Recent Progress and Open Problems in Algorithmic Convex Geometry*

Santosh S. Vempala

School of Computer Science
Algorithms and Randomness Center
Georgia Tech, Atlanta, USA 30332
vempala@gatech.edu

Abstract

This article is a survey of developments in algorithmic convex geometry over the past decade. These include algorithms for sampling, optimization, integration, rounding and learning, as well as mathematical tools such as isoperimetric and concentration inequalities. Several open problems and conjectures are discussed on the way.

Contents

* This work was supported by NSF awards AF-0915903 and AF-0910584.

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Algorithmic convex geometry is the study of the algorithmic aspects of convexity. While convex geometry, closely related to finite-dimensional functional analysis, is a rich, classical field of mathematics, it has enjoyed a resurgence since the late 20th century, coinciding with the development of theoretical computer science. These two fields started coming together in the 1970's with the development of geometric algorithms for convex optimization, notably the Ellipsoid method. The ellipsoid algorithm uses basic notions in convex geometry with profound consequences for computational complexity, including the polynomial-time solution of linear programs. This development heralded the use of more geometric ideas, including Karmarkar's algorithm and interior point methods. It brought to the forefront fundamental geometric questions such as *rounding*, i.e., applying affine transformations to sets in Euclidean space to make them easier to handle. The ellipsoid algorithm also resulted in the striking application of continuous geometric methods for the solution of discrete optimization problems such as submodular function minimization.

A further startling development occurred in 1989 with the discovery of a polynomial method for estimating the volume of a convex body. This was especially surprising in the light of an exponential lower bound for any deterministic algorithm for volume estimation. The crucial ingredient in the volume algorithm was an efficient procedure to sample nearly uniform random points from a convex body. Over the past two decades the algorithmic techniques and analysis tools for sampling and volume estimation have been greatly extended and refined. One noteworthy aspect here is the development of isoperimetric inequalities that are of independent interest as purely geometric properties and lead to new directions in the study of convexity. Efficient sampling has also lead to alternative polynomial algorithms for convex optimization.

Besides optimization, integration and sampling, our focus problems in this survey are *rounding* and *learning*. Efficient algorithms for the first three problems rely crucially on rounding. Approximate rounding can be done using the ellipsoid method and with a tighter guarantee using random sampling. Learning is a problem that generalizes integration. For example, while we know how to efficiently compute the volume of a polytope, learning one efficiently from random samples is an open problem. For general convex bodies, volume computation is efficient, while learning requires a superpolynomial number of samples (volume computation uses a membership oracle while learning gets only random samples).

This survey is not intended to be comprehensive. There are numerous interesting developments in convex geometry, related algorithms and their growing list of applications and connections to other areas (e.g., data privacy, quantum computing, statistical estimators etc.) that we do not cover here, being guided instead by our five focus problems.

1.1 Basic definitions

Recall that a subset S of $\mathbb{R}^n$ is convex if for any two points $x, y \in S$, the interval $[x, y] \subseteq S$. A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is said to be *logconcave* if for any two points $x, y \in \mathbb{R}^n$ and any $\lambda \in [0, 1]$,

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

The indicator function of a convex set and the density function of a Gaussian distribution are two canonical examples of logconcave functions.

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is said to be *isotropic*, if its centroid is the origin, and its covariance matrix is the identity matrix. In terms of the associated random variable X,

this means that

$$\mathsf{E}(X) = 0 \quad \text{and} \quad \mathsf{E}(XX^T) = I.$$

This condition is equivalent to saying that for every unit vector $v \in \mathbb{R}^n$,

$$\int_{\mathbb{R}^n} (v^T x)^2 f(x)\, dx = 1.$$

We say that f is C-isotropic, if

$$\frac{1}{C} \leq \int_{\mathbb{R}^n} (v^T x)^2 f(x)\, dx \leq C$$

for every unit vector v. A convex body is said to be isotropic if the uniform density over it is isotropic. For any full-dimensional distribution D with bounded second moments, there is an affine transformation of space that puts it in isotropic position, namely, if

$$z = \mathsf{E}_D(X) \text{ and } A = \mathsf{E}((X - z)(X - z)^T)$$

then $y = A^{-\frac{1}{2}}(X - z)$ has an isotropic density.

