Reducing Isotropy and Volume to KLS: An $O^*(n^3\psi^2)$ Volume Algorithm

He Jia* Aditi Laddha* Yin Tat Lee† Santosh S. Vempala*

August 6, 2020

Abstract

We show that the the volume of a convex body in $\mathbb{R}^n$ in the general membership oracle model can be computed with $\widetilde{O}(n^3\psi^2/\varepsilon^2)$ oracle queries, where ψ is the KLS constant[1]. With the current bound of $\psi \lesssim n^{\frac{1}{4}}$, this gives an $\widetilde{O}(n^{3.5}/\varepsilon^2)$ algorithm, the first general improvement on the Lovász-Vempala $\widetilde{O}(n^4/\varepsilon^2)$ algorithm from 2003. The main new ingredient is an $\widetilde{O}(n^3\psi^2)$ algorithm for isotropic transformation, following which we can apply the $\widetilde{O}(n^3/\varepsilon^2)$ volume algorithm of Cousins and Vempala for well-rounded convex bodies. A positive resolution of the KLS conjecture would imply an $\widetilde{O}(n^3/\epsilon^2)$ volume algorithm. We also give an efficient implementation of the new algorithm for convex polytopes defined by m inequalities in $\mathbb{R}^n$: polytope volume can be estimated in time $\widetilde{O}(mn^c/\varepsilon^2)$ where $c < 3.7$ depends on the current matrix multiplication exponent and improves on the the previous best bound.

1 Introduction

Computing the volume is a fundamental problem from antiquity, playing a central role in the development of fields such as integral calculus, thermodynamics and fluid dynamics among others. Over the past several decades, numerical estimation of the volumes of high-dimensional bodies that arise in applications has been of great interest. To mention one example from systems biology, volume has been proposed as a promising parameter to distinguish between the metabolic networks of normal and abnormal individuals [8] where the networks are modeled as very high dimensional polytopes.

As an algorithmic problem for convex bodies in $\mathbb{R}^n$, volume computation has a 4-decade history. Early results by Barány and by Barány and Füredi showed that any deterministic algorithm for the task is doomed to have an exponential complexity, even to approximate the volume to within an exponentially large factor. Then came the stunning breakthrough of Dyer, Frieze, and Kannan showing that with randomization, the problem can be solved in full generality (the membership oracle model) to an arbitrary relative error ε in time polynomial in n and $1/\varepsilon$. They used the Markov chain Monte Carlo method and reduced the volume problem to sampling uniformly from a sequence of convex bodies, and showed that the sampling itself can be done in polynomial-time. Subsequent progress on the complexity of volume computation has been accompanied by the discovery of several techniques of independent interest, summarized below. The current best complexity of $\widetilde{O}(n^4/\epsilon^2)$ for general convex bodies is achieved by the 2003 algorithm of Lovász and Vempala.

The main subroutine for volume computation is sampling. Sampling is done by random walks, notably the ball walk and hit-and-run. The rate of convergence of random walks is determined by their conductance. For a Markov chain with state space Ω, transition function $P_u(.)$ and stationary density Q, the conductance is

$$\phi = \min_{A\subset\Omega: Q(A)\leq\frac{1}{2}} \frac{\int_{u\in A} P_u(\Omega\setminus A)\, dQ(u)}{Q(A)}$$

*Georgia Tech, {hjia36, aladdha6, vempala}@gatech.edu
†University of Washington and Microsoft Research, yintat@uw.edu

[1] $\widetilde{O}$ suppresses polylogarithmic terms. O^*suppresses dependence on error parameters as well as polylogarithmic terms.

Year/Authors	New ingredients	Steps
1989/Dyer-Frieze-Kannan [6]	Everything	n^{23}
1990/Lovász-Simonovits [18]	Better isoperimetry	n^{16}
1990/Lovász [17]	Ball walk	n^{10}
1991/Applegate-Kannan [2]	Logconcave sampling	n^{10}
1990/Dyer-Frieze [5]	Better error analysis	n^8
1993/Lovász-Simonovits [19]	Localization lemma	n^7
1997/Kannan-Lovï¿œsz-Simonovits [11]	Speedy walk, isotropy	n^5
2003/Lovász-Vempala [20]	Annealing, hit-and-run	n^4
2015/Cousins-Vempala [3] (well-rounded)	Gaussian Cooling	n^3

Table 1: The complexity of volume estimation. The complexity of all algorithms have an ϵ^{-2} factor. Each step uses $\widetilde{O}(n)$ bit of randomness and $\widetilde{O}(n^2)$ arithmetic operations. The last algorithm is for well-rounded convex bodies.

i.e., the minimum conditional escape probability with the stationary density (probability of crossing from a set to its complement starting from the stationary density in the set). The analysis of the ball walk done in [11] shows that the conductance can be bounded in terms of the isoperimetry of the stationary distribution, a purely geometric parameter. For an n-dimensional measure ν, its KLS constant (the isoperimetric or Cheeger constant is the reciprocal) is

$$\frac{1}{\psi_\nu} = \inf_{A\subset\mathbb{R}:Q(A)\leq\frac{1}{2}} \frac{\nu_{n-1}(\partial A)}{\nu(A)}$$

where ∂A is the boundary of the set A, and ν_{n-1} is the induced $(n-1)$-dimensional measure. The conductance of the Markov chain reduces to the isoperimetry of its stationary density[2]:

$$\phi \gtrsim \frac{1}{\psi \cdot n}$$

with ψ being the KLS constant of the stationary density of the Markov chain. This implies a mixing rate of $O(n^2\psi^2)$. Thus bounding the KLS constant becomes critical, and this consideration originally motivated the KLS conjecture. With many unexpected connections and applications since its formulation, the conjecture has become a central part of asymptotic convex geometry and functional analysis.

Conjecture 1.1 (KLS Conjecture [10])**.** *The KLS constant of any isotropic logconcave density in any dimension is bounded by an absolute constant. Equivalently, for a logconcave density q with covariance matrix A, $\psi_q \lesssim \|A\|_{\text{op}}^{1/2}$.*

Another equivalent formulation of the conjecture is that for any logconcave density, a halfspace induced subset achieves the extremal isoperimetry up to an absolute constant. The paper [10] also showed that

$$\psi_q = O\left(\sqrt{\text{Tr}A}\right)$$

which is $O(\sqrt{n})$ for isotropic logconcave densities. This implies mixing of the ball walk (and hence sampling) in n^3 steps from a warm start in an isotropic convex body K containing a unit ball. But how to find an isotropic transformation and maintain a warm start? They do this with two essential ingredients: (i) a bound of $O(n^2R^2)$ on the mixing time where $R^2 = \mathbb{E}_K(\|x-\bar{x}\|^2)$ (ii) interleaving the volume algorithm with isotropic transformation: the algorithm starts with a simple isotropic body (like a ball) and then goes through a sequence of $\widetilde{O}(n)$ convex bodies, maintaining well-roundedness, i.e., $R = \widetilde{O}(\sqrt{n})$, and computing ratios of volumes of consecutive bodies as well as isotropic transformation, via sampling; a random sample from the current phase serves as a warm start for the next phase. As $\widetilde{O}(n)$ samples are needed to estimate each ratio and for the isotropic transformation, this implies a volume/rounding algorithm of complexity $n^3 \cdot n \cdot n = O^*(n^5)$.

The algorithm of [20] improves on this as follows: (a) they separated the isotropic transformation from volume estimation by giving an $\widetilde{O}(n^4)$ algorithm for isotropic transformation and (b) they replaced the sequence of convex bodies with a sequence of logconcave densities ("simulated annealing"), specifically exponential densities restricted to convex bodies; they showed that a sequence of $\widetilde{O}(\sqrt{n})$ densities suffices, while still maintaining a warm start.

[2] We use $\gtrsim$ to denote "LHS greater than a constant factor times RHS".

This reduced the overall complexity to $n^3 \cdot \sqrt{n} \cdot \sqrt{n} = O^*(n^4)$. (Note that by a simple variance analysis, the number of samples needed per phase grows linearly with the number of phases [5]). The total number of samples used in the algorithm is $\widetilde{O}(n)$, so further improvements would require faster sampling. The concluding remark from [20] says:

"There is one possible further improvement on the horizon. This depends on a difficult open problem in convex geometry, a variant of the "Slicing Conjecture" [10]. If this conjecture is true... could perhaps lead to an $O^*(n^3)$ volume algorithm. But besides the mixing time, a number of further problems concerning achieving isotropic position would have to be solved."

Since then there has been progress on the KLS conjecture.

Theorem 1.2 ([16]). *The KLS constant for any logconcave density with covariance A is bounded by $O(\|A\|_F^{1/2})$. For isotropic logconcave densities in $\mathbb{R}^n$, $\psi \lesssim n^{1/4}$. The mixing time of the ball walk for isotropic logconcave densities from a warm start is $\widetilde{O}(n^{2.5})$.*

Unfortunately, despite these improvements to the KLS constant and the mixing rate of the ball walk, the complexity of volume computation remained at n^4. Even outputting the first random point needs n^4 oracle queries, since the mixing rate improvement is only from a warm start in an isotropic body.

Progress in a different line, without using the KLS conjecture came in 2015. Cousins and Vempala gave a volume algorithm with complexity $O^*(n^3)$ for any *well-rounded* convex body, i.e., they assume the input body contains the unit ball and is mostly contained in a ball of radius $R = \widetilde{O}(\sqrt{n})$. This is a weaker condition than (approximate) isotropic position, which requires that the covariance matrix of K is (close to) the identity. Their Gaussian Cooling algorithm uses a sequence of Gaussians restricted to the body, starting with a Gaussian of small variance almost entirely contained in the body and flattening it to a near-uniform distribution. Notably, they bypass the KLS conjecture, needing isoperimetry only for the special case of the Gaussian density restricted to a convex body, for which $\psi = O(1)$. The main open problem from their work is to find a faster algorithm to make the body well-rounded (or isotropic). An improved rounding algorithm would directly imply a faster volume algorithm.

1.1 Main results

We give a new algorithm for the isotropic transformation of any given convex body. The complexity of the algorithm is $O^*(n^3\psi^2)$, directly determined by the best possible KLS constant ψ for isotropic logconcave densities in $\mathbb{R}^n$. This is the first direct reduction of the complexity of volume computation to the KLS constant. With the current bound of $\psi = O(n^{1/4})$, the complexity is $O^*(n^{3.5})$ and implies volume computation in the same complexity for any convex body. If the conjecture turns out to be true, then the complexity becomes $O^*(n^3)$. No further improvement is in sight for the general setting.

Theorem 1.3 (Rounding). *There is a randomized algorithm that takes as input a convex body $K \subset \mathbb{R}^n$ given by a membership oracle with initial point $x_0 \in K$, bounds $r, R¿0$ s.t.,*

$$x_0 + rB_n \subseteq K \subseteq RB_n$$

and with probability $1-\delta$, computes an affine transformation T s.t., TK is in near-isotropic position, i.e., for x sampled uniform from TK,

$$I \preccurlyeq \mathrm{Cov}(xx^\top) \preccurlyeq 2I.$$

The algorithm takes $O(n^3\psi^2 \log^c(Rn/r)\log(1/\delta))$ oracle queries and $O(n^2)$ time per oracle query.

