Strong Self-Concordance and Sampling

Aditi Laddha
aladdha6@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Yin Tat Lee
yintat@uw.edu
University of Washington and
Microsoft Research
USA

Santosh Vempala
vempala@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

Motivated by the Dikin walk, we develop aspects of the interior-point theory for sampling in high dimension. Specifically, we introduce the notions of strong self-concordance and symmetry for a barrier. These properties imply that the Dikin walk defined using a strongly self-concordant barrier with symmetry parameter $\bar{\nu}$ mixes in $\tilde{O}(n\bar{\nu})$ steps from a warm start for a convex body in $\mathbb{R}^n$. For many natural barriers, $\bar{\nu}$ is roughly bounded by ν, the standard self-concordance parameter. We also show that these properties hold for the Lee-Sidford barrier. As a consequence, we obtain the first walk that mixes in $\tilde{O}(n^2)$ steps for an arbitrary polytope in $\mathbb{R}^n$. Strong self-concordance for other barriers leads to an interesting (and unexpected) connection — for the universal and entropic barriers, it is implied by the KLS conjecture.

CCS CONCEPTS

• **Theory of computation** → **Random walks and Markov chains**; • **Mathematics of computing** → *Markov-chain Monte Carlo convergence measures.*

KEYWORDS

markov chains, sampling, polytopes, self-concordance

ACM Reference Format:
Aditi Laddha, Yin Tat Lee, and Santosh Vempala. 2020. Strong Self-Concordance and Sampling. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC '20), June 22–26, 2020, Chicago, IL, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3357713.3384272

1 INTRODUCTION

The interior-point method is one of the major successes of optimization, in theory and practice [10, 29, 33]. It has led to the currently asymptotically fastest algorithms for solving linear and semidefinite programs and is a popular method for the accurate solution of medium to large-sized instances. The results of Nesterov and Nemirovski [27] demonstrate that $\nu = O(n)$ is possible for any convex set using their universal barrier, where ν is the self-concordance parameter of the barrier. For linear programming with feasible region $\{x : Ax \geq b\}$, the simple logarithmic barrier $g(x) = -\sum_i \ln((Ax - b)_i)$ has $\nu = O(m)$ for an $m \times n$ constraint matrix A, and is efficiently computable (the universal barrier is poly-time to estimate, but requires the computation of volume of a convex body). Over the past decade, Lee and Sidford [12–14] introduced a barrier for linear programming that achieves $\nu = O(n \log^{O(1)}(m))$ while being efficiently computable. The interior-point method has also directly influenced the design of combinatorial algorithms, leading to faster methods for maxflow/mincut and other optimization problems [4, 11, 24, 25, 28, 31, 32].

Sampling convex bodies is a fundamental problem that has close connections to convex optimization. Indeed, convex optimization can be reduced to sampling [1]. The most general methods that lead to polynomial-time sampling algorithms are the ball walk and hit-and-run, both requiring only membership oracle access to the convex set being sampled. These methods are not affine-invariant, i.e., their complexity depends on the affine position of the convex set. A tight bound on their complexity is $O^*\left(n^2R^2/r^2\right)$ where the convex body contains a ball of radius r and is mostly contained in a ball of radius R [8, 20, 21, 23]. The ratio R/r can be made $O(\sqrt{n})$ for any convex body by a suitable affine transformation. This effectively makes the complexity $O^*(n^3)$. However, the rounding (e.g., by near-isotropic transformation) is an expensive step, and its current best complexity is $O^*(n^4)$ [22]. Even for polytopes, this the rounding/isotropic step takes $O(mn^{4.5})$ total time for a polytope with m inequalities using an improved amortized analysis of the per-step complexity [26].

Interior-point theory offers an alternative sampling method with no need for rounding. A convex barrier function, via its Hessian, naturally defines an ellipsoid centered at each interior point of a convex body, the *Dikin* ellipsoid, which is always contained in the body. The Dikin walk, at each step, picks a uniformly random point in the Dikin ellipsoid around the current point. To ensure a uniform stationary density, the new point is accepted with a probability that depends on the ratio of the volumes of the Dikin ellipsoids at the two points, see Algorithm 1 below. Kannan and Narayanan [9] showed that the mixing rate of this walk with the standard logarithmic barrier is $O(mn)$ for a polytope in $\mathbb{R}^n$ defined using m inequalities. Each step of the walk involves computing the determinant and can be done in time $O(mn^{\omega-1})$, leading to an overall arithmetic complexity of $O(m^2n^{\omega})$ (see also [30] for a shorter proof of a Gaussian variant). Using a different more continuous approach, where each step is the solution of an ODE (rather than a straight-line step), Lee and Vempala [17] showed that the Riemannian Hamiltonian Monte Carlo improves the mixing rate for polytopes to $O(mn^{2/3})$ while keeping the same per-step complexity. This leads to the following basic questions:

- What is the fastest possible mixing rate of a Dikin walk?

STOC '20, June 22–26, 2020, Chicago, IL, USA

ACM ISBN 978-1-4503-6979-4/20/06…$15.00
https://doi.org/10.1145/3357713.3384272

- Is a mixing rate of $O(n)$ possible while keeping each step efficient (say matrix multiplication time or less)?

These are the natural analogies to the progress in optimization, where for the first, Nesterov and Nemirovski show a convergence rate to the optimum of $O(\sqrt{n})$, and for the second, Lee and Sidford show $\tilde{O}(\sqrt{n})$ for linear programming while maintaining efficiency.

These questions, in the context of sampling, lead to new challenges. Whereas for optimization, one step can be viewed as moving to the optimum of the objective in the current Dikin ellipsoid (a Newton step), for sampling, the next step is a random point in the Dikin ellipsoid; and since these ellipsoids have widely varying volumes, maintaining the correct stationary distribution takes some work.

To address these challenges, we use the symmetric self-concordance parameter $\bar{\nu}$. It is the smallest number such that for any point x in a convex body K, with unit Dikin ellipsoid E_x, we have $E_x \subseteq K \cap (2x - K) \subseteq \sqrt{\bar{\nu}} E_x$. In general $\bar{\nu}$ can be as high as ν^2 but for some important barriers, it is bounded as $O(\nu)$. This includes the logarithmic barrier, and, as we show, the Lee-Sidford(LS) barrier. This definition and parameter allows us to show that the isoperimetric (Cheeger) constant for the Dikin distance is asymptotically at least $1/\sqrt{\bar{\nu}}$.

We need a further, important refinement. The notion of self-concordance itself bounds the rate of change of the Hessian of the barrier (i.e., the Dikin matrix) with respect to the local metric in the spectral norm, i.e., the maximum change in any direction. We define *strong* self-concordance as the requirement that this derivative is bounded in *Frobenius* norm. Again, the logarithmic barrier satisfies this property, and we show that the Lee-Sidford barrier does as well.

Our main general result then is that the Dikin walk defined using any symmetric, strongly self-concordant barrier with convex Hessian mixes in $O(n\bar{\nu})$ steps. We prove that the LS barrier satisfies all these conditions with $\bar{\nu} = \tilde{O}(n)$ and so has a mixing rate of $\tilde{O}(n^2)$ for polytopes, completely answering the second question, and improving on several existing bounds in [3, 6]. We also show that the Dikin walk with the standard logarithmic barrier can be implemented in time $O(nnz(A) + n^2)$ where $nnz(A)$ is the number of nonzero entries in the constraint matrix A. This answers the open question posed in [9, 13]. These results along with earlier work on sampling polytopes are collected in Table 1. We note that while for the Dikin walk with a logarithmic barrier, there are simple examples showing that the mixing rate of $O(mn)$ is tight (take a hypercube and duplicate one of its facets $m-n$ times), for the Dikin walk with the LS barrier, we are not aware of a tight example or one with mixing rate greater than $\tilde{O}(n)$. There is the tantalizing possibility that it mixes in nearly linear time. Thus, the overall arithmetic complexity for sampling a polytope is reduced to $m \cdot \min\left\{\text{nnz}(A) \cdot n + n^3, n^{\omega+1}\right\}$ which improves the state of the art for all ranges of m.

We also study the notions of symmetry and strong self-concordance introduced in this paper for three well-studied barriers, namely, the classical universal barrier [27], the entropic barrier [2] and the canonical barrier [7]. While these barriers are not particularly efficient to evaluate, they are interesting because all of them achieve the best (or nearly best) possible self-concordance parameter values for arbitrary convex sets and convex cones (for the canonical barrier), and have played an important role in shaping the theory of interior-point methods for optimization. For the canonical barrier, the work of Hildebrand already establishes the convexity of the log determinant function (by definition of the barrier), and strong self-concordance [7]. For the entropic and universal barriers, we present an unexpected connection: the strong self-concordance is implied by the KLS isoperimetry conjecture! This suggests the possibility of more fruitful connections yet to be discovered using the notion of strong self-concordance.

[1] These entries are for general convex bodies presented by oracles, with R/r measuring the *roundness* of the input body; this can be made $O(\sqrt{n})$ with a rounding procedure that takes n^4 steps (membership queries). *After* rounding, the amortized per-step complexity of the ball walk in a polytope is $\tilde{O}(m)$.

Table 1: The complexity of uniform polytope sampling from a warm start.

