Randomized interior point methods for sampling and optimization

Hariharan Narayanan
Laboratory for Information and Decision Systems
MIT
har@mit.edu

Abstract

We present a Markov chain (Dikin walk) for sampling from a convex body equipped with a self-concordant barrier, whose mixing time from a "central point" is strongly polynomial in the description of the convex set. The mixing time of this chain is invariant under affine transformations of the convex set, thus eliminating the need for first placing the body in an isotropic position. This strengthens previous results of [11] for polytopes and generalizes these results to arbitrary convex sets. In the case of a convex set K defined by a semidefinite constraint of rank at most α and at most m additional linear constraints, our results specialize to the following statement.

Let $s \geq \frac{|p|}{|q|}$ for any chord $\overline{pq}$ of K passing through a point $x \in K$. Then, after

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

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

On every convex set of dimension n, there exists a self-concordant barrier whose "complexity" is polynomially bounded. Consequently, a rapidly mixing Markov chain of the kind we describe can be defined 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. We show that this random walk reaches an approximately optimal point in polynomial time with high probability and that the corresponding objective values converge with probability 1 to the optimal objective value as the number of steps tends to infinity. One technical contribution is a family of lower bounds for the isoperimetric constants of (weighted) Riemannian manifolds on which, interior point methods perform a kind of steepest descent. Using results of Barthe [2] and Bobkov and Houdré [5], on the isoperimetry of products of (weighted) Riemannian manifolds, we obtain sharper upper bounds on the mixing time of Dikin walk on products of convex sets than the bounds obtained from a direct application of the Localization Lemma, on which, the analyses of all random walks on convex sets have relied since (Lovász and Simonovits, 1993).

1 Introduction

Sampling from a nearly uniform distribution on a high dimensional convex set is an important ingredient in several computational tasks, including computing its volume [6, 17] and sampling from lattice points in it [12]. The usual strategy for doing so is to design a rapidly mixing Markov chain whose stationary distribution is the uniform distribution, and then run it for sufficiently long and pick the final point as a sample. The mixing times of all known Markov chains for sampling generic convex sets depend on the aspect ratio of the set, a measure of which is the ratio between its diameter and width.

Generalizing prior work with Kannan on polytopes [11], we present a Markov chain for sampling from a convex body defined using a combination of linear, hyperbolic and self-concordant constraints (by which we mean constraints corresponding to which there is a logarithmic, hyperbolic or self-concordant barrier respectively). When restricted to the case of polytopes, the bounds we present imply (upto universal constants), the bounds in (Kannan and Narayanan, [11]). The mixing time of this chain is an affine invariant, thus eliminating the need for first placing the body in an isotropic position. A self-concordant barrier F 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, let $D^2F(x)$ be the Hessian matrix of F at x. We define the transition measure P_x corresponding to x to roughly be a Gaussian whose covariance matrix is a fixed multiple of $\left(D^2F(x)\right)^{-1}$. The properties of the barrier function cause the random walk to avoid the boundary, but at the same time take relatively large steps. For example, let K be the 2-dimensional Euclidean ball $\{x : \|x\| \le R\}$ 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 $\sqrt{R^2 - \|x\|^2}$. We see 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.

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 [11]. The Markov chain considered in [11] was ergodic, while the one we use here is not. The analysis of the non-ergodic Markov chain hinges upon the fact that it can be viewed as a limit of ergodic Markov chains.

1.1 Barrier oracle model for convex sets

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 [11]) 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 with parameter ν (see Section 2).

2. We are given a real number s such that for any chord $\overline{pq}$ of K through the origin, $\frac{|p|}{|q|} \le s$.

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

1.2 Implementing the barrier oracle in the linear and semidefinite cases

The most frequently encountered barrier functions encountered are the logarithmic barrier for polytopes and the log det barrier for convex sets defined by semidefinite constraints (See Section 2). 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 $Ax \le \mathbf{1}$. Then, $H(x) = A^T D(x)^2 A$ where $D(x)$ is the diagonal matrix whose i^{th} diagonal entry $d_{ii}(x) = \frac{1}{1 - a_i^T x}$.

By results of Baur and Strassen [3], the complexity of solving linear equations and of computing the determinant of an $n \times n$ matrix is $O(n^\gamma)$. The computation of $A^T D(x)^2 A$ can be achieved using

$mn^{\gamma-1}$ arithmetic operations, by partitioning a padded extension of A^TD into $\leq \frac{m+n-1}{n}$ square matrices. Thus, the complexity of the barrier oracle is $O(mn^{\gamma-1})$ arithmetic operations where $\gamma < 2.377$ is the exponent for matrix multiplication.

In the case of a semidefinite constraint of rank ν, 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, [25]. We have replaced an exponent 3 in [25] 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.

1.3 Presentation of the convex set K

For the definitions of logarithmic, hyperbolic and self-concordant barriers, we refer to Section 2. 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 $\bar{F}$, K_h is a convex set accompanied with a hyperbolic barrier $\breve{F}$ with parameter ν_h, and K_s is a convex set accompanied by a ν_s-self-concordant barrier $\tilde{F}$. Although their intersection is bounded, each of these convex sets may be unbounded. Define the self-concordant barrier function

$$F := \bar{F} + n\breve{F} + n^2\tilde{F},$$

and define

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

to be the *complexity parameter* of F (which is different from its self-concordance parameter; this being $m + \sqrt{n}\nu_h + n\nu_s$). 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

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

The random walk we use here is a variation of the Dikin walk defined in [11], in which instead of picking the next point from a Dikin ellipsoid, one picks it from a Gaussian having that covariance.

1.4 Related Work

Let $B(x, \tau)$ be defined to be the $n-$dimensional Euclidean ball of radius τ centered at x and suppose K is a n-dimensional convex set such that $B(0, r) \subseteq K \subseteq B(0, R)$. The Markov chain known as the "Ball walk" [17, 10] 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. The mixing time of this walk from a warm start (i. e. a density that is bounded above by $O(1)$ times the stationary density) in order to achieve a constant total variation distance to stationarity is $O^*\left(\frac{n^2R^2}{r^2}\right)$. However, for no single pre-specified point (such as the center of mass, as opposed to a random one) is it known to mix in polynomial time. More recently, a random walk known as Hit-and-Run, was analyzed in [16, 18]. 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 $\{x_i + \lambda v \big| \lambda \in \mathbb{R}\} \cap K$. 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 $O\left(\frac{n^2R^2}{r^2}\right)$, and its mixing time from a fixed point at a distance d from the boundary is $O\left(n^3(\frac{R}{r})^2 \ln \frac{R}{d}\right)$. The mixing time of the Dikin walk in two cases of interest are as follows (details are provided in Theorem 2). Let $x \in K$ and for all chords $\overline{pq}$ passing through x, $\frac{|p-x|}{|q-x|} \leq s$. We will call such a point x $s-$central. Suppose K is

(S) a slice of the semidefinite cone $S^{\nu\times\nu}$ of $\nu \times \nu$ matrices endowed with the hyperbolic barrier $F(x) = -\ln\det x$ or

(Q) the intersection of $m = \frac{\nu}{2}$ ellipsoids, $A_1B \cap A_2B \cap \cdots \cap A_mB$ where A_i are non-singular affine transformations and B is the Euclidean Ball. In this case, the hyperbolic barrier is $F(x) = -\sum \ln(1 - \|A_i^{-1}(x)\|^2)$.

Then the mixing time starting at x is $O(n^2\nu(n\ln(\nu s) + \ln\frac{1}{\epsilon}))$. Whether or not the bodies are in isotropic position, in the above cases (S) and (Q) corresponding to semidefinite and quadratic programs, when $\nu = O(n^{1-\epsilon})$, the mixing time bounds are an improvement over the existing bounds for Hit-and-Run [18].

If K is defined by semidefinite 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 semidefiniteness 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 semidefinite 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 [28]. 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 semidefinite 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.

Lovász [16] proved a lower bound of $\Omega(n^2p^2)$ 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.

More generally, we can compare the upper bounds on the number of (barrier) oracle calls needed to generate a point from an convex set K whose total variation distance from the uniform is ϵ, when the starting point is at a distance η from the boundary. We assume that the barrier F whose complexity parameter is ν. For Hit-and-Run, the number of oracle calls is

$$O\left(n^3\frac{R^2}{r^2}\ln\left(\frac{n}{\eta\epsilon}\right)\right).$$

For Dikin walk, this is

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

The ratio between the bounds for number of oracle calls for Dikin walk and the number of oracle calls for Hit-and-Run is $O^*\left(\frac{\nu r^2}{nR^2}\right)$.

