

## Tilburg University

### Complexity analysis of a sampling-based interior point method for convex optimization

Badenbroek, Riley; de Klerk, Etienne

*Published in:*  
Mathematics of Operations Research

*Publication date:*  
2020

*Document Version*  
Peer reviewed version

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
Badenbroek, R., & de Klerk, E. (Accepted/In press). Complexity analysis of a sampling-based interior point method for convex optimization. *Mathematics of Operations Research*.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

# Complexity Analysis of a Sampling-Based Interior Point Method for Convex Optimization

Riley Badenbroek, Etienne de Klerk

Tilburg University, r.m.badenbroek@uvt.nl, e.deklerk@uvt.nl

We develop a short-step interior point method to optimize a linear function over a convex body assuming that one only knows a membership oracle for this body. The approach is based on Abernethy and Hazan's sketch of a universal interior point method using the so-called entropic barrier [ICML 2016 Proceedings]. It is well-known that the gradient and Hessian of the entropic barrier can be approximated by sampling from Boltzmann-Gibbs distributions, and the entropic barrier was shown to be self-concordant by Bubeck and Eldan [Math of OR, 2018]. The analysis of our algorithm uses properties of the entropic barrier, mixing times for hit-and-run random walks by Lovász and Vempala [Foundations of Computer Science, 2006], approximation quality guarantees for the mean and covariance of a log-concave distribution, and results from De Klerk, Glineur and Taylor on inexact Newton-type methods [arXiv 1709.0519, 2017].

*Key words:* interior point method, convex optimization, hit-and-run sampling, entropic barrier

---

**1. Introduction** The interior point revolution, to used a phrase coined by Wright [30], was the introduction of polynomial-time logarithmic barrier methods for linear programming, and their subsequent extension to convex programming. In their seminal work on the extension of interior point methods to convex programming, Nesterov and Nemirovskii [22] proved that every open convex set that does not contain an affine subspace is the domain of a self-concordant barrier, called the universal barrier. While this is an important theoretical result, its practical applicability is limited to cases where a barrier is known in closed form, and where its gradient and Hessian may be computed efficiently.

The most practical interior point software deals with self-dual cones, where self-concordant barriers are known; e.g. MOSEK [21], SDPT3 [29], and SeDuMi [28]. A promising recent development for more general cones is the primal-dual algorithm developed by Skajaa and Ye [26] and implemented in the software *alfonso* by Papp and Yıldız [23, 24], which only requires an efficiently computable self-concordant barrier of the primal cone. However, there are many convex bodies where one can solve the membership problem in polynomial time, but where no efficiently computable self-concordant barrier is known, such as the subtour elimination polytope.

Abernethy and Hazan [1] recently connected the field of simulated annealing (Kirkpatrick et al. [12]) to the study of interior point methods using the entropic barrier of Bubeck and Eldan [5]. One important property of this barrier is that its gradient and Hessian may be approximated through sampling, even though the barrier may not be known in closed form. This opens up the possibility of using interior point methods on sets for which we do not know efficiently computable self-concordant barriers. Interestingly, Güler [7] showed that the universal and entropic barriers coincide (up to an additive constant) if the domain of the barrier is a homogeneous cone, i.e. a cone with a transitive automorphism group.

The interior point method Abernethy and Hazan proposed was shown to converge in polynomial time by De Klerk, Glineur and Taylor [6], provided one can approximate the gradients and Hessians of the barrier sufficiently well. Our aim is to investigate when the approximations that may be obtained through sampling satisfy the requirements from [6]. In other words, we aim to show that we may approximate the gradient and Hessian sufficiently well in polynomial time, with high probability. The sampling algorithm that we will use is the Markov chain Monte Carlo method known as hit-and-run sampling, first suggested by Smith [27]. The mixing properties of this method, established by Lovász and Vempala [17], allow us to show the gradients and Hessians of the entropic barrier can be approximated to the desired accuracy in polynomial time, with high probability. Our hope is that this analysis will contribute to an extension of interior point methods to convex bodies where the membership problem is ‘easy’, but no efficiently computable barrier is known.

In their main text, Abernethy and Hazan connect the field of interior point methods to the simulated annealing algorithm by Kalai and Vempala [8] (for a generalization of Kalai and Vempala’s algorithm to non-linear objectives, see Lovász and Vempala [17]). To be precise, they show that the means of the family of distributions Kalai and Vempala sample from – which Abernethy and Hazan call the *heat path* – coincide with the central path associated with the entropic barrier. We also follow this central path, so our algorithm generates iterates that lie close to Kalai and Vempala’s heat path. Kalai and Vempala claim their algorithm terminates after  $O^*(n^{4.5})$  oracle calls, although an analysis using the tools in this paper shows it could be somewhat worse [2]. Regardless, Kalai and Vempala’s algorithm is perhaps the closest relative of our method.

Kalai and Vempala’s algorithm remained the state-of-the-art method for convex optimization in the membership oracle setting for a number of years in terms of computational complexity. Then, Lee, Sidford, and Vempala [14] proposed a method requiring  $O^*(n^2)$  oracle calls (and  $O^*(n^3)$  arithmetic operations). Their approach uses  $O^*(n)$  calls to the membership oracle to generate a separating hyperplane, and then invokes an algorithm such as Lee, Sidford, and Wong [15] for convex optimization in the separation oracle setting, requiring  $O^*(n)$  calls. As far as we are aware, there have been no further improvements to Lee, Sidford, and Vempala’s work since.

The number of oracle calls in our algorithm (see Theorem 8) is considerably worse than  $O^*(n^2)$  or even  $O^*(n^{4.5})$ . The main contribution of this work is therefore not to propose a state-of-the-art algorithm for convex optimization in the membership oracle setting. Instead, we aim to provide tools for the rigorous analysis of interior point methods in such settings.

**Outline of this paper** The outline of this paper is as follows. After some preliminary definitions in Section 2, we prove some useful properties of the entropic barrier in Section 3. Then, we give a hit-and-run mixing time theorem using self-concordance in Section 4. With this result established, we can show in Section 5 how hit-and-run might be applied to approximate means and covariances of Boltzmann distributions. Such approximations can then be used in Section 6 to analyze the aforementioned interior point method.

The main contributions of this work are in Sections 3-5, where the tools are developed to analyze interior point methods using the entropic barrier. These tools include

- a non-trivial lower bound on the spectrum of the covariance matrix of Boltzmann distributions, obtained through the theory of self-concordance (Section 3.1).
- a way to relate the distance between the parameters of two Boltzmann distributions to the distance between their means, and vice versa (Section 3.2).
- a careful analysis of a hit-and-run random walk applied to a Boltzmann distribution  $\theta_1$ , starting from a point that follows a Boltzmann distribution with parameter  $\theta_0$ , provided that  $\theta_1$  lies in a suitable ball around  $\theta_0$  (Section 4). Note that Kalai and Vempala [8] only show this for collinear  $\theta_0$  and  $\theta_1$ , and their argument does not extend to our wider setting.

- a rigorous analysis of the number of hit-and-run samples and walk lengths required to approximate the mean and covariance of a Boltzmann distribution to desired accuracy with high probability (Section 5). This analysis covers the pitfalls that the samples do not exactly follow the correct distribution, and are only near-independent.

Section 6 shows that these tools are necessary to formally analyze an interior point method using the entropic barrier, and that it is possible to establish polynomial time convergence of such a method.

Much of the analysis presented here is of a technical nature, and the reader may wish to skip the proofs on a first reading.

## 2. Preliminaries

We are interested in the problem

$$\min_{x \in K} \langle c, x \rangle, \quad (1)$$

where  $K \subseteq \mathbb{R}^n$  is a convex body, and  $\langle \cdot, \cdot \rangle$  is a reference inner product on  $\mathbb{R}^n$ . We may assume that  $\|c\| = \sqrt{\langle c, c \rangle} = 1$ , and that  $K$  contains a ball of radius  $r > 0$  and is contained in a ball of radius  $R \geq r$ . For any self-adjoint, positive definite linear operator  $A$ , we can define the inner product  $\langle \cdot, \cdot \rangle_A$  by  $\langle x, y \rangle_A := \langle x, Ay \rangle$ . The reference inner product induces the norm  $\|\cdot\|$ , and the inner product  $\langle \cdot, \cdot \rangle_A$  induces the norm  $\|\cdot\|_A$ .

**2.1. Self-Concordant Functions and the Entropic Barrier** The following discussion is condensed from Renegar [25]. Let  $\langle \cdot, \cdot \rangle$  be any inner product on  $\mathbb{R}^n$ . A function  $f$  from  $\text{dom } f \subseteq \mathbb{R}^n$  to  $\mathbb{R}$  is differentiable at  $\theta \in \text{dom } f$  if there exists a vector  $g(\theta) \in \mathbb{R}^n$  such that

$$\lim_{\|\Delta\theta\| \rightarrow 0} \frac{f(\theta + \Delta\theta) - f(\theta) - \langle g(\theta), \Delta\theta \rangle}{\|\Delta\theta\|} = 0.$$

The vector  $g(\theta)$  is called the gradient of  $f$  at  $\theta$  with respect to  $\langle \cdot, \cdot \rangle$ .

Furthermore, the function  $f$  is twice differentiable at  $\theta \in \text{dom } f$  if it is continuously differentiable at  $f$  and there exists a linear operator  $H(\theta) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that

$$\lim_{\|\Delta\theta\| \rightarrow 0} \frac{\|g(\theta + \Delta\theta) - g(\theta) - H(\theta)\Delta\theta\|}{\|\Delta\theta\|} = 0.$$

The linear operator  $H(\theta)$  is called the Hessian of  $f$  at  $\theta$  with respect to  $\langle \cdot, \cdot \rangle$ .

We denote the gradient and Hessian with respect to some other inner product  $\langle \cdot, \cdot \rangle_A$  by  $g_A$  and  $H_A$  respectively, and it can be shown that  $g_A(\theta) = A^{-1}g(\theta)$  and  $H_A(\theta) = A^{-1}H(\theta)$  (see e.g. Theorems 1.2.1 and 1.3.1 in Renegar [25]). For brevity, define  $\langle \cdot, \cdot \rangle_\theta := \langle \cdot, \cdot \rangle_{H(\theta)}$  and  $\|\cdot\|_\theta := \|\cdot\|_{H(\theta)}$  for any  $\theta \in \mathbb{R}^n$  and let  $g_\theta$  and  $H_\theta$  be the gradient and Hessian of  $f$  with respect to the local inner product  $\langle \cdot, \cdot \rangle_\theta$ .

The class of self-concordant functions plays an important role in the theory of interior point methods. We use the definition by Renegar [25].

**DEFINITION 1.** A function  $f$  is *self-concordant* if for all  $\theta_0 \in \text{dom } f$  and  $\theta_1 \in \mathbb{R}^n$  such that  $\|\theta_1 - \theta_0\|_{\theta_0} < 1$ , we have  $\theta_1 \in \text{dom } f$  and the following inequalities hold for all non-zero  $v \in \mathbb{R}^n$ :

$$1 - \|\theta_1 - \theta_0\|_{\theta_0} \leq \frac{\|v\|_{\theta_1}}{\|v\|_{\theta_0}} \leq \frac{1}{1 - \|\theta_1 - \theta_0\|_{\theta_0}}. \quad (2)$$

Some self-concordant functions have the additional property that the local norm of their gradients are bounded. Such functions are called barriers.

DEFINITION 2. A function  $f$  is a *barrier* if it is self-concordant and

$$\vartheta := \sup_{\theta \in \text{dom } f} \|g_\theta(\theta)\|_\theta^2 < \infty.$$

The value  $\vartheta$  is called the *complexity parameter* of  $f$ .

In the rest of this paper, the function  $f$  will always denote the log partition function associated with  $K$ , which is defined for any  $\theta \in \mathbb{R}^n$  by

$$f(\theta) := \ln \int_K e^{\langle \theta, x \rangle} dx.$$

Moreover, denote the expectation of a Boltzmann distribution over  $K$  with parameter  $\theta \in \mathbb{R}^n$  by

$$\mathbb{E}_\theta[X] = \frac{\int_K x e^{\langle \theta, x \rangle} dx}{\int_K e^{\langle \theta, x \rangle} dx}.$$

It is not hard to see that for all  $v \in \mathbb{R}^n$ ,

$$g(\theta) = \mathbb{E}_\theta[X], \quad H(\theta)v = \mathbb{E}_\theta[\langle X - \mathbb{E}_\theta[X], v \rangle (X - \mathbb{E}_\theta[X])]. \quad (3)$$

If  $\langle \cdot, \cdot \rangle$  were the Euclidean inner product,  $H(\theta)$  can be represented by the covariance matrix  $\mathbb{E}_\theta[(X - \mathbb{E}_\theta[X])(X - \mathbb{E}_\theta[X])^\top]$  of a Boltzmann distribution with parameter  $\theta$  over  $K$ . To emphasize this fact, we will write  $\Sigma(\theta)$  instead of  $H(\theta)$  where appropriate.

It was shown by Bubeck and Eldan [5] that  $f$  is self-concordant. In this case, Definition 1 guarantees that for all  $\theta_0, \theta_1 \in \mathbb{R}^n$  such that  $\|\theta_1 - \theta_0\|_{\theta_0} < 1$ , the following inequalities hold for all non-zero  $v \in \mathbb{R}^n$ :

$$1 - \|\theta_1 - \theta_0\|_{\theta_0} \leq \frac{\|v\|_{\theta_1}}{\|v\|_{\theta_0}} \leq \frac{1}{1 - \|\theta_1 - \theta_0\|_{\theta_0}}. \quad (4)$$

Moreover, let  $f^*$  be the Fenchel conjugate of  $f$ , defined in the usual manner:

$$f^*(x) = \sup_{\theta \in \mathbb{R}^n} \{\langle \theta, x \rangle - f(\theta)\},$$

where  $x \in \text{int } K$ . (The reason that  $\text{dom } f^* = \text{int } K$  will be discussed shortly.) The function  $f^*$  is called the *entropic barrier* for  $K$ . Again borrowing notation from Renegar [25], let  $g^*$  and  $H^*$  be the gradient and Hessian of  $f^*$  (all linear operators in this paper are self-adjoint, so we will not use an asterisk to refer to an adjoint). Define  $\langle \cdot, \cdot \rangle_x^* := \langle \cdot, \cdot \rangle_{H^*(x)}$  for all  $x \in \text{int } K$ , and let  $\|\cdot\|_x^*$  be the local norm induced by this inner product.

Since it was shown by Bubeck and Eldan [5] that  $f$  is self-concordant,  $f^*$  is self-concordant as well. Here, Definition 1 shows that for all  $x_0 \in \text{int } K$  and  $x_1 \in \mathbb{R}^n$  such that  $\|x_1 - x_0\|_{x_0}^* < 1$ , it holds that  $x_1 \in \text{int } K$  and for all non-zero  $v \in \mathbb{R}^n$ ,

$$1 - \|x_1 - x_0\|_{x_0}^* \leq \frac{\|v\|_{x_1}^*}{\|v\|_{x_0}^*} \leq \frac{1}{1 - \|x_1 - x_0\|_{x_0}^*}. \quad (5)$$

The following is known about the domain of the conjugate of self-concordant functions.

LEMMA 1 (**Proposition 3.3.3 in [25]**). *For any self-concordant  $f$ , we have  $\text{dom } f^* = \{g(\theta) : \theta \in \mathbb{R}^n\}$ .*

For the log partition function specifically, we have an explicit description of  $\{g(\theta) : \theta \in \mathbb{R}^n\}$ .

LEMMA 2 (**Lemma 3.1 in [13]**). *If  $f$  is the log partition function associated with a convex body  $K$ , then  $\{g(\theta) : \theta \in \mathbb{R}^n\} = \text{int } K$ .*

Hence,  $\text{dom } f^* = \text{int } K$ , as one would expect from a barrier for  $K$ . The following is known about the derivatives of  $f^*$ .

LEMMA 3 (Theorem 3.3.4 in [25]). *For any self-concordant  $f$ , we have for all  $\theta \in \mathbb{R}^n$ ,*

$$g^*(g(\theta)) = \theta, \quad H^*(g(\theta)) = H(\theta)^{-1}.$$

To clarify that  $g$  assigns to a  $\theta \in \mathbb{R}^n$  a point  $x \in K$ , we will sometimes write  $x(\theta)$  for  $g(\theta)$ . Lemmas 1, 2 and 3 imply that  $g^*$  assigns to every  $x \in \text{int } K$  a vector  $\theta \in \mathbb{R}^n$  such that  $g(\theta) = x(\theta) = x$ . For this reason, we will often write  $\theta(x)$  for  $g^*(x)$  to keep the notation intuitive. The notation from this section is summarized in Table 1.

TABLE 1. Overview of the properties of  $f$  and  $f^*$ .

Log partition function $f$	Entropic barrier $f^*$
$f(\theta) = \ln \int_K e^{\langle \theta, x \rangle} dx$	$f^*(x) = \sup_{\theta \in \mathbb{R}^n} \{\langle \theta, x \rangle - f(\theta)\}$
$g(\theta) = \mathbb{E}_\theta[X] = x(\theta)$	$g^*(x) = \theta(x)$ such that $g(\theta(x)) = x$
$H(\theta) = \Sigma(\theta)$	$H^*(x) = H(g^*(x))^{-1} = \Sigma(\theta(x))^{-1}$
Domain: $\theta \in \mathbb{R}^n$	Domain: $x \in \text{int } K$

Finally, it was shown by Bubeck and Eldan [5] that  $\vartheta = n + o(n)$ , where  $\vartheta$  is the complexity parameter of the entropic barrier defined in Definition 2.

**2.2. Interior Point Method Sketched by Abernethy-Hazan** The short-step interior point method proposed by De Klerk, Glineur, and Taylor [6], which was based on a sketch by Abernethy and Hazan [1], is sketched in Algorithm 1.

---

**Algorithm 1** Sketch of the interior point method by De Klerk, Glineur, and Taylor [6] (based on a sketch by Abernethy and Hazan [1])

---

**Input:** Tolerances  $\epsilon, \tilde{\epsilon}, \bar{\epsilon} > 0$ ; entropic barrier parameter  $\vartheta \leq n + o(n)$ ; objective  $c \in \mathbb{R}^n$ ; central path proximity parameter  $\delta > 0$ ; an  $x_0 \in K$  and  $\eta_0 > 0$  such that  $\|x_0 - x(-\eta_0 c)\|_{x(-\eta_0 c)}^* \leq \frac{1}{2}\delta$ ; growth rate  $\beta > 0$ ; sample size  $N \in \mathbb{N}$ .

**Output:**  $x_k$  such that  $\langle c, x_k \rangle - \min_{x \in K} \langle c, x \rangle \leq \bar{\epsilon}$  at terminal iteration  $k$ .

---

```

1:  $\hat{\epsilon} \leftarrow \tilde{\epsilon} \sqrt{\frac{1+\epsilon}{1-\epsilon}} + \sqrt{\frac{2\epsilon}{1-\epsilon}}$ 
2:  $\gamma \leftarrow \frac{2(1-\delta)^4 - \tilde{\epsilon}(1+(1-\delta)^4)}{(1-\tilde{\epsilon})(1-\delta)^2(1+(1-\delta)^4)}$ 
3:  $k \leftarrow 0$ 
4: while  $\vartheta(1 + \frac{1}{2}\delta)/\eta_k > \bar{\epsilon}$  do
5:   Generate samples  $Y^{(1)}, \dots, Y^{(N)}$  from the Boltzmann distribution with parameter  $-\eta_k c$ 
6:   Find approximation  $\widehat{\Sigma}(-\eta_k c)$  of  $\Sigma(-\eta_k c)$  using samples  $Y^{(1)}, \dots, Y^{(N)}$ 
7:   Find approximation  $\hat{\theta}(x_k)$  of  $\theta(x_k)$  by minimizing  $\Psi_k(\theta) = f(\theta) - \langle \theta, x_k \rangle$  over  $\mathbb{R}^n$ 
8:    $x_{k+1} \leftarrow x_k - \gamma \widehat{\Sigma}(-\eta_k c)[\eta_{k+1} c + \hat{\theta}(x_k)]$   $\triangleright x_{k+1} \approx x(-\eta_{k+1} c)$ 
9:    $\eta_{k+1} \leftarrow (1 + \frac{\beta}{\sqrt{\vartheta}})\eta_k$ 
10:   $k \leftarrow k + 1$ 
11: end while
12: return  $x_k$ 

```

---

In every iteration  $k$ , we would like to know (an approximation of)  $H^*(x_k)^{-1}$ . Because the function  $f^*$  is self-concordant,  $H^*(x_k)$  is not too different from  $H^*(x(-\eta_k c))$  if  $x_k$  is close to  $x(-\eta_k c)$  in some well-defined sense. Thus, the algorithm can also be proven to work if we find (an approximation of)  $H^*(x(-\eta_k c))^{-1} = \Sigma(-\eta_k c)$ . In practice, we can find an approximation of  $\Sigma(-\eta_k c)$  by generating sufficiently many samples from the Boltzmann distribution with parameter  $-\eta_k c$  and computing the empirical covariance matrix.

To find  $g^*(x_k) = \theta(x_k)$ , we will use the approach proposed by Abernethy and Hazan [1]. Note that the function  $\Psi_k(\theta) = f(\theta) - \langle \theta, x_k \rangle$  has Fréchet derivatives

$$D\Psi_k(\theta) = g(\theta) - x_k = \mathbb{E}_\theta[X] - x_k, \quad D^2\Psi_k(\theta) = \Sigma(\theta). \quad (6)$$

In other words,  $\Psi_k$  is a convex function which is minimized at the  $\theta \in \mathbb{R}^n$  such that  $g(\theta) = x_k$ , which is equal to  $\theta(x_k)$ . Therefore, to approximate  $\theta(x_k)$ , it suffices to minimize  $\Psi_k$  over  $\theta \in \mathbb{R}^n$ . As is clear from (6), the gradient and Hessian of  $\Psi_k$  can be approximated at some particular  $\theta \in \mathbb{R}^n$  by generating sufficiently many samples of the Boltzmann distribution with parameter  $\theta$ , and consequently computing the empirical mean c.q. covariance matrix.

One might wonder what quality guarantees should be satisfied by the approximations  $\widehat{\Sigma}(-\eta_k c)$  and  $\widehat{\theta}(x_k)$  such that the algorithm still provably works. This question was answered by the following theorem from De Klerk, Glineur and Taylor [6].

**THEOREM 1 (Theorem 7.6 in [6]).** *Consider Algorithm 1 with the input settings  $\beta = \frac{1}{32}$  and  $\delta = \frac{1}{4}$ , and let  $\epsilon, \bar{\epsilon} > 0$  such that  $\widehat{\epsilon} = \bar{\epsilon} \sqrt{\frac{1+\epsilon}{1-\epsilon}} + \sqrt{\frac{2\epsilon}{1-\epsilon}} \leq \frac{1}{6}$ . Suppose that in every iteration of Algorithm 1, the approximation  $\widehat{\Sigma}(-\eta_k c)$  satisfies*

$$\begin{aligned} (1-\epsilon)y^\top \widehat{\Sigma}(-\eta_k c)y &\leq y^\top \Sigma(-\eta_k c)y \leq (1+\epsilon)y^\top \widehat{\Sigma}(-\eta_k c)y & \forall y \in \mathbb{R}^n, \\ (1-\epsilon)y^\top \widehat{\Sigma}(-\eta_k c)^{-1}y &\leq y^\top \Sigma(-\eta_k c)^{-1}y \leq (1+\epsilon)y^\top \widehat{\Sigma}(-\eta_k c)^{-1}y & \forall y \in \mathbb{R}^n, \end{aligned}$$

and that the approximation  $\widehat{\theta}(x_k)$  satisfies

$$\left\| \widehat{\theta}(x_k) - \theta(x_k) \right\|_{\Sigma(-\eta_{k+1}c)} \leq \bar{\epsilon} \left\| -\eta_{k+1}c - \theta(x_k) \right\|_{\Sigma(-\eta_{k+1}c)}.$$

If the algorithm is initialized with an  $x_0 \in K$  and  $\eta_0 > 0$  such that  $\|x_0 - x(-\eta_0 c)\|_{x(-\eta_0 c)}^* \leq \frac{1}{2}\delta$ , then it terminates after

$$k = \left\lceil 40\sqrt{\vartheta} \ln \left( \frac{\vartheta(1 + \frac{1}{2}\delta)}{\eta_0 \bar{\epsilon}} \right) \right\rceil$$

iterations. The result is an  $x_k$  such that

$$\langle c, x_k \rangle - \min_{x \in K} \langle c, x \rangle \leq \bar{\epsilon}.$$

The main purpose of this paper is to give a detailed description of how one can approximate  $\Sigma(-\eta_k c)$  and  $\theta(x_k)$  in practice such that quality requirements similar to the ones in the theorem above are satisfied. In order to do this, we need some results from probability theory.