For a density function $f : \mathbb{R}^n \to \mathbb{R}_+$, we let π_f be the measure associated with it. The following notions of distance between two distributions P and Q will be used: The *Total variation distance* of P and Q is

$$d_{tv}(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|.$$

The χ-*squared distance* of P with respect to Q is

$$\chi^2(P, Q) = \int_{\mathbb{R}}^n \left(\frac{dP(u) - dQu}{dQ(u)} \right)^2 dQ(u) = \int_{\mathbb{R}}^n \frac{dP(u)}{dQ(u)}\, dP(u) - 1$$

The first term in the last expression is also called the L_2 distance of P w.r.t. Q. Finally, P is said to be M-*warm* w.r.t. Q if

$$M = \sup_{A \in \mathcal{A}} \frac{P(A)}{Q(A)}.$$

A discrete-time *Markov chain* is defined using a triple $(K, \mathcal{A}, \{P_u : u \in K\})$ along with a starting distribution Q_0, where K is the state space, $\mathcal{A}$ is a set of measurable subsets of K and P_u is a measure over K, as a sequence of elements of K, $w_0, w_1, \ldots$, where w_0 is chosen from Q_0 and each subsequent w_i is chosen from $P_{w_{i-1}}$. Thus, the choice of w_{i+1} depends only on w_i and is independent of $w_0, \ldots, w_{i-1}$. It is easy to verify that a distribution Q is stationary iff for every $A \in \mathcal{A}$,

$$\int_A P_u(K \setminus A)\, dQ(u) = \int_{K \setminus A} P_u(A)\, dQ(u).$$

The *conductance* of a subset A is defined as

$$\phi(A) = \frac{\int_A P_u(K \setminus A)\, dQ(u)}{\min\{Q(A), Q(K \setminus A)\}}$$

and the conductance of the Markov chain is

$$\phi = \min_A \phi(A) = \min_{0 < Q(A) \leq \frac{1}{2}} \frac{\int_A P_u(K \setminus A)\, dQ(u)}{Q(A)}.$$

The following weaker notion of conductance will also be useful. For any $0 \leq s < \frac{1}{2}$, the *s-conductance* of a Markov chain is defined as

$$\phi_s = \min_{A: s < Q(A) \leq \frac{1}{2}} \frac{\int_A P_u(K \setminus A)\, dQ(u)}{Q(A) - s}.$$

We let $\mathsf{Var}_f(g)$ denote the variance of a real-valued function g w.r.t. a density function f. The unit ball in $\mathbb{R}^n$ by B_n. We use $O^*(.)$ to suppress error parameters and terms that are polylogarithmic in the leading terms.

2 Problems

We now state our focus problems more precisely and mention the current bounds on their complexity. Algorithms achieving these bounds are discussed in a Section 4. As input to these problems, we typically assume a general function oracle, i.e., access to a real-valued function as a blackbox. The complexity will be measured by both the number of oracle calls and the number of arithmetic operations.

2.1 Optimization

Input: an oracle for a function $f : \mathbb{R}^n \to \mathbb{R}_+$; a point x_0 with $f(x_0) > 0$; an error parameter ε.
Output: a point x such that