In the specific case where the constraints are either semidefinite or linear, we can compare the upper bounds on the number of arithmetic operations needed in Hit-and-Run and Dikin walk. Suppose K as above that K is an convex set and the starting point is at a distance η from the boundary, and that it is defined by m linear constraints and additionally, semidefinite constraints of total rank α (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

$$O\left(mn^{\gamma-1} + n^2\alpha^2 + n\alpha^\gamma\right),$$

by the discussion in Subsection 1.2. For Hit-and-Run, the number of arithmetic steps needed to make one move in a naive implementation is $O^*\left(\log(R/r)(mn + n\alpha^2 + \alpha^\gamma)\right)$, (since the natural way of certifying positive semidefiniteness is to take a Cholesky factorization, which has a complexity $O(\alpha^\gamma)$, computing the new semidefinite matrix after one step has a complexity $n(\alpha^2)$ (Section 11.3, [25]) and testing containment in the region defined by linear constraints takes $O(nm)$ operations). We see that

1. If $m < n\alpha^2 + \alpha^\gamma$, 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 $m \geq n\alpha^2 + \alpha^\gamma$, 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 $O^*\left(\frac{(m+\alpha n)r^2}{R^2}\right)$ if $m < n\alpha^2 + \alpha^\gamma$ and $O^*\left(\frac{(m+\alpha n)r^2}{R^2 n^{0.62}}\right)$ if $m \geq n\alpha^2 + \alpha^\gamma$.

2 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 $int(K)$ denote its interior. For any function F on $int(K)$ having continuous derivatives of order k, for vectors $h_1, \dots, h_k \in \mathbb{R}^n$ and $x \in int(K)$, for $k \geq 1$, we recursively define

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

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

1. (Convex, Smooth) F is a convex thrice continuously differentiable function on $int(K)$.

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

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

   (a) $D^2F(x)[h,h]$ is 2-Lipschitz continuous with respect to the local norm, which is equivalent to
   $$D^3F(x)[h,h,h] \leq 2(D^2F(x)[h,h])^{\frac{3}{2}}.$$

4. $F(x)$ is ν-Lipschitz continuous with respect to the local norm defined by F,
   $$|D[F](x)[h]|^2 \leq \nu D^2[F](x)[h,h].$$

   We call the smallest positive integer ν for which this holds 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 non-singular 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 semidefinite cones. Polytopes and the intersections of ellipsoids can be expressed as sections of semidefinite cones.
2. Sections of ℓ_p balls.
3. Convex sets defined by the epigraphs of matrix norms (see page 199 of [23]).

For other examples and methods of constructing barriers for new convex sets by combining existing barriers, see Chapter 5 of [23].

2.1 Hyperbolic barriers

We refer the reader to [9] for the definition of a hyperbolic barrier. For the concrete applications in this paper, it suffices to note that on the semidefinite 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 non-singular affine transformations and B is the Euclidean Ball, $-\sum \ln(1 - \|A_i^{-1}(x)\|^2)$ is a hyperbolic barrier with parameter $2m$.

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

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

2.2 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 halfspaces defined by these constraints, and is given by

$$F(x) = -\sum_{i=1}^m \ln\left(1 - a_i^T x\right).$$

2.3 Dikin Ellipsoids

Around any point $x \in K$, the *Dikin ellipsoid* (of radius r) is defined to be

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

Fact 1. *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 non-singular 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 $G(y) := F(T^{-1}y)$.*

Fact 2. *For any y such that*

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

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

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

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 [23].

The following was proved in a more general context by Nesterov and Todd in Theorem 4.1, [26].

Theorem 1 (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 $p + \frac{|p-x|}{|p-y|}(z-p) \in D_x$.*

3 The Dikin walk

For $x \in int(K)$, let G_x denote the Gaussian density function given by

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

where

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

3.1 Algorithm 1

Let $0 = x_0 \in int(K)$. 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
   $$x_i := \begin{cases} z, & \text{with probability } \min\left(1, \frac{G_z(x_{i-1})}{G_{x_{i-1}}(z)}\right) \\ x_{i-1} & \text{otherwise.} \end{cases}$$
   (c) If $z \notin K$, let $x_i := x_{i-1}$.

Theorem 2 (Sampling)**.** *Let $K \ni 0$ be an $n-$dimensional convex set accompanied by a barrier F as in Subsection 1.3, with complexity parameter ν. Let $s \geq \frac{|p|}{|q|}$ for any chord $\overline{pq}$ of K containing the origin. Then, the number of steps $x_1, \dots, x_t$ that need to be taken before the total variation distance and the $\mathcal{L}_2$ distance of the density $\rho(x_t)$ of x_t to the uniform density is less than ϵ is $O\left(n\nu\left(n\ln(\nu s) + \ln\frac{1}{\epsilon}\right)\right)$. The number of steps needed from a warm start, i. e. when the $\mathcal{L}_\infty$ norm of $\rho(x_0)(\text{vol}(K))$ is $O(1)$, is $O\left(n\nu\ln\frac{1}{\epsilon}\right)$.*

In particular, if K is

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

(Q) the intersection of r ellipsoids, $A_1B \cap A_2B \cap \cdots \cap A_\tau B$ where A_i are non-singular affine transformations and B is the Euclidean Ball. In this case, $F(x) = -\sum \ln(1 - \|A_i^{-1}(x)\|^2)$,

the mixing time from a fixed "$s-$central" point or a warm start, respectively, are $O\left(n^2\tau\left(n\ln(n\tau s) + \ln\frac{1}{\epsilon}\right)\right)$ and $O\left(n^2\tau\ln\frac{1}{\epsilon}\right)$.

The mixing bounds in this paper are obtained by relating the Markov chain to the metric of Riemannian manifold studied in operations research [22, 27], rather than the Hilbert metric [11, 16]. This metric possesses several potentially useful characteristics. For example, when the convex set

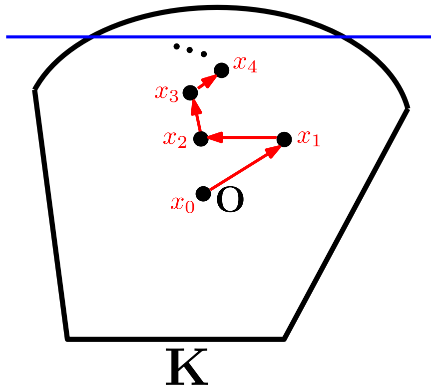

Figure 1: Trajectory of a Dikin walk for optimizing a linear function on a convex set

is a direct product of polytopes, 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 polytopes, 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 [17]). Even if K is a direct product of polytopes, the Dikin Markov chain itself does not factor into a product of Dikin Markov chains and Theorem 3 does not follow from a direct use of the Localization Lemma.

Theorem 3. *If an $n-$dimensional convex set $K := K_1 \times \cdots \times K_h$ is the direct product of polytopes K_i, each of which individually has a function F_i with a complexity parameter (defined in Equation 1) at most κ, then, the mixing time of Dikin walk from a warm start on K defined using the function $\sum_{i=1}^h F_i$ is $O(\kappa n)$.*

When there are $\Omega(n)$ factors, each of which is a polytope with κ faces, the total number of faces of K is $\Omega(n\kappa)$. In this case, the results of [11] using the logarithmic barrier, give a bound of $O(\kappa n^2)$ while Theorem 3 gives a bound of $O(\kappa n)$.

4 Convex programming

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.

Unlike other random walk based algorithms ([4]) this algorithm does not proceed in phases, but consists of a single random walk $x_0, x_1, \dots$ on K. The algorithm here is a Las Vegas algorithm rather than a Monte Carlo algorithm as was the case in [11]. It is also different 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

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

is bounded, for any chord $\overline{pq}$ of Q passing through the origin, $\frac{|p|}{|q|} \leq s$ and $\epsilon, > 0$ (if $B(0, r) \subseteq K \subseteq B(0, R)$, then $s \leq \frac{R}{r}$). Then, the algorithm is required to do the following.

- If $\exists x \in K$ such that $c^T x \geq 1$,
- Output $x' \in K$ such that $c^T x' \geq 1 - \epsilon$.