Corollary 1.4 (Volume). *The oracle complexity of computing the volume of a convex body $K \subset \mathbb{R}^n$ given by a well-defined membership oracle with probability at least $1-\delta$ to within relative error ε is*

$$O(n^3(\psi^2/\varepsilon^2)\log^{O(1)}(Rn/r\epsilon)\log(1/\delta)) = O(n^{3.5}/\varepsilon^2 \log^{O(1)}(Rn/r\epsilon))$$

which is $\widetilde{O}(n^{3.5}/\varepsilon^2)$ for the current KLS constant bound of $\psi = O(n^{1/4})$. The time complexity is $O(n^2)$ per oracle query.

For an explicit convex polytope $Ax \leq b$, with $A \in R^{m \times n}$, since each oracle query takes $O(mn)$ time, this immediately gives a bound of $O^*(mn^{4.5})$, matching the current best bound for large m [23] (there is a different approach [14] giving a time bound of $O^*(m^2 n^{\omega - \frac{1}{3}})$, which is better when the number of facets is close to linear in the dimension). Here we give an efficient implementation of our new algorithm that improves significantly on the runtime. Let $\omega(a, b, c)$ be the exponent of complexity of multiplying an $n^a \times n^b$ matrix with an $n^b \times n^c$ matrix, e.g., $\omega(1, 1, 1) \leq 2.3728639$ [12].

Theorem 1.5 (Polytope Volume). *The volume of a convex polytope $\{x : Ax \leq b\} \subset \mathbb{R}^n$ defined by m inequalities can be computed with high probability to within relative error ε using fast matrix multiplication in time $\widetilde{O}(mn^c/\varepsilon^2)$ where $c = \omega(1, 1, 3.5) < 3.7$. If the KLS conjecture is true, then the running time is $\widetilde{O}(mn^{\omega(1,1,3)-1})$ where $\omega(1, 1, 3) < 4.2$. Without fast matrix multiplication, polytope volume can be computed in time $\widetilde{O}(mn^4/\varepsilon^2)$.*

1.2 Approach

The algorithm is based on the following ideas. First, if the covariance matrix is skewed (i.e., some eigenvalues are much larger than others), this can be detected via a small random sample. The roundness can then be improved by scaling up the subspace of small eigenvalues. But how to sample a highly skewed convex body? To break this chicken-and-egg problem, we use a sequence of convex bodies obtained by intersecting the original set with balls of increasingly larger radii:

$$K_t = K \cap B(0, t),$$

with t starting at r and doubling till it reaches R. In each outer iteration, we compute a transformation that makes the set K_t (nearly) isotropic. When K_t is isotropic, we show that K_{2t} is well-rounded (Lemma 3.4). Hence, the trace of the covariance of K_{2t} is $O(n)$. However, its eigenvalues could be widely varying. To make K_{2t} isotropic, we will estimate the larger eigenvalues using a small random sample (Lemma A.2). The second idea is that scaling up the small eigenvalues nearly doubles the size of the ball contained inside (Lemma 3.2), while having a mild effect on the higher norms of the covariance (Lemma 3.1). The latter concept is where KLS comes in. We show that if the KLS constant for isotropic logconcave distributions in $\mathbb{R}^n$ is bounded by $\psi_n^2 \lesssim n^{\frac{1}{p}}$ for all n, then for any logconcave density q with covariance matrix A (not necessarily identity), the KLS constant is bounded as $\psi_q^2 \lesssim \|A\|_p \log n$ (Theorem B.12). As we outline below, this improves the sampling time in each inner iteration.

In the beginning, when the ball contained inside K_t is small, we will use a few samples to get a coarse estimate of the larger eigenvalue directions and scale up the orthogonal subspace. The sampling time is higher since the roundness parameter of K_t is higher. The time per sample is roughly $n^2 \frac{\|A\|_p}{r^2}$ when r is the radius of the ball inside and A is the covariance matrix for the uniform density q on K_t. As we increase r, by a constant factor in each step, the norm of the covariance grows much more slowly, and so the sampling time decreases. Meanwhile, we use more samples in each step, roughly r^2, and this trade-off keeps the overall time at $n^{3+\frac{1}{p}} \simeq n^3 \psi^2$. During the process we need a warm start for each phase; we achieve this in $\widetilde{O}(n^3)$ steps using the Gaussian Cooling algorithm. It is known that $\psi_q^2 \lesssim \|A\|_p$ holds for $p = 2$ [16], giving the $\widetilde{O}(n^{3.5})$ complexity for rounding. Our general reduction establishes that any future improvement to higher p would imply faster rounding and volume estimation.

2 Preliminaries

For a positive definite matrix $n \times n$ matrix A, we will use the operator or spectral norm, $\|A\|_{\text{op}}$, its Frobenius norm, $\|A\|_F \stackrel{\text{def}}{=} \text{Tr}(A^2)^{1/2}$, and its $p'th$ norm for $p \geq 1$,defined as

$$\|A\|_p \stackrel{\text{def}}{=} \left(\sum_{i=1}^{n} |\lambda_i|^p \right)^{1/p}$$

with $p = 1$ being $\text{Tr}A$ and $p = 2$ being $\|A\|_F$.

Definition 2.1. A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *logconcave* if its logarithm is concave along every line, i.e., for any $x, y \in \mathbb{R}^n$ and any $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}. \tag{2.1}$$

Many common probability distributions have logconcave densities, e.g., Gaussian, exponential, logistic, and gamma distributions; indicator functions of convex sets are also logconcave. Logconcavity is preserved by product, min and convolution (in particular marginals of logconcave densities are logconcave).

Definition 2.2. A distribution D is said to be *isotropic* if

$$\mathbb{E}_D(x) = 0 \quad \text{and} \quad \mathbb{E}_D(xx^\top) = I.$$

We say a convex body is isotropic if the uniform distribution over it is isotropic. Any distribution with bounded second moments can be brought to an isotropic position by an affine transformation. We say that a convex body K is (r, R)-*rounded* if it contains a ball of radius r, and its covariance matrix A satisfies

$$\mathrm{Tr}A = \mathbb{E}_K(\|x - \bar{x}\|^2) = R^2.$$

For an isotropic body we have $A = I$ and hence $R^2 = n$. We will say K is well-rounded if $r = 1$ and $R = \widetilde{O}(\sqrt{n})$. Clearly, any isotropic convex body is also well-rounded, but not vice versa. A distribution is said to be α-isotropic if $I \preceq \mathbb{E}_D(xx^\top) \preceq \alpha I$.

Random points from isotropic logconcave densities have strong concentration properties. We mention three that we will use.

Lemma 2.3 ([21, Theorem 5.17])**.** *For any $t \geq 1$, and any logconcave density p in $\mathbb{R}^n$ with covariance matrix A,*

$$\mathbb{P}_{x\sim p}\left(\|x - \mathbb{E}x\|_2 \geq t \cdot \sqrt{\mathrm{Tr}A}\right) \leq \exp(-t + 1).$$

Lemma 2.4 ([24, Theorem 5.17])**.** *For any $t \geq 1$, and any isotropic logconcave density p in $\mathbb{R}^n$, there is a constant c such that*

$$\mathbb{P}_{x\sim p}(\|x\|_2 \geq t\sqrt{n}) \leq \exp(-c\sqrt{n}t).$$

Lemma 2.5 ([1, 25])**.** *For any isotropic logconcave distribution ν in $\mathbb{R}^n$,with probability at least $1-\delta$, the empirical mean and covariance*

$$\overline{x} = (1/N)\sum_{i=1}^{N} x_i, \quad \text{and} \quad X = \frac{1}{N}\sum_{i=}^{N}(x_i - \overline{x}_i)(x_i - \overline{x}_i)^T$$

of $N = O(n\log(1/\delta)/\varepsilon^2)$ i.i.d. random samples $x_i \sim \nu$ satisfy

$$\|\overline{x}\|_2 \leq \varepsilon, \quad \|X - I\|_{\mathrm{op}} \leq \varepsilon.$$

The convergence of Markov chains is established by showing the the t-th step distribution Q_t approaches the steady state distribution Q. We will use the total variation distance d_{TV} for this. We also need a notion of a warm start.

Definition 2.6 (Warm start)**.** We say that a starting distribution Q_0 is M-warm for a Markov chain with unique stationary distribution Q if its χ-squared distance is bounded by M:

$$\chi^2(Q_0, Q) = \mathbb{E}_Q\left(\frac{dQ_0(u)}{dQ(u)} - 1\right)^2 \leq M.$$

Note that $\chi^2(Q_0, Q) + 1 = \mathbb{E}_Q\left(\frac{dQ_0(u)}{dQ(u)}\right)^2 = \mathbb{E}_{Q_0}\frac{dQ_0(u)}{dQ(u)}$.

Our algorithms will use the ball walk for sampling. In a convex body K, the *ball walk* with step-size δ is defined as follows: At the current point, $x \in K$, pick y uniformly in $B(x, \delta)$, the ball of radius δ around x; if $y \in K$, go to y (else, do nothing).

The next theorem is a fast sampler for distributions with KLS constant ψ_q given a warm start.

Theorem 2.7 ([11])**.** *Let K be a convex body containing the unit ball. Using the ball walk with step size $\frac{1}{\sqrt{n}}$ in K from an M-warm start Q_0, the number of steps to generate a nearly independent point within TV distance or χ^2-distance ε of the uniform stationary density Q in $\mathbb{R}^n$ is $O(n^2\psi_q^2\log^{O(1)}(nM/\epsilon))$.*

The next lemma connects the KLS constant for isotropic distributions to that of general distributions. We give a proof in Section (B).

Lemma 2.8. *Let ψ_n be a bound on the KLS constant for isotropic logconcave densities and ψ_q be the bound for a logconcave density q (not necessarily isotropic) with covariance matrix A_q, both in $\mathbb{R}^n$. If $\psi_n^2 = \Theta\left(n^{1/p}\right)$ for all n, then for any logconcave density q with covariance A, we have $\psi_q^2 = O(\|A\|_p \log^{O(1)} n)$.*

We will also use a fast sampler for well-rounded convex bodies that does not require a warm start.

Theorem 2.9 (Sampling a well-rounded convex body [3, 4])**.** *There is an algorithm that, for any $\varepsilon > 0, p > 0$, and any convex body K in $\mathbb{R}^n$ that contains the unit ball and has $\mathbb{E}_K(\|X\|^2) = R^2$, with probability $1-\delta$, generates random points from a density ν that is within total variation distance ε from the uniform distribution on K. In the membership oracle model, the complexity of each random point, including the first, is,*

$$O\left((R^2n^2 + n^3)\log n \log^2 \frac{n}{\varepsilon} \log \frac{1}{\delta}\right) = \widetilde{O}(n^3 + R^2n^2).$$

Finally, we use the algorithm from [4, 3] for computing the volume of a well-rounded convex body.

Theorem 2.10 (Volume of a well-rounded convex body)**.** *[3, 4] There is an algorithm that, for any $\varepsilon > 0, \delta > 0$ and convex body K in $\mathbb{R}^n$ that contains the unit ball and has $\mathbb{E}_K(\|X\|^2) = O(n)$, with probability $1-\delta$, approximates the volume of K to within relative error ε and has complexity*

$$O\left(\frac{n^3}{\varepsilon^2} \cdot \log^2 n \log^2 \frac{1}{\varepsilon} \log^2 \frac{n}{\varepsilon} \log \frac{1}{\delta}\right) = \widetilde{O}(n^3\epsilon^{-2})$$

in the membership oracle model.

