RANDOMIZED INTERIOR POINT METHODS FOR SAMPLING AND OPTIMIZATION

By Hariharan Narayanan

University of Washington

We present a Markov chain, "Dikin walk," for sampling from a convex body equipped with a self-concordant barrier. This Markov chain corresponds to a natural random walk with respect to a Riemannian metric defined using the Hessian of the barrier function.

For every convex set of dimension n, there exists a self-concordant barrier whose self-concordance parameter is $O(n)$. Consequently, a rapidly mixing Markov chain of the kind we describe can be defined (but not always be efficiently implemented) on any convex set. We use these results to design an algorithm consisting of a single random walk for optimizing a linear function on a convex set. Using results of Barthe [*Geom. Funct. Anal.* **12** (2002) 32–55] and Bobkov and Houdré [*Ann. Probab.* **25** (1997) 184–205], on the isoperimetry of products of weighted Riemannian manifolds, we obtain sharper upper bounds on the mixing time of a Dikin walk on products of convex sets than the bounds obtained from a direct application of the localization lemma. The results in this paper generalize previous results of Kannan and Narayanan [In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing* (2009) 561–570 ACM] from polytopes to spectrahedra and beyond, and improve upon those results in a special case when the convex set is a direct product of lower-dimensional convex sets. This Markov chain like the chain described in [In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing* (2009) 561–570 ACM] is affine-invariant.

1. Introduction. The task of sampling from distributions supported in high-dimensional Euclidean space arises frequently in statistics. In particular, the question of sampling from a nearly uniform distribution supported on high-dimensional convex set arises naturally in the task of computing the volume [6] of the latter. Markov chains that are rapidly mixing are a tool for sampling. The usual strategy for sampling from distributions supported on high-dimensional Euclidean space is to design a rapidly mixing Markov chain whose stationary distribution is the desired distribution, run it for sufficiently long and then pick the final point as a sample.

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 which when supplied with a point in $\mathbb{R}^n$ answers "yes" if the point is in K and "no" otherwise. If these balls are not provided, the convex set could be either extremely small or extremely large, making it impossible to sample the convex set in any fixed time frame.

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. A long series of works directed to improving the complexity culminated in the work of Lovász and Vempala achieve a mixing time of $O^*(n^4)$ in [19]. Another random walk for sampling convex sets, known as the ball walk, does the following. Suppose the current point is x_i. A point 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 y; otherwise $x_{i+1} = x_i$. After many successive improvements over several papers, it was shown in [11] that if $\delta < \frac{r}{\sqrt{n}}$ a ball walk mixes in $O^*(n\frac{R^2}{\delta^2})$ steps from a warm start (i.e., a distribution whose density with respect to the uniform measure is bounded above by a universal constant). A third random walk analyzed more recently by Lovász and Vempala is known as Hit-and-Run [17, 20]. This walk mixes in $O(n^3(\frac{R}{r})^2 \ln \frac{R}{d\varepsilon})$ steps from a point at a distance d from the boundary [20], where ε is the desired variation distance to stationarity.

There has been previous work converting sampling algorithms into algorithms for optimization, notably the work of Bertsimas and Vempala [4, 33]. Their algorithm can be viewed as a randomized analogue of Vaidya's algorithm [32], wherein the role of the analytic center is played by the center of mass. In our work, we instead borrow the central idea of Karmarkar's algorithm, namely projective rescaling, to obtain a randomized algorithm for convex optimization. We need to perform this step just once, although affine rescaling occurs at every step. Generalizing prior work with Kannan on polytopes [12], in Theorem 1 we present a Markov chain for sampling from a convex body defined by logarithmic, hyperbolic and self-concordant barriers (see Section 3). When restricted to the case of polytopes, the bounds we present imply, up to universal constants, the bounds in [12]. In our case, the "Dikin random walk" uses Gaussian steps having ellipsoidal covariance to make its individual steps, and is described in Section 5.1.

There is also work on sampling lattice points in polytopes based on sampling real points from the normalized Lebesgue measure in a polytope [13].

We believe that in cases where the solution is NP-hard, for example, integer programming, randomized algorithms hold promise. While we are not aware of any advantage that our method possesses for the task of minimizing a linear form over a convex set over using an interior point algorithm associated with the barrier, we believe in the utility of our randomized optimization scheme for integer programming. Huang and Mehrotra [10] have used a variety of geometric random walks for integer programming, including our random walk. The Dikin random walk (and a

long-step variant of it) were implemented alongside the so-called ball walk and Hit-and-Run random walks by the authors of [10]. In several cases, our methods are competitive. In their approach, instead of rounding the solution of an LP, they round the result of a biased random walk. Although deterministic algorithms perform better for solving LPs, it seems conceivable that the randomized algorithms produce better starting points for heuristics for solving integer programs. Another application of geometric random walks that we have not discussed is volume computation. The results of Huang and Mehrotra suggest that integer programming, or more generally, nonconvex optimization are areas that could benefit from the use of geometric random walks.

In joint work with Alexander Rakhlin, we used the Dikin walk to obtain a *low regret* algorithm for online convex optimization (see [23]). A property of the Dikin walk that was crucially used in that paper was that the bound on the conductance is "scale-free," when a Metropolis rule is overlayed for sampling from exponential distributions. This feature is not known to be shared by other random walks on convex sets, although the conductance of Hit-and-Run is known to be scale-free by the work of Lovász and Vempala [20] when sampling from the *uniform* distribution.

A self-concordant barrier F (see [31]) on a convex set K, is a convex function whose domain is the interior of K, that tends to infinity as one approaches its boundary, and whose second derivative at a point along any unit direction is large in a suitable sense compared to its first and third derivatives along the same vector. In order to convey the basic idea of Theorem 1, let $D^2 F(x)$ be the Hessian matrix of F at x. We define the transition measure P_x corresponding to x to be roughly a Gaussian whose covariance matrix is a fixed multiple of $(D^2 F(x))^{-1}$. The properties of the barrier function cause the random walk to avoid the boundary, but at the same time cause it to take relatively large steps. For example, let K be the 2-dimensional Euclidean ball $\{x : \|x\| \leq R\}$ and $F(x) := -\ln(R^2 - \|x\|^2)$ be a self-concordant barrier for it. Then, for $x \in K$, up to constants, the expected magnitude of the component of a step in the radial direction is $R - \|x\|$, while the expected magnitude of the component in the transverse direction is roughly $\sqrt{R^2 - \|x\|^2}$. It can be shown that the mixing time from 0 is independent of the diameter $2R$, and that the size and typical orientation of a step vary according to the local geometry.

In Theorem 3, we use this random walk to design an algorithm for optimization, which essentially consists of doing such a random walk on a projectively transformed version of K. This transformation preferentially dilates regions corresponding to a larger objective value, causing them to occupy more space, and hence become the target of a random walk. In the case of polytopes, a slightly different version of this appeared in [12]. The Markov chain considered in [12] was ergodic, while the one we use here is not. The analysis of the nonergodic Markov chain hinges upon the fact that it can be viewed as a limit of ergodic Markov

chains. Suppose K is defined by a semi-definite constraint of rank at most ν_h and at most m additional linear constraints, that is,

$$K := \{\mathbf{x}|\mathbf{A}\mathbf{x} \leq \mathbf{b}\} \cap \left\{\mathbf{x}\,\middle|\, \sum_i x_i \mathbf{A_i} \preceq \mathbf{B}\right\},$$

where $\mathbf{x} := (x_1, \ldots, x_n)^T$, $\mathbf{A}$ is an $m \times n$ matrix, and each $\mathbf{A_i}$ is a $\nu_h \times \nu_h$ symmetric positive definite matrix, as is $\mathbf{B}$. The symbol $\preceq$ indicates domination in the symmetric positive definite cone. Our results specialize to the following statement.

Let $s \geq \frac{|p|}{|q|}$ for any chord $\overline{pq}$ of K passing through a point $x \in K$. Let ρ_t stand for the density of the tth point x_t of the Dikin walk. Then, after

$$t = O\left(n(m + n\nu_h)\left(n \ln((m + n\nu_h)s) + \ln\frac{1}{\varepsilon}\right)\right)$$

steps are taken by a Dikin walk starting at x, the total variation distance and the $\mathcal{L}_2$ distance of the density $\rho_t(\cdot)$ of the point to the uniform density are less than ε.

One technical contribution of this paper is a family of lower bounds (see Theorem 8) for the isoperimetric constants of weighted Riemannian manifolds on which interior point methods perform a kind of steepest descent. While the results in [12] used the isoperimetric properties of the "Hilbert metric," we use a Riemannian metric. It is not clear whether the Hilbert metric can be used to obtain the results in this paper, because we consider convex sets defined by a combination of three barriers of increasingly general type—logarithmic, hyperbolic and self-concordant. These barriers are defined in Section 3. The logarithmic, hyperbolic and self-concordant barriers give rise to metrics possessing progressively weaker regularity properties. The net metric that we are (seemingly) forced to use is a weighted sum of the three metrics, where the weights take care of this progressively weakening regularity. It is unclear how the somewhat disparate issues that arise from the use of a combination of these different kinds of metrics could be dealt with in a unified way if one directly used the Hilbert metric. The proof of our mixing bounds follows the broad outline of most results on mixing in geometric random walks. We first prove a purely geometric isoperimetric inequality for a "Dikin manifold." We then obtain a bound on the probabilistic distance between the one-step densities starting from two nearby points.

We end this section with an outline of the paper. Section 2 describes past work on sampling convex sets. Section 3 describes self-concordant barriers, which are needed to define the Dikin walk. Section 4 defines the oracle model that will form the framework of our algorithm. Section 5 describes the three main theorems: Theorems 1, 2 and 3. Section 6.1 introduces a "Hessian metric" based on a self-concordant barrier. Section 7 describes the proof of Theorem 1. Section 8 describes the proof of Theorem 2. Section 9 describes the proof of Theorem 3. We have provided the proofs of lemmas in the Appendix.

2. Comparison with related work.
We begin with some definitions that are used in related work.

DEFINITION 1 (Warm start). Let ρ_0 be a starting distribution for a Markov chain $x_0, x_1, \ldots$ whose stationary distribution is the uniform distribution μ on a convex set K. We say ρ_0 is a warm start for the chain if there is a fixed universal constant C such that

$$\sup_A \frac{\rho_0(A)}{\mu(A)} < C,$$

where the supremum is taken over all measurable subsets A.

DEFINITION 2 (Mixing time). Let ρ_0 be a starting distribution for a Markov chain $x_0, x_1, \ldots$ whose stationary distribution is the uniform distribution μ on a convex set K. Let ρ_t be the distribution of the tth point of the Dikin walk. We define the mixing time in total variation distance $T_{\text{TV}}(\varepsilon, \rho_0)$ to be the smallest t such that the total variation distance of $\rho_t(\cdot)$ to the uniform distribution on K is less than ε.

For every bounded f, let $\|f\|_{2,\mu}$ denote $\sqrt{\int_K f(x)^2 \, d\mu(x)}$. Let $f_t(x) := \rho_t(x) - (\text{vol}(K))^{-1}$. We define the mixing time in $\mathcal{L}_2$ distance $T_2(\varepsilon, \rho(x_0))$ to be the smallest t such that

$$\|f_t\|_{2,\mu} < \varepsilon.$$

Let $B(x, \tau)$ be defined to be the n-dimensional Euclidean ball of radius τ centered at x and suppose K is an n-dimensional convex set such that $B(0, r) \subseteq K \subseteq B(0, R)$. The Markov chain known as the "Ball walk" [11, 18] is defined as follows. If the random walker is at a point x_i in a convex body K at time step i, a random point z is picked in $B(x_i, O(\frac{1}{\sqrt{n}}))$, and x_{i+1} is set to z if it lies in K, otherwise the move is rejected and x_{i+1} is set to x_i. More recently, a random walk known as Hit-and-Run, was analyzed in [17] and [20]. If the random walker is at a point x_i in a convex body K at time step i, a vector is picked from the uniform distribution on the sphere and through x_i, and x_{i+1} is chosen from the uniform measure on the chord $\{x_i + \lambda v | \lambda \in \mathbb{R}\} \cap K$. Unlike the Ball walk, this walk provably mixes rapidly from any interior point, with a weak (logarithmic) dependence on the distance of the starting point from the boundary. From a warm start, the mixing time of Hit-and-Run is $O(\frac{n^2 R^2}{r^2})$, and its mixing time from a fixed point at a distance d from the boundary is $O(n^3(\frac{R}{r})^2 \ln \frac{R}{d})$.

The mixing time of the Dikin walk in two cases of interest are as follows (details are provided in Theorem 1). Let $x \in K$ and for all chords $\overline{pq}$ passing through x, $\frac{|p-x|}{|q-x|} \le s$. We will call such a point x, s-central. Suppose $K \in \mathbb{R}^n$ is:

1. The intersection of an n-dimensional affine subspace (identified with $\mathbb{R}^n$) with the semi-definite cone $S^{\nu_h \times \nu_h}$ of $\nu_h \times \nu_h$ matrices endowed with the hyperbolic barrier $F(x) = -\ln \det x$ or

2. The intersection of $m = \frac{\nu_h}{2}$ ellipsoids, $A_1 B \cap A_2 B \cap \cdots \cap A_m B$ where A_i are nonsingular affine transformations and B is the Euclidean ball. In this case, the hyperbolic barrier is $F(x) = -\sum_i \ln(1 - \|A_i^{-1}(x)\|^2)$.

Then we define the complexity parameter of the barriers as $\nu = n\nu_h$ and the mixing time starting at x is $O(n\nu(n \ln(\nu s) + \ln \frac{1}{\varepsilon}))$. Whether or not the bodies are in isotropic position, in the above cases 1 and 2 corresponding to semi-definite and quadratic programs, when $\nu_h = O(n^{1-\varepsilon})$ [and so $\nu = O(n^{2-\varepsilon})$ and the mixing time from a warm start is $O(n^{3-\varepsilon})$], the mixing time bounds are an improvement over the existing bound for Hit-and-Run [20], which is $O(n^3)$ [assuming $R/r = \Omega(\sqrt{n})$].

Lovász [17] proved a lower bound of $\Omega(n^2 p^2)$ on the mixing time of Hit-and-Run in a cylinder $B_n \times [-p, p]$ from a warm start, where B_n is the unit ball in n-dimensions. Dikin walk has a mixing time of $O(n^2)$ from a warm start. Thus, for a cylinder with $p = \omega(1)$, the lower bound on the number of steps needed for Hit-and-Run to mix (without rescaling the body) is larger than the upper bound on the number of steps for Dikin walk.

In the specifc case where the constraints are either semi-definite or linear, we can compare the upper bounds on the number of arithmetic operations needed for one step in Hit-and-Run and Dikin walk. This has been done in Section 10.

3. Self-concordant barriers. Let K be a convex subset of $\mathbb{R}^n$ that is not contained in any $(n-1)$-dimensional affine subspace and $\text{int}(K)$ denote its interior. For any function F on $\text{int}(K)$ having continuous derivatives of order k, for vectors $h_1, \ldots, h_k \in \mathbb{R}^n$ and $x \in \text{int}(K)$, for $k \geq 1$, we recursively define

$$D^k F(x)[h_1, \ldots, h_k]$$
$$:= \lim_{\varepsilon \to 0} \frac{D^{k-1} F(x + \varepsilon h_k)[h_1, \ldots, h_{k-1}] - D^{k-1} F(x)[h_1, \ldots, h_{k-1}]}{\varepsilon},$$

where $D^0 F(x) := F(x)$. Following Nesterov and Nemirovskii [27], we call a real-valued function $F : \text{int}(K) \to \mathbb{R}$, a regular self-concordant barrier if it satisfies the conditions stated below. For convenience, if $x \notin \text{int}(K)$, we define $F(x) = \infty$.