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

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

and let $\hat{F}$ be a barrier for $\hat{K} := T(Q)$. Such a barrier can be easily constructed from F; details follow Theorem 4.

Our algorithm for convex programming consists simply of doing a modified Dikin walk on $\hat{K}$ for a sufficient number of steps that depends on the desired accuracy ϵ and confidence $1 - \delta$. We define $\hat{G}_t$ using $\hat{F}$ in the same way that G_t was defined using F; details appear below.

4.1 Algorithm 2

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

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

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

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

Theorem 4 (Las Vegas algorithm for optimization). *Let K, F, s and r be as in Theorem 2. In the cases where F is a ν-barrier or a hyperbolic barrier with parameter ν, let $\tau(\epsilon, \delta)$ be set to $O\left(n\nu\left(\ln\frac{1}{\delta} + \left(n \ln\frac{s\nu}{\epsilon}\right)\right)\right)$. If $\{c^T x \geq 1\} \cap K$ is nonempty and $x_0, x_1, \ldots$ is the modified Dikin walk in Algorithm 2, then*

$$\mathbb{P}\left[\forall i > \tau(\epsilon, \delta),\, c^T x_i \geq 1 - \epsilon\right] \geq 1 - \delta. \tag{3}$$

Corollary 1. *For any $\kappa > 0$, with probability 1,*

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

4.2 Constructing barriers

The construction in [21], provides us with a barrier $\hat{F}_s$ on $\hat{K}$, given by

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

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

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

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

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

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

where

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

For $x \notin int(\hat{K})$, 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.

5 Metric defined by a barrier

For any smooth strictly convex function G, the Hessian D^2G is positive definite. Given the barrier F, for every $x \in supp(F)$ and $u, v \in \mathbb{R}^n$,

$$< u, v >_x := D^2F(x)[u, v]$$

is bilinear, and $\|u\|_x = \sqrt{< u, u >_x}$ is a norm. In addition, we define

1. $\breve{<} u, v >_x := D^2\breve{F}(x)[u, v]$ and $\breve{\|} u\|_x := \sqrt{\breve{<} u, u >_x}$.
2. $\tilde{<} u, v >_x := D^2\tilde{F}(x)[u, v]$ and $\tilde{\|} u\|_x := \sqrt{\tilde{<} u, u >_x}$ and
3. $\bar{<} u, v >_x := D^2\bar{F}(x)[u, v]$ and $\bar{\|} u\|_x := \sqrt{\bar{<} u, u >_x}$.

We define

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

where the infimum is taken over all rectifiable paths Γ from x to y. Let $\mathcal{M}$ be the metric space whose point set is K and metric is d. d_ℓ, d_h and d_s are defined analogously in terms of the respective norms $\bar{\|} \cdot \|$, $\breve{\|} \cdot \|$ and $\tilde{\|} \cdot \|$.

Lemma 2 (Nesterov-Todd (Lemma 3.1 [27])). *If* $\|x - y\|_x < 1$ *then,*

$$\|x - y\|_x - \|x - y\|_x^2 \leq d(x, y) \leq -\ln(1 - \|x - y\|_x).$$

While some of the presented bounds can be obtained from the isoperimetric bounds for the "Hilbert metric" (Theorem 6) proved by Lovász, we can prove stronger results for sampling certain classes of convex sets such as the direct product of a number of polytopes, by using results of Barthe [2] and Bobkov and Houdré [5] on the isoperimetry of product spaces. In particular, for a direct product of an arbitrary number of polytopes, each defined by $O(\kappa)$ constraints, this allows us to show a 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 [27] and by Nesterov and Nemirovski in [22], and Karmarkar studied the properties of a related metric [14] that underlay his celebrated algorithm [13]. For other work on sampling Riemannian manifolds motivated by statistical applications, see [15, 20], Chapter 8 [19].

5.1 Isoperimetry

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

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

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 $\beta_{fat}(\delta, \mathcal{M}, \mu)$ is defined as

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

Given interior points x, y in $int(K)$, suppose p, q are the ends of the chord in K containing x, y and p, x, y, q lie in that order. Denote by d_H the Hilbert (projective) metric defined by

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

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

Theorem 5. *If F is the self-concordant barrier of K with complexity parameter ν, presented in the format of Subsection 1.3,*

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

Proof.

Lemma 3. *1. $d_s(x, y) \leq 2(1 + 3\nu_s) n d_H(x, y)$.*

2. $d_h(x, y) \leq \sqrt{n\nu_h} d_H(x, y)$.

3. $d_\ell(x, y) \leq \sqrt{m} d_H(x, y)$

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

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

and a direct computation shows that

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

Lemma 3 follows from Theorems 7 and 8. □

Theorem 5 follows from Theorem 6 and Lemma 3. □

Theorem 6 (Lovász, [16]). *Let S_1 and S_2 be measurable subsets of K. Then,*

$$\mathrm{vol}(K \setminus S_1 \setminus S_2)\mathrm{vol}(K) \geq \left(e^{d_H(S_1,S_2)} - 1\right) \mathrm{vol}(S_1)\mathrm{vol}(S_2).$$

For $x \in K$ and a vector v, $|v|_x$ is defined to be $\sup_\alpha \{x \pm \alpha v \in K\}$.

Theorem 7 (Theorem 2.3.2 (iii), [23]). *Let F be a self-concordant barrier whose self-concordance parameter is ν_s as defined in Section 2. Then, for all $h \in \mathbb{R}^n$ and $x \in int(K)$*

$$|h|_x \leq \|h\|_x \leq 2(1 + 3\nu_s)|h|_x.$$

The following result is implicit in [9].

Theorem 8 (Güler, [9]). *Let $-\ln p(x)$ be a hyperbolic barrier for K, where p has degree ν_h. Then, for all $h \in \mathbb{R}^n$ and $x \in int(K)$,*

$$|h|_x \leq \sqrt{D^2 \breve{F}(x)[h,h]} \leq \sqrt{\nu_h}|h|_x.$$

6 Analysis of the mixing time

We denote the marginal 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. 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. This Lemma is used in the proof of Claim 3.

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,*

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

If F is a self-concordant barrier, and

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

when $k = 3$, this simplifies to

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

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

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

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

$$\sigma\left(p > M + \epsilon\right) \leq e^{-\frac{n\epsilon^2}{2}}. \tag{4}$$

When F is a self-concordant barrier, the second claim in the lemma follows because

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

□

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 $\alpha \geq 0$,*

$$\mathbb{P}\left[\min(\|p\|, \|q\|) \leq \sqrt{\frac{n}{2(\alpha + 2\ln n)}}\right] \leq 2e^{-\alpha}. \tag{5}$$

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

$$I \succeq \sum_i a_i a_i^T. \tag{6}$$

Examining the trace and the norm on both sides of (6), we obtain

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

and

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

We note that

$$\min(\|p\|, \|q\|) = \left(\min_i |a_i^T x|\right)^{-1}. \tag{7}$$

Thus, it is sufficient to show that

$$\mathbb{P}\left[\exists_i (a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right] \leq 2e^{-\alpha}, \tag{8}$$

which we proceed to do. Let S be the subset of $[m] := \{1, \dots, m\}$, 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 (6), $|S| \leq n^2$. Thus, by the union bound,

$$\mathbb{P}\left[\exists_i (a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right] \quad \leq \quad n^2 \sup_i \mathbb{P}\left[(a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right]. \tag{9}$$

We note that, by (4), for any vector w with norm less or equal to 1,

$$\mathbb{P}\left[a_i^T w \geq \sqrt{\frac{2(\alpha + 2\ln n)}{n}}\right] \leq \frac{e^{-\alpha}}{n^2}, \tag{10}$$

and so

$$\mathbb{P}\left[\exists_i (a_i^T v)^2 \geq \frac{2(\alpha + 2\ln n)}{n}\right] \leq 2e^{-\alpha}. \tag{11}$$

□

Proof of Theorem 2. In order to obtain mixing time bounds, we will first prove 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 marginals P_x and P_y is $< 1 - \Omega(1)$.

Without loss of generality, let x be the origin 0 (which is achievable by translation), and for any v, let $D^2F(0)[v, v] = \|v\|^2$ (which is achievable by an affine transformation of K).

For $x \neq y$,

$$1 - d_{TV}(P_x, P_y) = \mathbb{E}_z\left[\min\left(1, \frac{G_y(z)}{G_x(z)}, \frac{G_z(x)}{G_x(z)}, \frac{G_z(y)}{G_x(z)}\right)\right], \tag{12}$$

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, [25]) with D^kF in the place of M.