Markov Chain	Mixing Rate	Per step cost
Ball Walk[1][8]	n^2R^2/r^2	mn
Hit-and-Run[1][21]	n^2R^2/r^2	mn
Dikin [9]	mn	$mn^{\omega-1}$
RHMC [17]	$mn^{\frac{2}{3}}$	$mn^{\omega-1}$
Geodesic Walk[16]	$mn^{\frac{3}{4}}$	$mn^{\omega-1}$
John's Walk[6]	n^7	$mn^4 + n^8$
Vaidya Walk[3]	$m^{\frac{1}{2}}n^{\frac{3}{2}}$	$mn^{\omega-1}$
Approximate John Walk[3]	$n^{2.5}$	$mn^{\omega-1}$
Dikin (this paper)	mn	$\text{nnz}(A) + n^2$
Weighted Dikin (this paper)	n^2	$mn^{\omega-1}$

1.1 Dikin Walk

The general Dikin walk is defined as follows. For a convex set P with a positive definite matrix $\mathbf{H}(u)$ for each point $u \in P$, let

$$E_u(r) = \left\{x : (x-u)^\top \mathbf{H}(u)(x-u) \le r^2\right\}.$$

Algorithm 1: `DikinWalk`

input : starting point x_0 in a polytope $P = \{x : \mathbf{A}x \ge b\}$
output: x_T
Set $r = \frac{1}{512}$
for $t \leftarrow 1$ **to** T **do**
  $x_t \leftarrow x_{t-1}$
  Pick y from $E_{x_t}(r)$
  $x_t \leftarrow y$ with probability $\min\left\{1, \frac{\text{vol}(E_{x_t}(r))}{\text{vol}(E_y(r))}\right\}$
end

1.2 Strong Self-Concordance

We define some properties for matrix functions. Usually but not necessarily, these matrices come from the Hessian of some convex function.

Definition 1 (Self-concordance)**.** For any convex set $K \subseteq \mathbb{R}^n$, we call a matrix function $\mathbf{H} : K \to \mathbb{R}^{n\times n}$ self-concordant if for any $x \in K$, we have

$$-2\|h\|_{\mathbf{H}(x)}\mathbf{H}(x) \preceq \frac{d}{dt}\mathbf{H}(x+th) \preceq 2\|h\|_{\mathbf{H}(x)}\mathbf{H}(x).$$

Definition 2 ($\bar{\nu}$-Symmetry)**.** For any convex set $K \subseteq \mathbb{R}^n$, we call a matrix function $\mathbf{H} : K \to \mathbb{R}^{n\times n}$ $\bar{\nu}$-symmetric if for any $x \in K$, we have

$$E_x(1) \subseteq K \cap (2x - K) \subseteq E_x(\sqrt{\bar{\nu}}).$$

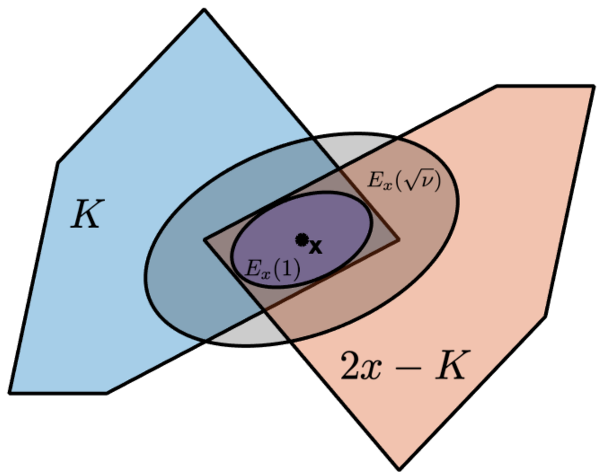

Figure 1: $E_u(1) \subseteq K \cap (2u - K) \subseteq E_u(\sqrt{\bar{\nu}})$.

The following lemma shows that self-concordant matrix functions also enjoy a similar regularity as the usual self-concordant functions.

Lemma 1.1. *Given any self-concordant matrix function* $\mathbf{H}$ *on* $K \subseteq \mathbb{R}^n$*, we define* $\|v\|_x^2 = v^\top \mathbf{H}(x)v$*. Then, for any* $x, y \in K$ *with* $\|x - y\|_x < 1$*, we have*

$$(1 - \|x-y\|_x)^2\, \mathbf{H}(x) \preceq \mathbf{H}(y) \preceq \frac{1}{(1-\|x-y\|_x)^2}\mathbf{H}(x).$$

Proof in A.1. Many natural barriers, including the logarithmic barrier and the LS-barrier, satisfy a much stronger condition than self-concordance, which we define here.

Definition 3 (Strong Self-Concordance)**.** For any convex set $K \subseteq \mathbb{R}^n$, we say a matrix function $\mathbf{H} : K \to \mathbb{R}^{n\times n}$ is strongly self-concordant if for any $x \in K$, we have

$$\left\|\mathbf{H}(x)^{-1/2} D\mathbf{H}(x)[h]\mathbf{H}(x)^{-1/2}\right\|_F \le 2\,\|h\|_x$$

where $D\mathbf{H}(x)[h]$ is the directional derivative of $\mathbf{H}$ at x in the direction h.

Similar to Lemma 1.1, we have a global version of strong self-concordance.

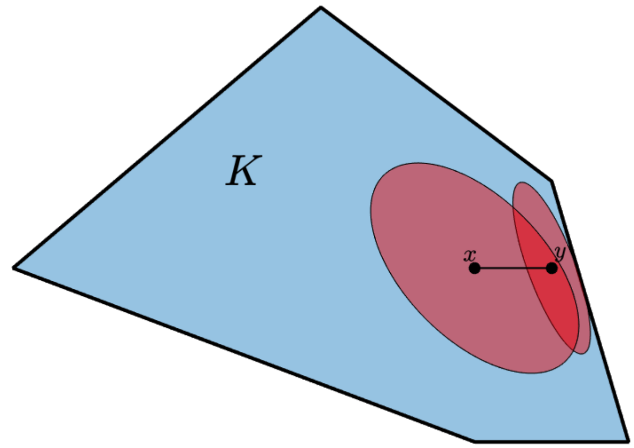

Figure 2: Strong self-concordance measures the rate of change of Hessian of a barrier in the Frobenius norm

Lemma 1.2. *Given any strongly self-concordant matrix function* $\mathbf{H}$ *on* $K \subset \mathbb{R}^n$*. For any* $x, y \in K$ *with* $\|x - y\|_x < 1$*, we have*

$$\|\mathbf{H}(x)^{-\frac{1}{2}}(\mathbf{H}(y) - \mathbf{H}(x))\mathbf{H}(x)^{-\frac{1}{2}}\|_F \le \frac{\|x-y\|_x}{(1-\|x-y\|_x)^2}.$$

Proof in A.2. We note that strong self-concordance is stronger than self-concordance since the Frobenius norm is always larger or equal to the spectral norm. As an example, we will verify that the conditions hold for the standard log barrier (Lemma 4.1).

1.3 Results

Our first theorem is the following.

Theorem 1.3. *The mixing rate of the Dikin walk for a* $\bar{\nu}$*-symmetric, strongly self-concordant matrix function with convex log determinant is* $O(n\bar{\nu})$*.*

This implies faster mixing and sampling for polytopes using the LS barrier (see Sec. 3.1 for the definition).

Theorem 1.4. *The mixing rate of the Dikin walk based on the LS matrix for any polytope in* $\mathbb{R}^n$ *is* $\tilde{O}(n^2)$ *and each step can be implemented in* $\tilde{O}(mn^{\omega-1})$[2] *arithmetic operations.*

On a related note, we show that each step of the standard Dikin walk is fast, and does not need matrix multiplication.

Theorem 1.5. *The Dikin walk with the logarithmic barrier for a polytope* $\{\mathbf{A}x \ge b\}$ *can be implemented in time* $O(\mathrm{nnz}(\mathbf{A}) + n^2)$ *per step while maintaining the mixing rate of* $O(mn)$*. See 4.*

The next lemma results from studying strong self-concordance for classical barriers. The KLS constant below is conjectured to be $O(1)$ and known to be $O(n^{\frac{1}{4}})$ [15].

Lemma 1.6. *Let* ψ_n *be the KLS constant of isotropic logconcave densities in* $\mathbb{R}^n$*, namely, for any isotropic logconcave density* p *and any set* $S \subset \mathbb{R}^n$*, we have*

$$\int_{\partial S} p(x)dx \ge \frac{1}{\psi_n}\min\left\{\int_S p(x)dx, \int_{\mathbb{R}^n\setminus S} p(x)dx\right\}.$$

[2] We use $\tilde{O}$ to hide factors polylogarithmic in n, m.

Let $\mathbf{H}(x)$ *be the Hessian of the universal or entropic barriers. Then, we have*

$$\left\|\mathbf{H}(x)^{-1/2}D\mathbf{H}(x)[h]\mathbf{H}(x)^{-1/2}\right\|_F = O(\psi_n)\,\|h\|_x\,.$$

In short, the universal and entropic barriers in $\mathbb{R}^n$ *are strongly self-concordant up to a scaling factor depending on* ψ_n.

In fact, our proof(see Section 5) shows that up to a logarithmic factor the strong self-concordance of these barriers is *equivalent* to the KLS conjecture.

2 MIXING WITH STRONG SELF-CONCORDANCE

To prove fast mixing, we will show that with a large probability, the Dikin ellipsoids at the current point and proposed next point have volumes within a constant factor; this would imply that a standard Metropolis filter succeeds with large probability and there is no "local" conductance bottleneck. For global convergence, the two important ingredients are showing that one-step distributions from nearby points have a large overlap and a suitable isoperimetric inequality. Both parts depart significantly from the Euclidean set-up as the notion of distance is defined by local Dikin ellipsoids. A key ingredient of the proof of Theorem 1.3 is the following lemma.

Lemma 2.1. *For two points* $x, y \in P$*, with* $\|x-y\|_x \le \frac{1}{512\sqrt{n}}$*, we have* $d_{TV}(P_x, P_y) \le \frac{3}{4}$.