$$f(x) \geq \max f - \varepsilon.$$

Linear optimization over a convex set is the special case when $f = e^{-c^T x}\chi^K(x)$ where $c^T x$ is a linear function to be minimized and K is the convex set. This is equivalent to a membership oracle for K in this case. Given only access to the function f, for any logconcave f, the complexity of optimization is $\mathrm{poly}(n, \log(1/\varepsilon))$ by either an extension of the Ellipsoid method to handle membership oracles [21], or Vaidya's algorithm [60] or a reduction to sampling [6, 25, 45]. This line of work began with the breakthrough result of Khachiyan [34] showing that the Ellipsoid method proposed by Yudin and Nemirovskii [65] is polynomial for explicit linear programs. The current best time complexity is $O^*(n^{4.5})$ achieved by a reduction to logconcave sampling [45]. For optimization of an explicit function over a convex set, the oracle is simply membership in the convex set. A stronger oracle is typically available, namely a *separation* oracle that gives a hyperplane that separates the query point from the convex set, in the case when the point lies outside. Using a separation oracle, the ellipsoid algorithm has complexity $O^*(n^2)$ while Vaidya's algorithm and that of Bertsimas and Vempala have complexity $O^*(n)$.

The current frontier for polynomial-time algorithms is when f is a logconcave function in $\mathbb{R}^n$. Slight generalizations, e.g., linear optimization over a star-shaped body is NP-hard, even to solve approximately [49, 9].

2.2 Integration/Counting

Input: an oracle for a function $f : \mathbb{R}^n \to \mathbb{R}_+$; a point x_0 with $f(x_0) > 0$; an error parameter ε.
Output: a real number A such that

$$(1-\varepsilon)\int f \leq A \leq (1+\varepsilon)\int f.$$

Dyer, Frieze and Kannan [15, 16] gave a polynomial-time randomized algorithm for estimating the volume of a convex body (the special case above when f is the indicator function of a convex body) with complexity $\text{poly}(n, 1/\varepsilon, \log(1/\delta))$, where δ is the probability that the algorithm's output is incorrect. This was extended to integration of logconcave functions by Applegate and Kannan [3] with an additional dependence of the complexity on the Lipschitz parameter of f. Following a long line of improvements, the current best complexity is $O^*(n^4)$ using a variant of simulated annealing, quite similar to the algorithm for optimization [45].

In the discrete setting, the natural general problem is when f is a distribution over the integer points in a convex set, e.g., f is 1 only for integer points in a convex body K. While there are many special cases that are well-studied, e.g., perfect matchings of a graph [23], not much is known in general except a roundness condition [30].

2.3 Learning

Input: random points drawn from an unknown distribution in $\mathbb{R}^n$ and their labels given by an unknown function $f : \mathbb{R}^n \to \{0, 1\}$; an error parameter ε.
Output: A function $g : \mathbb{R}^n \to \{0, 1\}$ such that $\mathsf{P}(g(x) \neq f(x)) \leq \varepsilon$ over the unknown distribution.

The special case when f is an unknown linear function can be solved by using any linear classifier that correctly labels a sample of size

$$\frac{C}{\varepsilon}\left(n \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}\right),$$

where C is a constant. Such a classifier can be learned by using any efficient algorithm for linear programming. The case when f is a polynomial threshold function of bounded degree can also be learned efficiently essentially by reducing to the case of linear thresholds via a linearization of the polynomial.

2.4 Sampling

Input: an oracle for a function $f : \mathbb{R}^n \to \mathbb{R}_+$; a point x_0 with $f(x_0) > 0$; an error parameter ε.
Output: a random point x whose distribution D is within ε total variation distance of the distribution with density proportional to f.

As the key ingredient of their volume algorithm [16], Dyer, Frieze and Kannan gave a polynomial-time algorithm for the case when f is the indicator function over a convex body. The complexity of the sampler itself was $\text{poly}(n, 1/\varepsilon)$. Later, Applegate and Kannan [3] generalized this to smooth logconcave functions with complexity polynomial in $n, 1/\varepsilon$ and the Lipschitz parameter of the logconcave function. In [45] the complexity was improved to $O^*(n^4)$ and the dependence on the Lipschitz parameter was removed.