**2.3. Log-Concavity and Divergence of Probability Distributions** The probability density function of a Boltzmann distribution belongs to the well-studied class of log-concave functions. We start by recalling the definition of log-concavity.

**DEFINITION 3.** A function  $h: \mathbb{R}^n \rightarrow \mathbb{R}_+$  is *log-concave* if for any two  $x, y \in \mathbb{R}^n$  and  $\lambda \in (0, 1)$ ,

$$h(\lambda x + (1-\lambda)y) \geq h(x)^\lambda h(y)^{1-\lambda}.$$

In other words, if  $h$  is strictly positive, then it is log-concave if and only if  $x \mapsto \log(h(x))$  is concave.

We will need the following concentration result for log-concave distributions. Note that it is stronger than Chebyshev's inequality.

LEMMA 4 (**Lemma 3.3 from [19]**). *Let  $X$  be a random variable with a log-concave distribution, and let  $\|\cdot\|$  be the Euclidean norm. Denote  $\mathbb{E}[\|X - \mathbb{E}[X]\|^2] =: \sigma^2$ . Then for all  $t > 1$ ,*

$$\mathbb{P}\{\|X - \mathbb{E}[X]\| > t\sigma\} \leq e^{1-t}.$$

Next, we define level sets for general probability density functions.

DEFINITION 4. Let  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  be a probability density function supported on  $K \subseteq \mathbb{R}^n$ . Then, the level set  $L_p$  of (the distribution with density)  $h$  is  $\{x \in K : h(x) \geq \alpha_p\}$ , where  $\alpha_p$  is chosen such that  $\int_{L_p} h(x) dx = p$ .

Note that the level sets of log-concave distributions are convex (see e.g. Section 3.5 in Boyd and Vandenberghe [4]).

We will generate samples from the family of Boltzmann distributions with a random walk method known as hit-and-run sampling, to be defined later. Hit-and-run sampling is only guaranteed to work if the distribution of the starting point and the distribution one would like to sample from are "close". Even then, the result of the hit-and-run walk does not have the correct distribution, but a distribution which is again "close" to the desired distribution. To make these statements exact, we will need two measures of divergence between probability distributions. Before we can define them, we recall the definition of absolute continuity.

DEFINITION 5. Let  $(K, \mathcal{E})$  be a measurable space, and let  $\nu$  and  $\mu$  be measures on this space. Then,  $\nu$  is *absolutely continuous* with respect to  $\mu$  if  $\mu(A) = 0$  implies  $\nu(A) = 0$  for all  $A \in \mathcal{E}$ . We write this property as  $\nu \ll \mu$ .

The first measure of divergence between probability distributions is the  $L_2$ -norm.

DEFINITION 6. Let  $(K, \mathcal{E})$  be a measurable space. Let  $\nu$  and  $\mu$  be two probability distributions over this space, such that  $\nu \ll \mu$ . Then, the  $L_2$ -norm of  $\nu$  with respect to  $\mu$  is

$$\|\nu/\mu\| := \int_K \frac{d\nu}{d\mu} d\nu = \int_K \left(\frac{d\nu}{d\mu}\right)^2 d\mu,$$

where  $\frac{d\nu}{d\mu}$  is the Radon-Nikodym derivative of  $\nu$  with respect to  $\mu$ .

If  $K \subseteq \mathbb{R}^n$ , and  $\nu$  and  $\mu$  have probability densities  $h_\nu$  and  $h_\mu$ , respectively, with respect to the Lebesgue measure, it can be shown that

$$\|\nu/\mu\| = \int_K \frac{h_\nu(x)}{h_\mu(x)} h_\nu(x) dx.$$

The second way in which we will measure distance between probability distributions is by total variation distance.

DEFINITION 7. Let  $(K, \mathcal{E})$  be a measurable space. For two probability distributions  $\mu$  and  $\nu$  over this space, their *total variation distance* is

$$\|\mu - \nu\| := \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

A useful property of the total variation distance is that it allows coupling of random variables, as the following lemma asserts.

LEMMA 5 (**e.g. Proposition 4.7 in [16]**). *Let  $X$  be a random variable on  $K \subseteq \mathbb{R}^n$  with distribution  $\mu$ , and let  $\nu$  be a different probability distribution on  $K$ . If  $\|\mu - \nu\| = \alpha$ , we can construct another random variable  $Y$  on  $K$  distributed according to  $\nu$  such that  $\mathbb{P}\{X = Y\} = 1 - \alpha$ .*



**2.4. Near-Independence** The end point of a random walk depends on the starting point of the walk, but as the walk length increases, this dependence starts to vanish. We will use the notion of near-independence to quantify this.

DEFINITION 8. Two random variables  $X$  and  $Y$  taking values in measurable space  $(K, \mathcal{E})$  are *near-independent* or *q-independent* if for all  $A, B \in \mathcal{E}$ ,

$$|\mathbb{P}\{X \in A \wedge Y \in B\} - \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\}| \leq q.$$

Before we can analyze the near-independence of starting and end points of a random walk, we need the formal machinery of Markov kernels. Intuitively, a Markov kernel assigns to any point in  $K$  a probability distribution over  $K$ . Its analogue for discrete space Markov chains is a transition probability matrix.

DEFINITION 9. Let  $(K, \mathcal{E})$  be a measurable space, and let  $\mathcal{B}[0, 1]$  be the Borel- $\sigma$ -algebra over  $[0, 1]$ . A *Markov kernel* is a map  $Q: K \times \mathcal{E} \rightarrow [0, 1]$  with the properties

- (i) For every  $x \in K$ , the map  $B \mapsto Q(x, B)$  for  $B \in \mathcal{E}$  is a probability measure on  $(K, \mathcal{E})$ ;
- (ii) For every  $B \in \mathcal{E}$ , the map  $x \mapsto Q(x, B)$  for  $x \in K$  is  $(\mathcal{E}, \mathcal{B}[0, 1])$ -measurable.

Since  $Q(x, \cdot)$  is a measure for any fixed  $x \in K$ , we can integrate a function  $\phi$  over  $K$  with respect to this measure. This integral will be denoted by  $\int_K \phi(y)Q(x, dy)$ . We emphasize again that for any  $x \in K$ , this expression is just a Lebesgue integral.

Suppose the Markov kernel  $Q$  corresponds to one step of a random walk, i.e. after one step from  $x \in K$ , the probability of ending up in  $B \in \mathcal{E}$  is  $Q(x, B)$ . The probability that after  $m \geq 1$  steps a random walk starting at  $x \in K$  ends up in  $B \in \mathcal{E}$  is then given by

$$Q^m(x, B) := \int_K Q(y, B)Q^{m-1}(x, dy),$$

where  $Q^1 := Q$ . Another interpretation of  $Q^m(x, B)$  is the probability of a random walk ending in  $B$ , conditional on the random starting point  $X$  of the walk taking value  $x$ . If moreover the starting point of the random walk is not fixed, but follows a probability distribution  $\nu$ , then the end point of the random walk after  $m$  steps follows distribution  $\nu Q^m$ , defined by

$$(\nu Q^m)(B) = \int_K Q^m(x, B) d\nu(x),$$

for all  $B \in \mathcal{E}$ .

The following lemma connects total variation distance to near-independence. It will ensure that if the distribution of the end point  $Y$  of a random walk approaches some fixed desired distribution  $\mu$ , then the start point  $X$  of this random walk and  $Y$  are near-independent. A similar relation was established by Lovász and Vempala [19], but we will use a version that does not assume  $Y$  follows the desired distribution  $\mu$ .

LEMMA 6 (cf. Lemma 4.3(a) in [19]). *Fix a probability distribution  $\mu$  over a set  $K \subseteq \mathbb{R}^n$ . Let  $Q$  be a Markov kernel on  $K$ , and let  $\ell: \mathbb{R}_+ \rightarrow \mathbb{N}$ . Suppose that for any  $\bar{M} \geq 0$ ,  $\bar{q} > 0$  and any distribution  $\bar{\nu}$  satisfying  $\bar{\nu} \ll \mu$  and  $\|\bar{\nu}/\mu\| \leq \bar{M}$ , it holds that  $\|\bar{\nu}Q^{\ell(\bar{M}/\bar{q}^2)} - \mu\| \leq \bar{q}$ . Let  $M \geq 0, q > 0$ , and let  $\nu$  be a distribution such that  $\nu \ll \mu$  and  $\|\nu/\mu\| \leq M$ . If  $X$  is a random variable with distribution  $\nu$ , and  $Y$  is a random variable with distribution conditional on  $X = x$  given by  $Q^{\ell(M/q^2)}(x, \cdot)$  for any  $x \in K$ , then  $X$  and  $Y$  are  $3q$ -independent.*

*Proof.* Let  $A$  and  $B$  be measurable subsets of  $K$ . As noted in Lovász and Vempala [19, relation (4)], one has the elementary relation

$$\left| \mathbb{P}\{Y \in B \wedge X \in A\} - \mathbb{P}\{Y \in B\}\mathbb{P}\{X \in A\} \right| = \left| \mathbb{P}\{Y \in B \wedge X \notin A\} - \mathbb{P}\{Y \in B\}\mathbb{P}\{X \notin A\} \right|.$$

We may therefore assume  $\mathbb{P}\{X \in A\} = \nu(A) \geq \frac{1}{2}$ .

The marginal distribution of  $Y$  satisfies

$$\mathbb{P}\{Y \in B\} = \int_K Q^{\ell(M/q^2)}(x, B) d\nu(x) = \nu Q^{\ell(M/q^2)}(B). \quad (7)$$

Consider the restriction  $\nu_A$  of  $\nu$  to  $A$ , scaled to be a probability measure. Then,

$$\begin{aligned} \mathbb{P}\{Y \in B|X \in A\} &= \frac{\mathbb{P}\{Y \in B \wedge X \in A\}}{\mathbb{P}\{X \in A\}} \\ &= \frac{\int_A Q^{\ell(M/q^2)}(x, B) d\nu(x)}{\nu(A)} \\ &= \int_K Q^{\ell(M/q^2)}(x, B) d\nu_A(x) \\ &= \nu_A Q^{\ell(M/q^2)}(B). \end{aligned} \quad (8)$$

Since  $\nu(A) \geq \frac{1}{2}$ , we have  $\frac{d\nu_A}{d\nu}(x) \leq 2$  for  $\nu$ -almost all  $x \in K$ . Then,

$$\|\nu_A/\mu\| = \int_K \left(\frac{d\nu_A}{d\mu}\right)^2 d\mu = \int_K \left(\frac{d\nu_A}{d\nu}\right)^2 \left(\frac{d\nu}{d\mu}\right)^2 d\mu \leq 4\|\nu/\mu\| \leq 4M.$$

Therefore,  $\|\nu_A Q^{\ell(M/q^2)} - \mu\| = \|\nu_A Q^{\ell(4M/(2q)^2)} - \mu\| \leq 2q$  by assumption. Since  $\|\nu/\mu\| \leq M$ , we also have  $\|\nu Q^{\ell(M/q^2)} - \mu\| \leq q$  by assumption. Hence, by combining (7) and (8) with the triangle inequality and Definition 7, it follows that

$$\begin{aligned} \left| \mathbb{P}\{Y \in B|X \in A\} - \mathbb{P}\{Y \in B\} \right| &= \left| \nu_A Q^{\ell(M/q^2)}(B) - \nu Q^{\ell(M/q^2)}(B) \right| \\ &\leq \left| \nu_A Q^{\ell(M/q^2)}(B) - \mu(B) \right| + \left| \mu(B) - \nu Q^{\ell(M/q^2)}(B) \right| \\ &\leq \|\nu_A Q^{\ell(M/q^2)} - \mu\| + \|\nu Q^{\ell(M/q^2)} - \mu\| \\ &\leq 2q + q = 3q. \end{aligned}$$

Multiplying both sides of the outermost inequality by  $\mathbb{P}\{X \in A\}$  shows

$$\left| \mathbb{P}\{Y \in B \wedge X \in A\} - \mathbb{P}\{Y \in B\}\mathbb{P}\{X \in A\} \right| \leq 3q\mathbb{P}\{X \in A\} \leq 3q,$$

which completes the proof.  $\square$

Having shown that the start and end point of a random walk are near-independent, we continue by proving the near-independence of the result of two independent random walks with the same starting point.

**LEMMA 7.** *Let  $Y_1$  and  $Y_2$  be random variables that are both  $q$ -independent of a random variable  $X$ . Assume that  $Y_1$  and  $Y_2$  are conditionally independent given  $X$  and that for all measurable events  $\{Y_1 \in A\}$  and  $\{Y_2 \in B\}$ , the following sets are measurable:*

$$\begin{aligned} &\{x \in K : \mathbb{P}\{Y_1 \in A|X = x\} \geq \mathbb{P}\{Y_1 \in A\}\} \\ &\{x \in K : \mathbb{P}\{Y_2 \in B|X = x\} \geq \mathbb{P}\{Y_2 \in B\}\}. \end{aligned}$$

*Then,  $Y_1$  and  $Y_2$  are  $2q$ -independent.*

*Proof.* [Proof of Lemma 7] Denote the probability distribution of  $X$  by  $\mu$ . We want to bound the following term.

$$\begin{aligned} & \left| \mathbb{P}\{Y_1 \in A \wedge Y_2 \in B\} - \mathbb{P}\{Y_1 \in A\}\mathbb{P}\{Y_2 \in B\} \right| \\ &= \left| \int_K (\mathbb{P}\{Y_1 \in A \wedge Y_2 \in B|X=x\} - \mathbb{P}\{Y_1 \in A\}\mathbb{P}\{Y_2 \in B\}) d\mu(x) \right| \\ &= \left| \int_K (\mathbb{P}\{Y_1 \in A|X=x\}\mathbb{P}\{Y_2 \in B|X=x\} - \mathbb{P}\{Y_1 \in A\}\mathbb{P}\{Y_2 \in B\}) d\mu(x) \right|, \end{aligned} \quad (9)$$

where the last equality holds by the conditional independence of  $Y_1$  and  $Y_2$ . We will use the identity  $ab - cd = (a - c)(b - d) + (a - c)d + (b - d)c$ , where  $a, b, c, d \in \mathbb{R}$ . Select

$$a = \mathbb{P}\{Y_1 \in A|X=x\}, \quad b = \mathbb{P}\{Y_2 \in B|X=x\}, \quad c = \mathbb{P}\{Y_1 \in A\}, \quad d = \mathbb{P}\{Y_2 \in B\}.$$

This allows us to expand (9) in an obvious manner. The triangle inequality then gives

$$\begin{aligned} & \left| \mathbb{P}\{Y_1 \in A \wedge Y_2 \in B\} - \mathbb{P}\{Y_1 \in A\}\mathbb{P}\{Y_2 \in B\} \right| \\ & \leq \left| \int_K (\mathbb{P}\{Y_1 \in A|X=x\} - \mathbb{P}\{Y_1 \in A\})(\mathbb{P}\{Y_2 \in B|X=x\} - \mathbb{P}\{Y_2 \in B\}) d\mu(x) \right| \\ & + \left| \int_K (\mathbb{P}\{Y_1 \in A|X=x\} - \mathbb{P}\{Y_1 \in A\})\mathbb{P}\{Y_2 \in B\} d\mu(x) \right| \\ & + \left| \int_K (\mathbb{P}\{Y_2 \in B|X=x\} - \mathbb{P}\{Y_2 \in B\})\mathbb{P}\{Y_1 \in A\} d\mu(x) \right|. \end{aligned} \quad (10)$$

We will upper bound each of these terms.

For the first term, we can use Hölder's inequality as follows.

$$\begin{aligned} & \left| \int_K (\mathbb{P}\{Y_1 \in A|X=x\} - \mathbb{P}\{Y_1 \in A\})(\mathbb{P}\{Y_2 \in B|X=x\} - \mathbb{P}\{Y_2 \in B\}) d\mu(x) \right| \\ & \leq \sqrt{\int_K (\mathbb{P}\{Y_1 \in A|X=x\} - \mathbb{P}\{Y_1 \in A\})^2 d\mu(x)} \sqrt{\int_K (\mathbb{P}\{Y_2 \in B|X=x\} - \mathbb{P}\{Y_2 \in B\})^2 d\mu(x)}. \end{aligned} \quad (11)$$

Define

$$C := \{x \in K : \mathbb{P}\{Y_1 \in A|X=x\} \geq \mathbb{P}\{Y_1 \in A\}\}.$$

Since both  $\mathbb{P}\{Y_1 \in A|X=x\}$  and  $\mathbb{P}\{Y_1 \in A\}$  lie in  $[0, 1]$ , the square of their difference can be upper bounded by their absolute difference. Therefore,

$$\begin{aligned} & \int_K (\mathbb{P}\{Y_1 \in A|X=x\} - \mathbb{P}\{Y_1 \in A\})^2 d\mu(x) \\ & \leq \int_K \left| \mathbb{P}\{Y_1 \in A|X=x\} - \mathbb{P}\{Y_1 \in A\} \right| d\mu(x) \\ & = \int_C (\mathbb{P}\{Y_1 \in A|X=x\} - \mathbb{P}\{Y_1 \in A\}) d\mu(x) + \int_{K \setminus C} (\mathbb{P}\{Y_1 \in A\} - \mathbb{P}\{Y_1 \in A|X=x\}) d\mu(x) \\ & = \mathbb{P}\{Y_1 \in A \wedge X \in C\} - \mathbb{P}\{Y_1 \in A\}\mathbb{P}\{X \in C\} + \mathbb{P}\{Y_1 \in A\}\mathbb{P}\{X \notin C\} - \mathbb{P}\{Y_1 \in A \wedge X \notin C\} \\ & \leq 2q, \end{aligned}$$

since  $X$  and  $Y_1$  are near-independent. Because the same holds for  $\mathbb{P}\{Y_2 \in B|X=x\}$  and  $\mathbb{P}\{Y_2 \in B\}$ , (11) is upper bounded by  $2q$ .

For the second term in (10), observe that

$$\begin{aligned} & \left| \int_K (\mathbb{P}\{Y_1 \in A|X = x\} - \mathbb{P}\{Y_1 \in A\})\mathbb{P}\{Y_2 \in B\} d\mu(x) \right| \\ &= \mathbb{P}\{Y_2 \in B\} \left| \int_K (\mathbb{P}\{Y_1 \in A|X = x\} - \mathbb{P}\{Y_1 \in A\}) d\mu(x) \right| \\ &= \mathbb{P}\{Y_2 \in B\} \left| \mathbb{P}\{Y_1 \in A\} - \mathbb{P}\{Y_1 \in A\} \right| = 0. \end{aligned}$$

The same clearly holds for the third term in (10). Hence,

$$\left| \mathbb{P}\{Y_1 \in A \wedge Y_2 \in B\} - \mathbb{P}\{Y_1 \in A\}\mathbb{P}\{Y_2 \in B\} \right| \leq 2q.$$

□

For near-independent vector-valued random variables, the products of some of the entries in the respective vectors are also near-independent, as the following lemma shows.

**LEMMA 8.** *Let  $X = (X_1, \dots, X_n)$  and  $Y = (Y_1, \dots, Y_n)$  be  $q$ -independent random variables with values in  $\mathbb{R}^n$ , and let  $\mathcal{S} \subseteq \{1, \dots, n\}$ . Suppose the function  $x = (x_1, \dots, x_n) \mapsto \prod_{i \in \mathcal{S}} x_i$  is measurable. Then, the random variables  $\prod_{i \in \mathcal{S}} X_i$  and  $\prod_{i \in \mathcal{S}} Y_i$  are  $q$ -independent.*

*Proof.* The result follows from Lemma 3.5 in Lovász and Vempala [19] applied to  $x \mapsto \prod_{i \in \mathcal{S}} x_i$ . □

The measurability conditions in Lemmas 7 and 8 are satisfied for sufficiently detailed  $\sigma$ -algebras. We assume these conditions to hold in the remainder of this paper.

To close this section, we cite the following result on the expectation of the product of near-independent real-valued random variables.

**LEMMA 9 (Lemma 2.7 from [10]).** *Let  $X$  and  $Y$  be  $q$ -independent random variables such that  $|X| \leq a$  and  $|Y| \leq b$ . Then*

$$|\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]| \leq 4qab.$$

**3. Entropic Barrier Properties** The self-concordance of  $f$  and  $f^*$  can be used to show two results that we will need for the analysis of an interior point method that uses the entropic barrier. First, it will turn out that we will need a lower bound on  $\|\theta\|_\theta$  for all  $\theta \in \mathbb{R}^n$ , which requires an investigation of the spectrum (with respect to the Euclidean inner product) of the covariance matrix of a Boltzmann distribution. Second, we will show that if  $x, y \in \text{int } K$  are close, then  $\theta(x)$  and  $\theta(y)$  are also close, in a well-defined sense.

**3.1. Spectra of Boltzmann Covariance Matrices** To analyze the spectra of the Boltzmann covariance matrices, we will need information about the spectrum of the covariance matrix of the uniform distribution. We will denote the smallest and largest eigenvalue of a self-adjoint linear operator  $A$  with respect to the reference inner product by  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$ . Recall that for positive semidefinite linear operators  $A$ , we have  $\lambda_{\min}(A) = \min_{v: \|v\|=1} \langle v, Av \rangle$  and  $\lambda_{\max}(A) = \max_{v: \|v\|=1} \langle v, Av \rangle = \|A\|$ .

One should note that an upper bound of the spectrum of  $\Sigma(\theta)$  is trivial to derive for any  $\theta \in \mathbb{R}^n$ . If  $K$  is contained in a ball with radius  $R$ , i.e. the diameter of  $K$  is at most  $2R$ ,

$$\lambda_{\max}(\Sigma(\theta)) = \max_{v: \|v\|=1} \mathbb{E}_\theta[\langle X - \mathbb{E}_\theta[X], v \rangle^2] \leq (2R)^2, \tag{12}$$

where the equality uses (3). Thus, we will focus on bounding the smallest eigenvalue of  $\Sigma(\theta)$  from below. Our starting point is the following result from Kannan, Lovász and Simonovits [9].

LEMMA 10 (**Theorem 4.1 in [9]**). *Let  $K \subseteq \mathbb{R}^n$  be a convex body, and recall that  $\Sigma(0)$  denotes the covariance matrix of the uniform distribution over  $K$ . If  $\Sigma(0) = I$ , then  $K$  is contained in a Euclidean ball with radius  $n + 1$ .*

We can use this result to bound the spectrum of  $\Sigma(0)$  from below.

