Random Walks on Polytopes and an Affine Interior Point Method for Linear Programming

Ravi Kannan
Algorithms Research Group
Microsoft Research Labs, India
kannan@microsoft.com

Hariharan Narayanan
Department of Computer Science
University of Chicago
hari@cs.uchicago.edu

Abstract

Let K be a polytope in $\mathbb{R}^n$ defined by m linear inequalities. We give a new Markov Chain algorithm to draw a nearly uniform sample from K. The underlying Markov Chain is the first to have a mixing time that is strongly polynomial when started from a "central" point x_0. If s is the supremum over all chords $\overline{pq}$ passing through x_0 of $\frac{|p-x_0|}{|q-x_0|}$ and ϵ is an upper bound on the desired total variation distance from the uniform, it is sufficient to take $O\left(mn\left(n\log(sm)+\log\frac{1}{\epsilon}\right)\right)$ steps of the random walk. We use this result to design an affine interior point algorithm that does a *single* random walk to solve linear programs approximately. More precisely, suppose $Q=\{z\big|Bz\leq \mathbf{1}\}$ contains a point z such that $c^Tz\geq d$ and $r:=\sup_{z\in Q}\|Bz\|+1$, where B is an $m\times n$ matrix. Then, after $\tau=O\left(mn\left(n\ln\left(\frac{mr}{\epsilon}\right)+\ln\frac{1}{\delta}\right)\right)$ steps, the random walk is at a point x_τ for which $c^Tx_\tau\geq d(1-\epsilon)$ with probability greater than $1-\delta$. The fact that this algorithm has a run-time that is provably polynomial is notable since the analogous deterministic affine algorithm analyzed by Dikin has no known polynomial guarantees.

1 Introduction

We use ideas from interior point algorithms to define a random walk on a polytope. We call this walk *Dikin walk*. The Markov Chain defining Dikin walk is invariant under affine transformations of the polytope. Consequently, the complex interleaving of rounding and sampling present in previous sampling algorithms for convex sets (see [6, 7, 16]) is unnecessary. The following are notable features of Dikin walk.

1. The measures defined by the transition probabilities of Dikin walk are affine invariants, so there is no dependence on R/r (where R is the radius of the smallest ball containing the polytope K and r is the radius of the largest ball contained in K).

2. If K is an n-dimensional polytope defined by m linear constraints, the mixing time of the Dikin walk is $O(nm)$ from a warm start (i. e. if the starting distribution has a density bounded above by a constant).

3. If the walk is started at the "analytic center" (which can be found efficiently by interior point methods [20, 21]), it achieves a variation distance of ϵ in $O\left(mn\left(n\log m+\log\frac{1}{\epsilon}\right)\right)$ steps. This is strongly polynomial in the description of the polytope.

Previous sampling algorithms were applicable to convex sets specified in the following way. The input consists of an n-dimensional convex set K circumscribed around and inscribed in balls of radius r and R respectively. The algorithm has access to an oracle that when supplied with a point in $\mathbb{R}^n$ answers "yes" if the point is in K and "no" otherwise.

The first polynomial time algorithm for sampling convex sets appeared in [6]. It did a random walk on a sufficiently dense grid. The dependence of its mixing time on the dimension was $O^*(n^{23})$. It resulted in the first randomized polynomial time algorithm to approximate the volume of a convex set.

Another random walk that has been analyzed for sampling convex sets is known as the ball walk, which does the following. Suppose the current point is x_i. y is chosen uniformly at random from a ball of radius δ centered at x_i. If $y \in K$, x_{i+1} is set to K; otherwise $x_{i+1} = x_i$. After many successive improvements over several papers, it was shown in [7] that a ball walk mixes in $O^*(n\frac{R^2}{\delta^2})$ steps from a warm start if $\delta < \frac{r}{\sqrt{n}}$. A ball walk has not been proved to mix rapidly from any single point. A third random walk analyzed recently is known as Hit-and-Run [12, 14]. This walk mixes in $O\left(n^3(\frac{R}{r})^2 \ln \frac{R}{d\epsilon}\right)$ steps from a point at a distance d from the boundary [14], where ϵ is the desired variation distance to stationarity. Dikin walk is similar to ball walk except that Dikin ellipsoids (defined later) are used instead of balls. Dikin walk is the first walk to mix in *strongly polynomial* time from a central point such as the center of mass (for which s, as defined below, is $O(n)$) and the analytic center (for which $s = O(m)$). Our main result related to the Dikin walk is the following.

Theorem 1. *Let n be greater than some universal constant. Let K be an n-dimensional polytope defined by m linear constraints and $x_0 \in K$ be a point such that s is the supremum over all chords $\overline{pq}$ passing through x_0 of $\frac{|p-x_0|}{|q-x_0|}$ and $\epsilon > 0$ be the desired variation distance to the uniform distribution. Let $\tau > 7 \times 10^8 \times mn\left(n \ln(20\, s\sqrt{m}) + \ln\left(\frac{32}{\epsilon}\right)\right)$ and $x_0, x_1, \ldots$ be a Dikin walk. Then, for any measurable set $S \subseteq K$, the distribution of x_τ satisfies $\left|\mathbb{P}[x_\tau \in S] - \frac{\mathrm{vol}(S)}{\mathrm{vol}(K)}\right| < \epsilon$.*

1.0.1 Running times

The mixing time for Hit-and-Run from a warm start is $O\left(\frac{n^2R^2}{r^2}\right)$, while for Dikin walk this is $O(mn)$. Hit-and-Run takes more random walk steps to provably mix on any class of polytopes where $m = o\left(\frac{nR^2}{r^2}\right)$. For polytopes with polynomially many faces, R/r cannot be $O\left(n^{\frac{1}{2}-\epsilon}\right)$ (but can be arbitrarily larger). Thus, $m = o(n\left(\frac{R}{r}\right)^2)$ holds true for some important classes of polytopes, such as those arising from the question of sampling contingency tables with fixed row and column sums (where $m = O(n)$). Each step of Dikin walk can be implemented using $O(mn^{\gamma-1})$ arithmetic operations, $\gamma < 2.376$ being the exponent of matrix multiplication (see 2.1.1). One step of Hit-and-Run implemented naively would need $O(mn)$ arithmetic operations. Evaluating costs in this manner, Hit-and-Run takes more random walk steps to provably mix on any class of polytopes where $m^\gamma = o\left(\frac{n^2R^2}{r^2}\right)$. A sufficient condition for $m = o\left(\frac{n^{3-\gamma}R^2}{r^2}\right)$ to hold is $m = o(n^{4-\gamma})$.

1.1 Applications

1.1.1 Sampling lattice points in polytopes

While polytopes form a restricted subclass of the set of all convex bodies, algorithms for sampling polytopes have numerous applications. It was shown in [8] that if an n dimensional polytope defined by m inequalities contains a ball of radius $\Omega(n\sqrt{\log m})$, then it is possible to sample the lattice points inside it in polynomial time by sampling the interior of the polytope and picking a nearby lattice point. Often, combinatorial structures can be encoded as lattice points in a polytope, leading in this way to algorithms for sampling them. Contingency tables are two-way tables that are used by statisticians to represent bivariate data. A solution proposed in [4] to the frequently encountered problem of testing the independence of two characteristics of empirical data involves sampling uniformly from the set of two-way tables having fixed row and column sums. It was shown in [17] that under some conditions, this can be achieved in polynomial time by quantizing random points from an associated polytope.

1.1.2 Linear Programming

We use this result to design an affine interior point algorithm that does a *single* random walk to solve linear programs approximately. In this respect, our algorithm differs from existing randomized algorithms for linear programming such as that of Lovász and Vempala [15], which solves more general convex programs. While optimizing over a polytope specified as in the previous subsection, if $m = O(n^{2-\epsilon})$, the number of random steps taken by our algorithm is less than that of [15]. Given a polytope Q containing the origin and a linear objective c, our aim is to find with probability $> 1 - \delta$, a point $y \in Q$ such that $c^T y \geq 1 - \epsilon$ if there exists a point $z \in Q$ such that $c^T z \geq 1$. We first truncate Q using a hyperplane $c^T y = 1 - \hat{\epsilon}$, for $\hat{\epsilon} << \epsilon$ and obtain $Q_{\hat{\epsilon}} = Q \cap \{y | c^T y \leq 1 - \hat{\epsilon}\}$. We then projectively transform $Q_{\hat{\epsilon}}$ to "stretch" it into a new polytope $\gamma(Q_{\hat{\epsilon}})$ where $\gamma : y \mapsto \frac{y}{1 - c^T y}$. Finally, we do a simplified Dikin walk (without the Metropolis filter) on $\gamma(Q_{\hat{\epsilon}})$ which approaches close to the optimum in polynomial time. This algorithm is purely affine after one preliminary projective transformation, in the sense that Dikin ellipsoids are used that are affine invariants but not projective invariants. This is an important distinction in the theory of interior point methods and the fact that our algorithm is polynomial time is notable since the corresponding deterministic affine algorithm analyzed by Dikin [5, 23] has no known polynomial guarantees on its run-time. Its projective counterpart, the algorithm of Karmarkar however does [9]. In related work [2], Belloni and Freund have explored the use of randomization for preconditioning. While there is no "local" potential function that is improved upon in each step, our analysis may be interpreted as using the $\mathcal{L}_{2,\mu}$ norm (μ being the appropriate stationary measure) of the probability density of the k^{th} point as a potential, and showing that this reduces at each step by a multiplicative factor of $(1 - \frac{\Phi^2}{2})$ where Φ is the conductance of the walk on the transformed polytope. We use the $\mathcal{L}_{2,\mu}$ norm rather than variation distance because this allows us to give guarantees of exiting the region where the objective function is low before the relevant Markov Chain has reached approximate stationarity. The main result related to algorithm (Dikin) is the following.

Theorem 2. *Let n be larger than some universal constant. Given a system of inequalities* $By \leq \mathbf{1}$, *a linear objective c such that the polytope*

$$Q := \{y : By \leq \mathbf{1} \text{ and } |c^T y| \leq 1\}$$

is bounded, and $\epsilon, \delta > 0$, *the following is true. If* $\exists\, z$ *such that* $Bz \leq \mathbf{1}$ *and* $c^T z \geq 1$, *then* y, *the output of* Dikin , *satisfies*

$$By \leq \mathbf{1}$$
$$c^T y \geq 1 - \epsilon$$

with probability greater than $1 - \delta$.