Proof. Let $\mathcal{E}(x, \mathbf{A})$ denote the uniform distribution over an ellipsoid centered at x with covariance matrix $\mathbf{A}$ and radius $r = \frac{1}{512}$. Then,

$$d_{\text{TV}}(P_x, P_y) \le \frac{1}{2}\text{rej}_x + \frac{1}{2}\text{rej}_y + d_{\text{TV}}(\mathcal{E}(x, \mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(y))) \quad (1)$$

where rej_x and rej_y are the rejection probabilities at x and y. We break the proof into 2 parts. First we bound the rejection probability at x. Consider the algorithm picks a point z from $E_x(r)$. Let $f(z) = \ln\det \mathbf{H}(z)$. The acceptance probability of the sample z is

$$\min\left\{1, \frac{\text{vol}(E_x(r))}{\text{vol}(E_z(r))}\right\} = \min\left\{1, \sqrt{\frac{\det(\mathbf{H}(z))}{\det(\mathbf{H}(x))}}\right\}. \quad (2)$$

By our assumption f is a convex function, and hence

$$\ln\frac{\det(\mathbf{H}(z))}{\det(\mathbf{H}(x))} = f(z) - f(x) \ge \langle \nabla f(x), z - x\rangle. \quad (3)$$

$$\langle \nabla f(x), z-x\rangle = \langle \mathbf{H}(x)^{-\frac{1}{2}}\nabla f(x), \mathbf{H}(x)^{-\frac{1}{2}}(z-x)\rangle \quad (4)$$

where $z' = \mathbf{H}(x)^{-\frac{1}{2}}z$ is sampled from a ball of radius r centered at $x' = \mathbf{H}(x)^{-\frac{1}{2}}x$, and hence we know that

$$\Pr(v^\top(z'-x') \ge -\epsilon r\|v\|_2) \ge 1 - e^{-n\epsilon^2/2}.$$

In particular, with probability at least 0.99 in z, we have

$$\langle \nabla f(x), z-x\rangle \ge -\frac{4r}{\sqrt{n}}\|\mathbf{H}(x)^{-\frac{1}{2}}\nabla f(x)\|_2. \quad (5)$$

To compute $\|\mathbf{H}(x)^{-\frac{1}{2}}\nabla f(x)\|_2^2$, it is easier to compute the directional derivative of ∇f. Note that

$$\begin{aligned}
\|\mathbf{H}(x)^{-\frac{1}{2}}\nabla f(x)\|_2 &= \max_{\|v\|_2=1}\left(\mathbf{H}(x)^{-\frac{1}{2}}\nabla f(x)\right)^\top v \\
&= \max_{\|v\|_2=1}\text{Tr}(\mathbf{H}(x)^{-1}D\mathbf{H}(x)[\mathbf{H}(x)^{-\frac{1}{2}}v]) \\
&= \max_{\|u\|_x=1}\text{Tr}\left(\mathbf{H}(x)^{-\frac{1}{2}}D\mathbf{H}(x)[u]\mathbf{H}(x)^{-\frac{1}{2}}\right) \\
&\le \max_{\|u\|_x=1}\sqrt{n}\|\mathbf{H}(x)^{-\frac{1}{2}}D\mathbf{H}(x)[u]\mathbf{H}(x)^{-\frac{1}{2}}\|_F \\
&\le \max_{\|u\|_x=1} 2\sqrt{n}\|u\|_x \le 2\sqrt{n} \quad (6)
\end{aligned}$$

where the first inequality follows from $\left|\sum_{i=1}^n \lambda_i\right| \le \sqrt{n}\sqrt{\sum_{i=1}^n \lambda_i^2}$ and the second inequality follows from the definition of strong self-concordance.

Combining (2), (3), (5) and (6), we see that with probability at least 0.99 in z, the acceptance probability of the sample z is

$$\min\left\{1, \frac{\text{vol}(E_x(r))}{\text{vol}(E_z(r))}\right\} \ge e^{-4r} \ge 0.9922 \quad (7)$$

where we used that $r = \frac{1}{512}$. Hence, the rejection probability rej_x (and similarly rej_y) satisfies

$$\text{rej}_x \le 0.0039 \qquad \text{and} \qquad \text{rej}_y \le 0.0039. \quad (8)$$

To bound the second term, note that d_{TV} follows the triangle inequality. So, we can bound the second term in (1) as

$$\begin{aligned}
d_{\text{TV}}(\mathcal{E}(x, \mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(y))) \le\ & d_{\text{TV}}(\mathcal{E}(x, \mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(x))) \quad (9) \\
&+ d_{\text{TV}}(\mathcal{E}(y, \mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(y)))
\end{aligned}$$

By definition of d_{TV},

$$d_{\text{TV}}(\mathcal{E}(x, (\mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(y))) = \frac{1}{2}\frac{\text{vol}(E_x \backslash E_y)}{\text{vol}(E_x)} + \frac{1}{2}\frac{\text{vol}(E_y \backslash E_x)}{\text{vol}(E_y)} \quad (10)$$

The first term is a ratio of volumes and hence is invariant under the transformation $z \to \mathbf{H}(x)^{1/2}z$, after which it becomes the total variation distance between 2 balls of radius r whose centers are at a distance at most $\frac{r}{\sqrt{n}}$. To bound this, we use lemma 3.2 from [8],

$$d_{\text{TV}}(\mathcal{E}(x, \mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(x)) \le \frac{e}{e+1} \quad (11)$$

Now, we bound $d_{\text{TV}}(\mathcal{E}(y, \mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(y)))$. Let $Y_x = \{z : (z-y)^\top \mathbf{H}(x)(z-y) \le r^2\}$ and $Y_y = \{z : (z-y)^\top \mathbf{H}(y)(z-y) \le r^2\}$. Then,

$$\begin{aligned}
d_{\text{TV}}(\mathcal{E}(y, (\mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(y))) &= \frac{1}{2}\frac{\text{vol}(Y_x \backslash Y_y)}{\text{vol}(x)} + \frac{1}{2}\frac{\text{vol}(Y_y \backslash Y_x)}{\text{vol}(Y_y)} \quad (12) \\
&= 1 - \frac{1}{2}\frac{\text{vol}(Y_x \cap Y_y)}{\text{vol}(Y_x)} - \frac{1}{2}\frac{\text{vol}(Y_x \cap Y_y)}{\text{vol}(Y_y)} \quad (13)
\end{aligned}$$

We bound the total variation distance by bounding the fraction of volume in the intersection of the ellipsoids having the same center.

Again, we can assume that $\mathbf{H}(y) = \mathbf{I}$ and that $y = 0$. Then, strong self-concordance and Lemma 1.2 show that

$$\|\mathbf{I} - \mathbf{H}(x)^{-1}\|_F \le 2\|x - y\|_x \le \frac{1}{256\sqrt{n}}. \tag{14}$$

In particular, we have that

$$\frac{255}{256}\mathbf{I} \preceq \mathbf{H}(x)^{-1} \preceq \frac{257}{256}\mathbf{I}. \tag{15}$$

We partition the inverse eigenvalues, $\{\lambda_i\}_{i\in[n]}$ of $\mathbf{H}(x)$ into those with values at least 1 and the rest. Then consider the ellipsoid $\mathcal{I}$ whose inverse eigenvalues are $\min\{1, \lambda_i\}$ along the eigenvectors of $\mathbf{H}(x)$. This is contained in both Y_x and Y_y. We will see that $\mathrm{vol}(\mathcal{I})$ is a constant fraction of the volume of both Y_x and Y_y. First, we compare $\mathcal{I}$ and Y_y.

$$\begin{aligned}\frac{\mathrm{vol}(Y_x \cap Y_y)}{\mathrm{vol}(Y_y)} &\ge \frac{\mathrm{vol}(\mathcal{I})}{\mathrm{vol}(Y_y)} = \left(\prod_{i:\lambda_i<1} \lambda_i\right)^{1/2} \\ &= \left(\prod_{i:\lambda_i<1} (1 - (1 - \lambda_i))\right)^{1/2} \\ &\ge \exp\left(-\sum_{i:\lambda_i<1} (1 - \lambda_i)\right)\end{aligned} \tag{16}$$

where we used that $1 - x \ge \exp(-2x)$ for $0 \le x \le \frac{1}{2}$ and $\lambda_i \ge \frac{1}{2}$ (15). From the inequality (14), it follows that

$$\sqrt{\sum_i (\lambda_i - 1)^2} \le \frac{1}{256\sqrt{n}}.$$

Therefore, $\sum_{i:\lambda_i<1} |\lambda_i - 1| \le \frac{1}{256}$. Putting it into (16), we have

$$\frac{\mathrm{vol}(Y_x \cap Y_y)}{\mathrm{vol}(Y_y)} = \frac{\mathrm{vol}(\mathcal{I})}{\mathrm{vol}(Y_y)} \ge e^{-\frac{1}{256}}. \tag{17}$$

Similarly, we have

$$\begin{aligned}\frac{\mathrm{vol}(Y_x \cap Y_y)}{\mathrm{vol}(Y_x)} &\ge \left(\frac{\prod_{i:\lambda_i<1} \lambda_i}{\prod_{i:\lambda_i} \lambda_i}\right)^{1/2} = \left(\frac{1}{\prod_{i:\lambda_i>1} \lambda_i}\right)^{1/2} \\ &\ge \left(\frac{1}{\exp(\sum_{i:\lambda_i>1}(\lambda_i - 1))}\right)^{1/2} \ge e^{-\frac{1}{512}}.\end{aligned} \tag{18}$$

Putting (17) and (18) into (13), we have

$$d_{\mathrm{TV}}(\mathcal{E}(y, \mathbf{H}(x)), \mathcal{E}(y, \mathbf{H}(y))) \le 1 - \frac{e^{-\frac{1}{256}}}{2} - \frac{e^{-\frac{1}{512}}}{2} \tag{19}$$

Putting (8), (11) and (19) into (1), we have

$$d_{\mathrm{TV}}(P_x, P_y) \le \frac{0.0039}{2} + \frac{0.0039}{2} + 1 - \frac{e^{-\frac{1}{256}}}{2} - \frac{e^{-\frac{1}{512}}}{2} + \frac{e}{e+1} \le \frac{3}{4}$$

□

The next lemma establishes isoperimetry and only needs the symmetric containment assumption. This isoperimetry is for the cross-ratio distance. For a convex body K, and any two points $x, y \in K$, suppose that p, q are the endpoints of the chord through x, y in K, so that these points occur in the order p, x, y, q. Then, the *cross-ratio* distance between x and y is defined as

$$d_K(x, y) = \frac{\|x - y\|_2 \|p - q\|_2}{\|p - x\|_2 \|y - q\|_2}.$$

This distance enjoys the following isoperimetric inequality.