1 (Convex, smooth). F is a convex thrice continuously differentiable function on $\text{int}(K)$.

2 (Barrier). For every sequence of points $\{x_i\} \in \text{int}(K)$ converging to a point $x \notin \text{int}(K)$, $\lim_{i \to \infty} f(x_i) = \infty$.

3 (Differential inequalities). For all $h \in \mathbb{R}^n$ and all $x \in \text{int}(K)$, the following inequalities hold.

(a) $D^2F(x)[h, h]$ is 2-Lipschitz continuous with respect to the local norm, which is equivalent to

$$D^3F(x)[h, h, h] \leq 2\big(D^2F(x)[h, h]\big)^{3/2}.$$

4. $F(x)$ is ν-Lipschitz continuous with respect to the local norm defined by F,

$$\big|DF(x)[h]\big|^2 \leq \nu D^2F(x)[h, h].$$

We call the smallest nonnegative real number ν for which this holds for all $h \in \mathbb{R}^n$ and x in the interior of K, the self-concordance parameter of the barrier.

It follows from these conditions that if F is a self-concordant barrier for K and A is a nonsingular affine transformation, then $F_A(x) := F(A^{-1}x)$ is a self-concordant barrier for AK. This fact is responsible for the affine-invariance of Dikin walk. Some examples of convex sets for which explicit barriers are known are:

1. Convex sets defined by hyperbolic constraints. This set includes sections of semi-definite cones. Polytopes and the intersections of ellipsoids can be expressed as sections of semi-definite cones.

2. Convex sets defined by the epigraphs of matrix norms (see page 199 of [27]).

For other examples and methods of constructing barriers for new convex sets by combining existing barriers; see Chapter 5 of [27].

3.1. *Generic self-concordant barrier.* We refer by F_s to a generic self-concordant barrier.

3.2. *Hyperbolic barriers.*

DEFINITION 3. A homogenous polynomial $p : \mathbb{C}^n \to \mathbb{C}$ is called hyperbolic with respect to a direction $d \in \mathbb{R}^n$ if $p(d) \neq 0$ and there exists a constant $t_0 \in \mathbb{R}$ such that $p(x + itd) \neq 0$ if $x \in \mathbb{R}^n$ and $t < t_0$.

Associated with such a polynomial is a cone of hyperbolicity (see [9]). The function

$$F_h(x) = -\log p(x)$$

is called a hyperbolic barrier.

For the concrete applications in this paper, it suffices to note that on the semi-definite cone $S^{m \times m}$, $-\ln \det x$ is a hyperbolic barrier with parameter m, and that on the intersection of ellipsoids, $A_1B \cap A_2B \cap \cdots \cap A_mB$ where A_i are nonsingular affine transformations and B is the Euclidean ball, $-\sum \ln(1 - \|A_i^{-1}(x)\|^2)$ is a hyperbolic barrier with self-concordance parameter m.

LEMMA 1 (Theorem 4.2, Güler [9]). *If F is a hyperbolic barrier,*

$$\big|D^4F(x)[h, h, h, h]\big| \leq 6\big(D^2F(x)[h, h]\big)^2.$$

3.3. *Logarithmic barrier of a polytope.* Given any set of linear constraints $\{a_i^T x \leq 1\}_{i=1}^m$, the logarithmic barrier is a real valued function defined on the intersection of the half-spaces defined by these constraints, and is given by

$$F_\ell(x) = -\sum_{i=1}^m \ln(1 - a_i^T x).$$

3.4. *Dikin ellipsoids.* Around point $x \in K$, given a self-concordant barrier F the *Dikin ellipsoid* of radius r is defined to be

$$D_x := \{y : D^2 F(x)[x - y, x - y] \leq r^2\}.$$

Dikin ellipsoids are affine invariants in that, if the Dikin ellipsoid of radius r around a point $x \in K$ is D_x^r and T is a nonsingular affine transformation of K, the Dikin ellipsoid of radius r centered at the point Tx for $T(K)$ is $T(D_x^r)$, as long as the new barrier that is used is $H(y) := F(T^{-1}y)$.

The following fact will be used in the proof of Lemmas 17 and 18.

FACT 1.   For any y such that

$$D^2 F(x)[x - y, x - y] = r^2 < 1,$$

for any vector $h \in \mathbb{R}^n$,

$$(1 - r)^2 D^2 F(x)[h, h] \leq D^2 F(y)[h, h] \leq \frac{1}{(1 - r)^2} D^2 F(x)[h, h].$$

Also, the Dikin ellipsoid centered at x, having radius 1, is contained in K. This has been shown in Theorem 2.1.1 of Nesterov and Nemirovskii [27].

The following was proved in a more general context by Nesterov and Todd in Theorem 4.1, [28].

FACT 2 (Nesterov–Todd).   Let $\overline{pq}$ be a chord of a polytope P and x, y be interior points on it so that p, x, y, q are in order. Let D_x be the Dikin ellipsoid of unit radius at x with respect to a point x. Then $z \in D_y$ implies that $p + \frac{|p-x|}{|p-y|}(z - p) \in D_x$.

4. Oracle model.   There are two standard information models for convex sets in the operations research literature, the separation model and the (self-concordant) barrier model (see Freund [7], page 2). Existing work on sampling convex sets, with the exception of Kannan and Narayanan [12] has focussed on the separation model and a weaker model known as the membership oracle model. The self-concordant barrier model we will consider is the following:

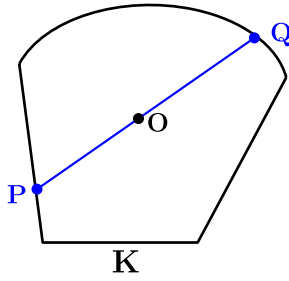1. We are guaranteed that the origin belongs to K and that K has a self-concordant barrier F (see Section 3).

FIG. 1. *A chord $\overline{pq}$ in a convex body K.*

2. We are given a real number s such that for any chord $\overline{pq}$ of K such that $\frac{|p|}{|q|} \geq 1$ through the origin, $\frac{|p|}{|q|} \leq s$ (see Figure 1).

3. On querying a point $x \in \mathbb{R}^n$, we are returned a positive semi-definite matrix corresponding to the Hessian of F at x, if $x \in K$ and returned "No" if $x \notin K$.

4.1. *Presentation of the convex set K.* For the definitions of logarithmic, hyperbolic and self-concordant barriers, we refer to Section 3. We will assume that the convex set K is specified as the set of points that satisfy a family of constraints

$$K := \bigcap_{i=1}^{m} \{F_i(x) < \infty\},$$

where the F_i are either logarithmic, hyperbolic or arbitrary self-concordant functions. Without loss of generality, we may aggregate these barriers and may assume that $K := K_\ell \cap K_h \cap K_s$, where K_ℓ is a polytope with m faces accompanied by the logarithmic barrier F_ℓ, K_h is a convex set accompanied by a hyperbolic barrier F_h with self-concordance parameter ν_h, and K_s is a convex set accompanied by a self-concordant barrier F_s whose self-concordance parameter is ν_s. Although their intersection is bounded, each of these convex sets may be unbounded.

DEFINITION 4. Given F_ℓ, F_h, F_s as above, we define the self-concordant barrier function

$$F := F_\ell + n F_h + n^2 F_s,$$

and define

(1) $$\nu := m + n\nu_h + (n\nu_s)^2$$

to be the *complexity parameter* of F (which is different from its *self-concordance parameter* the latter being bounded above by $m + n\nu_h + n^2\nu_s$).

Let C be a sufficiently large universal constant. We define the radius of a Dikin step, r to be $1/C$. For a point $x \in K$ and $v \in \mathbb{R}^n$, we define

$$\|v\|_x^2 := D^2 F(x)[v, v].$$

The random walk we use here is a variation of the Dikin walk defined in [12]. Instead of picking the next point from a Dikin ellipsoid, here we pick it from a Gaussian having that covariance.

5. Main results. In this section, we present our main results. Theorem 1 shows that given a convex body accompanied with a barrier, it is possible to use the Hessian of the barrier to define a rapidly mixing random walk. This allows us to efficiently pick a nearly random point from the uniform measure on the convex set.

Theorem 2 shows that in a special case where the convex set is a direct product of convex sets, the mixing time is governed by the worst factor. While we are not aware of any algorithmic application of this result, since it is easy to sample from a direct product, given samples from the factors, this illustrates that one can do better than to apply the localization lemma [18] in a natural special case.

The mixing results can be adapted to give a random walk based polynomial-time Las Vegas algorithm for optimizing a linear function $c^T x$ on certain convex sets K. The complexity of this algorithm is roughly the same as that of the sampling algorithm. This is the content of Theorem 3.

5.1. *Sampling through Dikin walk.* For $x \in \text{int}(K)$, let G_x denote the Gaussian density function given by

$$G_x(y) := \left(\frac{n}{\pi r^2}\right)^{n/2} \exp\left(-\frac{n\|x-y\|_x^2}{r^2} + V(x)\right),$$

where

$$V(x) = \left(\tfrac{1}{2}\right) \ln \det D^2 F(x).$$

Let r be a sufficiently small absolute constant (asymptotically less or equal to $\frac{1}{\sqrt{2}}$) so that the following tail estimate for the Gaussian holds.

FACT 3. Suppose $x \to z$ is a transition of the Dikin walk given by

$$G_x(z) := \left(\frac{n}{\pi r^2}\right)^{n/2} \exp\left(-\frac{n\|x-z\|_x^2}{r^2} + V(x)\right),$$

then

$$\mathbb{P}\big[\|x-z\|_x \leq \tfrac{1}{2}\big] \geq 1 - 10^{-3}.$$

The Dikin walk algorithm is given above (see Algorithm 1).
We prove the following in Section 7.3.

Algorithm 1 Dikin walk algorithm

Let $0 = x_0 \in \text{int}(K)$. For $i \geq 1$, given x_{i-1}:

1. Toss a fair coin. If `Heads` let $x_i := x_{i-1}$.
2. Else:
   (a) Choose z from the density $G_{x_{i-1}}$.
   (b) If $z \in K$, let

$$x_i := \begin{cases} z, & \text{with probability } \min\left(1, \dfrac{G_z(x_{i-1})}{G_{x_{i-1}}(z)}\right), \\ x_{i-1}, & \text{otherwise.} \end{cases}$$

   (c) If $z \notin K$, let $x_i := x_{i-1}$.

THEOREM 1 (Sampling). *Let $K \ni 0$ be an n-dimensional convex set accompanied by a barrier F as in Section* 4.1, *with complexity parameter v. Let $s \geq \frac{|p|}{|q|}$ for any chord $\overline{pq}$ of K containing the origin, with $\frac{|p|}{|q|} \geq 1$. Let e be the time of the first nontrivial move of the Markov chain. Then the number of steps x_t after e, before both the distances—total variation distance and the $\mathcal{L}_2$ distance of the density $\rho_t(x)$ of x_t to the uniform density—are less than ε is $O(nv(n\ln(vs) + \ln\frac{1}{\varepsilon}))$. The number of steps needed from a warm start is $O(nv\ln\frac{1}{\varepsilon})$. The time of the first nontrivial move e has a geometric distribution whose mean is bounded above by a universal constant.*

In particular, suppose K is:

(S) a slice of the semi-definite cone $S^{v_h \times v_h}$ of $v_h \times v_h$ matrices with $F(x) = -\ln \det x$ or

(Q) the intersection of $\frac{v_h}{2}$ ellipsoids, $A_1 B \cap A_2 B \cap \cdots \cap A_{v_h/2} B$ where A_i are nonsingular affine transformations and B is the Euclidean ball. In this case, $F(x) = -\sum \ln(1 - \|A_i^{-1}(x)\|^2)$.

In each case, the complexity parameter v is nv_h and the mixing time from a fixed "s-central" point or a warm start, respectively, are $O(nv(n\ln(vs) + \ln\frac{1}{\varepsilon}))$ and $O(nv\ln\frac{1}{\varepsilon})$.

The mixing bounds in this paper are obtained by relating the Markov chain to the metric of a Riemannian manifold studied in operations research [26] and [29], rather than the Hilbert metric [12] and [17]. The aforementioned Riemannian metric possesses several potentially useful characteristics. For example, when the convex set is a direct product of convex sets, this metric factors in a natural way into a product of the metrics corresponding to the individual convex sets, which is not the case for the Hilbert metric. Using results of Barthe [2] and Bobkov and Houdré [5] on the isoperimetry on product manifolds, this leads to an improved

upper bound on the mixing time when K is a direct product of convex sets, and opens up the future possibility of using differential-geometric techniques for proving isoperimetric bounds, in addition to relying on the localization lemma, which underlies the analysis of all Markov chains on convex sets ever since it was introduced in (Lovász and Simonovits [18]). Even if K is a direct product of convex sets, the Dikin Markov chain itself does not factor into a product of Dikin Markov chains and Theorem 2 does not follow from a direct use of the localization lemma.

THEOREM 2.   *If an n-dimensional convex set $K := K_1 \times \cdots \times K_h$ is the direct product of convex sets K_i, each of which individually has a function F_i with a complexity parameter [defined in equation (1)] at most ν, then the mixing time of Dikin walk from a warm start on K defined using the function $\sum_{i=1}^{h} F_i$ is $O(\nu n)$.*

When there are $h = \Omega(n)$ factors, each of which is a polytope with κ faces, the total number of faces of K is $\nu_\ell = \Omega(n\kappa)$. In this case, the results of [12] give a bound of $O(n\nu_\ell)$ while Theorem 2 gives a bound of $O(\nu_\ell)$.

5.2. *Convex programming.*   Our algorithm for convex optimization in this paper is a Las Vegas algorithm rather than a Monte Carlo algorithm (as was the case in [12]). It is also different from [12] in that the Markov chain used here does not depend on ε, the error tolerance.

We will consider convex programs specified as follows. Suppose we are given a convex set K containing the origin as an interior point and a linear objective c, such that

$$Q := K \cap \{y : c^T y \le 1\}$$

is bounded, for any chord $\overline{pq}$ of Q passing through the origin, $\frac{|p|}{|q|} \le s$ and $\varepsilon > 0$ [if $B(0, r) \subseteq K \subseteq B(0, R)$, then $s \le \frac{R}{r}$]. Then the algorithm is required to do the following. If $\exists x \in K$ such that $c^T x \ge 1$, output $x' \in K$ such that $c^T x' \ge 1 - \varepsilon$.

Let $T : Q \to \mathbb{R}^n$ be defined by

$$T(x) = \frac{x}{1 - c^T x},$$

and let $\hat{F}$ be a barrier for $\hat{K} := T(Q)$. Such a barrier can be easily constructed from F; details appear in Section 5.2.1.