1.1.3 Strong Polynomiality

Let us call a point x central if $\ln s$, where s is the function of x defined in Theorem 1, is polynomial in m. The mixing time of Dikin walk both from a warm start, and from a starting point that is central, is strongly polynomial in that the number of arithmetic operations depends only on m and n. Previous Markov Chains for sampling convex sets (and hence polytopes) do not possess either of these characteristics. In the setting of approximate Linear Programming that we have considered, the numbers of iterations taken by known interior point methods such as those of Karmarkar [9],

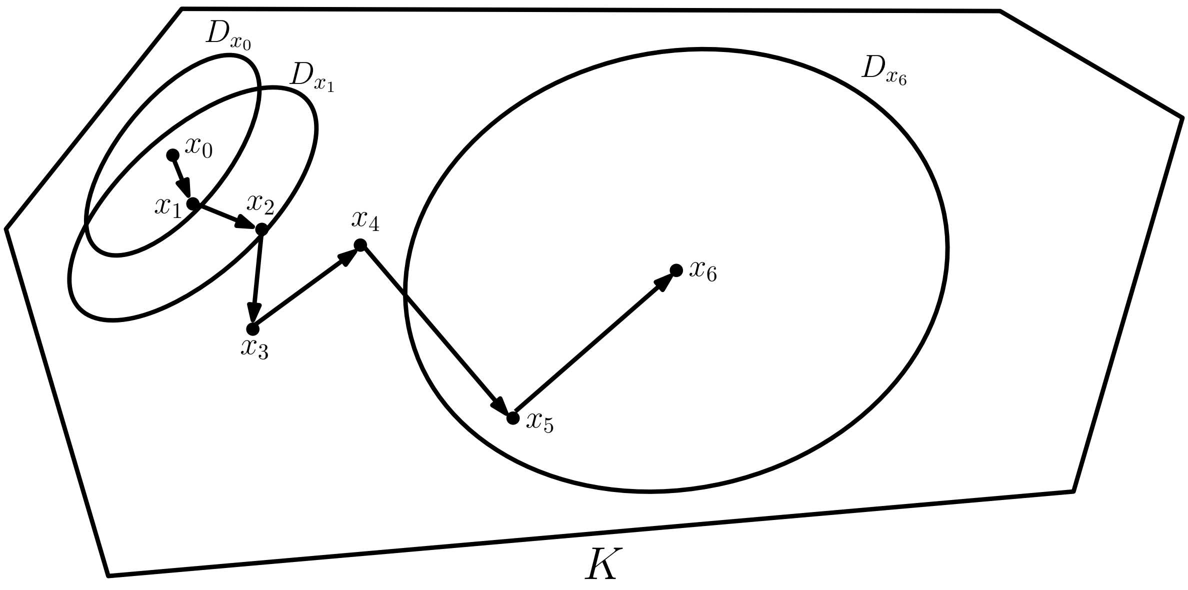

Figure 1: A realization of Dikin walk. Dikin ellipsoids D_{x_0}, D_{x_1} and D_{x_6} have been depicted.

Renegar [20], Vaidya [21] etc are strongly polynomial when started from a point that is central in the above sense. The algorithm Dikin presented here is no different in this respect. The fact that Dikin walk has a mixing time that is strongly polynomial from a central point such as the center of mass, is related to two properties of Dikin ellipsoids listed below.

1.1.4 Dikin ellipsoids and their virtues

Let K be a polytope in $n-$dimensional Euclidean space given as the intersection of m halfspaces $a_i^T x \leq 1$, $1 \leq i \leq m$. Defining A to be the $m \times n$ matrix whose i^{th} row is a_i^T, the polytope can be specified by $Ax \leq \mathbf{1}$. Let $x_0 \in int(K)$ belong to the interior of K. Let

$$H(x) = \sum_{1 \leq i \leq m} \frac{a_i a_i^T}{(1 - a_i^T x)^2}$$

and $\|z - x\|_x^2 := (z - x)^T H(x)(z - x)$. The *Dikin* ellipsoid D_x^r of radius r for $x \in K$ is the ellipsoid containing all points z such that

$$\|z - x\|_x \leq r.$$

Fact 1. *(1) Dikin ellipsoids are affine invariants in that if T is an affine transformation and $x \in K$, the Dikin ellipsoid of radius r centered at the point Tx for the polytope $T(K)$ is $T(D_x^r)$. This is easy to verify from their definition.*

(2) For any interior point x, the Dikin ellipsoid centered at x, having radius 1, is contained in K. This has been shown in Theorem 2.1.1 of [18]. Also, the Dikin ellipsoid at x having radius $\sqrt{m}$ contains $Sym_x(K) := K \cap \{y \big| 2x - y \in K\}$. This can be derived by an argument along the lines of Theorem 4.

2 Randomly Sampling Polytopes

2.1 Preliminaries

For two vectors v_1, v_2, let $\langle v_1, v_2 \rangle_x = v_1^T H(x) v_2$. For $x \in K$, we denote by D_x, the Dikin ellipsoid of radius $\frac{3}{40}$ centered at x. Dikin ellipsoids have been studied in the context of optimization [5] and have recently been used in online learning [1]. The second property mentioned in the subsection below implies that the Dikin walk does not leave K.

The "Dikin walk" is a "Metropolis" type walk which picks a move and then decides whether to "accept" the move and go there or "reject" and stay. The transition probabilities of the Dikin walk are listed below. When at x, one step of the walk is made as follows.

---

1. Flip an unbiased coin. If `Heads`, stay at x.
2. If `Tails` pick a random point y from D_x.
3. If $x \notin D_y$, then reject y (stay at x);
   if $x \in D_y$, then accept y with probability
   $\min\left(1, \frac{\mathrm{vol}(D_x)}{\mathrm{vol}(D_y)}\right) = \min\left(1, \sqrt{\frac{\det H(y)}{\det H(x)}}\right)$.

Therefore,

$$\mathbb{P}[x \to y] = \begin{cases} \min\left(\frac{1}{2\,\mathrm{vol}(D_x)}, \frac{1}{2\,\mathrm{vol}(D_y)}\right), \\ \text{if } y \in D_x \text{ and } x \in D_y; \\ 0, \text{otherwise.} \end{cases}$$

and $\mathbb{P}[x \to x] = 1 - \int_y d\mathbb{P}[x \to y]$.

2.1.1 Implementation of a Dikin step

Let K be the set of points satisfying the system of inequalities $Ax \leq \mathbf{1}$. $H(x) = A^T D(x)^2 A$ where $D(x)$ is the diagonal matrix whose i^{th} diagonal entry $d_{ii}(x) = \frac{1}{1 - a_i^T x}$.

We can generate a Gaussian vector v such that $\mathbb{E}[vv^T] = (A^T D^2 A)^{-1}$ by the following procedure. Let u be a random m-vector from a Gaussian distribution whose covariance matrix is Id. Find v that satisfies the linear equations:

$$DAv = z$$

$$A^T D(z - u) = 0,$$

or equivalently,

$$A^T D^2 A v = A^T D u.$$

Allowing $(DA)^\dagger$ to be the Moore-Penrose pseudo-inverse of DA,

$$(DA)^\dagger (z - u) = 0 \Leftrightarrow (z - u) \perp \mathrm{column\,span}(DA)$$

$$\Leftrightarrow A^T D(z - u) = 0.$$

Thus, $\mathbb{E}vv^T = (DA)^\dagger \mathbb{E}zz^T (DA)^{\dagger T}$. z is the orthogonal projection of u onto the column span of DA,

therefore $(DA)^\dagger \mathbb{E}zz^T (DA)^{\dagger T} = H(x)^{-1}$. We can now generate a random point from the Dikin ellipsoid by scaling $v/\|v\|_x$ appropriately. The probability of accepting a Dikin step, is either 0 or the minimum of 1 and ratio of two determinants. Two matrix-vector products suffice to test whether the original point lies in the Dikin ellipsoid of the new one. By results of Baur and Strassen [3], the complexity of solving linear equations and of computing the determinant of an $n \times n$ matrix is $O(n^\gamma)$. The most expensive step, the computation of $A^T D(x)^2 A$ can be acheived using $mn^{\gamma-1}$, by partitioning a padded extension of $A^T D$ into $\leq \frac{m+n-1}{n}$ square matrices. Thus, all the operations needed for one step of Dikin walk can be computed using $O(mn^{\gamma-1})$ arithmetic operations where $\gamma < 2.377$ is the exponent for matrix multiplication.

2.2 Isoperimetric inequality

Given interior points x, y in a polytope K, suppose p, q are the ends of the chord in K containing x, y and p, x, y, q lie in that order. Then we denote $\frac{|x-y||p-q|}{|p-x||q-y|}$ by $\sigma(x, y)$. $\ln(1 + \sigma(x, y))$ is a metric known as the Hilbert metric, and given four collinear points a, b, c, d, $(a : b : c : d) = \frac{(a-c)\cdot(b-d)}{(a-d)\cdot(b-c)}$ is known as the cross ratio.

The theorem below was proved by Lovász in [12].

Theorem 3 (Lovász)**.** *Let S_1 and S_2 be measurable subsets of K. Then,*

$$\mathrm{vol}(K \setminus S_1 \setminus S_2)\,\mathrm{vol}(K) \geq \sigma(S_1, S_2)\,\mathrm{vol}(S_1)\,\mathrm{vol}(S_2).$$

2.3 Dikin norm and Hilbert metric

Theorem 4 relates the Dikin norm to the Hilbert metric. The Dikin norms can be used to define a Riemannian manifold by using the associated bilinear form $< \cdot, \cdot >_x$ to construct a metric tensor. Dikin walk is a random walk analogous to the "ball walk" on such a manifold.

Observation 1. *The isoperimetric properties of this manifold can be deduced from those of the Hilbert metric, and in fact, Theorem 3 and Theorem 4 together imply that the weighted Cheeger constant of this manifold is bounded below by $\frac{1}{2\sqrt{m}}$.*

Theorem 4. *Let x, y be interior points of K. Then,*

$$\sigma(x, y) \geq \frac{\|x - y\|_x}{\sqrt{m}}.$$

Proof. It is easy to see that we can restrict attention to the line ℓ containing x, y. We may also assume that $x = 0$ after translation. So now $b_i \geq 0$. Let c_i be the component of a_i along ℓ; we may view c_i, y as real numbers with ℓ as the real line now. $K \cap \ell = \{y : c_i y \leq b_i\}$ (where b_i had been taken to be 1). Dividing constraint i by $|c_i|$, we may assume that $|c_i| = 1$. After renumbering constraints so that $b_1 = \min\{b_i \big| c_i = -1\}$ and $b_2 = \min\{b_i \big| c_i = 1\}$, we have $K \cap \ell = [-b_1, b_2]$. Also

$$\|x - y\|_x^2 = y^2 \sum_i \frac{1}{b_i^2}.$$

Without loss of generality, assume that $y \geq 0$. [The proof is symmetric for $y \leq 0$.] Then, $\sigma(x, y) = \frac{y(b_1+b_2)}{b_1(b_2-y)}$, which is $\geq y \max_i(1/|b_i|)$. This is in turn $\geq \frac{\|x-y\|_x}{\sqrt{m}}$. □

2.4 Geometric and probabilistic distance

Let the Lebesgue measure be denoted λ. The total variation distance between two distributions π_1 and π_2 is $d(\pi_1, \pi_2) := \sup_S |\pi_1(S) - \pi_2(S)|$ where S ranges over all measurable sets. Let the marginal distributions of transition probabilities starting from a point u be denoted P_u. Let us fix $r := 3/40$ for the remainder of this chapter. The main lemma of this section is stated below.