Theorem 2.2 ([18]). *For any convex body K, and disjoint subsets S_1, S_2 of it, and $S_3 = K \setminus S_1 \setminus S_2$, we have*

$$\mathrm{vol}(S_3) \ge d_K(S_1, S_2) \frac{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}{\mathrm{vol}(K)}.$$

We now relate the cross-ratio distance to the ellipsoidal norm.

Lemma 2.3. *For any $x, y \in K$, $d_K(x, y) \ge \frac{\|x-y\|_x}{\sqrt{\bar{\nu}}}$.*

Proof. Consider the Dikin ellipsoid at x. For the chord $[p, q]$ induced by x, y with these points in the order p, x, y, q, suppose that $\|p - x\|_2 \le \|y - q\|_2$. Then by Lemma 1.3, $p \in K \cap (2x - K)$. And hence $\|p - x\|_x \le \sqrt{\bar{\nu}}$. Therefore,

$$\begin{aligned}d_K(x, y) = \frac{\|x - y\|_2 \|p - q\|_2}{\|p - x\|_2 \|y - q\|_2} &\ge \frac{\|x - y\|_2}{\|p - x\|_2} \\ &= \frac{\|x - y\|_x}{\|p - x\|_x} \ge \frac{\|x - y\|_x}{\sqrt{\bar{\nu}}}.\end{aligned}$$

□

We can now prove the main conductance bound.

Theorem 1.3. *The mixing rate of the Dikin walk for a $\bar{\nu}$-symmetric, strongly self-concordant matrix function with convex log determinant is $O(n\bar{\nu})$.*

Proof. We follow the standard high-level outline [35]. Consider any measurable subset $S_1 \subseteq K$ and let $S_2 = K \setminus S_1$ be its complement. Define the points with low escape probability for these subsets as

$$S_i' = \left\{x \in S_i : P_x(K \setminus S_i) < \frac{1}{8}\right\}$$

and $S_3' = K \setminus S_1' \setminus S_2'$. Then, for any $u \in S_1'$, $v \in S_2'$, we have $d_{TV}(P_u, P_v) > 1 - \frac{1}{4}$. Hence, by Lemma 2.1, we have $\|u - v\|_u \ge \frac{1}{512\sqrt{n}}$. Therefore, by Lemma 2.3,

$$d_K(u, v) \ge \frac{1}{512\sqrt{n} \cdot \sqrt{\bar{\nu}}}.$$

We can now bound the conductance of S_1. We may assume that $\mathrm{vol}(S_i') \ge \mathrm{vol}(S_i)/2$; otherwise, it immediately follows that the conductance of S_1 is $\Omega(1)$. Assuming this, we have

$$\begin{aligned}\int_{S_1} P_x(S_2)\, dx &\ge \int_{S_3'} \frac{1}{8} dx \ge \frac{1}{8}\mathrm{vol}(S_3') \\ &\ge \frac{1}{8} d_K(S_1', S_2') \frac{\mathrm{vol}(S_1')\mathrm{vol}(S_2')}{\mathrm{vol}(P)} \qquad \text{(from Thm 2.2)} \\ &\ge \frac{1}{32768\sqrt{n\bar{\nu}}} \min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}.\end{aligned}$$

□

It is well-known that the inverse squared conductance of a Markov Chain is a bound on its mixing rate, e.g., in the following form.

Theorem 2.4. *[19] Let Q_t be the distribution of the current point after t steps of a Markov chain with stationary distribution Q and conductance at least ϕ, starting from initial distribution Q_0. Then, with $M = \sup_A \frac{Q_0(A)}{Q(A)}$,*

$$d_{TV}(Q_t, Q) \le \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t$$

where $d_{TV}(Q_t, Q)$ is the total variation distance between Q_t and Q.

3 FAST POLYTOPE SAMPLING WITH THE LS BARRIER

3.1 LS Barrier

In this section, we assume the convex set is a polytope $P = \{x \in \mathbb{R}^n | \mathbf{A}x > b\}$. For any $x \in \text{int}P$, let $\mathbf{S}_x = \mathbf{Diag}(\mathbf{A}x - b)$ and $\mathbf{A}_x = \mathbf{S}_x^{-1}\mathbf{A}$. We state the definition of the Lee-Sidford barrier [14], henceforth referred to as the LS barrier.

Definition 4 (LS Barrier)**.** The LS barrier is defined as

$$\psi(x) = \max_{w \in \mathbb{R}^m : w \ge 0} \frac{1}{2} f(x, w)$$

where

$$f(x, w) = \ln\det\left(\mathbf{A}_x \mathbf{W}^{1-\frac{2}{q}} \mathbf{A}_x\right) - \left(\frac{1}{2} - \frac{1}{q}\right)\sum_{i=1}^m w_i$$

and $\mathbf{W} = \mathbf{Diag}(w)$, and $q = 2(1 + \ln m)$.

We follow the notation in [14]:

Definition 5. For any $x \in P$, we define $w_x = \arg\max_{w \ge 0} f(x, w)$, $\mathbf{W}_x = \mathbf{Diag}(w_x)$, $s_x = \mathbf{A}x - b$, $\mathbf{S}_x = \mathbf{Diag}(s_x)$, $\mathbf{A}_x = \mathbf{S}_x^{-1}\mathbf{A}$, $\mathbf{P}_x = \mathbf{W}_x^{\frac{1}{2}-\frac{1}{q}} \mathbf{A}_x \left(\mathbf{A}_x \mathbf{W}_x^{1-\frac{2}{q}} \mathbf{A}_x\right)^{-1} (\mathbf{W}_x^{\frac{1}{2}-\frac{1}{q}} \mathbf{A}_x)^\top$, $\sigma_x = \text{diag}(\mathbf{P}_x)$, $\Sigma_x = \mathbf{Diag}(\sigma_x)$, $\mathbf{P}_x^{(2)} = \mathbf{P}_x \circ \mathbf{P}_x$, $\Lambda_x = \Sigma_x - \mathbf{P}_x^{(2)}$, $\bar{\Lambda}_x = \Sigma_x^{-1/2}\Lambda_x\Sigma_x^{-1/2}$, and $\mathbf{N}_x = 2\bar{\Lambda}_x(\mathbf{I} - (1 - \frac{2}{q})\bar{\Lambda}_x)^{-1}$.

3.2 Properties of LS Barrier

Lemma 3.1 ([14])**.** *The function $\psi(x)$ has the following properties:*

1. *(Lemma 23) $\psi(x)$ is convex.*
2. *(Lemma 47.2)*
$$\mathbf{P}_x^{(2)} \preceq \Sigma_x \tag{20}$$
3. *(Lemma 31)*
$$0 \le \sigma_{x,i} = w_{x,i} \le 1 \tag{21}$$
$$\mathbf{A}_x^\top \mathbf{W}_x \mathbf{A}_x \preceq \nabla^2\psi(x) \preceq (1+q)\mathbf{A}_x^\top \mathbf{W}_x \mathbf{A}_x \tag{22}$$
4. *(Lemma 33) For any $x_t = x + th$ and $s_t = \mathbf{A}x_t - b$, we have*
$$\|\mathbf{S}_t^{-1}\frac{d}{dt}s_t\|_{\mathbf{W}_t} \le \|h\|_{\nabla^2\psi(x_t)} \tag{23}$$
5. *(Lemma 34) For any $x_t = x + th$ and $w_t = w_{x_t}$, we have*
$$\|\mathbf{W}_t^{-1}\frac{d}{dt}w_t\|_{\mathbf{W}_t} \le q\|h\|_{\nabla^2\psi(x_t)} \tag{24}$$

3.3 Mixing Rate

Definition 6. The LS matrix for a point $x \in P$ is defined as

$$\mathbf{H}(x) = (1+q^2)(1+q)\cdot \mathbf{A}^\top \mathbf{S}_x^{-1} \mathbf{W}_x^{1-\frac{2}{q}} \mathbf{S}_x^{-1}\mathbf{A}.$$

We establish the strong self-concordance of LS Matrix in the next lemma.

Lemma 3.2 (Strong Self Concordance)**.** *The LS matrix is strongly self-concordant, i.e., for any $x_t \in P$ given by $x_t = x + th$ and $\mathbf{H}_t = \mathbf{H}(x_t)$, we have*

$$\|\mathbf{H}_t^{-1/2}(\frac{d}{dt}\mathbf{H}_t)\mathbf{H}_t^{-1/2}\|_F \le 2\|h\|_{\mathbf{H}_t}.$$

Proof. We redefine

$$\overline{\mathbf{H}}_t = \mathbf{A}^\top \mathbf{V}_t \mathbf{A}$$

with $\mathbf{V}_t = \mathbf{S}_t^{-1}\mathbf{W}_t^{1-2/q}\mathbf{S}_t^{-1}$, $\mathbf{P}_t = \sqrt{\mathbf{V}_t}\mathbf{A}(\mathbf{A}^\top\mathbf{V}_t\mathbf{A})^{-1}\mathbf{A}^\top\sqrt{\mathbf{V}_t}$. Note that $\mathbf{V}_t$ is a diagonal matrix and that $\overline{\mathbf{H}}_t$ and $\mathbf{H}_t$ are just off by a scaling factor. Hence, we have

$$\begin{aligned}
\|\mathbf{H}_t^{-1/2}(\frac{d}{dt}\mathbf{H}_t)\mathbf{H}_t^{-1/2}\|_F^2 &= \|\overline{\mathbf{H}}_t^{-1/2}(\frac{d}{dt}\overline{\mathbf{H}}_t)\overline{\mathbf{H}}_t^{-1/2}\|_F^2 \\
&= \text{Tr}\overline{\mathbf{H}}_t^{-1}(\frac{d}{dt}\overline{\mathbf{H}}_t)\overline{\mathbf{H}}_t^{-1}(\frac{d}{dt}\overline{\mathbf{H}}_t) \\
&= \text{Tr}\left((\mathbf{A}^\top\mathbf{V}_t\mathbf{A})^{-1}\mathbf{A}^\top(\frac{d}{dt}\mathbf{V}_t)\mathbf{A}\right)^2 \\
&= \text{Tr}\mathbf{P}_t\frac{d\ln\mathbf{V}_t}{dt}\mathbf{P}_t\frac{d\ln\mathbf{V}_t}{dt} \\
&= \frac{d\ln v_t}{dt}^\top \mathbf{P}_t^{(2)}\frac{d\ln v_t}{dt}.
\end{aligned}$$