LEMMA 11. *Let  $K \subseteq \mathbb{R}^n$  be a convex body that contains a Euclidean ball of radius  $r$ . Then,  $\lambda_{\min}(\Sigma(0)) \geq \frac{1}{4} \left(\frac{r}{n+1}\right)^2$ .*

*Proof.* The convex body  $K' = \Sigma(0)^{-1/2}K$  has the property that the uniform distribution over  $K'$  has identity covariance. By Lemma 10,  $K'$  is contained in a ball of radius  $n + 1$ .

Let  $x \in K$  be the center of the ball with radius  $r$  contained in  $K$ , and let  $v$  be a unit vector such that  $\Sigma(0)^{-1/2}v = \lambda_{\max}(\Sigma(0)^{-1/2})v$ . Since  $v$  is a unit vector, the point  $x + rv$  lies in  $K$ . Because  $\Sigma(0)^{-1/2}x$  and  $\Sigma(0)^{-1/2}(x + rv)$  lie in  $K'$ , we find  $\|\Sigma(0)^{-1/2}((x + rv) - x)\| \leq 2(n + 1)$ , where  $2(n + 1)$  is the diameter of a ball containing  $K'$ . In conclusion,

$$2(n + 1) \geq \|r\Sigma(0)^{-1/2}v\| = r\lambda_{\max}(\Sigma(0)^{-1/2}) = \frac{r}{\sqrt{\lambda_{\min}(\Sigma(0))}},$$

which proves  $\lambda_{\min}(\Sigma(0)) \geq \frac{1}{4} \left(\frac{r}{n+1}\right)^2$ .  $\square$

With the spectrum of the uniform covariance matrix bounded, we can continue to analyze  $\Sigma(\theta)$ , where  $\theta \in \mathbb{R}^n$ . Using Lemma 3, we get for every  $\theta \in \mathbb{R}^n$ ,

$$\|\theta\|_{\vartheta}^2 = \langle g^*(g(\theta)), H(\theta)g^*(g(\theta)) \rangle = \langle g_{g(\theta)}^*(g(\theta)), H^*(g(\theta))g_{g(\theta)}^*(g(\theta)) \rangle = (\|g_{g(\theta)}^*(g(\theta))\|_{g(\theta)}^*)^2 \leq \vartheta, \quad (13)$$

where the inequality follows from Lemma 2 and the definition of the complexity parameter  $\vartheta$  from Renegar [25] (see Definition 2). With this inequality, we can now prove a bound on the smallest eigenvalue of  $\Sigma(\theta)$  for all  $\theta \in \mathbb{R}^n$ .

THEOREM 2. *Let  $K \subseteq \mathbb{R}^n$  be a convex body that contains a Euclidean ball of radius  $r$  and is contained in a Euclidean ball of radius  $R$ . Define  $f$  as the log partition function  $f(\theta) = \ln \int_K e^{\langle \theta, x \rangle} dx$ , where  $\langle \cdot, \cdot \rangle$  is the Euclidean inner product, and denote its Hessian by  $\Sigma(\theta)$ . Let  $f^*$  be the entropic barrier for  $K$  with complexity parameter  $\vartheta$ . Let  $\lambda_{\min}(\Sigma(\theta))$  be the smallest eigenvalue of  $\Sigma(\theta)$ . Then, for any  $\theta \in \mathbb{R}^n$  with  $\|\theta\| \leq \frac{1}{4R}$ ,*

$$\lambda_{\min}(\Sigma(\theta)) \geq \frac{1}{16} \left(\frac{r}{n+1}\right)^2,$$

and for all  $\theta \in \mathbb{R}^n$  with  $\|\theta\| > \frac{1}{4R}$ ,

$$\lambda_{\min}(\Sigma(\theta)) \geq \frac{1}{64} \left(\frac{1}{4R\|\theta\|}\right)^{4\sqrt{\vartheta}+2} \left(\frac{r}{n+1}\right)^2.$$

*Proof.* We want to find a lower bound on  $\|v\|_{\theta}$ , where  $\|v\| = 1$ . The idea is to use the self-concordance properties of  $f$  to move from  $\Sigma(\theta)$  to the covariance matrix of the uniform distribution, and then apply Lemma 11.

If  $\|\theta\| \leq \frac{1}{4R}$ , then (12) shows

$$\|\theta - 0\|_0 \leq \sqrt{\lambda_{\max}(\Sigma(0))}\|\theta\| \leq 2R\|\theta\| \leq \frac{1}{2} < 1,$$

and thus we may apply the first inequality in (4) and Lemma 11 to show that

$$\|v\|_{\theta} \geq \|v\|_0(1 - \|\theta - 0\|_0) \geq \frac{1}{2} \sqrt{\lambda_{\min}(\Sigma(0))}\|v\| \geq \frac{1}{4} \frac{r}{n+1} \|v\|. \quad (14)$$

It then follows from (14) that

$$\lambda_{\min}(\Sigma(\theta)) = \min_{v: \|v\|=1} \|v\|_{\theta}^2 \geq \frac{1}{16} \left( \frac{r}{n+1} \right)^2.$$

Next, suppose that  $\|\theta\| > \frac{1}{4R}$ . Let  $\theta_0 = \theta$  and recursively define  $\theta_k = (1 - \frac{1}{2\sqrt{\vartheta+1}})\theta_{k-1}$ . Observe that by (13), for all  $k$ ,

$$\|\theta_{k-1} - \theta_k\|_{\theta_k} = \frac{\|\theta_{k-1}\|_{\theta_k}}{2\sqrt{\vartheta+1}} = \frac{\|\theta_k\|_{\theta_k}}{2\sqrt{\vartheta}} \leq \frac{\sqrt{\vartheta}}{2\sqrt{\vartheta}} = \frac{1}{2} < 1.$$

Since  $\theta_k$  and  $\theta_{k-1}$  are close in the sense above, we can apply self-concordance. By the first inequality of (4), for all  $k$ ,  $\|v\|_{\theta_{k-1}} \geq (1 - \|\theta_{k-1} - \theta_k\|_{\theta_k}) \|v\|_{\theta_k} \geq \frac{1}{2} \|v\|_{\theta_k}$ . Thus, after  $m$  steps, we have

$$\|v\|_{\theta} = \|v\|_{\theta_0} \geq 2^{-m} \|v\|_{\theta_m}. \quad (15)$$

Setting

$$m = \left\lceil \frac{\log_2(\frac{1}{4R\|\theta\|})}{\log_2(1 - \frac{1}{2\sqrt{\vartheta+1}})} \right\rceil,$$

we obtain

$$\|\theta_m\| = \left(1 - \frac{1}{2\sqrt{\vartheta+1}}\right)^m \|\theta_0\| \leq \frac{1}{4R\|\theta\|} \|\theta_0\| = \frac{1}{4R}.$$

We may now apply (14) to see that  $\|v\|_{\theta_m} \geq \frac{1}{4} \frac{r}{n+1} \|v\|$ . Combined with (15), it follows that

$$\|v\|_{\theta} \geq 2^{-m} \|v\|_{\theta_m} \geq \frac{2^{-m}}{4} \frac{r}{n+1} \|v\| = 2^{-m-2} \frac{r}{n+1} \|v\|. \quad (16)$$

Because  $m$  is an integer, we arrive at the following lower bound for  $2^{-m-2}$ :

$$2^{-m-2} \geq \frac{1}{8} (4R\|\theta\|)^{1/\log_2(1 - \frac{1}{2\sqrt{\vartheta+1}})}.$$

Since  $4R\|\theta\| > 1$  by assumption, and  $1/\log_2(1-t) \geq -1/t$  for all  $t \in (0,1)$ , this bound can be developed to

$$2^{-m-2} \geq \frac{1}{8} (4R\|\theta\|)^{1/\log_2(1 - \frac{1}{2\sqrt{\vartheta+1}})} \geq \frac{1}{8} (4R\|\theta\|)^{-2\sqrt{\vartheta}-1},$$

and we can conclude from (16) that

$$\lambda_{\min}(\Sigma(\theta)) = \min_{v: \|v\|=1} \|v\|_{\theta}^2 \geq \frac{1}{64} \left( \frac{1}{4R\|\theta\|} \right)^{4\sqrt{\vartheta}+2} \left( \frac{r}{n+1} \right)^2.$$

□

Note that this lower bound is exponential in  $\vartheta = n + o(n)$ . For our analysis in Section 6, we will need a stronger lower bound on  $\|\theta\|_{\theta} = \sqrt{\langle \theta, \Sigma(\theta)\theta \rangle}$  than the one obtained from Theorem 2 by setting  $v = \theta/\|\theta\|$ . The following lemma gives such a lower bound that is not exponential in  $n$ .

**LEMMA 12.** *Let  $K \subseteq \mathbb{R}^n$  be a convex body that contains a Euclidean ball of radius  $r$ , and let  $\langle \cdot, \cdot \rangle$  be the Euclidean inner product. Then, it holds for all  $\theta \in \mathbb{R}^n$  that*

$$\|\theta\|_{\theta} \geq \frac{r\|\theta\|}{2(n+1) + r\|\theta\|}. \quad (17)$$

*Proof.* Note that the right hand side of (17) is always strictly smaller than one. The claim therefore holds automatically for all  $\theta$  with  $\|\theta\|_\theta \geq 1$ , and we can assume in the remainder that  $\|\theta\|_\theta < 1$ . By Lemma 11, we have that  $\lambda_{\min}(\Sigma(0)) \geq \frac{1}{4}(\frac{r}{n+1})^2$ . The second inequality in (4) then gives us

$$\frac{1}{2} \frac{r}{n+1} \|\theta\| \leq \|\theta\|_0 \leq \frac{\|\theta\|_\theta}{1 - \|\theta\|_\theta},$$

or equivalently,

$$\|\theta\|_\theta \geq \frac{\frac{1}{2} \frac{r}{n+1} \|\theta\|}{1 + \frac{1}{2} \frac{r}{n+1} \|\theta\|} = \frac{r \|\theta\|}{2(n+1) + r \|\theta\|}.$$

□

**3.2. Parameter Proximity** Next, we show that if  $x, y \in \text{int } K$  are “close”, then so are  $\theta(x)$  and  $\theta(y)$ , and vice versa.

LEMMA 13. *Let  $K$  be a convex body, and let  $x, y, z \in \text{int } K$ . If  $\|\theta(x) - \theta(y)\|_{\theta(z)} + \|\theta(y) - \theta(z)\|_{\theta(z)} < 1$ , then*

$$\|x - y\|_z^* \leq \frac{1}{1 - \|\theta(y) - \theta(z)\|_{\theta(z)}} \left( \frac{\|\theta(x) - \theta(y)\|_{\theta(z)}}{1 - \|\theta(x) - \theta(y)\|_{\theta(z)} - \|\theta(y) - \theta(z)\|_{\theta(z)}} \right).$$

Similarly, if  $\|x - y\|_z^* + \|y - z\|_z^* < 1$ , then

$$\|\theta(x) - \theta(y)\|_{\theta(z)} \leq \frac{1}{1 - \|y - z\|_z^*} \left( \frac{\|x - y\|_z^*}{1 - \|x - y\|_z^* - \|y - z\|_z^*} \right).$$

*Proof.* We have

$$\begin{aligned} \|x - y\|_z^* &= \|g(\theta(x)) - g(\theta(y))\|_{H^*(z)} \\ &= \|g(\theta(x)) - g(\theta(y))\|_{H(\theta(z))^{-1}} \\ &= \|g_{\theta(z)}(\theta(x)) - g_{\theta(z)}(\theta(y))\|_{H(\theta(z))} \\ &= \left\| \int_0^1 H_{\theta(z)}(\theta(y) + t[\theta(x) - \theta(y)])[\theta(x) - \theta(y)] dt \right\|_{\theta(z)}, \end{aligned} \quad (18)$$

by the fundamental theorem of calculus (see e.g. Theorem 1.5.6 in Renegar [25]). We have the following upper bound on (18):

$$\begin{aligned} &\left\| \int_0^1 H_{\theta(z)}(\theta(y) + t[\theta(x) - \theta(y)])[\theta(x) - \theta(y)] dt \right\|_{\theta(z)} \\ &\leq \|\theta(x) - \theta(y)\|_{\theta(z)} \max_{u \in \mathbb{R}^n} \frac{\langle u, \int_0^1 H_{\theta(z)}(\theta(y) + t[\theta(x) - \theta(y)])u dt \rangle_{\theta(z)}}{\|u\|_{\theta(z)}^2} \\ &\leq \|\theta(x) - \theta(y)\|_{\theta(z)} \int_0^1 \max_{u \in \mathbb{R}^n} \frac{\langle u, H_{\theta(z)}(\theta(y) + t[\theta(x) - \theta(y)])u \rangle_{\theta(z)}}{\|u\|_{\theta(z)}^2} dt \\ &= \|\theta(x) - \theta(y)\|_{\theta(z)} \int_0^1 \max_{u \in \mathbb{R}^n} \frac{\|u\|_{\theta(y) + t[\theta(x) - \theta(y)]}^2}{\|u\|_{\theta(z)}^2} dt. \end{aligned} \quad (19)$$

If  $\|\theta(x) - \theta(z)\|_{\theta(z)} < 1$  and  $\|\theta(y) - \theta(z)\|_{\theta(z)} < 1$ , then by the triangle inequality,

$$\|\theta(z) - \theta(y) - t[\theta(x) - \theta(y)]\|_{\theta(z)} = \|t[\theta(z) - \theta(x)] + (1-t)[\theta(z) - \theta(y)]\|_{\theta(z)} < 1,$$

for all  $t \in [0, 1]$ . We can therefore apply the second inequality of (4) as follows:

$$\begin{aligned} \int_0^1 \max_{u \in \mathbb{R}^n} \frac{\|u\|_{\theta(y)+t[\theta(x)-\theta(y)]}^2}{\|u\|_{\theta(z)}^2} dt &\leq \int_0^1 \left( \frac{1}{1 - \|\theta(z) - \theta(y) - t[\theta(x) - \theta(y)]\|_{\theta(z)}} \right)^2 dt \\ &\leq \int_0^1 \left( \frac{1}{1 - \|\theta(z) - \theta(y)\|_{\theta(z)} - t\|\theta(x) - \theta(y)\|_{\theta(z)}} \right)^2 dt \\ &= \frac{1}{1 - \|\theta(z) - \theta(y)\|_{\theta(z)}} \left( \frac{1}{1 - \|\theta(z) - \theta(y)\|_{\theta(z)} - \|\theta(x) - \theta(y)\|_{\theta(z)}} \right). \end{aligned} \quad (20)$$

The upper bound on  $\|x - z\|_z^*$  thus follows from combining (18), (19) and (20). The upper bound on  $\|\theta(x) - \theta(z)\|_{\theta(z)}$  can be derived in the same manner as the above by interchanging  $x$  and  $\theta(x)$ ,  $y$  and  $\theta(y)$ ,  $z$  and  $\theta(z)$ , and  $f$  and  $f^*$ .  $\square$

We will not always need this general lemma with three points  $x$ ,  $y$  and  $z$ . For easy reference, we will state the following corollary that only considers two points  $x$  and  $z$ .

**COROLLARY 1.** *Let  $K$  be a convex body, and let  $x, z \in \text{int } K$ . If  $\|\theta(x) - \theta(z)\|_{\theta(z)} < 1$ , then*

$$\|x - z\|_z^* \leq \frac{\|\theta(x) - \theta(z)\|_{\theta(z)}}{1 - \|\theta(x) - \theta(z)\|_{\theta(z)}} \quad \text{and} \quad \frac{\|x - z\|_z^*}{1 + \|x - z\|_z^*} \leq \|\theta(x) - \theta(z)\|_{\theta(z)}. \quad (21)$$

Similarly, if  $\|x - z\|_z^* < 1$ , then

$$\|\theta(x) - \theta(z)\|_{\theta(z)} \leq \frac{\|x - z\|_z^*}{1 - \|x - z\|_z^*} \quad \text{and} \quad \frac{\|\theta(x) - \theta(z)\|_{\theta(z)}}{1 + \|\theta(x) - \theta(z)\|_{\theta(z)}} \leq \|x - z\|_z^*. \quad (22)$$

*Proof.* Substitution of  $y = z$  in Lemma 13 gives the upper bounds on  $\|x - z\|_z^*$  and  $\|\theta(x) - \theta(z)\|_{\theta(z)}$ . These can be rewritten as lower bounds on  $\|\theta(x) - \theta(z)\|_{\theta(z)}$  and  $\|x - z\|_z^*$ , respectively.  $\square$

**4. Hit-and-Run Sampling** The procedure we will use to generate samples is called *hit-and-run* sampling. This routine was introduced for the uniform distribution by Smith [27] and later generalized to absolutely continuous distributions (see for example Bélisle et al. [3]). We will use the version in Algorithm 2, based on Lovász and Vempala [17].

---

**Algorithm 2** The hit-and-run sampling procedure

---

**Require:** probability density  $h : \mathbb{R}^n \rightarrow \mathbb{R}_+$  with respect to the Lebesgue measure of the distribution to sample from (i.e. the target distribution); covariance matrix  $\Sigma \in \mathbb{R}^{n \times n}$ ; starting point  $x \in K$ ; number of hit-and-run steps  $\ell \in \mathbb{N}$ .

- 1:  $X_0 \leftarrow x$
  - 2: Sample directions  $D_1, \dots, D_\ell$  i.i.d. from a  $\mathcal{N}(0, \Sigma)$ -distribution
  - 3: Sample  $P_1, \dots, P_\ell$  i.i.d. from a uniform distribution over  $[0, 1]$ , independent from  $D_1, \dots, D_\ell$
  - 4: **for**  $i \in \{1, \dots, \ell\}$  **do**
  - 5: Determine the two end points  $Y_i$  and  $Z_i$  of the line segment  $K \cap \{X_{i-1} + tD_i : t \in \mathbb{R}\}$
  - 6: Determine  $s \in [0, 1]$  such that  $\int_0^s h(Y_i + t(Z_i - Y_i)) dt = P_i \int_0^1 h(Y_i + t(Z_i - Y_i)) dt$
  - 7:  $X_i \leftarrow Y_i + s(Z_i - Y_i)$
  - 8: **end for**
  - 9: **return**  $X_\ell$
-



This procedure samples a random direction  $D_i$  from a normal distribution, and samples the next iterate  $X_i$  from the desired distribution restricted to the line through  $X_{i-1}$  in the direction  $D_i$ , intersected with  $K$ . Effectively, this reduces a high-dimensional sampling problem to a sequence of one-dimensional sampling problems.

The following theorem from Lovász and Vempala [17] is the starting point of our analysis.

**THEOREM 3 (Theorem 1.1 in [17]).** *Let  $\mu$  be a log-concave probability distribution supported on a convex body  $K \subseteq \mathbb{R}^n$ , and let  $q > 0$ . Consider a hit-and-run random walk as in Algorithm 2 with respect to the target distribution  $\mu$  from a random starting point with distribution  $\nu$  supported on  $K$ . Assume that the following holds:*

- (i) *the level set of  $\mu$  with probability  $\frac{1}{8}$  (see Definition 4) contains a ball of radius  $s$  with respect to  $\|\cdot\|$ ;*
- (ii)  *$\frac{d\nu}{d\mu}(x) \leq M'$  for all  $x \in K \setminus A$  for some set  $A \subseteq K$  with  $\nu(A) \leq q$ ;*
- (iii)  *$\mathbb{E}_\mu[\|X - \mathbb{E}_\mu[X]\|^2] \leq S^2$ .*

*Let  $\nu^{(\ell)}$  be the distribution of the current hit-and-run point after  $\ell$  steps of hit-and-run sampling applied to  $\mu$ , where the directions are chosen from a  $\mathcal{N}(0, I)$ -distribution. Then, after*

$$\ell = \left\lceil 10^{30} \frac{n^2 S^2}{s^2} \ln^2 \left( \frac{2M'nS}{sq} \right) \ln^3 \left( \frac{2M'}{q} \right) \right\rceil,$$

*hit-and-run steps, we have  $\|\nu^{(\ell)} - \mu\| \leq q$ .*

Suppose that rather than (ii), we know  $\|\nu/\mu\| \leq M$ , i.e.  $\int_K \frac{d\nu}{d\mu}(x) d\nu(x) \leq M$ . If  $A = \{x \in K : \frac{d\nu}{d\mu}(x) > M/q\}$ , then

$$M \geq \int_A \frac{d\nu}{d\mu}(x) d\nu(x) \geq \frac{M}{q} \nu(A),$$

and thus we have  $\nu(A) \leq q$ . (This construction was also applied by Lovász and Vempala [18, page 10].) We can therefore set  $M' = M/q$  in the theorem. If one additionally considers a transformation  $x \mapsto \Sigma^{-1/2}x$  for some invertible matrix  $\Sigma$  applied to  $K$ , before Theorem 3 is applied, we arrive at the following corollary.

**COROLLARY 2.** *Let  $\mu$  be a log-concave probability distribution supported on a convex body  $K \subseteq \mathbb{R}^n$ , and let  $q > 0$ . Consider a hit-and-run random walk as in Algorithm 2 with respect to the target distribution  $\mu$  from a random starting point with distribution  $\nu$  supported on  $K$ . Assume that the following holds for some invertible matrix  $\Sigma$ :*

- (i) *the level set of  $\mu$  with probability  $\frac{1}{8}$  (see Definition 4) contains a ball of radius  $s$  with respect to  $\|\cdot\|_{\Sigma^{-1}}$ ;*
- (ii)  *$\|\nu/\mu\| \leq M$ ;*
- (iii)  *$\mathbb{E}_\mu[\|X - \mathbb{E}_\mu[X]\|_{\Sigma^{-1}}^2] \leq S^2$ .*

*Let  $\nu^{(\ell)}$  be the distribution of the hit-and-run point after  $\ell$  steps of hit-and-run sampling applied to  $\mu$ , where the directions are drawn from a  $\mathcal{N}(0, \Sigma)$ -distribution. Then, after*

$$\ell = \left\lceil 10^{30} \frac{n^2 S^2}{s^2} \ln^2 \left( \frac{2MnS}{sq^2} \right) \ln^3 \left( \frac{2M}{q^2} \right) \right\rceil, \quad (23)$$

*hit-and-run steps, we have  $\|\nu^{(\ell)} - \mu\| \leq q$ .*

This corollary can be used to show that two hit-and-run samples with the same starting point are near-independent.

**LEMMA 14.** *Let  $\mu$  be a log-concave probability distribution supported on a convex body  $K \subseteq \mathbb{R}^n$ , and let  $q > 0$ . Suppose the conditions of Corollary 2 are satisfied for some  $\Sigma$ ,  $s$ ,  $M$ , and  $S$ . Let  $X$  be a random variable with distribution  $\nu$  supported on  $K$ . Consider two hit-and-run random walks*

as in Algorithm 2, applied to  $\mu$ , both starting from the same realization of  $X$ . Let the number of steps  $\ell$  of both walks be given by (23), and call the resulting end points  $Y_1$  and  $Y_2$ . Then,  $Y_1$  and  $Y_2$  are  $6q$ -independent.

*Proof.* Let  $Q$  be the Markov kernel of a hit-and-run step, where directions are chosen from  $\mathcal{N}(0, \Sigma)$ , and the iterates are drawn from  $\mu$  restricted to appropriate line segments, as defined in Algorithm 2. Note that the only dependence of (23) on  $M$  and  $q$  is through the fraction  $M/q^2$ . Thus, the conditions in Lemma 6 are satisfied. It follows that  $X$  and  $Y_1$  are  $3q$ -independent, and  $X$  and  $Y_2$  are  $3q$ -independent. Since the  $D_i$  and  $P_i$  in the random walks are independent,  $Y_1$  and  $Y_2$  are conditionally independent given  $X$ . Therefore, Lemma 7 shows the result.  $\square$

In the remainder of this section, we aim to show that the conditions of Corollary 2 are satisfied if  $\nu$  and  $\mu$  are Boltzmann distributions with parameters  $\theta_0$  and  $\theta_1$ , respectively, such that  $\|\theta_1 - \theta_0\|_{\theta_0}$  is sufficiently small. Note that Kalai and Vempala [8] only show these conditions to be satisfied if  $\theta_0$  and  $\theta_1$  are collinear. In studying interior point methods, we are also interested in (small) deviations from the central path, so it is important to know that the mixing conditions can be shown to hold for these cases.