Lemma 1. *Let x, y be points such that $\sigma(x,y) \le \frac{3}{400\sqrt{mn}}$. Then, the total variation distance between P_x and P_y is less than $1 - \frac{13}{200} + o(1)$.*

Proof. Let us fix the convention that $\frac{dP_y}{dP_x}(x) := 0$ and $\frac{dP_y}{dP_x}(y) := +\infty$. If $x \to w$ is one step of the Dikin walk,

$$d(P_x, P_y) = 1 - \mathbb{E}_w \left[\min\left(1, \frac{dP_y}{dP_x}(w)\right)\right].$$

It follows from Lemma 2 that

$$\mathbb{E}_w \left[\min\left(1, \frac{dP_y}{dP_x}(w)\right)\right] \ge \min\left(1, \frac{\mathrm{vol} D_x}{\mathrm{vol}(D_y)}\right) \mathbb{P}\left[(y \in D_w) \wedge (w \in D_y \setminus \{x\})\right].$$

It follows from Lemma 4 that

$$\min\left(1, \frac{\mathrm{vol}(D_x)}{\mathrm{vol} D_y}\right) \mathbb{P}\left[(y \in D_w) \wedge (w \in D_y \setminus \{x\})\right] \ge \tag{1}$$

$$e^{-\frac{r}{5}} \mathbb{P}\left[(y \in D_w) \wedge (w \in D_y \setminus \{x\})\right]. \tag{2}$$

Let E_x denote the event that

$$0 < \max\left(\|x-w\|_w^2, \|x-w\|_x^2\right) \le r^2\left(1 - \frac{1}{n}\right),$$

E_y denote the event that $\max\left(\|y-w\|_w, \|y-w\|_y\right) \le r$ and E_{vol} denote the event that $\mathrm{vol}(D_w) \ge e^{4r}\,\mathrm{vol}(D_x)$. The complement of an event E shall be denoted $\overline{E}$.

The probability of E_y when $x \to w$ is a transition of Dikin walk can be bounded from below by $\left(\frac{e^{-4r}}{2}\right) \mathbb{P}\left[E_y \wedge E_x \wedge \overline{E_{vol}}\right]$ where w is chosen uniformly at random from D_x. It thus suffices to find a lower bound for $\mathbb{P}\left[E_y \wedge E_x \wedge \overline{E_{vol}}\right]$ where w is chosen uniformly at random from D_x, which we proceed to do. Let $\mathrm{erf}(x)$ denote the well known error function $\frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ and $\mathrm{erfc}(x) := 1 - \mathrm{erf}(x)$.

$$\mathbb{P}\left[E_y \wedge E_x \wedge \overline{E_{vol}}\right] \ge \tag{3}$$

$$\mathbb{P}\left[E_y \wedge E_x\right] - \mathbb{P}\left[E_{vol}\right]. \tag{4}$$

Lemma 3 implies that $\mathbb{P}\left[E_{vol}\right] \le \frac{\mathrm{erfc}(2)}{2} + o(1)$. Let E_x^1 be the event that

$$\|x-w\|_x^2 \le r^2\left(1 - \frac{1}{n}\right).$$

As a consequence of Lemma 5,

$$\begin{aligned} \mathbb{P}\left[E_x\right] + o(1) \;&\ge\; \left(\frac{1 - 3\sqrt{2}r}{2}\right) \mathbb{P}\left[E_x^1\right] \\ &\ge\; \left(\frac{1 - 3\sqrt{2}r}{2\sqrt{e}}\right) - o(1). \end{aligned} \tag{5}$$

Lemma 6 and Lemma 7 together tell us that

$$\mathbb{P}\left[E_y \middle| E_x\right] \;\ge\; 1 - \left(\frac{4r^2 + \mathrm{erfc}(2) + o(1)}{1 - 3\sqrt{2}r}\right) - \left(\frac{4r^2 + \mathrm{erfc}(3/2) + o(1)}{1 - 3\sqrt{2}r}\right) \tag{6}$$

$$= 1 - \left(\frac{8r^2 + \mathrm{erfc}(2) + \mathrm{erfc}(\frac{3}{2}) + o(1)}{1 - 3\sqrt{2}r}\right). \tag{7}$$

Putting (5) and (7) together gives us that

$$\mathbb{P}\left[E_y \wedge E_x\right] = \mathbb{P}\left[E_y \middle| E_x\right] \mathbb{P}\left[E_x\right] \quad (8)$$

$$\geq \frac{1 - 3\sqrt{2}r}{2\sqrt{e}} - \left(\frac{8r^2 + \mathrm{erfc}(2) + \mathrm{erfc}(\frac{3}{2})}{2\sqrt{e}}\right) - o(1). \quad (9)$$

Putting together (2), (4) and (9), we see that if $x \to w$ is a transition of the Dikin walk,

$$\mathbb{E}_w\left[\min\left(1, \frac{dP_y}{dP_x}(w)\right)\right] \geq \frac{e^{-\frac{21r}{5}}}{4\sqrt{e}}\left(1 - (3\sqrt{2}r + 8r^2 + \mathrm{erfc}(2)(1+\sqrt{e}) + \mathrm{erfc}(\frac{3}{2}))\right) - o(1).$$

For our choice of $r = 3/40$, this evaluates to more than $\frac{13}{200} - o(1)$. □

Since Dikin ellipsoids are affine-invariant, we shall assume without loss of generality that x is the origin and the Dikin ellipsoid at x is the Euclidean unit ball of radius r. This also means that in system of coordinates, the local norm $\|\cdot\|_x = \|\cdot\|_o$ is the Euclidean norm $\|\cdot\|$ and the local inner product $\langle\cdot,\cdot\rangle_x = \langle\cdot,\cdot\rangle_o$ is the usual inner product $\langle\cdot,\cdot\rangle$. On occasion we have used $a \cdot b$ to signify $\langle a, b\rangle$.

Lemma 2. *Let $w \in \mathrm{supp}(P_x) \setminus \{x, y\}$ and $y \in D_w$ and $w \in D_y$. Then,*

$$\frac{dP_y}{dP_x}(w) \geq \min\left(1, \frac{\mathrm{vol}(D_x)}{\mathrm{vol}(D_y)}\right).$$

Proof. Under the hypothesis of the lemma,

$$\frac{dP_y}{dP_x}(w) = \frac{\min\left(\frac{1}{\mathrm{vol}(D_y)}, \frac{1}{\mathrm{vol} D_w}\right)}{\min\left(\frac{1}{\mathrm{vol}(D_x)}, \frac{1}{\mathrm{vol} D_w}\right)}$$

$$= \frac{\min\left(\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_y)}, 1\right)}{\min\left(\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_x)}, 1\right)}.$$

The above expression can be further simplified by considering two cases.

1. Suppose $\min\left(\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_y)}, 1\right) = 1$, then

$$\frac{\min\left(\frac{\mathrm{vol} D_w}{\mathrm{vol}(D_y)}, 1\right)}{\min\left(\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_x)}, 1\right)} \geq 1.$$

2. Suppose $\min\left(\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_y)}, 1\right) = \frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_y)}$, then

$$\frac{\min\left(\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_y)}, 1\right)}{\min\left(\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_x)}, 1\right)} \geq \frac{\mathrm{vol}(D_x)}{\mathrm{vol}(D_y)}.$$

Therefore,

$$\frac{dP_y}{dP_x}(w) \geq \min\left(1, \frac{\mathrm{vol}(D_x)}{\mathrm{vol} D_y}\right).$$

□

Lemma 3. *Let w be chosen uniformly at random from D_x. The probability that* $\mathrm{vol}(D_x) \leq e^{2rc}\,\mathrm{vol}(D_w)$ *is greater or equal to* $1 - \frac{\mathrm{erfc}(c)}{2} - o(1)$*, i. e.*

$$\mathbb{P}\left[\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_x)} \leq e^{2rc}\right] \quad \geq \quad 1 - \frac{\mathrm{erfc}\,(c)}{2} - o(1).$$

Proof. By Lemma 13, $\ln(\frac{1}{\mathrm{vol}(D_x)})$ is a convex function. Therefore,

$$\ln\,\mathrm{vol}(D_w) - \ln\,\mathrm{vol}(D_x) \quad \leq \quad \nabla \ln(\frac{1}{\mathrm{vol}(D_x)}) \cdot (w - x).$$

By Lemma 12, $\|\nabla \ln(\frac{1}{\mathrm{vol}(D_x)})\| \leq 2\sqrt{n}$. Therefore,

$$\nabla \ln(\frac{1}{\mathrm{vol}(D_x)}) \cdot (w - x) \leq 2r \left(\frac{\sqrt{n}\,\nabla \ln(\frac{1}{\mathrm{vol}(D_x)}) \cdot (w - x)}{\|\nabla \ln(\frac{1}{\mathrm{vol}(D_x)})\|\|w - x\|}\right)$$

As stated in Theorem 5, when the dimension $n \to \infty$,

$$\frac{\sqrt{n}\,\nabla \ln(\frac{1}{\mathrm{vol}(D_x)}) \cdot (w - x)}{\|\nabla \ln(\frac{1}{\mathrm{vol}(D_x)})\|\|w - x\|}$$

converges in distribution to a standard Gaussian random variable whose mean is 0 and variance is 1. Therefore,

$$\mathbb{P}\left[\frac{\sqrt{n}\,\nabla \ln(\frac{1}{\mathrm{vol}(D_x)}) \cdot (w - x)}{\|\nabla \ln(\frac{1}{\mathrm{vol} D_x})\|\|w - x\|} \leq c\right] \quad \geq \quad \frac{1 + \,\mathrm{erf}(c)}{2} - o(1).$$

This implies that

$$\begin{aligned}\mathbb{P}\left[\frac{\mathrm{vol}(D_w)}{\mathrm{vol}(D_x)} \leq e^{c}\right] \quad &\geq \quad \mathbb{P}\left[\nabla \ln(\frac{1}{\mathrm{vol}(D_x)}) \cdot (w - x) \leq c\right] \\ &\geq \quad \left(\frac{1 + \,\mathrm{erf}\,(\frac{c}{2r})}{2}\right) - o(1).\end{aligned}$$

□

Lemma 4.

$$\ln\left(\frac{\mathrm{vol}(D_y)}{\mathrm{vol}(D_x)}\right) \leq n\sigma(x, y).$$

Proof. Suppose $\overline{pq}$ is a chord and p, x, y, q appear in that order. By Theorem 8,

$$\begin{aligned}\ln\left(\frac{\mathrm{vol}(D_y)}{\mathrm{vol}(D_x)}\right) \quad &\leq \quad \ln\left(\frac{|p - y|^n}{|p - x|^n}\right) \\ &\leq \quad n\sigma(x, y).\end{aligned}$$

□

Lemma 5. *Let w be chosen uniformly at random from D_x. Then,*

$$\begin{aligned}\mathbb{P}\left[\|x - w\|_w^2 \leq r^2\left(1 - \frac{1}{n}\right) \Big| \|x - w\|_x^2 \leq r^2\left(1 - \frac{1}{n}\right)\right] \\ \geq \frac{1 - 3\sqrt{2}r}{2} - o(1).\end{aligned}$$

Proof. Let E_x^1 be the event that

$$\|x-w\|_x^2 \le r^2\left(1-\frac{1}{n}\right).$$

We set c to $3\sqrt{2}r$ in Lemma 8 and see that

$$\mathbb{P}\left[\|x-w\|_w^2+\|x-w\|_{2x-w}^2 \ge 2r^2\left(1-\frac{1}{n}\right)\Big|E_x^1\right] \le 3\sqrt{2}r+o(1).$$