Note that $\mathbf{P}_t^{(2)} \preceq \Sigma_t$, by (20). Therefore,

$$\begin{aligned}
\|\mathbf{H}_t^{-1/2}(\frac{d}{dt}\mathbf{H}_t)\mathbf{H}_t^{-1/2}\|_F^2 &\le \frac{d\ln v_t}{dt}^\top \Sigma_t \frac{d\ln v_t}{dt} \\
&= \sum_{i=1}^m \sigma_{t,i}\left(\frac{d\ln s_{t,i}^{-2}w_{t,i}^{1-2/q}}{dt}\right)^2 \\
&\le 4\sum_{i=1}^m \sigma_{t,i}\left(\left(\frac{d\ln s_{t,i}}{dt}\right)^2 + \left(\frac{d\ln w_{t,i}}{dt}\right)^2\right) \\
&= 4\sum_{i=1}^m \sigma_{t,i}\left(\left(\frac{1}{s_{t,i}}\frac{ds_{t,i}}{dt}\right)^2 + \left(\frac{1}{w_{t,i}}\frac{dw_{t,i}}{dt}\right)^2\right) \\
&\le 4(1+q^2)\|h\|_{\nabla^2\psi(x_t)}^2
\end{aligned}$$

where we used $\sigma_t = w_t$ (21) in the second last equation and equations (23) and (24) for the last inequality.

Finally, (22) shows that $\nabla^2\psi(x_t) \preccurlyeq (1+q)\mathbf{A}_t^\top\mathbf{W}_t\mathbf{A}_t$. Since $0 \le w_t = \sigma_t \le 1$ by the property of leverage score, we have

$$\nabla^2\psi(x) \preceq (1+q)\mathbf{A}_t^\top\mathbf{W}_t\mathbf{A}_t \preceq (1+q)\mathbf{A}_t^\top\mathbf{W}_t^{1-2/q}\mathbf{A}_t = (1+q)\overline{\mathbf{H}}_t.$$

Thus, $\|h\|_{\nabla^2\psi(x_t)}^2 \le (1+q)\|h\|_{\overline{\mathbf{H}}_t}^2$. Hence, we have

$$\|\mathbf{H}_t^{-1/2}(\frac{d}{dt}\mathbf{H}_t)\mathbf{H}_t^{-1/2}\|_F^2 \le 4(1+q^2)(1+q)\|h\|_{\overline{\mathbf{H}}_t}^2 \le 4\|h\|_{\mathbf{H}_t}^2$$

where we used that $\mathbf{H}_t = (1+q^2)(1+q)\overline{\mathbf{H}}_t$. □

Lemma 3.3. *The LS-ellipsoid matrix has the following properties:*

1. *$\ln\det\mathbf{H}(x)$ is convex.*
2. *$\mathbf{H}$ is $\bar{\nu}$-symmetric matrix function with $\bar{\nu} = O(n\log^3 m)$.*

Proof. For any $x \in \text{int}P$, (21) shows that

$$\sum_i w_{x,i} = \sum_i \sigma_{x,i} = \text{Tr}\mathbf{W}_x^{\frac{1}{2}-\frac{1}{q}}\mathbf{A}_x\left(\mathbf{A}_x\mathbf{W}_x^{1-\frac{2}{q}}\mathbf{A}_x\right)^{-1}(\mathbf{W}_x^{\frac{1}{2}-\frac{1}{q}}\mathbf{A}_x)^\top = \text{Tr}\mathbf{I}_{n\times n} = n.$$

Hence, the LS barrier can be restated as

$$\psi(x) = \frac{1}{2}\ln\det(\mathbf{A}_x^\top\mathbf{W}_x^{1-2/q}\mathbf{A}_x) - \left(\frac{1}{2}-\frac{1}{q}\right)n = \frac{1}{2}\ln\det\frac{1}{(1+q^2)(1+q)}\mathbf{H}(x) - \left(\frac{1}{2}-\frac{1}{q}\right)n$$

where w_x is the maximizer of $f(x,w)$. Since $\psi(x)$ is convex, so is $\ln\det\mathbf{H}(x)$.

Next, we prove that $\bar{\nu} = O(n\log^3 m)$. For any $x \in P$ and any $y \in E_x(1)$, $(y-x)^\top\mathbf{A}_x\mathbf{W}_x^{1-2/q}\mathbf{A}_x(y-x) \le \frac{1}{(1+q^2)(1+q)}$ and hence

$$\begin{aligned}
&\|\mathbf{A}_x(y-x)\|_\infty^2 \\
&= \max_{i\in[m]}\left(e_i^\top\mathbf{A}_x(\mathbf{A}_x\mathbf{W}_x^{1-2/q}\mathbf{A}_x)^{-1/2}(\mathbf{A}_x\mathbf{W}_x^{1-2/q}\mathbf{A}_x)^{1/2}(y-x)\right)^2 \\
&\le \frac{1}{(1+q^2)(1+q)}\max_{i\in[m]} e_i^\top\mathbf{A}_x(\mathbf{A}_x\mathbf{W}_x^{1-2/q}\mathbf{A}_x)^{-1}\mathbf{A}_x e_i \\
&\le \max_{i\in[m]}\frac{\sigma_{x,i}}{w_{x,i}^{1-2/q}} \le \max_{i\in[m]}\frac{\sigma_{x,i}}{w_{x,i}} = 1
\end{aligned}$$

since $w_{x,i} \le 1$. So, $E_x \subseteq P\cap(2x-P)$ for all $x \in P$.

For any $y \in P\cap(2x-P)$, we have $\|\mathbf{S}_x^{-1}\mathbf{A}(x-y)\|_\infty \le 1$. Hence,

$$\begin{aligned}
\frac{(x-y)^T\mathbf{H}(x)(x-y)}{(1+q^2)(1+q)} &= (x-y)^T\mathbf{A}^\top\mathbf{S}_x^{-1}\mathbf{W}_x^{1-2/q}\mathbf{S}_x^{-1}\mathbf{A}(x-y) \\
&= \sum_{i=1}^m w_{x,i}^{1-2/q}(\mathbf{S}_x^{-1}\mathbf{A}(x-y))_i^2 \le \sum_{i=1}^m w_{x,i}^{1-2/q} \\
&\le \left(\sum_{i=1}^m \left(w_{x,i}^{1-2/q}\right)^{\frac{1}{1-(2/q)}}\right)^{1-2/q}\left(\sum_{i=1}^m 1^{q/2}\right)^{2/q} \\
&\le \left(\sum_{i=1}^m w_{x,i}\right)^{1-2/q} m^{2/q} \le n^{1-2/q}m^{2/q} \le en.
\end{aligned}$$

□

Lemmas 3.2 and 3.3 imply that mixing time of Dikin walk with LS matrix is $\tilde{O}(n^2)$ from a warm start. Implementing each step of this walk involves the following tasks:

(1) Compute $\mathbf{H}(x)^{-1/2}v$ for some vector v
(2) Compute the ratio $\det(\mathbf{H}(y)^{-1}\mathbf{H}(x))$ for points x, y.

Given w_x, w_y, computing $\mathbf{H}(x)$, its inverse and its determinant can all be done in time $\tilde{O}\left(mn^{\omega-1}\right)$. w_x can be updated in $\tilde{O}(mn^{\omega-1})$ per step as shown in [14, Theorem 46]. Using this, each step of Dikin walk with LS Matrix can be implemented in time $O(mn^{\omega-1})$ This means that the total time to sample a polytope from a warm start is $\tilde{O}(mn^{\omega+1})$ as claimed in Theorem 1.4.

4 FAST IMPLEMENTATION OF DIKIN WALK

Lemma 4.1 (Strong Self-Concordance). *The matrix function* $\mathbf{H}(x) = \mathbf{A}^\top\mathbf{S}_x^{-2}\mathbf{A}$ *which is the Hessian of the log barrier function* $\phi(x) = -\sum_{i=1}^m \log(A_i x - b_i)$*, is strongly self-concordant.*

Proof. Let $x_t = x+th$ for some fixed vector h. Let $\mathbf{S}_t = \mathbf{Diag}(\mathbf{A}x_t - b)$, $\mathbf{A}_t = \mathbf{S}_t^{-1}\mathbf{A}$, $\mathbf{P}_t = \mathbf{A}_t(\mathbf{A}_t^\top\mathbf{A}_t)^{-1}\mathbf{A}_t^\top$, $\sigma_t = \text{diag}(\mathbf{P}_t)$, $\Sigma_t = \mathbf{Diag}(\sigma_t)$, and $\mathbf{P}_t^{(2)} = \mathbf{P}_t\circ\mathbf{P}_t$. By [14, Lemma 47.2], $\mathbf{P}_t^{(2)} \preccurlyeq \Sigma_t \preceq \mathbf{I}$. We are now ready to prove strong self-concordance.

$$\begin{aligned}
&\|\mathbf{H}_t^{-1/2}(\frac{d}{dt}\mathbf{H}_t)\mathbf{H}_t^{-1/2}\|_F^2 \\
&= \text{Tr}\mathbf{H}_t^{-1}(\frac{d}{dt}\mathbf{H}_t)\mathbf{H}_t^{-1}(\frac{d}{dt}\mathbf{H}_t) = \text{Tr}\mathbf{P}_t\frac{d\ln s_t^{-2}}{dt}\mathbf{P}_t\frac{d\ln s_t^{-2}}{dt} \\
&= \frac{d\ln s_t^{-2}}{dt}^\top\mathbf{P}_t^{(2)}\frac{d\ln s_t^{-2}}{dt} \le \sum_{i=1}^m\left(\frac{d\ln s_t^{-2}}{dt}\right)^2 \\
&= \sum_{i=1}^m 4s_{t,i}^{-2}\left(a_i^\top h\right)^2 = 4h^\top\mathbf{A}^\top\mathbf{S}_t^{-2}\mathbf{A}h = 4\|h\|_{\mathbf{H}_t}^2.
\end{aligned}$$

□

The function $\log\det\mathbf{A}^\top\mathbf{S}_x^{-2}\mathbf{A}$ is called the volumetric barrier and is known to be convex.