Computational Model. We use the most general membership oracle model for convex bodies, which comes with bounds r, R guaranteeing that the input convex body K is (r, R)-rounded and allows membership queries to K. As established in the literature on volume computation, the number of arithmetic operations is bounded by $O(n^2)$ per oracle query, and all arithmetic operations can be done using only a polylogarithmic number of additional bits. We mention a few familiar technical difficulties whose solution is well-documented in the literature. First, the sampling algorithms produce points from *approximately* the correct distribution. Second, samples produced in a sequence are not completely independent. Third, all our algorithms are randomized (they have to be in the oracle model). For the first two issues, we refer to [22]; briefly, the approximate distribution is handled by a trick called "divine intervention", where one can view the sampled distribution as being the correct one with large probability and an incorrect one with a small failure probability. Since the complexity dependence on proximity to the target is logarithmic, this leads to a controllable overall failure probability. The near-independence is handled as follows: first, we maintain parallel independent threads of samples, which start as completely random points (e.g., from a Gaussian) and remain independent throughout. For estimates computed in different sequential phases (e.g., ratios of integrals, or affine transformations), the degree of dependence is explicitly bounded using the tools developed in [22]. For the failure probability, we note that the overhead for failure probability δ is $O(\log(1/\delta))$. In the rest of the paper, for convenience, we say WHP to mean with probability $1-\delta$ incurring only $O(\log(1/\delta))$ overhead in the complexity.

3 Algorithm and Analysis

The algorithm considers a sequence of balls of doubling radii, and in each iteration makes the intersection of the current ball with the convex body nearly isotropic.

Algorithm 1 Iterative Isotropization

procedure IterativeIsotropization($K \subset \mathbb{R}^n$, $r > 0$, $R > 0$)
  Assumption: $B(0, r) \subseteq K \subseteq B(0, R)$
  Define $K_t = K \cap B(0, t)$..
  $t \leftarrow r, x \leftarrow 0, T \leftarrow \frac{1}{r} I$ ▷ Current slice t and normalization
  while $t < R$ **do**
    $(x, T) \leftarrow$ Isotropization$(K_t - x, T)$
    $t \to 2t$
  end while
end procedure

The main subroutine is a procedure to compute an isotropic transformation of a well-rounded body. We address this in the next section, then come back to the general analysis.

3.1 Inner loop: Isotropic transformation of a well-rounded body

We begin with the algorithm. In each iteration, the inner ball radius grows by a constant factor, while the p-norm of the covariance grows much more slowly. As a result, we can afford to sample more points with each iteration and thereby get progressively better approximations to the isotropy: in the first step the number of samples in only $\widetilde{O}(1)$ while by the end the number of samples is $\widetilde{O}(n)$.

Algorithm 2 Isotropize

procedure Isotropize($K \subset \mathbb{R}^n$, $T_0 \in \mathbb{R}^{n \times n}$)
  Assumption: $B(0, \frac{1}{4}) \subset T_0 K$ and $\mathbb{E}_{x \sim T_0 K} \|x\|^2 = O(n)$.
  Use Gaussian cooling to sample a point $x \in TK$.
  $r \leftarrow \frac{1}{4}, T \leftarrow T_0$. ▷ We maintain $\{x : \|x\|_2 \le r\} \subseteq TK$
  while $2^{10} r^2 \log^2 n \le n$ **do**
    $k \leftarrow c \cdot r^2 \log^5 n$ for a constant c.
    Sample points $x_1, \cdots, x_{2k}$ from TK using the ball walk with initial point x.
    $\widehat{A} \leftarrow \frac{1}{k} \sum_{i=1}^{k} (x_i - \widehat{x})(x_i - \widehat{x})^\top$ where $\widehat{x} = \frac{1}{k} \sum_{i=1}^{k} x_{i+k}$.
    $M \leftarrow I_n + P_{\widehat{A}}$ where $P_{\widehat{A}}$ is the projection to the subspace spanned by eigenvectors of $\widehat{A}$ with eigenvalue at most $\lambda = n$.
    $T \leftarrow MT$, $r \leftarrow 2(1 - \frac{1}{\log n}) \cdot r$.
  end while
  Sample $O(n)$ points to compute the mean $\widehat{x}$ and empirical covariance matrix $\widehat{A}$ such that $\widehat{A}^{-\frac{1}{2}}(K - \widehat{x})$ is 2-isotropic.
  Return $(\widehat{x}, \widehat{A}^{-\frac{1}{2}})$.
end procedure

First, we show that the p-th norm of the covariance matrix $\|A\|_p$ remains bounded. Although we only use an eigenvalue threshold of $\lambda = n$ in the algorithm, we will prove a slightly more general statement below, assuming the eigenvalue threshold is $\lambda = n/r^\alpha$.

Lemma 3.1. *Let A_j be the covariance matrix of TK and r_j be the inner radius at the j-th iteration of algorithm* Isotropize*. For $\alpha \in [0, 2/(p-1)]$, we have*

- *the number of iterations is $O(\log n)$.*
- $\mathrm{Tr} A_j = O(r_j^2 \mathrm{Tr} A_0)$ *for all j.*
- $\|A_j\|_p \le 24 n r_j^{(\frac{2+\alpha}{p} - \alpha)} \log n$ *for all j.*

Proof. In each iteration, the algorithm increases r_j by a factor of $2(1 - \frac{1}{\log n})$. Note that the algorithm starts with $r_0 = \frac{1}{4}$ and ends before $r_j \log n \le \sqrt{n}$. Hence, it takes less than $O(\log n)$ iterations.

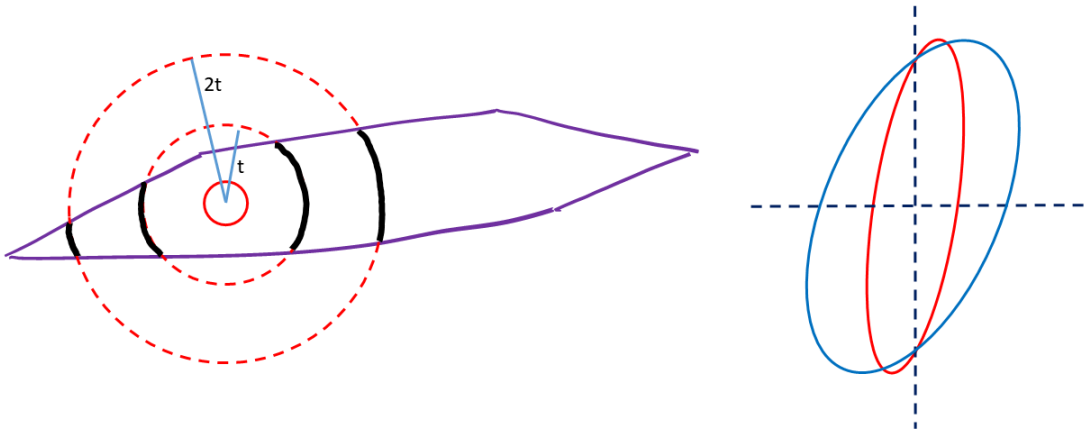

Figure 3.1: (a) Algorithm IterativeIsotropization uses balls of doubling radii (b) Algorithm Isotropize scales up the estimated "small eigenvalue" subspace in each iteration.

To bound $\mathrm{Tr}A_j$, we let M_j be the transformation at j-th iteration. We have

$$A_j = M_j A_{j-1} M_j \tag{3.1}$$

Hence, we have

$$\mathrm{Tr}A_j = \mathrm{Tr}A_{j-1}^{\frac{1}{2}} M_j^2 A_{j-1}^{\frac{1}{2}} \leq 5\mathrm{Tr}A_{j-1}.$$

Hence, $\mathrm{Tr}A_j$ increases by a factor of at most 4. Moreover, since r_j increases by a factor of $2(1-\frac{1}{\log n})$ we have that $\frac{\mathrm{Tr}A_j}{r_j^2}$ increases by a factor of $(1-\frac{1}{\log n})^{-2}$ per iteration. Since there are $O(\log n)$ iterations, it increases by at most $O(1)$ in total. Hence, we have $\mathrm{Tr}A_j = O(r_j^2 \mathrm{Tr}A_0) = O(r_j^2 n)$ for all j.

To bound $\|A_j\|_p$, we note that initially $\|A_0\|_p \leq \mathrm{Tr}A_0 \leq 4n$. Let P_j be the projection at j-th iteration. By (3.1), we have

$$\begin{aligned}\|A_j\|_p &= \|M_j A_{j-1} M_j\|_p \\ &= \|A_{j-1}^{1/2} M_j^2 A_{j-1}^{1/2}\|_p \\ &= \|A_{j-1}^{1/2}(I+3P_j)A_{j-1}^{1/2}\|_p.\end{aligned}$$

Hence, we have

$$\|A_j\|_p \leq \|A_{j-1}\|_p + 3\|A_{j-1}^{1/2} P_j A_{j-1}^{1/2}\|_p.$$

Let $\lambda \stackrel{\mathrm{def}}{=} \frac{n}{r_j^\alpha}$. Since P_j is the projection of the eigenspace of $\widehat{A}$ with eigenvalues less than λ, we have that

$$P_j \preceq 2\lambda(\widehat{A} + \lambda \cdot I)^{-1}.$$

(To see this, we note that sides have the same eigenspace and hence it follows from $1_{x\leq\lambda} \leq \frac{2\lambda}{x+\lambda}$.) By Lemma A.2 with $\epsilon = \frac{1}{2}$, we have that $\frac{1}{2}A_j - O(\frac{\log^3 n \cdot \mathrm{Tr}A_j}{k}) \cdot I \preceq \widehat{A}$. Using that $\mathrm{Tr}A_j = O(r_j^2 n)$ and $k = \Omega(r_j^{2+\alpha}\log^3 n)$ samples, we have

$$A_{j-1} \preceq 2\widehat{A} + \frac{n}{r_j^\alpha} \cdot I \preceq 2\widehat{A} + \lambda \cdot I.$$

Hence, we have $P_j \preceq 4\lambda(A_{j-1} + \lambda \cdot I)^{-1}$ and

$$\|A_j\|_p \leq \|A_{j-1}\|_p + 12\lambda \|A_{j-1}^{1/2}(A_{j-1} + \lambda \cdot I)^{-1} A_{j-1}^{1/2}\|_p$$

Let λ_i be the eigenvalues of A_{j-1}. Then,

$$\|A_{j-1}^{1/2}(A_{j-1} + \lambda \cdot I)^{-1} A_{j-1}^{1/2}\|_p^p = \sum_{i=1}^{n} \left(\frac{\lambda_i}{\lambda_i + \lambda}\right)^p \leq \frac{1}{\lambda^p} \sum_{\lambda_i \leq \lambda} \lambda_i^p + \sum_{\lambda_i \geq \lambda} 1.$$

The first term $\sum_{\lambda_i \leq \lambda} \lambda_i^p$ is maximized when the small eigenvalues have values exactly $\lambda_i = \lambda$. In this case, there are

$$\frac{\mathrm{Tr} A_{j-1}}{\lambda} \leq \frac{O(r_{j-1}^2 n)}{\lambda}$$

small eigenvalues, i.e., of value at most λ. Hence, we have

$$\frac{1}{\lambda^p} \sum_{\lambda_i \leq \lambda} \lambda_i^p \leq \frac{O(r_{j-1}^2 n)}{\lambda}.$$

For the second term,

$$\sum_{\lambda_i \geq \lambda} 1 \leq \frac{\mathrm{Tr} A_{j-1}}{\lambda} \leq \frac{O(r_{j-1}^2 n)}{\lambda}.$$

Hence, $\|A_j\|_p$ is increased by

$$O(\lambda) \cdot \left(\frac{r_{j-1}^2 n}{\lambda}\right)^{1/p} = O\left(\frac{n}{r_j^\alpha}\right) \cdot \left(r_{j-1}^2 r_j^\alpha\right)^{1/p} \leq O\left(n r^{\frac{2+\alpha}{p} - \alpha}\right).$$

Since there are at most $O(\log n)$ iterations, $\|A_j\|_p$ is at most $O(n r^{(\frac{2+\alpha}{p} - \alpha)} \log n)$. □

Next, we show that the inner radius increases by almost a factor of 2 every iteration, again for a general choice of eigenvalue threshold $\lambda = n/r^\alpha$.