If $\|x-w\|_w^2+\|x-w\|_{2x-w}^2 \le 2r^2\left(1-\frac{1}{n}\right)$, then either $\|x-w\|_w^2$ or $\|x-w\|_{2x-w}^2$ must be less or equal to $r^2\left(1-\frac{1}{n}\right)$. □

Lemma 6. *Let $\sigma(x,y) \le \frac{3}{400\sqrt{mn}}$. Then, if w is chosen uniformly at random from D_x,*

$$\mathbb{P}\left[\|y-w\|_y \ge r\Big|\max\left(\|x-w\|_x^2, \|x-w\|_w^2\right) \le r^2\left(1-\frac{1}{n}\right)\right] \le \frac{4r^2+\operatorname{erfc}(2)+o(1)}{1-3\sqrt{2}r}.$$

Proof. It follows from Lemma 10, after substituting 1 for η and 2 for η_1 that

$$\mathbb{P}\left[\|y-w\|_y \ge r\Big|\|x-w\|_x^2 \le r^2\left(1-\frac{1}{n}\right)\right] \le 2r^2+\frac{\operatorname{erfc}(2)}{2}+o(1).$$

This lemma follows using the upper bound from Lemma 5 for

$$\mathbb{P}\left[\|x-w\|_w^2 \le r^2\left(1-\frac{1}{n}\right)\Big|\|x-w\|_x^2 \le r^2\left(1-\frac{1}{n}\right)\right].$$

An application of Theorem 4 completes the proof. □

Lemma 7. *Suppose $\sigma(x,y) \le \frac{3}{400\sqrt{mn}}$. Let w be chosen uniformly at random from D_x. Then,*

$$\mathbb{P}\left[\|y-w\|_w \ge r\Big|\max(\|x-w\|_w^2, \|x-w\|_x^2) \le r^2\left(1-\frac{1}{n}\right)\right] \le \frac{4r^2+\operatorname{erfc}(3/2)+o(1)}{1-3\sqrt{2}r}.$$

Proof. Substituting $c=1$ in Lemma 9, we see that

$$\mathbb{P}\left[\|y-w\|_w^2-\|x-w\|_w^2 \ge \psi_1\Big|\|x-w\|_x^2 \le r^2(1-\frac{c}{n})\right] \le 2r^2+\frac{\operatorname{erfc}(3/2)}{2}+o(1).$$

This implies that

$$\mathbb{P}\left[\|y-w\|_w^2-\|x-w\|_w^2 \ge \frac{r}{n}\Big|\|x-w\|_x^2 \le r^2\left(1-\frac{1}{n}\right)\right] \le 2r^2+\frac{\operatorname{erfc}(3/2)}{2}+o(1).$$

This lemma follows using the lower bound from Lemma 5 for

$$\mathbb{P}\left[\|x-w\|_w^2 \le r^2\left(1-\frac{1}{n}\right)\Big|\|x-w\|_x^2 \le r^2\left(1-\frac{1}{n}\right)\right].$$

□

The following theorem has the geometric interpretation that the probability distribution obtained by orthogonally projecting a random vector v_n from an n-dimensional ball of radius $\sqrt{n}$ onto a line converges in distribution to the standard mean zero, variance 1, normal distribution $N[0,1]$. This was known to Poincaré, and is a fact often mentioned in the context of measure concentration phenomena, see for example [11].

Theorem 5 (Poincaré)**.** *Let v_n be any n-dimensional vector and h_n be a random vector chosen uniformly from the n-dimensional unit Euclidean ball. Then, as $n \to \infty$, $\frac{\sqrt{n}<v_n,h_n>}{\|v_n\|\|h_n\|}$ converges in distribution to a zero-mean Gaussian whose variance is 1, i. e. $N[0,1]$.*

Let

$$\psi_1 := \frac{\|y-x\|_x^2}{(1-r)^2} + \frac{(3+2\sqrt{6})r\|y-x\|_x}{\sqrt{n}}.$$

Lemma 8. *Let v be chosen uniformly at random from D_x and c be a positive constant. Then,*

$$\mathbb{P}\left[\|x-v\|_v^2 + \|x-v\|_{2x-v}^2 \ge 2r^2\left(1-\frac{(c-\frac{18r^2}{c})}{n}\right)\right] \le c + o(1).$$

Proof. Let the i^{th} constraint be $a_i^T x \le 1$ for all $i \in \{1,\dots,m\}$. Let $x-v$ be denoted h. In the present frame, for any vector v, $\|v\|_x = \|v\|$.

$$\|x-v\|_v^2 + \|x-v\|_{2x-v}^2 = \sum_i \frac{(a_i^T h)^2}{(1-a_i^T h)^2} + \sum_i \frac{(a_i^T h)^2}{(1+a_i^T h)^2} \tag{10}$$

In the present coordinate frame $\sum_i a_i a_i^T = I$. Consequently for each i,

$$\begin{aligned}\mathbb{E}[(a_i^T h)^2] &= \frac{\|a_i\|^2\mathbb{E}[\|h\|^2]}{n} &(11)\\ &\le \frac{r^2}{n}. &(12)\end{aligned}$$

$$\begin{aligned}\sum_i\left(\frac{(a_i^T h)^2}{2(1-a_i^T h)^2} + \frac{(a_i^T h)^2}{2(1+a_i^T h)^2}\right) &= \sum_i (a_i^T h)^2\left(\frac{1+(a_i^T h)^2}{(1-(a_i^T h)^2)^2}\right) &(13)\\ &= \sum_i\left((a_i^T h)^2 + \frac{3(a_i^T h)^4-(a_i^T h)^6}{(1-(a_i^T h)^2)^2}\right)\\ &= \|h\|_x^2 + \sum_i \frac{3(a_i^T h)^4-(a_i^T h)^6}{(1-(a_i^T h)^2)^2}. &(14)\end{aligned}$$

In the present coordinate frame $\sum_i a_i a_i^T = I$. Consequently for each i,

$$\mathbb{E}\left[\frac{(a_i^T h)^2}{\|a_i\|^2\|h\|^2}\right] = \frac{1}{n}. \tag{15}$$

By Theorem 5, the probability that $|a_i^T h| \geq n^{-\frac{1}{4}}$ is $O(e^{-\sqrt{n}/2})$. $|a_i^T h|$ is $\leq \|a_i^T\| r$, which is less than $\frac{1}{2}$. This allows us to write

$$\mathbb{E}\left[\frac{3(a_i^T h)^4 - (a_i^T h)^6}{(1-(a_i^T h)^2)^2}\right] = 3\mathbb{E}[(a_i^T h)^4](1+o(1)), \tag{16}$$

and so

$$\mathbb{E}\left[\sum_i \frac{3(a_i^T h)^4 - (a_i^T h)^6}{(1-(a_i^T h)^2)^2}\right] = \sum_i 3\mathbb{E}[(a_i^T h)^4](1+o(1)). \tag{17}$$

Next, we shall find an upper bound on $\mathbb{E}[\sum_i (a_i^T h)^4]$. The length of h and its direction are independent, therefore

$$\mathbb{E}\left[\sum_i (a_i^T h)^4\right] = \sum_i \|a_i\|^4 \mathbb{E}[\|h\|^4] \mathbb{E}\left[\frac{(a_i^T h)^4}{\|a_i\|^4 \|h\|^4}\right]. \tag{18}$$

A direct integration by parts tells us that if the distribution of X is $N[0,1]$, then $\mathbb{E}[X^4] = 3$. Therefore,

$$\mathbb{E}\left[\frac{(a_i^T h)^4}{\|a_i\|^4 \|h\|^4}\right] = \frac{3+o(1)}{n^2}. \tag{19}$$

$\mathbb{E}[\|h\|^4]$ is equal to $r^4(1+o(1))$ and so

$$\mathbb{E}\left[\sum_i (a_i^T h)^4\right] = \sum_i \left(\frac{3+o(1)}{n^2}\right) \|a_i\|^4 r^4. \tag{20}$$

This implies that

$$\begin{aligned}
\mathbb{E}\left[\sum_i \frac{3(a_i^T h)^4}{(1-(a_i^T h)^2)^2}\right] &= \frac{9+o(1)}{n^2} \sum_i \|a_i\|^4 r^4 && (21)\\
&\leq \frac{9+o(1)}{n^2} \sum_i \|a_i\|^2 r^4 && (22)\\
&= \frac{(9+o(1))r^4}{n}. && (23)
\end{aligned}$$

In (22), we used the fact that $\sum_i a_i a_i^T = I$ and so $\|a_i\|^2 \leq 1$ for each i. Together, Markov's inequality and (23) yield the following.

$$\begin{aligned}
\mathbb{P}\left[\sum_i \frac{3(a_i^T h)^4 - (a_i^T h)^6}{(1-(a_i^T h)^2)^2} \geq \frac{c_2 r^4}{n}\right] &\leq \mathbb{P}\left[\sum_i \frac{3(a_i^T h)^4}{(1-(a_i^T h)^2)^2} \geq \frac{c_2 r^4}{n}\right] && (24)\\
&\leq \frac{9+o(1)}{c_2}. && (25)
\end{aligned}$$

Also,

$$\begin{aligned}
\mathbb{P}[\|h\|_x^2 \geq r^2(1-\frac{c_1}{n})] &= \mathbb{P}[\|h\|_x^n \geq r^n(1-\frac{c_1}{n})^{n/2}] && (26)\\
&\leq 1 - e^{-\frac{c_1}{2}} + o(1). && (27)
\end{aligned}$$

We infer from (25) and (27) that

$$\begin{aligned}\mathbb{P}\left[\|h\|_x^2 + \sum_i \frac{3(a_i^T h)^4 - (a_i^T h)^6}{(1-(a_i^T h)^2)^2} \geq r^2(1 - \frac{c_1 - c_2 r^2}{n})\right] &\leq 1 - e^{-\frac{c_1}{2}} + \frac{9}{c_2} + o(1) \\ &\leq \frac{c_1}{2} + \frac{9}{c_2} + o(1). \quad (28)\end{aligned}$$

Setting c_1 to c and c_2 to $\frac{18}{c}$ proves the lemma. □

Let E_x^c be the event that $\|x-w\|_x^2 \leq r^2(1-\frac{c}{n})$.

Lemma 9. *Let w be a point chosen uniformly at random from D_x. Then, for any positive constant c, independent of n,*

$$\mathbb{P}\left[\|y-w\|_w^2 - \|x-w\|_w^2 \geq \psi_1 \middle| E_x^c\right]$$
$$\leq 2r^2 + \frac{\mathrm{erfc}(3/2)}{2} + o(1).$$

Proof. $\|y\|_w^2$ can be bounded above in terms of $\|y\|_o$ as follows.

$$\|y\|_w^2 \leq y^T\left(\sum_i \frac{a_i a_i^T}{(1-a_i^T w)^2}\right) y \quad (29)$$

$$\leq \left(\sup_i \frac{1}{(1-a_i^T w)^2}\right) \sum_i y^T a_i a_i^T y. \quad (30)$$

For each i, $\|a_i\| \leq 1$, therefore

$$\left(\sup_i \frac{1}{(1-a_i^T w)^2}\right) \sum_i y^T a_i a_i^T y \leq \frac{\|y\|_o^2}{(1-r)^2}. \quad (31)$$

Let E_w^c be the event that $\|w\|_o^2 \leq 1 - \frac{c}{n}$.