5.2.1. *Constructing barriers.*   Let $T : Q \to \mathbb{R}^n$ be defined by $T(x) = \frac{x}{1 - c^T x}$ and $\hat{K}$ be the projective image TQ of Q as defined in Section 5.2. The construction in [25], provides us with a barrier $\hat{F}_s$ on $\hat{K}$. The barrier $\hat{F}_s$ is given by

$$\hat{F}_s(y) := \left(\frac{8}{3\sqrt{3}} + \frac{1}{2\sqrt{\nu_s}} \left(\frac{7}{3} \right)^{3/2} \right)^2 \left(F_s \left(\frac{y}{1 + c^T y} \right) + 2\nu_s \ln(1 + c^T y) \right),$$

whose self-concordance parameter is $\leq (3.08\sqrt{\nu} + 3.57)^2$. If $F_h = -\ln p(x)$ where p is a hyperbolic polynomial of degree ν_h, $\hat{F}_h$ is defined simply by

$$\hat{F}_h(y) := F_h\left(\frac{y}{1 + c^T y}\right) + \nu_h \ln(1 + c^T y),$$

and has the same self-concordance parameter ν_h. This applies to the special case of the logarithmic barrier as well. For any point $x \in \text{int}(\hat{K})$, we use the Hessian matrix $D^2 \hat{F}$ to define a norm

$$\|\hat{v}\|_x := (v^T D^2 \hat{F} v)^{1/2},$$

$$\hat{G}_x(y) := \left(\frac{n}{\pi r^2}\right)^{n/2} \exp\left(-\frac{n\|\widehat{x - y}\|_x^2}{r^2} + \hat{V}(x)\right),$$

where

$$\hat{V}(x) = \left(\tfrac{1}{2}\right) \ln \det D^2 \hat{F}(x).$$

For $x \notin \text{int}(\hat{K})$, for any $y \neq x$, $\hat{G}_x(y) := 0$. Let

$$s := \sup_{\overline{pq} \ni O} \frac{|p|}{|q|},$$

where the supremum is taken over all chords of K containing the origin.

Our algorithm for convex optimization consists simply of doing a modified Dikin walk (see Figure 2) on $\hat{K}$ for a sufficient number of steps that depends on the desired accuracy ε and confidence $1 - \delta$. Note that we define $\hat{G}_t$ using $\hat{F}$ in the same way that G_t was defined using F.

5.2.2. *Las Vegas algorithm.*   We state the Las Vegas algorithm above (see Algorithm 2).

THEOREM 3.   *Let K, F and s be as in Theorem 1. In the cases where F is a ν-barrier or a hyperbolic barrier with parameter ν, let $\tau(\varepsilon, \delta)$ be set to*
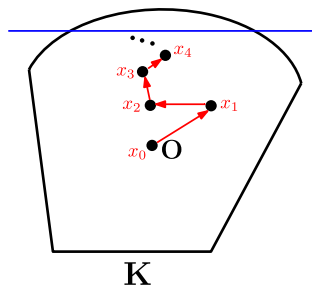

FIG. 2.   *Trajectory of a Dikin walk for optimizing a linear function on a convex set.*

Algorithm 2 Las Vegas algorithm for optimization

Let $x_0 = 0$. While $c^T x_{i-1} < 1 - \varepsilon$:

1. Toss a fair coin. If Heads, set $x_i := x_{i-1}$.
2. Else:
   (a) Choose z from the density $\hat{G}_{T(x_{i-1})}$.
   (b) If $z \in \hat{K}$, let

   $$x_i := \begin{cases} T^{-1}(z), & \text{with probability } \min\left(1, \dfrac{\hat{G}_z(T(x_{i-1}))}{\hat{G}_{T(x_{i-1})}(z)}\right), \\ x_{i-1}, & \text{otherwise.} \end{cases}$$

   (c) If $z \notin \hat{K}$, let $x_i := x_{i-1}$.

$O(n\nu(\ln \frac{1}{\delta} + (n \ln \frac{sn\nu}{\varepsilon})))$. *If $\{c^T x \geq 1\} \cap K$ is nonempty and $x_0, x_1, \ldots$ is the mod-ified Dikin walk in Las Vegas algorithm, then*

$$\mathbb{P}[\forall i > \tau(\varepsilon, \delta), c^T x_i \geq 1 - \varepsilon] \geq 1 - \delta.$$

We remind the reader that for a Las Vegas algorithm, correctness is guaranteed, but there are no absolute bounds on the run time. Consider the stopping rule under which the random walk is terminated the first time that it reaches $c^T x_i \geq 1 - \varepsilon$. The following corollary shows that the resulting algorithm is a Las Vegas algorithm.

COROLLARY 4 (Las Vegas algorithm for optimization). *For any $\kappa > 0$, with probability* 1,

$$\lim_{i \to \infty} e^{i^{1-\kappa}} (1 - c^T x_i) = 0.$$

6. Isoperimetry.

6.1. *Metric defined by a barrier.* The isoperimetric properties of a certain met-ric measure space govern the mixing bounds in the present work. In this section, we provide preliminary definitions and properties needed in our proofs.

For any smooth strictly convex function G, the Hessian $D^2 G$ is positive definite.

DEFINITION 5. Given the barrier F, for every $x \in \text{supp}(F)$ and $u, v \in \mathbb{R}^n$, $D^2 F(x)[u, v]$ is bilinear, and $\sqrt{\langle_F u, u\rangle_x}$ is a norm. We define:

   (i) $\langle_F u, v\rangle_x := D^2 F(x)[u, v]$ and $\|_F u\|_x := \sqrt{\langle_F u, u\rangle_x}$,
   (ii) $\langle_\ell u, v\rangle_x := D^2 F_\ell(x)[u, v]$ and $\|_\ell u\|_x := \sqrt{\langle_\ell u, u\rangle_x}$,
   (iii) $\langle_h u, v\rangle_x := D^2 F_h(x)[u, v]$ and $\|_h u\|_x := \sqrt{\langle_h u, u\rangle_x}$ and
   (iv) $\langle_s u, v\rangle_x := D^2 F_s(x)[u, v]$ and $\|_s u\|_x := \sqrt{\langle_s u, u\rangle_x}$.

DEFINITION 6. We define

$$d_F(x, y) = \inf_{\Gamma} \int_z \|_F d\Gamma\|_z \, dz,$$

where the infimum is taken over all rectifiable paths Γ from x to y. d_ℓ, d_h and d_s are defined analogously in terms of the respective norms $\|_\ell \cdot \|$, $\|_h \cdot \|$ and $\|_s \cdot \|$, which we refer to as Dikin norms:

(i) $d_\ell(x, y) = \inf_{\Gamma} \int_z \|_\ell d\Gamma\|_z \, dz.$
(ii) $d_h(x, y) = \inf_{\Gamma} \int_z \|_h d\Gamma\|_z \, dz.$
(iii) $d_s(x, y) = \inf_{\Gamma} \int_z \|_s d\Gamma\|_z \, dz.$

Where F is clear from context, the subscript F will be skipped.

Let $\mathcal{M}$ be the metric space associated with F whose point set is K and metric is d_F.

The following lemma is needed to relate the Riemannian metric $d_s(x, y)$ to the Dikin norm $\|_s x - y\|$.

LEMMA 2 (Nesterov–Todd, Lemma 3.1, [29]). *If* $\|_s x - y\|_x < 1$ *then,*

$$\|_s x - y\|_x - \|_s x - y\|_x^2 \le d_s(x, y) \le -\ln(1 - \|_s x - y\|_x).$$

Since any logarithmic or hyperbolic barrier is also a self-concordant barrier, this implies that if $\|_F x - y\|_x < 1$ then

$$\|_F x - y\|_x - \|_F x - y\|_x^2 \le d_F(x, y) \le -\ln(1 - \|_F x - y\|_x).$$

While some of the presented bounds can be obtained from the isoperimetric bounds for the "Hilbert metric" (Theorem 7) proved by Lovász, we can prove stronger results for sampling certain classes of convex sets such as the direct product of an arbitrary number of convex sets, by using results of Barthe [2] and Bobkov and Houdré [5] on the isoperimetry of product spaces, which do not seem to follow directly from the Hilbert metric. In particular, for a direct product of an arbitrary number of polytopes, each defined by $O(\kappa)$ constraints, this allows us to show an upper bound on the mixing time from a warm start of $O(\kappa n)$. The bound obtained using the Hilbert metric in the obvious way is $O(\kappa n^2)$, since the Hilbert metric on a direct product does not decompose conveniently into factors as does the Riemannian metric.

Riemannian metrics defined in this way have been studied because of their importance in convex optimization, for example, by Nesterov and Todd in [29] and by Nesterov and Nemirovski in [26], and Karmarkar studied the properties of a related metric [15] that underlay his celebrated algorithm [14]. For other work on sampling Riemannian manifolds motivated by statistical applications, see [16] and [22], and Chapter 8 of [21] and the references therein.

6.2. *Results on isoperimetry.*   Let $\mathcal{M}$ be a metric space endowed with distance function d and μ be a probability measure on it. We term $(S_1, \mathcal{M} \setminus (S_1 \cup S_2), S_2)$ a δ-partition of $\mathcal{M}$, if

$$\delta \leq d(S_1, S_2) := \inf_{x \in S_1, y \in S_2} d(x, y),$$

where S_1, S_2 are measurable subsets of $\mathcal{M}$. Let $\mathcal{P}_\delta$ be the set of all δ-partitions of $\mathcal{M}$. The isoperimetric constant $\beta_{\text{fat}}(\delta, \mathcal{M}, \mu)$ is defined as

$$\inf_{\mathcal{P}_\delta} \frac{\mu(\mathcal{M} \setminus (S_1 \cup S_2))}{\mu(S_1)\mu(S_2)}.$$

Given interior points x, y in $\text{int}(K)$, suppose p, q are the ends of the chord in K containing x, y and p, x, y, q lie in that order.

DEFINITION 7.   Denote by d_H the Hilbert (projective) metric defined by

$$d_H(x, y) := \ln\left(1 + \frac{|x - y||p - q|}{|p - x||q - y|}\right).$$

DEFINITION 8.   For $x \in K$ and a vector v, $|v|_x$ is defined to be

$$\sup_\alpha\{x \pm \alpha v \in K\}.$$

Let $\beta_{\text{fat}} := \beta_{\text{fat}}(\delta, \mathcal{M}, \mu)$, where $\delta = \frac{1}{\sqrt{n}}$.

THEOREM 5 (Theorem 2.3.2(iii), [27]).   *Let F be a self-concordant barrier whose self-concordance parameter is v_s as defined in Section* 3. *Then, for all $h \in \mathbb{R}^n$ and $x \in \text{int}(K)$*

$$|h|_x \leq \|{}_s h\|_x \leq 2(1 + 3v_s)|h|_x.$$

The following result is implicit in [9].

THEOREM 6 (Güler, [9]).   *Let $-\ln p(x)$ be a hyperbolic barrier for K, where p has degree v_h. Then, for all $h \in \mathbb{R}^n$ and $x \in \text{int}(K)$,*

$$|h|_x \leq \sqrt{D^2 F_h(x)[h, h]} \leq \sqrt{v_h}|h|_x.$$

LEMMA 3.   1. $d_s(x, y) \leq 2(1 + 3v_s)\, d_H(x, y).$
2. $d_h(x, y) \leq \sqrt{v_h}d_H(x, y).$
3. $d_\ell(x, y) \leq \sqrt{m}d_H(x, y).$
4. $d_F(x, y) \leq O(\sqrt{v})\, d_H(x, y).$

PROOF.   For any z on the segment $\overline{xy}$, $d_\mathrm{H}(x, z) + d_\mathrm{H}(z, y) = d_\mathrm{H}(x, y)$. Therefore, it suffices to prove the result infinitesimally. By Lemma 2,

$$\lim_{y \to x} \frac{d_F(x, y)}{\|_F x - y\|_x} = 1,$$

and a direct computation shows that

$$\lim_{y \to x} \frac{d_\mathrm{H}(x, y)}{|x - y|_x} \geq 1.$$

Parts 1 and 2 of Lemma 3 follow from Theorems 5 and 6. Part 3 is a special case of part 2. Part 4 follows from adding up parts 1 to 3.   □

THEOREM 7 (Lovász, [17]).   *Let S_1 and S_2 be measurable subsets of K. Let μ be the uniform probability measure on K. Then*

$$\mu\big(K \setminus (S_1 \cup S_2)\big) \geq \big(e^{d_\mathrm{H}(S_1, S_2)} - 1\big)\mu(S_1)\mu(S_2).$$

THEOREM 8.   *If F is a self-concordant barrier of K with complexity parameter ν, presented in the format of Section* 4.1, *then*

$$\beta_{\mathrm{fat}} = \Omega\left(\frac{1}{\sqrt{n\nu}}\right).$$

PROOF.   Theorem 8 follows from Theorem 7 and Lemma 3.   □

7. Analysis of the mixing time.

7.1. *Preliminaries.*   We denote the conditional distribution of x_{i+1} given $x_i = x$ by P_x. Lemma 4 is a statement about the concentration of derivatives of odd order in high dimension. It will be used in the proof of Lemma 6, which relates the Riemannian distance between two points x and y to the total variation distance between P_x and P_y. Lemma 5 states that if the unit Dikin ellipsoid around a point contains the unit ball, then the points at which a random line through x chosen from the distribution induced by the uniform measure on the unit sphere intersects the boundary are, with high probability, at a distance $\Omega^*(\sqrt{n})$ from x.