Fact 1. *Let $M[h_1, \dots, h_k]$ be a symmetric k-linear form on $\mathbb{R}^n$. Then,*

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

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

6.1 Relating the Markov Chain to the manifold

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

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

We will use the following fact repeatedly:

Fact 3. *Suppose $x \to z$ is a transition of the Dikin walk, then,*

$$\mathbb{P}\left[\|x - z\| \leq \frac{1}{2}\right] \geq 1 - 10^{-3}.$$

This can be ensured by setting the value of r to be a sufficiently small constant.

Finally, we will frequently make use of the facts from (Theorem 2.1.1, [23]) 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

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

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

$$(1 - r)^2 D^2F(x)[h, h] \leq D^2F(y)[h, h] \leq \frac{1}{(1 - r)^2} D^2F(x)[h, h]. \tag{13}$$

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

For two probability distributions P_x and P_y, let $d_{TV}(P_x, P_y)$ represent the total variation distance between them.

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

Proof. Without loss of generality, we may assume that $\breve{F}$, $\bar{F}$ and $\tilde{F}$ 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 $\tilde{F}$,) fourth order derivatives uniformly over K. Due to affine invariance, without loss of generality, let $< u, v >_x := < u, v >$, the usual dot product. As defined in Section 3, for any $z \in K$,

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

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 $d_{TV}(P_x, P_y) = 1 - \Omega(1)$. Without loss of generality, we assume x is the origin and we drop this subscript at times to simplify notation.

$$1 - d_{TV}(P_x, P_y) = \mathbb{E}_z\left[\min\left(1, \frac{G_y(z)}{G_x(z)}, \frac{G_z(x)}{G_x(z)}, \frac{G_z(y)}{G_x(z)}\right)\right], \tag{14}$$

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

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

We will prove the following lemmas.

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

Lemma 8.

$$\mathbb{P}\left[-V(z) < O(1)\right] \quad > \quad \frac{9}{10}. \tag{15}$$

Next, we will prove the following probabilistic upper bound, thereby completing the proof.

Proposition 1.

$$\mathbb{P}\left[\max\left(\|y-z\|_y^2, \|z\|_z^2, \|z-y\|_z^2\right) - \|z\|^2 < O(\frac{1}{n})\right] > \frac{199}{1000}. \tag{16}$$

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

$$\mathbb{P}\left[\max\left(\|z\|_y^2 - \|z\|^2, < y, z >_y, \|z\|_z^2 - \|z\|^2, < y, z >_z\right) < O(\frac{1}{n})\right] \quad > \quad \frac{2}{10}. \tag{17}$$

This fact follows from the following three lemmas and the union bound. The proof of Lemma 9 would go through if $\frac{4}{10}$ were replaced by $\frac{1}{2} - \Omega(1)$.

Lemma 9.

$$\mathbb{P}\left[\max\left(\bar{\|}z\|_y^2 - \bar{\|}z\|^2, \bar{<}y, z >_y, \bar{\|}z\|_z^2 - \bar{\|}z\|^2, \bar{<}y, z >_z\right) < O(\frac{1}{n})\right] \quad > \quad \frac{4}{10}. \tag{18}$$

Lemma 10.

$$\mathbb{P}\left[\max\left(\breve{\|}z\|_y^2 - \breve{\|}z\|^2, \breve{<}y, z >_y, \breve{\|}z\|_z^2 - \breve{\|}z\|^2, \breve{<}y, z >_z\right) < O(\frac{1}{n^2})\right] > \frac{9}{10}. \tag{19}$$

Lemma 11.

$$\mathbb{P}\left[\max\left(\tilde{\|}z\|_y^2 - \tilde{\|}z\|^2, \tilde{<}y, z >_y, \tilde{\|}z\|_z^2 - \tilde{\|}z\|^2, \tilde{<}y, z >_z\right) < O(\frac{1}{n^3})\right] \quad > \quad \frac{9}{10}. \tag{20}$$

□

□

Proof of Lemma 7. Let $\breve{\bar{F}} := F - \bar{F}$. For $w \in K$, let $X(w) := D^2\bar{F}(w) - D^2\bar{F}(0)$. By Lemma 12 in [11], for any point $w \in K$, the gradient of $\mathbf{Tr}X(w)$ measured using $\bar{\|} \cdot \|_w$ is $\leq 2\sqrt{n}$. Therefore, the gradient of $\mathbf{Tr}X(w)$ measured using $\bar{\|}\cdot\|$ is $\leq O(\sqrt{n})$. $X(0) = 0$, therefore, $|\mathbf{Tr}X(y)| \leq O(\|y\|\sqrt{n}) = O(1)$.

$\|y\| = O(1/\sqrt{n})$, therefore, $\|X\| = O(\|y\| \sup_{\|w\| \leq \|y\|} \|D^3\bar{F}(w)\|) = O(1/\sqrt{n})$.

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

Then

$$\|Y(y)\| = D^2\breve{\bar{F}}(y) - D^2\breve{\bar{F}}(0) = O(\|y\| \sup_{\|w\| \leq \|y\|} \|D^3\breve{\bar{F}}(w)\|) = O(1/n).$$

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

$$-V(y) \quad = \quad -\ln\det(I + X + Y) \tag{21}$$

$$= \quad -\mathbf{Tr}(X + Y + R), \tag{22}$$

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. □

Proof of Lemma 8. Let $\breve{\breve{F}} := F - \bar{F}$. For $w \in K$, let $W(w) := D^2\bar{F}(w)$, and let $Z(w) := D^2\breve{\breve{F}}(w)$.

$$\begin{aligned} V(z) &= \frac{1}{2}\ln\det(W(z)+Z(z)) \\ &= \frac{1}{2}\left(\mathbf{Tr}\ln(W(z)+Z(z))\right) \text{ with probability} \geq 1-10^{-3} \text{ (i.e. if } \|W(z)+Z(z)-I\|<1) \\ &= -\frac{1}{2}\left(\mathbf{Tr}(W(z)-W(0))+\mathbf{Tr}(Z(z)-Z(0))\right) \quad (23) \\ &+ \frac{1}{4}\left(\mathbf{Tr}\,(W(z)+Z(z)-I)^2\right) \quad (24) \\ &- O\left(\left|\mathbf{Tr}\,(I-(W(z)+Z(z)))^3\right|\right) \quad \text{with probability} \geq 1-10^{-3}. \quad (25) \end{aligned}$$

The lemma will follow from the following claims:

Claim 1.

$$\mathbb{P}\left[\mathbf{Tr}(Z(z)-Z(0)) \leq O(1)\right] \geq \frac{99}{100}.$$

Claim 2.

$$\mathbb{P}\left[\mathbf{Tr}(W(z)-W(0)) \leq O(1)\right] \geq \frac{99}{100}.$$

Claim 3.

$$\mathbb{P}\left[\left|\mathbf{Tr}\,(I-(W(z)+Z(z)))^3\right| \leq O(1)\right] \geq \frac{99}{100}. \quad (26)$$

□

Proof of Claim 1. Let $Z_h(w) := D^2\breve{F}(w)$ and $Z_s = D^2\tilde{F}(w)$. Then, $Z(w) = nZ_h + n^2 Z_s$. Next,

$$\mathbf{Tr}Z_h(z) - \mathbf{Tr}Z_h(0) = D\mathbf{Tr}Z_h(0)[z] + \frac{D^2\mathbf{Tr}Z_h(z')[z,z]}{2}, \quad (27)$$

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

$$\begin{aligned} |D\mathbf{Tr}Z_h(0)[z]| &\leq O\left(n\left(\sup_{\|v\|=1}|D^3\breve{F}[v,v,v]|\right)\right) \text{ with probability} \geq 1-10^{-3} \quad (28) \\ &\leq O\left(n\left(\sup_{\breve{\|}v\|=1/\sqrt{n}}|D^3\breve{F}[v,v,v]|\right)\right) \text{ with probability} \geq 1-10^{-3} \quad (29) \\ &\leq O(1/\sqrt{n}) \text{ with probability} \geq 1-10^{-3}, \quad (30) \end{aligned}$$

which is at most $O(\sqrt{n})$. Applying Lemma 4, we see that

$$\mathbb{P}[-D\mathbf{Tr}Z_h(0)[z] < O(1/n)] > \frac{998}{1000}. \quad (31)$$

$$D^2\mathbf{Tr}Z(z')[z,z] \;=\; D^2\mathbf{Tr}D^2\breve{F}(z')[z,z] \quad (32)$$