By Theorem 5,

$$\mathbb{P}\left[(-2\langle y,w\rangle_o) \geq \frac{2r\eta_1\|y\|_o}{\sqrt{n}} \middle| E_w^c\right] \leq \frac{1-\mathrm{erf}(\eta_1)}{2} + o(1). \quad (32)$$

$(\langle y,w\rangle_o - \langle y,w\rangle_w)^2$ can be bounded above using the Cauchy-Schwarz inequality as follows.

$$\begin{aligned}(\langle y,w\rangle_o - \langle y,w\rangle_w)^2 &= \left(w^T\left(1 - \sum_i \frac{a_i a_i^T}{(1-a_i^T w)^2}\right) y\right)^2 \\ &= \left(\sum_i \frac{w^T a_i((1-a_i^T w)^2 - 1) a_i^T y}{(1-a_i^T w)^2}\right)^2 \\ &\leq \left(\sum_i \frac{\left(w^T a_i((1-a_i^T w)^2-1)\right)^2}{(1-a_i^T w)^4}\right)\left(\sum_i (a_i^T y)^2\right).\end{aligned}$$

Let κ be a standard one-dimensional Gaussian random variable whose variance is 1 and mean is 0 (i.e. having distribution $N[0,1]$). Since $r < \frac{1}{2}$ and each $\|a_i\| = \|a_i\|_o$ is less or equal to 1, it follows from Theorem 5 that conditional on E_w^c,

$$\frac{\left(nw^T a_i((1-a_i^T w)^2 - 1)\right)^2}{4r^2\|a_i\|^2(1-a_i^T w)^4}$$

converges in distribution to the distribution of κ^4, whose expectation can be shown using integration by parts to be 3. So,

$$\begin{aligned}\mathbb{E}\left[\sum_i \frac{\left(w^T a_i((1-a_i^T w)^2-1)\right)^2}{(1-a_i^T w)^4}\Big| E_w^c\right] &\leq \sum_i \left(\frac{4}{n^2}\right)\|a_i\|_o^4 r^4(3+o(1)) \\ &\leq \left(\frac{12+o(1)}{n^2}\right) r^4 \sum_i \|a_i\|_o^2 \\ &= \frac{(12+o(1))r^4}{n}.\end{aligned}$$

Thus by Markov's inequality,

$$\mathbb{P}\left[\sum_i \frac{\left(w^T a_i((1-a_i^T w)^2-1)\right)^2}{(1-a_i^T w)^4} \geq \frac{12\eta_2 r^4}{n}\Big| E_w^c\right] \leq \frac{1+o(1)}{\eta_2}. \tag{33}$$

$\sum_i (a_i^T y)^2$ is equal to $\|y\|_o^2$. Therefore (33) implies that

$$\mathbb{P}\left[(\langle y,w\rangle_o - \langle y,w\rangle_w)^2 \geq \frac{12\eta_2 r^4\|y\|_o^2}{n}\right] \leq \frac{1+o(1)}{\eta_2}. \tag{34}$$

Putting (32) and (34) together, we see that

$$\mathbb{P}\left[-2\langle y,w\rangle_w \geq \frac{2r\eta_1\|y\|_o}{\sqrt{n}} + 2\sqrt{\frac{12\eta_2 r^4\|y\|_o^2}{n}}\,\Big| E_w^c\right] \leq \frac{1-\operatorname{erf}(\eta_1)}{2} + \frac{1+o(1)}{\eta_2} \tag{35}$$

Conditional on E_w^c, $\|w\|_w^2$ is less or equal to $r(1-\frac{c}{n})$.

Therefore, using erfc(x) to denote $1-\operatorname{erf}(x)$,

$$\mathbb{P}\left[\|y-w\|_w^2 - \|w\|_w^2 \geq \frac{\|y\|_o^2}{(1-r)^2} + \frac{2r\|y\|_o}{\sqrt{n}}\left(\eta_1 + r\sqrt{12\eta_2}\right)\Big| E_w^c\right] \leq \eta_2^{-1} + \frac{\operatorname{erfc}(\eta_1)}{2} + o(1).$$

Setting $\eta_1 = 3/2$ and $\eta_2 = \frac{1}{2r^2}$, gives

$$\mathbb{P}\left[\|y-w\|_w^2 - \|w\|_w^2 \geq \Big| E_w^c\right] \leq 2r^2 + \frac{\operatorname{erfc}(3/2)}{2} + o(1). \tag{36}$$

□

Lemma 10. *Let c be a positive constant. Let*

$$\psi_2 := \|y-x\|_y^2 + \frac{2r\eta_1\|y-x\|_x}{\sqrt{n}} + \frac{2\eta\|y-x\|_x}{\sqrt{n}}\left(\sqrt{3}r + \|y-x\|_x\right).$$

If w is a point chosen uniformly at random from D_x, for any positive constants η and η_1, Then,

$$\mathbb{P}\left[\|y-w\|_y^2 - \|x-w\|_x^2 \geq \psi_2\Big| E_w^c\right]$$

$$\leq \frac{2r^2}{\eta^2} + \frac{\operatorname{erfc}(\eta_1)}{2} + o(1).$$

Proof.

$$\begin{aligned}
\|y-w\|_y^2 &= \|y\|_y^2 + \|w\|_y^2 - 2\langle w, y\rangle_y & (37)\\
&\le \|y\|_y^2 + \|w\|_o^2 & (38)\\
&+ \sqrt{(\|w\|_y^2 - \|w\|_o^2)^2} - 2\langle w, y\rangle_o + 2\sqrt{(\langle w, y\rangle_o - \langle w, y\rangle_y)^2}. & (39)
\end{aligned}$$

We shall obtain probabilistic upper bounds on each term in (39).

$$\begin{aligned}
(\|w\|_y^2 - \|w\|_o^2)^2 &= \left(w^T\left(\sum_i a_i a_i^T\left(\frac{1-(1-a_i^Ty)^2}{(1-a_i^Ty)^2}\right)\right)w\right)^2 & (40)\\
&\le \left(\sum_i (w^Ta_i)^4\right)\left(\sum_i\left(\frac{1-(1-a_i^Ty)^2}{(1-a_i^Ty)^2}\right)^2\right) & (41)\\
&= \left(\sum_i (w^Ta_i)^4\right)\left(\sum_i 4\left(a_i^Ty\right)^2(1+o(1))\right) & (42)\\
&= (4+o(1))\,\|y\|_o^2\sum_i (w^Ta_i)^4. & (43)
\end{aligned}$$

In inferring (42) from (41) we have used the fact that $\|y\|_o$ is $O(\frac{1}{\sqrt{n}})$ which is $o(1)$. As was stated in (19) in slightly different terms,

$$\mathbb{E}\left[(w^Ta_i)^4\right] = \frac{\|a_i\|^4 r^4(3+o(1))}{n^2}.$$

Therefore by Markov's inequality, for any constant c,

$$\begin{aligned}
\mathbb{E}\left[\sum_i (w^Ta_i)^4 \Big| \|w\|_o^2 \le r^2(1-\frac{c}{n})\right] &= \sum_i \frac{\|a_i\|^4 r^4(3+o(1))}{n^2}\\
&\le \frac{r^4(3+o(1))}{n^2}\sum_i \|a_i\|^2\\
&= \frac{r^4(3+o(1))}{n}.
\end{aligned}$$

Therefore,

$$\mathbb{P}\left[(\|w\|_y^2 - \|w\|_o^2)^2 \ge \eta^2\frac{12\|y\|_o^2 r^4}{n}\right] \le \frac{1+o(1)}{\eta^2}. \qquad (44)$$

By Theorem 5, as $n \to \infty$, the distribution of $\frac{\sqrt{n}\langle w,y\rangle_o}{r\|y\|_o}$ converges in distribution to $N[0,1]$. Therefore

$$\mathbb{P}\left[(-2\langle w, y\rangle_o) \ge \frac{2\eta_1 r\|y\|_o}{\sqrt{n}}\Big| \|w\|_o^2 \le r^2(1-\frac{c}{n})\right] \le \frac{\mathrm{erfc}(\eta_1)}{2} + o(1). \qquad (45)$$

Finally, we need similar tail bounds for $(\langle w,y\rangle_o - \langle w,y\rangle_y)^2$. Note that

$$
\begin{aligned}
(\langle w,y\rangle_o - \langle w,y\rangle_y)^2 &= \left(w^T\left(\sum_i a_i a_i^T\left(\frac{1-(1-a_i^Ty)^2}{(1-a_i^Ty)^2}\right)\right)y\right)^2 && (46)\\
&\le \left(\sum_i (w^Ta_ia_i^Ty)^2\right)\left(\sum_i\left(\frac{1-(1-a_i^Ty)^2}{(1-a_i^Ty)^2}\right)^2\right) && (47)\\
&= \left(\sum_i (w^Ta_ia_i^Ty)^2\right)\left(\sum_i (4+o(1))(a_i^Ty)^2\right) && (48)\\
&= (4+o(1))\left(\sum_i (w^Ta_ia_i^Ty)^2\right)\|y\|_o^2. && (49)
\end{aligned}
$$

It suffices now to obtain a tail bound on $\sum_i (w^Ta_ia_i^Ty)^2$. By Theorem 5,

$$
\begin{aligned}
\mathbb{E}\left[\sum_i (w^Ta_ia_i^Ty)^2 \Big| \|w\|_o^2 \le r^2(1-\frac{c}{n})\right] &\le \left(\sum_i \|a_ia_i^Ty\|^2\right)\frac{r^2(1+o(1))}{n}\\
&\le \left(\sum_i (a_i^Ty)^2\right)\frac{r^2(1+o(1))}{n}\\
&\le \frac{\|y\|_o^2 r^2(1+o(1))}{n}.
\end{aligned}
$$

Therefore,

$$
\mathbb{P}\left[(\langle w,y\rangle_o - \langle w,y\rangle_y)^2 \le \frac{4\eta^2\|y\|_o^4r^2}{n}\right] \le \frac{1+o(1)}{\eta^2}. \qquad (50)
$$

Putting together (44), (45) and (50), we see that

$$
\mathbb{P}\left[\|y-w\|_y^2 - \|w\|_o^2 \ge \|y\|_y^2 + \frac{2\eta\|y\|_o}{\sqrt{n}}\left(\sqrt{3}r + \frac{r\eta_1}{\eta} + \|y\|_o\right)\Big| E_w^c\right] \le \frac{2r^2}{\eta^2} + \frac{\operatorname{erfc}(\eta_1)}{2} + o(1).
$$

□

The following is a generalization of the Cauchy-Schwarz inequality that takes values in a cone of semidefinite matrices where inequality is replaced by dominance in the semidefinite cone. It will be used to prove Lemma 12 and may be of independent interest.