We begin with condition (i) from Corollary 2.

**LEMMA 15.** *Let  $\theta_0, \theta_1 \in \mathbb{R}^n$  and  $p \in (0, 1)$ . Let  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  be the density of the Boltzmann distribution with parameter  $\theta_1$  over a convex body  $K \subseteq \mathbb{R}^n$ . Let  $L$  be the level set of  $h$  with probability  $p$ . Then,  $L$  contains a closed  $\|\cdot\|_{\Sigma(\theta_0)^{-1}}$ -ball with radius*

$$\frac{p}{e} \left(1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*\right).$$

*Proof.* Lemma 5.13 from Lovász and Vempala [20] shows that  $L$  contains a  $\langle \cdot, \cdot \rangle_{\Sigma(\theta_1)^{-1}}$ -ball with radius  $p/e$ . In other words, there exists some  $z \in L$  such that for all  $y \in \mathbb{R}^n$  with  $\|y - z\|_{\Sigma(\theta_1)^{-1}} \leq p/e$  it holds that  $y \in L$ . Thus, for all  $y \in \mathbb{R}^n$  with  $\|y - z\|_{\Sigma(\theta_0)^{-1}} \leq (1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*)p/e$ , the second inequality in (5) and Lemma 3 show

$$\|y - z\|_{\Sigma(\theta_1)^{-1}} = \|y - z\|_{x(\theta_1)}^* \leq \frac{\|y - z\|_{x(\theta_0)}^*}{1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*} = \frac{\|y - z\|_{\Sigma(\theta_0)^{-1}}}{1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*} \leq \frac{p}{e},$$

which proves that all such  $x$  lie in  $L$ .  $\square$

Next, we prove upper and lower bounds on the  $L_2$  norm of two Boltzmann distributions. This corresponds to (ii) in Corollary 2.

**LEMMA 16.** *Let  $\mu_0$  and  $\mu_1$  be Boltzmann distributions supported on a convex body  $K \subseteq \mathbb{R}^n$  with parameters  $\theta_0$  and  $\theta_1$  respectively. Then, if  $\|\theta_1 - \theta_0\|_{\theta_0} < 1$ ,*

$$\exp\left(\|\theta_1 - \theta_0\|_{\theta_0}^2 \left[1 + \frac{1}{6}\|\theta_1 - \theta_0\|_{\theta_0}^2 - \frac{2}{3}\|\theta_1 - \theta_0\|_{\theta_0}\right]\right) \leq \|\mu_0/\mu_1\| \leq \frac{\exp(-2\|\theta_1 - \theta_0\|_{\theta_0})}{(1 - \|\theta_1 - \theta_0\|_{\theta_0})^2}.$$

*Proof.* For ease of notation, let  $\theta := \theta_0$  and  $u := \theta_1 - \theta_0$ . By definition,

$$\|\mu_0/\mu_1\| = \mathbb{E}_{\theta_0} \left[ \frac{d\mu_0}{d\mu_1} \right] = \int_K \frac{e^{\langle 2\theta, x \rangle}}{e^{\langle \theta+u, x \rangle}} dx \frac{\int_K e^{\langle \theta+u, x \rangle} dx}{\left(\int_K e^{\langle \theta, x \rangle} dx\right)^2} = \frac{\int_K e^{\langle \theta-u, x \rangle} dx}{\int_K e^{\langle \theta, x \rangle} dx} \frac{\int_K e^{\langle \theta+u, x \rangle} dx}{\int_K e^{\langle \theta, x \rangle} dx}. \quad (24)$$

The key observation is that the natural logarithm of this expression equals

$$\begin{aligned} & f(\theta + u) - f(\theta) + f(\theta - u) - f(\theta) \\ &= \int_0^1 \langle g(\theta + tu), u \rangle dt - \int_0^1 \langle g(\theta - tu), u \rangle dt \\ &= \int_0^1 \left\langle g(\theta) + \int_0^1 H(\theta + stu)(tu) ds, u \right\rangle dt - \int_0^1 \left\langle g(\theta) + \int_0^1 H(\theta - stu)(-tu) ds, u \right\rangle dt \\ &= \int_0^1 t \int_0^1 \langle H(\theta + stu)u, u \rangle ds dt + \int_0^1 t \int_0^1 \langle H(\theta - stu)u, u \rangle ds dt, \end{aligned} \quad (25)$$

where we used the fundamental theorem of calculus twice. By the second inequality in (4),

$$\langle H(\theta + stu)u, u \rangle = \|u\|_{\theta+stu}^2 \leq \frac{\|u\|_{\theta}^2}{(1 - st\|u\|_{\theta})^2},$$

and the same upper bound holds for  $\langle H(\theta - stu)u, u \rangle$ . Then, (25) can be bounded above by

$$2 \int_0^1 t \int_0^1 \frac{\|u\|_{\theta}^2}{(1 - st\|u\|_{\theta})^2} ds dt = -2[\|u\|_{\theta} + \ln(1 - \|u\|_{\theta})],$$

which is non-negative for  $0 \leq \|u\|_{\theta} < 1$ . Since (25) is the natural logarithm of (24),

$$\|\mu_0/\mu_1\| \leq \exp(-2[\|u\|_{\theta} + \ln(1 - \|u\|_{\theta})]) = \frac{\exp(-2\|u\|_{\theta})}{(1 - \|u\|_{\theta})^2}.$$

The lower bound on  $\|\mu_0/\mu_1\|$  follows similarly after noting

$$\langle H(\theta + stu)u, u \rangle = \|u\|_{\theta+stu}^2 \geq \|u\|_{\theta}^2(1 - st\|u\|_{\theta})^2.$$

□

Both the lower and upper bound in Lemma 16 have Taylor approximations  $1 + \|\theta_1 - \theta_0\|_{\theta_0}^2 + O(\|\theta_1 - \theta_0\|_{\theta_0}^3)$  at  $\|\theta_1 - \theta_0\|_{\theta_0} = 0$ .

The bounds in this theorem are more general than the ones used in Kalai and Vempala [8, Lemma 4.4]. They consider the case where  $\theta_1 = (1 + \alpha)\theta_0$  for some  $\alpha \in (-1, 1)$ . By using the log-concavity of the Boltzmann distribution, they show

$$\|\mu_0/\mu_1\| \leq \frac{1}{(1 + \alpha)^n(1 - \alpha)^n}, \quad (26)$$

where  $\mu_0$  and  $\mu_1$  are Boltzmann distributions with parameters  $\theta_0$  and  $\theta_1$ , respectively. This bound may outperform Lemma 16, but that comes at the cost of generality.

Finally, we show that condition (iii) in Corollary 2 holds.

LEMMA 17. *Let  $\langle \cdot, \cdot \rangle$  be the Euclidean inner product, and suppose  $\theta_0, \theta_1 \in \mathbb{R}^n$  satisfy  $\|\theta_0 - \theta_1\|_{\theta_0} < 1$ . Then,*

$$\mathbb{E}_{\theta_1} \left[ \|X - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_0)^{-1}}^2 \right] \leq \frac{n}{(1 - \|\theta_0 - \theta_1\|_{\theta_0})^2}.$$

*Proof.* By using the cyclic permutation invariance of the trace,

$$\begin{aligned} \mathbb{E}_{\theta_1} \left[ \|X - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_0)^{-1}}^2 \right] &= \mathbb{E}_{\theta_1} \left[ \text{tr} \left[ (X - \mathbb{E}_{\theta_1}[X])^{\top} \Sigma(\theta_0)^{-1} (X - \mathbb{E}_{\theta_1}[X]) \right] \right] \\ &= \text{tr} \left[ \Sigma(\theta_1) \Sigma(\theta_0)^{-1} \right] \\ &= \text{tr} \left[ \Sigma(\theta_0)^{-1/2} \Sigma(\theta_1) \Sigma(\theta_0)^{-1/2} \right]. \end{aligned}$$

We can upper bound  $\text{tr}[\Sigma(\theta_0)^{-1/2} \Sigma(\theta_1) \Sigma(\theta_0)^{-1/2}] = \text{tr}[H(\theta_0)^{-1/2} H(\theta_1) H(\theta_0)^{-1/2}]$  by

$$n \max_{u \in \mathbb{R}^n} \frac{\langle H(\theta_0)^{-1/2} u, H(\theta_1) H(\theta_0)^{-1/2} u \rangle}{\|u\|^2} = n \max_{v \in \mathbb{R}^n} \frac{\langle v, H(\theta_1) v \rangle}{\|H(\theta_0)^{1/2} v\|^2} = n \max_{v \in \mathbb{R}^n} \frac{\|v\|_{\theta_1}^2}{\|v\|_{\theta_0}^2}.$$

The claim now follows from (4). □

It is interesting to compare the upper bound in Lemma 17 with the one Kalai and Vempala [8] arrive at through a near-isotropy argument. It is shown by [8, Lemma 4.2 and 4.3] that

$$\mathbb{E}_{\theta_1} \left[ \|X - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_0)^{-1}}^2 \right] \leq 16n \|\mu_1/\mu_0\| \max_{v \in \mathbb{R}^n} \frac{\mathbb{E}_{\theta_0} \left[ \langle v, X - \mathbb{E}_{\theta_0}[X] \rangle_{\Sigma(\theta_0)^{-1}}^2 \right]}{\|v\|_{\Sigma(\theta_0)^{-1}}^2}, \quad (27)$$

where  $\mu_0$  and  $\mu_1$  are Boltzmann distributions with parameters  $\theta_0$  and  $\theta_1$ , respectively. Observe that for all  $v \in \mathbb{R}^n$ ,

$$\frac{\mathbb{E}_{\theta_0} [\langle \Sigma(\theta_0)^{-1}v, X - \mathbb{E}_{\theta_0}[X] \rangle^2]}{\|v\|_{\Sigma(\theta_0)^{-1}}^2} = \frac{\langle \Sigma(\theta_0)^{-1}v, \Sigma(\theta_0)\Sigma(\theta_0)^{-1}v \rangle}{\|v\|_{\Sigma(\theta_0)^{-1}}^2} = 1,$$

and therefore the right hand side of (27) is just  $16n\|\mu_1/\mu_0\|$ . If we upper bound this norm by Lemma 16, we find

$$\mathbb{E}_{\theta_1} \left[ \|X - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_0)^{-1}}^2 \right] \leq \frac{16n \exp(-2\|\theta_1 - \theta_0\|_{\theta_1})}{(1 - \|\theta_1 - \theta_0\|_{\theta_1})^2}. \quad (28)$$

By the second inequality in (4), we have

$$\frac{n}{(1 - \|\theta_0 - \theta_1\|_{\theta_0})^2} \leq \frac{n}{\left(1 - \frac{\|\theta_0 - \theta_1\|_{\theta_1}}{1 - \|\theta_0 - \theta_1\|_{\theta_1}}\right)^2} \leq \frac{16n \exp(-2\|\theta_1 - \theta_0\|_{\theta_1})}{(1 - \|\theta_1 - \theta_0\|_{\theta_1})^2},$$

where the second inequality holds for  $\|\theta_1 - \theta_0\|_{\theta_1} \leq 0.438$ . In this case, the bound in Lemma 17 is stronger than (28). Alternatively, if  $\theta_0 = (1 + \alpha)\theta_1$  for some  $\alpha \in (-1, 1)$ , (26) shows that (27) can be bounded by

$$\mathbb{E}_{\theta_1} \left[ \|X - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_0)^{-1}}^2 \right] \leq \frac{16n}{(1 + \alpha)^n(1 - \alpha)^n}. \quad (29)$$

Since (13) shows  $\|\theta_1 - \theta_0\|_{\theta_0} = \alpha\|\theta_0\|_{\theta_0} \leq \alpha\sqrt{\vartheta} = \alpha\sqrt{n + o(n)}$ , the upper bound from Lemma 17 is better than (29) for sufficiently large  $n$  and  $\alpha\sqrt{\vartheta} < 1$ .

The results from this section can be summarized as follows.

**THEOREM 4.** *Let  $K \subseteq \mathbb{R}^n$  be a convex body, and let  $\langle \cdot, \cdot \rangle$  be the Euclidean inner product. Let  $q > 0$ , and  $\theta_0, \theta_1 \in \mathbb{R}^n$  such that  $\Delta\theta := \|\theta_1 - \theta_0\|_{\theta_0} < 1$  and  $\Delta x := \|x(\theta_1) - x(\theta_0)\|_{x(\theta_0)}^* < 1$ . Pick  $\epsilon \in [0, 1)$ , and suppose we have an invertible matrix  $\widehat{\Sigma}(\theta_0)$  such that*

$$(1 - \epsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \leq y^\top \Sigma(\theta_0)^{-1}y \leq (1 + \epsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \quad \forall y \in \mathbb{R}^n. \quad (30)$$

*Consider a hit-and-run random walk as in Algorithm 2 applied to the Boltzmann distribution  $\mu$  with parameter  $\theta_1$  from a random starting point drawn from a Boltzmann distribution with parameter  $\theta_0$ . Let  $\nu^{(\ell)}$  be the distribution of the hit-and-run point after  $\ell$  steps of hit-and-run sampling applied to  $\mu$ , where the directions are drawn from a  $\mathcal{N}(0, \widehat{\Sigma}(\theta_0))$ -distribution. Then, after*

$$\ell = \left\lceil \frac{1 + \epsilon}{1 - \epsilon} \frac{64e^2 n^3 10^{30}}{(1 - \Delta\theta)^2 (1 - \Delta x)^2} \ln^2 \left( \sqrt{\frac{1 + \epsilon}{1 - \epsilon}} \frac{16en\sqrt{n} \exp(-2\Delta\theta)}{q^2 (1 - \Delta\theta)^3 (1 - \Delta x)} \right) \ln^3 \left( \frac{2 \exp(-2\Delta\theta)}{q^2 (1 - \Delta\theta)^2} \right) \right\rceil, \quad (31)$$

*hit-and-run steps, we have  $\|\nu^{(\ell)} - \mu\| \leq q$ .*

*Proof.* We will apply Corollary 2 with respect to  $\widehat{\Sigma}(\theta_0)$ .

By Lemma 15, the level set of  $\mu$  with probability  $\frac{1}{8}$  contains a  $\|\cdot\|_{\Sigma(\theta_0)^{-1}}$ -ball with radius  $\frac{1}{8e}(1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*)$ . Denote the center of this ball by  $z \in K$ . Then, for all  $y \in K$  with  $\|y - z\|_{\widehat{\Sigma}(\theta_0)^{-1}} \leq \frac{1}{8e\sqrt{1+\epsilon}}(1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*)$ , it can be seen from (30) that

$$\|y - z\|_{\Sigma(\theta_0)^{-1}} \leq \sqrt{1+\epsilon}\|y - z\|_{\widehat{\Sigma}(\theta_0)^{-1}} \leq \frac{1}{8e}(1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*),$$

and thus  $y$  lies in the level set. Therefore, the level set of  $\mu$  with probability  $\frac{1}{8}$  contains a  $\|\cdot\|_{\widehat{\Sigma}(\theta_0)^{-1}}$ -ball with radius  $\frac{1}{8e\sqrt{1+\epsilon}}(1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*)$ .

Moreover, (30) and Lemma 17 show

$$\mathbb{E}_{\theta_1}[\|X - \mathbb{E}_{\theta_1}[X]\|_{\widehat{\Sigma}(\theta_0)^{-1}}^2] \leq \frac{1}{1-\epsilon}\mathbb{E}_{\theta_1}[\|X - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_0)^{-1}}^2] \leq \frac{n}{(1-\epsilon)(1 - \|\theta_0 - \theta_1\|_{\theta_0})^2}.$$

Using  $s = \frac{1}{8e\sqrt{1+\epsilon}}(1 - \|x(\theta_0) - x(\theta_1)\|_{x(\theta_0)}^*)$ ,  $S^2 = \frac{n}{(1-\epsilon)(1 - \|\theta_0 - \theta_1\|_{\theta_0})^2}$ , and Lemma 16, Corollary 2 now proves the result.  $\square$

Since Theorem 4 is essentially an application of Corollary 2, Lemma 14 also holds in this setting. For ease of reference, we state this result below.

**LEMMA 18.** *Let  $K \subseteq \mathbb{R}^n$  be a convex body, and let  $q > 0$ . Let  $\theta_0, \theta_1 \in \mathbb{R}^n$  such that the conditions of Theorem 4 are satisfied for some  $\epsilon$  and  $\widehat{\Sigma}(\theta_0)$ . Let  $X$  be a random variable following a Boltzmann distribution supported on  $K$  with parameter  $\theta_0$ . Consider two hit-and-run random walks as in Algorithm 2, applied to the Boltzmann distribution with parameter  $\theta_1$ , both starting from the realization of  $X$ , where all  $D_i$  and  $P_i$  in one random walk are independent of all  $D_i$  and  $P_i$  in the other random walk. Let the number of steps  $\ell$  of both walks be given by (31), and call the resulting end points  $Y_1$  and  $Y_2$ . Then,  $Y_1$  and  $Y_2$  are  $6q$ -independent.*

**5. Sampling Quality Guarantees** In this section, we provide probabilistic guarantees on the quality of the empirical mean and covariance estimates of a log-concave distribution  $\mu$ . We will repeatedly use that for random variables  $Y$  and  $Z$  taking values in a set  $K$  and a function  $\phi$  on  $K$ ,

$$\mathbb{E}[\phi(Y)] = \mathbb{E}[\phi(Z)] + \mathbb{E}[\phi(Y) - \phi(Z)] = \mathbb{E}[\phi(Z)] + \mathbb{E}[\phi(Y) - \phi(Z) | Y \neq Z] \mathbb{P}\{Y \neq Z\}. \quad (32)$$

We start by analyzing the quality of the mean estimate.

**THEOREM 5.** *Let  $K \subseteq \mathbb{R}^n$  be a convex body, and let  $\langle \cdot, \cdot \rangle$  be the Euclidean inner product. Suppose  $K$  is contained in a Euclidean ball with radius  $R > 0$ . Let  $\alpha > 0$ ,  $p \in (0, 1)$ , and  $\epsilon \in [0, 1)$ . Let  $\theta_0, \theta_1 \in \mathbb{R}^n$  such that  $\Delta\theta := \|\theta_1 - \theta_0\|_{\theta_0} < 1$  and  $\Delta x := \|x(\theta_1) - x(\theta_0)\|_{x(\theta_0)}^* < 1$ . Suppose we have an invertible matrix  $\widehat{\Sigma}(\theta_0)$  such that*

$$(1 - \epsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \leq y^\top \Sigma(\theta_0)^{-1}y \leq (1 + \epsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \quad \forall y \in \mathbb{R}^n. \quad (33)$$

*Pick*

$$N \geq \frac{2n}{p\alpha^2}, \quad q \leq \frac{p \min\{1, \alpha^2\}}{204nR^2} \lambda_{\min}(\Sigma(\theta_1)).$$

*Let  $X_0$  be a random starting point drawn from a Boltzmann distribution with parameter  $\theta_0$ . Let  $Y^{(1)}, \dots, Y^{(N)}$  be the end points of  $N$  hit-and-run random walks applied to the Boltzmann distribution with parameter  $\theta_1$  having starting point  $X_0$ , where the directions are drawn from a  $\mathcal{N}(0, \widehat{\Sigma}(\theta_0))$ -distribution, and each walk has length  $\ell$  given by (31). (Note that  $\ell$  depends on  $\epsilon$ ,  $n$ ,  $q$ ,  $\Delta\theta$ , and  $\Delta x$ .) Then, the empirical mean  $\widehat{Y} := \frac{1}{N} \sum_{j=1}^N Y^{(j)}$  satisfies*

$$\mathbb{P}\left\{\|\widehat{Y} - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_1)^{-1}} \leq \alpha\right\} \geq 1 - p. \quad (34)$$

*Proof.* Theorem 4 ensures that the distributions of the samples  $Y^{(1)}, \dots, Y^{(N)}$  all have a total variation distance to the Boltzmann distribution with parameter  $\theta_1$  of at most  $q$ . By Lemma 18, the samples are pairwise  $6q$ -independent. It therefore remains to be shown that  $N$  pairwise  $6q$ -independent samples with total variation distance to the Boltzmann distribution with parameter  $\theta_1$  of at most  $q$  are enough to guarantee (34).

We start by investigating an expression resembling the variance of  $\widehat{Y}$  in the norm induced by  $\Sigma(\theta_1)^{-1}$ :

$$\begin{aligned} \mathbb{E}[\|\widehat{Y} - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_1)^{-1}}^2] &= \frac{1}{N^2} \sum_{j=1}^N \mathbb{E}[(Y^{(j)} - \mathbb{E}_{\theta_1}[X])^\top \Sigma(\theta_1)^{-1} (Y^{(j)} - \mathbb{E}_{\theta_1}[X])] \\ &\quad + \frac{1}{N^2} \sum_{j=1}^N \sum_{k \neq j} \mathbb{E}[(Y^{(j)} - \mathbb{E}_{\theta_1}[X])^\top \Sigma(\theta_1)^{-1} (Y^{(k)} - \mathbb{E}_{\theta_1}[X])]. \end{aligned} \quad (35)$$

The first term of (35) can be bounded if one notes that Lemma 5 guarantees that for each  $Y^{(j)}$  there exists a  $Z^{(j)}$  with Boltzmann distribution with parameter  $\theta_1$  and  $\mathbb{P}\{Y^{(j)} = Z^{(j)}\} \geq 1 - q$ . Using (32), we have for all  $j \in \{1, \dots, N\}$ ,

$$\begin{aligned} &\mathbb{E}[(Y^{(j)} - \mathbb{E}_{\theta_1}[X])^\top \Sigma(\theta_1)^{-1} (Y^{(j)} - \mathbb{E}_{\theta_1}[X])] \\ &\leq \mathbb{E}[(Z^{(j)} - \mathbb{E}_{\theta_1}[X])^\top \Sigma(\theta_1)^{-1} (Z^{(j)} - \mathbb{E}_{\theta_1}[X])] + q \lambda_{\max}(\Sigma(\theta_1)^{-1}) \|Y^{(j)} - \mathbb{E}_{\theta_1}[X]\|^2 \\ &= \text{tr}(\mathbb{E}[(Z^{(j)} - \mathbb{E}_{\theta_1}[X])(Z^{(j)} - \mathbb{E}_{\theta_1}[X])^\top] \Sigma(\theta_1)^{-1}) + q \frac{\|Y^{(j)} - \mathbb{E}_{\theta_1}[X]\|^2}{\lambda_{\min}(\Sigma(\theta_1))} \\ &\leq n + q \frac{(2R)^2}{\lambda_{\min}(\Sigma(\theta_1))}. \end{aligned}$$

To bound the second term of (35), note that since  $Y^{(j)}$  and  $Y^{(k)}$  are  $6q$ -independent, so are  $\Sigma(\theta_1)^{-1/2}(Y^{(j)} - \mathbb{E}_{\theta_1}[X])$  and  $\Sigma(\theta_1)^{-1/2}(Y^{(k)} - \mathbb{E}_{\theta_1}[X])$ . By Lemma 8 and Lemma 9, we have for all  $j \neq k$ ,

$$\begin{aligned} &\mathbb{E}[(Y^{(j)} - \mathbb{E}_{\theta_1}[X])^\top \Sigma(\theta_1)^{-1} (Y^{(k)} - \mathbb{E}_{\theta_1}[X])] \\ &= \sum_{i=1}^n \mathbb{E}[(\Sigma(\theta_1)^{-1/2}(Y^{(j)} - \mathbb{E}_{\theta_1}[X]))_i (\Sigma(\theta_1)^{-1/2}(Y^{(k)} - \mathbb{E}_{\theta_1}[X]))_i] \\ &\leq \sum_{i=1}^n [(\mathbb{E}[\Sigma(\theta_1)^{-1/2}(Y^{(j)} - \mathbb{E}_{\theta_1}[X]))_i] (\mathbb{E}[\Sigma(\theta_1)^{-1/2}(Y^{(k)} - \mathbb{E}_{\theta_1}[X]))_i] \\ &\quad + 4(6q)(2R\lambda_{\max}(\Sigma(\theta_1)^{-1/2}))^2]. \end{aligned}$$

Using Lemma 5 and (32) in the same manner as before, we get for all  $j$ ,

$$|(\mathbb{E}[\Sigma(\theta_1)^{-1/2}(Y^{(j)} - \mathbb{E}_{\theta_1}[X]))_i]| \leq 0 + q(2R)\lambda_{\max}(\Sigma(\theta_1)^{-1/2}) = \frac{2Rq}{\sqrt{\lambda_{\min}(\Sigma(\theta_1))}}.$$

In conclusion,

$$\begin{aligned} \mathbb{E}[(Y^{(j)} - \mathbb{E}_{\theta_1}[X])^\top \Sigma(\theta_1)^{-1} (Y^{(j)} - \mathbb{E}_{\theta_1}[X])] &\leq \frac{n}{N} + \frac{4qR^2}{N\lambda_{\min}(\Sigma(\theta_1))} + \frac{96qnR^2}{\lambda_{\min}(\Sigma(\theta_1))} + \frac{4nR^2q^2}{\lambda_{\min}(\Sigma(\theta_1))} \\ &\leq p\alpha^2. \end{aligned}$$

The proof is completed by applying Markov's inequality:

$$\mathbb{P}\left\{\|\widehat{Y} - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_1)^{-1}}^2 > \alpha^2\right\} \leq \frac{\mathbb{E}[\|\widehat{Y} - \mathbb{E}_{\theta_1}[X]\|_{\Sigma(\theta_1)^{-1}}^2]}{\alpha^2} \leq p.$$

□

If  $Y^{(1)}, \dots, Y^{(N)}$  are random variables, we define the associated empirical covariance matrix as

$$\widehat{\Sigma} := \frac{1}{N} \sum_{j=1}^N Y^{(j)} (Y^{(j)})^\top - \left( \frac{1}{N} \sum_{j=1}^N Y^{(j)} \right) \left( \frac{1}{N} \sum_{j=1}^N Y^{(j)} \right)^\top. \quad (36)$$

Before we can prove a result similar to Theorem 5 for the empirical covariance, it will be helpful to prove that, without loss of generality, we can assume that the underlying distribution has identity covariance and mean zero.