In order to bound the above quantity, 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 $\breve{\|}z'\| = O(1/\sqrt{n})$ with probability $\geq 1 - 10^{-3}$. Let D_A be the differential operator whose action on a function G is determined by the relation

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

Thus $D_A^2 \breve{F}(z') = Z_h(0)$. Now,

$$\breve{\|}u\| = O(1) \Rightarrow \forall \breve{\|}v\| = 1, D^4 \breve{F}(z')[u, u, v, v] \leq O(1). \tag{33}$$

$$\begin{aligned} D^2 \left(\mathbf{Tr} Z_h(z') \right) [z, z] &= D^2 \left(\mathbf{Tr} D_A^2 \breve{F}(0) \right) [z, z] \\ &\leq n \sup_{\|Av\|=1} O(D^4 F(z')[v, v, z, z]) \text{ with probability } \geq 1 - 10^{-3} \\ &\leq n^2 \sup_{\breve{\|}v\|=1/\sqrt{n}} D^4[v, v, v, v] \breve{\|}z\|^2 \quad \text{(by Fact 1)} \\ &= O(1/n) \quad \text{with probability } \geq 1 - 10^{-3}. \end{aligned} \tag{34}$$

Therefore, by Equations 27, 31 and 34, we have

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

with probability $\geq 1 - 10^{-3}$, $\tilde{\|}z\| = O(1/n)$, therefore with probability $\geq 1 - 10^{-3}$,

$$\begin{aligned} \mathbf{Tr} Z_s(z) - \mathbf{Tr} Z_s(0) &= O(\mathbf{Tr} Z_s(0) \tilde{\|}z\|) &(35) \\ &= O(1/n^2). &(36) \end{aligned}$$

The claim follows from the last two sentences, since $Z = nZ_h + n^2 Z_s$.

□

Proof of Claim 2.

$$\mathbf{Tr} W(z) - \mathbf{Tr} W(0) = D \mathbf{Tr} W(0)[z] + \frac{D^2 \mathbf{Tr} W(z')}{2}, \tag{37}$$

for some $z' \in [0, z]$. Lemma 12 in (Kannan and Narayanan, [11]) shows that $\bar{\|}\nabla \mathbf{Tr} W\| \leq 2\sqrt{n}$. Since for all vectors v, $\|v\| \geq \bar{\|}v\|$, this implies that $\|\nabla \mathbf{Tr} W\| \leq 2\sqrt{n}$. By Lemma 4, this implies that

$$\mathbb{P}\left[D \mathbf{Tr} W(0)[z] < O(1) \right] \geq 1 - 10^{-3}. \tag{38}$$

By Lemma 13 in (Kannan and Narayanan, [11]), $\frac{D^2 \mathbf{Tr} W(z')}{2} \leq 0$, thereby completing the proof. □

Proof of Claim 3. In order to prove that

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

it suffices to show that

$$\mathbb{P}\left[\|(W(z) - W(0))\| \leq O(n^{-1/3}) \right] \geq 1 - 10^{-3} \tag{39}$$

and that

$$\mathbb{P}\left[\|(Z(z) - Z(0))\| \leq O(n^{-1/3}) \right] \geq 1 - 10^{-3}. \tag{40}$$

From Lemma 5 and Theorem 1, we obtain (39). We obtain (40) from (13). □

6.2 Regularity of the metric defined by the Logarithmic barrier

Proof of Lemma 9. Let $\bar{F}(w) := -\sum_{i=1}^m \ln(1- < a_i, v >)$, for any $v \in K$. Thus

$$\left(\sum_{i=1}^m < v, a_i >^2\right) = \bar{\|}v\|^2 \leq \|v\|^2.$$

Fixing an orthonormal basis with respect to $< \cdot, \cdot >$,

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

where $X \preceq Y$ signifies that Y dominates X in the semidefinite cone.

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

Lemma 12.

$$\mathbb{P}\left[\max\left(\bar{\|}z\|_y^2 - \bar{\|}z\|^2, \bar{<}y, z >_y, \bar{<}y, z >_z\right) < O(\frac{1}{n})\right] \quad > \quad \frac{19}{20}. \tag{41}$$

Lemma 13.

$$\mathbb{P}\left[\bar{\|}z\|_z^2 - \bar{\|}z\|^2 < O(\frac{1}{n})\right] \quad > \quad \frac{9}{20}. \tag{42}$$

□

Proof of Lemma 12.

$$\begin{aligned} \bar{\|}z\|_y^2 - \bar{\|}z\|^2 &= D^2\bar{F}(y)[z, z] - D^2\bar{F}(0)[z, z] \\ &= D^3\bar{F}(w)[y, z, z], \end{aligned}$$

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

$$D^3\bar{F}(w)[y, z, z] = (1 + o(1))D^3\bar{F}(0)[y, z, z]. \tag{43}$$

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{i=1}^m (y^T a_i a_i^T z)(a_i^T z)\right)^2\right] &\leq \mathbb{E}\left[\left(\sum_{i=1}^m (y^T a_i a_i^T z)^2\right)\left(\sum_i (a_i^T z)^2\right)\right] \\ &= \mathbb{E}\left[\|y(\sum_i a_i a_i^T)\|^2 \|z\|^4/n\right] \\ &= O(1/n^2). \end{aligned} \tag{44}$$

Therefore,

$$\begin{aligned} \mathbb{P}\left[D^3\bar{F}(0)[y, z, z] < O(1/n)\right] &= \mathbb{P}\left[-2\sum_{i=1}^m (a_i^T y)(a_i^T z)^2 < 1/n)\right] \\ &\geq 1 - O\left(n^2\mathbb{E}\left[\left(\sum_{i=1}^m (y^T a_i a_i^T z)(a_i^T z)\right)^2\right]\right) \\ &\geq 1 - 10^{-3} \quad \text{By (44)}. \end{aligned}$$

Proceeding to the next term,

$$\begin{aligned}\mathbb{P}\left[\bar{<}y,z>_y = O(1/n)\right] &= \mathbb{P}\left[\bar{<}y,z> + (\bar{<}y,z>_y - \bar{<}y,z>) = O(1/n)\right] \\ &\bar{<}y,z>_y - \bar{<}y,z> = O(\|z\|/n)\end{aligned}$$

by (13), and

$$\mathbb{E}[\bar{<}y,z>^2] \le O(1/n^2), \tag{45}$$

so we obtain

$$\mathbb{P}\left[\bar{<}y,z>_y = O(1/n)\right] > 1 - 10^{-3}. \tag{46}$$

Finally,

$$\bar{<}y,z>_z - \bar{<}y,z> = D^3\bar{F}(w)[y,z,z] \tag{47}$$

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

From Equations (47), (45) and (45), it follows that

$$\mathbb{P}\left[\bar{<}y,z>_z = O(1/n)\right] > 1 - 2(10^{-3}). \tag{48}$$

Lemma 12 follows from Equations (45), (46) and (48). □

Proof of Lemma 13. In order to prove that

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

it suffices to show that

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

because the distribution of z is symmetric about the origin.

$$\begin{aligned}\sum_i \left(\frac{(a_i^Tz)^2}{2(1-a_i^Tz)^2} + \frac{(a_i^Tz)^2}{2(1+a_i^Tz)^2}\right) &= \sum_i (a_i^Tz)^2\left(\frac{1+(a_i^Tz)^2}{(1-(a_i^Tz)^2)^2}\right) &(49)\\ &= \sum_i \left((a_i^Tz)^2 + \frac{3(a_i^Tz)^4-(a_i^Tz)^6}{(1-(a_i^Tz)^2)^2}\right) \\ &= \bar{\|}z\|^2 + \sum_i \frac{3(a_i^Tz)^4-(a_i^Tz)^6}{(1-(a_i^Tz)^2)^2}. &(50)\end{aligned}$$

The probability that $|a_i^Tz| \ge n^{-\frac{1}{4}}$ is $O(e^{-\sqrt{n}/2})$. $|a_i^Tz|$ is $\le \|a_i^T\|r$, which is less than $\frac{1}{2}$. This allows us to write with probability $\ge 1 - 10^{-3}$

$$\mathbb{E}\left[\frac{3(a_i^Tz)^4-(a_i^Tz)^6}{(1-(a_i^Tz)^2)^2}\right] \quad = \quad 3\mathbb{E}[(a_i^Tz)^4](1+o(1)), \tag{51}$$

which is $O(\bar{\|}a_i\|^4/n^2)$. Since $\sum_i a_ia_i^T \preceq I$, $\forall i$, $\bar{\|}a_i\| \le 1$ and $\sum_i \bar{\|}a_i\|^2 \le n$.