Lemma 3.2. *Algorithm* Isotropization *maintains the invariant* $\{\|x\|_2 \leq r\} \subseteq TK$.

Proof. First, we describe the idea of the proof. When the algorithm modifies the covariance matrix A, it estimates the subspace of directions with variance less than n/r^α and doubles them. Lemma A.2 shows that

$$\widehat{A} = (1 \pm \frac{1}{2})A \pm \frac{n}{2r^\alpha} \cdot I,$$

this means, roughly speaking, that any direction that was not doubled satisfies

$$\lambda_i \geq \sigma^2 = \frac{n}{2r^\alpha}.$$

So the current body contains an ellipsoid whose axis lengths along the non-doubled directions (call this subspace V) are at least σ and at least r in all directions (the body contains a ball of radius r). Now the body is stretched by a factor of $\sqrt{2}$ in the subspace $V^\perp$. We will argue that the resulting body contains a ball of radius nearly $2r$. To see this, we argue that any point in a ball of this radius is in the convex hull of the ball of radius σ in the subspace V and a ball of radius $\sqrt{2}r$ in the subspace $V^\perp$.

Now we proceed to the formal proof by induction. Initially, we have $\{\|x\|_2 \leq r\} \subset TK$ by the assumption on K. It suffices to show this is maintained after each iteration. Let T be the old transformation and T' be the new transformation during one iteration. Let r be the old radius during that iteration. By the invariant, we know that $\{\|x\|_2 \leq r\} \subset TK$ and hence

$$\{\|M^{-1}x\|_2 \leq r\} \subset T'K$$

with $M = I + P_{\widehat{A}}$ (this only scales up the body by a factor of 2 in some directions). Let A be the covariance matrix of TK. Then $(I + P_{\widehat{A}})A(I + P_{\widehat{A}})$ is the covariance matrix of $T'K$. Our goal is to show that $T'K$ contains a ball of radius $2(1 - \frac{1}{\log n})r$. From above, we have

$$\{x^\top A^{-1}x \le 1\} \subset \{x^\top((I + P_{\widehat{A}})A(I + P_{\widehat{A}}))^{-1}x \le 1\} \subset T'K.$$

Hence, we know that $\mathrm{Conv}(\Omega_1 \cup \Omega_2) \subset T'K$ where

$$\Omega_1 = \{x^\top(\frac{1}{r^2}I - \frac{3}{4r^2}P_{\widehat{A}})x \le 1\} \text{ and } \Omega_2 = \{x^\top A^{-1}x \le 1\}.$$

where for Ω_1 we used that $(I + P)^{-2} = I - \frac{3}{4}P$ for any projection matrix P.

To prove that $T'K$ contains a larger ball, we take x and write it as a convex combination of x_1 and x_2 where $x_1 \in \Omega_1$ and $x_2 \in \Omega_2$. To do this, we define P_A be the projection to the subspace spans by eigenvectors of A with eigenvalues at most $\lambda/2^8 \log^2 n$ with

$$\lambda \stackrel{\text{def}}{=} \frac{n}{r^\alpha}.$$

For any x with $\|x\|_2 \le 2(1 - \frac{1}{\log n})r$, we will show that $x \in \mathrm{Conv}(\Omega_1 \cup \Omega_2)$. To do this, we let $t \stackrel{\text{def}}{=} \frac{\|P_A x\|_2^2}{\|x\|_2^2}$ and write x as the convex combination,

$$x = t\left(\frac{1}{t}P_A x\right) + (1 - t)\left(\frac{1}{1 - t}(I - P_A)x\right). \tag{3.2}$$

For the second term in (3.2), we have

$$\left(\frac{1}{1 - t}(I - P_A)x\right)^\top A^{-1}\left(\frac{1}{1 - t}(I - P_A)x\right) \le \frac{\|(I - P_A)x\|_2^2}{(1 - t)^2\lambda/(2^8 \log^2 n)} = \frac{2^8\|x\|_2^2 \log^2 n}{\lambda}$$

Since $\|x\|_2^2 \le 4r^2$ and $\lambda = \frac{n}{r^\alpha}$, we have

$$\frac{2^8\|x\|_2^2 \log^2 n}{\lambda} \le \frac{2^{10}r^2 \log^2 n}{\frac{n}{r^\alpha}} = \frac{2^{10}r^{2+\alpha} \log^2 n}{n} \le 1.$$

Hence,

$$\frac{1}{1 - t}(I - P_A)x \in \Omega_2. \tag{3.3}$$

For the first term in (3.2), we note that for any $\beta > 0$, we have

$$P_{\widehat{A}} \succeq \beta\lambda(\beta\lambda + \widehat{A})^{-1} - \frac{\beta}{1 + \beta}I \tag{3.4}$$

(To see this, we note that both sides have the same eigenspace and hence it follows from $1_{x \le \lambda} \ge \frac{\beta\lambda}{x + \beta\lambda} - \frac{\beta}{1+\beta}$.)

By Lemma A.2 with $\epsilon = 1$, we have

$$\widehat{A} \preceq 2A + O\left(\frac{\log^3 n \cdot \mathrm{Tr}A}{k}\right) \cdot I.$$

Using $\mathrm{Tr}A = O(r^2 n)$ (Lemma 3.1) and $k = \Omega(r^{2+\alpha} \log^5 n)$ (the algorithm description), we have

$$\widehat{A} \preceq 2A + \frac{n}{2^7 r^\alpha \log^2 n} \cdot I \preceq 2A + \frac{\lambda}{2^7 \log^2 n} \cdot I.$$

Putting this into (3.4), we have

$$P_{\widehat{A}} \succeq \beta\lambda(\beta\lambda + \frac{\lambda}{2^7 \log^2 n} + 2A)^{-1} - \frac{\beta}{1 + \beta}I.$$

On the range of P_A, we have $A \preceq \frac{\lambda}{2^8 \log^2 n}$ and hence

$$(P_A x)^\top P_{\widehat{A}}(P_A x) \geq \left(\frac{\beta\lambda}{\beta\lambda + \frac{\lambda}{2^6 \log^2 n}} - \frac{\beta}{1+\beta} \right) \|P_A x\|_2^2.$$

Using this, we have

$$\begin{aligned} f(x) &\stackrel{\text{def}}{=} \frac{1}{t^2 r^2} (P_A x)^\top (I - \frac{3}{4} P_{\widehat{A}})(P_A x) \\ &\leq \left(1 - \frac{3}{4} \frac{\beta\lambda}{\beta\lambda + \frac{\lambda}{2^6 \log^2 n}} + \frac{3}{4} \frac{\beta}{1+\beta} \right) \frac{\|x\|_2^2}{r^2} \\ &\leq \left(1 - \frac{3}{4} \frac{\beta}{\beta + \frac{1}{64 \log^2 n}} + \beta \right) \cdot 4 \left(1 - \frac{1}{\log n} \right)^2 \end{aligned}$$

where we used $\|x\|_2 \leq 2(1 - \frac{1}{\log n})r$ at the end. Putting $\beta = \frac{1}{8 \log n}$, we have

$$\begin{aligned} f(x) &\leq \left(1 - \frac{3}{4} \cdot \frac{8 \log n}{8 \log n + 1} + \frac{3}{4} \cdot \frac{1}{1 + 8 \log n} \right) \cdot 4 \left(1 - \frac{1}{\log n} \right) \\ &= \left(\frac{7 + 8 \log n}{8 \log n + 1} \right) \cdot \left(1 - \frac{1}{\log n} \right) \\ &\leq \left(1 + \frac{1}{\log n} \right) \cdot \left(1 - \frac{1}{\log n} \right) \\ &\leq 1 \end{aligned}$$

Hence, from the definition of Ω_1,we have

$$\frac{1}{t} P_A x \in \Omega_1. \tag{3.5}$$

Combining (3.5) and (3.3), we have

$$x \in \text{Conv}(\Omega_1 \cup \Omega_2) \subset T' K.$$

Hence, $T'K$ contains a ball of radius $2(1 - \frac{1}{\log n})r$. This completes the induction. □

With the bound on the inner radius and the p-norm of the covariance matrix, we can apply Theorem 2.7 and Lemma 2.8 to bound the mixing time and the complexity of the algorithm Isotropize.

Theorem 3.3. *The algorithm* Isotropize *applied to a well-rounded input convex body K satisfying $B(0,1) \subseteq K$ with $\mathbb{E}_K \|x\|^2 = O(n)$, with high probability, finds a transformation T using $\widetilde{O}(n^3 \psi^2)$ oracle calls, s.t. TK is 2-isotropic.*

Proof. Theorem 2.9 shows that it takes $\widetilde{O}(n^3)$ to get the first sample from a well-rounded body with Gaussian Cooling. (Note that a uniform point from K_t would be a very bad warm start for K_{2t}.) This gives a warm start for all subsequent steps.

Let A_j be the covariance matrix at the j-th iteration of the algorithm. By Lemma 3.1 with $\alpha = 0$, we have that

$$\|A_j\|_p \leq O(n r^{2/p} \log n).$$

In the j-th iteration the algorithm samples $k = c \cdot r_j^2 \log^5 n$ points. To bound the sampling cost, first note that for distribution q with KLS constant ψ_q, the complexity per sample is $\widetilde{O}(n^2 \psi_q^2 / r_j^2)$ since by Lemma 3.2, the algorithm maintains a ball of radius r_j inside the body. Suppose that the KLS constant for isotropic distributions is ψ. We choose p so that $\psi^2 = O(n^{1/p})$. By Lemma 2.8 with $\beta = 1/2p$, this implies that the KLS constant even for non-isotropic distributions with covariance A satisfies $\psi_q^2 = O(\|A\|_p \log n)$. Lemma 2.7 shows that the total complexity of the j-th iteration is at most

$$c \cdot r_j^2 \cdot \widetilde{O}\left(\frac{n^2 \cdot \|A_j\|_p}{r_j^2} \right) \leq \widetilde{O}\left(n^3 r_j^{2/p} \right) = \widetilde{O}\left(n^{3 + \frac{1}{p}} \right).$$

since we stop with $r^2 = O(n)$. By the definition of p, the complexity is $\widetilde{O}(n^3\psi^2)$. For the cost of computing covariance matrix and the mean, Lemma 2.5 shows that WHP $O(n)$ samples suffice for a constant factor estimate of the covariance. The total cost is

$$n \cdot \widetilde{O}(n^2 \cdot \|A\|_p r^{-2}) = r^2 \cdot \widetilde{O}(n^2 \cdot \|A\|_p r^{-2}) = \widetilde{O}\left(n^{3+\frac{1}{p}}\right)$$

where we used that $n = \widetilde{O}(r^2)$ at the end of the algorithm. □

3.2 Outer loop: the general case

For the general case, we first show that the next body is well-rounded after each iteration and hence satisfies the condition of the algorithm Isotropize. The proof is an adaptation of a proof from [22, Lemma 5.4].

