K^\circ)) \qquad \text{(since } Q' \text{ was subdivided)} \\ &\geq \lambda \operatorname{dist}_{\ell_p}(\operatorname{center}(Q), \mathbb{R}^n \setminus (K^\circ)) - \lambda \operatorname{dist}_{\ell_p}(\operatorname{center}(Q'), \operatorname{center}(Q)) \\ &\geq \lambda \operatorname{dist}_{\ell_p}(\operatorname{center}(Q), \mathbb{R}^n \setminus (K^\circ)) - (\lambda/2) \operatorname{diam}_{\ell_p}(Q),\end{aligned}$$

where the last inequality uses the fact that $\operatorname{center}(Q')$ is a vertex of Q. This implies

$$\operatorname{dist}_{\ell_p}(\operatorname{center}(Q), \mathbb{R}^n \setminus (K^\circ)) < \left(\frac{2}{\lambda} + \frac{1}{2}\right) \operatorname{diam}_{\ell_p}(Q) = \frac{9}{2} \operatorname{diam}_{\ell_P}(Q). \tag{167}$$

4. Set $x = \mathrm{center}(Q)$. Also let $z_x, z_y \in \mathbb{R}^n \setminus (K^\circ)$ such that $\mathrm{dist}_{\ell_p}(x, z_x) = \mathrm{dist}_{\ell_p}(x, \mathbb{R}^n \setminus (K^\circ))$ and $\mathrm{dist}_{\ell_p}(y, z_y) = \mathrm{dist}_{\ell_p}(y, \mathbb{R}^n \setminus (K^\circ))$. Then,
$$\mathrm{dist}_{\ell_p}(y, z_y) \geq \mathrm{dist}_{\ell_p}(x, z_y) - \mathrm{dist}_{\ell_p}(x, y) \geq \mathrm{dist}_{\ell_p}(x, z_x) - \mathrm{dist}_{\ell_p}(x, y),$$
and similarly,
$$\mathrm{dist}_{\ell_p}(y, z_y) \leq \mathrm{dist}_{\ell_p}(y, z_x) \leq \mathrm{dist}_{\ell_p}(x, z_x) + \mathrm{dist}_{\ell_p}(x, y).$$
Using the upper and lower bounds on $\mathrm{dist}_{\ell_p}(x, z_x)$ derived in item 3 along with the above inequalities and the bound $\mathrm{dist}_{\ell_p}(x, y) \leq (1/2)\,\mathrm{diam}_{\ell_p}(Q)$ yields the claimed bounds.

5. Let $Q_1, Q_2 \in \mathcal{F}$ be abutting cubes with $\mathrm{diam}(Q_1) > \mathrm{diam}(Q_2)$, and let $y \in Q_1 \cap Q_2$. Item 4 applied to both Q_1 and Q_2 then gives
$$\frac{3}{2}\,\mathrm{diam}_{\ell_p}(Q_1) \leq \mathrm{dist}_{\ell_p}(y, \mathbb{R}^n \setminus (K^\circ)) \leq 5\,\mathrm{diam}_{\ell_p}(Q_2).$$
This implies
$$1 > \frac{\mathrm{diam}_{\ell_p}(Q_2)}{\mathrm{diam}_{\ell_p}(Q_1)} = \frac{\mathrm{sidelength}\,(Q_2)}{\mathrm{sidelength}\,(Q_1)} \geq \frac{3}{10}.$$
Since the ratio of sidelengths of any two Whitney cubes is an integral power of two, this forces the ratio of the sidelengths of Q_2 and Q_1 to be $1/2$. We conclude that if two cubes $Q_1, Q_2 \in \mathcal{F}$ are abutting, then the ratio of their sidelengths is an element of the set $\{1/2, 1, 2\}$. □

B Some results used in proofs

B.1 The isoperimetric inequality of Kannan, Lovász and Montenegro

In the proof of Theorem 5.1, an isoperimetric inequality due to Kannan, Lovász and Montenegro [KLM06] was used. In their paper, Kannan, Lovász and Montenegro state their inequality only when the distance between the sets and the diameter of the convex body are both measured using the ℓ_2-norm. However, since their proof uses the localization lemma framework of Lovász and Simonovits [LS93] to reduce the problem to the setting of a line segment, it applies without any changes even when the corresponding quantities are measured in any other ℓ_p norm, where $1 \leq p \leq \infty$ (see, e.g., the statement of Corollary 2.7 of [LS93]). For completeness, we reproduce the statement of the isoperimetric inequality of Kannan, Lovász, and Montenegro in this more general form, and also provide a short sketch of how their proof applies also in this setting.