Therefore,

$$\begin{aligned}\mathbb{P}\left[\left(\bar{\|}z\|_z^2 + \bar{\|}z\|_{-z}^2\right)/2 < \bar{\|}z\|^2 + \frac{100}{n})\right] &= \mathbb{P}\left[\sum_i \frac{3(a_i^Tz)^4-(a_i^Tz)^6}{(1-(a_i^Tz)^2)^2} < 10^2/n\right] &(52)\\ &\le n\sum_i \bar{\|}a_i\|^4/(100n^2) &(53)\\ &\le \bar{\|}a_i\|^2/(100n) &(54)\\ &\le 1/100. &(55)\end{aligned}$$

□

6.3 Regularity of the metric defined by the Hyperbolic Barrier

Proof of Lemma 10. We will prove upper bounds on each of (a) $\breve{\|}z\|_y^2 - \breve{\|}z\|^2$, (b) $< y, z >_y$, (c) $\breve{\|}z\|_z^2 - \breve{\|}z\|^2$ and (d) $\breve{<}y, z >_z$ that hold with constant probability, and then use the union bound. We will repeatedly use the observation (that holds from Fact 2) that for any point w such that $\breve{\|}w\| = o(1)$,

$$\breve{\|}y\|_w \leq O(n^{-1/2}\|y\|) \leq O\left(\frac{1}{n}\right) \tag{56}$$

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

$$\breve{\|}z\|_w \leq O(n^{-1/2}\|z\|) \leq O(1/\sqrt{n}). \tag{57}$$

(a)

$$\begin{aligned} \breve{\|}z\|_y^2 - \breve{\|}z\|^2 &= D^2\breve{F}(y)[z,z] - D^2\breve{F}(0)[z,z] \\ &= D^3\breve{F}[w][y,z,z], \end{aligned}$$

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

By Fact 1,

$$\mathbb{P}\left[D^3\breve{F}(w)[y,z,z] < O(n^{-2})\right] \quad \geq \quad \mathbb{P}\left[\breve{\|}y\|_w(\breve{\|}z\|_w)^2 \leq O(\frac{1}{n^2})\right]. \tag{58}$$

Since $\|z\|^2 = \bar{\|}z\|^2 + n\breve{\|}z\|^2 + n^2\tilde{\|}z\|^2$, we have

$$\breve{\|}z\|_w = O(\|z\|_w/\sqrt{n}).$$

Also, $\breve{\|}w\| = O(\|y\|/\sqrt{n}) = O(1/n)$, and so with probability $\geq 1 - 10^{-3}$,

$$O(\breve{\|}z\|_w) = O(\|z\|/\sqrt{n}) = O(1/\sqrt{n}).$$

Thus,

$$\mathbb{P}\left[\breve{\|}y\|_w\breve{\|}z\|_w \leq O(\frac{1}{n\sqrt{n}})\right] \quad \geq \quad 1 - 10^{-3}. \tag{59}$$

and

$$\mathbb{P}\left[\breve{\|}y\|_w(\breve{\|}z\|_w)^2 \leq O(\frac{1}{n^2})\right] \quad \geq \quad 1 - 10^{-3}. \tag{60}$$

(b)

$$\begin{aligned} \breve{<}y, z >_y &= \breve{<}y, z > + (\breve{<}y, z >_y - \breve{<}y, z >) \\ &= \breve{<}y, z > + D^3\breve{F}(w)[y,y,z] \quad \text{(for } w \in [0,y]) \qquad (61) \\ &= O(\frac{\breve{\|}y\|}{n}) + O(\breve{\|}y\|_w^2\breve{\|}z\|_w) \ \text{ with probability} > 99/100. \qquad (62) \end{aligned}$$

In going from (61) to (62), we used Fact 1 and Fact 2. 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 (59), (60) and (62),

$$\begin{aligned} \mathbb{P}\left[(\breve{<}y, z >_y) < O(\frac{1}{n})\right] \quad &\geq \quad \mathbb{P}\left[\frac{\breve{\|}y\|\breve{\|}z\|}{\sqrt{n}}) + \breve{\|}y\|_w^2\breve{\|}z\|_w = O\left(\frac{1}{n}\right)\right] \qquad (63) \\ &> \quad 98/100. \end{aligned}$$

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

$$|\breve{\|}z\|_z^2 - \breve{\|}z\|^2 - D^3\breve{F}(0)[z,z,z]| \leq \sup_{w\in[0,y]} |\frac{D^4\breve{F}(w)[z,z,z,z]}{2}| \quad (64)$$

By Lemma 4,

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

$\mathbb{P}[\breve{\|}z\| = O(1/\sqrt{n})] > 99/100$, therefore, each term in (64) is $O(1/n^2)$ with probability $\frac{99}{100}$, and so

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

(d)

$$\begin{aligned} \breve{<}y,z>_z &= \breve{<}y,z> + (\breve{<}y,z>_z - \breve{<}y,z>) \\ &\leq \breve{<}y,z> + \sup_{w\in[0,z]} \left|D^3\breve{F}(w)[y,z,z]\right| \\ &= O(\frac{\breve{\|}y\|}{n}) + \breve{\|}y\|_w\breve{\|}z\|_w^2 \quad \text{with probability} > 99/100. \end{aligned}$$

By Equations (65) and (66),

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

Therefore,

$$\mathbb{P}\left[(\breve{<}y,z>_z) < O(\frac{1}{n^2})\right] \quad > \quad \frac{98}{100}.$$

□

6.4 Regularity of the metric defined by the self-concordant barrier

Proof of Lemma 11. 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) $\tilde{\|}z\|_y^2 - \tilde{\|}z\|^2$ (b) $\tilde{<}y,z>_y$, (c) $\tilde{\|}z\|_z^2 - \|z\|^2$ and (d) $\tilde{<}y,z>_z$ that hold with constant probability separately, and then use the union bound. We will repeatedly use the observation (that holds from Fact 2) that for any point w such that $\tilde{\|}w\| = o(1)$,

$$\tilde{\|}y\|_w \leq O(n^{-1}\|y\|) \leq O\left(\frac{1}{n\sqrt{n}}\right) \quad (65)$$

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

$$\tilde{\|}z\|_w \leq O(n^{-1}\|z\|) \leq O(1/n). \quad (66)$$

(a)

$$\begin{aligned} \tilde{\|} z\|_y^2 - \tilde{\|} z\|^2 &= D^2\tilde{F}(y)[z,z] - D^2\tilde{F}(0)[z,z] \\ &= D^3\tilde{F}(w)[y,z,z], \end{aligned}$$

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

$$\begin{aligned} D^3\tilde{F}(w)[y,z,z] &\leq \tilde{\|} y\|_w \tilde{\|} z\|_w^2 \quad \text{by Fact 1} && (67) \\ &\leq (n^{-3/2})(n^{-1})^2 \quad \text{with probability} > 1 - 10^{-3} && (68) \end{aligned}$$

(b)

$$\begin{aligned} \tilde{<} y,z>_y &= \tilde{<} y,z> + (\tilde{<} y,z>_y - \tilde{<} y,z>) && (69) \\ &= \tilde{<} y,z> + D^3\tilde{F}(w)[y,y,z] && (70) \\ &= O\left(\frac{\tilde{\|} y\|}{n\sqrt{n}}\right) + O\left(\|y\|_w^2\|z\|_w\right) \quad \text{with probability} > 99/100. && (71) \end{aligned}$$

In going from (61) to (62), we used Fact 1 and Fact 2. We see that

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

Therefore,

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

(c)

$$\begin{aligned} \tilde{\|} z\|_z^2 - \tilde{\|} z\|^2 &= D^2\tilde{F}(z)[z,z] - D^2\tilde{F}(0)[z,z] \\ &= D^3\tilde{F}(w)[z,z,z] \end{aligned}$$

for some w on the line segment $[0, z]$. By Fact 1,

$$\begin{aligned} D^3\tilde{F}(w)[z,z,z] &\leq \tilde{\|} z\|_w \tilde{\|} z\|_w^2 && (72) \\ &\leq O\left(\frac{1}{n^3}\right) \quad \text{with probability} \geq 1 - 10^{-3}. && (73) \end{aligned}$$

(d)

$$\begin{aligned} \tilde{<} y,z>_z &= \tilde{<} y,z> + (\tilde{<} y,z>_z - \tilde{<} y,z>) \\ &\leq \tilde{<} y,z> + \sup_{w\in[0,z]} D^3\tilde{F}(w)[y,z,z] \\ &= O(\frac{\tilde{\|} y\|}{n\sqrt{n}}) + \tilde{\|} y\|_w \tilde{\|} z\|_w^2 \quad \text{with probability} > 99/100. \end{aligned}$$

By Equations (65) and (66),

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

Therefore,

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

□

6.5 Bound on Conductance

Lemma 14 (Bound on Conductance)**.** *Let μ be the uniform distribution on K. The conductance*

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

of the Markov Chain in Algorithm 1 is $\Omega(\beta_{fat})$.