2.5 Rounding.

We first state this problem for a special case.
Input: a membership oracle for a convex body K; a point $x_0 \in K$; an error parameter ε.
Output: a linear transformation A and a point z such that $K' = A(K - z)$ satisfies one of the following:

- Exact sandwiching:

$$B_n \subseteq K' \subseteq RB_n. \tag{1}$$

- Second moment sandwiching: For every unit vector u,

$$1 - \varepsilon \leq \mathsf{E}_{K'}((u \cdot x)^2) \leq 1 + \varepsilon.$$

- Approximate sandwiching:

$$\mathrm{vol}(B_n \cap K') \geq \frac{1}{2}\mathrm{vol}(B_n) \text{ and } \mathrm{vol}(K' \cap RB_n) \geq \frac{1}{2}\mathrm{vol}(K'). \tag{2}$$

When the input is a general density function, the second moment condition extends readily without any change to the statement.

The classical Löwner-John theorem says that for any convex body K, the ellipsoid E of maximal volume contained in K has the property that $K \subseteq nE$ (and $K \subseteq \sqrt{n}E$ if K is centrally symmetric). Thus by using the transformation that makes this ellipsoid a ball, one achieves $R = n$ in the first condition above (exact sandwiching). Moreover, this is the best possible as shown by the simplex. However, computing the ellipsoid of maximal volume is hard. Lovász [41] showed how to compute an ellipsoid that satisfies the containment with $R = n^{3/2}$ using the Ellipsoid algorithm. This remains the best-known deterministic algorithm. A randomized algorithm can achieve $R = n(1 + \varepsilon)$ for any $\varepsilon > 0$ using a simple reduction to random sampling [27]. In fact, all that one needs is $n \cdot C(\varepsilon)$ random samples [8, 57, 54, 1], and then the transformation to be computed is the one that puts the sample in isotropic position.

3 Geometric inequalities and conjectures

We begin with some fundamental inequalities. For two subsets A, B of $\mathbb{R}^n$, their Minkowski sum is

$$A + B = \{x + y : x \in A, y \in B\}.$$

The Brunn-Minkowski inequality says that if A, B and $A + B$ are measurable, compact subsets of $\mathbb{R}^n$, then

$$\mathrm{vol}(A + B)^{\frac{1}{n}} \geq \mathrm{vol}(A)^{\frac{1}{n}} + \mathrm{vol}(B)^{\frac{1}{n}}. \tag{3}$$

The following extension is the Prékopa-Leindler inequality: for any three functions $f, g, h : \mathbb{R}^n \to \mathbb{R}_+$, satisfying

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

for every $\lambda \in [0, 1]$,

$$\int_{\mathbb{R}^n} h(x)\,dx \geq \left(\int_{\mathbb{R}^n} f(x)\,dx\right)^\lambda \left(\int_{\mathbb{R}^n} g(x)\,dx\right)^{1-\lambda}. \tag{4}$$

The product and the minimum of two logconcave functions are also logconcave. We also have the following fundamental properties [12, 39, 55, 56].

▶ **Theorem 1.** *All marginals and the distribution function of a logconcave function are logconcave. The convolution of two logconcave functions is logconcave.*

In the rest of this section, we describe inequalities about rounding, concentration of mass and isoperimetry. These inequalities play an important role in the analysis of algorithms for our focus problems. Many of the inequalities apply to logconcave functions. The latter represent the limit to which methods that work for convex bodies can be extended. This is true for all the algorithms we will see with a few exceptions, e.g., an extension of isoperimetry and sampling to star-shaped bodies. We note here that Milman [50] has shown a general approximate equivalence between concentration and isoperimetry over convex domains.

3.1 Rounding

We next describe a set of properties related to affine transformations. The first one below is from [27].

▶ **Theorem 2.** *[27] Let K be a convex body in $\mathbb{R}^n$ in isotropic position. Then,*

$$\sqrt{\frac{n+1}{n}}B_n \subseteq K \subseteq \sqrt{n(n+1)}B_n.$$

Thus, in terms of the exact sandwiching condition (1), isotropic position achieves a factor of n, and this ratio of the radii of the outer and inner ball, n, is the best possible as shown by the n-dimensional simplex. A bound of $O(n)$ was earlier established by Milman and Pajor [51].