LEMMA 4 (Concentration bound).   *Let h be chosen uniformly at random from the unit sphere $S^n = \{v | \|v\| = 1\}$. Then, for any odd k,*

$$\mathbb{P}\left[D^k F(x)[h, \ldots, h] > k\varepsilon \sup_{\|v\|_x \leq 1} D^k F(x)[v, \ldots, v]\right] \leq \exp\left(\frac{-n\varepsilon^2}{2}\right).$$

If F is a self-concordant barrier, and

$$\forall v, \qquad \|v\|^2 \geq D^2 F(x)[v, v]$$

when $k = 3$, this simplifies to

$$\mathbb{P}[D^3 F(x)[h, h, h] > 3\varepsilon] \leq \exp\left(\frac{-n\varepsilon^2}{2}\right).$$

PROOF.   The "Bernstein inequality" of Gromov (Section 8.5, [8]) which applies to multivariate polynomials restricted to S^n, states that for any polynomial p on S^n of degree k,

$$\sup_{h \in S^n} \|\operatorname{grad} p(h)\| \leq k \sup_{h \in S^n} |p(h)|.$$

For any fixed x, $D^k F(x)[h, \ldots, h]$ is a polynomial in h of degree k. Therefore,

$$\frac{D^k F(x)[h, \ldots, h]}{k \sup_{\|v\|_x \leq 1} |D^k F(x)[v, \ldots, v]|}$$

is 1-Lipschitz on S^n. If k is odd, $D^k F(x)[h, \ldots, h] = -D^k F(x)[-h, \ldots, -h]$ and, therefore, its median with respect to the uniform measure σ on the unit sphere is 0. The first part of the lemma follows from the measure concentration properties of Lipschitz functions on the sphere (page 44 in [1]); namely, if f is an 1-Lipschitz function on the unit sphere and M is its median, then

(2) $$\sigma(p > M + \varepsilon) \leq e^{-(n\varepsilon^2)/2}.$$

When F is a self-concordant barrier, the second statement follows because

$$\sup_{\|v\| \leq 1} D^3 F(x)[v, v, v] \leq \sup_{\|v\|_x \leq 1} D^3 F(x)[v, v, v] \leq 1. \qquad \square$$

LEMMA 5.   *Let P be a polytope and x a point in it. Let the Dikin ellipsoid at x with respect to the logarithmic barrier at x contain the unit ball. Let v be chosen uniformly at random from the unit ball centered at x and ℓ be the line through x and v, and p and q be the two points of intersection of ℓ with the boundary ∂P. Then, for any constant $\alpha \geq 0$,*

$$\mathbb{P}\left[\min(\|p\|, \|q\|) \leq \sqrt{\frac{n}{2(\alpha + 2\ln n)}}\right] \leq 2e^{-\alpha}.$$

PROOF.   Without loss of generality, we may assume x to be the origin. The unit ball is contained in the Dikin ellipsoid and so P can be expressed as $\bigcap_{i=1}^{m}\{a_i^T x \leq 1\}$, where

(3) $$I \succeq \sum_{i=1}^{m} a_i a_i^T.$$

Examining the trace and the norm on both sides of (3), we obtain

$$\forall i, \qquad \|a_i\| \leq 1$$

and

$$\sum_{i=1}^{m} \|a_i\|^2 \leq n.$$

We note that

$$\min(\|p\|, \|q\|) = \left(\min_i |a_i^T x|\right)^{-1}.$$

Thus, it is sufficient to show that

$$\mathbb{P}\left[\exists_i (a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right] \leq 2e^{-\alpha},$$

which we proceed to do. Let S be the subset of $[m] := \{1, \ldots, m\}$, consisting of those i for which $\|a_i\| \geq \frac{1}{\sqrt{n}}$. Clearly, if for some i, $(a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}$, then $i \in S$. By (3), $|S| \leq n^2$. Thus, by the union bound,

$$\mathbb{P}\left[\exists_{i \in S} (a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right] \leq n^2 \sup_{i \in S} \mathbb{P}\left[(a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right].$$

We note that, by (2), for any vector w with norm less or equal to 1,

$$\mathbb{P}\left[a_i^T w \geq \sqrt{\frac{2(\alpha + 2\ln n)}{n}}\right] \leq \frac{e^{-\alpha}}{n^2},$$

and so

$$\mathbb{P}\left[\exists_i (a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right] \leq 2e^{-\alpha}. \qquad \square$$

In order to obtain mixing time bounds, we will first prove in Lemma 6 that if two points x and y are nearby in that $d(x, y) \leq O(\frac{1}{\sqrt{n}})$, then the total variation distance between the corresponding distributions P_x and P_y is $< 1 - \Omega(1)$.

For $x \neq y$,

$$1 - d_{\text{TV}}(P_x, P_y) = \mathbb{E}_z\left[\min\left(1, \frac{G_y(z)}{G_x(z)}, \frac{G_z(x)}{G_x(z)}, \frac{G_z(y)}{G_x(z)}\right)\right],$$

where the expectation is taken over a random point z from the density G_x and

$$\min\left(1, \frac{G_y(z)}{G_x(z)}, \frac{G_z(x)}{G_x(z)}, \frac{G_z(y)}{G_x(z)}\right)$$

is defined to be 0 if $z \notin K$.

We will use the following fact (see Section 2.2, [24]) with $D^k F$ in the place of M.

FACT 4. Let $M[h_1, \ldots, h_k]$ be a symmetric k-linear form on $\mathbb{R}^n$. Then

$$M[h_1, \ldots, h_k] \leq \|h_1\| \|h_2\| \cdots \|h_k\| \sup_{\|v\| \leq 1} M[v, \ldots, v].$$

FACT 5. Let the eigenvalues of the covariance matrix of an n-dimensional Gaussian g be bounded above by λ. Let $\langle \cdot, \cdot \rangle$ be an inner product and $v \in \mathbb{R}^n$. Then $\mathbb{E}[\langle v, g \rangle^2] \leq \lambda \langle v, v \rangle$.

7.2. *Relating the Markov chain to the manifold.* In this subsection, we find an upper bound on the total variation distance between the distributions of one-step transition probabilities corresponding at two points at a Riemannian distance of $O(1/\sqrt{n})$. We also describe two results used in the proof of Theorem 1. We will frequently make statements of the form

$$\mathbb{P}[g(x) > O(f(x))] \leq c.$$

By this we mean, there exists a universal constant C such that

$$\mathbb{P}[g(x) > Cf(x)] \leq c.$$

Finally, we will frequently make use of the facts from Theorem 2.1.1, [27], stated below that Dikin ellipsoids vary smoothly, and that they are contained in the convex set.

- Given any self-concordant barrier F, for any y such that

$$D^2 F(x)[x - y, x - y] = r^2 < 1,$$

  for any vector $h \in \mathbb{R}^n$,

$$(4) \qquad (1 - r)^2 D^2 F(x)[h, h] \leq D^2 F(y)[h, h] \leq \frac{1}{(1 - r)^2} D^2 F(x)[h, h].$$

- The Dikin ellipsoid centered at x, having radius 1, is contained in K.

For two probability distributions P_x and P_y, let $d_{\mathrm{TV}}(P_x, P_y)$ represent the total variation distance between them. Without loss of generality, let x be the origin 0 (which is achievable by translation), and for any v, let $D^2 F(0)[v, v] = \|v\|^2$ (which is achievable by an affine transformation of K).

LEMMA 6 (Relating d to Markov chain). *If $x, y \in K$ and $d(x, y) \leq \frac{1}{C\sqrt{n}}$, then $d_{\mathrm{TV}}(P_x, P_y) = 1 - \Omega(1)$.*

PROOF. Without loss of generality, we may assume that F_h, F_ℓ and F_s are strictly convex. In case any one is not, we can add the strictly convex logarithmic barrier of a sufficiently large cube, thereby making an arbitrarily small change to its second, third and if it is not F_s, fourth order derivatives uniformly over K. Due

to affine invariance, without loss of generality, let $\langle u, v \rangle_x := \langle u, v \rangle$, the usual dot product. As defined in Section 5.1, for any $z \in K$,

$$V(z) = \tfrac{1}{2} \det D^2 F.$$

By Lemma 2, it suffices to prove that there is an absolute constant C such that if $x, y \in K$ and $\|x - y\|_x \leq \frac{1}{C\sqrt{n}}$, then $d_{\mathrm{TV}}(P_x, P_y) = 1 - \Omega(1)$. Without loss of generality, we assume x is the origin and we drop this subscript at times to simplify notation.

$$1 - d_{\mathrm{TV}}(P_x, P_y) = \mathbb{E}_z \left[\min \left(1, \frac{G_y(z)}{G_x(z)}, \frac{G_z(x)}{G_x(z)}, \frac{G_z(y)}{G_x(z)} \right) \right],$$

where the expectation is taken over a random point z having density G_x. Thus, it suffices to prove the existence of some absolute constant c such that

$$\mathbb{P}\left[\min \left(\frac{G_y(z)}{G_x(z)}, \frac{G_z(x)}{G_x(z)}, \frac{G_z(y)}{G_x(z)} \right) > c \right] = \Omega(1).$$

This translates to

$$\mathbb{P}\big[\max\big(n\|y - z\|_y^2 - r^2 V(y), n\|z\|_z^2 - r^2 V(z), n\|z - y\|_z^2 - r^2 V(z) \big)$$
$$- n\|z\|^2 < O(r^2) \big] = \Omega(1).$$

Now we use the following lemmas whose proofs are in the Appendix.

LEMMA 7. $-V(y) < O(1)$.

LEMMA 8.

(5) $$\mathbb{P}\big[-V(z) < O(1) \big] > \tfrac{9}{10}.$$

LEMMA 9.

$$\mathbb{P}\left[\max\big(\|y - z\|_y^2, \|z\|_z^2, \|z - y\|_z^2 \big) - \|z\|^2 < O\left(\frac{1}{n} \right) \right] > \frac{199}{1000}.$$

Substituting the statements of the last three lemmas into the preceding expression of probability completes the proof. $\square$

The following two results are needed in the proof of Theorem 1. Lemma 10 is proved in the Appendix and the proof is based on that of a theorem in [17].

LEMMA 10 (Bound on conductance). *Let μ be the uniform distribution on K. The conductance*

$$\Phi := \inf_{\mu(S_1) \leq 1/2} \frac{\int_{S_1} P_x(K \setminus S_1) \, d\mu(x)}{\mu(S_1)}$$

of the Markov chain in Dikin walk algorithm is $\Omega(\beta_{\mathrm{fat}})$.

THEOREM 9 (Lovász–Simonovits [18]). *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 kth step. Let $M := \sup_S \frac{\mu_0(S)}{\mu(S)}$ where the supremum is over all measurable subsets S of K. For every bounded f, let $\|f\|_{2,\mu}$ denote $\sqrt{\int_K f(x)^2 \, d\mu(x)}$. For any fixed f, let Ef be the map that takes x to $\int_K f(y) \, dP_x(y)$. Then:*

1. *For all S,*

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

2. *If $\int_K f(x) \, d\mu(x) = 0$,*

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

7.3. *Proof of Theorem* 1. We first give a summary of the proof. By Theorem 8, β_{fat} is bounded below by $\Omega(\frac{1}{\sqrt{n\nu}})$. Lemma 10 states that the conductance Φ of the Markov chain is bounded below by $\Omega(\beta_{\text{fat}})$. We substitute this in Theorem 9 to recover the fact that the mixing time from a warm start is $O(n\nu \ln \frac{1}{\varepsilon})$.

To recover the fact that the mixing time from an "s-central" point is

$$O\left(n\nu\left(n \ln(\nu s) + \ln \frac{1}{\varepsilon}\right)\right).$$

We first get a bound of $\exp(n \ln(\nu s))$ on the $\mathcal{L}_2$ norm of the starting density, which we then use in Theorem 9. This concludes our summary—we now proceed with the proof.

Let e be the first time the Markov chain escapes $x_0 = 0$. Thus, e is an integer valued random variable defined by the event that y_e is the first point in $y_1, \ldots$ that is not equal to y_0. Let the density of y_e be ρ_e. We know that $\forall t$, $\mathbb{P}[e \ge t+1 | e \ge t] \le 1 - \Omega(1)$, therefore,

$$\sup_x \frac{\rho_e(x)}{G_0(x)} = O(1).$$

Let $\mu(S)$ denote the measure assigned by the uniform measure on K on a measurable set S. Therefore,

$$\int_K \left(\text{vol}(K)\rho_e(x)\right)^2 d\mu(x) = \exp\left(O\left(n \ln(s\nu)\right)\right).$$

To see the last step, consider a linear transformation that maps the Dikin ellipsoid at the origin to the unit ball. Then the convex set is contained inside a ball of radius $O(s\nu)$, making $(\text{vol}(K))$ bounded above by $(\frac{O(s^2\nu^2)}{n})^{n/2}$ because the volume of a

ball in n-dimensions of radius R is $(\frac{O(R^2)}{n})^{n/2}$. Also, ρ_e is bounded above in sup-norm by $\exp((n/2)\ln(O(n)))$, because $G_0(x)$ is so bounded. Therefore

$$\int_K \rho_e(x)^2\, dx < \sup_K \rho_e(x) \int_K \rho_e(x)\, dx < \big(O(n)\big)^{n/2}.$$

This concludes our explanation.

For every bounded f, let $\|f\|_{2,\mu}$ denote $\sqrt{\int_K (\mathrm{vol}(K)f(x))^2\, d\mu(x)}$. Let $f_e(x) := \rho_e(x) - (\mathrm{vol}(K))^{-1}$. We then see from the above computation that

$$(6) \qquad \|f_e\|_{2,\mu} \le \|\rho_e(x)\|_{2,\mu} + \|(\mathrm{vol}(K))^{-1}\mathbf{1}_K\|_{2,\mu} < \exp\big(O(n\ln(sv))\big).$$

Theorem 9 states that

$$\forall S, \qquad \big|\mu_k(S) - \mu(S)\big| \le \sqrt{M}\left(1 - \frac{\Phi^2}{2}\right)^k.$$

Therefore, if

$$k > \frac{\ln\sqrt{M/\varepsilon^2}}{-\ln(1-(\Phi^2/2))} = O\left(\ln\!\left(\frac{1}{\varepsilon}\right)nv\right),$$

then

$$\forall S, \qquad \big|\mu_k(S) - \mu(S)\big| \le \varepsilon.$$

This completes the proof of mixing from a warm start, that is, when $M = O(1)$.

To recover the proof of mixing from a s-central point, we see by Theorem 9, Lemma 10 and (6) that

$$(7) \qquad \|E^k f_e\|_{2,\mu} \le \left(1 - \frac{\Phi^2}{2}\right)^k \|f_e\|_{2,\mu}$$

$$(8) \qquad \le \left(1 - \Omega\!\left(\frac{1}{nv}\right)\right)^k \exp\big(O(n\ln(sv))\big).$$

Therefore, for $k = \Omega(nv(\ln(1/\varepsilon) + n\ln(sv)))$, we have $\|E^k f_e\|_{2,\mu} \le \varepsilon$. This completes the proof of Theorem 1.

8. Mixing in a direct product of convex sets.

8.1. *Preliminaries.* Our analysis of mixing in a direct product of convex sets in Theorem 2 hinges upon a lower bound on the Cheeger constant β_{fat}, which is obtained by comparing the isoperimetry of the weighted manifold $\mathcal{M}$ obtained by equipping K with the metric from the Hessian of F, with the isoperimetry of the Dikin metric on the h-dimensional cube $[-\frac{\pi}{2}, \frac{\pi}{2}]^h$, with respect to the barrier $F_\square(x_1, \ldots, x_h) := -\sum_i \ln\cos(x_i)$ (see Section 6.2, [29]).

For a manifold $\mathcal{N}$ equipped with a measure μ and metric d, let the Minkowski outer measure of a (measurable) set A be defined as

$$\mu^+(A) := \liminf_{\varepsilon \to 0^+} \frac{\mu(A_\varepsilon) - \mu(A)}{\varepsilon},$$

where $A_\varepsilon := \{x | d_{\mathcal{M}}(x, A) < \varepsilon\}$.

DEFINITION 9.   The (infinitesimal) Cheeger constant of the weighted manifold $(\mathcal{N}, \mu)$ is

$$\beta_{\mathcal{N}} = \inf_{A \subset \mathcal{M}, \mu(A) \leq 1/2} \frac{\mu^+(A)}{\mu(A)},$$

where the infimum is taken over measurable subsets.

The isoperimetric function of μ is the largest function I_μ such that $\mu^+(A) \geq I_\mu(\mu(A))$ holds for all Borel sets.

Let the h-fold product space $\mathcal{N} \times \cdots \times \mathcal{N}$ be denoted $\mathcal{N}^h$, where the distance between points $(x_1, \ldots, x_h)$ and $(y_1, \ldots, y_h)$ is $\sqrt{\sum_i d(x_i, y_i)^2}$.

We will need the following theorem of Bobkov and Houdré (Theorem 1.1, [5]).

THEOREM 10.   *For any triple $(\mathcal{N}, d, \mu)$ as above,*

(9) $$I_{\mu^{\otimes h}} \geq \frac{1}{2\sqrt{6}} I_\mu.$$

We will also need the following theorem, which is a modification of Barthe (Theorem 10, [2]), obtained by scaling the metric on $\mathbb{R}$ by $\frac{1}{\sqrt{\nu}}$.