Theorem B.1 (Kannan, Lovász, and Montenegro [KLM06, Theorem 4.3]). *Fix p satisfying $1 \leq p \leq \infty$. Let K be a convex body, and let S_1, S_2 be disjoint measurable subsets of K. Define $S_3 := K \setminus (S_1 \cup S_2)$. Let $\epsilon, D > 0$ be such that for any two points $x, y \in K$, $\mathrm{dist}_{\ell_p}(x, y) \leq D$, and such that $\mathrm{dist}_{\ell_p}(S_1, S_2) \geq \epsilon$. Then*
$$\mathrm{vol}\,(S_3) \geq \frac{\epsilon}{D} \cdot \frac{\mathrm{vol}\,(S_1)\,\mathrm{vol}\,(S_2)}{\mathrm{vol}\,(K)} \cdot \log\left(1 + \frac{\mathrm{vol}\,(K)^2}{\mathrm{vol}\,(S_1)\,\mathrm{vol}\,(S_2)}\right). \tag{168}$$

Proof sketch. The theorem above is stated and proved by Kannan, Lovász and Montenegro for the case $p = 2$ in Section 6.4 of [KLM06]. To prove the result for other p, simply repeat the same proof (with ϵ and D defined with respect to ℓ_p instead of ℓ_2). The only point one has to note is that when [KLM06] apply their Lemma 6.1 in the last paragraph of their proof, they only need to consider ratios of lengths which lie along the same line segment, and such a ratio does not depend upon which ℓ_p-norm is used to measure the corresponding lengths. □

B.2 Vertex expansion of the hypercube graph

In this section, we elaborate on the bound on the vertex expansion used in Case 2(b) of the proof of Theorem 6.2, and restate the relevant results in [AGK76]. We begin by restating some of the notation from [AGK76].

Let $\mathscr{Y}$ be a finite set, W a probability distribution on $\mathscr{Y}$, and define the product probability distribution W^n on $\mathscr{Y}^n$ as

$$W^n(y) = \prod_{i=1}^{n} W(y_i)$$

for $y \in \mathscr{Y}^n$. For $y, y' \in \mathscr{Y}^n$, introduce the Hamming distance

$$d(y, y') := |\{1 \le i \le n \,:\, y_i \ne y'_i\}|.$$

For a set $\mathscr{B} \subseteq \mathscr{Y}^n$, define the Hamming neighbourhood $\Gamma\mathscr{B}$ of $\mathscr{B}$ as

$$\Gamma\mathscr{B} := \{y \in \mathscr{Y}^n \,:\, d(y, y') \le 1 \text{ for some } y' \in \mathscr{B}\}$$

and the inner boundary $\partial\mathscr{B}$ of $\mathscr{B}$ as

$$\partial\mathscr{B} := \Gamma\mathscr{B}^c \cap \mathscr{B}.$$

Also set

$$\varphi(t) = (2\pi)^{-1/2} e^{-t^2/2},$$
$$\Phi(t) = \int_{-\infty}^{t} \varphi(x)dx, \text{ and}$$
$$f(s) = \varphi(\Phi^{-1}(s)).$$

Setting $\mathscr{X} = \{0\}$ in Theorem 5 of [AGK76], we obtain the following.

Theorem B.2 ([**AGK76**]). *There is a constant a depending only on W such that for any $\mathscr{B} \subseteq \mathscr{Y}^n$,*

$$W^n(\partial\mathscr{B}) \ge an^{-1/2} f(W^n(\mathscr{B})).$$

Corollary B.3. *Let Q_n be the n-dimensional hypercube graph (V, E), where $V = \{0, 1\}^n$ and vertices u, v are adjacent if and only if their Hamming distance is 1. Set μ to be the uniform distribution on $\{0, 1\}^n$. For any $S \subseteq V$ with $\mu(S) \le (1/2)$, there exists a universal constant c such that*

$$\frac{\mu(\Gamma(S^c) \cap S)}{\mu(S)} \ge cn^{-1/2},$$

where $\Gamma(S)$ denotes the neighbourhood of vertices in S.

Proof. Setting $\mathscr{Y} = \{0, 1\}$, $\mathscr{B} = S$, and W as the uniform distribution on $\{0, 1\}$ in the previous theorem, we get that

$$\mu(\Gamma(S^c) \cap S) \ge cn^{-1/2} f(\mu(S)).$$

To conclude, we note that for $\mu(S) \le (1/2)$, $f(\mu(S)) \ge c'\mu(S)$ for a universal constant c. ◻

References

[AGK76] R. Ahlswede, P. Gács, and J. Körner. Bounds on conditional probabilities with applications in multi-user communication. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 34(2):157–177, June 1976.

[AK91] D. Applegate and R. Kannan. Sampling and integration of near log-concave functions. In *Proc. 23rd ACM Symposium on Theory of Computing (STOC)*, pages 156–163. ACM, January 1991.

[BC13] P. Bürgisser and F. Cucker. *Condition*. Springer, 2013.

[BDJ98] R. Bubley, M. Dyer, and M. Jerrum. An elementary analysis of a procedure for sampling points in a convex body. *Random Structures & Algorithms*, 12(3):213–235, 1998.

[CDWY18] Y. Chen, R. Dwivedi, M. J. Wainwright, and B. Yu. Fast MCMC sampling algorithms on polytopes. *JMLR*, 19:86, 2018.

[CE22] Y. Chen and R. Eldan. Localization schemes: A framework for proving mixing bounds for Markov chains, March 2022. To appear in *Proc. 63rd IEEE Symposium on Foundations of Computer Science (FOCS), 2022*. arXiv:2203.04163.