Sandwiching was extended to logconcave functions in [48] as follows.

▶ **Theorem 3.** *[48] Let f be a logconcave density in isotropic position. Then for any $t > 0$, the level set $L(t) = \{x : f(x) \geq t\}$ contains a ball of radius $\pi_f(L(t))/e$. Also, the point $z = \arg\max f$ satisfies $\|z\| \leq n + 1$.*

For general densities, it is convenient to define a rounding parameter as follows. For a density function f, let $R^2 = \mathsf{E}_f(\|X - \mathsf{E}X\|^2)$ and r be the radius of the largest ball contained in the level set of f of measure $1/8$. Then R/r is the rounding parameter and we say a function is *well-rounded* if $R/r = O(\sqrt{n})$. By the above theorem, any logconcave density in isotropic position is well-rounded.

An isotropic transformation can be estimated for any logconcave function using random samples drawn from the density proportional to f. In fact, following the work of Bourgain [8] and Rudelson [57], Adamczak et al [1] showed that $O(n)$ random points suffice to achieve, say, 2-isotropic position with high probability.

▶ **Theorem 4.** *[1] Let $x_1, \ldots x_m$ be random points drawn from an isotropic logconcave density f. Then for any $\varepsilon \in (0,1), t \geq 1$, there exists $C(\varepsilon, t)$ such that if $m \geq C(\varepsilon, t)n$,*

$$\left\| \frac{1}{m}\sum_{i=1}^{m} x_i x_i^T - I \right\|_2 \leq \varepsilon$$

with probability at least $1 - e^{-ct\sqrt{n}}$. Moreover,

$$C(\varepsilon, t) = C\frac{t^4}{\varepsilon^2}\log^2\frac{2t^2}{\varepsilon^2}$$

for absolute constants c, C suffices.

The following conjecture was alluded to in [44].

▶ Conjecture 5. There exists a constant C such that for any convex body K in $\mathbb{R}^n$, there exists an ellipsoid E that satisfies approximate sandwiching:

$$\mathrm{vol}(E \cap K) \geq \frac{1}{2}\mathrm{vol}(E); \mathrm{vol}(K \cap C \log n E) \geq \frac{1}{2}\mathrm{vol}(K).$$

3.2 Measure and concentration

The next inequality was proved by Grünbaum [22] (for the special case of the uniform density over a convex body).

▶ **Lemma 6.** *[22] Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a logconcave density function, and let H be any halfspace containing its centroid. Then*

$$\int_H f(x)\,dx \geq \frac{1}{e}.$$

This was extended to hyperplanes that come close to the centroid in [6, 48].

▶ **Lemma 7.** *[6, 48] Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an isotropic logconcave density function, and let H be any halfspace within distance t from its centroid. Then*

$$\int_H f(x)\,dx \geq \frac{1}{e} - t.$$

The above lemma easily implies a useful concentration result.

▶ **Theorem 8.** *Let f be a logconcave density in $\mathbb{R}^n$ and z be the average of m random points from π_f. If H is a halfspace containing z,*

$$\mathsf{E}\left(\pi_f(H)\right) \geq \left(\frac{1}{e} - \sqrt{\frac{n}{m}}\right).$$

A strong radial concentration result was shown by Paouris [54].

▶ **Lemma 9.** *[54] Let K be an isotropic convex body. Then, for any $t \geq 1$,*

$$\mathsf{P}(\|X\| \geq ct\sqrt{n}) \leq e^{-t\sqrt{n}}$$

where c is an absolute constant.

This implies that all but an exponentially small fraction of the mass of an isotropic convex body lies in a ball of radius $O(\sqrt{n})$.

The next inequality is a special case of a more general inequality due to Brascamp and Lieb. It also falls in a family called Poincaré-type inequalities. Recall that $\mathsf{Var}_\mu(f)$ denote the variance of a real-valued function f w.r.t. a density function μ.