THEOREM 11.   *Let $k \geq 2$ be an integer. For $i = 1, \ldots, k$, let (X_i, d_i, μ_i) be a Riemannian manifold, with its geodesic distance and an absolutely continuous Borel measure of probability and let λ_i be a probability measure on $\mathbb{R}$ with even log-concave density. If $I_{\mu_i} \geq \frac{I_{\lambda_i}}{\eta}$ for $i = 1, \ldots, k$, then*

$$I_{\mu_1 \otimes \cdots \otimes \mu_k} \geq \frac{I_{\lambda_1 \otimes \cdots \otimes \lambda_k}}{\eta}.$$

The following lemma allows us to relate the "fat" Cheeger constant β_{fat} with the infinitesimal version $\beta_{\mathcal{M}}$.

LEMMA 11.   *Let $A, B \subseteq \mathcal{M}$ and $d_{\mathcal{M}}(A, B) \geq \delta_{\mathcal{M}}$. Then*

$$\mu(\mathcal{M} \setminus \{A \cup B\}) \geq 2 \min(\mu(A), \mu(B))(e^{(\beta_{\mathcal{M}} \delta_{\mathcal{M}})/2} - 1).$$

PROOF. We will consider two cases.

First, suppose that $\max(\mu(A), \mu(B)) < \frac{1}{2}$. Without loss of generality, we assume that $\mu(A) \leq \mu(B)$. Then, let

$$\delta_1 := \sup_{\mu(A_\delta) < 1/2} \delta.$$

We proceed by contradiction to show that

$$\mu(\mathcal{M} \setminus \{A \cup B\}) \geq 2\mu(A)(e^{(\beta_\mathcal{M} \delta_\mathcal{M})/2} - 1).$$

Suppose for some $\beta < \beta_\mathcal{M}$,

(10) $$\exists \delta \in [0, \delta_1), \qquad \mu(A_\delta) < e^{\beta \delta} \mu(A).$$

Let δ' be the infimum of such δ. Note that since $\mu(A_\delta)$ is a monotonically increasing function of δ,

$$\mu(A_{\delta'}) \geq e^{\beta \delta'} \mu(A).$$

However, we know that

$$\mu^+(\partial A_{\delta'}) := \liminf_{\varepsilon \to 0^+} \frac{\mu(A_\varepsilon) - \mu(A)}{\varepsilon} \geq \beta_\mathcal{M},$$

which contradicts the fact that in any right neighborhood of δ', there is a δ for which (10) holds. This proves that for all $\delta \in [0, \delta_1)$, $\mu(A_\delta) \geq e^{\delta \beta_\mathcal{M}} \mu(A)$. We note that $A_{\delta_1} \cap B_{\delta_\mathcal{M} - \delta_1} = \varnothing$, therefore, $\mu(B_{\delta_\mathcal{M} - \delta_1}) \leq \frac{1}{2}$. So the same argument tells us that

(11) $$\mu(B_{\delta_\mathcal{M} - \delta_1} \setminus B) \geq (e^{\beta_\mathcal{M}(\delta_\mathcal{M} - \delta_1)} - 1)\mu(B).$$

Thus, $\mu(\mathcal{M} \setminus \{A \cup B\}) \geq \mu(A)(e^{\beta_\mathcal{M}(\delta_\mathcal{M} - \delta_1)} + e^{\beta_\mathcal{M} \delta_1} - 2)$. This implies that

$$\mu(\mathcal{M} \setminus \{A \cup B\}) \geq 2\mu(A)(e^{(\beta_\mathcal{M} \delta_\mathcal{M})/2} - 1).$$

Next, consider the second case; suppose $\mu(B) \geq \frac{1}{2}$. We then set $\delta_1 := \delta_\mathcal{M}$, and see that the arguments from (10) to (11) carry through verbatim. Thus, in this case, $\mu(\mathcal{M} \setminus \{A \cup B\}) \geq \mu(A)(e^{\beta_\mathcal{M} \delta_\mathcal{M}} - 1) \geq 2\mu(A)(e^{(\beta_\mathcal{M} \delta_\mathcal{M})/2} - 1)$. $\quad\square$

This immediately leads to the following corollary.

COROLLARY 12.

$$\beta_{\text{fat}} = \Omega\left(\frac{\beta_\mathcal{M}}{\sqrt{n}}\right).$$

Nesterov and Todd show in Lemma 4.1, [29], that the Riemannian metric $\mathcal{M}$ on the direct product of convex sets induced by $\sum_{i=1}^{h} F_i$ is the same as the direct product of the Riemannian metrics $\mathcal{M}_i$ induced by individual F_i on the respective convex sets K_i.

Theorem 2 illustrates the utility of using the Riemannian metric in proving isoperimetric bounds in one specific case, namely when the convex set of interest is a product of convex sets having smaller dimension. A large number of combinatorial optimization questions can be viewed as optimizing a nonconvex quadratic function over a unit cube. This is one scenario where we hope the improved bounds on the mixing time may lead to interesting results in the future.

8.2. *Proof of Theorem* 2.   In order to show that the mixing time is $O(n\nu)$, by Theorem 9, it suffices to show that the conductance Φ is bounded below by $\Omega(\frac{1}{\sqrt{n\nu}})$. By Lemma 10, it is in turn enough to bound β_{fat} from below by $\frac{1}{\sqrt{n\nu}}$. By Corollary 12, it suffices to show that $\beta_{\mathcal{M}} = \Omega(\frac{1}{\sqrt{\nu}})$. We will show this using Theorems 11 and 10. Consider the h-dimensional cube $[-\frac{\pi}{2}, \frac{\pi}{2}]^h$, and the metric from the Hessian of the barrier $F_\square(x_1, \ldots, x_h) := -\sum_i \ln \cos(x_i)$ (see Section 6.2, [29]). The map

$$\psi : (x_1, \ldots, x_h) \mapsto \big(\ln(\sec(x_1) + \tan(x_1)), \ldots, \ln(\sec(x_h) + \tan(x_h))\big),$$

maps the cube with the Hessian metric isometrically onto Euclidean space, and the push-forward of the uniform density on the cube is a density $\phi \otimes \cdots \otimes \phi$ on $\mathbb{R}^h$, where

$$\phi(x) = \frac{\cos(2 \arctan((e^x - 1)/(e^x + 1)))}{\pi},$$

and

$$\frac{d^2 \ln \phi}{dx^2} = -\frac{4e^{2x}}{(1 + e^{2x})^2} < 0,$$

and the density is even (thus meeting the conditions of Theorem 11). In the 1-dimensional case, it is easy to check that the barrier is 1-self-concordant (Section 6.2, [29]). Therefore, in the 1-dimensional case, I_ϕ is bounded above and below by fixed constants. This, together with Theorem 8 implies that for each $\mathcal{M}_i$ and the uniform measure μ_i (on K_i), the isoperimetric profile I_{μ_i} of $(\mathcal{M}_i, \mu_i)$ satisfies $I_{\mu_i} \geq \Omega(\frac{1}{\sqrt{\nu}}) I_\phi$. Now applying Theorems 11 and 10 in succession, we see that

$$I_{\mu_1 \otimes \cdots \otimes \mu_h} \geq \Omega\left(\frac{1}{\sqrt{\nu}}\right) I_{\phi \otimes \cdots \otimes \phi} \geq \Omega\left(\frac{1}{\sqrt{\nu}}\right) I_\phi = \Omega\left(\frac{1}{\sqrt{\nu}}\right) \mathbf{1},$$

where $\mathbf{1}$ is the constant function taking the value 1. Therefore,

$$\beta_{\mathcal{M}} = \Omega\left(\frac{1}{\sqrt{\nu}}\right).$$

9. Analysis of Las Vegas algorithm.

PROOF OF THEOREM 3.   Note that this Markov chain is not ergodic and has no stationary probability distribution. We will analyze its behavior up to time t by relating this to the limiting behavior of Dikin walks on a family of convex sets $\{\hat{K}^j\}_{j \geq 1}$ each contained in the next, such that $\hat{K}^j = \hat{K} \cap \{x | c^T x \leq j\}$. Note that for any fixed j, our mixing results from Theorem 1 apply since $\hat{K}^j$ is bounded. Let $\hat{F}_j = \hat{F} + \ln(j - c^T x)$. By known properties of barriers [27], the self-concordance parameter of $\hat{F}_j$ is at most 1 more than that of $\hat{F}$. As j tends to ∞, $D^2 \hat{F}_j$ converges uniformly to $D^2 \hat{F}$ in the $\ell_2^n \to \ell_2^n$ operator norm on Hessian matrices on any compact subset of $\hat{K}$. Therefore, for any t, the distribution of the t-tuple $(T(x_0), T(x_1), \ldots, T(x_t))$ is the limit in total variation distance of the distributions of t-tuples $(T(x_0^j), T(x_1^j), \ldots, T(x_t^j))$, where $(x_0^j, \ldots, x_t^j)$ is a random walk on $\hat{K}^j$ starting at 0.

We will now give an upper bound for $\mathbb{P}[c^T(x_i) \leq 1 - \varepsilon]$. Let $y_i := T(x_i)$. Let e be the first time the Markov chain $y_0, y_1, \ldots$ escapes y_0. Thus e is an integer valued random variable defined by the event that y_e is the first point in $y_1, \ldots$ that is not equal to y_0. Let the density of y_e be ρ_e. $\forall t$, we know from Lemma 6 (applied when $y \to x$) that $\mathbb{P}[e \geq t + 1 | e \geq t] \leq 1 - \Omega(1)$, therefore,

$$\sup_x \frac{\rho_e(x)}{\hat{G}_0(x)} = O(1).$$

Without loss of generality, in the rest of this proof, we assume that for all v, $D^2 \hat{F}(0)[v, v] = \|v\|^2$.

Therefore,

$$\int_{\hat{K}} \rho_e(x)^2 \, dx = O\left(\int_{\mathbb{R}^n} \hat{G}_0(x)^2 \, dx\right) \tag{12}$$

$$= \exp(O(n \ln(n))). \tag{13}$$

Let $\rho_t(x)$ be the density of x_t. Then, by Theorem 1 and the fact that as far as total variation distance is concerned, a random walk on $\hat{K}$ can be viewed as the limit of random walks on the $\hat{K}^j$ as $j \to \infty$,

$$\frac{\int_{\hat{K}} \rho_t(x)^2 \, dx}{\int_{\hat{K}} \rho_e(x)^2 \, dx} = O\left(\left(1 - \frac{\Phi^2}{2}\right)^t\right). \tag{14}$$

By Lemma 10, this is less than $\exp(-t\Omega(\beta_{\text{fat}}^2))$, using the fact that $1 - x \leq e^{-x}$.

The following lemma is proved in the Appendix.

LEMMA 12.   Let $\hat{K}_\varepsilon = T(K \cap \{c^T x \leq 1 - \varepsilon\})$. Let ρ be a density supported on $\hat{K}$ such that

$$\int_{\hat{K}_\varepsilon} \rho(x) \, dx \geq \delta.$$

Then,

$$\int_{\hat{K}} \rho(x)^2 \, dx \geq \delta^2 \exp\left(-O\left(n \ln\left(\frac{s\nu}{\varepsilon}\right)\right)\right).$$

Using (13), (14), and Lemma 12, if $\mathbb{P}[c^T x_t \leq 1 - \varepsilon] = \delta$, we have that there exists an absolute constant $C > 0$ such that

$$\frac{\delta^2 \exp(-C(n \ln(s\nu/\varepsilon)))}{\exp(C(n \ln n))} < e^{-t(\beta_{\text{fat}})^2/C}.$$

Therefore,

$$\mathbb{P}[c^T x_t \leq 1 - \varepsilon] \leq \exp\left(C\left(n \ln\left(\frac{ns\nu}{\varepsilon}\right)\right) - (t(\beta_{\text{fat}})^2)/C\right).$$

This gives the desired upper bound on $\mathbb{P}[c^T x_t \leq 1 - \varepsilon]$. We next proceed to get an expression for $\tau(\varepsilon, \delta)$. We have

$$\mathbb{P}\left[(\exists t \geq \tau) c^T x_\tau \leq 1 - \varepsilon\right] \leq \sum_{t \geq \tau} \exp\left(C\left(n \ln\left(\frac{ns\nu}{\varepsilon}\right)\right) - t(\beta_{\text{fat}}^2)/C\right)$$

$$\leq \frac{\exp(C(n \ln((ns\nu)/\varepsilon)) - \tau(\beta_{\text{fat}}^2)/C)}{1 - \exp(-(\beta_{\text{fat}})^2/C)}.$$

Therefore, for any δ, $\mathbb{P}[(\exists t \geq \tau) c^T x_\tau \leq 1 - \varepsilon] < \delta$ for

$$\tau > C'\left(\frac{1}{(\beta_{\text{fat}})^2}\left(\ln\frac{1}{\delta} + \left(n \ln\frac{ns\nu}{\varepsilon}\right)\right)\right),$$

for a universal constant C'. Together with the fact that $\beta_{\text{fat}} = \Omega(\frac{1}{\sqrt{n\nu}})$ (from Theorem 8) this completes the proof. $\square$

10. Implementing the barrier oracle in the linear and semi-definite cases.
The most frequently encountered barrier functions are the logarithmic barrier for polytopes and the log det barrier for convex sets defined by semi-definite constraints (see Section 3). We discuss the implementation of the barrier oracle for the logarithmic barrier below, in the case where x is in the set. Let K_ℓ be the set of points satisfying the system of inequalities $Ax \leq \mathbf{1}$. Then $H(x) = A^T D(x)^2 A$ where $D(x)$ is the diagonal matrix whose ith diagonal entry $d_{ii}(x) = \frac{1}{1 - a_i^T x}$.

By the results of Baur and Strassen [3], the complexity of solving linear equations and of computing the determinant of an $n \times n$ matrix is $O(n^\gamma)$ where $\gamma < 2.377$ is the exponent for matrix multiplication. The computation of $A^T D(x)^2 A$ can be achieved using $mn^{\gamma-1}$ arithmetic operations, by partitioning a padded extension of $A^T D$ into $\leq \frac{m+n-1}{n}$ square matrices. Thus, the complexity of the barrier oracle is $O(mn^{\gamma-1})$ arithmetic operations.

In case the convex set is defined by a semi-definite constraint of rank ν, the log det barrier is a hyperbolic barrier and has a self-concordance parameter of ν

and a *complexity parameter* [defined in (1)] of $n\nu$. the number of arithmetic steps needed for computing the Hessian of the log det barrier is $O(n^2\nu^2 + n\nu^\gamma)$ (see Section 11.3, [24]. We have replaced an exponent 3 in [24] with γ). Given the Hessian, it can be inverted in (n^γ) arithmetic steps. This is needed to implement one step of the Dikin walk. Suppose K, as above, is a convex set that is defined by m linear constraints and additionally, semi-definite constraints of total rank ν_h [which can be as low as $O(1)$, e.g., for the intersection of a constant number of ellipsoids]. Then, the number of arithmetic steps for implementing one Dikin step is

$$O(mn^{\gamma-1} + n^2\nu_h^2 + n\nu_h^\gamma).$$

For Hit-and-Run, the number of arithmetic steps needed to make one move in a naive implementation is

$$O^*(\log(R/r)(mn + n\nu_h^2 + \nu_h^\gamma))$$

(since the natural way of certifying positive semi-definiteness is to take a Cholesky factorization, which has a complexity $O(\nu_h^\gamma)$, computing the new semi-definite matrix after one step has a complexity $n(\nu_h^2)$ (Section 11.3, [24]) and testing containment in the region defined by linear constraints takes $O(nm)$ operations). We see that:

1. If $m < n\nu_h^2 + \nu_h^\gamma$, then the ratio between the number of arithmetic steps for one move of Dikin walk and one move of Hit-and-Run is not more than $O^*(n)$.