**LEMMA 19.** *Let  $K$  be a convex body and let  $\widehat{\Sigma}$  be the empirical covariance matrix of a distribution  $\mu$  over  $K$  with density  $h$  based on samples  $X^{(1)}, \dots, X^{(N)}$  as in (36). Denote the set  $\{Ax + b : x \in K\}$  by  $AK + b$ , where  $A \in \mathbb{R}^{n \times n}$  is of full rank. Let  $h'(y) = \det(A^{-1})h(A^{-1}(y - b))$  be a probability density over  $AK + b$  with induced distribution  $\mu'$ . Let  $\widehat{\Sigma}'$  be the empirical covariance matrix of  $\mu'$  based on the samples  $Y^{(j)} = AX^{(j)} + b$  for  $j \in \{1, \dots, N\}$  as in (36). Let the true covariance matrix of  $\mu$  be  $\Sigma$ , and the covariance matrix of  $\mu'$  be  $\Sigma'$ . Then, for any  $\varepsilon \in [0, 1]$ ,*

$$(1 - \varepsilon)u^\top \widehat{\Sigma} u \leq u^\top \Sigma u \leq (1 + \varepsilon)u^\top \widehat{\Sigma} u \quad \forall u \in \mathbb{R}^n, \quad (37)$$

if and only if

$$(1 - \varepsilon)v^\top \widehat{\Sigma}' v \leq v^\top \Sigma' v \leq (1 + \varepsilon)v^\top \widehat{\Sigma}' v \quad \forall v \in \mathbb{R}^n. \quad (38)$$

*Proof.* Let  $\widehat{X} = \frac{1}{N} \sum_{j=1}^N X^{(j)}$  and  $\widehat{\Sigma} = \frac{1}{N} \sum_{j=1}^N (X^{(j)} - \widehat{X})(X^{(j)} - \widehat{X})^\top$ . The empirical covariance matrix of  $\mu'$  is

$$\begin{aligned} \widehat{\Sigma}' &= \frac{1}{N} \sum_{j=1}^N (AX^{(j)} + b - (A\widehat{X} + b))(AX^{(j)} + b - (A\widehat{X} + b))^\top \\ &= A \left( \frac{1}{N} \sum_{j=1}^N (X^{(j)} - \widehat{X})(X^{(j)} - \widehat{X})^\top \right) A^\top = A\widehat{\Sigma}A^\top. \end{aligned}$$

Similarly, if  $X$  has distribution  $\mu$  and  $Y$  has distribution  $\mu'$ ,

$$\begin{aligned} \Sigma' &= \int_{AK+b} (y - \mathbb{E}[Y])(y - \mathbb{E}[Y])^\top h'(y) dy \\ &= \int_K (Ax + b - (A\mathbb{E}[X] + b))(Ax + b - (A\mathbb{E}[X] + b))^\top h(x) dx = A\Sigma A^\top. \end{aligned}$$

The equivalence of (37) and (38) follows by taking  $u = Av$ .  $\square$

With this lemma, we are ready to bound the number of samples required to find an approximation of the covariance matrix of  $\mu$  satisfying a certain quality criterion.

**THEOREM 6.** *Let  $K \subseteq \mathbb{R}^n$  be a convex body, and let  $\langle \cdot, \cdot \rangle$  be the Euclidean inner product. Suppose  $K$  is contained in a Euclidean ball with radius  $R > 0$ . Let  $\varepsilon, p \in (0, 1)$ , and  $\epsilon \in [0, 1]$ . Let  $\theta_0, \theta_1 \in \mathbb{R}^n$  such that  $\Delta\theta := \|\theta_1 - \theta_0\|_{\theta_0} < 1$  and  $\Delta x := \|x(\theta_1) - x(\theta_0)\|_{x(\theta_0)}^* < 1$ . Suppose we have an invertible matrix  $\widehat{\Sigma}(\theta_0)$  such that*

$$(1 - \epsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \leq y^\top \Sigma(\theta_0)^{-1}y \leq (1 + \epsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \quad \forall y \in \mathbb{R}^n.$$

Pick

$$N \geq \frac{490n^2}{p\varepsilon^2}, \quad q \leq \frac{p\varepsilon^2}{49980n^2R^4} \lambda_{\min}(\Sigma(\theta_1))^2. \quad (39)$$

Let  $X_0$  be a random starting point drawn from a Boltzmann distribution with parameter  $\theta_0$ . Let  $Y^{(1)}, \dots, Y^{(N)}$  be the end points of  $N$  hit-and-run random walks applied to the Boltzmann distribution

with parameter  $\theta_1$  having starting point  $X_0$ , where the directions are drawn from a  $\mathcal{N}(0, \widehat{\Sigma}(\theta_0))$ -distribution, and each walk has length  $\ell$  given by (31). (Note that  $\ell$  depends on  $\epsilon$ ,  $n$ ,  $q$ ,  $\Delta\theta$ , and  $\Delta x$ .) Then, the empirical covariance matrix  $\widehat{\Sigma} \approx \Sigma(\theta_1)$  as defined in (36) satisfies

$$\mathbb{P} \left\{ (1 - \epsilon)v^\top \widehat{\Sigma} v \leq v^\top \Sigma(\theta_1) v \leq (1 + \epsilon)v^\top \widehat{\Sigma} v \quad \forall v \in \mathbb{R}^n \right\} \geq 1 - p. \quad (40)$$

*Proof.* By the same argument as in Theorem 5,  $Y^{(1)}, \dots, Y^{(N)}$  are pairwise  $6q$ -independent samples, each with a distribution that has total variation distance to the Boltzmann distribution with parameter  $\theta_1$  of at most  $q$ .

The remainder of the proof uses an approach similar to Theorem 5.11 from Kannan, Lovász and Simonovits [10], although their result only applies to the uniform distribution. As in Theorem 5, define  $\widehat{Y} = \frac{1}{N} \sum_{j=1}^N Y^{(j)}$ . Lemma 19 shows that applying an affine transformation to  $K$  does not affect the statement. We can therefore assume that the Boltzmann distribution with parameter  $\theta_1$  is isotropic, i.e. has identity covariance, and that the mean of this distribution is the origin. However, the support of this distribution is now contained in a ball of radius  $\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R$ .

We want to prove that with probability at least  $1 - p$ , for every  $v \in \mathbb{R}^n$ ,

$$(1 - \epsilon)v^\top \widehat{\Sigma} v \leq \|v\|^2 \leq (1 + \epsilon)v^\top \widehat{\Sigma} v,$$

or equivalently,

$$\frac{1}{1 + \epsilon} \leq \frac{v^\top \widehat{\Sigma} v}{\|v\|^2} \leq \frac{1}{1 - \epsilon}. \quad (41)$$

We may therefore assume in the remainder that  $\|v\| = 1$ . Letting  $S := \frac{1}{N} \sum_{j=1}^N Y^{(j)}(Y^{(j)})^\top$ , (41) is equivalent to

$$\frac{1}{1 + \epsilon} + (v^\top \widehat{Y})^2 \leq v^\top S v \leq \frac{1}{1 - \epsilon} + (v^\top \widehat{Y})^2. \quad (42)$$

We will use that

$$v^\top S v = v^\top v + v^\top (S - I)v = 1 + v^\top (S - I)v.$$

We continue by showing that  $\mathbb{P}\{\rho(S - I) > \frac{34}{35}\epsilon/(1 + \epsilon)\}$  is small, where  $\rho(S - I)$  is the spectral radius of  $S - I$ . It is known that

$$\rho(S - I) = \sqrt{\lambda_{\max}([S - I]^2)} \leq \sqrt{\text{tr}([S - I]^2)}. \quad (43)$$

To apply Markov's inequality, we will bound  $\mathbb{E}[\text{tr}([S - I]^2)] = \mathbb{E}[\text{tr}(S^2 - 2S)] + n$ . By Lemma 5, for each  $Y^{(j)}$  we can find a  $Z^{(j)}$  following a Boltzmann distribution with parameter  $\theta_1$  such that  $\mathbb{P}\{Z^{(j)} \neq Y^{(j)}\} \leq q$ . Then, using (32),

$$\begin{aligned} \mathbb{E}[\text{tr}(S)] &= \mathbb{E} \left[ \sum_{r=1}^n \frac{1}{N} \sum_{j=1}^N (Y_r^{(j)})^2 \right] \\ &\geq \mathbb{E} \left[ \sum_{r=1}^n \frac{1}{N} \sum_{j=1}^N (Z_r^{(j)})^2 \right] - qn (2R\lambda_{\max}(\Sigma(\theta_1)^{-1/2}))^2 \\ &= n - qn (2R\lambda_{\max}(\Sigma(\theta_1)^{-1/2}))^2, \end{aligned} \quad (44)$$

where the last inequality uses that the origin is contained in the support of the Boltzmann distribution with parameter  $\theta_1$ , since we assumed the origin is the mean of this distribution.

Finding a bound on  $\mathbb{E}[\text{tr}(S^2)]$  requires more work. Note that

$$\mathbb{E}[\text{tr}(S^2)] = \frac{1}{N^2} \mathbb{E} \left[ \sum_{j=1}^N \sum_{k=1}^N ((Y^{(j)})^\top Y^{(k)})^2 \right] = \frac{1}{N^2} \sum_{j=1}^N \mathbb{E} [\|Y^{(j)}\|^4] + \frac{1}{N^2} \mathbb{E} \left[ \sum_{j=1}^N \sum_{k \neq j} ((Y^{(j)})^\top Y^{(k)})^2 \right].$$



By another application of Lemma 5 and (32),  $\mathbb{E}[\|Y^{(j)}\|^4]$  can be bounded as

$$\mathbb{E}[\|Y^{(j)}\|^4] \leq \mathbb{E}[\|Z^{(j)}\|^4] + q(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^4.$$

Observe that since the Boltzmann distribution with parameter  $\theta_1$  is isotropic,  $\mathbb{E}[\|Z^{(j)}\|^2] = n$  for all  $j$ . Hence, by Lemma 4, we have  $\mathbb{P}\{\|Z^{(j)}\| > t\sqrt{n}\} \leq e^{1-t}$  for all  $t > 1$ . By a change of variables  $s = n^2 t^4$ ,

$$\begin{aligned} \mathbb{E}[\|Z^{(j)}\|^4] &= \int_0^\infty \mathbb{P}\{\|Z^{(j)}\|^4 > s\} ds \\ &= \int_0^\infty \mathbb{P}\{\|Z^{(j)}\| > t\sqrt{n}\} 4t^3 n^2 dt \\ &\leq \int_0^1 4t^3 n^2 dt + \int_1^\infty 4t^3 n^2 e^{1-t} dt = 65n^2. \end{aligned}$$

It remains to bound  $\mathbb{E}[(Y^{(j)})^\top Y^{(k)}]^2$  for  $k \neq j$ . It follows from the  $6q$ -independence of  $Y^{(j)}$  and  $Y^{(k)}$ , combined with Lemma 8 and Lemma 9, that

$$\begin{aligned} \mathbb{E}[(Y^{(j)})^\top Y^{(k)}]^2 &= \mathbb{E}\left[\sum_{r=1}^n \sum_{s=1}^n (Y_r^{(j)} Y_r^{(k)}) (Y_s^{(j)} Y_s^{(k)})\right] \\ &\leq \sum_{r=1}^n \sum_{s=1}^n \left(\mathbb{E}[Y_r^{(j)} Y_s^{(j)}] \mathbb{E}[Y_r^{(k)} Y_s^{(k)}] + 4(6q)(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^2\right). \end{aligned}$$

Applying the now familiar Lemma 5 and (32) once more, we find for all  $j$  that

$$\mathbb{E}[Y_r^{(j)} Y_s^{(j)}] \leq \mathbb{E}[Z_r^{(j)} Z_s^{(j)}] + q(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^2.$$

Note that because  $Z^{(j)}$  is isotropic,  $\mathbb{E}[Z_r^{(j)} Z_s^{(j)}]$  is one if  $r = s$  and zero otherwise. Therefore,

$$\begin{aligned} &\sum_{r=1}^n \sum_{s=1}^n \mathbb{E}[Y_r^{(j)} Y_s^{(j)}] \mathbb{E}[Y_r^{(k)} Y_s^{(k)}] \\ &\leq n \left(1 + q(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^2\right)^2 + n(n-1) \left(q(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^2\right)^2. \end{aligned}$$

In summary, the values of  $N$  and  $q$  in (39) give us

$$\begin{aligned} \mathbb{E}[\text{tr}(S^2)] &\leq \frac{1}{N} \left(65n^2 + q(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^4\right) \\ &\quad + \frac{N(N-1)}{N^2} \left[ n \left(1 + q(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^2\right)^2 \right. \\ &\quad \left. + n(n-1) \left(q(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^2\right)^2 + 4n^2(6q)(2\lambda_{\max}(\Sigma(\theta_1)^{-1/2})R)^2 \right] \\ &\leq n + 0.14p\varepsilon^2. \end{aligned}$$

Combined with (44), the above yields the following upper bound on  $\mathbb{E}[\text{tr}([S - I]^2)] = \mathbb{E}[\text{tr}(S^2 - 2S)] + n$ :

$$\mathbb{E}[\text{tr}([S - I]^2)] \leq n + 0.14p\varepsilon^2 - 2n + 2qn(2R\lambda_{\max}(\Sigma(\theta_1)^{-1/2}))^2 + n \leq \frac{p\varepsilon^2}{7}.$$

By (43) and Markov's inequality,

$$\mathbb{P}\left\{\rho(S-I) > \frac{34}{35} \frac{\varepsilon}{1+\varepsilon}\right\} \leq \frac{\mathbb{E}[\text{tr}([S-I]^2)]}{\left(\frac{34}{35} \frac{\varepsilon}{1+\varepsilon}\right)^2} \leq \frac{p\varepsilon^2/7}{\left(\frac{34}{35} \frac{\varepsilon}{1+\varepsilon}\right)^2} = p(1+\varepsilon)^2 \frac{175}{1156}.$$

Using Theorem 5 for  $\alpha^2 = \frac{\varepsilon}{35(1+\varepsilon)}$ , we find that the  $N$  and  $q$  in (39) yield the quality guarantee

$$\mathbb{P}\left\{\|\widehat{Y}\| > \sqrt{\frac{\varepsilon}{35(1+\varepsilon)}}\right\} \leq \frac{2p\varepsilon^2}{490n \frac{\varepsilon}{35(1+\varepsilon)}} \leq \frac{p\varepsilon(1+\varepsilon)}{7}.$$

Thus, we have for all  $\varepsilon \in (0, 1)$ ,

$$\mathbb{P}\left\{\rho(S-I) \leq \frac{34}{35} \frac{\varepsilon}{1+\varepsilon} \wedge \|\widehat{Y}\| \leq \sqrt{\frac{\varepsilon}{35(1+\varepsilon)}}\right\} \geq 1 - p(1+\varepsilon)^2 \frac{175}{1156} - p \frac{\varepsilon(1+\varepsilon)}{7} \geq 1 - p.$$

We can now verify the inequalities in (42) to complete the proof. With probability at least  $1 - p$ ,

$$\frac{1}{1+\varepsilon} + \|\widehat{Y}\|^2 \leq \frac{1+\varepsilon/35}{1+\varepsilon} = 1 - \frac{34}{35} \frac{\varepsilon}{1+\varepsilon} \leq 1 - \rho(S-I),$$

thereby verifying the first inequality in (42) for all unit vectors  $v$ . The second inequality in (42) can be shown by noting that with probability at least  $1 - p$ ,

$$1 + \rho(S-I) \leq 1 + \frac{\varepsilon}{1+\varepsilon} \leq \frac{1}{1-\varepsilon} \leq \frac{1}{1-\varepsilon} + (v^\top \widehat{Y})^2,$$

for all  $v \in \mathbb{R}^n$ .  $\square$

One might expect that if all eigenvalue-eigenvector pairs are approximated well enough for (40) to hold, this also implies a similar statement relating  $\widehat{\Sigma}^{-1}$  and  $\Sigma(\theta_1)^{-1}$ . This is confirmed by the following 'folklore' result from linear algebra.

LEMMA 20. *Let  $\Sigma$  and  $\widehat{\Sigma}$  be two symmetric invertible matrices such that*

$$(1-\varepsilon)v^\top \widehat{\Sigma}v \leq v^\top \Sigma v \leq (1+\varepsilon)v^\top \widehat{\Sigma}v \quad \forall v \in \mathbb{R}^n,$$

for some constant  $0 \leq \varepsilon < \frac{1}{2}(\sqrt{5}-1) \approx 0.618$ . Then, for all  $u \in \mathbb{R}^n$ ,

$$\left(1 - \frac{2\varepsilon^2 + \varepsilon}{1 + \varepsilon^2 + \varepsilon}\right) u^\top \widehat{\Sigma}^{-1}u \leq u^\top \Sigma^{-1}u \leq \left(1 + \frac{2\varepsilon^2 + 3\varepsilon}{1 - \varepsilon^2 - \varepsilon}\right) u^\top \widehat{\Sigma}^{-1}u.$$

We can combine this lemma with Theorem 6 to bound the number of samples required to approximate both  $\Sigma(\theta_1)$  and  $\Sigma(\theta_1)^{-1}$  to a desired accuracy.

COROLLARY 3. *Let  $K \subseteq \mathbb{R}^n$  be a convex body, and let  $\langle \cdot, \cdot \rangle$  be the Euclidean inner product. Suppose  $K$  is contained in a Euclidean ball with radius  $R > 0$ . Let  $p \in (0, 1)$ ,  $\varepsilon \in [0, 1)$ , and  $0 < \varepsilon_1 \leq \sqrt{13} - 3 \approx 0.606$ . Let  $\theta_0, \theta_1 \in \mathbb{R}^n$  such that  $\Delta\theta := \|\theta_1 - \theta_0\|_{\theta_0} < 1$  and  $\Delta x := \|x(\theta_1) - x(\theta_0)\|_{x(\theta_0)}^* < 1$ . Suppose we have an invertible matrix  $\widehat{\Sigma}(\theta_0)$  such that*

$$(1-\varepsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \leq y^\top \Sigma(\theta_0)^{-1}y \leq (1+\varepsilon)y^\top \widehat{\Sigma}(\theta_0)^{-1}y \quad \forall y \in \mathbb{R}^n.$$

Pick

$$N \geq \frac{7840n^2}{p\varepsilon_1^2}, \quad q \leq \frac{p\varepsilon_1^2}{799680n^2R^4} \lambda_{\min}(\Sigma(\theta_1))^2.$$

TABLE 2. Overview of the tolerances in Section 6

Symbol	Usage
$\delta$	$\ x_0 - z(\eta_0)\ _{z(\eta_0)}^* \leq \frac{1}{2}\delta$ at start of Algorithm 3
$\epsilon$	Quality of approximation $\widehat{\Sigma}(-\eta_{k+1}c)$ of $\Sigma(-\eta_{k+1}c)$ , as in (47) and (48)
$\tilde{\epsilon}$	Quality of approximation $\widehat{g}_{z(\eta_{k+1})}^{*,\eta_{k+1}}(x_k)$ of $g_{z(\eta_{k+1})}^{*,\eta_{k+1}}(x_k)$
$\widehat{\epsilon}$	Quality of approximation $\widehat{\Sigma}(-\eta_k c)[\eta_{k+1}c + \widehat{\theta}(x_k)]$ of $\Sigma(-\eta_k c)[\eta_{k+1}c + \theta(x_k)]$
$\bar{\epsilon}$	$\langle c, x_k \rangle - \min_{x \in K} \langle c, x \rangle \leq \bar{\epsilon}$ at termination of Algorithm 3
$\epsilon'$	Quality of approximation $\widehat{\Sigma}(\bar{\theta})$ of $\Sigma(\bar{\theta})$ , as in (47) and (48)
$\tilde{\epsilon}'$	Quality of approximation $\widehat{x}(\theta_i)$ of $x(\theta_i)$
$\tilde{\epsilon}''$	Quality of approximation $\widehat{\Sigma}(\bar{\theta})^{-1}[\widehat{x}(\theta_i) - x]$ of $\Sigma(\bar{\theta})^{-1}[x(\theta_i) - x]$
$B$	$\ \theta_0 - \theta(x)\ _{\bar{\theta}} \leq B$ at start of Algorithm 4
$C$	$\ x - x(\bar{\theta})\ _{\Sigma(\bar{\theta})^{-1}} \leq C$ at start of Algorithm 4
$b$	$\ \theta_i - \theta(x)\ _{\bar{\theta}} \leq b$ at termination of Algorithm 4

Let  $X_0$  be a random starting point drawn from a Boltzmann distribution with parameter  $\theta_0$ . Let  $Y^{(1)}, \dots, Y^{(N)}$  be the end points of  $N$  hit-and-run random walks applied to the Boltzmann distribution with parameter  $\theta_1$  having starting point  $X_0$ , where the directions are drawn from a  $\mathcal{N}(0, \widehat{\Sigma}(\theta_0))$ -distribution, and each walk has length  $\ell$  given by (31). (Note that  $\ell$  depends on  $\epsilon, n, q, \Delta\theta$ , and  $\Delta x$ .) Then, with probability  $1 - p$ , the empirical covariance matrix  $\widehat{\Sigma} \approx \Sigma(\theta_1)$  as defined in (36) satisfies

$$(1 - \epsilon_1)v^\top \widehat{\Sigma} v \leq v^\top \Sigma(\theta_1)v \leq (1 + \epsilon_1)v^\top \widehat{\Sigma} v \quad \forall v \in \mathbb{R}^n, \quad (45)$$

$$(1 - \epsilon_1)v^\top \widehat{\Sigma}^{-1} v \leq v^\top \Sigma(\theta_1)^{-1} v \leq (1 + \epsilon_1)v^\top \widehat{\Sigma}^{-1} v \quad \forall v \in \mathbb{R}^n. \quad (46)$$

*Proof.* We apply Theorem 6 with  $\varepsilon = \frac{1}{4}\epsilon_1 < \epsilon_1$ . Thus, (45) holds. For  $\epsilon_1 \leq \sqrt{13} - 3$ ,

$$\frac{1}{4}\epsilon_1 \leq \frac{\sqrt{(3 + \epsilon_1)^2 + 4\epsilon_1(2 + \epsilon_1)} - (3 + \epsilon_1)}{2(2 + \epsilon_1)},$$

where the right hand side is chosen such that Lemma 20 shows that (46) holds.  $\square$

**6. Short-Step IPM Using the Entropic Barrier** From now on, let the reference inner product  $\langle \cdot, \cdot \rangle$  be the Euclidean dot product. Before we show how the results from the previous sections may be applied to interior point methods, we have to fix some notation. With  $c$  as in (1), define

$$f^{*,\eta}(x) = \eta \langle c, x \rangle + f^*(x),$$

and let  $z(\eta)$  be the minimizer of  $f^{*,\eta}$ , that is,  $g^*(z(\eta)) = -\eta c$ . Moreover, it follows from Lemma 3 that  $H^*(z(\eta)) = [H(g^*(z(\eta)))]^{-1} = \Sigma(-\eta c)^{-1}$ . We collect the various tolerances used in this section in Table 2 for the reader's convenience.