Lemma 4.2 ([34, Lemma 3]). $f(x) = \log\det\mathbf{A}^\top\mathbf{S}_x^{-2}\mathbf{A}$ *is a convex function in* x.

The main result of this section is to give a faster implementation of log barrier based Dikin Walk by noting that we can avoid computing $\mathbf{H}(x)$ explicitly or its inverse or determinant for the Dikin walk with log barrier. This resolves an open problem posed in [9, 13].

The main challenge is to avoid computing the determinant of $\mathbf{H}(x)$. Instead, an unbiased estimator of the ratio of two such determinants suffices. We reduce this, first to estimating a log-det, and then to an inverse maintenance problem in the next two lemmas.

To calculate rejection probability for the Dikin Walk, we calculate an unbiased estimator of $\frac{\det\mathbf{H}(x)}{\det\mathbf{H}(y)}$. We first find an unbiased estimator, Y of the term $\log\det\mathbf{H}(x) - \log\det\mathbf{H}(y)$ which can be calculated in $\widetilde{O}\left(\text{nnz}(\mathbf{A})+n^2\right)$ time using lemma 4.4. We then find an unbiased estimator, X of the determinant of $\mathbf{H}(x)$ using lemma 4.3 which describes an algorithm to find an unbiased estimator of a value r given access to an unbiased estimator of $\log r$.

Lemma 4.3 (Determinant). *Given a random variable* Y *with* $\mathbb{E}(Y) = \log r$*, the random variable* X *defined as*

$$X = e\cdot\prod_{j=1}^i Y_j \text{ with probability } \frac{1}{e\cdot i!}$$

with Y_j *being i.i.d. copies of* Y *has* $\mathbb{E}(X) = r$.

Proof. We know that

$$r = \sum_{i=0}^\infty\frac{(\log(r))^i}{i!}.$$

Using $X = e\cdot\prod_{j=1}^i Y_j$ with probability $\frac{1}{e\cdot i!}$ where Y_j are i.i.d. random variables with $\mathbb{E}(Y_j) = \log r$. Then,

$$\mathbb{E}[X] = \sum_{i=0}^\infty\frac{\mathbb{E}(Y)^i}{i!} = e^{\log(r)} = r.$$

□

Lemma 4.4 (Log Determinant). *Define* $\mathbf{H}(t) = \mathbf{A}^\top\mathbf{A} + t(\mathbf{A}^\top\mathbf{W}\mathbf{A} - \mathbf{A}^\top\mathbf{A}) = \mathbf{A}^\top(\mathbf{I} + t(\mathbf{W} - \mathbf{I}))\mathbf{A}$. *Let* $v \sim N(0, I)$ *and* t *be uniform in* $[0, 1]$ *and*

$$Y = v^\top\mathbf{H}(t)^{-1}\mathbf{A}^\top(\mathbf{H} - I)\mathbf{A}v + \log\det\mathbf{A}^\top\mathbf{A}.$$

Then, $\mathbb{E}(Y) = \log\det\mathbf{A}^\top\mathbf{W}\mathbf{A}$.

Proof. We have

$$\begin{aligned}
&\log\det(\mathbf{H}(1)) - \log\det(\mathbf{A}^\top\mathbf{A})\\
&= \int_0^1 \frac{d\log\det\mathbf{H}(t)}{dt}dt\\
&= \int_0^1 \mathrm{Tr}(\mathbf{H}(t)^{-1}\frac{d\mathbf{H}(t)}{dt})dt\\
&= \int_0^1 \mathrm{Tr}(\mathbf{H}(t)^{-1}\mathbf{A}^\top(\mathbf{H} - \mathbf{I})\mathbf{A}^\top)dt\\
&= \mathbb{E}_{v\sim N(0,I)}[v^\top\int_0^1 \mathrm{Tr}(\mathbf{H}(t)^{-1}\mathbf{A}^\top(\mathbf{H} - \mathbf{I})\mathbf{A})dt \cdot v]\\
&= \int_0^1 \mathbb{E}_{v\sim N(0,I)}[v^\top\mathbf{H}(t)^{-1}\mathbf{A}^T(\mathbf{H} - \mathbf{I})\mathbf{A}v]dt
\end{aligned}$$

□

Note that given $\mathbf{H}(t)^{-1}$, we can estimate the last expression as the sum of $\mathbb{E}_{v\sim N(0,I)}[v^\top\mathbf{H}(t)^{-1}\mathbf{A}^\top(\mathbf{H} - \mathbf{I})\mathbf{A}v]$. Maintaining $\mathbf{H}(t)^{-1}$ reduces to the inverse maintainence problem for $\mathbf{H}$. It is shown in [13] that a matrix inverse can be maintained efficiently in the following sense. Suppose we have a sequence of matrices of the form $\mathbf{A}^\top\mathbf{D}^{(k)}\mathbf{A}$ where each $\mathbf{D}^{(k)}$ is a slowly-changing diagonal matrix. Then for each matrix in the sequence, its inverse times any given vector v can be computed in time $\widetilde{O}\left(\mathrm{nnz}(\mathbf{A}) + n^2\right)$. We use $\mathbf{W} = \mathbf{S_x}^{-2}\mathbf{S_y}^2$ to calculate as unbiased estimate of $\log\det\mathbf{H}(x) - \log\det\mathbf{H}(y)$.

Lemma 4.5 ([13, Theorem 13]). *Suppose that a sequence of matrices* $\mathbf{A}^\top\mathbf{D}^{(k)}\mathbf{A}$ *for the inverse maintenance problem satisfies the*

$$\sum_i\left(\frac{d_i^{(k+1)} - d_i^{(k)}}{d_i^{(k)}}\right)^2 = O(1).$$

Then there is an algorithm that with high probability maintains an $\widetilde{O}\left(\mathrm{nnz}(\mathbf{A}) + n^2\right)$*-time linear system solver for* r *rounds in total time* $\widetilde{O}\left(r(\mathrm{nnz}(\mathbf{A}) + n^2 + n^\omega)\right)$

We note that the condition $\sum_i\left(\frac{d_i^{(k+1)} - d_i^{(k)}}{d_i^{(k)}}\right)^2 = O(1)$ is satisfied since

$$\begin{aligned}
\sum_i\left(\frac{d_i^{(k+1)} - d_i^{(k)}}{d_i^{(k)}}\right)^2 &= \sum_i\left(\frac{(s_i^{(k+1)})^{-2} - (s_i^{(k)})^{-2}}{(s_i^{(k)})^{-2}}\right)^2\\
&= O\left(\sum_i\left(\frac{s_i^{(k+1)} - s_i^{(k)}}{s_i^{(k)}}\right)^2\right)\\
&= O\left(\|x^{(k+1)} - x^{(k)}\|^2_{x^{(k)}}\right).
\end{aligned}$$

Putting these together we have the following unbiased estimator for $\sqrt{\det\mathbf{H}(x)/\det\mathbf{H}(y)}$:

Compute $X = \frac{e}{2}\cdot\prod_{j=1}^i Y_j$ with probability $\dfrac{1}{e\cdot i!}$ where each Y_j is an i.i.d. sample generated as follows:

(1) Pick $v \sim N(0, I)$ and t uniformly in $[0, 1]$.
(2) Set $\mathbf{W} = \mathbf{S_x}^{-2}\mathbf{S_y}^2$.
(3) Compute $Y = v^\top\mathbf{H}(t)^{-1}\mathbf{A}^\top(\mathbf{H}(1) - \mathbf{I})\mathbf{A}v$ where $\mathbf{H}(t) = \mathbf{A}^\top(\mathbf{I} + t(\mathbf{W} - \mathbf{I}))\mathbf{A}$ using efficient inverse maintenance.

We need one more trick. In the algorithm, at each step we need to compute $\min\left\{1, \frac{p(y\to x)}{p(x\to y)}\right\}$. While we can approximate the ratio inside the min, this might make the overall probability incorrect due to the min function not being smooth. So instead we propose a smoother filter. This might have other applications.

Lemma 4.6 (Smooth Metropolis filter). *Let the probability of selecting the state* y *from the state* x *of an ergodic Markov chain be* $p(x \to y)$. *Then accepting the step* $x \to y$ *with probability* $\dfrac{p(y\to x)}{p(y\to x) + p(x\to y)}$ *gives the uniform stationary distribution.*

Proof. Let $\tilde{p}(x \to y)$ be the probability of taking a step from x to y. Then, $\tilde{p}$ satisfies detailed balance.

$$\begin{aligned}
\tilde{p}(x\to y) &= p(x\to y)\cdot\frac{p(y\to x)}{p(y\to x) + p(x\to y)}\\
&= \frac{p(x\to y)p(y\to x)}{p(y\to x) + p(x\to y)}\\
&= p(y\to x)\cdot\frac{p(x\to y)}{p(y\to x) + p(x\to y)}\\
&= \tilde{p}(y\to x)
\end{aligned}$$

So, $\tilde{p}(x \to y) = \tilde{p}(y \to x)$ for all x and y. Hence the stationary distribution of this Markov Chain is uniform. □

For the Dikin walk, $\frac{p(y\to x)}{p(x\to y)} = \sqrt{\frac{\det(\mathbf{H}_y)}{\det(\mathbf{H}_x)}}$. Note that the rejection probability function $\frac{p(y\to x)}{p(y\to x)+p(x\to y)} = \frac{\frac{p(y\to x)}{p(x\to y)}}{1+\frac{p(y\to x)}{p(x\to y)}}$ is increasing in $\frac{p(y\to x)}{p(x\to y)}$. As Dikin barrier is strongly self-concordant (Lemma 4.1) and by (7), we get that with probability at least 0.99, for y randomly drawn from E_x, $\frac{\mathrm{vol}(E_x(r))}{\mathrm{vol}(E_y(r))} \geq 0.9922$ from equation (7). Hence, the probability of not rejecting at each step at least 0.498 with large probability.