2. If $m \geq n\nu_h^2 + \nu_h^\gamma$, then the ratio between the number of arithmetic steps for one move of Dikin walk and one move of Hit-and-Run is not more than $O^*(n^{\gamma-2}) < O^*(n^{0.38})$.

Combining the arithmetic complexity of implementing one step of Hit-and-Run with the mixing time, the ratio between the number of arithmetic steps needed to produce one random point using Dikin walk to the number of arithmetic steps needed for producing one random point using Hit-and-Run is $O^*(\frac{(m+\nu_h n)r^2}{R^2})$ if $m < n\nu_h^2 + \nu_h^\gamma$ and $O^*(\frac{(m+\nu_h n)r^2}{R^2 n^{0.62}})$ if $m \geq n\nu_h^2 + \nu_h^\gamma$.

10.1. *Implementing one Dikin step.* If K is defined by semi-definite constraints, from a point $x \in K$, one step for Hit-and-Run requires $\Omega(\log(R/d))$ membership operations, each of which requires testing the semi-definiteness of a $\nu \times \nu$ matrix [which takes $O(\nu^\gamma)$ arithmetic steps], where R is the radius of a circumscribing ball, and d is the distance of x to the boundary of K. Convex sets defined by semi-definite programs can be very ill-conditioned, and the best possible a priori upper bound on $\log \frac{R}{d}$ is not less than e^L where L is the total bit-length of rational data defining K and the point [30]. In the general setting, the number of arithmetic operations needed for implementing a Dikin step would be independent of R/r, but would depend on two affine-invariant quantities—the parameter

associated with the barrier and $\log s$, where the starting point is s-central. In ill-conditioned semi-definite programs, $\log s$ can be exponential in the bitlength, but for special points it can be much smaller; for example, for the center of mass and or the analytic center, it is $O(\log n)$ and $O(\log \nu)$, respectively.

11. Concluding remarks. We developed randomized analogues of barrier-based interior point methods, and demonstrated their use in sampling convex sets and optimizing a linear function over a convex set. One potential application of these methods is to integer programming as shown by Huang and Mehrotra [10]. It remains to be seen whether a more efficient algorithm for computing the volume of a polytope can be constructed using Dikin walk.

APPENDIX: PROOFS OF LEMMAS

PROOF OF LEMMA 7. Let $\breve{F} := F - F_\ell$. For $w \in K$, let $X(w) := D^2 F_\ell(w) - D^2 F_\ell(0)$. By Lemma 12 in [12], for any point $w \in K$, the gradient of $\mathbf{Tr} X$ at w measured using $\|\ell \cdot \|_w$ is $\leq 2\sqrt{n}$. Therefore, the gradient of $\mathbf{Tr} X$ at 0 measured using $\|\ell \cdot \|$ is $\leq O(\sqrt{n})$. $X(0) = 0$, therefore, $|\mathbf{Tr} X(y)| \leq O(\|y\|\sqrt{n}) = O(1)$, since $\|y\| = O(1/\sqrt{n})$. For the same reason, $\|X(y)\| = O(\sup_{\|w\| \leq \|y\|} \|D^3 F_\ell(w)[y, \cdot, \cdot]\|) = O(1/\sqrt{n})$.

For $w \in K$, let $Y(w) := D^2 \breve{F}(w) - D^2 \breve{F}(0)$. Then

$$\|Y(y)\| = D^2 \breve{F}(y) - D^2 \breve{F}(0) = O\left(\sup_{\|w\| \leq \|y\|} \|D^3 \breve{F}(w)[y, \cdot, \cdot]\| \right) = O(1/n).$$

Therefore, $|\mathbf{Tr} Y(y)| = O(1)$.

$$-V(y) = -\ln \det(I + X + Y)$$
$$= -\mathbf{Tr}(X + Y + R),$$

where R is a matrix whose $\|\cdot\| \to \|\cdot\|$ norm $\|R\|$ is bounded above by $O(\max(\|X\|^2, \|Y\|^2)) = O(1)$. Thus, $|V(y)| = O(1)$, and Lemma 7 is proved. $\square$

PROOF OF LEMMA 8. We need to show that

$$\mathbb{P}[-V(z) < O(1)] > \tfrac{9}{10}.$$

Let $\breve{F} := F - F_\ell$. For $w \in K$, let $W(w) := D^2 F_\ell(w)$, and let $Z(w) := D^2 \breve{F}(w)$.

$$V(z) = \tfrac{-1}{2} \ln \det(W(z) + Z(z))$$
$$= \tfrac{-1}{2} (\mathbf{Tr} \ln(W(z) + Z(z)))$$

$$\text{with probability} \geq 1 - 10^{-3} \ (\text{i.e., if } \|W(z) + Z(z) - I\| < 1)$$

$$= \tfrac{-1}{2}\big(\mathbf{Tr}\big(W(z) - W(0)\big) + \mathbf{Tr}\big(Z(z) - Z(0)\big)\big)$$

$$+ \tfrac{1}{4}\big(\mathbf{Tr}\big(W(z) + Z(z) - I\big)^2\big)$$

$$- O\big(\big|\mathbf{Tr}\big(I - (W(z) + Z(z))\big)^3\big|\big) \qquad \text{expanding by power series.}$$

This holds with probability $\geq 1 - 10^3$. The lemma is a consequence of the following lemmas.

LEMMA 13.
$$\mathbb{P}\big[-\big(\mathbf{Tr}\big(Z(z) - Z(0)\big)\big) \leq O(1)\big] \geq \tfrac{99}{100}.$$

LEMMA 14.
$$\mathbb{P}\big[\mathbf{Tr}\big(-\big(W(z) - W(0)\big)\big) \leq O(1)\big] \geq \tfrac{99}{100}.$$

LEMMA 15. *The following two probabilistic inequalities hold*:
$$\mathbb{P}\big[\big|\mathbf{Tr}\big(I - (W(z) + Z(z))\big)^3\big| \leq O(1)\big] \geq \tfrac{99}{100},$$
$$\mathbb{P}\big[\big|\mathbf{Tr}\big(I - (W(z) + Z(z))\big)^2\big| \leq O(1)\big] \geq \tfrac{99}{100}. \qquad \square$$

PROOF OF LEMMA 9. Since $\|y\| < O(\tfrac{1}{\sqrt{n}})$ and $\|z\| < \tfrac{1}{3}$ with probability greater than $1 - 10^{-3}$, $\|y\|_y$ and $\|y\|_z$ are $O(\tfrac{1}{\sqrt{n}})$. So it suffices to show that

$$\mathbb{P}\bigg[\max\big(\|z\|_y^2 - \|z\|^2, \langle y, z\rangle_y, \|z\|_z^2 - \|z\|^2, \langle y, z\rangle_z\big) < O\Big(\tfrac{1}{n}\Big)\bigg] > \frac{2}{10}.$$

This fact follows from the following three lemmas and the union bound. The proof of Lemma 16 would go through if $\tfrac{4}{10}$ were replaced by $\tfrac{1}{2} - \Omega(1)$.

LEMMA 16.
$$\mathbb{P}\bigg[\max\big(\|\ell z\|_y^2 - \|\ell z\|^2, \langle \ell y, z\rangle_y, \|\ell z\|_z^2 - \|\ell z\|^2, \langle \ell y, z\rangle_z\big) < O\Big(\tfrac{1}{n}\Big)\bigg] > \frac{4}{10}.$$

LEMMA 17.
$$\mathbb{P}\bigg[\max\big(\|h z\|_y^2 - \|h z\|^2, \langle h y, z\rangle_y, \|h z\|_z^2 - \|h z\|^2, \langle h y, z\rangle_z\big) < O\Big(\tfrac{1}{n^2}\Big)\bigg] > \frac{9}{10}.$$

LEMMA 18.
$$\mathbb{P}\bigg[\max\big(\|s z\|_y^2 - \|s z\|^2, \langle s y, z\rangle_y, \|s z\|_z^2 - \|s z\|^2, \langle s y, z\rangle_z\big) < O\Big(\tfrac{1}{n^3}\Big)\bigg] > \frac{9}{10}. \qquad \square$$

PROOF OF LEMMA 10.   Let S_1 be a measurable subset of K such that $\mu(S_1) \le \frac{1}{2}$ and $S_2 := K \setminus S_1$ be its complement. For any $x \ne y \in K$,

$$\frac{dP_y}{d\mu}(x) = \frac{dP_x}{d\mu}(y).$$

Let $S_1' = S_1 \cap \{x \mid P_x(S_2) = o(1)\}$ and $S_2' = S_2 \cap \{y \mid P_y(S_1) = o(1)\}$. By the reversibility of the chain, which is easily checked,

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

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

$$d_{\mathrm{TV}}(P_x, P_y) := 1 - \int_K \min\left(\frac{dP_x}{d\mu}(w), \frac{dP_y}{d\mu}(w)\right) d\mu(w) = 1 - o(1).$$

Lemma 6 states that for an absolute constant C, if $d(x, y) \le \frac{1}{C\sqrt{n}}$, then $d_{\mathrm{TV}}(P_x, P_y) = 1 - \Omega(1)$. Therefore, Theorem 8 implies that

$$\mu\left((K \setminus S_1') \setminus S_2'\right) \ge \Omega(\beta_{\mathrm{fat}}) \min(\mu(S_1'), \mu(S_2')).$$

First, suppose $\mu(S_1') \ge (1 - \Omega(1))\mu(S_1)$ and $\mu(S_2') \ge (1 - \Omega(1))\mu(S_2)$. Then

$$\int_{S_1} P_x(S_2) \, d\mu(x) \ge \mu\left((K \setminus S_1') \setminus S_2'\right)$$

$$\ge \Omega(\beta_{\mathrm{fat}})\mu(S_1')$$

$$\ge \Omega(\beta_{\mathrm{fat}}) \min(\mu(S_1'), \mu(S_2'))$$

and we are done. Otherwise, without loss of generality, suppose $\mu(S_1') \le (1 - \Omega(1))\mu(S_1)$. Then

$$\int_{S_1} P_x(S_2) \, d\mu(x) \ge \Omega(\mu(S_1))$$

and we are done. $\square$

PROOF OF LEMMA 12.   We remind the reader that we are given a convex set K containing the origin as an interior point and a linear objective c, such that

$$Q := K \cap \{y : c^T y \le 1\}$$

is bounded, for any chord $\overline{pq}$ of Q passing through the origin, $\frac{|p|}{|q|} \le s$ and $\varepsilon > 0$. Also, $T : Q \to \mathbb{R}^n$ is defined by

$$T(x) = \frac{x}{1 - c^T x}.$$

Let K_ε be $K \cap \{c^T x \le 1 - \varepsilon\}$ and $\hat{K}_\varepsilon := TK_\varepsilon$. Given four collinear points a, b, c, d, $(a : b : c : d) := \frac{(a-c)\cdot(b-d)}{(a-d)\cdot(b-c)}$ is called the the cross ratio. Let $\overline{p'q'} \ni 0$ be

a chord of $\hat{K}_\varepsilon$, and $p = T^{-1}(p')$ and $q = T^{-1}(q')$. If $c^T p' \leq c^T q'$, then $\frac{|p'|}{|q'|} \leq \frac{|p|}{|q|} \leq s$. On the other hand, if $c^T p' \geq c^T q'$, let r be the intersection of $\overline{pq}$ with $\{x : c^T x = 1\}$. By the projective invariance of the cross ratio (see, e.g., Lemma 14 in [12])

$$(\infty : 0 : p' : q') = (r : 0 : p : q).$$

Therefore

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

$$\leq (s)\left(\frac{s+1}{\varepsilon}\right).$$

Thus

(15)
$$\sup_{\overline{p'q'} \ni 0} \ln \frac{|p'|}{|q'|} = O\left(\ln \frac{s}{\varepsilon}\right),$$

where the supremum is taken over all chords of K containing the origin. By (15) and Theorems 5 and 6, it follows that

$$\sup_{h \in \hat{K}_\varepsilon} \ln \|h\| \leq O\left(\ln\left(\frac{s\nu}{\varepsilon}\right)\right),$$

and, therefore,

(16)
$$\ln \operatorname{vol} \hat{K}_\varepsilon \leq O\left(n \ln\left(\frac{s\nu}{\varepsilon}\right)\right).$$

Let ρ be a density supported on $\hat{K}$ such that

$$\int_{\hat{K}_\varepsilon} \rho(x)\, dx \geq \delta.$$

Then

$$\int_{\hat{K}} \rho(x)^2\, dx$$

$$\geq \int_{\hat{K}_\varepsilon} \rho(x)^2\, dx$$

$$\geq \frac{\delta^2}{\operatorname{vol} \hat{K}_\varepsilon} \qquad \text{by convexity of the squared function and Jensen's inequality}$$

$$\geq \delta^2 \exp\left(-O\left(n \ln\left(\frac{s\nu}{\varepsilon}\right)\right)\right) \qquad \text{by (16).} \qquad \square$$

PROOF OF LEMMA 13.   Let $Z_h(w) := D^2 F_h(w)$ and $Z_s = D^2 F_s(w)$. Then $Z(w) = n Z_h + n^2 Z_s$. Next,

$$(17) \qquad \mathbf{Tr} Z_h(z) - \mathbf{Tr} Z_h(0) = D \mathbf{Tr} Z_h(0)[z] + \frac{D^2 \mathbf{Tr} Z_h(z')[z,z]}{2},$$

for some $z' \in [0, z]$.

$$\left| D \mathbf{Tr} Z_h(0)[z] \right|$$

$$\leq O\left(n \left(\sup_{\|v\|=1} \left| D^3 F_h[v, v, v] \right| \right) \right) \qquad \text{with probability} \geq 1 - 10^{-3}$$

$$\leq O\left(n \left(\sup_{\|h v\|=1/\sqrt{n}} \left| D^3 F_h[v, v, v] \right| \right) \right) \qquad \text{with probability} \geq 1 - 10^{-3}$$

$$\leq O(1/\sqrt{n}) \qquad \text{with probability} \geq 1 - 10^{-3}.$$

Applying Lemma 4, we see that

$$(18) \qquad \mathbb{P}\left[-D \mathbf{Tr} Z_h(0)[z] < O(1/n) \right] > \tfrac{998}{1000}.$$

Next,

$$(19) \qquad D^2 \mathbf{Tr} Z(z')[z, z] = D^2 \mathbf{Tr} D^2 \left(n F_h + n^2 F_s \right)(z')[z, z].$$

In order to bound (19), let A be an invertible matrix such that $Z(z') = A^T Z(0) A$. Such a matrix A exists for which $\|A - I\| = O(1/\sqrt{n})$ with probability $\geq 1 - 10^{-3}$ because $\|h z'\| = O(1/\sqrt{n})$ with probability $\geq 1 - 10^{-3}$. Let D_A be the differential operator whose action on a function G is determined by the relation

$$\forall v \in \mathbb{R}^n, \qquad D_A G(w)[v] := DG(w)[Av].$$

Thus, $D_A^2 F_h(z') = Z_h(0)$. Now,

$$(20) \quad \|h u\| = O(1) \quad \Rightarrow \quad \forall \|h v\| = 1, \qquad D^4 F_h(z')[u, u, v, v] \leq O(1),$$

$$D^2 \left(\mathbf{Tr} Z_h(z') \right)[z, z]$$

$$= D^2 \left(\mathbf{Tr} D_A^2 F_h(0) \right)[z, z]$$

$$(21) \qquad \leq n \sup_{\|A v\|=1} O\left(D^4 F(z')[v, v, z, z] \right) \qquad \text{with probability} \geq 1 - 10^{-3}$$

$$\leq n^2 \sup_{\|h v\|=1/\sqrt{n}} D^4 F(z')[v, v, v, v] \|h z\|^2 \qquad \text{(by Fact 4)}$$

$$= O(1/n) \qquad \text{with probability} \geq 1 - 10^{-3}.$$

The last line here uses Lemma 1.