Lemma 3.4. *Suppose that for a convex body K, the set $K \cap B(0,t)$ is 2-isotropic. Then $K \cap B(0,2t)$ is well-rounded.*

Proof. Let $K_t = K \cap B(0,t)$ and $K_{2t} = K \cap B(0,2t)$. Since K_t is 2-isotropic, K_t has mean 0 with $I \preceq \mathbb{E}_{x \in K_t} xx^\top \preceq 2I$. Hence, K_t contains the unit ball of radius 1. Let $t_{\text{small}} = 2\sqrt{n}$ and $K_{t_{\text{small}}} = K \cap B(0, t_{\text{small}})$. If $2t \le 3t_{\text{small}}$ then clearly $\mathbb{E}_{K_{2t}} \|x\|^2 \le (6\sqrt{n})^2 = 36n$ and we are done. Otherwise, $t > 3t_{\text{small}}/2$. Let λ such that

$$t = \lambda t_{\text{small}} + (1-\lambda)2t.$$

Note that $\lambda = \frac{2t-t}{2t-t_{\text{small}}} \le \frac{t}{2t} = \frac{1}{2}$. By the Brunn-Minkowski inequality, we have

$$\text{vol}(K_t)^{1/n} \ge \lambda \text{vol}(K_{t_{\text{small}}})^{1/n} + (1-\lambda)\text{vol}(K_{2t})^{1/n}. \tag{3.6}$$

Next, by the choice of t_{small}, since K_t is 2-isotropic, using Paouris' inequality (Lemma 2.4), we have

$$\text{vol}(K_t) \le \text{vol}(K_{t_{\text{small}}})(1 + \exp(-c\sqrt{n}))$$

for some constant c. Using this with (3.6), we have

$$\begin{aligned}
\text{vol}(K_{2t}) &\le \frac{1}{1-\lambda}\left(\text{vol}(K_t)^{1/n} - \lambda \text{vol}(K_{t_{\text{small}}})^{1/n}\right)^n \\
&\le \frac{1}{1-\lambda}\left(\text{vol}(K_{t_{\text{small}}})^{1/n}((1+\exp(-c\sqrt{n}))^{1/n} - \lambda \text{vol}(K_{t_{\text{small}}})^{1/n}\right)^n \\
&\le \text{vol}(K_{t_{\text{small}}})\frac{1}{1-\lambda}\left(1 + \frac{1}{n}\exp(-c\sqrt{n}) - \lambda\right)^n \\
&\le \text{vol}(K_{t_{\text{small}}})\left(1 + \frac{2}{n}\exp(-c\sqrt{n})\right)^n. \\
&\le \text{vol}(K_{t_{\text{small}}})\left(1 + 4\exp(-c\sqrt{n})\right).
\end{aligned}$$

Thus, the fraction of the volume of K_{2t} that lies outside a ball of radius t_{small} is exponentially small. Moreover, since K_t is 2-isotropic, all of K_t lies in a ball of radius $2\sqrt{n}$ and $0 \in K_t$. Moreover, $0 \in K_{2t}$ and since $K_{2t} = K \cap B(0,2t) \subseteq 2(K \cap B(0,t)) = 2K_t$, we have K_{2t} lies in a ball of radius $4n$.

$$\mathbb{E}_{K_{2t}} \|x\|^2 \le t_{\text{small}}^2 + (4n)^2 \cdot 4\exp(-c\sqrt{n}) = 5n.$$

□

Now we are ready to prove the main theorem about rounding.

Proof of Theorem 1.3. The initial body K_r is a ball, and at the end of the first iteration, the body is 2-isotropic. Let K_t be the convex body in some iteration. Assume that K_t is 2-isotropic. By Lemma 3.4 K_{2t} is well-rounded. Since $(K_t - x)$ contains a unit ball, so does $(K_{2t} - x)$. Then by Theorem 3.3 the inner loop to make K_{2t} near-isotropic takes $\tilde{O}(n^3\psi^2)$ oracle queries. Since the number of iterations of the outer loop is only $\log(R/r)$, the theorem follows. □

4 Polytope Volume

In this section, we consider the special case of convex polytopes. We assume that a polytope $P = \{x : Ax \leq b\}$ is given explicitly by $A \in R^{m\times n}, b \in R^m$. A naive implementation of the our general membership oracle algorithm would take $O(mn)$ arithmetic operations per oracle call, leading to a time bound of $\widetilde{O}(mn^{4.5})$, where the first term is the time complexity from the oracle queries and the second term is from the additional arithmetic operations per oracle query. This is now the number of arithmetic operations. This already matches the current best time complexity for polytopes using earlier volume algorithms and an amortization trick introduced in [23]. Here we give two significantly faster implementations. The first algorithm is based on a very simple application of fast matrix multiplication. Since fast matrix multiplication is practical only for very high dimension, we also give a different implementation extending the ideas of [23].

4.1 An efficient implementation using Fast Matrix Multiplication

We can replace every ball walk sampling step with the following algorithm.

Algorithm 3 Polytope Ball Walk

procedure PolytopeBallWalk($A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, x_0 \in \mathbb{R}^n, \delta > 0, k \in \mathbb{N}$)
  Assumption: $Ax_0 \leq b$.
  Sample points $z_1, \ldots, z_k$ from $B_n(0, \delta)$.
  $Z \leftarrow \begin{pmatrix} z_1 & z_2 & \cdots & z_k \end{pmatrix}, Y \leftarrow AZ$.
  $x \leftarrow x_0, y \leftarrow Ax_0$.
  for $i \leftarrow 1 : k$ **do**
    if $y + Y_i \leq b$. **then**
      $x \leftarrow x + z_i, y \leftarrow y + Y_i$.
    end if
  end for
  Return (x).
end procedure

Next, we show that this algorithm will improve the running time of the ball walk.

Lemma 4.1. *Given a polytope $\{x \in \mathbb{R}^n : Ax \leq b\}$ where $A \in \mathbb{R}^{m\times n}$, k steps of the ball walk in the polytope can be implemented in time $O(C(m, n, k))$ where $C(m, n, k)$ denotes the minimum number of arithmetic operations needed to multiply an $m \times n$ matrix by an $n \times k$ matrix.*

Proof. We can see that Algorithm 3 does the same computations as the ball walk algorithm with step size δ and run for k steps. The only difference is that we generate the random vectors from $B(0, \delta)$ first and do a preprocessing step to avoid multiplying Az_i at each step. So the resulting point will be the same as the ball walk algorithm and we reduce the running time of each step.

Let Z, Y, x, y be the same as in Algorithm 3. Generating k random vectors will cost $O(nk)$. For each step, we compute $y + Y_i$ in $O(m)$ time and $x + z_i$ in $O(n)$ time. So the total time taken is

$$O(nk) + C(m, n, k) + O((m + n)k) = O(C(m, n, k)).$$

□

We will apply this speedup to the polytope volume computation. From Theorem 3.3, in the j-th iteration of Isotropiztion, we run ball walk for $k = \widetilde{O}(n^{3+1/p})$ steps with the same transformation matrix T, and hence we can use the above algorithm to improve running time.

For getting the first sample, we use the Gaussian Cooling algorithm which in turn runs ball walk in phases with different ball radii and target distributions. However, the total number of ball walk steps in the Gaussian Cooling algorithm for a well-rounded body is $\tilde{O}(n^3)$ and we also know the maximum number of steps needed for each phase.

Therefore, we can construct the matrix Z in Algorithm 3 with $k = \tilde{O}(n^3)$. The following lemma gives the current best matrix bounds from [27], [13], and [12]. For $k > 0$, define the exponent of the rectangular matrix multiplication as follows:

$$\omega(1, 1, k) = \inf\{\tau \in \mathbb{R} \mid C(n, n, \lfloor n^k \rfloor) = O(n^\tau)\}.$$

Fact 4.2 ([27],[13],[12]). $\omega(1, 1, 3.5) = 4.68546$, $\omega(1, 1, 3) = 4.199712$.

To prove the first part of Theorem 1.5, use Algorithm 3. For $p = 2$, the total time is $\widetilde{O}(mn^{\omega(1,1,3.5)-1} + mn^{\omega(1,1,3)-1}) = \widetilde{O}(mn^{3.68546})$. If the KLS conjecture is true, then the total time is $\widetilde{O}(mn^{\omega(1,1,3)-1}) = \widetilde{O}(mn^{3.199712})$.

4.2 Implementation with an amortized Ball Walk

We can also speed up the algorithm by using an implementation of the ball walk for polytopes introduced by [23]. For a body in near-isotropic position, it requires only an expected $O(m)$ operations per step from a warm start.

The main idea of the algorithm is when moving from a point $x \in K$ to y, instead of checking the distance between the point y and every constraint (a_j, b_j) of the polytope at every step, we check the distance from a constraint at only a fraction of the steps. Firstly, with high probability, the step size along any constraint direction a_j is bounded by $O^*(\frac{\eta}{\sqrt{n}}) = O^*(\frac{1}{n})$. This can be used to determine which constraint to check at the current step by maintaining a probabilistic lower bound on $h_j(x) = b_j - a_j^\top x$, the distance between the current point and the j-th constraint. Then, a constraint is checked at the current step only if the corresponding lower bound is $O(\frac{1}{n})$. The last ingredient of their analysis is an anti-concentration bound which states that the probability of the distance between a uniform point from an isotropic convex body and a hyperplane being less than ϵ is $O(\epsilon)$. This ensures that we need to check only $O(m/n)$ constraints in expectation at every step for an isotropic polytope.

However, in our algorithm, the intermediate polytopes are not near-isotropic but well-rounded. The aforementioned probability is inversely proportional to the standard deviation of the body in that direction, which can be as small as $\frac{r}{n}$. We extend the anti-concentration bound to convex polytopes containing a ball of radius r, i.e., the probability that a uniform point from such a polytope is at a distance less than ϵ from any hyperplane is $O(\frac{n\epsilon}{r})$ which ensures that we need to check only $O\left(\frac{m}{r}\right)$ constraints in expectation at every step of this modified ball walk. Since the number of points sampled in the j-th iteration of Isotropize is $\tilde{O}(r^2)$, this still gives runtime savings because we sample fewer points for smaller values of r and save asymptotically.

Algorithm 4 Amortized Ball Walk for Polytopes[23]

procedure MODIFIED BALL WALK($A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, x_0 \in \mathbb{R}^n, \rho > 0, \delta > 0, N \in \mathbb{N}$)
  Assumption: $Ax_0 \le b$ and $\|x\|_2 < \rho$.
  $d_j = b_j - a_j^\top x_0$ for $j \in [m]$.
  Sort d_j's in ascending order in an array H.
  for $i \leftarrow 0 : N$ **do**
    $y_{i+1} \sim B(x_i, \delta)$.
    if $\|y_{i+1}\|_2 > \rho$ **then**
      $x_{i+1} = x_i$
    end if
    Let j^* be the largest j such that $d_{j^*} \le \alpha \cdot \frac{\eta}{\sqrt{n}} \cdot i$ ▷ $\alpha = 4\log\left(\frac{2mN}{\epsilon}\right)$
    for $k = 1 \to j^*$ **do**
      $d_k = b_k - a_k^\top y_{i+1}$.
      if $d_k < 0$ **then**
        $x_{i+1} = x_i$.
      end if
      Insert $d_k + \alpha \cdot \frac{\eta}{\sqrt{n}} \cdot i$ in H in sorted order.
    end for
    $x_{i+1} = y_{i+1}$.
  end for
  Return x_N.
end procedure

Before proving the anti-concentration lemma, we need a result relating the radius of a ball inside a convex body K and the minimum eigenvalue of the covariance matrix of K.

