



# The mixing time of the Dikin walk in a polytope—A simple proof



Sushant Sachdeva<sup>a,\*</sup>, Nisheeth K. Vishnoi<sup>b</sup>

<sup>a</sup> Yale University, New Haven, CT, USA

<sup>b</sup> École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

## ARTICLE INFO

### Article history:

Received 7 August 2015

Received in revised form

5 July 2016

Accepted 5 July 2016

Available online 14 July 2016

### Keywords:

Polytopes

Sampling

Volume computation

Random walks

Interior point methods

## ABSTRACT

We study the mixing time of the Dikin walk in a polytope—a random walk based on the log-barrier from the interior point method literature. This walk, and a close variant, were studied by Narayanan (2016) and Kannan–Narayanan (2012). Bounds on its mixing time are important for algorithms for sampling and optimization over polytopes. Here, we provide a simple proof of their result that this random walk mixes in time  $O(mn)$  for an  $n$ -dimensional polytope described using  $m$  inequalities.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Sampling a point from the uniform distribution on a polytope  $K \subseteq \mathbb{R}^n$  is an extensively-studied problem and is a crucial ingredient in several computational tasks involving convex bodies. Towards this, typically, one sets up an ergodic and reversible random walk inside  $K$  whose stationary distribution is uniform over  $K$ . The mixing time of such a walk determines its efficacy, and, in turn, depends on the isoperimetric constant of  $K$  with respect to the transition function of the walk. Starting with the influential work of Dyer et al. [3], there has been a long line of work on faster and faster algorithms for generating an approximately uniform point from a convex body. Moreover, since convex bodies show up in a variety of areas, there is a wide body of work connecting random walks and isoperimetry in convex bodies to several areas in mathematics and optimization.

One such important connection to the interior point method literature was presented in the works of Kannan and Narayanan [6] and Narayanan [9] who proposed the *Dikin walk* in a polytope. Roughly, the uniform version of the Dikin walk, considered by [6], when at a point  $x \in K$ , computes the *Dikin ellipsoid* at  $x$ , and moves to a random point in it after a suitable Metropolis filter. The Metropolis step ensures that the walk is ergodic and reversible. The Gaussian version of the Dikin walk, considered by [9], picks

the new point from a Gaussian distribution centered at  $x$  with its covariance given by the Dikin ellipsoid at  $x$ , and applies a suitable Metropolis filter. The Dikin ellipsoid at a point  $x$  is the ellipsoid described by the Hessian of the log-barrier function at  $x$ . It was introduced by Dikin in the first interior point method for linear programming [2].

Several virtues of the Dikin ellipsoid (see [10,11,7]) were used by [6,9] to prove that the mixing time of the Dikin walk is  $O(mn)$  starting from a warm start, when  $K$  is described by  $m$  inequality constraints. Recall that a distribution over  $K$  is said to be a warm start if its density is bounded from above by a constant relative to the uniform distribution on  $K$ . Roughly, the proof (for either walk) consists of two parts: (1) an isoperimetric inequality, proved by Lovász [8], for convex bodies in terms of a distance introduced by Hilbert, and (2) a bound on the changes in the sampling distributions of the Dikin walk in terms of the Hilbert distance. The bound in (2) was the key technical contribution of [6,9] towards establishing the mixing time of the Dikin walk. We present a simple proof of this bound for the Gaussian Dikin walk implying that it mixes in time  $O(mn)$ . Our proof uses well-known facts about Gaussians, and concentration of Gaussian polynomials.

### 1.1. Dikin walk on polytopes

Suppose  $K \subseteq \mathbb{R}^n$  is a bounded polytope with a non-empty interior, described by  $m$  inequalities,  $a_i^\top x \geq b_i$ , for  $i \in [m]$ . We use the notation  $x \in K$  to denote that  $x$  is in the interior of  $K$ . The log-barrier function for  $K$  at  $x \in K$  is  $F(x) := -\sum_{i \in [m]} \log(a_i^\top x - b_i)$ . Let

\* Corresponding author.

E-mail addresses: [sachdeva@cs.yale.edu](mailto:sachdeva@cs.yale.edu) (S. Sachdeva), [nisheeth.vishnoi@epfl.ch](mailto:nisheeth.vishnoi@epfl.ch) (N.K. Vishnoi).

<http://dx.doi.org/10.1016/j.orl.2016.07.005>

0167-6377/© 2016 Elsevier B.V. All rights reserved.