Therefore, by equations (17), (18) and (21), we have

$$\mathbb{P}\left[-\left(\mathbf{Tr} Z_h(z) - \mathbf{Tr} Z_h(0) \right) < O(1/n) \right] > \tfrac{997}{1000}.$$

Also with probability $\geq 1 - 10^{-3}$, $\|_s z\| = O(1/n)$. Therefore, with probability $\geq 1 - 10^{-3}$,

$$\mathbf{Tr}Z_s(z) - \mathbf{Tr}Z_s(0) = O(\mathbf{Tr}Z_s(0)\|_s z\|) \tag{22}$$

$$= O(1/n^2). \tag{23}$$

The statement follows from the last two sentences, since $Z = nZ_h + n^2 Z_s$. $\quad\square$

PROOF OF LEMMA 14.

$$-\mathbf{Tr}W(z) + \mathbf{Tr}W(0) = -\left(D\mathbf{Tr}W(0)[z] + \frac{D^2\mathbf{Tr}W(z')[z, z]}{2}\right), \tag{24}$$

for some $z' \in [0, z]$. Lemma 12 in [12] shows that $\|_\ell \nabla \mathbf{Tr}W\| \leq 2\sqrt{n}$. Since for all vectors v, $\|v\| \geq \|_\ell v\|$, this implies that $\|\nabla \mathbf{Tr}W\| \leq 2\sqrt{n}$. By Lemma 4, this implies that

$$\mathbb{P}[|D\mathbf{Tr}W(0)[z]| < O(1)] \geq 1 - 10^{-3}.$$

By Lemma 13 in [12], $\frac{-D^2\mathbf{Tr}W(z')}{2} \leq 0$, thereby completing the proof. $\quad\square$

PROOF OF LEMMA 15. Recall that $W(0) + Z(0) = I$.
In order to prove that

$$\mathbb{P}[|\mathbf{Tr}(I - (W(z) + Z(z)))^3| \leq O(1)] \geq \tfrac{99}{100},$$

it suffices to show that

$$\mathbb{P}[\|(W(z) - W(0))\| \leq O(n^{-1/3})] \geq 1 - 10^{-3} \tag{25}$$

and that

$$\mathbb{P}[\|(Z(z) - Z(0))\| \leq O(n^{-1/2})] \geq 1 - 10^{-3}. \tag{26}$$

From Lemma 5 and Fact 2, we obtain (25). We obtain (26) from (4).
In order to prove that

$$\mathbb{P}[|\mathbf{Tr}(I - (W(z) + Z(z)))^2| \leq O(1)] \geq \tfrac{99}{100},$$

it suffices to show that

$$\mathbb{P}[\mathbf{Tr}(W(z) - W(0))^2\| \leq O(1)] \geq 1 - 10^{-3} \tag{27}$$

and that

$$\mathbb{P}[\mathbf{Tr}(Z(z) - Z(0))^2 \leq O(1)] \geq 1 - 10^{-3}, \tag{28}$$

since by Cauchy–Schwarz,

$$|\mathbf{Tr}(W(z) - W(0))(Z(z) - Z(0))|^2 \leq \mathbf{Tr}((W(z) - W(0))^2)\mathbf{Tr}((Z(z) - Z(0))^2).$$

Inequality (28) follows from (26). We will prove (27) below. We have

$$\mathbf{Tr}\big((W(z) - W(0))^2\big)$$

$$= \mathbf{Tr}\bigg(\sum_{i \le m} a_i a_i^T \bigg(\frac{1}{(1 - a_i^T z)^2} - 1\bigg)\bigg)^2$$

$$\le \mathbf{Tr}\bigg(\sum_{i \le m} a_i a_i^T \big((2a_i^T z) + O(|a_i^T z|^2)\big)\bigg)^2$$

$$\le O\bigg(\mathbf{Tr}\bigg(\sum_i \sum_j (a_i a_i^T)(a_j a_j^T)\big((2a_i^T z)(2a_j^T z) + O(n^{-5/4})\big)\bigg)\bigg)$$

$$\le \sum_i \sum_j \langle a_i, a_j \rangle^2 \big((2a_i^T z)(2a_j^T z) + O(n^{-5/4})\big).$$

(The last three lines above are true with probability $> 1 - 10^{-4}$.) We proceed to obtain a bound on

$$\mathbb{E} \sum_i \sum_j \langle a_i, a_j \rangle^2 \big((2a_i^T z)(2a_j^T z) + O(n^{-5/4})\big)$$

and then the lemma follows from Markov's inequality:

$$\mathbb{E} \sum_i \sum_j \langle a_i, a_j \rangle^2 \big((2a_i^T z)(2a_j^T z) + O(n^{-5/4})\big)$$

is less than equal to

$$\sum_i \sum_j \langle a_i, a_j \rangle^2 \Big(\sqrt{\mathbb{E}\big((a_i^T z)^2\big)\mathbb{E}\big((a_j^T z)^2\big)} + O(n^{-5/4})\Big)$$

$$\le \sum_i \sum_j \langle a_i, a_j \rangle^2 \big(O(1/n)\big)$$

$$= O(1/n)\mathbf{Tr} \sum_i a_i a_i^T \sum_j a_j a_j^T$$

$$= O(1). \qquad \qquad \square$$

PROOF OF LEMMA 16.   Let $F_\ell(w) := -\sum_{i=1}^m \ln(1 - \langle a_i, v \rangle)$, for any $v \in K$. Thus

$$\bigg(\sum_{i=1}^m \langle v, a_i \rangle^2\bigg) = \|\ell v\|^2 \le \|v\|^2.$$

Fixing an orthonormal basis with respect to $\langle \cdot, \cdot \rangle$,

$$\sum_{i=1}^m a_i a_i^T \preceq I,$$

where $X \preceq Y$ signifies that Y dominates X in the semi-definite cone.

Recall that for any v such that $\|v\| = 1$, $\mathbb{E}(\langle v, z \rangle^2) = r^2 < 1/C$ for some sufficiently large constant C. It suffices to prove the following two inequalities.

LEMMA 19.
$$\mathbb{P}\left[\max(\|\ell z\|_y^2 - \|\ell z\|^2, \langle \ell y, z \rangle_y, \langle \ell y, z \rangle_z) < O\left(\frac{1}{n}\right)\right] > \frac{19}{20}.$$

LEMMA 20.
$$\mathbb{P}\left[\|\ell z\|_z^2 - \|\ell z\|^2 < O\left(\frac{1}{n}\right)\right] > \frac{9}{20}.$$
□

PROOF OF LEMMA 17. We will prove upper bounds on each of (a) $\|hz\|_y^2 - \|hz\|^2$, (b) $\langle y, z \rangle_y$, (c) $\|hz\|_z^2 - \|hz\|^2$ and (d) $\langle hy, z \rangle_z$ that hold with constant probability, and then use the union bound. We will repeatedly use the observation (that holds from Fact 1) that for any point w such that $\|hw\| = o(1)$,

(29)
$$\|hy\|_w \leq O(n^{-1/2}\|y\|) \leq O\left(\frac{1}{n}\right)$$

and with probability $\geq 1 - 10^{-3}$

(30)
$$\|hz\|_w \leq O(n^{-1/2}\|z\|) \leq O(1/\sqrt{n}).$$

(a)
$$\|hz\|_y^2 - \|hz\|^2 = D^2 F_h(y)[z, z] - D^2 F_h(0)[z, z]$$
$$= D^3 F_h[w][y, z, z],$$

for some w on the line segment $[0, y]$.

By Fact 4,
$$\mathbb{P}[D^3 F_h(w)[y, z, z] < O(n^{-2})] \geq \mathbb{P}\left[\|hy\|_w(\|hz\|_w)^2 \leq O\left(\frac{1}{n^2}\right)\right].$$

Since $\|z\|^2 = \|\ell z\|^2 + n\|hz\|^2 + n^2\|sz\|^2$, we have
$$\|hz\|_w = O(\|z\|_w/\sqrt{n}).$$

Also, $\|hw\| = O(\|y\|/\sqrt{n}) = O(1/n)$, and so with probability $\geq 1 - 10^{-3}$,
$$O(\|hz\|_w) = O(\|z\|/\sqrt{n}) = O(1/\sqrt{n}).$$

Thus,

(31)
$$\mathbb{P}\left[\|hy\|_w\|hz\|_w \leq O\left(\frac{1}{n\sqrt{n}}\right)\right] \geq 1 - 10^{-3}$$

and

(32)
$$\mathbb{P}\left[\|hy\|_w(\|hz\|_w)^2 \leq O\left(\frac{1}{n^2}\right)\right] \geq 1 - 10^{-3}.$$

(b)

$$\langle_h y, z \rangle_y = \langle_h y, z \rangle + \left(\langle_h y, z \rangle_y - \langle_h y, z \rangle \right)$$

$$(33) \qquad\qquad = \langle_h y, z \rangle + D^3 F_h(w)[y, y, z] \qquad (\text{for } w \in [0, y])$$

$$(34) \qquad\qquad = O\left(\frac{\|_h y\|}{n} \right) + O\left(\|_h y\|_w^2 \|_h z\|_w \right) \qquad \text{with probability} > 99/100.$$

In going from (33) to (34), we used Facts 4 and 5. In the above calculation, to ensure that w is well-defined, we take it to be the candidate with the least norm. Thus by equations (31), (32) and (34),

$$\mathbb{P}\left[\left(\langle_h y, z \rangle_y \right) < O\left(\frac{1}{n} \right) \right] \geq \mathbb{P}\left[\frac{\|_h y\| \|_h z\|}{\sqrt{n}} + \|_h y\|_w^2 \|_h z\|_w = O\left(\frac{1}{n} \right) \right]$$

$$> 98/100.$$

(c) For some $w \in [0, z]$,

$$(35) \qquad\qquad \left| \|_h z\|_z^2 - \|_h z\|^2 - D^3 F_h(0)[z, z, z] \right|$$

$$(36) \qquad\qquad \leq \sup_{w \in [0, y]} \left| \frac{D^4 F_h(w)[z, z, z, z]}{2} \right|.$$

By Lemma 4,

$$\mathbb{P}\left[D^3 F_h(0)[z, z, z] = O\left(\frac{\sup_{\|_h v\| \leq 1} D^3 F_h(0)[v, v, v] \|_h z\|^3}{\sqrt{n}} \right) \right] > 99/100.$$

$\mathbb{P}[\|_h z\| = O(1/\sqrt{n})] > 99/100$, therefore, each term in (36) is $O(1/n^2)$ with probability $\frac{99}{100}$ by Lemma 1, and so

$$\mathbb{P}\left[\|_h z\|_z^2 - \|_h z\|^2 < O\left(\frac{1}{n^2} \right) \right] > \frac{98}{100}.$$

(d)

$$\langle_h y, z \rangle_z = \langle_h y, z \rangle + \left(\langle_h y, z \rangle_z - \langle_h y, z \rangle \right)$$

$$\leq \langle_h y, z \rangle + \sup_{w \in [0, z]} \left| D^3 F_h(w)[y, z, z] \right|$$

$$= O\left(\frac{\|_h y\|}{n} \right) + 2 \|_h y\|_w \|_h z\|_w^2 \qquad \text{with probability} > 99/100.$$

By equations (29) and (30),

$$\mathbb{P}\left[\frac{\|_h y\|}{n} + \|_h y\|_w \|_h z\|_w^2 = O\left(\frac{1}{n^2} \right) \right] > 99/100.$$

Therefore,

$$\mathbb{P}\left[\left(\langle_h y, z \rangle_z \right) < O\left(\frac{1}{n^2} \right) \right] > \frac{98}{100}. \qquad\qquad \square$$

PROOF OF LEMMA 18. We trace the same steps involved in the proof of the last lemma, the only difference being that of scale. We proceed to prove upper bounds of $O(1/n^3)$ on each of the terms (a) $\|_s z\|_y^2 - \|_s z\|^2$, (b) $\langle_s y, z\rangle_y$, (c) $\|_s z\|_z^2 - \|z\|^2$ and (d) $\langle_s y, z\rangle_z$ that hold with constant probability separately, and then use the union bound. We will repeatedly use the observation (that holds from Fact 1) that for any point w such that $\|_s w\| = o(1)$,

$$\|_s y\|_w \leq O(n^{-1}\|y\|) \leq O\left(\frac{1}{n\sqrt{n}}\right) \tag{37}$$

and with probability $\geq 1 - 10^{-3}$

$$\|_s z\|_w \leq O(n^{-1}\|z\|) \leq O(1/n). \tag{38}$$

(a)

$$\|_s z\|_y^2 - \|_s z\|^2 = D^2 F_s(y)[z, z] - D^2 F_s(0)[z, z]$$
$$= D^3 F_s(w)[y, z, z],$$

for some w on the line segment $[0, y]$.

$$D^3 F_s(w)[y, z, z] \leq \|_s y\|_w \|_s z\|_w^2 \qquad \text{by Fact 4}$$
$$\leq (n^{-3/2})(n^{-1})^2 \qquad \text{with probability} > 1 - 10^{-3}.$$

(b)

$$\langle_s y, z\rangle_y = \langle_s y, z\rangle + (\langle_s y, z\rangle_y - \langle_s y, z\rangle)$$
$$= \langle_s y, z\rangle + D^3 F_s(w)[y, y, z] \tag{39}$$
$$= O\left(\frac{\|_s y\|}{n\sqrt{n}}\right) + O(\|y\|_w^2 \|z\|_w) \qquad \text{with probability} > 99/100. \tag{40}$$

In going from (39) to (40), we used Facts 4 and 5. We see that

$$\mathbb{P}\left[\frac{\|_s y\|}{n\sqrt{n}} + 2\|_s y\|_w^2 \|_s z\|_w = O\left(\frac{1}{n^3}\right)\right] > 99/100.$$

Therefore,

$$\mathbb{P}\left[(\langle_s y, z\rangle_y) < O\left(\frac{1}{n^3}\right)\right] > \frac{98}{100}.$$