Lemma 11 (Semidefinite Cauchy-Schwartz). *Let $\alpha_1,\ldots,\alpha_m$ be reals and $A_1,\ldots,A_m$ be $r\times n$ matrices. Let $B \preccurlyeq C$ signify that B is dominated by C in the semidefinite cone. Then*

$$
\left(\sum_{i=1}^m \alpha_iA_i\right)\left(\sum_{i=1}^m \alpha_iA_i\right)^T \preccurlyeq \left(\sum_{i=1}^m \alpha_i^2\right)\left(\sum_{i=1}^m A_iA_i^T\right). \qquad (51)
$$

Proof. For each i and j,

$$
0 \preccurlyeq (\alpha_jA_i - \alpha_iA_j)(\alpha_jA_i - \alpha_iA_j)^T
$$

Therefore,

$$
\begin{aligned}
0 &\preccurlyeq \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_j A_i - \alpha_i A_j)(\alpha_j A_i - \alpha_i A_j)^T \\
&= \left(\sum_{i=1}^{m}\alpha_i^2\right)\left(\sum_{i=1}^{m}A_i A_i^T\right) - \left(\sum_{i=1}^{m}\alpha_i A_i\right)\left(\sum_{i=1}^{m}\alpha_i A_i\right)^T
\end{aligned}
$$

□

We shall obtain an upper bound of $2\sqrt{n}$ on

$$\|\nabla \ln(\frac{1}{\mathrm{vol} D_x})\|\Big|_{x=o} = \|\nabla \ln \det H\|\Big|_o.$$

Lemma 12. $\|\nabla \ln \det H|_x\|_x \le 2\sqrt{n}$.

Proof. In our frame,

$$\sum a_i a_i^T = I, \tag{52}$$

where I is the $n \times n$ identity matrix, and for any vector v,

$$\|v\|_o = \|v\|. \tag{53}$$

If X is a matrix whose $\ell_2 \to \ell_2$ norm is less than 1, $\log(I + X)$ can be assigned a unique value by equating it with the power series

$$\sum_{i=1}^{\infty}(-1)^{i-1}\frac{X^i}{i}.$$

Using this formalism when y is in a small neighborhood of the identity.

$$\ln \det H(y) \quad = \quad \mathrm{trace} \ln H(y). \tag{54}$$

In order to obtain an upper bound on $\|\nabla \ln \det H\|$ at o, it suffices to uniformly bound $\left|\frac{\partial \ln \det H}{\partial h}\right|$ along all unit vectors h, since

$$\|\nabla \ln \det H\| = \sup_{\|h\|=1}\left|\frac{\partial}{\partial h}\mathrm{trace} \ln H\right|. \tag{55}$$

$$\left[\frac{\partial}{\partial h}\mathrm{trace} \ln H\right]\Bigg|_o$$

$$
\begin{aligned}
&= \lim_{\delta \to 0}\frac{\left(\mathrm{trace} \ln \left(\sum \frac{a_i a_i^T}{(1-\delta a_i^T h)^2}\right) - \ln I\right)}{\delta} && (56)\\
&= \sum_i 2(a_i^T h)(\mathrm{trace}\, a_i a_i^T) && (57)\\
&= 2\sum_i \|a_i\|^2 a_i^T h. && (58)
\end{aligned}
$$

The Semidefinite Cauchy-Schwarz inequality from Lemma 11 gives us the following.

$$(\sum_i \|a_i\|^2 a_i)(\sum_i \|a_i\|^2 a_i^T) \preccurlyeq (\sum_i \|a_i\|^4)(\sum_i a_i a_i^T) \tag{59}$$

$\sum_i a_i a_i^T = I$, so the magnitude of each vector a_i must be less or equal to 1, and $\sum_i \|a_i\|^2$ must equal n.

Therefore

$$\begin{aligned}(\sum_i \|a_i\|^4)(\sum_i a_i a_i^T) &= (\sum_i \|a_i\|^4)I & (60)\\ &\preccurlyeq (\sum_i \|a_i\|^2)I & (61)\\ &= nI & (62)\end{aligned}$$

(59) and (62) imply that

$$(\sum_i \|a_i\|^2 a_i)(\sum_i \|a_i\|^2 a_i^T) \preccurlyeq nI. \tag{63}$$

(55), (58) and (63) together imply that

$$\|\nabla \ln \det H\| \leq 2\sqrt{n}. \tag{64}$$

□

The following is due to P. Vaidya [22].

Lemma 13. $\ln \det H$ *is a convex function.*

Proof. Let $\frac{\partial}{\partial h}$ denote partial differentiation along a unit vector h. Recall that $\sum_i a_i a_i^T = I$.

$$\frac{\partial^2 \ln \det H}{(\partial h)^2}\Big|_o$$

$$\begin{aligned}
&= \lim_{\delta \to 0} \frac{1}{\delta^2} \operatorname{trace} \ln \left(\left(\sum \frac{a_i a_i^T}{(1 - \delta a_i^T h)^2} \right) \left(\sum \frac{a_i a_i^T}{(1 + \delta a_i^T h)^2} \right) \right) \\
&= \lim_{\delta \to 0} \frac{\operatorname{trace} \left(\ln \left(\sum_i a_i a_i^T (\sum_{j \geq 0} (j+1)(\delta a_i^T h)^j) \right) \right)}{\delta^2} \\
&+ \frac{\operatorname{trace} \left(\ln \left(\sum_i a_i a_i^T (\sum_{j \geq 0} (j+1)(-\delta a_i^T h)^j) \right) \right)}{\delta^2} \\
&= \lim_{\delta \to 0} \frac{\operatorname{trace} \sum_{k \geq 1} \frac{(-1)^{k-1}}{k} \left(\sum_i a_i a_i^T (\sum_{j \geq 1} (j+1)(\delta a_i^T h)^j) \right)^k}{\delta^2} \\
&+ \frac{\operatorname{trace} \sum_{k \geq 1} \frac{(-1)^{k-1}}{k} \left(\sum_i a_i a_i^T (\sum_{j \geq 1} (j+1)(-\delta a_i^T h)^j) \right)^k}{\delta^2}.
\end{aligned}$$

The only terms in the numerators of the above limit that matter are those involving δ^2. So this simplifies to

$$\begin{aligned} 2\sum_i \operatorname{trace} a_i a_i^T (a_i^T h)^2 &= 2\sum_i \|a_i\|^2 (a_i^T h)^2 \\ &\geq 2\sum_i (a_i^T h)^4 \\ &\geq \frac{2\left(\sum_i (a_i^T h)^2\right)^2}{m} \\ &= \frac{2}{m}. \end{aligned}$$

This proves the lemma. □

2.5 Conductance and mixing time

The proof of the following theorem is along the lines of Theorem 11 in [12].

Theorem 6. *Let n be greater than some universal constant. Let S_1 and $S_2 := K \setminus S_1$ be measurable subsets of K. Then,*

$$\int_{S_1} P_x(S_2) d\lambda(x) \geq \frac{6}{10^5 \sqrt{mn}} \min\left(\operatorname{vol}(S_1), \operatorname{vol}(S_2)\right).$$

Proof. Let ρ be the density of the uniform distribution on K. We shall use ρ in some places where it is seemingly unnecessary because, then, most of this proof transfers verbatim to a proof of Theorem 11 as well. For any $x \neq y \in K$,

$$\rho(y)\frac{dP_y}{d\lambda}(x) = \rho(x)\frac{dP_x}{d\lambda}(y),$$

therefore ρ is the stationary density of the Markov chain. Let $\delta = \frac{3}{400\sqrt{mn}}$ and $\epsilon = \frac{13}{200}$. Let $S_1' = S_1 \cap \{x \big| \rho(x)P_x(S_2) \leq \frac{\epsilon}{2\operatorname{vol}(K)}\}$ and $S_2' = S_2 \cap \{y \big| \rho(y)P_y(S_1) \leq \frac{\epsilon}{2\operatorname{vol}(K)}\}$. By the reversibility of the chain, which is easily checked,

$$\int_{S_1} \rho(x)P_x(S_2)d\lambda(x) = \int_{S_2} \rho(y)P_y(S_1)d\lambda(y).$$

If $x \in S_1'$ and $y \in S_2'$ then

$$\int_K \min\left(\rho(x)\frac{dP_x}{d\lambda}(w), \rho(y)\frac{dP_y}{d\lambda}(w)\right) d\lambda(w) < \frac{\epsilon}{\operatorname{vol}(K)}.$$

For sufficiently large n, Lemma 1 implies that $\sigma(S_1', S_2') \geq \delta$. Therefore Theorem 3 implies that

$$\pi(K \setminus S_1' \setminus S_2') \geq \delta\pi(S_1')\pi(S_2').$$

First suppose $\pi(S_1') \geq (1-\delta)\pi(S_1)$ and $\pi(S_2') \geq (1-\delta)\pi(S_2)$. Then,

$$\begin{aligned} \int_{S_1} P_x(S_2)d\rho(x) &\geq \frac{\epsilon\pi(K \setminus S_1' \setminus S_2')}{2} \\ &\geq \frac{\epsilon\delta\pi(S_1')\pi(S_2')}{2} \\ &\geq \left(\frac{(1-\delta)^2\epsilon\delta}{8}\right)\min(\pi(S_1), \pi(S_2)) \end{aligned}$$

and we are done. Otherwise, without loss of generality, suppose $\pi(S_1') \leq (1-\delta)\pi(S_1)$. Then

$$\int_{S_1} P_x(S_2) d\rho(x) \geq \frac{\epsilon\delta}{2}\pi(S_1)$$

and we are done. □

The following theorem was proved in [13].

Theorem 7 (Lovász-Simonovits)**.** *Let μ_0 be the initial distribution for a lazy reversible ergodic Markov chain whose conductance is Φ and stationary measure is μ, and μ_k be the distribution of the k^{th} step. Let $M := \sup_S \frac{\mu_0(S)}{\mu(S)}$ where the supremum is over all measurable subsets S of K. Then, for all such S,*

$$|\mu_k(S) - \mu(S)| \leq \sqrt{M}\left(1 - \frac{\Phi^2}{2}\right)^k.$$

We now in a position to prove the main theorem regarding Dikin walk, Theorem 1.

of Theorem 1. Let t be the time when the first proper move is made. $\mathbb{P}[t \geq t' \big| t \geq t'-1] \leq 1 - \frac{13}{200} + o(1)$ by Lemma 1 applied when $x = x_0$ and y approaches x_0. Therefore when n is sufficiently large,

$$\mathbb{P}\left[t < \frac{\ln(\frac{\epsilon}{2})}{\ln(1 - \frac{6}{100})}\right] \geq 1 - \frac{\epsilon}{2}.$$

Let μ_k be the distribution of x_k and μ be the stationary distribution, which is uniform. Let ρ_k and ρ likewise be the density of μ_k and $\rho = \frac{1}{\text{vol}(K)}$ the density of the uniform distribution. We shall now find an upper bound for $\frac{\rho_{k+t}}{\rho}$. For any $x \in K$, $\rho_t(x) \geq \frac{100}{6\,\text{vol}(D_x)}$ by Lemma 1, applied when $x = x_0$ and y approaches x_0. By (2) in Fact 1 $\frac{\text{vol}(D_x)}{\text{vol}(K)} \geq \left(\frac{r}{\sqrt{2m}\,s}\right)^n$, which implies that

$$\sup_{S \subseteq K} \frac{\mu_t(S)}{\mu(S)} = \sup_{x \in K} \frac{\rho_t(x)}{\rho} \quad (65)$$

$$\leq \left(\frac{\sqrt{2m}\,s}{r}\right)^n \left(\frac{100}{6}\right). \quad (66)$$

The theorem follows by plugging in Equation 66 and the lower bound on the conductance of Dikin walk given by Theorem 6 into Theorem 7. □

3 Affine algorithm for linear programming

We shall consider problems of the following form. Given a system of inequalities $By \leq \mathbf{1}$, a linear objective c such that the polytope

$$Q := \{y : By \leq \mathbf{1} \text{ and } |c^T y| \leq 1\}$$

is bounded, and $\epsilon, \delta > 0$ the algorithm is required to do the following.