$H(x)$  denote the Hessian of  $F$  at  $x$ , i.e.,  $H(x) := \sum_{i \in [m]} \frac{1}{(a_i^\top x - b_i)^2} a_i a_i^\top$ . For all  $x \in K$ ,  $H(x)$  is a positive definite matrix, and defines the local norm at  $x$ , denoted  $\|\cdot\|_x$ , as  $\|v\|_x^2 := v^\top H(x)v$ . The ellipsoid  $\{z : \|z - x\|_x \leq 1\}$  is known as the Dikin ellipsoid at  $x$ .

From a point  $x \in K$ , the next point  $z$  in the Dikin walk is sampled from the Dikin ellipsoid at  $x$ . The uniform Dikin walk, considered by [6], sampled the new point  $z$  from the uniform distribution in this ellipsoid. In the Gaussian Dikin walk, considered by [9],  $z$  is sampled from  $g_x$ , a multivariate Gaussian distribution centered at  $x$  with covariance matrix  $\frac{r^2}{n} H(x)^{-1}$ , where  $r$  is a constant. Thus, the density of the distribution is given by

$$g_x(z) = \sqrt{\det H(x)} \left(\frac{n}{2\pi r^2}\right)^{n/2} \cdot \exp\left(-\frac{n}{2r^2} \cdot \|z - x\|_x^2\right).$$

Equivalently, the next point  $z$  is given by

$$z = x + \frac{r}{\sqrt{n}} (H(x))^{-1/2} g,$$

where  $g$  is an  $n$ -dimensional vector with each coordinate of  $g$  sampled as an independent standard Gaussian  $\mathcal{N}(0, 1)$ .

In order to convert this into a random walk that stays inside  $K$ , with its stationary distribution as the uniform distribution on  $K$ , we apply the Metropolis filter to obtain the transition probability density  $p_x$  of the Gaussian Dikin walk:  $\forall z \neq x$ , if  $z \in K$ ,  $p_x(z) = \min\{g_x(z), g_z(x)\}$  (the walk stays at  $x$  with the remaining probability).

### 1.2. Hilbert metric, isoperimetry, and mixing time

We introduce the distance function which plays an important role in establishing the mixing time of the Dikin walk. Given two points  $x, y \in K$ , let  $p, q$  be the end points of the chord in  $K$  passing through  $x, y$ , such that the points lie in the order  $p, x, y, q$ . We define  $\sigma(x, y) := \frac{|xy| |pq|}{|px| |qy|}$ , where  $|xy|$  denotes the length of the line segment  $xy$ .  $\log(1 + \sigma(x, y))$  is a metric on  $K$ , known as Hilbert metric.

Lovász proved the following theorem for any random walk on  $K$ : Suppose for any two initial points  $x, y \in K$  that are close in  $\sigma$  distance, the statistical distance of the distributions after one step of the walk each from  $x$  and  $y$ , is bounded away from 1. Then, the lazy version of the random walk (where we stay at the current point with probability  $1/2$  at each step) mixes rapidly.

**Theorem 1** (Lovász [8]). Consider a reversible random walk in  $K$  with its stationary distribution being uniform on  $K$ . Suppose  $\exists \Delta > 0$  such that for all  $x, y \in K$  with  $\sigma(x, y) \leq \Delta$ , we have  $\|p_x - p_y\|_1 \leq 1 - \Omega(1)$ , where  $p_x$  denotes the distribution after one step of the random walk from  $x$ . Then, after  $O(\Delta^{-2})$  steps, the lazy version of the walk from a warm start is within  $1/4$  total variation distance from the uniform distribution on  $K$ .

Kannan and Narayanan proved that the transition function of the uniform Dikin walk,  $p_x$ , for  $x \in \text{int}(K)$ , satisfies the hypothesis of the theorem above with  $\Delta = \Omega\left(\frac{1}{\sqrt{mn}}\right)$ , thus implying that it mixes in  $O(mn)$  steps from a warm start. An analogous result for the Gaussian Dikin walk is implicit in the work of Narayanan. Our main contribution is an alternative and simple proof of their main technical contributions. In particular, we prove the following theorem.

**Theorem 2.** Let  $\varepsilon \in (0, 1/2]$ . For the Gaussian Dikin walk on  $K$  with  $r \leq \frac{\varepsilon}{400} (\log \frac{200}{\varepsilon})^{-3/2}$ , for any two points  $x, y \in K$  such that  $\|x - y\|_x \leq \frac{r}{\sqrt{n}}$ , we have  $\|p_x - p_y\|_1 \leq \varepsilon$ .