Proof. Let S_1 be a measurable subset of K such that $\mu(S_1) \leq \frac{1}{2}$ and $S_2 := K\setminus S_1$ be its complement. For any $x \neq y \in K$,

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

Let $S_1' = S_1 \cap \{x \big| P_x(S_2) = o(1)\}$ and $S_2' = S_2 \cap \{y \big| \rho(y)P_y(S_1) = o(1)\}$. 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

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

Lemma 6 implies that for an absolute constant C, if $d(x, y) \leq \frac{r}{C\sqrt{n}}$, then $d_{TV}(P_x, P_y) = 1 - \Omega(1)$. Therefore Theorem 5 implies that

$$\mu((K \setminus S_1') \setminus S_2') \geq \Omega(\beta_{fat}) \min(\mu(S_1'), \mu(S_2')).$$

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

$$\begin{aligned}\int_{S_1} P_x(S_2)d\mu(x) \quad &\geq \quad \mu(K \setminus S_1' \setminus S_2') \\ &\geq \quad \Omega(\beta_{fat})\mu(S_1') \\ &\geq \quad \Omega(\beta_{fat}) \min(\mu(S_1'), \mu(S_2'))\end{aligned}$$

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

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

and we are done. □

6.6 Mixing Bounds

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. $\forall t$, we know that $\mathbb{P}[e \geq t + 1 | e \geq t] \leq 1 - \Omega(1)$, therefore

$$\sup_x \frac{\rho_e(x)}{G_0(x)} \quad = \quad O(1). \tag{74}$$

Therefore

$$\int_K \rho_e(x)^2 dx = O(\int_{\mathbb{R}^n} G_e(x)^2 dx) \tag{75}$$

$$= \exp\left(O(n\ln(n\nu))\right). \tag{76}$$

Together with the following theorem, setting $f(x) := \rho(x_e) - (\mathrm{vol}(K))^{-1}$ this completes the proof of mixing from a fixed point. The proof of mixing from a warm start follows similarly.

Theorem 9 (Lovász-Simonovits [17]). *Let μ_0 be the initial distribution for a lazy reversible ergodic Markov chain whose conductance is Φ and stationary measure is μ, and μ_k be the distribution of the k^{th} step. Let $M := \sup_S \frac{\mu_0(S)}{\mu(S)}$ where the supremum is over all measurable subsets S of K. 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,*

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

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

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

□

7 Mixing in a direct product of polytopes

Our analysis 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 $F_\square(x_1, \dots, x_h) := -\sum_i \ln\cos(x_i)$ (see Section 6.2, [27]).

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

$$\mu^+(A) := \liminf_{\epsilon \to 0^+} \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon}, \tag{77}$$

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

Definition 1. *The (infinitesimal) Cheeger constant of the weighted manifold $(\mathcal{N}, \mu)$ is*

$$\beta_{\mathcal{N}} = \inf_{A \subset \mathcal{M}, \mu(A) \leq \frac{1}{2}} \frac{\mu^+(A)}{\mu(A)}, \tag{78}$$

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 $(x_1, \dots, x_h)$ and $(y_1, \dots, y_h)$ 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 1. *For any triple $(\mathcal{N}, d, \mu)$ as above,*

$$I_{\mu^{\otimes h}} \geq \frac{1}{2\sqrt{6}} I_\mu. \tag{79}$$

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 2. *Let $k \geq 2$ be an integer. For $i = 1, \dots, k$, 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 $i = 1, \dots, k$, 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 15. *Let $A, B \subseteq \mathcal{M}$ and $d_{\mathcal{M}}(A, B) \geq \delta_{\mathcal{M}}$. Then,*

$$\mu(\mathcal{M} \setminus \{A \cup B\}) \geq 2 \min(\mu(A), \mu(B))(e^{\frac{\beta_{\mathcal{M}} \delta_{\mathcal{M}}}{2}} - 1). \tag{80}$$

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

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

We proceed by contradiction. Suppose for some $\beta < \beta_{\mathcal{M}}$,

$$\exists \delta \in [0, \delta_1), \mu(A_\delta) < e^{\beta\delta} \mu(A). \tag{81}$$

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

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

However, we know that

$$\mu^+(\partial A_{\delta'}) := \liminf_{\epsilon \to 0^+} \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon} \geq \beta_{\mathcal{M}}, \tag{82}$$

which contradicts the fact that in any right neighborhood of δ', there is a δ for which (81) 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 $A_{\delta_1} \cap B_{\delta_{\mathcal{M}} - \delta_1} = \emptyset$, therefore $\mu(\delta_{\mathcal{M}} - \delta_1) \leq \frac{1}{2}$. So the same argument tells us that

$$\mu(B_{\delta_{\mathcal{M}} - \delta_1}) \geq e^{\beta_{\mathcal{M}}(\delta_{\mathcal{M}} - \delta_1)} \mu(B). \tag{83}$$

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

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

Next, suppose $\mu(B) \geq \frac{1}{2}$. We then set $\delta_1 := \delta_{\mathcal{M}}$, and see that the arguments from (81) to (83) carry through verbatim. Thus, in this case, $\mu(\mathcal{M} \setminus \{A \cup B\}) \geq \min(\mu(A), \mu(B))(e^{\beta_{\mathcal{M}} \delta_{\mathcal{M}}} - 1)$.

□

This immediately leads to the following corollary:

Corollary 1.

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

Nesterov and Todd show in (Lemma 4.1, [27]) that the Riemannian metric $\mathcal{M}$ on the direct product of polytopes 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.

Proof of Theorem 3. By Corollary 1, it suffices to show that $\beta_{\mathcal{M}} = \Omega(\frac{1}{\sqrt{\nu}})$. We will show this using Theorem 2 and Theorem 1. Consider the $h-$dimensional cube $[-\frac{\pi}{2}, \frac{\pi}{2}]^h$, and the metric from the Hessian of the barrier $F_\Box(x_1, \ldots, x_h) := -\sum_i \ln\cos(x_i)$ (see Section 6.2, [27]). The map

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

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

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

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 2). In the $1-$dimensional case, it is easy to check that the barrier is $1-$self-concordant (Section 6.2 [27]). Therefore, in the $1-$dimensional case, I_ϕ is bounded above and below by fixed constants. This, together with Theorem 5 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 Theorem 2 and Theorem 1 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,

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

□

8 Analysis of Las Vegas Algorithm

Proof of Theorem 4. In contrast with the Markov chain used for linear programming in [11], 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 2 apply since $\hat{K}^j$ is bounded. Let $\hat{F}_j = \hat{F} + \ln(j - c^T x)$. By known properties of barriers ([23]), 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 $2 \to 2$ operator norm 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 $\mathbb{P}[c^T(x_i) \le 1-\epsilon]$. Let $y_i := T(x_i)$. Let e be the first time the Markov chain $y_0, y_1, \dots$ escapes y_0. Thus e is an integer valued random variable defined by the event that y_e is the first point in $y_1, \dots$ that is not equal to y_0. Let the density of y_e be ρ_e. $\forall t$, we know from Lemma 6 that $\mathbb{P}[e \ge t+1 | e \ge t] \le 1 - \Omega(1)$, therefore

$$\sup_x \frac{\rho_e(x)}{\hat{G}_0(x)} \quad = \quad O(1). \tag{84}$$

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

$$\int_{\hat{K}} \rho_e(x)^2 dx \quad = \quad O(\int_{\mathbb{R}^n} \hat{G}_0(x)^2 dx) \tag{85}$$

$$\quad = \quad \exp\left(O(n \ln(n))\right). \tag{86}$$

Let $\rho_t(x)$ be the density of x_t. Then, by Theorem 2 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 $j \to \infty$,

$$\frac{\int_{\hat{K}} \rho_t(x)^2 dx}{\int_{\hat{K}} \rho_e(x)^2 dx} = O((1 - \frac{\Phi^2}{2})^t). \tag{87}$$

By Lemma 14, this is $O\left(\exp\left(-\frac{t(\beta_{fat})^2}{2}\right)\right)$.