- If $\exists y$ such that $By \leq \mathbf{1}$ and $c^T y \geq 1$, output y such that $By \leq \mathbf{1}$ and $c^T y \geq 1 - \epsilon$ with probability greater than $1 - \delta$.

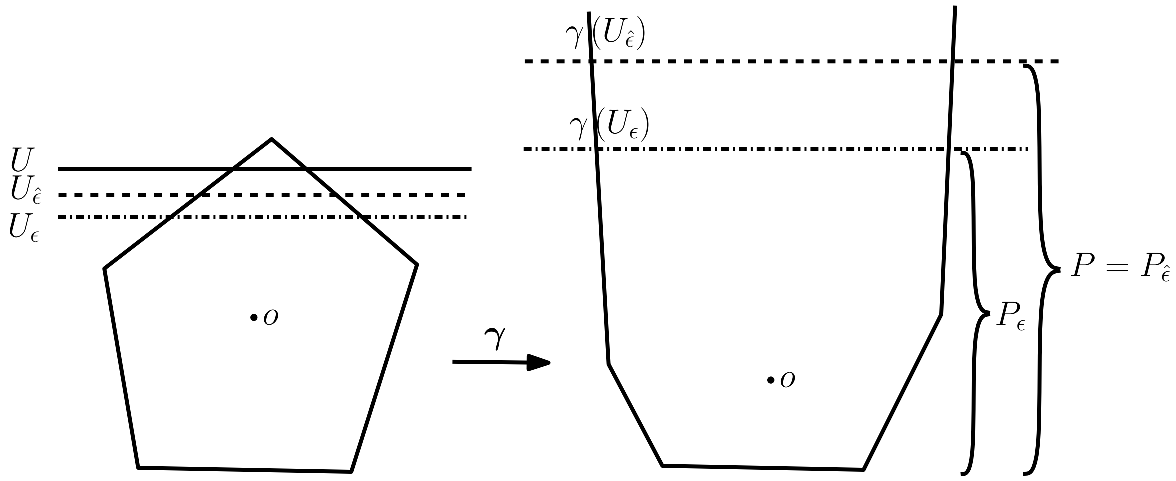

Figure 2: The effect of the projective transformation γ.

Any linear program can be converted to such a form, either by the sliding objective method or by combining the primal and dual problems and using the duality gap added to an appropriate slack variable as the new objective (see [10] and references therein). Before the iterative stage of the algorithm which is purely affine, we need to transform the problem using a projective transformation. Let $s \geq \sup_{y \in Q} \|By\| + 1$, and

$$\tau := \left\lceil 4 \times 10^8 \times mn \left(n \ln \left(\frac{4ms^2}{\epsilon^2} \right) + 2 \ln \left(\frac{2}{\delta} \right) \right) \right\rceil. \tag{67}$$

Let γ be the projective transformation $\gamma : y \mapsto \frac{y}{1 - c^T y}$, and γ^{-1} the inverse map, $\gamma^{-1} : x \mapsto \frac{x}{1 + c^T x}$. For any $\epsilon' > 0$, let $Q_{\epsilon'} := Q \cap \{y \big| c^T y \leq 1 - \epsilon'\}$ and $U_{\epsilon'}$ be the hyperplane $\{y \big| c^T y = 1 - \epsilon'\}$. Let $\hat\epsilon = \frac{\epsilon\delta}{4n}$ and $K_\epsilon := \gamma(Q_\epsilon)$. Let $K := K_{\hat\epsilon} = \gamma\left(Q_{\hat\epsilon}\right)$. For $x \in K$, let D_x denote the Dikin ellipsoid (with respect to K) of radius $r := \frac{3}{40}$, centered at x.

4 Algorithm

---

1. Choose x_0 uniformly at random from $r^{-1}D_o$, where o is the origin.
2. While $i < \tau$ and $c^T\gamma^{-1}(x_i) < 1 - \epsilon$, choose x_{i+1} using the rule below.
   (a) Flip an unbiased coin. If `Heads`, set x_{i+1} to x_i.
   (b) If `Tails` pick a random point y from D_{x_i}.
   (c) If $x_i \notin D_y$, then reject y and set x_{i+1} to x_i; if $x_i \in D_y$, then set x_{i+1} to y.
3. If $c^T\gamma^{-1}(x_\tau) \geq 1 - \epsilon$ output $\gamma^{-1}(x_\tau)$, otherwise declare that there is no y such that $By \leq \mathbf{1}$ and $c^Ty \geq 1$.

---

5 Analysis

For any bounded $f : K \to \mathbb{R}$, we define

$$\|f\|_2 := \sqrt{\int_K f(x)^2 \rho(x) d\lambda(x)}$$

where $\rho(x) = \frac{\mathrm{vol}(D_x)}{\int_K \mathrm{vol}(D_x)d\lambda(x)}$. The following lemma shows that cross ratio is a projective invariant.

Lemma 14. *Let $\gamma : \mathbb{R}^n \to \mathbb{R}^n$ be a projective transformation. Then, for any 4 collinear points a, b, c and d, $(a : b : c : d) = (\gamma(a) : \gamma(b) : \gamma(c) : \gamma(d))$.*

Proof. Let $\{e_1, \dots, e_n\}$ be a basis for $\mathbb{R}^n$. Without loss of generality, suppose that $a, b, c, d \in \mathbb{R}e_1$. γ can be factorized as $\gamma = \gamma_2 \circ \gamma_1$ where $\gamma_1 : \mathbb{R}^n \to \mathbb{R}^n$ is a projective transformation and maps $\mathbb{R}e_1$ to $\mathbb{R}e_1$ and $\gamma_2 : \mathbb{R}^n \to \mathbb{R}^n$ is an affine transformation. Affine transformations clearly preserve the cross ratio, so the problem reduces to showing that $(a : b : c : d) = (\gamma_1(a) : \gamma_1(b) : \gamma_1(c) : \gamma_1(d))$, which is a 1-dimensional question. In 1-dimension, the group of projective transformations is generated by translations ($x \mapsto x + \beta$), scalar multiplication ($x \mapsto \alpha x$) and inversion ($x \mapsto x^{-1}$), where $\alpha, \beta \in \mathbb{R} \setminus \{0\}$. In each of these cases the equality is easily checked. □

The following was proved in a more general context by Nesterov and Todd in Theorem 4.1, [19].

Theorem 8 (Nesterov-Todd)**.** *Let $\overline{pq}$ be a chord of K and x, y be interior points on it so that p, x, y, q are in order. Then $z \in D_y$ implies that $p + \frac{|p-x|}{|p-y|}(z - p) \in D_x$.*

The following theorem is from [13].

Theorem 9 (Lovász-Simonovits)**.** *Let M be a lazy reversible ergodic Markov chain on $K \subseteq \mathbb{R}^n$ with conductance Φ, whose stationary distribution is μ. For every bounded f, let $\|f\|_{2,\mu}$ denote $\sqrt{\int_K f(x)^2 d\mu(x)}$. For any fixed f, let Mf be the function that takes x to $\int_K f(y)dP_x(y)$. Then if $\int_K f(x)d\mu(x) = 0$,*

$$\|M^k f\|_{2,\mu} \leq \left(1 - \frac{\Phi^2}{2}\right)^k \|f\|_{2,\mu}.$$

We shall now prove the main theorem regarding Algorithm Dikin , Theorem 2.

of Theorem 2. Let $\overline{pq}$ be a chord of the polytope K_ϵ containing the origin o such that $c^T(\gamma^{-1}(p)) \geq c^T(\gamma^{-1}(q))$. Let $p' = \gamma^{-1}(p)$, $q' = \gamma^{-1}(q)$ and r' be the intersection of the chord $\overline{p'q'}$ with the hyperplane $U := \{y|c^T y = 1\}$. Then, $\frac{|q-o|}{|p-o|} \leq \frac{|q'-o|}{|p'-o|} \leq s$. $\frac{|p-o|}{|q-o|}$ is equal to $|(\infty : o : q : p)|$. By Lemma 14, the cross ratio is a projective invariant. Therefore,

$$\frac{|p-o|}{|q-o|} = \left(\frac{|p'-o|}{|p'-r'|}\right)\left(\frac{|r'-q'|}{|q'-o|}\right) \tag{68}$$

$$\leq \left(\frac{1}{\epsilon}\right)(s). \tag{69}$$

Therefore, for any chord $\overline{pq}$ of K_ϵ through o, $\frac{|p|}{|q|} \leq \frac{s}{\epsilon}$.

Let $D = \int_K \text{vol}(D_y) d\lambda(y)$. Let

$$\rho_o(x) = \begin{cases} \frac{1}{\text{vol}(D_o)}, & x \in D_o; \\ 0, & \text{otherwise}, \end{cases}$$

be the density of x_o and likewise ρ_τ be the density of the distribution of x_τ. Let $f_0(x) = \frac{\rho_0(x)}{\rho(x)}$ and $f_\tau(x) = \frac{\rho_\tau(x)}{\rho(x)}$.

$$\begin{aligned} \|f_0\|_2^2 &= \int_{D_o} \left(\frac{\rho_0(x)}{\rho(x)}\right)^2 \rho(x) d\lambda(x) \\ &\leq \frac{D}{\text{vol}(D_o) \inf_{x \in D_o} \text{vol}(D_x)} \end{aligned}$$