[CZ52] A. P. Calderón and A. Zygmund. On the existence of certain singular integrals. *Acta Mathematica*, 88:85 – 139, 1952.

[DF91] M. Dyer and A. Frieze. Computing the volume of convex bodies: a case where randomness provably helps. In B. Bollobás, editor, *Proceedings of Symposia in Applied Mathematics*, volume 44, pages 123–169. American Mathematical Society, 1991.

[DFK91] M. Dyer, A. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, January 1991.

[EG15] L. Evans and R. Gariepy. *Measure Theory and Fine Properties of Functions, Revised Edition*. Chapman and Hall/CRC, 2015.

[Fed96] H. Federer. *Geometric measure theory*. Classics in mathematics. Springer, 1996. Reprint of the 1969 edition.

[Fef05] C. L. Fefferman. A sharp form of Whitney's extension theorem. *Annals of Mathematics*, 161(1):509–577, 2005.

[FK09] C. L. Fefferman and B. Klartag. Fitting a C^m-Smooth Function to Data II. *Revista Matemática Iberoamericana*, 25(1):49 – 273, 2009.

[FSA20] S. Fallahi, H. J. Skaug, and G. Alendal. A comparison of Monte Carlo sampling methods for metabolic network models. *PLOS ONE*, 15(7):e0235393, July 2020.

[Har66] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory, Series A*, 1:385–393, 1966.

[HCT+17] H. S. Haraldsdóttir, B. Cousins, I. Thiele, R. M. T. Fleming, and S. S. Vempala. CHRR: Coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33(11):1741–1743, June 2017.

[JS88] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for Markov chains: the approximation of permanent resolved. In *Proc. 20th ACM Symposium on Theory of Computing (STOC)*, pages 235–244. ACM, January 1988.

[KLM06] R. Kannan, L. Lovász, and R. Montenegro. Blocking conductance and mixing in random walks. *Combinatorics, Probability and Computing*, 15(4):541–570, July 2006.

[KN12] R. Kannan and H. Narayanan. Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37(1):1–20, 2012.

[LK99] L. Lovász and R. Kannan. Faster mixing via average conductance. In *Proc. 31st Annual ACM Symposium on Theory of Computing (STOC)*, pages 282–287. ACM, 1999.

[LLV20] A. Laddha, Y. T. Lee, and S. S. Vempala. Strong self-concordance and sampling. In *Proc. 52nd ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1212–1222. ACM, June 2020. arXiv:1911.05656.

[Lov90] L. Lovász. Geometric algorithms and algorithmic geometry. In *Proceedings of the International Congress of Mathematicians (ICM)*, volume 1, pages 139–154, 1990.

[Lov99] L. Lovász. Hit-and-run mixes fast. *Math. Program.*, 86(3):443–461, December 1999.

[LS90] L. Lovász and M. Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *Proc. 31st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 346–354 vol. 1. IEEE, October 1990.

[LS93] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures & Algorithms*, 4(4):359–412, 1993.

[LV06] L. Lovász and S. S. Vempala. Hit-and-run from a corner. *SIAM J. Comput.*, 35(4):985–1005, January 2006.

[LV17] Y. T. Lee and S. S. Vempala. Geodesic walks in polytopes. In *Proc. 49th ACM Symposium on Theory of Computing (STOC)*, pages 927–940. ACM, June 2017.

[LV21] A. Laddha and S. S. Vempala. Convergence of Gibbs sampling: Coordinate hit-and-run mixes fast. In *Proc. 37th International Symposium on Computational Geometry (SOCG)*, pages 51:1–51:12, 2021. arXiv:2009.11338.

[LV22] Y. T. Lee and S. S. Vempala. The manifold joys of sampling. In *Proc. 49th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 229 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 4:1–4:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. Invited talk.

[MS01] R. Montenegro and J.-B. Son. Edge isoperimetry and rapid mixing on matroids and geometric Markov chains. In *Proc. 33rd ACM Symposium on Theory of Computing (STOC)*, STOC '01, pages 704–711, New York, NY, USA, July 2001. Association for Computing Machinery.

[MV19] O. Mangoubi and N. K. Vishnoi. Faster polytope rounding, sampling, and volume computation via a sub-linear ball walk. In *Proc. 60th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1338–1357. IEEE, November 2019.

[Nar16] H. Narayanan. Randomized interior point methods for sampling and optimization. *The Annals of Applied Probability*, 26(1):597–641, February 2016.

[NS22] H. Narayanan and P. Srivastava. On the mixing time of coordinate hit-and-run. *Combinatorics, Probability and Computing*, 31(2):320–332, 2022. arXiv:2009.14004.

[Smi84] R. L. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):1296–1308, 1984.

[Ste70] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions (PMS-30)*. Princeton University Press, 1970.

[TV17] E. Tsukerman and E. Veomett. Brunn-Minkowski theory and Cauchy's surface area formula. *American Mathematical Monthly*, 124(10):922–929, 2017.

[Whi34] H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934.