A more detailed description of the algorithm by De Klerk, Glineur and Taylor [6] is given in Algorithm 3. We note that an approximation of  $\Sigma(-\eta_0 c)$  can be obtained by hit-and-run sampling using the algorithm by Kalai and Vempala [8]. This algorithm also generates an  $X_0$  following the Boltzmann distribution with parameter  $-\eta_0 c$ .

The assumptions in De Klerk, Glineur and Taylor [6] include that one can find an estimate  $\widehat{\Sigma}(-\eta_{k+1}c)$  of  $\Sigma(-\eta_{k+1}c)$  such that

$$(1 - \epsilon)y^\top \widehat{\Sigma}(-\eta_{k+1}c)y \leq y^\top \Sigma(-\eta_{k+1}c)y \leq (1 + \epsilon)y^\top \widehat{\Sigma}(-\eta_{k+1}c)y \quad \forall y \in \mathbb{R}^n, \quad (47)$$

$$(1 - \epsilon)y^\top \widehat{\Sigma}(-\eta_{k+1}c)^{-1}y \leq y^\top \Sigma(-\eta_{k+1}c)^{-1}y \leq (1 + \epsilon)y^\top \widehat{\Sigma}(-\eta_{k+1}c)^{-1}y \quad \forall y \in \mathbb{R}^n. \quad (48)$$

---

**Algorithm 3** Short-step IPM using the entropic barrier from De Klerk, Glineur and Taylor [6]

---

**Input:** Tolerances  $\tilde{\epsilon}, \bar{\epsilon} > 0$  and  $\epsilon \in (0, \sqrt{13} - 3)$ ; entropic barrier parameter  $\vartheta \leq n + o(n)$ ; objective  $c \in \mathbb{R}^n$ ; radius  $r > 0$  of Euclidean ball contained in  $K$ ; radius  $R > 0$  of Euclidean ball containing  $K$ ; central path proximity parameter  $\delta > 0$ ; an  $x_0 \in K$  and  $\eta_0 > 0$  such that  $\|x_0 - z(\eta_0)\|_{z(\eta_0)}^* \leq \frac{1}{2}\delta$ ; approximation  $\widehat{\Sigma}(-\eta_0 c)$  of  $\Sigma(-\eta_0 c)$  satisfying (47) and (48);  $X_0 \in K$  drawn randomly from a Boltzmann distribution with parameter  $-\eta_0 c$ ; growth rate  $\beta > 0$ .

**Output:**  $x_k$  such that  $\langle c, x_k \rangle - \min_{x \in K} \langle c, x \rangle \leq \bar{\epsilon}$  at terminal iteration  $k$ .

- 1:  $\widehat{\epsilon} \leftarrow \tilde{\epsilon} \sqrt{\frac{1+\epsilon}{1-\epsilon}} + \sqrt{\frac{2\epsilon}{1-\epsilon}}$
- 2:  $d \leftarrow \left( \frac{1}{2}\delta + \beta + \frac{3\beta^2}{(1-\beta)^3} \right) / \left( 1 - \left( \beta + \frac{3\beta^2}{(1-\beta)^3} \right) \right)$
- 3:  $\gamma \leftarrow \frac{2(1-d)^4 - \widehat{\epsilon}(1+(1-d)^4)}{(1-\widehat{\epsilon})(1-d)^2((1-d)^4+1)}$
- 4:  $m \leftarrow \left\lceil \log \left( \frac{\vartheta(1+\delta/2)}{\eta_0 \widehat{\epsilon}} \right) / \log \left( 1 + \frac{\beta}{\sqrt{\vartheta}} \right) \right\rceil$
- 5:  $N \leftarrow \left\lceil \frac{15680n^2 m}{p\epsilon^2} + 2 \right\rceil$
- 6:  $q \leftarrow \min \left\{ \frac{p}{mN}, \frac{N-2}{N} \frac{p\epsilon^2}{1599360mn^2 R^4} \frac{1}{4096} \left( \frac{r}{n+1} \right)^4 \left( \frac{\bar{\epsilon}}{4R\vartheta(1+\delta/2)(1+\beta/\sqrt{\vartheta})} \right)^{8\sqrt{\vartheta}+4} \right\}$
- 7:  $k \leftarrow 0$
- 8: **while**  $\vartheta(1 + \frac{1}{2}\delta)/\eta_k > \bar{\epsilon}$  **do**
- 9:     Generate hit-and-run samples  $Y^{(1)}, \dots, Y^{(N)}$ , and  $X_{k+1}$  from the Boltzmann distribution with parameter  $-\eta_{k+1}c$ , starting from  $X_k$ , drawing directions from  $\mathcal{N}(0, \widehat{\Sigma}(-\eta_k c))$ , using  $\ell$  steps as in (31) with bounds on  $\Delta x$  and  $\Delta \theta$  given in (63) and (64)
- 10:     Compute the empirical covariance  $\widehat{\Sigma}(-\eta_{k+1}c)$  based on  $Y^{(1)}, \dots, Y^{(N)}$  as in (36), which approximates  $\Sigma(-\eta_{k+1}c)$  as in (47) and (48).
- 11:     Find approximation  $\widehat{\theta}(x_k)$  of  $\theta(x_k)$  satisfying (49) using  $X_{k+1}$
- 12:      $x_{k+1} \leftarrow x_k - \gamma \widehat{\Sigma}(-\eta_k c)[\eta_{k+1}c + \widehat{\theta}(x_k)]$   $\triangleright x_{k+1} \approx z(\eta_{k+1})$
- 13:      $\eta_{k+1} \leftarrow (1 + \frac{\beta}{\sqrt{\vartheta}})\eta_k$
- 14:      $k \leftarrow k + 1$
- 15: **end while**
- 16: **return**  $x_k$

---

Note that the number of samples required to find such a  $\widehat{\Sigma}(-\eta_{k+1}c)$  is given by Corollary 3. Moreover, at some point  $x_k \in \text{int } K$  we want to find an approximate gradient  $\widehat{g}_{z(\eta_{k+1})}^{*, \eta_{k+1}}(x_k)$  of  $f^{*, \eta_{k+1}}$  in the sense that

$$\left\| \widehat{g}_{z(\eta_{k+1})}^{*, \eta_{k+1}}(x_k) - g_{z(\eta_{k+1})}^{*, \eta_{k+1}}(x_k) \right\|_{z(\eta_{k+1})}^* \leq \tilde{\epsilon} \left\| g_{z(\eta_{k+1})}^{*, \eta_{k+1}}(x_k) \right\|_{z(\eta_{k+1})}^*,$$

for some  $\tilde{\epsilon} > 0$ . Equivalently, we are looking for an approximation  $\widehat{\theta}(x_k)$  of  $\theta(x_k)$  such that

$$\left\langle \widehat{\theta}(x_k) - \theta(x_k), H^*(z(\eta_{k+1}))^{-1}[\widehat{\theta}(x_k) - \theta(x_k)] \right\rangle \leq (\tilde{\epsilon})^2 \left\langle \eta_{k+1}c + \theta(x_k), H^*(z(\eta_{k+1}))^{-1}[\eta_{k+1}c + \theta(x_k)] \right\rangle.$$

Since  $\theta(z(\eta_{k+1})) = -\eta_{k+1}c$ , it will suffice to find a  $\widehat{\theta}(x_k)$  such that

$$\left\| \widehat{\theta}(x_k) - \theta(x_k) \right\|_{\Sigma(-\eta_{k+1}c)} \leq \tilde{\epsilon} \left\| \theta(z(\eta_{k+1})) - \theta(x_k) \right\|_{\Sigma(-\eta_{k+1}c)}. \quad (49)$$

**6.1. Gradient Approximation** One immediate problem with finding a  $\widehat{\theta}(x_k)$  satisfying (49) is that the right hand side of (49) is unknown: it depends on  $\theta(x_k)$ , the very vector we are trying

to approximate. Hence, we will first present a gradient descent algorithm to find a  $\widehat{\theta}(x_k)$  such that  $\|\widehat{\theta}(x_k) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}e)} \leq b$  for any given  $b > 0$ . In the next section, we will then lower bound the right hand side of (49) by known values.

To approximate  $\theta(x_k)$  in Algorithm 3, we will use the approach proposed by Abernethy and Hazan [1], which was also discussed in Section 2.2. Let  $x \in \text{int } K$ , and consider the unconstrained minimization problem

$$\min_{\theta \in \mathbb{R}^n} \Psi(\theta), \quad \text{where} \quad \Psi(\theta) = f(\theta) - \langle \theta, x \rangle.$$

Since the Fréchet derivative of  $\Psi$  is  $D\Psi(\theta) = \mathbb{E}_\theta[X] - x$ , the gradient of  $\Psi$  is zero at the  $\theta \in \mathbb{R}^n$  such that  $g(\theta) = \mathbb{E}_\theta[X] = x$ , which is by definition  $\theta(x)$ . Moreover,  $D^2\Psi(\theta) = \Sigma(\theta)$ , and thus  $\Psi$  is strictly convex, meaning we can use unconstrained minimization techniques to approximate  $\theta(x)$ . Note that because the Hessians of  $f$  and  $\Psi$  are the same,  $\Psi$  is self-concordant. Although the gradient and Hessian of  $\Psi$  are not readily available, they can be estimated through sampling. We can then again apply the inexact gradient descent results from De Klerk, Glineur and Taylor [6]. This is formalized in Algorithm 4. To distinguish parameters in Algorithm 4 from those with a similar function in Algorithm 3, we add a prime to those parameters in Algorithm 4 (e.g. we use  $\epsilon'$  in stead of  $\epsilon$ ). Among other things, Algorithm 4 depends on two vectors  $\theta_0$  and  $\bar{\theta}$ . Intuitively,  $\theta_0$  is our initial estimate for  $\theta(x)$ , while  $\bar{\theta}$  is used to fix an inner product. (As we will see in Section 6.3, we cannot take  $\bar{\theta} = \theta_0$  to analyze Algorithm 3.)

The following theorem shows that Algorithm 4 produces the desired result with high probability.

**THEOREM 7.** *Let  $p' \in (0, 1)$ . With probability at least  $1 - p'$ , Algorithm 4 produces a  $\widehat{\theta}(x)$  such that  $\|\widehat{\theta}(x) - \theta(x)\|_{\bar{\theta}} \leq b$ . The number of required samples hit-and-run steps is  $m'N'\ell'$ , if*

$$\ell' = \left\lceil \frac{1 + \epsilon'}{1 - \epsilon'} \frac{64e^2 n^3 10^{30}}{(1 - 2d')^2} \ln^2 \left( \sqrt{\frac{1 + \epsilon'}{1 - \epsilon'}} \frac{16en\sqrt{n} \exp(-2d')}{(q')^2(1 - d')^2(1 - 2d')} \right) \ln^3 \left( \frac{2 \exp(-2d')}{(q')^2(1 - d')^2} \right) \right\rceil, \quad (50)$$

and  $m'$ ,  $N'$ ,  $\epsilon'$ ,  $q'$ , and  $d'$  are defined in Algorithm 4, where  $m'$ ,  $q'$ , and  $N'$  depend on  $p'$  and  $b$  as in lines 3 -5 of Algorithm 4.

*Proof.* First, we want to show that all iterates  $\theta_i$  lie close to  $\bar{\theta}$  in the sense that  $\|\theta_i - \bar{\theta}\|_{\bar{\theta}} \leq d'$  for all  $i$ . The argument is inductive: if  $\|\theta_i - \bar{\theta}\|_{\bar{\theta}} \leq d'$ , then the remainder of this proof will show that  $\|\theta_{i+1} - \theta(x)\|_{\bar{\theta}} \leq \|\theta_i - \theta(x)\|_{\bar{\theta}}$ , i.e. we are making progress to the minimizer, and therefore

$$\begin{aligned} \|\theta_{i+1} - \bar{\theta}\|_{\bar{\theta}} &\leq \|\theta_{i+1} - \theta(x)\|_{\bar{\theta}} + \|\theta(x) - \theta_0\|_{\bar{\theta}} + \|\theta_0 - \bar{\theta}\|_{\bar{\theta}} \\ &\leq 2\|\theta(x) - \theta_0\|_{\bar{\theta}} + \|\theta_0 - \bar{\theta}\|_{\bar{\theta}} \leq 2B + \|\theta_0 - \bar{\theta}\|_{\bar{\theta}} \leq d'. \end{aligned} \quad (51)$$

We want to invoke Theorem 5 for  $\theta_0 = \bar{\theta}$  and  $\theta_1 = \theta_i$  to show that we can approximate the mean of the Boltzmann distribution with parameter  $\theta_i$ . Before we can do that, we need appropriate bounds on  $\Delta\theta = \|\theta_i - \bar{\theta}\|_{\bar{\theta}}$ ,  $\Delta x = \|x(\theta_i) - x(\bar{\theta})\|_{x(\bar{\theta})}^*$  and  $\lambda_{\min}(\Sigma(\theta_i))$ . Note that  $\Delta\theta = \|\theta_i - \bar{\theta}\|_{\bar{\theta}} \leq d' < 1$  by assumption, and Corollary 1 shows

$$\Delta x = \|x(\theta_i) - x(\bar{\theta})\|_{x(\bar{\theta})}^* \leq \frac{\|\theta_i - \bar{\theta}\|_{\bar{\theta}}}{1 - \|\theta_i - \bar{\theta}\|_{\bar{\theta}}} \leq \frac{d'}{1 - d'} < 1.$$

By Theorem 2, for all  $i$ ,

$$\|\theta_i\| \leq \|\bar{\theta}\| + \|\theta_i - \bar{\theta}\| \leq \|\bar{\theta}\| + \frac{\|\theta_i - \bar{\theta}\|_{\bar{\theta}}}{\sqrt{\lambda_{\min}(\Sigma(\bar{\theta}))}} \leq \|\bar{\theta}\| + \frac{8d'(n+1)}{r} \max\{1, (4R\|\bar{\theta}\|)^{2\sqrt{\bar{\theta}}+1}\}, \quad (52)$$

---

**Algorithm 4** Approximation routine for  $\theta(x)$

---

**Input:**  $x \in \text{int } K$ ;  $0 < b < B$ ;  $\theta_0, \bar{\theta} \in \mathbb{R}^n$  such that  $\|\theta_0 - \theta(x)\|_{\bar{\theta}} \leq B$ ;  $d' < \frac{1}{3}$  such that  $2B + \|\theta_0 - \bar{\theta}\|_{\bar{\theta}} \leq d'$ ; approximation  $\widehat{\Sigma}(\bar{\theta})^{-1}$  of  $\Sigma(\bar{\theta})^{-1}$  and  $\epsilon' \in (0, 1)$  satisfying (47) and (48);  $\bar{X} \in K$  drawn randomly from a Boltzmann distribution with parameter  $\bar{\theta}$ ; radius  $r > 0$  of Euclidean ball contained in  $K$ ; radius  $R > 0$  of Euclidean ball containing  $K$ ; relative gradient error  $\tilde{\epsilon}' > 0$  such that  $\tilde{\epsilon}' \sqrt{\frac{1+\epsilon'}{1-\epsilon'}} + \sqrt{\frac{2\epsilon'}{1-\epsilon'}} < \frac{2(1-d')^4}{1+(1-d')^4}$ ;  $C \in (0, 1)$  such that  $\|x - x(\bar{\theta})\|_{\Sigma(\bar{\theta})^{-1}} \leq C$ ; failure probability  $p' \in (0, 1)$ .

**Output:**  $\theta_i$  such that  $\|\theta_i - \theta(x)\|_{\bar{\theta}} \leq b$  at terminal iteration  $i$ .

```

1:  $\tilde{\epsilon}' \leftarrow \tilde{\epsilon}' \sqrt{\frac{1+\epsilon'}{1-\epsilon'}} + \sqrt{\frac{2\epsilon'}{1-\epsilon'}}$ 
2:  $\gamma' \leftarrow \frac{2(1-d')^4 - \tilde{\epsilon}'(1+(1-d')^4)}{(1-\tilde{\epsilon}')(1-d')^2((1-d')^4+1)}$ 
3:  $m' \leftarrow \left\lceil \log(b/B) / \log\left(\frac{1-(1-d')^4}{1+(1-d')^4} + \tilde{\epsilon}'\right) \right\rceil$ 
4:  $N' \leftarrow \left\lceil 2n \frac{m'}{p'} \left( \frac{1+b(1-C)}{b(1-C)^2} \left( 1 + \frac{1+\tilde{\epsilon}'}{\tilde{\epsilon}'} \sqrt{\frac{1+\epsilon'}{1-\epsilon'}} \frac{1-2d'}{1-3d'} \right)^2 \right) \right\rceil$ 
5:  $q' \leftarrow \frac{p'}{204m'nR^2} \left( \frac{b(1-C)^2}{1+b(1-C)} \frac{\tilde{\epsilon}'\sqrt{1-\epsilon'}}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \frac{1-3d'}{1-2d'} \right)^2 \times$ 
    $\frac{1}{64} \left( \frac{r}{n+1} \right)^2 \min \left\{ 1, \left( 4R\|\bar{\theta}\| + \frac{32Rd'(n+1)}{r} \max\{1, (4R\|\bar{\theta}\|)^{2\sqrt{\bar{\theta}+1}}\} \right)^{-4\sqrt{\bar{\theta}-2}} \right\}$ 
6:  $i \leftarrow 0$ 
7: while  $B \left( \frac{1-(1-d')^4}{1+(1-d')^4} + \tilde{\epsilon}' \right)^i > b$  do
8:   Generate hit-and-run samples  $Y^{(1)}, \dots, Y^{(N')}$  from the Boltzmann distribution with parameter  $\theta_i$ , starting from  $\bar{X}$ , drawing directions from  $\mathcal{N}(0, \widehat{\Sigma}(\bar{\theta}))$ , using  $\ell'$  steps as in (50)
9:   Compute sample mean:  $\widehat{x}(\theta_i) \leftarrow \frac{1}{N'} \sum_{j=1}^{N'} Y^{(j)}$ 
10:  if  $\|\widehat{x}(\theta_i) - x\|_{\widehat{\Sigma}(\bar{\theta})^{-1}} \leq \frac{b(1-C)^2}{1+b(1-C)} \frac{1+\tilde{\epsilon}'}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}}$  then
11:    return  $\theta_i$   $\triangleright \|\theta_i - \theta(x)\|_{\bar{\theta}} \leq b$ 
12:  else
13:     $\theta_{i+1} \leftarrow \theta_i - \gamma' \widehat{\Sigma}(\bar{\theta})^{-1} (\widehat{x}(\theta_i) - x)$ 
14:     $i \leftarrow i + 1$ 
15:  end if
16: end while
17: return  $\theta_i$ 

```

---

and therefore another application of Theorem 2 yields

$$\begin{aligned} \lambda_{\min}(\Sigma(\theta_i)) &\geq \frac{1}{64} \left( \frac{r}{n+1} \right)^2 \min \left\{ 1, (4R\|\theta_i\|)^{-4\sqrt{\bar{\theta}-2}} \right\} \\ &\geq \frac{1}{64} \left( \frac{r}{n+1} \right)^2 \min \left\{ 1, \left( 4R\|\bar{\theta}\| + \frac{32Rd'(n+1)}{r} \max\{1, (4R\|\bar{\theta}\|)^{2\sqrt{\bar{\theta}+1}}\} \right)^{-4\sqrt{\bar{\theta}-2}} \right\}, \end{aligned}$$

where the second inequality is from (52). We will now apply Theorem 5 with  $p = p'/m'$  and

$$\alpha = \left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{\tilde{\epsilon}'\sqrt{1-\epsilon'}}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right) \left( \frac{1-3d'}{1-2d'} \right).$$

Note that the values of  $N'$  and  $q'$  in Algorithm 4 are chosen such that Theorem 5 now shows that in each iteration  $i$ , it holds with probability at least  $1 - p'/m'$  that the sample mean  $\widehat{x}(\theta_i)$  satisfies

$$\|\widehat{x}(\theta_i) - x(\theta_i)\|_{x(\theta_i)}^* \leq \left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{\tilde{\epsilon}'\sqrt{1-\epsilon'}}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right) \left( \frac{1-3d'}{1-2d'} \right), \quad (53)$$

where we recall that Lemma 3 demonstrates  $\|\widehat{x}(\theta_i) - x(\theta_i)\|_{\Sigma(\theta_i)-1} = \|\widehat{x}(\theta_i) - x(\theta_i)\|_{x(\theta_i)}^*$ . We want to show that it follows from (53) that the  $\|\cdot\|_{x(\bar{\theta})}^*$ -norm of  $\widehat{x}(\theta_i) - x(\theta_i)$  is also small. Since  $\|\bar{\theta} - \theta_i\|_{\bar{\theta}} \leq d' < \frac{1}{3}$  by (51), the second inequality in (4) shows

$$\|\bar{\theta} - \theta_i\|_{\theta_i} \leq \frac{\|\bar{\theta} - \theta_i\|_{\bar{\theta}}}{1 - \|\bar{\theta} - \theta_i\|_{\bar{\theta}}} \leq \frac{d'}{1-d'} < \frac{1}{2}.$$

Therefore, the first inequality in (21) implies that

$$\|x(\bar{\theta}) - x(\theta_i)\|_{x(\theta_i)}^* \leq \frac{\|\bar{\theta} - \theta_i\|_{\theta_i}}{1 - \|\bar{\theta} - \theta_i\|_{\theta_i}} \leq \frac{\frac{d'}{1-d'}}{1 - \frac{d'}{1-d'}} = \frac{d'}{1-2d'} < 1. \quad (54)$$

Consequently, we can now use the second inequality in (5), along with (53) and (54) to see that

$$\begin{aligned} \|\widehat{x}(\theta_i) - x(\theta_i)\|_{x(\bar{\theta})}^* &\leq \frac{\|\widehat{x}(\theta_i) - x(\theta_i)\|_{x(\theta_i)}^*}{1 - \|x(\bar{\theta}) - x(\theta_i)\|_{x(\theta_i)}^*} \\ &\leq \frac{\left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{\tilde{\epsilon}'\sqrt{1-\epsilon'}}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right) \left( \frac{1-3d'}{1-2d'} \right)}{1 - \frac{d'}{1-2d'}} \\ &= \left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{\tilde{\epsilon}'\sqrt{1-\epsilon'}}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right). \end{aligned} \quad (55)$$

Algorithm 4 distinguishes the following two cases in lines 10 and 12:

- (a)  $\|\widehat{x}(\theta_i) - x\|_{\widehat{\Sigma}(\bar{\theta})-1} \leq \left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{1+\tilde{\epsilon}'}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right)$ ;
- (b)  $\|\widehat{x}(\theta_i) - x\|_{\widehat{\Sigma}(\bar{\theta})-1} > \left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{1+\tilde{\epsilon}'}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right)$ .