Proof of Theorem 1.5. Implementing Dikin walk requires maintaining matrices $\mathbf{H}_t = \mathbf{A}^\top\mathbf{S}_t^{-2}\mathbf{A}$ corresponding to point x_t. 4.5 shows that this can be done in $\widetilde{O}\left(n^\omega + r(nnz(\mathbf{A}) + n^2)\right)$ time where r is the number of steps in the chain. Additionally, each step requires calculating the rejection probability which is a smooth function in $\dfrac{\det(\mathbf{H}_t)}{\det(\mathbf{H}_{t+1})}$ and hence can be calculated in $\widetilde{O}\left(nnz(\mathbf{A}) + n^2\right)$ amortized time using lemmas 4.3 and 4.4. □

5 STRONG SELF-CONCORDANCE OF OTHER BARRIERS

Here we analyze the strong self-concordance of the universal and entropic barriers.

Proof of Lemma 1.6. The entropic barrier is the dual of

$$f(\theta)=\log(\int_{x\in K}\exp(\theta^\top x)dx).$$

Its first three derivatives are the moments [2]:

$$\begin{aligned}Df(\theta)[h_1]&=\frac{\int_{x\in K}x^\top h_1\exp(\theta^\top x)dx}{\int_{x\in K}\exp(\theta^\top x)dx}\\&=\mathbb{E}_{x\sim p_\theta}x^\top h_1.\end{aligned}$$

where p_θ is the corresponding exponential distribution with support K.

$$\begin{aligned}D^2f(\theta)[h_1,h_2]=&\frac{\int_{x\in K}x^\top h_1x^\top h_2\exp(\theta^\top x)dx}{\int_{x\in K}\exp(\theta^\top x)dx}\\&-\frac{\prod_{i=1}^2\int_{x\in K}x^\top h_i\exp(\theta^\top x)dx}{\left(\int_{x\in K}\exp(\theta^\top x)dx\right)^2}\\=&\,\mathbb{E}_{x\sim p_\theta}h_2^\top xx^\top h_1-h_2^\top\mu\mu^\top h_1\\=&\,\mathbb{E}_{x\sim p_\theta}(x-\mu)^\top h_1\cdot(x-\mu)^\top h_2\end{aligned}$$

Next, we note that

$$\begin{aligned}D\mu[h]&=D\frac{\int_{x\in K}x\exp(\theta^\top x)dx}{\int_{x\in K}\exp(\theta^\top x)dx}[h]\\&=\frac{\int_{x\in K}xx^\top h\exp(\theta^\top x)dx}{\int_{x\in K}\exp(\theta^\top x)dx}\\&\quad-\frac{\int_{x\in K}x\exp(\theta^\top x)dx}{\int_{x\in K}\exp(\theta^\top x)dx}\cdot\frac{\int_{x\in K}x^\top h\exp(\theta^\top x)dx}{\int_{x\in K}\exp(\theta^\top x)dx}\\&=\mathbb{E}_{x\sim p_\theta}xx^\top h-\mu\mu^\top h\\&=\mathbb{E}_{y\sim p_\theta}(y-\mu)(y-\mu)^\top h.\end{aligned}$$

So, we have

$$\begin{aligned}&D^3f(\theta)[h_1,h_2,h_3]\\=&\mathbb{E}_{x\sim p_\theta}(-\mathbb{E}_{y\sim p_\theta}(y-\mu)(y-\mu)^\top h_3)^\top h_1\cdot(x-\mu)^\top h_2\\&+\mathbb{E}_{x\sim p_\theta}(x-\mu)^\top h_1\cdot(-\mathbb{E}(y-\mu)(y-\mu)^\top h_3)^\top h_2\\&+\mathbb{E}_{x\sim p_\theta}(x-\mu)^\top h_1\cdot(x-\mu)^\top h_2\cdot(x-\mu)^\top h_3\\=&\mathbb{E}_{x\sim p_\theta}(x-\mu)^\top h_1\cdot(x-\mu)^\top h_2\cdot(x-\mu)^\top h_3.\end{aligned}$$

By [27, (2.15)], we have that

$$D^2f^*(x_\theta)[h_1,h_2]=h_1^\top\nabla^2f(\theta)^{-1}h_2$$

and

$$\begin{aligned}&D^3f^*(x_\theta)[h_1,h_2,h_3]\\&\quad=-D^3f(\theta)[\nabla^2f(\theta)^{-1}h_1,\nabla^2f(\theta)^{-1}h_2,\nabla^2f(\theta)^{-1}h_3]\end{aligned}$$

where $x_\theta=\nabla f(\theta)$. Hence, we have

$$\begin{aligned}&\nabla^2f^*(x_\theta)^{-\frac12}D^3f^*(x_\theta)[h]\nabla^2f^*(x_\theta)^{-\frac12}\\&=-\mathbb{E}_{x\sim p_\theta}\nabla^2f(\theta)^{-\frac12}(x-\mu)(x-\mu)^\top\nabla^2f(\theta)^{-\frac12}\\&\qquad\cdot(x-\mu)^\top\nabla^2f(\theta)^{-1}h\\&=-\mathbb{E}_{x\sim\tilde p_\theta}xx^\top\cdot x^\top\nabla^2f(\theta)^{-\frac12}h\end{aligned}$$

where $x\sim p_\theta$ and $\tilde p_\theta$ is the distribution given by $\nabla^2f(\theta)^{-\frac12}(x-\mu)$. Note that $\tilde p_\theta$ is isotropic and [5, Fact 6.1] shows that

$$\max_{\|v\|_2=1}\left\|\mathbb{E}_{x\sim\tilde p_\theta}xx^T(x^Tv)\right\|_F=O(\psi_n). \tag{25}$$

Hence, we have that

$$\begin{aligned}&\left\|\nabla^2f^*(x_\theta)^{-\frac12}D^3f^*(x_\theta)[h]\nabla^2f^*(x_\theta)^{-\frac12}\right\|_F\\&=O(\psi_n)\left\|\nabla^2f^*(x_\theta)^{-\frac12}h\right\|_2=O(\psi_n)\|h\|_{x_\theta}.\end{aligned}$$

This proves the lemma for the entropic barrier (recall that the entropic barrier is f^* instead of f).

For the universal barrier, first we recall that the polar of a convex set K is $K^\circ(x)=\left\{z:z^\top(y-x)\le1\quad\forall y\in K\right\}$ and the barrier function is

$$\Phi(x)=\log\mathrm{vol}(K^\circ(x)).$$

Its derivatives have the following identities [27, Page 52]. Here the random point y is drawn uniformly from the polar $K^\circ(x)$.

$$\begin{aligned}\nabla^2\Phi(x)=&(n+2)(n+1)\mathbb{E}yy^\top-(n+1)^2\mathbb{E}y\mathbb{E}y^\top,\\D\nabla^2\Phi(x)[h]=&-(n+1)(n+2)(n+3)\mathbb{E}yy^\top(y^\top h)\\&+(n+1)^2(n+2)\mathbb{E}yy^\top\cdot\mathbb{E}y^\top h\\&+2(n+1)^2(n+2)\mathbb{E}y(y^\top h)\cdot\mathbb{E}y^\top\\&-2(n+1)^3\mathbb{E}y\cdot\mathbb{E}y^\top\cdot\mathbb{E}y^\top h\end{aligned}$$

Let $\mu=\mathbb{E}y$, we can re-write the derivatives as follows:

$$\begin{aligned}\nabla^2\Phi(x)=&(n+2)(n+1)\mathbb{E}(y-\mu)(y-\mu)^\top+(n+1)\mu\mu^\top\\D\nabla^2\Phi(x)[h]=&-\Pi_{i=1}^3(n+i)\mathbb{E}(y-\mu)(y-\mu)^\top(y-\mu)^\top h\\&-2(n+2)(n+1)(\mathbb{E}(y-\mu)(y-\mu)^\top\mu^\top h\\&+\mathbb{E}\mu(y-\mu)^\top(y-\mu)^\top h+\mathbb{E}(y-\mu)\mu^\top(y-\mu)^\top h)\\&-2(n+1)\mu\mu^\top\mu^\top h.\end{aligned}$$

Without loss of generality, we assume $\nabla^2\Phi(x)=I$. Then, we have

$$(n+2)(n+1)\mathbb{E}(y-\mu)(y-\mu)^\top\preceq I\qquad\text{and}\qquad(n+1)\mu\mu^\top\preceq I.$$

For the first term, (25) shows that

$$\|(n+1)(n+2)(n+3)\mathbb{E}(y-\mu)(y-\mu)^\top(y-\mu)^\top h\|_F=O(\psi_n).$$

The Frobenius norm of the next three terms are bounded by

$$2\left|\mu^\top h\right|\left\|(n+2)(n+1)\mathbb{E}(y-\mu)(y-\mu)^\top\right\|_F\le2\sqrt{n}\,\|\mu\|\le2$$

and so is the last term:

$$2\left\|(n+1)\mu\mu^\top\right\|_F\left|\mu^\top h\right|\le2.$$

□

To conclude this section, we remark that the universal and entropic barriers do *not* satisfy our symmetry condition. Consider a rotational cone $C=\left\{x:\sum_{i=2}^nx_i^2\le x_1^2,0\le x_1\le1\right\}$ and any point $x=(x_1,0,\ldots,0)$. Then symmetric body around x, namely $K=C\cap(x-C)$ has the property that (a) the John ellipsoid satisfies $E\subset K\subset\sqrt{n}C$ (as it does for any symmetric convex body) and (b) the inertial ellipsoid has a sandwiching ratio of n, proving that $\bar\nu\ge n=\Omega(\nu^2)$. For the entropic barrier, we have a similar result because multiplying the indicator function of this symmetric convex body with an exponential function of the form e^{-c^Tx} still has the same property for the inertial ellipsoid. This example highlights the advantages of barriers with John-like ellipsoids (log barrier, LS barrier) vs Inertia-like ellipsoids (universal, entropic).