(c)

$$\|_s z\|_z^2 - \|_s z\|^2 = D^2 F_s(z)[z, z] - D^2 F_s(0)[z, z]$$
$$= D^3 F_s(w)[z, z, z]$$

for some w on the line segment $[0, z]$. By Fact 4,

$$D^3 F_s(w)[z, z, z] \leq 2\|_s z\|_w \|_s z\|_w^2$$

$$\leq O\left(\frac{1}{n^3}\right) \qquad \text{with probability} \geq 1 - 10^{-3}.$$

(d)

$$\langle_s y, z \rangle_z = \langle_s y, z \rangle + (\langle_s y, z \rangle_z - \langle_s y, z \rangle)$$

$$\leq \langle_s y, z \rangle + \sup_{w \in [0, z]} D^3 F_s(w)[y, z, z]$$

$$= O\left(\frac{\|_s y\|}{n\sqrt{n}}\right) + 2\|_s y\|_w \|_s z\|_w^2 \qquad \text{with probability} > 99/100.$$

By equations (37) and (38),

$$\mathbb{P}\left[\frac{\|_s y\| \|_s z\|}{\sqrt{n}} + 2\|_s y\|_w \|_s z\|_w^2 = O\left(\frac{1}{n^3}\right)\right] > 99/100.$$

Therefore,

$$\mathbb{P}\left[(\langle_s y, z \rangle_z) < O\left(\frac{1}{n^3}\right)\right] > \frac{98}{100}. \qquad \square$$

PROOF OF LEMMA 19.   (a)

$$\|_\ell z\|_y^2 - \|_\ell z\|^2 = D^2 F_\ell(y)[z, z] - D^2 F_\ell(0)[z, z]$$

$$= D^3 F_\ell(w)[y, z, z],$$

for some $w \in [0, y]$, and consequently $\|_\ell w\| = O(1/\sqrt{n})$, and hence

$$D^3 F_\ell(w)[y, z, z] = D^3 F_\ell(0)[y, z, z] + D^4 F_\ell(w_1)[y, y, z, z].$$

The term $D^4 F_\ell(w_1)[y, y, z, z]$ is bounded above $O(1/n)$ with probability $1 - 10^{-3}$. We apply Cauchy–Schwarz below:

$$\mathbb{E}\left[\left(\sum_{i=1}^m (y^T a_i a_i^T z)(a_i^T z)\right)^2\right] \leq \mathbb{E}\left(\sum_{i=1}^m (y^T a_i a_i^T z)^2\right) \mathbb{E}\left(\sum_i (a_i^T z)^2\right)$$

(41)

$$\leq \mathbb{E}\left(\sum_i \|(a_i a_i^T) y\|^2 \|z\|^2 / n\right) \mathbb{E}(\|z\|^2)$$

$$= O(1/n^2), \qquad \text{since } \left\|\sum_i a_i a_i^T\right\| \leq 1.$$

Therefore,

$$\mathbb{P}[D^3 F_\ell(0)[y, z, z] < O(1/n)]$$

$$= \mathbb{P}\left[2 \sum_{i=1}^{m} (a_i^T y)(a_i^T z)^2 < O(1/n)\right]$$

(42)
$$\geq 1 - O\left(n^2 \mathbb{E}\left[\left(\sum_{i=1}^{m} (y^T a_i a_i^T z)(a_i^T z)\right)^2\right]\right)$$

(43)
$$\geq 1 - 10^{-3} \qquad \text{rescaling } y \text{ by a universal constant.}$$

(b) Proceeding to the next term,

$$\mathbb{P}[\langle \ell y, z \rangle_y = O(1/n)] = \mathbb{P}[\langle \ell y, z \rangle + (\langle \ell y, z \rangle_y - \langle \ell y, z \rangle) = O(1/n)].$$

Also,

$$\langle \ell y, z \rangle_y - \langle \ell y, z \rangle = O(\|z\|/n)$$

by (4), and

(44)
$$\mathbb{E}[\langle \ell y, z \rangle^2] \leq O(1/n^2),$$

so by Markov's inequality we obtain

(45)
$$\mathbb{P}[\langle \ell y, z \rangle_y = O(1/n)] > 1 - 10^{-3}.$$

(c) Finally, we obtain a probabilistic upper bound

(46)
$$\mathbb{P}[\langle \ell y, z \rangle_z - \langle \ell y, z \rangle < O(1/n)] > 1 - 10^{-3},$$

as follows. Note that

$$\langle \ell y, z \rangle_z - \langle \ell y, z \rangle = y^T \sum_i \left(\frac{a_i a_i^T z}{(1 - a_i^T z)^2} - a_i^T z\right).$$

This is equal to

(47)
$$\sum_i y a_i^T \left(\frac{2(a_i^T z)^2 - (a_i^T z)^3}{(1 - a_i^T z)^2}\right).$$

Let the a_i be listed in order of non-increasing length. For $i = 1$ to n, the probability that $|a_i^T z| \geq n^{-1/4}$ is $O(e^{-\sqrt{n}/2})$. For $i > n$, $|a_i^T z| \geq n^{-1/4}$ is $O(e^{-\sqrt{i}/2})$. It is true with probability $1 - 10^{-3}$ that $\forall_i |a_i^T z|$ is $\leq \|a_i^T\| r$, which is less than $\frac{1}{2}$. Therefore, with probability $> 1 - 10^{-3}$, every term $\frac{2(a_i^T z)^2 - (a_i^T z)^3}{(1 - a_i^T z)^2}$ is bounded above by $3(a_i^T z)^2$:

$$\sum_i y a_i^T \left(\frac{2(a_i^T z)^2 - (a_i^T z)^3}{(1 - a_i^T z)^2}\right) < \|y\| \left\|\sum_i a_i \left(\frac{2(a_i^T z)^2 - (a_i^T z)^3}{(1 - a_i^T z)^2}\right)\right\|.$$

Next, we apply the semi-definite Cauchy–Schwarz inequality from [12] and take operator norms on both sides. With probability $> 1 - 10^{-3}$,

$$\left\| \sum_i a_i \left(\frac{2(a_i^T z)^2 - (a_i^T z)^3}{(1 - a_i^T z)^2} \right) \right\|^2 \le \left\| \sum_i a_i a_i^T \right\| \sum_i 9(a_i^T z)^4 \le O\left(\sum_i (a_i^T z)^4 \right).$$

Next

$$\mathbb{E}\left[\sum_i (a_i^T z)^4 \right] = O\left(\frac{\sum_i \|a_i\|^4}{n^2} \right) = O\left(\frac{1}{n} \right).$$

The last step uses the fact that each $\|a_i\| \le 1$ and $\sum_i \|a_i\|^2 \le n$. Thus

$$\mathbb{P}\left[\sum_i y a_i^T \left(\frac{2(a_i^T z)^2 - (a_i^T z)^3}{(1 - a_i^T z)^2} \right) < O(1/n) \right] > 1 - 10^{-3}.$$

It follows that

(48) $$\mathbb{P}\left[\langle \ell y, z \rangle_z - \langle \ell y, z \rangle = O(1/n) \right] > 1 - 10^{-3}.$$

Lemma 19 follows. $\square$

PROOF OF LEMMA 20.   In order to prove that

$$\mathbb{P}\left[\|\ell z\|_z^2 - \|\ell z\|^2 < O\left(\frac{1}{n} \right) \right] > \frac{9}{20},$$

it suffices to show that

$$\mathbb{P}\left[(\|\ell z\|_z^2 + \|\ell z\|_{-z}^2)/2 < \|\ell z\|^2 + O\left(\frac{1}{n} \right) \right] > \frac{9}{10},$$

because the distribution of z is symmetric about the origin.

(49)
$$\begin{aligned}
\sum_i \left(\frac{(a_i^T z)^2}{2(1 - a_i^T z)^2} + \frac{(a_i^T z)^2}{2(1 + a_i^T z)^2} \right) &= \sum_i (a_i^T z)^2 \left(\frac{1 + (a_i^T z)^2}{(1 - (a_i^T z)^2)^2} \right) \\
&= \sum_i \left((a_i^T z)^2 + \frac{3(a_i^T z)^4 - (a_i^T z)^6}{(1 - (a_i^T z)^2)^2} \right) \\
&= \|\ell z\|^2 + \sum_i \frac{3(a_i^T z)^4 - (a_i^T z)^6}{(1 - (a_i^T z)^2)^2}.
\end{aligned}$$

Let the a_i be listed in order of nonincreasing length. For $i = 1$ to n, the probability that $|a_i^T z| \ge n^{-1/4}$ is $O(e^{-\sqrt{n}/2})$. For $i > n$, $|a_i^T z| \ge n^{-1/4}$ is $O(e^{-\sqrt{i}/2})$. It is true with probability $1 - 10^{-3}$ that $\forall_i |a_i^T z|$ is $\le \|a_i^T\| r$, which is less than $\frac{1}{2}$. This allows us to write, with probability $\ge 1 - 10^{-3}$,

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

which is $O(\|_\ell a_i\|^4/n^2)$. Since $\sum_i a_i a_i^T \preceq I$, therefore, $\forall i, \|_\ell a_i\| \leq 1$ and $\sum_i \|_\ell a_i\|^2 \leq n$.

Therefore,

$$\mathbb{P}\left[(\|_\ell z\|_z^2 + \|_\ell z\|_{-z}^2)/2 > \|_\ell z\|^2 + O\left(\frac{100}{n}\right) \right]$$

$$= \mathbb{P}\left[\sum_i \frac{3(a_i^T z)^4 - (a_i^T z)^6}{(1 - (a_i^T z)^2)^2} > O\left(10^2/n\right) \right]$$

$$\leq n \sum_i \|_\ell a_i\|^4/(100n^2)$$

$$\leq \sum_i \|_\ell a_i\|^2/(100n)$$

$$\leq 1/100. \qquad \square$$

REFERENCES

[1] BALL, K. M. (1997). An elementary introduction to modern convex geometry. In *Flavors of Geometry* (S. Levy, ed.). *Math. Sci. Res. Inst. Publ.* **31** 1–58. Cambridge Univ. Press, Cambridge. MR1491097

[2] BARTHE, F. (2002). Log-concave and spherical models in isoperimetry. *Geom. Funct. Anal.* **12** 32–55. MR1904555

[3] BAUR, W. and STRASSEN, V. (1983). The complexity of partial derivatives. *Theoret. Comput. Sci.* **22** 317–330. MR0693063

[4] BERTSIMAS, D. and VEMPALA, S. (2004). Solving convex programs by random walks. *J. ACM* **51** 540–556. MR2147847

[5] BOBKOV, S. G. and HOUDRÉ, C. (1997). Isoperimetric constants for product probability measures. *Ann. Probab.* **25** 184–205. MR1428505

[6] DYER, M., FRIEZE, A. and KANNAN, R. (1991). A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM* **38** 1–17. MR1095916

[7] FREUND, R. M. and VERA, J. R. (2009). Equivalence of convex problem geometry and computational complexity in the separation oracle model. *Math. Oper. Res.* **34** 869–879. MR2573500

[8] GROMOV, M. (2003). Isoperimetry of waists and concentration of maps. *Geom. Funct. Anal.* **13** 178–215. MR1978494

[9] GÜLER, O. (1997). Hyperbolic polynomials and interior point methods for convex programming. *Math. Oper. Res.* **22** 350–377. MR1450796

[10] HUANG, K. and MEHROTRA, S. (2013). An empirical evaluation of walk-and-round heuristics for mixed-integer linear programs. *Comput. Optim. Appl.* **55** 545–570.

[11] KANNAN, R., LOVÁSZ, L. and SIMONOVITS, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms* **11** 1–50. MR1608200

[12] KANNAN, R. and NARAYANAN, H. (2009). Random walks on polytopes and an affine interior point method for linear programming. In *STOC'09—Proceedings of the* 2009 *ACM International Symposium on Theory of Computing* 561–570. ACM, New York. MR2780102

[13] KANNAN, R. and VEMPALA, S. (1999). Sampling lattice points. In *STOC'97* (*El Paso*, *TX*) 696–700. ACM, New York. MR1753411

[14] KARMARKAR, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica* **4** 373–395. MR0779900

[15] KARMARKAR, N. K. (1990). Riemannian geometry underlying interior point methods. In *Mathematical Developments Arising from Linear Programming. Contemporary Mathematics* **144** 51–75. Amer. Math. Soc., Providence, RI. MR1097865

[16] LEBEAU, G. and MICHEL, L. (2010). Semiclassical analysis of a random walk on a manifold. *Ann. Probab.* **38** 277–315.

[17] LOVÁSZ, L. (1999). Hit-and-run mixes fast. *Math. Program.* **86** 443–461. MR1733749

[18] LOVÁSZ, L. and SIMONOVITS, M. (1993). Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms* **4** 359–412. MR1238906

[19] LOVÁSZ, L. and VEMPALA, S. (2006). Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comput. System Sci.* **72** 392–417. MR2205290

[20] LOVÁSZ, L. and VEMPALA, S. (2006). Hit-and-run from a corner. *SIAM J. Comput.* **35** 985–1005 (electronic). MR2203735

[21] NARAYANAN, H. (2009). Diffusion in computer science and statistics. Ph.D thesis, Univ. Chicago. MR2713584

[22] NARAYANAN, H. and NIYOGI, P. (2008). Sampling hypersurfaces through diffusion. In *Approximation*, *Randomization and Combinatorial Optimization. Lecture Notes in Computer Science* **5171** 535–548. Springer, Berlin. MR2538813

[23] NARAYANAN, H. and RAKHLIN, A. (2010). Random walk approach to regret minimization. In *Advances in Neural Information Processing Systems* (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta, eds.) **23** 1777–1785. Curran Associates, Red Hook, NY.

[24] NEMIROVSKI, A. (1994). Interior point polynomial time methods in convex programming (Lecture Notes). Available at http://www2.isye.gatech.edu/~nemirovs/Lect_IPM.pdf.

[25] NESTEROV, Y. (2006). Constructing self-concordant barriers for convex cones. CORE discussion paper No. 2006/30.

[26] NESTEROV, Y. and NEMIROVSKI, A. (2008). Primal central paths and Riemannian distances for convex sets. *Found. Comput. Math.* **8** 533–560. MR2443088

[27] NESTEROV, Y. and NEMIROVSKII, A. (1994). *Interior-Point Polynomial Algorithms in Convex Programming. SIAM Studies in Applied Mathematics* **13**. SIAM, Philadelphia, PA. MR1258086

[28] NESTEROV, YU. E. and TODD, M. J. (1997). Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.* **22** 1–42. MR1436572

[29] NESTEROV, YU. E. and TODD, M. J. (2002). On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.* **2** 333–361. MR1930942

[30] RAMANA, M. V. (1997). An exact duality theory for semidefinite programming and its complexity implications: Semidefinite programming. *Math. Program.* **77** 129–162. MR1461379

[31] RENEGAR, J. (1988). A polynomial-time algorithm, based on Newton's method, for linear programming. *Math. Program.* **40** 59–93. MR0923697

[32] VAIDYA, P. M. (1996). A new algorithm for minimizing convex functions over convex sets. *Math. Program.* **73** 291–341. MR1393123

[33] VEMPALA, S. (2005). Geometric random walks: A survey. In *Combinatorial and Computational Geometry. Math. Sci. Res. Inst. Publ.* **52** 577–616. Cambridge Univ. Press, Cambridge. MR2178341

DEPARTMENT OF STATISTICS
    AND DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE 98105
USA
E-MAIL: harin@uw.edu