Lemma 16. *Let ρ be a density supported on $\hat{K}$ such that*

$$\int_{\hat{K}_\epsilon} \rho(x) dx \ge \delta.$$

Then,

$$\int_{\hat{K}} \rho(x)^2 dx \quad \ge \quad \delta^2 \exp\left(-O\left(n \ln\left(\frac{s\nu}{\epsilon}\right)\right)\right). \tag{88}$$

Proof. Let K_ϵ be $K \cap \{c^T x \le 1\}$ and $\hat{K}_\epsilon := TK_\epsilon$. 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 $\hat{K}_\epsilon$, and $p = T^{-1}(p')$ and $q = T^{-1}(q')$. If $c^T p' \le c^T q'$, then $\frac{|p'|}{|q'|} \le \frac{|p|}{|q|} \le s$. On the other hand, if $c^T p' \ge c^T q'$, let r be the intersection of $\overline{pq}$ with $\{c^T x \le 1\}$. By the projective invariance of the cross ratio (see for example, Lemma 14 in [11])

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

Therefore

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

$$\quad \le \quad (s)\left(\frac{s+1}{\epsilon}\right).$$

Thus

$$\sup_{\overline{p'q'} \ni 0} \ln \frac{|p'|}{|q'|} = O\left(\ln \frac{s}{\epsilon}\right), \tag{89}$$

where the supremum is taken over all chords of K containing the origin. By (89) and Theorem 7 and Theorem 8, it follows that

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

and therefore

$$\ln \mathrm{vol}\hat{K}_\epsilon \leq O\left(n\ln\left(\frac{s\nu}{\epsilon}\right)\right). \tag{90}$$

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

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

Then,

$$\begin{aligned}
\int_{\hat{K}} \rho(x)^2 dx &\geq \int_{\hat{K}_\epsilon} \rho(x)^2 dx & (91)\\
&\geq \frac{\delta^2}{\mathrm{vol}\hat{K}_\epsilon} & (92)\\
&\geq \delta^2 \exp\left(-O\left(n\ln\left(\frac{s\nu}{\epsilon}\right)\right)\right). \quad \text{By (90).} & (93)
\end{aligned}$$

□

Using the above Lemma 16, if $\delta \geq \mathbb{P}\left[c^T x_t \geq 1-\epsilon\right]$, we have

$$t(\beta_{fat})^2 \leq O\left(n\ln\left(\frac{s\nu}{\epsilon}\right)\right) - \ln \mathbb{P}\left[c^T x_t \geq 1-\epsilon\right].$$

Therefore,

$$\mathbb{P}\left[c^T x_t \leq 1-\epsilon\right] \leq \exp\left(O\left(n\ln\left(\frac{s\nu}{\epsilon}\right)\right) - t(\beta_{fat})^2\right).$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left[(\exists t \geq \tau)\, c^T x_\tau \leq 1-\epsilon\right] &\leq \sum_{t\geq\tau} \exp\left(O\left(n\ln\left(\frac{s\nu}{\epsilon}\right)\right) - t(\beta_{fat})^2\right) & (94)\\
&\leq \frac{\exp\left(O\left(n\ln\left(\frac{s\nu}{\epsilon}\right)\right) - \tau(\beta_{fat})^2\right)}{1-\exp\left(-(\beta_{fat})^2\right)}. & (95)
\end{aligned}$$

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

$$\tau = O\left(\frac{1}{(\beta_{fat})^2}\left(\ln\frac{1}{\delta} + \left(n\ln\frac{s\nu}{\epsilon}\right)\right)\right),$$

which together with Theorem 5 completes the proof. □

9 Acknowledgements

I thank Ravi Kannan for stimulating conversations and Partha Niyogi for insights that motivated the view of Dikin walk as a random walk on a manifold.

References

[1] K. M. Ball, "An elementary introduction to modern convex geometry. In S. Levy (Ed.)," *Flavors of Geometry*, Number **31** in MSRI Publications, pp. 158, New York: Cambridge U. Press, 1997.

[2] F. Barthe, "Log-concave and spherical models in isoperimetry," *GAFA,* Vol. 12 (2002), 32–55

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

[4] D. Bertsimas and S. Vempala, "Solving convex programs by random walks," *Journal of the ACM (JACM)*, 2004, **51(4)**.

[5] S. G. Bobkov and C. Houdré, "Isoperimetric constants for product probability measures," *The Annals of Probability,* (1997), Vol. 25, No. 1, 184–205

[6] M. Dyer, A. Frieze and R. Kannan, "A random polynomial time algorithm for approximating the volume of convex sets," 1991, *Journal of the Association for Computing Machinary,* **38**, pp. 1-17,

[7] , R. Freund and J. Vera, "Equivalence of convex problem geometry and computational complexity in the separation oracle model," *Mathematics of Operations Research*, Vol. 34, No. 4, November 2009, pp. 869-879

[8] M. Gromov, "Isoperimetry of waists and concentration of maps," *Geom. Funct. Anal. 13,* 2003, **No. 1**, pp. 178215.

[9] O. Güler, "Hyperbolic polynomials and interior point methods for convex programming," *Mathematics of Operations Research,* May 1997, **v.22 n.2**, pp. 350-377.

[10] R. Kannan, L. Lovász and M. Simonovits, "Random walks and an $O^*(n^5)$ volume algorithm for convex bodies," *Random Structures and Algorithms*, August 1997, **11(1)**, pp. 1-50.

[11] R. Kannan and H. Narayanan, "Random walks on polytopes and an affine interior point algorithm for linear programming," *Proceedings of the ACM Symposium on Theory of Computing,* 2009, pp. 561-570.

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

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

[14] N. K. Karmarkar, "On the Riemannian Geometry underlying interior point methods," Mathematical Developments Arising from linear programming. *Contemporary Mathematics, Amer. Math. Soc.,* 1990, **144**.

[15] G. Lebeau, L. Michel, "Semiclassical analysis of a random walk on a manifold," http://arxiv.org/abs/0802.0644"

[16] L. Lovász, "Hit-and-run mixes fast," *Math. Programming, series A,* 1999, **86**, pp. 443-461.

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

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

[19] H. Narayanan, "Diffusion in Computer Science and Statistics," *Ph.D thesis*, University of Chicago, August 2009

[20] H. Narayanan and P. Niyogi, "Sampling Hypersurfaces through Diffusion," *12th Intl. Workshop on Randomization and Computation (RANDOM),* August 2008

[21] Y. Nesterov, "Constructing Self-concordant Barriers for Convex Cones," *CORE Discussion Paper No. 2006/30*, March 2006.

[22] Y. Nesterov and A Nemirovski, "Primal Central Paths and Riemannian Distances for Convex Sets," *Foundations of Computational Mathematics* 8(5), 533-560 (2008)

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

[24] Yu. Nesterov and A. Nemirovski, "Multi-parameter surfaces of analytic centers and long-step surface-following interior point methods," *Mathematics of Operations Research* v. 23 (1998), 1-38.

[25] A. Nemirovski, "Interior Point Polynomial Time Methods in Convex Programming (Lecture notes)," `http://www2.isye.gatech.edu/~nemirovs/Lect_IPM.pdf` (1994)

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

[27] Y. E. Nesterov and M.J. Todd, "On the Riemannian geometry defined by self-concordant barriers and interior-point methods," *Foundations of Computational Mathematics*, **2** (2002), 333–361, (Foundations of Computational Mathematics, 2 (2002), 333–361,

[28] M.V. Ramana, An exact duality theory for semidefinite programming and its complexity implications. *Mathematical Programming*, 1997, **77**, 129-162.

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

[30] S. Vempala, "Geometric Random Walks: A Survey,", *Combinatorial and Computational Geometry*, MSRI Publications Volume 52, 2005