A PROOFS

A.1 Proof of Lemma 1.1

Proof. Let $h = y - x$, $x_t = x + th$ and $\phi(t) = h^\top \mathbf{H}(x_t)h$. Then,

$$\left|\phi'(t)\right| = \left|h^\top \frac{d}{dt}\mathbf{H}(x_t)h\right| \leq 2\|h\|_{x_t}^3 = 2\phi(t)^{3/2}.$$

Hence, we have $\left|\frac{d}{dt}\frac{1}{\sqrt{\phi(t)}}\right| \leq 1$. Therefore, $\frac{1}{\sqrt{\phi(t)}} \geq \frac{1}{\sqrt{\phi(0)}} - t$ and,

$$\phi(t) \leq \frac{\phi(0)}{(1 - t\sqrt{\phi(0)})^2}. \tag{26}$$

Now we fix any v and define $\psi(t) = v^\top \mathbf{H}(x_t)v$. Then,

$$\left|\psi'(t)\right| = \left|v^\top \frac{d}{dt}\mathbf{H}(x_t)v\right| \leq 2\|h\|_{x_t}\|v\|_{x_t}^2 = 2\phi(t)\psi(t).$$

Using (26) at the end, we have

$$\left|\frac{d}{dt}\ln\psi(t)\right| \leq \frac{2\sqrt{\phi(0)}}{(1 - t\sqrt{\phi(0)})}.$$

Integrating both sides from 0 to 1,

$$\left|\ln\frac{\psi(1)}{\psi(0)}\right| \leq \int_0^1 \frac{2\sqrt{\phi(0)}}{(1 - t\sqrt{\phi(0)})}dt = 2\ln(\frac{1}{1 - \sqrt{\phi(0)}}).$$

The result follows from this with $\psi(1) = v^\top \mathbf{H}(y)v$, $\psi(0) = v^\top \mathbf{H}(x)v$, and $\phi(0) = \|x - y\|_x^2$. □

A.2 Proof of Lemma 1.2

Proof. Let $x_t = (1 - t)x + ty$. Then, we have

$$\begin{aligned}&\|\mathbf{H}(x)^{-\frac{1}{2}}(\mathbf{H}(y) - \mathbf{H}(x))\mathbf{H}(x)^{-\frac{1}{2}}\|_F \\ &= \int_0^1 \|\mathbf{H}(x)^{-\frac{1}{2}}\frac{d}{dt}\mathbf{H}(x_t)\mathbf{H}(x)^{-\frac{1}{2}}\|_F dt.\end{aligned}$$

We note that $\mathbf{H}$ is self-concordant. Hence, Lemma 1.1 shows that

$$\begin{aligned}&\|\mathbf{H}(x)^{-\frac{1}{2}}\frac{d}{dt}\mathbf{H}(x_t)\mathbf{H}(x)^{-\frac{1}{2}}\|_F^2 \\ &= \mathrm{Tr}\mathbf{H}(x)^{-1}\left(\frac{d}{dt}\mathbf{H}(x_t)\right)\mathbf{H}(x)^{-1}\left(\frac{d}{dt}\mathbf{H}(x_t)\right) \\ &\leq \frac{1}{(1 - \|x - x_t\|_x)^4}\mathrm{Tr}\mathbf{H}(x_t)^{-1}\left(\frac{d}{dt}\mathbf{H}(x_t)\right)\mathbf{H}(x_t)^{-1}\left(\frac{d}{dt}\mathbf{H}(x_t)\right) \\ &\leq \frac{4}{(1 - \|x - x_t\|_x)^4}\|x - x_t\|_{x_t}^2 \\ &\leq \frac{4}{(1 - \|x - x_t\|_x)^6}\|x - x_t\|_x^2\end{aligned}$$

where we used the strong self-concordance in the second inequality and Lemma 1.1 again for the last inequality. Hence,

$$\begin{aligned}\|\mathbf{H}(x)^{-\frac{1}{2}}(\mathbf{H}(y) - \mathbf{H}(x))\mathbf{H}(x)^{-\frac{1}{2}}\|_F &\leq \int_0^1 \frac{2\|x - x_t\|_x}{(1 - \|x - x_t\|_x)^3}dt \\ &= \int_0^1 \frac{2t\|x - y\|_x}{(1 - t\|x - y\|_x)^3}dt \\ &= \frac{\|x - y\|_x}{(1 - \|x - y\|_x)^2}.\end{aligned}$$

□

REFERENCES

[1] Dimitris Bertsimas and Santosh Vempala. 2004. Solving convex programs by random walks. *Journal of the ACM (JACM)* 51, 4 (2004), 540–556.
[2] Sébastien Bubeck and Ronen Eldan. 2014. The entropic barrier: a simple and optimal universal self-concordant barrier. *arXiv preprint arXiv:1412.1587* (2014).
[3] Yuansi Chen, Raaz Dwivedi, Martin J Wainwright, and Bin Yu. 2018. Fast MCMC sampling algorithms on polytopes. *The Journal of Machine Learning Research* 19, 1 (2018), 2146–2231.
[4] Paul Christiano, Jonathan A Kelner, Aleksander Madry, Daniel A Spielman, and Shang-Hua Teng. 2011. Electrical flows, laplacian systems, and faster approximation of maximum flow in undirected graphs. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 273–282.
[5] Ronen Eldan. 2013. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis* 23, 2 (2013), 532–569.
[6] Adam Gustafson and Hariharan Narayanan. 2018. John's walk. *arXiv preprint arXiv:1803.02032* (2018).
[7] Roland Hildebrand. 2014. Canonical barriers on convex cones. *Mathematics of operations research* 39, 3 (2014), 841–850.
[8] Ravi Kannan, László Lovász, and Miklós Simonovits. 1997. Random walks and an o*(n5) volume algorithm for convex bodies. *Random Structures & Algorithms* 11, 1 (1997), 1–50.
[9] Ravindran Kannan and Hariharan Narayanan. 2012. Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research* 37, 1 (2012), 1–20.
[10] Narendra Karmarkar. 1984. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*. 302–311.
[11] Jonathan A Kelner, Yin Tat Lee, Lorenzo Orecchia, and Aaron Sidford. 2014. An almost-linear-time algorithm for approximate max flow in undirected graphs, and its multicommodity generalizations. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 217–226.
[12] Yin Tat Lee and Aaron Sidford. 2014. Path finding methods for linear programming: Solving linear programs in o (vrank) iterations and faster algorithms for maximum flow. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 424–433.
[13] Yin Tat Lee and Aaron Sidford. 2015. Efficient inverse maintenance and faster algorithms for linear programming. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE, 230–249.
[14] Yin Tat Lee and Aaron Sidford. 2019. Solving Linear Programs with Sqrt (rank) Linear System Solves. *arXiv preprint arXiv:1910.08033* (2019).
[15] Yin Tat Lee and Santosh Srinivas Vempala. 2017. Eldan's Stochastic Localization and the KLS Hyperplane Conjecture: An Improved Lower Bound for Expansion. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 998–1007.
[16] Yin Tat Lee and Santosh S Vempala. 2017. Geodesic walks in polytopes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on theory of Computing*. 927–940.
[17] Yin Tat Lee and Santosh S Vempala. 2018. Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 1115–1121.
[18] László Lovász. 1999. Hit-and-run mixes fast. *Mathematical Programming* 86, 3 (1999), 443–461.
[19] László Lovász and Miklós Simonovits. 1993. Random walks in a convex body and an improved volume algorithm. *Random structures & algorithms* 4, 4 (1993), 359–412.
[20] László Lovász and Santosh Vempala. 2006. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE, 57–68.
[21] László Lovász and Santosh Vempala. 2006. Hit-and-run from a corner. *SIAM J. Comput.* 35, 4 (2006), 985–1005.
[22] László Lovász and Santosh Vempala. 2006. Simulated annealing in convex bodies and an O*(n4) volume algorithm. *J. Comput. System Sci.* 72, 2 (2006), 392–417.

[23] László Lovász and Santosh Vempala. 2007. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms* 30, 3 (2007), 307–358.

[24] Aleksander Madry. 2010. Fast approximation algorithms for cut-based problems in undirected graphs. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, 245–254.

[25] Aleksander Madry. 2013. Navigating central path with electrical flows: From flows to matchings, and back. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 253–262.

[26] Oren Mangoubi and Nisheeth K Vishnoi. 2019. Faster Polytope Rounding, Sampling, and Volume Computation via a Sub-Linear Ball Walk. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 1338–1357.

[27] Yurii Nesterov and Arkadii Nemirovskii. 1994. *Interior-point polynomial algorithms in convex programming*. Vol. 13. Siam.

[28] Richard Peng. 2016. Approximate undirected maximum flows in o (m polylog (n)) time. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 1862–1867.

[29] James Renegar. 1988. A polynomial-time algorithm, based on Newton's method, for linear programming. *Mathematical programming* 40, 1-3 (1988), 59–93.

[30] Sushant Sachdeva and Nisheeth K Vishnoi. 2016. The mixing time of the Dikin walk in a polytope—a simple proof. *Operations Research Letters* 44, 5 (2016), 630–634.

[31] Jonah Sherman. 2013. Nearly maximum flows in nearly linear time. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 263–269.

[32] Jonah Sherman. 2017. Area-convexity, l regularization, and undirected multicommodity flow. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. 452–460.

[33] Pravin M Vaidya. 1989. Speeding-up linear programming using fast matrix multiplication. In *30th annual symposium on foundations of computer science*. IEEE, 332–337.

[34] Pravin M Vaidya. 1996. A new algorithm for minimizing convex functions over convex sets. *Mathematical programming* 73, 3 (1996), 291–341.

[35] Santosh Vempala. 2005. Geometric random walks: a survey. *Combinatorial and computational geometry* 52, 573-612 (2005), 2.