By Fact 1 and the fact that the Dikin ellipsoid of radius r with respect to K_ϵ is contained in the Dikin ellipsoid of the same radius with respect to K, $\sqrt{2m} D_o \supseteq Sym_o(K_\epsilon)$. (69) implies that $Sym_o(K_\epsilon) \supseteq \left(\frac{\epsilon}{s}\right) K_\epsilon$. We see from Theorem 8 that $\inf_{x \in D_o} \text{vol}(D_x) \geq \text{vol}((1-r)D_o)$. Therefore,

$$\begin{aligned} \|f_0\|_2^2 &\leq \frac{D}{\text{vol}(D_o) \inf_{x \in D_o} \text{vol}(D_x)} \\ &\leq \left(\frac{2m(\frac{s}{\epsilon})^2}{1-r}\right)^n \left(\frac{D}{\int_{K_\epsilon} \text{vol}(D_y) d\lambda(y)}\right) \\ &= \left(\frac{2m(\frac{s}{\epsilon})^2}{1-r}\right)^n \left(\frac{1}{\pi(K_\epsilon)}\right), \end{aligned} \tag{70}$$

where π is the stationary distribution. For a line $\ell \perp U$, let π_ℓ and ρ_ℓ be interpreted as the induced measure and density respectively. Let ℓ intersect the facet of K that belongs to $U_{\hat{\epsilon}}$ at u. Then by Theorem 8, for any $x, y \in \ell \cap K$ such that $|x-u| > |y-u|$, $\frac{\rho_\ell(x)}{|u-x|^n} \leq \frac{\rho_\ell(y)}{|u-y|^n}$. By integrating over such 1-dimensional fibres ℓ perpendicular to U, we see that

$$\begin{aligned} \pi(K_\epsilon) &= \frac{\int_{\ell \perp U} \pi_\ell(\ell \cap K_\epsilon) du}{\int_{\ell \perp U} \pi_\ell(\ell) du} \\ &\leq \sup_{\ell \perp U} \frac{\pi_\ell(\ell \cap K_\epsilon)}{\pi_\ell(\ell)} \\ &\leq \left(\frac{(1-1/\hat{\epsilon})^{n+1} - (1/\epsilon - 1/\hat{\epsilon})^{n+1}}{(1/\epsilon - 1/\hat{\epsilon})^{n+1}}\right) \\ &\lesssim \exp\left(\frac{\delta}{4}\right) - 1 \quad \text{as } n \to \infty. \end{aligned} \tag{71}$$

The relationship between conductance Φ and decay of the $\mathcal{L}_2$ norm from Theorem 9 tells us that

$$\begin{aligned}
\|f_\tau - \mathbb{E}_\rho f_\tau\|_2^2 &\leq \|f_0 - \mathbb{E}_\rho f_0\|_2^2 \, e^{-\tau\Phi^2} \\
&= \left(\|f_0\|_2^2 - \|(\mathbb{E}_\rho f_0)\mathbf{1}\|_2^2\right) \, e^{-\tau\Phi^2} \\
&\leq \left(\frac{2m(\frac{s}{\epsilon})^2}{1-r}\right)^n \left(\frac{e^{-\tau\Phi^2}}{\pi(K_\epsilon)}\right) \text{ (from (70))}
\end{aligned}$$

which is less than $\frac{\delta^2}{4\pi(K_\epsilon)}$, when we substitute Φ from Theorem 11 and the value of τ from (67).

$$\begin{aligned}
\frac{\delta^2}{4\pi(K_\epsilon)} &\geq \int_{K_\epsilon} (f_\tau(x) - \mathbb{E}_\rho f_\tau)^2 \rho(x) d\lambda(x) \\
&\geq \frac{\left(\int_{K_\epsilon} (f_\tau(x) - \mathbb{E}_\rho f_\tau)\rho(x) d\lambda(x)\right)^2}{\int_{K_\epsilon} \rho(x) d\lambda(x)} \\
&= \frac{(\mathbb{P}[x_\tau \in K_\epsilon] - \pi(K_\epsilon))^2}{\pi(K_\epsilon)}.
\end{aligned}$$

which together with (71) implies that $\mathbb{P}[x_\tau \in K_\epsilon] \lesssim \delta$ and completes the proof. □

The following generalization of Theorem 3 was proved in [16].

Theorem 10 (Lovász-Vempala). *Let S_1 and S_2 be measurable subsets of K and μ a measure supported on K that possesses a density whose logarithm is concave. Then,*

$$\mu(K \setminus S_1 \setminus S_2)\mu(K) \geq \sigma(S_1, S_2)\mu(S_1)\mu(S_2).$$

The proof of the following lemma is along the lines of Lemma 1 and is provided below.

Lemma 15. *Let x, y be points such that $\sigma(x, y) \leq \frac{3}{400\sqrt{mn}}$. Then, the overlap*

$$\int_{\mathbb{R}^n} \min\left(\text{vol}(D_x)P_x, \, \text{vol}(D_y)P_y\right) d\lambda(x)$$

between $\text{vol}(D_x)P_x$ *and* $\text{vol}(D_y)P_y$ *in algorithm* Dikin *is greater than* $(\frac{9}{100} - o(1))\,\text{vol}(D_x)$.

Proof. If $x \to w$ is one step of Dikin ,

$$\int_{\mathbb{R}^n} \min\left(\text{vol}(D_x)P_x, \, \text{vol}(D_y)P_y\right) d\lambda(x) = \mathbb{E}_w\left[\min\left(\text{vol}(D_x), \, \text{vol}(D_y)\frac{dP_y}{dP_x}(w)\right)\right].$$

$$\mathbb{E}_w\left[\min\left(\text{vol}(D_x), \, \text{vol}(D_y)\frac{dP_y}{dP_x}(w)\right)\right] = \text{vol}(D_x)\mathbb{P}\left[(y \in D_w) \wedge (w \in D_y \setminus \{x\})\right].$$

Let E_x denote the event that
$0 < \max\left(\|x - w\|_w^2, \|x - w\|_x^2\right) \leq r^2\left(1 - \frac{1}{n}\right)$ and
E_y denote the event that $\max\left(\|y - w\|_w, \|y - w\|_y\right) \leq r$. The probability of E_y when $x \to w$ is a

transition of Dikin is greater or equal to $\frac{\mathbb{P}[E_y \wedge E_x]}{2}$ when w is chosen uniformly at random from D_x. Thus, using Lemmas 5, 6 and 7,

$$\int_{\mathbb{R}^n} \min\left(\,\mathrm{vol}(D_x)P_x,\ \mathrm{vol}(D_y)P_y\right) d\lambda(x) \geq$$

$$\mathrm{vol}(D_x)\frac{\mathbb{P}\left[E_y \middle| E_x\right]\mathbb{P}\left[E_x\right]}{2} \geq$$

$$\frac{\mathrm{vol}(D_x)(1 - 3\sqrt{2}r - 8r^2 - \ \mathrm{erfc}(2) - \ \mathrm{erfc}(\frac{3}{2}) - o(1))}{4\sqrt{e}}.$$

When $r = 3/40$, this evaluates to more than
$\mathrm{vol}(D_x)(\frac{9}{100} - o(1))$. □

The proof of the following theorem closely follows that of Theorem 4.

Theorem 11. *If K is a bounded polytope, the conductance of the Markov chain in Algorithm Dikin is bounded below by $\frac{8}{10^5\sqrt{mn}}$.*

Proof. For any $x \neq y \in K$, $\mathrm{vol}(D_y)\frac{dP_y}{d\lambda}(x) =$
$\mathrm{vol}(D_x)\frac{dP_x}{d\lambda}(y)$, and therefore

$$\rho(x) := \frac{\mathrm{vol}(D_x)}{\int_K \mathrm{vol}(D_x) d\lambda(x)}$$

is the stationary density. Let $\delta = \frac{3}{400\sqrt{mn}}$ and $\epsilon = \frac{9}{100}$. Theorem 10 is applicable in our situation because by Lemma 13, the stationary density ρ is log-concave. The proof of Theorem 4 now applies verbatim apart from using Lemma 15 instead of Lemma 1, and Theorem 10 instead of Theorem 3. This gives us

$$\int_{S_1} P_x(S_2) d\rho(x) \quad \geq \quad \left(\frac{(1-\delta)^2\epsilon\delta}{8}\right)\min(\pi(S_1), \pi(S_2)).$$

Thus we are done. □

References

[1] J. Abernethy, E. Hazan and A. Rakhlin, " Competing in the Dark: An Efficient Algorithm for Bandit Linear Optimization," *21st Annual Conference on Learning Theory, (COLT)*, 2008 263-274

[2] A. Belloni and R. Freund, "Projective re-normalization for improving the behavior of a homogeneous conic linear system," *Mathematical Programming, Volume 118, Number 2*

[3] W. Baur and V. Strassen, "The Complexity of Partial Derivatives," Theoretical Computer Science, 22 (1983) 317-330

[4] P. Diaconis and B. Efron, "Testing for independence in a two-way table: new interpretations of the chi-square statistic," *Annals of Statistics* **13**, pp. 845–913, 1995.

[5] I. I. Dikin, "Iterative solutions of problems of linear and quadratic programming," *Soviet Math. Dokl* **8** (1967)

[6] M. Dyer, A. Frieze and R. Kannan, "A random polynomial-time algorithm for approximating the volume of convex bodies," *Journal of the ACM*, **1** (1991)

[7] R. Kannan, L. Lovász and M. Simonovits, "Random walks and an $O^*(n^5)$ volume algorithm for convex bodies." *Random Structures and Algorithms*, Volume 11, Issue 1, (August 1997)

[8] R. Kannan, S. Vempala, "Sampling Lattice points," *Proceedings of the ACM Symposium on Theory of Computing,* (1997)

[9] N. K. Karmarkar, "A new polynomial-time algorithm for linear programming," Combinatorica **4** (1984)

[10] M. Kojima and Y. Ye, "Recovering optimal dual solutions in Karmarkar's polynomial algorithm for linear programming," *Mathematical Programming,* **39** (1987) 305-317

[11] M. Ledoux, "The concentration of measure phenomenon," *Mathematical Surveys and Monographs* **89**, American Mathematical Society (2001).

[12] L. Lovász, "Hit-and-run mixes fast," *Math. Programming, series A,* **86** (1999).

[13] L. Lovász and M. Simonovits, "Random walks in a convex body and an improved volume algorithm," *Random structures and algorithms,* **4** (1993)

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

[15] L. Lovász and S. Vempala, "Fast Algorithms for Logconcave Functions: Sampling, Rounding, Integration and Optimization," *Proc. of the 47th IEEE Symposium on Foundations of Computer Science* (FOCS '06), 2006.

[16] L. Lovász and S. Vempala, "The geometry of logconcave functions and sampling algorithms," Random Structures and Algorithms, **3** (May 2007)

[17] B. Morris, "Improved bounds for sampling contingency tables," *Random Structures and Algorithms*, Volume 21, Issue 2, (Sep 2002)

[18] Y. E. Nesterov and A. S. Nemirovskii, "Interior point polynomial algorithms in convex programming," *SIAM Publications. SIAM, Philadelphia*, USA, (1994)

[19] Y. E. Nesterov and M. J. Todd, "Self-Scaled Barriers and Interior-Point Methods for Convex Programming," *Mathematics of Operations Research,* Vol. 22, No. 1, (Feb 1997), pp. 1-42

[20] J. Renegar, "A polynomial-time algorithm, based on Newton's method, for linear programming," *Mathematical Programming* **40** (1988) 59-93.

[21] P. Vaidya, "A new algorithm for minimizing convex functions over convex sets," *Mathematical Programming* **73** 291-341 (1996)

[22] P. Vaidya, "An algorithm for linear programming which requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations," *Mathematical Programming,* **47** (1990) 175-201.

[23] R. J. Vanderbei and J. C. Lagarias, "I. I. Dikin's convergence result for the affine-scaling algorithm," *Contemporary mathematics,* **114** (1990)