Lemma 4.3. *For any convex body in $\mathbb{R}^n$ containing a ball of radius r, the minimum eigenvalue of the covariance matrix is at least $\frac{r^2}{(n+1)^2}$.*

Proof. Let $\Sigma = \mathbb{E}_K[(x-\mu)(x-\mu)^\top]$ and λ_1 be the minimum eigenvalue of Σ. Then,

$$\begin{aligned}\lambda_1 &= \min_{z\neq 0, \|z\|_2=1} z^\top \Sigma z \\ &= \min_{z\neq 0, \|z\|_2=1} z^\top \mathbb{E}_K[(x-\mu)(x-\mu)^\top] z \\ &= \min_{z\neq 0, \|z\|_2=1} \mathbb{E}_K[(z^\top(x-\mu))^2] \\ &= \mathbb{E}_K[(y^\top(x-\mu))^2].\end{aligned}$$

Wlog, let $\mu = 0$ and let v be the point in K which maximizes $|y^\top v|$. Let $\phi(u)$ denote the largest real number t with $v + tu \in K$. Then we have,

$$vol(K) = \int_{\partial B}\int_0^{\phi(u)} t^{n-1} dt du = \frac{1}{n}\int_{\partial B} \phi(u)^n du$$

and

$$\begin{aligned}\sigma^2 &= \frac{1}{vol(K)}\int_K (y^T x) dx \\ &= \frac{1}{vol(K)}\int_{\partial B}\int_0^{\phi(u)} t^{n-1}(y^T(v+tu))^2 dt du \\ &= \frac{1}{vol(K)}\int_{\partial B}\left(\frac{\phi(u)^n}{n}(y^T v)^2 + 2\frac{\phi(u)^{n+1}}{n+1}(y^T v)(y^T u) + \frac{\phi(u)^{n+2}}{n+2}(y^T u)^2\right) du \\ &= \frac{1}{vol(K)}\int_{\partial B}\frac{\phi(u)^n}{n(n+1)^2}(y^T v)^2 + \frac{\phi(u)^n}{n}\left(\sqrt{\frac{n}{n+2}}\phi(u) y^T u + \frac{\sqrt{n(n+2)}}{n+1} y^T v\right)^2 du \\ &\geq \frac{1}{vol(K)}\int_{\partial B}\frac{\phi(u)^n}{n(n+1)^2}(y^T v)^2 du \\ \sigma^2 &= \frac{1}{(n+1)^2}(y^T v)^2\end{aligned}$$

Therefore,

$$r \leq |y^T v| \leq (n+1)\sigma.$$

□

Lemma 4.4 (Anti-concentration)**.** *Suppose that the uniform distribution π_K on convex body K has covariance matrix $\Sigma \succeq \sigma^2 I$ and that $X \sim \pi_K$ is a random vector uniformly distributed on K. Let H be any codimension-1 hyperplane. Then we have*

$$\Pr[\text{dist}(X, H) \leq \epsilon] \leq \frac{2\epsilon}{\sigma} \quad \forall \epsilon > 0.$$

Proof. Let the equation for H be $ax = b$ where $a \in \mathbb{R}^{1\times n}, \|a\|_2 = 1$ and $b \in \mathbb{R}$. Let $X \sim \pi_K$ is a random vector uniformly distributed on K. Let the distribution of aX be denoted by π_K^a. π_K^a is logconcave with variance $a\Sigma a^\top = \sigma_1^2 \geq \sigma^2$. Let $\hat{\pi}_K^a$ denote the distribution of $\frac{1}{\sigma_1}aX$. Then $\hat{\pi}_K^a$ is isotropic and logconcave. If x^* is the maximizer of $\hat{\pi}_K^a$, then

$$\hat{\pi}_K^a(x^*) \leq 1.$$

For any $\epsilon > 0$,

$$\begin{aligned}\Pr\left[|aX - b| \le \epsilon\right] &= \Pr\left[\frac{|aX-b|}{\sigma_1} \le \frac{\epsilon}{\sigma_1}\right] \\ &= \int_{\frac{b-\epsilon}{\sigma}}^{\frac{b+\epsilon}{\sigma}} \hat{\pi}_K^a(x)dx \le \int_{\frac{b-\epsilon}{\sigma}}^{\frac{b+\epsilon}{\sigma}} \hat{\pi}_K^a(x^*)dx \\ &\le \int_{\frac{b-\epsilon}{\sigma}}^{\frac{b+\epsilon}{\sigma}} dx = \frac{2\epsilon}{\sigma_1} \le \frac{2\epsilon}{\sigma}.\end{aligned}$$

□

Lemma 4.5 (Frequency of constraint checking)**.** *For a polytope K with covariance matrix $\Sigma \ge \sigma^2 I$ and Ball Walk with step size η and tolerance parameter α, suppose that the initial point X_0 is β-warm with respect to the uniform distribution for some $\beta > 0$. Let N_j be the number of steps (excluding the first step) of the modified Ball Walk Markov chain at which the algorithm checks inequality (A_j, b_j) and let N be the number of Markov chain steps. Let $F_j := \frac{N_j}{N}$. Then,*

$$\mathbb{E}[F_j] = 16\frac{\alpha\beta\eta}{\sqrt{n}\sigma}(1 + 2\log(n)) + \frac{\beta\epsilon}{N}.$$

This lemma follows directly from Lemma 4.5 in [23] using the anti-concentration bound from Lemma 4.4.

Lemma 4.6. *The algorithm* Isotropization *applied to a well-rounded input convex polytope K satisfying $B(0,1) \subseteq K$ with $\mathbb{E}_K \|x\|^2 = O(n)$, with high probability, finds a transformation T using $\tilde{O}(n^3\psi^2)$ oracle calls and $\tilde{O}(mn^4)$ arithmetic operations, s.t. TK is 2-isotropic.*

Proof. It takes $\tilde{O}(n^3)$ steps of the Gaussian Cooling algorithm to get the first sample from a well-rounded body where every step can be implemented in $O(mn)$ arithmetic operations. Let A_j and r_j be the covariance matrix and the inner radius at the j-th iteration of the algorithm respectively. From Theorem 3.3, the number of oracle calls in the j-th iteration of the algorithm is $\tilde{O}(n^3 r_j^{\frac{2}{p}})$. So, the total complexity of j-th iteration is at most

$$\tilde{O}(n^3 r_j^{\frac{2}{p}}) \cdot \tilde{O}(mnr_j^{-1}) \le \tilde{O}(mn^4 r^{\frac{2}{p}-1}).$$

Since we stop before $2^{10}r^2 \log n = n$, the complexity becomes

$$\tilde{O}\left(mn^4 + mn^{3.5+\frac{1}{p}}\right)$$

which is $\tilde{O}(mn^4)$ for $p = 2$. □

Given a convex polytope K, use Algorithm 4 to make it near-isotropic and then use the Gaussian Cooling algorithm [4] to compute volume. The second part of Theorem 1.5 (without fast matrix multiplication) follows.

References

[1] R. Adamczak, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, 23:535–561, 2010.

[2] D. Applegate and R. Kannan. Sampling and integration of near log-concave functions. In *STOC*, pages 156–163, 1991.

[3] B. Cousins and S. Vempala. Bypassing KLS: Gaussian cooling and an $O^*(n^3)$ volume algorithm. In *STOC*, pages 539–548, 2015.

[4] Ben Cousins and Santosh S. Vempala. Gaussian cooling and $o^{*(n^3)}$ algorithms for volume and gaussian volume. *SIAM J. Comput.*, 47(3):1237–1273, 2018.

[5] M. E. Dyer and A. M. Frieze. Computing the volume of a convex body: a case where randomness provably helps. In *Proc. of AMS Symposium on Probabilistic Combinatorics and Its Applications*, pages 123–170, 1991.

[6] M. E. Dyer, A. M. Frieze, and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. In *STOC*, pages 375–381, 1989.

[7] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

[8] Hulda S Haraldsdöttir, Ben Cousins, Ines Thiele, Ronan M.T Fleming, and Santosh Vempala. CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33(11):1741–1743, 01 2017.

[9] Haotian Jiang, Yin Tat Lee, and Santosh S Vempala. A generalized central limit conjecture for convex bodies. *arXiv preprint arXiv:1909.13127*, 2019.

[10] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

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

[12] François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th international symposium on symbolic and algebraic computation*, pages 296–303, 2014.

[13] François Le Gall and Florent Urrutia. Improved rectangular matrix multiplication using powers of the coppersmith-winograd tensor. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1029–1046. SIAM, 2018.

[14] Yin Tat Lee and Santosh S Vempala. Convergence rate of riemannian hamiltonian monte carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1115–1121. ACM, 2018.

[15] Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. *CoRR*, abs/1612.01507, 2016.

[16] Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *Proc. of IEEE FOCS*, 2017.

[17] L. Lovász. How to compute the volume? *Jber. d. Dt. Math.-Verein, Jubiläumstagung 1990*, pages 138–151, 1990.

[18] L. Lovász and M. Simonovits. Mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *ROCS*, pages 482–491, 1990.

[19] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. In *Random Structures and Alg.*, volume 4, pages 359–412, 1993.

[20] L. Lovász and S. Vempala. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comput. Syst. Sci.*, 72(2):392–417, 2006.

[21] L. Lovász and S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307–358, 2007.

[22] László Lovász and Santosh Vempala. Simulated annealing in convex bodies and an o*(n4) volume algorithm. *Journal of Computer and System Sciences*, 72(2):392–417, 2006.

[23] Oren Mangoubi and Nisheeth K Vishnoi. Faster polytope rounding, sampling, and volume computation via a sub-linear ball walk. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1338–1357. IEEE, 2019.

[24] G. Paouris. Concentration of mass on convex bodies. *Geometric and Functional Analysis*, 16:1021–1049, 2006.

[25] Nikhil Srivastava, Roman Vershynin, et al. Covariance estimation for distributions with 2+eps moments. *The Annals of Probability*, 41(5):3081–3111, 2013.

[26] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

[27] Jan van den Brand, Danupon Nanongkai, and Thatchaphol Saranurak. Dynamic matrix inverse: Improved algorithms and matching conditional lower bounds. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 456–480. IEEE, 2019.

A Empirical Covariance Matrix with Sublinear Sample Complexity

To bound the error of the empirical covariance matrix A, we use the following matrix Chernoff bound.

Lemma A.1 (Matrix Bernstein [26, Theorem 6.1])**.** *Consider a finite sequence $\{X_i\}$ of independent, random, self-adjoint matrices of dimension n. Assume that $\mathbb{E}X_i = 0$ and $\|X_i\|_{\mathrm{op}} \leq R$ almost surely. The, for all $t \geq 0$, we have*

$$\mathbb{P}\left(\|\sum_i X_i\|_{\mathrm{op}} \geq t\right) \leq 2n \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right)$$

where $\sigma^2 \stackrel{\mathrm{def}}{=} \|\sum_i \mathbb{E}(X_i^2)\|_{\mathrm{op}}$.

Lemma A.2. *Let p be a logconcave density in $\mathbb{R}^n$ with covariance A. Let $\widehat{A} \leftarrow \frac{1}{k}\sum_{i=1}^k (x_i - \widehat{x})(x_i - \widehat{x})^\top$ where $\widehat{x} = \frac{1}{k}\sum_{i=1}^k x_{i+k}$ and x_i are independent samples from p. With probability $1 - 1/n^{O(1)}$, for any $0 \leq \epsilon \leq 1$, we have*

$$(1-\epsilon)A - O\left(\frac{\log^3 n \mathrm{Tr}A}{\epsilon k}\right)\cdot I \preceq \widehat{A} \preceq (1+\epsilon)A + O\left(\frac{\log^3 n \cdot \mathrm{Tr}A}{\epsilon k}\right)\cdot I.$$

Remark. By a more careful tail analysis, one can get a bound of $O(\frac{\log n \cdot \mathrm{Tr}A}{\epsilon k})\cdot I$ and we speculate that the tight bound for the additive term might be $O(\frac{\mathrm{Tr}A}{\epsilon k})\cdot I$.