In Case (a), we have

$$\begin{aligned} \|x(\theta_i) - x\|_{\Sigma(\bar{\theta})-1} &\leq \|\widehat{x}(\theta_i) - x(\theta_i)\|_{\Sigma(\bar{\theta})-1} + \|\widehat{x}(\theta_i) - x\|_{\Sigma(\bar{\theta})-1} && \text{(triangle inequality)} \\ &\leq \|\widehat{x}(\theta_i) - x(\theta_i)\|_{\Sigma(\bar{\theta})-1} + \sqrt{1+\epsilon'}\|\widehat{x}(\theta_i) - x\|_{\widehat{\Sigma}(\bar{\theta})-1} && \text{(by (48))} \\ &\leq \frac{b(1-C)^2}{1+b(1-C)}. && \text{(by (55) and Case (a))} \end{aligned}$$

Since  $\|x - x(\bar{\theta})\|_{x(\bar{\theta})}^* \leq C < 1$  by assumption, it follows that

$$\|x - x(\theta_i)\|_{x(\bar{\theta})}^* + \|x - x(\bar{\theta})\|_{x(\bar{\theta})}^* \leq \frac{b(1-C)^2}{1+b(1-C)} + C = \frac{C+b(1-C)}{1+b(1-C)} < 1.$$

Thus, we may apply Lemma 13 to see that

$$\begin{aligned} \|\theta_i - \theta(x)\|_{\bar{\theta}} &\leq \frac{1}{1 - \|x - x(\bar{\theta})\|_{x(\bar{\theta})}^*} \frac{\|x - x(\theta_i)\|_{x(\bar{\theta})}^*}{1 - \|x - x(\theta_i)\|_{x(\bar{\theta})}^* - \|x - x(\bar{\theta})\|_{x(\bar{\theta})}^*} \\ &\leq \frac{1}{1-C} \frac{\frac{b(1-C)^2}{1+b(1-C)}}{1 - \frac{b(1-C)^2}{1+b(1-C)} - C} = b. \end{aligned}$$

In conclusion, the stopping criterion for Algorithm 4 is met in Case (a). In Case (b), we can expand (55) to

$$\begin{aligned}
 & \|\widehat{x}(\theta_i) - x(\theta_i)\|_{x(\bar{\theta})}^* \\
 & \leq \left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{\tilde{\epsilon}'\sqrt{1-\epsilon'}}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right) \\
 & = \frac{\tilde{\epsilon}'}{1+\tilde{\epsilon}'}\sqrt{1-\epsilon'} \left( \frac{b(1-C)^2}{1+b(1-C)} \right) \left( \frac{1+\tilde{\epsilon}'}{(1+\tilde{\epsilon}')\sqrt{1+\epsilon'}+\tilde{\epsilon}'\sqrt{1-\epsilon'}} \right) \\
 & < \frac{\tilde{\epsilon}'}{1+\frac{\tilde{\epsilon}'}{\epsilon'}}\sqrt{1-\epsilon'}\|\widehat{x}(\theta_i) - x\|_{\Sigma(\bar{\theta})^{-1}} && \text{(Case (b))} \\
 & \leq \frac{\tilde{\epsilon}'}{1+\frac{\tilde{\epsilon}'}{\epsilon'}}\|\widehat{x}(\theta_i) - x\|_{x(\bar{\theta})}^* && \text{(by (48))} \\
 & \leq \frac{\tilde{\epsilon}'}{1+\tilde{\epsilon}'} \left( \|x(\theta_i) - x\|_{x(\bar{\theta})}^* + \|\widehat{x}(\theta_i) - x(\theta_i)\|_{x(\bar{\theta})}^* \right) && \text{(triangle inequality),} \quad (56)
 \end{aligned}$$

where we have used  $\Sigma(\bar{\theta})^{-1} = H^*(x(\bar{\theta}))$ . Again using  $\Sigma(\bar{\theta})^{-1} = H^*(x(\bar{\theta}))$ , the outermost inequality in (56) can be rewritten as

$$\|\widehat{x}(\theta_i) - x(\theta_i)\|_{\Sigma(\bar{\theta})^{-1}} \leq \tilde{\epsilon}' \|x(\theta_i) - x\|_{\Sigma(\bar{\theta})^{-1}}.$$

Since we also have an approximation  $\widehat{\Sigma}(\bar{\theta}) \approx \Sigma(\bar{\theta})$  satisfying (47) and (48) for some  $\epsilon' > 0$ , one can show in a manner analogous to Corollary 7.5 in De Klerk, Glineur and Taylor [6] that

$$\left\| \widehat{\Sigma}(\bar{\theta})^{-1}(\widehat{x}(\theta_i) - x) - \Sigma(\bar{\theta})^{-1}(x(\theta_i) - x) \right\|_{\Sigma(\bar{\theta})} \leq \tilde{\epsilon}' \|\Sigma(\bar{\theta})^{-1}(x(\theta_i) - x)\|_{\Sigma(\bar{\theta})},$$

where  $\tilde{\epsilon}' = \tilde{\epsilon}'\sqrt{\frac{1+\epsilon'}{1-\epsilon'}} + \sqrt{\frac{2\epsilon'}{1-\epsilon'}}$ . Hence, we have an approximate gradient with respect to  $\langle \cdot, \cdot \rangle_{\bar{\theta}}$ , and we can apply approximate gradient descent. Recall that  $\Psi$  is self-concordant with minimizer  $\theta(x)$ . Note that Theorem 3.6 in [6] shows that the spectrum of  $H_{\bar{\theta}}(\theta_i)$  is contained in  $[(1-d')^2, \frac{1}{(1-d')^2}]$  for all  $i$ . Theorem 5.3 in [6] shows that if  $\tilde{\epsilon}' \leq \frac{2(1-d')^4}{1+(1-d')^4}$  and  $\gamma'$  takes the value it is given in Algorithm 4, then it holds for  $\theta_{i+1} = \theta_i - \gamma'\widehat{\Sigma}(\bar{\theta})^{-1}(\widehat{x}(\theta_i) - x)$  that

$$\|\theta_{i+1} - \theta(x)\|_{\bar{\theta}} \leq \left( \frac{1 - (1-d')^4}{1 + (1-d')^4} + \tilde{\epsilon}' \right) \|\theta_i - \theta(x)\|_{\bar{\theta}}.$$

Hence, if our starting point  $\theta_0$  satisfies  $\|\theta_0 - \theta(x)\|_{\bar{\theta}} \leq B$ , then after

$$m' = \left\lceil \frac{\log(b/B)}{\log\left(\frac{1-(1-d')^4}{1+(1-d')^4} + \tilde{\epsilon}'\right)} \right\rceil$$

iterations, we have  $\|\theta_{m'} - \theta(x)\|_{\bar{\theta}} \leq b$ . Recall that the probability of failure, where not finding a good enough approximation of  $x(\theta_i) = \mathbb{E}_{\theta_i}[X]$  constitutes a failure of the algorithm, is at most  $p'/m'$  in each iteration. By the union bound, the probability of success after  $m'$  iterations is therefore at least  $1 - p'$ .  $\square$

**6.2. Parameter Values** The algorithm in the previous section allows us to approximate  $\theta(x_k)$  up to any desired accuracy  $b > 0$ . However, the requirement (49) is stated in terms of the unknown expression  $\|\theta(z(\eta_{k+1})) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)}$ . The purpose of this section is to provide a lower bound on this expression, which may then be used in the role of  $b$  in Algorithm 4.



The triangle inequality shows that

$$\begin{aligned} \|\theta(z(\eta_{k+1})) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)} &\geq \|\theta(z(\eta_{k+1})) - \theta(z(\eta_k))\|_{\Sigma(-\eta_{k+1}c)} - \|\theta(z(\eta_k)) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)} \\ &\geq (\eta_{k+1} - \eta_k) \|c\|_{\Sigma(-\eta_{k+1}c)} - \frac{\|\theta(z(\eta_k)) - \theta(x_k)\|_{\Sigma(-\eta_k c)}}{1 - (\eta_{k+1} - \eta_k) \|c\|_{\Sigma(-\eta_k c)}}, \end{aligned} \quad (57)$$

where the second inequality uses (4). It follows from Corollary 1 that if  $\|x_k - z(\eta_k)\|_{z(\eta_k)}^* < 1$ , we can further bound (57) by

$$\|\theta(z(\eta_{k+1})) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)} \geq (\eta_{k+1} - \eta_k) \|c\|_{-\eta_{k+1}c} - \frac{\frac{\|x_k - z(\eta_k)\|_{z(\eta_k)}^*}{1 - \|x_k - z(\eta_k)\|_{z(\eta_k)}^*}}{1 - (\eta_{k+1} - \eta_k) \|c\|_{-\eta_k c}}. \quad (58)$$

For our main algorithm, we can configure how closely we want to follow the central path, i.e. how small  $\|x_k - z(\eta_k)\|_{z(\eta_k)}^*$  should be in each iteration  $k$ . Moreover, with Lemma 12 and (12), we have lower and upper bounds on  $\|c\|_{-\eta_{k+1}c}$  and  $\|c\|_{-\eta_k c}$ , respectively. The only remaining ingredient to finding an explicit lower bound on  $\|\theta(z(\eta_{k+1})) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)}$  is therefore to set explicit parameter values.

The parameter values in Algorithm 3 we still have to determine are the approximation error  $\hat{\epsilon}$ , the growth factor  $\beta$  and the central path proximity  $\delta$ . For most short-step interior point methods, this is a matter of keeping  $\beta$  relatively small and  $\delta$  large enough to guarantee that after one takes a step from the current iterate, the resulting point will lie within distance  $\frac{1}{2}\delta$  of the next minimizer on the central path. More formally, suppose  $\|x_k - z(\eta_k)\|_{z(\eta_k)}^* \leq \frac{1}{2}\delta$ , and we want to apply an approximate gradient descent step to estimate  $z(\eta_{k+1})$ . First note that, as in the proof of Theorem 7.6 in De Klerk, Glineur and Taylor [6],

$$\|z(\eta_{k+1}) - z(\eta_k)\|_{z(\eta_k)}^* \leq \beta + \frac{3\beta^2}{(1-\beta)^3}, \quad (59)$$

and therefore, by the second inequality in (5),

$$\|x_k - z(\eta_{k+1})\|_{z(\eta_{k+1})}^* \leq \frac{\|x_k - z(\eta_k)\|_{z(\eta_k)}^* + \|z(\eta_{k+1}) - z(\eta_k)\|_{z(\eta_k)}^*}{1 - \|z(\eta_{k+1}) - z(\eta_k)\|_{z(\eta_k)}^*} \leq \frac{\frac{1}{2}\delta + \beta + \frac{3\beta^2}{(1-\beta)^3}}{1 - \left(\beta + \frac{3\beta^2}{(1-\beta)^3}\right)}. \quad (60)$$

Supposing we have a  $\hat{g}^{*,\eta_{k+1}}(x_k)$  and  $\hat{\Sigma}(-\eta_{k+1}c)$  that satisfy

$$\left\| \hat{\Sigma}(-\eta_{k+1}c) \hat{g}^{*,\eta_{k+1}}(x_k) - g_{z(\eta_{k+1})}^{*,\eta_{k+1}}(x_k) \right\|_{z(\eta_{k+1})}^* \leq \hat{\epsilon} \left\| g_{z(\eta_{k+1})}^{*,\eta_{k+1}}(x_k) \right\|_{z(\eta_{k+1})}^*,$$

we can now apply Corollary 6.2 in [6]. After a damped gradient descent step, we have

$$\|x_{k+1} - z(\eta_{k+1})\|_{z(\eta_{k+1})}^* \leq \left( \frac{1-\kappa}{1+\kappa} + \hat{\epsilon} \right) \|x_k - z(\eta_{k+1})\|_{z(\eta_{k+1})}^* \leq \left( \frac{1-\kappa}{1+\kappa} + \hat{\epsilon} \right) \frac{\frac{1}{2}\delta + \beta + \frac{3\beta^2}{(1-\beta)^3}}{1 - \left(\beta + \frac{3\beta^2}{(1-\beta)^3}\right)}, \quad (61)$$

where

$$\kappa = \left( 1 - \frac{\delta + 2\beta + \frac{6\beta^2}{(1-\beta)^3}}{1 - \left(\beta + \frac{3\beta^2}{(1-\beta)^3}\right)} \right)^4.$$

Thus, (61) should be at most  $\frac{1}{2}\delta$  to maintain a distance of at most  $\frac{1}{2}\delta$  to the central path.

However, in our algorithm, we are also looking for a  $\widehat{\theta}(x_k)$  that satisfies (49). The right hand side of this condition depends on the unknown quantity  $\|\theta(z(\eta_{k+1})) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)}$ , and therefore we would like to bound it from below by some known, fixed quantity  $Q > 0$ . If we can find a  $\widehat{\theta}(x_k)$  such that  $\|\widehat{\theta}(x_k) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)} \leq \widetilde{\epsilon}Q$ , then (49) holds as well. Recall that we already made some progress in bounding  $\|\theta(z(\eta_{k+1})) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)}$  from below in (58). To make further developing this bound easier, one would like to increase  $\beta$  as much as possible (such that  $\eta_{k+1}$  and  $\eta_k$  are further from each other), and decrease  $\delta$  (such that  $x_k$  lies closer to  $z(\eta_k)$ ). This desire thus conflicts with our wish to make (61) smaller than or equal to  $\frac{1}{2}\delta$ .

The following lemma gives parameter values where both demands are met. These values depend on a number  $S$  with  $1/S = O(\frac{n\sqrt{\vartheta}}{r\eta_0}) = O(\frac{n\sqrt{n}}{r\eta_0})$ . We first show that (61) holds, which is a purely algebraic property.

LEMMA 21. *Let  $r, R > 0$  such that  $R \geq r$ . Let  $n, \vartheta \geq 1$ ,  $\eta_0 > 0$ , and  $\bar{\epsilon} \geq 0$ . Define the parameter*

$$S := \frac{1}{4} \frac{1}{\sqrt{\vartheta} + \frac{7}{176}} \frac{r\eta_0}{2(n+1) + r\eta_0},$$

and set

$$\widehat{\epsilon} = \frac{S}{32 + 7S}, \quad \beta = \min \left\{ \frac{21S}{640 + 332S}, \frac{7}{176} \frac{32 + 5S}{32 + 7S}, \frac{\bar{\epsilon}}{4R\sqrt{\vartheta}(1 + \frac{7}{1408})} \right\}, \quad \delta = S\beta.$$

Then,

$$\left( \frac{1 - \left(1 - \frac{\delta + 2\bar{\beta}}{1 - \bar{\beta}}\right)^4}{1 + \left(1 - \frac{\delta + 2\bar{\beta}}{1 - \bar{\beta}}\right)^4} + \widehat{\epsilon} \right) \frac{\bar{\beta} + \frac{1}{2}\delta}{1 - \bar{\beta}} \leq \frac{1}{2}\delta,$$

where  $\bar{\beta} := \beta + \frac{3\beta^2}{(1-\beta)^3}$ .

The proof is elementary and therefore omitted.

Next, we show that the right hand side (58) can be bounded away from zero for the parameter values in Lemma 21.

LEMMA 22. *Let  $K \subseteq \mathbb{R}^n$  be a convex body that contains a Euclidean ball of radius  $r > 0$  and is contained in a Euclidean ball of radius  $R \geq r$ . Define  $f$  as the log partition function  $f(\theta) = \ln \int_K e^{\langle \theta, x \rangle} dx$ , where  $\langle \cdot, \cdot \rangle$  is the Euclidean inner product, and denote its Hessian by  $\Sigma(\theta)$ . Let  $f^*$  be the entropic barrier for  $K$  with complexity parameter  $\vartheta \geq 1$ , and let  $\eta_0, \bar{\epsilon} > 0$ . Let the values of  $S$ ,  $\widehat{\epsilon}$ ,  $\beta$  and  $\delta$  be as in Lemma 21. For all positive integers  $k \leq \left\lceil \log\left(\frac{\vartheta(1+\delta/2)}{\eta_0\bar{\epsilon}}\right) / \log(1 + \beta/\sqrt{\vartheta}) \right\rceil$ , define  $\eta_k = (1 + \beta/\sqrt{\vartheta})^k \eta_0$ . Then, for all  $k \leq \left\lceil \log\left(\frac{\vartheta(1+\delta/2)}{\eta_0\bar{\epsilon}}\right) / \log(1 + \beta/\sqrt{\vartheta}) \right\rceil - 1$ ,*

$$(\eta_{k+1} - \eta_k) \|c\|_{-\eta_{k+1}c} - \frac{\frac{\delta/2}{1-\delta/2}}{1 - (\eta_{k+1} - \eta_k) \|c\|_{-\eta_k c}} \geq \frac{1}{2} \frac{\beta}{\sqrt{\vartheta} + \beta} \frac{r\eta_{k+1}}{2(n+1) + r\eta_{k+1}}.$$

*Proof.* Let  $k \leq \left\lceil \log\left(\frac{\vartheta(1+\delta/2)}{\eta_0\bar{\epsilon}}\right) / \log(1 + \beta/\sqrt{\vartheta}) \right\rceil - 1$ . We observe that  $\frac{\delta/2}{1-\delta/2} \leq \delta = S\beta$ , where the inequality holds for  $\delta \in [0, 1]$ . Using the definition of  $S$ , we get

$$\frac{\delta/2}{1-\delta/2} \leq \delta = \frac{1}{4} \frac{\beta}{\sqrt{\vartheta} + \frac{7}{176}} \frac{r\eta_0}{2(n+1) + r\eta_0} \leq \frac{1}{4} \frac{\beta}{\sqrt{\vartheta} + \beta} \frac{r\eta_{k+1}}{2(n+1) + r\eta_{k+1}}.$$

Through application of Lemma 12 and (12), the above thus gives us

$$(\eta_{k+1} - \eta_k) \|c\|_{-\eta_{k+1}c} - \frac{\frac{\delta/2}{1-\delta/2}}{1 - (\eta_{k+1} - \eta_k) \|c\|_{-\eta_k c}} \geq \frac{r(\eta_{k+1} - \eta_k)}{2(n+1) + r\eta_{k+1}} - \frac{\frac{1}{4} \frac{\beta}{\sqrt{\vartheta} + \beta} \frac{r\eta_{k+1}}{2(n+1) + r\eta_{k+1}}}{1 - 2R(\eta_{k+1} - \eta_k)}. \quad (62)$$

Since the value of  $k$  implies  $\eta_k \leq \vartheta(1 + \frac{1}{2}\delta)/\bar{\epsilon}$ , and  $\delta = S\beta \leq \frac{1}{4} \frac{7}{176} = \frac{7}{704}$ ,

$$2R(\eta_{k+1} - \eta_k) = \frac{2R\beta}{\sqrt{\vartheta}} \eta_k \leq \frac{2R\beta}{\sqrt{\vartheta}} \frac{\vartheta(1 + \delta/2)}{\bar{\epsilon}} \leq \frac{1}{2},$$

by our choice of  $\beta$ . With this fact, and  $\eta_{k+1} \leq (1 + \frac{\beta}{\sqrt{\vartheta}})\vartheta(1 + \frac{1}{2}\delta)/\bar{\epsilon}$ , we can further bound (62) by

$$\begin{aligned} & (\eta_{k+1} - \eta_k) \|c\|_{-\eta_{k+1}c} - \frac{\frac{\delta/2}{1-\delta/2}}{1 - (\eta_{k+1} - \eta_k) \|c\|_{-\eta_k c}} \\ & \geq \frac{r\eta_{k+1}(1 - \eta_k/\eta_{k+1})}{2(n+1) + r\eta_{k+1}} - \frac{1}{2} \frac{\beta}{\sqrt{\vartheta} + \beta} \frac{r\eta_{k+1}}{2(n+1) + r\eta_{k+1}} = \frac{1}{2} \frac{\beta}{\sqrt{\vartheta} + \beta} \frac{r\eta_{k+1}}{2(n+1) + r\eta_{k+1}}. \end{aligned}$$

□

It is important to note that the value of  $S$  is the same for  $K$  as for  $\alpha K$ , where  $\alpha > 0$ . To see this, we need to show that the product  $r\eta_0$  does not change as we scale  $K$ . Recall that we start Algorithm 3 with an  $x_0$  close to  $z(\eta_0)$ , where  $z(\eta_0)$  is the minimizer of  $x \mapsto \eta_0 \langle c, x \rangle + f^*(x)$ . Let  $\bar{f}$  be the log partition function associated with  $\alpha K$ , and  $\bar{f}^*$  the entropic barrier over  $\alpha K$ . We claim that  $\alpha z(\eta_0)$  is the minimizer of  $y \mapsto \frac{\eta_0}{\alpha} \langle c, y \rangle + \bar{f}^*(y)$ .

By definition,  $\bar{f}(\theta) = \ln \int_{\alpha K} e^{\langle \theta, y \rangle} dy = \ln \int_K e^{\langle \theta, \alpha x \rangle} \alpha^n dx$ . It follows that

$$\bar{g}(\theta) = \frac{\int_K \alpha x e^{\langle \theta, \alpha x \rangle} \alpha^n dx}{\int_K e^{\langle \theta, \alpha x \rangle} \alpha^n dx} = \alpha \frac{\int_K x e^{\langle \alpha \theta, x \rangle} dx}{\int_K e^{\langle \alpha \theta, x \rangle} dx} = \alpha g(\alpha \theta).$$

Using  $g(-\eta_0 c) = z(\eta_0)$ , it is easily seen that  $\bar{g}(-\frac{\eta_0}{\alpha} c) = \alpha g(-\eta_0 c) = \alpha z(\eta_0)$ , which shows that  $\alpha z(\eta_0)$  is the minimizer of  $y \mapsto \frac{\eta_0}{\alpha} \langle c, y \rangle + \bar{f}^*(y)$ . In other words,  $\eta_0$  is scaled by a factor  $\frac{1}{\alpha}$  if  $K$  is scaled by a factor  $\alpha$ . Since  $\alpha K$  contains a Euclidean ball of radius  $\alpha r$ , the term  $r\eta_0$  is invariant under scaling of  $K$ , and therefore  $S$  is scaling-invariant as well.

**6.3. Complexity Analysis** With the parameter values in Lemma 21, we can show that Algorithm 3 follows the central path with high probability.