In order to use this theorem along with Theorem 1 to obtain the claimed mixing time bound, one needs a simple fact that, for any  $x, y$  in a polytope  $K$ , which is described using  $m$  inequalities,  $\sigma(x, y) \geq \frac{r}{\sqrt{m}} \|x - y\|_x$ . A proof of this fact is given in Appendix; see Lemma 9.

The following two lemmas are the main ingredients in the proof of Theorem 2: (1) If two points  $x, y$  are close in the local norm, i.e.,  $\|x - y\|_x \leq \frac{r}{\sqrt{n}}$ , then the two Gaussian distributions  $g_x$  and  $g_y$  are close in statistical distance. (2) If  $r$  is small enough (as a function of  $\varepsilon$ ), then for all  $x$ ,  $p_x$  and  $g_x$  are  $\varepsilon$ -close in statistical distance.

**Lemma 3.** Let  $r \leq 1$ , and  $c \geq 0$  be such that  $c \leq \min\{r, 1/3\}$ . Let  $x, y \in K$ . If  $\|x - y\|_x \leq \frac{c}{\sqrt{n}}$ , then  $\|g_x - g_y\|_1 \leq 3c$ .

This lemma relies on a well-known fact about the Kullback–Leibler divergence between two multivariate Gaussian distributions, and Pinsker’s inequality that bounds the statistical distance between two distributions in terms of their divergence.

**Lemma 4.** Given  $\varepsilon \in [0, 1/2]$ , for  $r \leq \frac{\varepsilon}{100} (\log \frac{50}{\varepsilon})^{-3/2}$ , we have  $\|p_x - g_x\|_1 \leq \varepsilon$ .

This lemma, which shows that the Metropolis filter does not change the distribution much, relies on a result on the concentration of Gaussian polynomials, proved using hypercontractivity. Given the above lemmas, Theorem 2 follows by applying triangle inequality.

## 2. Statistical distance between Gaussians and the local norm

In this section, we present a proof of Lemma 3 that bounds the statistical distance between  $g_x$  and  $g_y$  for two points  $x, y$  that are close in the local norm. We need the following well-known fact about the Kullback–Leibler divergence between two multivariate Gaussian distributions.

**Fact 5.** Let  $G_1 = \mathcal{N}(\mu_1, \Sigma_1)$  and  $G_2 = \mathcal{N}(\mu_2, \Sigma_2)$  be two  $n$ -dimensional Gaussian distributions. Then,

$$D_{KL}(G_2 \parallel G_1) = \frac{1}{2} \left( \text{Tr}(\Sigma_1^{-1} \Sigma_2) - n + \log \frac{\det \Sigma_1}{\det \Sigma_2} + (\mu_1 - \mu_2)^\top \Sigma_1^{-1} (\mu_1 - \mu_2) \right),$$

where  $D_{KL}$  denotes the Kullback–Leibler divergence

$$D_{KL}(P \parallel Q) = \int \log \frac{P(x)}{Q(x)} dP(x).$$

In order to use this theorem, we have to bound the eigenvalues of  $H(x)H(y)^{-1}$ . For  $x, y$  that are close in the local norm, this follows since  $H(x) \approx H(y)$ .

**Proof of Lemma 3.** From the assumption, we have,

$$\frac{c^2}{n} \geq \|x - y\|_x^2 = \sum_{i \in [m]} \frac{(a_i^\top (x - y))^2}{(a_i^\top x - b_i)^2} \geq \max_{i \in [m]} \frac{(a_i^\top (x - y))^2}{(a_i^\top x - b_i)^2}.$$

Thus, for all  $i \in [m]$ , we have

$$\left(1 - \frac{c}{\sqrt{n}}\right) (a_i^\top x - b_i) \leq (a_i^\top y - b_i) \leq \left(1 + \frac{c}{\sqrt{n}}\right) (a_i^\top x - b_i).$$

By the definition of  $H$ , we get,

$$\left(1 - \frac{c}{\sqrt{n}}\right)^2 H(y) \leq H(x) \leq \left(1 + \frac{c}{\sqrt{n}}\right)^2 H(y).$$

Download English Version:

<https://daneshyari.com/en/article/1142029>

Download Persian Version:

<https://daneshyari.com/article/1142029>

[Daneshyari.com](https://daneshyari.com)