Proof. Let $\lambda \geq 0$ to some constant to be determined. By shifting the distribution, we can assume p has mean 0. Let $\widetilde{p}$ be the distribution given by p restricted to the ball

$$\{\|x\|_{(\lambda A + I)^{-1}} \leq 3s \cdot \sqrt{\mathrm{Tr}(\lambda A + I)^{-\frac{1}{2}} A (\lambda A + I)^{-\frac{1}{2}}}\}$$

for some $s = \Theta(\log n)$. Using the fact that p has mean 0 and the fact that $(\lambda A + I)^{-\frac{1}{2}} A (\lambda A + I)^{-\frac{1}{2}} \preceq A$, Lemma 2.3 shows that

$$\mathbb{P}_{x\sim p}(\|x\|_{(\lambda A+I)^{-1}} \leq 3s\cdot\sqrt{\mathrm{Tr}A}) \geq 1 - \frac{1}{n^{\Theta(1)}}.$$

Let A_i be the random matrices $(x_i - \widehat{x})(x_i - \widehat{x})^\top$ with x_i is sampled from p and $\widetilde{A}_i$ be the random matrices $(\widetilde{x}_i - \widehat{x})(\widetilde{x}_i - \widehat{x})^\top$ with $\widetilde{x}_i$ is sampled from $\widetilde{p}$. Note that when $k = \Omega(n)$, we have that WHP $\frac{1}{2}A \preceq \widehat{A} \preceq 2A$ 2.5. Hence, we can assume $k = O(n)$. We couple two matrices together such that $\widetilde{A}_i = A_i$ for $i = 1, 2, \cdots k$ with probability $1 - ke^{-s} = 1 - \frac{1}{n^{\Theta(1)}}$. Let $X_i = \frac{1}{k}(\lambda A + I)^{-\frac{1}{2}}(A_i - \mathbb{E}A_i)(\lambda A + I)^{-\frac{1}{2}}$ and $\widetilde{X}_i = \frac{1}{k}(\lambda A + I)^{-\frac{1}{2}}(\widetilde{A}_i - \mathbb{E}\widetilde{A}_i)(\lambda A + I)^{-\frac{1}{2}}$ for some λ to be chosen where the expectation $\mathbb{E}$ is conditional on $\hat{x}$. Note that

$$\begin{aligned}\mathbb{P}(\left\|(\lambda A + I)^{-\frac{1}{2}}(\widehat{A} - \mathbb{E}\widehat{A})(\lambda A + I)^{-\frac{1}{2}}\right\|_{\mathrm{op}} \geq t) &= \mathbb{P}(\|\sum_i X_i\|_{\mathrm{op}} \geq t) \\ &\leq \mathbb{P}(\|\sum_i \widetilde{X}_i\|_{\mathrm{op}} \geq t) + \frac{1}{n^{\Theta(1)}}. \end{aligned} \tag{A.1}$$

Hence, it suffices to study $\widetilde{X}_i$.

We will apply Lemma A.1. For the sup norm bound, note that

$$\begin{aligned}\|\widetilde{X}_i\|_{\mathrm{op}} &\le \frac{1}{k}\left(\|(\lambda A+I)^{-\frac{1}{2}}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}}\|_{\mathrm{op}} + \|(\lambda A+I)^{-\frac{1}{2}}\mathbb{E}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}}\|_{\mathrm{op}}\right)\\ &\le \frac{1}{k}\left(\|\widetilde{x}_i-\widehat{x}\|^2_{(\lambda A+I)^{-1}} + \mathbb{E}\|\widetilde{x}_i-\widehat{x}\|^2_{(\lambda A+I)^{-1}}\right)\\ &\le \frac{1}{k}\left(2\|\widetilde{x}_i\|^2_{(\lambda A+I)^{-1}} + 2\mathbb{E}\|\widetilde{x}_i\|^2_{(\lambda A+I)^{-1}} + 4\|\widehat{x}\|^2_{(\lambda A+I)^{-1}}\right)\\ &\le \frac{4}{k}\left(9s^2\mathrm{Tr}A + \|\widehat{x}\|^2_{(\lambda A+I)^{-1}}\right) \le R\end{aligned}$$

where the second inequality follows from the fact that non-zero eigenvalues of Y^TY and YY^T are the same for any rectangular matrix Y, and

$$R \stackrel{\mathrm{def}}{=} \frac{36}{k}\left(s^2\mathrm{Tr}A + \|\widehat{x}\|^2_{(\lambda A+I)^{-1}}\right) \tag{A.2}$$

For the variance bound, note that $(\lambda A+I)^{-\frac{1}{2}}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}} \preceq kR$ and hence (using again that $(a-b)^2 \le 2(a^2+b^2)$),

$$\begin{aligned}\mathbb{E}\widetilde{X}_i^2 &\preceq \mathbb{E}\frac{2}{k^2}(((\lambda A+I)^{-\frac{1}{2}}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}})^2 + (\mathbb{E}(\lambda A+I)^{-\frac{1}{2}}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}})^2)\\ &\preceq \frac{2kR}{k^2}\mathbb{E}((\lambda A+I)^{-\frac{1}{2}}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}} + \mathbb{E}((\lambda A+I)^{-\frac{1}{2}}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}})\\ &= \frac{4R}{k}\mathbb{E}(\lambda A+I)^{-\frac{1}{2}}\widetilde{A}_i(\lambda A+I)^{-\frac{1}{2}}\\ &\preceq \frac{8R}{k}\mathbb{E}(\lambda A+I)^{-\frac{1}{2}}A_i(\lambda A+I)^{-\frac{1}{2}}\\ &= \frac{8R}{k}(\frac{A}{\lambda A+I} + (\lambda A+I)^{-\frac{1}{2}}\widehat{x}\widehat{x}^\top(\lambda A+I)^{-\frac{1}{2}})\end{aligned}$$

where we used that $\frac{d\widetilde{p}}{dp} \le 2$ in the last inequality. Hence, we have

$$\sigma^2 \stackrel{\mathrm{def}}{=} \|\sum_i \mathbb{E}\widetilde{X}_i^2\|_{\mathrm{op}} \le 8R(\|\frac{A}{\lambda A+I}\|_{\mathrm{op}} + \|\widehat{x}\|^2_{(\lambda A+I)^{-1}}).$$

Apply Lemma A.1, with probability $1-\frac{1}{n^{\Theta(1)}}$, we have

$$\begin{aligned}\|\sum_i \widetilde{X}_i\|_{\mathrm{op}} &\lesssim \sigma\sqrt{\log n} + R\log n\\ &\lesssim \sqrt{(\|\frac{A}{\lambda A+I}\|_{\mathrm{op}} + \|\widehat{x}\|^2_{(\lambda A+I)^{-1}})}\sqrt{R\log n} + R\log n\end{aligned}$$

Using the value of R from equation A.2, for any $c \ge 1$, we get

$$\begin{aligned}\|\sum_i \widetilde{X}_i\|_{\mathrm{op}} &\lesssim \frac{1}{c}\|\frac{A}{\lambda A+I}\|_{\mathrm{op}} + \frac{1}{c}\|\widehat{x}\|^2_{(\lambda A+I)^{-1}} + \frac{c}{k}(\log^2 n\cdot\mathrm{Tr}A + \|\widehat{x}\|^2_{(\lambda A+I)^{-1}})\log n\\ &\lesssim \frac{1}{c\lambda} + (\frac{1}{c}+\frac{c\log n}{k})\cdot\|\widehat{x}\|^2_{(\lambda A+I)^{-1}} + \frac{c\log^3 n}{k}\cdot\mathrm{Tr}A\end{aligned}$$

Using this and equation (A.1) , we have

$$\left\|(\lambda A+I)^{-\frac{1}{2}}(\widehat{A}-\mathbb{E}\widehat{A})(\lambda A+I)^{-\frac{1}{2}}\right\|_{\mathrm{op}} \lesssim \frac{1}{c\lambda} + (\frac{1}{c}+\frac{c\log n}{k})\cdot\|\widehat{x}\|^2_{(\lambda A+I)^{-1}} + \frac{c\log^3 n}{k}\cdot\mathrm{Tr}A.$$

Finally, we note that $\widehat{x}$ follows a logconcave distribution with mean 0 and covariance matrix $\frac{1}{k}A$. By Lemma 2.3, we have that

$$\|\widehat{x}\|_{(\lambda A+I)^{-1}} \lesssim \log n\sqrt{\frac{1}{k}\mathrm{Tr}A}$$

with probability $1-1/n^{O(1)}$. This gives

$$\left\|(\lambda A+I)^{-\frac{1}{2}}(\widehat{A}-\mathbb{E}\widehat{A})(\lambda A+I)^{-\frac{1}{2}}\right\|_{\mathrm{op}} \lesssim \frac{1}{c\lambda}+(\frac{\log^2 n}{ck}+\frac{c\log^3 n}{k})\cdot \mathrm{Tr}A$$
$$\lesssim \frac{1}{c\lambda}+\frac{c\log^3 n}{k}\cdot \mathrm{Tr}A$$

Hence, we have

$$|\widehat{A}-\mathbb{E}\widehat{A}| \lesssim \frac{1}{c}A+\frac{1}{c\lambda}I+\frac{c\log^3 n}{k}\cdot \mathrm{Tr}A\cdot(\lambda A+I).$$

Taking $c=\Theta(\epsilon^{-1})$, $\lambda=\frac{k}{c^2\log^3 n\mathrm{Tr}A}$, we have

$$|\widehat{A}-\mathbb{E}\widehat{A}| \preceq \epsilon A+O(\frac{\log^3 n\mathrm{Tr}A}{\epsilon k})\cdot I.$$

The result follows using $\mathbb{E}\widehat{A}=A$. □

B Proof of the anisotropic KLS bound

Consider the following stochastic localization process.

Definition B.1. For a logconcave density p, we define the following stochastic differential equation:

$$c_0=0, \quad dc_t=dW_t+\mu_t dt, \tag{B.1}$$

where the probability density p_t, the mean μ_t and the covariance A_t are defined by

$$p_t(x)=\frac{e^{c_t^\top x-\frac{t}{2}\|x\|_2^2}p(x)}{\int_{\mathbb{R}^n}e^{c_t^\top y-\frac{t}{2}\|y\|_2^2}p(y)dy}, \quad \mu_t=\mathbb{E}_{x\sim p_t}x, \quad A_t=\mathbb{E}_{x\sim p_t}(x-\mu_t)(x-\mu_t)^\top.$$

The following lemma shows that one can upper bound the expansion ψ_p by upper bounding $\|A_t\|_{\mathrm{op}}$:

Lemma B.2 ([15, Lemma 31 in ArXiv ver 3]). *Given a logconcave density p, let A_t be as in Definition B.1 using initial density p. Suppose there is a $T>0$ such that*

$$\mathbb{P}\left(\int_0^T \|A_s\|_{\mathrm{op}}\,ds\le\frac{1}{64}\right)\ge\frac{3}{4}$$

Then, we have $\psi_p=O\left(T^{-1/2}\right)$.

To bound $\|A_t\|_{\mathrm{op}}$, we need a basic stochastic calculus rule about A_t.