LEMMA 23. *Let  $p \in (0, 1)$  and  $\eta_0 > 0$ , and the values of  $S$ ,  $\hat{\epsilon}$ ,  $\beta$  and  $\delta$  as in Lemma 21. Let  $\tilde{\epsilon} > 0$  and  $\epsilon \in (0, \sqrt{13} - 3)$  be such that  $\tilde{\epsilon} \sqrt{\frac{1+\epsilon}{1-\epsilon}} + \sqrt{\frac{2\epsilon}{1-\epsilon}} \leq \hat{\epsilon}$ . Set  $B$ ,  $b$ ,  $d'$  and  $C$  as follows:*

$$\begin{aligned} B &= \frac{\delta}{2 - \delta} \frac{1 - \beta(4 - \beta(3 - \beta)(1 - \beta))}{1 - \beta(5 - \beta(3 - \beta)(1 - 2\beta))}, & b &= \frac{\tilde{\epsilon}}{2} \frac{\beta}{\sqrt{\vartheta} + \beta} \frac{r\eta_0}{2(n+1) + r\eta_0}, \\ d' &= 2B + \frac{\beta + 3\beta^3 - \beta^4}{1 - \beta(5 - \beta(3 - \beta)(1 - 2\beta))}, & C &= \frac{\frac{1}{2}\delta + \beta + \frac{3\beta^2}{(1-\beta)^3}}{1 - \left(\beta + \frac{3\beta^2}{(1-\beta)^3}\right)}, \end{aligned}$$

and pick  $\epsilon', \tilde{\epsilon}' > 0$  such that  $\tilde{\epsilon}' \sqrt{\frac{1+\epsilon'}{1-\epsilon'}} + \sqrt{\frac{2\epsilon'}{1-\epsilon'}} < \frac{2(1-d')^4}{1+(1-d')^4}$  and  $\epsilon' \geq \epsilon$ . Let  $m$  and  $N$  be as defined in Algorithm 3. If  $\|x_k - z(\eta_k)\|_{z(\eta_k)}^* \leq \frac{1}{2}\delta$  in some iteration  $k$  of Algorithm 3, then, with probability at least  $1 - p/m$ , we have  $\|x_{k+1} - z(\eta_{k+1})\|_{z(\eta_{k+1})}^* \leq \frac{1}{2}\delta$ .

*Proof.* First, we need to find an approximation  $\widehat{\Sigma}(-\eta_{k+1}c)$  of  $\Sigma(-\eta_{k+1}c)$  satisfying (47) and (48). Note that by (59) and the fact that  $\beta \leq \frac{7}{176}$ ,

$$\Delta x = \|z(\eta_{k+1}) - z(\eta_k)\|_{z(\eta_k)}^* \leq \beta + \frac{3\beta^2}{(1-\beta)^3} < 0.0452. \quad (63)$$

Then, the first inequality in (22) from Corollary 1 shows that

$$\Delta\theta = \|- \eta_{k+1}c + \eta_k c\|_{-\eta_k c} \leq \frac{\Delta x}{1 - \Delta x} < 0.0473. \quad (64)$$

Since

$$\|-\eta_{k+1}c\| = \eta_{k+1} \leq \eta_m = \eta_0 \left(1 + \frac{\beta}{\sqrt{\vartheta}}\right)^{\left\lceil \frac{\log\left(\frac{\vartheta(1+\delta/2)}{\eta_0 \epsilon}\right)}{\log\left(1 + \frac{\beta}{\sqrt{\vartheta}}\right)} \right\rceil} \leq \frac{\vartheta(1+\delta/2)(1+\beta/\sqrt{\vartheta})}{\bar{\epsilon}},$$

by the value of  $m$ , Theorem 2 shows

$$\begin{aligned} \lambda_{\min}(\Sigma(-\eta_{k+1}c)) &\geq \frac{1}{64} \left(\frac{r}{n+1}\right)^2 \min\left\{1, (4R\eta_{k+1})^{-4\sqrt{\vartheta}-2}\right\} \\ &\geq \frac{1}{64} \left(\frac{r}{n+1}\right)^2 \min\left\{1, \left(\frac{\bar{\epsilon}}{4R\vartheta(1+\delta/2)(1+\beta/\sqrt{\vartheta})}\right)^{4\sqrt{\vartheta}+2}\right\}. \end{aligned}$$

We can now apply Corollary 3 for  $\theta_0 = -\eta_k c$ ,  $\theta_1 = -\eta_{k+1}c$ , and  $\epsilon_1 = \epsilon$ . It can be shown that values of  $N$  and  $q$  given in Algorithm 3 satisfy

$$N \geq \frac{7840n^2}{\epsilon^2 \frac{p}{m} \left(\frac{1}{2} - \frac{1}{N}\right)}, \quad q \leq \left(\frac{\epsilon^2 \lambda_{\min}(\Sigma(-\eta_{k+1}c))^2}{799680n^2 R^4}\right) \left(\frac{p}{m} \left(\frac{1}{2} - \frac{1}{N}\right)\right).$$

Therefore, Corollary 3 guarantees that  $\widehat{\Sigma}(-\eta_{k+1}c)$  satisfies (47) and (48) with probability at least

$$1 - \frac{p}{m} \left(\frac{1}{2} - \frac{1}{N}\right). \quad (65)$$

Note that Algorithm 3 moreover sets  $q \leq p/(mN)$ . Lemma 5 thus guarantees that with probability  $1 - q \geq 1 - p/(mN)$ , the sample  $X_{k+1}$  follows the Boltzmann distribution with parameter  $-\eta_{k+1}c$ .

Second, we need an approximation  $\widehat{\theta}(x_k)$  of  $\theta(x_k)$  satisfying (49). It will suffice to show that the chosen parameters satisfy the requirements for Algorithm 4, and that this algorithm yields the desired result. Thus, we will call Algorithm 4 with  $\theta = -\eta_{k+1}c$ . Regarding the first condition of Algorithm 4, using the second inequality in (4), Corollary 1 and (59), we find

$$\begin{aligned} \|\theta(x) - \theta_0\|_{\bar{\theta}} &\leq \frac{\|\theta(x) - \theta_0\|_{\theta_0}}{1 - \|\bar{\theta} - \theta_0\|_{\theta_0}} \leq \frac{\frac{\|x_k - z(\eta_k)\|_{z(\eta_k)}^*}{1 - \|x_k - z(\eta_k)\|_{z(\eta_k)}^*}}{1 - \frac{\|z(\eta_k) - z(\eta_{k+1})\|_{z(\eta_k)}^*}{1 - \|z(\eta_k) - z(\eta_{k+1})\|_{z(\eta_k)}^*}} \leq \frac{\frac{\delta/2}{1-\delta/2}}{1 - \frac{\beta + \frac{3\beta^2}{(1-\beta)^3}}{1-\beta - \frac{3\beta^2}{(1-\beta)^3}}} \\ &= \frac{\delta}{2-\delta} \frac{1 - \beta(4 - \beta(3 - \beta)(1 - \beta))}{1 - \beta(5 - \beta(3 - \beta)(1 - 2\beta))} = B. \end{aligned}$$

Furthermore, we need to show that  $2B + \|\theta_0 - \bar{\theta}\|_{\bar{\theta}} \leq d'$ . Again using (4), Corollary 1 and (59), we find

$$2B + \|\theta_0 - \bar{\theta}\|_{\bar{\theta}} \leq 2B + \frac{\|\theta_0 - \bar{\theta}\|_{\theta_0}}{1 - \|\theta_0 - \bar{\theta}\|_{\theta_0}} \leq 2B + \frac{\frac{\|z(\eta_k) - z(\eta_{k+1})\|_{z(\eta_k)}^*}{1 - \|z(\eta_k) - z(\eta_{k+1})\|_{z(\eta_k)}^*}}{1 - \frac{\|z(\eta_k) - z(\eta_{k+1})\|_{z(\eta_k)}^*}{1 - \|z(\eta_k) - z(\eta_{k+1})\|_{z(\eta_k)}^*}}$$

$$\leq 2B + \frac{\frac{\beta + \frac{3\beta^2}{(1-\beta)^3}}{1-\beta - \frac{3\beta^2}{(1-\beta)^3}}}{1 - \frac{\beta + \frac{3\beta^2}{(1-\beta)^3}}{1-\beta - \frac{3\beta^2}{(1-\beta)^3}}} = 2B + \frac{\beta + 3\beta^3 - \beta^4}{1 - \beta(5 - \beta(3 - \beta)(1 - 2\beta))} = d'.$$

Since  $\epsilon \leq \epsilon'$ , we can use the approximation  $\widehat{\Sigma}(-\eta_{k+1}c)$  of  $\Sigma(-\eta_{k+1}c)$  from the main algorithm in Algorithm 4. Similarly, we already have a sample  $\bar{X} = X_{k+1}$  from the main algorithm.

Finally, as was shown in (60),

$$\|x - x(\bar{\theta})\|_{\Sigma(\bar{\theta})^{-1}} = \|x_k - z(\eta_{k+1})\|_{z(\eta_{k+1})}^* \leq \frac{\frac{1}{2}\delta + \beta + \frac{3\beta^2}{(1-\beta)^3}}{1 - \left(\beta + \frac{3\beta^2}{(1-\beta)^3}\right)} = C.$$

Hence, Theorem 7 shows that Algorithm 4 returns an approximation  $\widehat{\theta}(x_k)$  of  $\theta(x_k)$  such that, with probability  $1 - p' = 1 - p/(2m)$ ,  $\|\widehat{\theta}(x_k) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)} \leq b$ . Note that by (58) and Lemma 22, the right hand side of (49) is lower bounded by our value for  $b$ . We conclude  $\widehat{\theta}(x_k)$  satisfies (49). Since  $\widehat{\Sigma}(-\eta_k c)$  satisfies (47) and (48), Corollary 7.5 in De Klerk, Glineur and Taylor [6] shows

$$\left\| \widehat{\Sigma}(-\eta_k c)[\eta_{k+1}c + \widehat{\theta}(x_k)] - g_{z(\eta_{k+1})}^{*,\eta_{k+1}}(x_k) \right\|_{z(\eta_{k+1})}^* \leq \widehat{\epsilon} \left\| g_{z(\eta_{k+1})}^{*,\eta_{k+1}}(x_k) \right\|_{z(\eta_{k+1})}^*,$$

which means the distance of  $x_{k+1}$  to  $z(\eta_{k+1})$  is bounded by (61). By Lemma 21, we have  $\|x_{k+1} - z(\eta_{k+1})\|_{z(\eta_{k+1})}^* \leq \frac{1}{2}\delta$ .

The probability that  $\widehat{\Sigma}(-\eta_{k+1}c)$  satisfies (47) and (48) was given by (65). Moreover, the sample  $X_{k+1}$  follows the Boltzmann distribution with parameter  $-\eta_{k+1}c$  with probability at least  $1 - p/(mN)$ . Finally,  $\|\widehat{\theta}(x_k) - \theta(x_k)\|_{\Sigma(-\eta_{k+1}c)} \leq b$  with probability at least  $1 - p/(2m)$ . Hence, the probability that these three events occur simultaneously is at least  $1 - p/m$  by the union bound.

□

Using this result, it is not hard to see that Algorithm 3 converges with high probability. By the union bound, the probability that  $\|x_k - z(\eta_k)\|_{z(\eta_k)}^* \leq \frac{1}{2}\delta$  for all  $k \in \{1, \dots, m\}$  is at least  $1 - p$ . After  $m$  iterations, it follows from Renegar [25, relation (2.14)] that  $\langle c, x_m \rangle - \min_{x \in K} \langle c, x \rangle \leq \bar{\epsilon}$ .

We now turn to a run-time analysis of Algorithm 3 by breaking down the complexity of its most crucial steps. The complexities of the other lines are straightforward, and not relevant for the algorithm's total complexity.

*Line 8 in Algorithm 3* By construction, the main loop in Algorithm 3 has  $m$  iterations. Since  $\vartheta = n + o(n)$ ,

$$m = O\left(\frac{\sqrt{\vartheta}}{\beta} \log\left(\frac{\sqrt{\vartheta}}{\eta_0 \bar{\epsilon}}\right)\right) = O\left(\frac{\sqrt{n}}{\beta} \log\left(\frac{\sqrt{n}}{\eta_0 \bar{\epsilon}}\right)\right).$$

Note that if  $\beta$  were a constant, this would essentially be the usual worst-case iteration complexity for interior point methods. However, we now have  $1/\beta = \max\{O(1/S), O(R\sqrt{n}/\bar{\epsilon})\}$ , where  $1/S = O(\frac{n\sqrt{\vartheta}}{r\eta_0}) = O(\frac{n\sqrt{n}}{r\eta_0})$ . Therefore,

$$m = O\left(\sqrt{n} \max\left\{\frac{n\sqrt{n}}{r\eta_0}, \frac{R\sqrt{n}}{\bar{\epsilon}}\right\} \log\left(\frac{\sqrt{n}}{\eta_0 \bar{\epsilon}}\right)\right).$$

*Line 9 in Algorithm 3* Since  $1/\epsilon = O((1/\widehat{\epsilon})^2) = O(1/S^2)$ , the number of hit-and-run samples  $N$  in every iteration satisfies

$$N = O\left(\frac{n^2 m}{p\epsilon^2}\right) = O\left(\frac{n^2 m}{pS^4}\right) = O\left(\frac{n^{8.5}}{p(r\eta_0)^4} \max\left\{\frac{n\sqrt{n}}{r\eta_0}, \frac{R\sqrt{n}}{\bar{\epsilon}}\right\} \log\left(\frac{\sqrt{n}}{\eta_0 \bar{\epsilon}}\right)\right).$$

Each of these samples originates from a random walk with  $\ell$  steps, where

$$\ell = O\left(n^3 \log^2\left(\frac{n\sqrt{n}}{q^2}\right) \log^3(1/q^2)\right), \quad \text{where} \quad \log(1/q^2) = O\left(\sqrt{n} \log\left(\frac{Rn}{\bar{\epsilon}}\right) + \log\left(\frac{m^2 n^{12} R^8}{p^2 r^8 S^8}\right)\right).$$

*Line 11 in Algorithm 3* The number of hit-and-run steps required in every iteration to find a suitable approximation of  $\theta(x_k)$  by Algorithm 4 was proven in Theorem 7 to be  $m'N'\ell'$ . Note that  $B = \Theta(\delta) = \Theta(S\beta)$  and  $b = o(\hat{\epsilon}\beta(S\sqrt{n})) = o(S^2\beta\sqrt{n})$ . Therefore, we have  $m' = O(\log(\frac{1}{S\sqrt{n}}))$ . It also follows that

$$N' = O\left(\frac{nm'}{p'b^2}\right) = O\left(\frac{m'}{pS^4\beta^2}\right) = O\left(\frac{n^6}{p(r\eta_0)^4} \max\left\{\frac{n^3}{(r\eta_0)^2}, \frac{R^2n}{\bar{\epsilon}^2}\right\} \log\left(\frac{1}{S\sqrt{n}}\right)\right).$$

The complexity of  $\ell'$  is the same as that for  $\ell$ , but with  $q$  replaced by  $q'$ . Since each time Algorithm 4 is called it holds that  $\|\bar{\theta}\| = \|\eta_{k+1}c\| \leq \eta_m = O(\vartheta/\bar{\epsilon})$ ,

$$\log(1/(q')^2) = O\left(\log\left(\frac{m^2(m')^2R^4n^4}{p^2b^4(\bar{\epsilon}')^4r^4}\right) + n \log\left(\frac{Rn}{\bar{\epsilon}} + \frac{R^2d'n^2}{r\bar{\epsilon}}\right)\right).$$

Let the  $O^*$  notation suppresses polylogarithmic terms in the problem parameters. It follows from the above that the quantity  $mm'N'\ell'$  determines the number of hit-and-run steps, and this quantity has complexity

$$mm'N'\ell' = O^*\left(\sqrt{n} \frac{n^6}{p(r\eta_0)^4} n^3 n^5 \max\left\{\frac{n^{4.5}}{(r\eta_0)^3}, \frac{R^3n^{1.5}}{\bar{\epsilon}^3}\right\}\right) = O^*\left(\frac{n^{14.5}}{p(r\eta_0)^4} \max\left\{\frac{n^{4.5}}{(r\eta_0)^3}, \frac{R^3n^{1.5}}{\bar{\epsilon}^3}\right\}\right).$$

The number of hit-and-run steps also determines the complexity of the total number of operations in Algorithm 3. The following theorem summarizes the preceding discussion.

**THEOREM 8.** *Let  $p \in (0, 1)$  and  $\eta_0 > 0$ , and the values of  $S$ ,  $\hat{\epsilon}$ ,  $\beta$  and  $\delta$  as in Lemma 21. Let  $\tilde{\epsilon} > 0$  and  $\epsilon \in (0, \sqrt{13} - 3)$  be such that  $\tilde{\epsilon}\sqrt{\frac{1+\epsilon}{1-\epsilon}} + \sqrt{\frac{2\epsilon}{1-\epsilon}} \leq \hat{\epsilon}$ . Set  $B$ ,  $b$ ,  $d'$  and  $C$  as in Lemma 23, and pick  $\epsilon', \tilde{\epsilon}' > 0$  such that  $\tilde{\epsilon}'\sqrt{\frac{1+\epsilon'}{1-\epsilon'}} + \sqrt{\frac{2\epsilon'}{1-\epsilon'}} < \frac{2(1-d')^4}{1+(1-d')^4}$  and  $\epsilon' \geq \epsilon$ . Let  $m$  and  $N$  be as defined in Algorithm 3.*

*Then, with probability at least  $1 - p$ , Algorithm 3 returns an  $x_m \in K$  such that  $\langle c, x_m \rangle - \min_{x \in K} \langle c, x \rangle \leq \bar{\epsilon}$  after  $m$  iterations. If  $\bar{\epsilon}/R$  is fixed, the total complexity of Algorithm 3 is determined by*

$$O^*\left(\frac{n^{19}}{p(r\eta_0)^7}\right), \tag{66}$$

*hit-and-run steps.*

**7. Concluding Remarks** Our algorithm uses the starting conditions that a point  $x_0$  is known with  $\|x_0 - z(\eta_0)\|_{z(\eta_0)}^*$  small, and an approximation  $\hat{\Sigma}(-\eta_0c)$  of  $\Sigma(-\eta_0c)$  is known. A natural question is then how to find such  $x_0$  and  $\hat{\Sigma}(-\eta_0c)$ . As was shown in Lovász and Vempala [18], it is possible to generate samples from the uniform distribution under some mild assumptions. Thus, one can approximate the mean and the covariance matrix of the uniform distribution. If  $\eta_0$  is sufficiently small, self-concordance ensures that these approximations can serve as  $x_0$  and  $\hat{\Sigma}(-\eta_0c)$  to start the algorithm.

Alternatively, if  $\eta_0$  is too large to use this approach, we could use the algorithm by Kalai and Vempala [8] to generate approximations of  $\Sigma(-\eta c)$  for increasing  $\eta > 0$ . Once  $\eta$  is large enough, the results in this paper show that one can apply hit-and-run sampling to the Boltzmann distribution with parameter  $-\eta_0c$ , and thereby approximate its mean to find a point  $x_0$  with  $\|x_0 - z(\eta_0)\|_{z(\eta_0)}^*$  as small as desired.

The complexity bound (66) is much worse than the bounds normally associated with interior point methods. The main cause is the high number of hit-and-run samples required to approximate the covariance matrices to sufficient accuracy. The practical applicability of sampling-based interior point methods will thus largely depend on the number of samples needed. In practice, possible approaches to reduce this workload could include mixing acceleration (see e.g. Kaufman and Smith [11]), or long-step and predictor-corrector methods (to reduce the number of interior point iterations). We hope the results in this paper may further motivate the development of such improved sampling methods.

**Acknowledgments** The authors would like to thank Nikolaus Schweizer for many insightful discussions and for proofreading parts of this manuscript. We also thank Daniel Dadush for sharing his expertise.

## References

- [1] Abernethy J, Hazan E (2016) Faster convex optimization: Simulated annealing with an efficient universal barrier. *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, 2520–2528.
- [2] Badenbroek R, de Klerk E (2019) Simulated annealing with hit-and-run for convex optimization: rigorous complexity analysis and practical perspectives for copositive programming. *arXiv preprint arXiv:1907.02368* .
- [3] Bélisle CJ, Romeijn HE, Smith RL (1993) Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research* 18(2):255–266.
- [4] Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge university press).
- [5] Bubeck S, Eldan R (2018) The entropic barrier: Exponential families, log-concave geometry, and self-concordance. *Mathematics of Operations Research* 44(1):264–276.
- [6] de Klerk E, Glineur F, Taylor A (2017) Worst-case convergence analysis of gradient and Newton methods through semidefinite programming performance estimation. *arXiv preprint arXiv:1709.05191* .
- [7] Güler O (1996) Barrier functions in interior point methods. *Mathematics of Operations Research* 21(4):860–885.
- [8] Kalai AT, Vempala S (2006) Simulated annealing for convex optimization. *Mathematics of Operations Research* 31(2):253–266.
- [9] Kannan R, Lovász L, Simonovits M (1995) Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry* 13(1):541–559.
- [10] Kannan R, Lovász L, Simonovits M (1997) Random walks and an  $O^*(n^5)$  volume algorithm for convex bodies. *Random Structures and Algorithms* 11(1):1–50.
- [11] Kaufman DE, Smith RL (1998) Direction choice for accelerated convergence in hit-and-run sampling. *Operations Research* 46(1):84–95.
- [12] Kirkpatrick S, Gelatt CD, Vecchi MP, et al. (1983) Optimization by simulated annealing. *Science* 220(4598):671–680.
- [13] Klartag B (2006) On convex perturbations with a bounded isotropic constant. *Geometric & Functional Analysis GFA* 16(6):1274–1290.
- [14] Lee YT, Sidford A, Vempala SS (2018) Efficient convex optimization with membership oracles. *Proceedings of Machine Learning Research*, volume 75, 1292–1294.
- [15] Lee YT, Sidford A, Wong SCw (2015) A faster cutting plane method and its implications for combinatorial and convex optimization. *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, 1049–1065 (IEEE).
- [16] Levin DA, Peres Y, Wilmer EL (2017) *Markov Chains and Mixing Times*, volume 107 (American Mathematical Society).
- [17] Lovász L, Vempala S (2006) Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, 57–68 (IEEE).

- [18] Lovász L, Vempala S (2006) Hit-and-run from a corner. *SIAM Journal on Computing* 35(4):985–1005.
- [19] Lovász L, Vempala S (2006) Simulated annealing in convex bodies and an  $O^*(n^4)$  volume algorithm. *Journal of Computer and System Sciences* 72(2):392–417.
- [20] Lovász L, Vempala S (2007) The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms* 30(3):307–358.
- [21] MOSEK ApS (2017) *The MOSEK optimization toolbox for MATLAB manual. Version 8.1*. URL <http://docs.mosek.com/8.1/toolbox/index.html>.
- [22] Nesterov Y, Nemirovskii A (1994) *Interior-Point Polynomial Algorithms in Convex Programming* (SIAM).
- [23] Papp D, Yildiz S (2018) *alfonso: ALgorithm FOr Non-Symmetric Optimization*. URL <https://github.com/dpapp-github/alfonso>.
- [24] Papp D, Yildiz S (2018) On “a homogeneous interior-point algorithm for non-symmetric convex conic optimization”. *arXiv preprint arXiv:1712.00492v2* .
- [25] Renegar J (2001) *A Mathematical View of Interior-Point Methods in Convex Optimization* (SIAM).
- [26] Skajaa A, Ye Y (2015) A homogeneous interior-point algorithm for nonsymmetric convex conic optimization. *Mathematical Programming* 150(2):391–422.
- [27] Smith RL (1984) Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research* 32(6):1296–1308.
- [28] Sturm JF (1999) Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software* 11(1-4):625–653.
- [29] Toh KC, Todd MJ, Tütüncü RH (1999) SDPT3 - a MATLAB software package for semidefinite programming, version 1.3. *Optimization Methods and Software* 11(1-4):545–581.
- [30] Wright MH (2005) The interior-point revolution in optimization: history, recent developments, and lasting consequences. *Bulletin of the American Mathematical Society* 42(1):39–56.