Lemma B.3 ([15, Lemma 27 in arXiv ver 3]). *The covariance A_t satisfies $dA_t=\int_{\mathbb{R}^n}(x-\mu_t)(x-\mu_t)^\top\left((x-\mu_t)^\top dW_t\right)p_t(x)dx-A_t^2dt$.*

We bound $\|A_t\|_{\mathrm{op}}$ using the potential $\Phi_t\stackrel{\mathrm{def}}{=}\mathrm{Tr}A_t^q$. We will use fractional q and there is no simple formula of $d\Phi_t$. The upper bound on $d\Phi_t$ is known (see [7]). For completeness, we give an alternative proof here. Our proof relies on the following lemma about the smoothness of the trace function.

Lemma B.4 (Prop 3.1 in 0809.0813). *Let f be a twice differentiable function on (α,β) such that for some $\theta_\pm,\mu_\pm\in\mathbb{R}$, for all $\alpha<a<b<\beta$, we have*

$$\theta_-\frac{f''(a)+f''(b)}{2}+\mu_-\le\frac{f'(b)-f'(a)}{b-a}\le\theta_+\frac{f''(a)+f''(b)}{2}+\mu_+.$$

Then, for any matrix X with eigenvalues lies between (α,β), we have

$$\theta_-\mathrm{Tr}(f''(X)H^2)+\mu_-\mathrm{Tr}H^2\le\left.\frac{\partial^2\mathrm{Tr}f(X)}{\partial X^2}\right|_{H,H}\le\theta_+\mathrm{Tr}(f''(X)H^2)+\mu_+\mathrm{Tr}H^2.$$

Now, we can use this to upper bound the derivative of Φ_t.

Lemma B.5. *For any $q > 1$, we have that*

$$\begin{aligned} d\Phi_t \leq & q\mathbb{E}_{x\sim p_t}(x-\mu_t)^\top A_t^{q-1}(x-\mu_t)(x-\mu_t)^\top dW_t \\ & + q(q-1)\cdot \mathbb{E}_{x,y\sim p_t}((x-\mu_t)^\top(y-\mu_t))^2(x-\mu_t)^\top A_t^{q-2}(y-\mu_t)dt. \end{aligned}$$

Proof. By Lemma B.3, we have that

$$\begin{aligned} dA_t &= \int_{\mathbb{R}^n}(x-\mu_t)(x-\mu_t)^\top\left((x-\mu_t)^\top dW_t\right)p_t(x)dx - A_t^2 dt \\ &= \sum_i Z_i\cdot dW_{t,i} - A_t^2 dt \end{aligned}$$

with $Z_i \stackrel{\text{def}}{=} \mathbb{E}_{x\sim p_t}(x-\mu_t)(x-\mu_t)^\top(x-\mu_t)_i$. By Itô's formula, we have that

$$d\Phi_t = \left.\frac{\partial \mathrm{Tr}A_t^q}{\partial A_t}\right|_{dA_t} + \frac{1}{2}\sum_i \left.\frac{\partial^2 \mathrm{Tr}A_t^q}{\partial A_t^2}\right|_{Z_i,Z_i} dt.$$

For the first-order term, we have

$$\begin{aligned} \left.\frac{\partial \mathrm{Tr}A_t^q}{\partial A_t}\right|_{dA_t} &= q\cdot \mathrm{Tr}A_t^{q-1}dA_t \\ &= q\mathbb{E}_{x\sim p_t}(x-\mu_t)^\top A_t^{q-1}(x-\mu_t)(x-\mu_t)^\top dW_t - q\mathrm{Tr}A_t^{q+1}. \end{aligned}$$

For the second-order term, we use Lemma B.4 with $f(x) = x^q$. It is easy to see that $\frac{f'(b)-f'(a)}{b-a} \leq f''(a) + f''(b)$. Hence, we can use Lemma B.4 with $\theta_+ = 2$ and $\mu_+ = 0$ to get

$$\left.\frac{\partial^2 \mathrm{Tr}A_t^q}{\partial A_t^2}\right|_{Z_i,Z_i} \leq 2q(q-1)\mathrm{Tr}(A_t^{q-2}Z_i^2).$$

Combining the first and second-order terms, we have

$$\begin{aligned} d\Phi_t \leq & q\mathbb{E}_{x\sim p_t}(x-\mu_t)^\top A_t^{q-1}(x-\mu_t)(x-\mu_t)^\top dW_t \\ & + q(q-1)\cdot\sum_i \mathrm{Tr}(A_t^{q-2}Z_i^2)dt \\ = & q\mathbb{E}_{x\sim p_t}(x-\mu_t)^\top A_t^{q-1}(x-\mu_t)(x-\mu_t)^\top dW_t \\ & + q(q-1)\cdot\mathbb{E}_{x,y\sim p_t}((x-\mu_t)^\top(y-\mu_t))^2(x-\mu_t)^\top A_t^{q-2}(y-\mu_t)dt. \end{aligned}$$

□

To analyze the stochastic inequality for $d\Phi_t$, we introduce a 3-Tensor.

Definition B.6 (3-Tensor)**.** For an isotropic logconcave distribution p in $\mathbb{R}^n$ and symmetric matrices A, B and C, define

$$T(A,B,C) = \sup_{\text{isotropic log-concave } p} \mathbb{E}_{x,y\sim p}(x^\top Ay)(x^\top By)(x^\top Cy).$$

Using the definition above, we can simplify the upper bound of Φ_t as follows:

$$d\Phi_t \leq q\mathbb{E}_{x\sim p_t}(x-\mu_t)^\top A_t^{q-1}(x-\mu_t)(x-\mu_t)^\top dW_t + q(q-1)\cdot T(A_t^{q-1}, A_t, A_t). \tag{B.2}$$

To further bounding $d\Phi_t$, we need following inequalities about logconcave distributions:

Lemma B.7 ([15, Lemma 32 in arXiv ver 3])**.** *Let p be a logconcave density with mean μ and covariance A. For any positive semi-definite matrix C, we have that*

$$\|\mathbb{E}_{x\sim p}(x-\mu)(x-\mu)^\top C(x-\mu)\|_2 \lesssim \|A\|_{\mathrm{op}}^{1/2}\mathrm{Tr}(A^{1/2}CA^{1/2}).$$

Lemma B.8 ([9, Lemma 41]). *For any $0 \le \alpha \le 1$, $A \succeq 0$, and $C \succeq 0$, we have $T(A^\alpha, A^{1-\alpha}, C) \le T(A, I, C)$.*

Lemma B.9 ([9, Lemma 40]). *Suppose that $\psi_n \le \alpha n^\beta$ for all n with some fixed $0 \le \beta \le \frac{1}{2}$ and $\alpha \ge 1$. For any two symmetric matrices A and B, we have*

$$T(A, B, I) \lesssim \alpha^2 \log n \cdot \|A\|_1 \cdot \|B\|_{1/(2\beta)}.$$

Using the lemmas above, we have the following:

Lemma B.10. *Suppose that $\psi_n = O(n^\beta)$ for all n with some fixed $0 \le \beta \le \frac{1}{2}$. For any $q = \frac{1}{2\beta}$, we have $d\Phi_t \le \delta_t dt + v_t^\top dW_t$ with $\delta_t \lesssim \frac{\log n}{\beta^2} \cdot \Phi_t^{1+2\beta}$ and $\|v_t\|_2 \lesssim \frac{1}{\beta} \cdot \Phi_t^{1+\beta}$.*

Proof. From (B.2), we have

$$\begin{aligned}
\delta_t &\le q(q-1) \cdot T(A_t^{q-1}, A_t, A_t) \\
&\le q(q-1) \cdot T(A_t^q, A_t, I) \\
&\lesssim q(q-1) \cdot \log n \cdot \mathrm{Tr} A_t^q \cdot (\mathrm{Tr} A_t^{1/(2\beta)})^{2\beta} \\
&\lesssim \frac{\log n}{\beta^2} \cdot \Phi_t^{1+2\beta}
\end{aligned}$$

where we used Lemma B.8 in the second inequality and Lemma B.9 at the end. From (B.2) and Lemma B.7, we have

$$\begin{aligned}
\|v_t\|_2 &\stackrel{\text{def}}{=} q \|\mathbb{E}_{x \sim p_t} (x - \mu_t)^\top A_t^{q-1} (x - \mu_t)(x - \mu_t)\|_2 \\
&\lesssim q \|A_t\|_{\mathrm{op}}^{1/2} \cdot \mathrm{Tr} A_t \lesssim \frac{1}{\beta} \cdot \Phi_t^{1+\beta}.
\end{aligned}$$

□

Finally, we use the following lemma to bound stochastic inequality.

Lemma B.11 ([9, Lemma 35]). *Let Φ_t be a stochastic process such that $\Phi_0 \le \frac{U}{2}$ and $d\Phi_t = \delta_t dt + v_t^\top dW_t$. Let $T > 0$ be some fixed time, $U > 0$ be some target upper bound, and f and g be some auxiliary functions such that for all $0 \le t \le T$*

1. *$\delta_t \le f(\Phi_t)$ and $\|v_t\|_2 \le g(\Phi_t)$,*
2. *Both $f(\cdot)$ and $g(\cdot)$ are non-negative non-decreasing functions,*
3. *$f(U) \cdot T \le \frac{U}{8}$ and $g(U) \cdot \sqrt{T} \le \frac{U}{8}$.*

Then, we have that $\mathbb{P}\left[\max_{t \in [0,T]} \Phi_t \ge U\right] \le 0.01$.

Using this lemma, we obtain the following bound.

Theorem B.12. *Suppose that $\psi_n = O(n^\beta)$ for all n with some fixed $0 \le \beta \le \frac{1}{2}$. Then, for any logconcave distribution p with covariance matrix A, we have that*

$$\psi_p \lesssim \frac{\sqrt{\log n}}{\beta} \|A\|_{\frac{1}{2\beta}}^{1/2}.$$

Proof. Consider the stochastic process starts with $p_0 = p$. Let $\Phi_t = \mathrm{Tr} A_t^{1/2\beta}$. Lemma B.10 shows that $d\Phi_t = \delta_t dt + v_t^\top dW_t$ with $\delta_t \le \frac{\log n}{\beta^2} \cdot \Phi_t^{1+2\beta}$ and $\|v_t\|_2 \le \frac{1}{\beta} \cdot \Phi_t^{1+\beta}$. Let $U = 2\mathrm{Tr} A^{1/2\beta}$, $f(\Phi) = c\frac{\log n}{\beta^2} \Phi^{1+2\beta}$, and $g(\Phi) = c\frac{1}{\beta} \cdot \Phi^{1+\beta}$ for some large enough constant c. Take $T = c' \frac{\beta^2}{U^{2\beta} \log n}$ with some small enough constant c'. Note that $f(U) \cdot T \le \frac{U}{8}$ and $g(U) \cdot \sqrt{T} \le \frac{U}{8}$. This verifies the conditions in Lemma B.11 and hence this shows that

$$\mathbb{P}\left[\max_{t \in [0,T]} \Phi_t \ge U\right] \le 0.01.$$

Hence, we have $\mathbb{P}\left[\max_{t\in[0,T]} \|A_t\|_{\mathrm{op}} \geq U^{2\beta}\right] \leq 0.01$ and

$$\mathbb{P}\left(\int_0^T \|A_s\|_{\mathrm{op}}\, ds \leq \frac{1}{64}\right) \geq \frac{3}{4}.$$

Using this, Lemma B.2 shows that

$$\psi_p \lesssim T^{-1/2} \lesssim \frac{1}{\beta} U^{\beta} \sqrt{\log n} \lesssim \frac{\sqrt{\log n}}{\beta} \|A\|_{\frac{1}{2\beta}}^{1/2}.$$

□