▶ **Theorem 10.** *Let γ be the standard Gaussian density in $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Then*

$$\mathsf{Var}_\gamma(f) \leq \int_{\mathbb{R}^n} \|\nabla f\|^2\,d\gamma.$$

A smooth function here is one that is locally Lipschitz.

An extension of this inequality to logconcave restrictions was developed in [63], eliminating the need for smoothness and deriving a quantitative bound. In particular, the lemma shows how variances go down when the standard Gaussian is restricted to a convex set.

▶ **Theorem 11.** *[63] Let g be the standard Gaussian density function in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}_+$ be any logconcave function. Define the function h to be the density proportional to their product, i.e.,*

$$h(x) = \frac{f(x)g(x)}{\int_{\mathbb{R}^n} f(x)g(x)\,dx}.$$

Then, for any unit vector $u \in \mathbb{R}^n$,

$$\mathsf{Var}_h(u \cdot x) \leq 1 - \frac{e^{-b^2}}{2\pi}$$

where the support of f along u is $[a_0, a_1]$ and $b = \min\{|a_0|, |a_1|\}$.

The next conjecture is called the *variance hypothesis* or *thin shell* conjecture.

▶ Conjecture 12 (Thin shell). For X drawn from a logconcave isotropic density f in $\mathbb{R}^n$,

$$\mathsf{Var}_f(\|X\|^2) \leq Cn$$

We note here that for an isotropic logconcave density,

$$\sigma_n^2 = \mathsf{E}_f((\|X\| - \sqrt{n})^2) \leq \frac{1}{n}\mathsf{Var}_f(\|X\|^2) \leq C\sigma_n^2,$$

i.e., the deviation of $\|X\|$ from $\sqrt{n}$ and the deviation of $\|X\|^2$ from $\mathsf{E}(\|X\|^2) = n$ are closely related. Thus, the conjecture implies that most of the mass of a logconcave distribution lies in shell of *constant* thickness.

A bound of $\sigma_n \leq \sqrt{n}$ is easy to establish for any isotropic logconcave density. Klartag showed that $\sigma_n = o(\sqrt{n})$ [36, 37], a result that can be interpreted as a central limit theorem (since the standard deviation is now a smaller order term than the expectation). The current best bound on σ_n is due to Fleury [19] who showed that $\sigma_n \leq n^{3/8}$. We state his result below.

▶ **Lemma 13.** *[19] For any isotropic logconcave measure μ,*

$$\mathsf{P}\left(|\|x\| - \sqrt{n}| \geq tn^{\frac{3}{8}}\right) \leq Ce^{-ct}.$$

where c, C are constants.

3.3 Isoperimetry

The following theorem is from [14], improving on a theorem in [43] by a factor of 2. For two subsets S_1, S_2 of $\mathbb{R}^n$, we let $d(S_1, S_2)$ denote the minimum Euclidean distance between them.

▶ **Theorem 14.** *[14] Let S_1, S_2, S_3 be a partition into measurable sets of a convex body K of diameter D. Then,*

$$\mathrm{vol}(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}.$$

A limiting (and equivalent) version of this inequality is the following: Let $\partial S \cap K$ denote the interior boundary of S w.r.t. K. For any subset S of a convex body of diameter D,

$$\mathrm{vol}_{n-1}(\partial S \cap K) \geq \frac{2}{D} \min\{\mathrm{vol}(S), \mathrm{vol}(K \setminus S)\},$$

i.e., the surface area of S inside K is large compared to the volumes of S and $K \setminus S$. This is in direct analogy with the classical isoperimetric inequality, which says that the surface area to volume ratio of any measurable set is at least the ratio for a ball. Henceforth, for a density function f, we use $\Phi_f(S)$ to denote the isoperimetric ratio of a subset S, i.e.,

$$\Phi_f(S) = \frac{\pi_f(\partial S)}{\min\{\pi_f(S), 1 - \pi_f(S)\}}$$

and

$$\Phi_f = \inf_{S \subseteq \mathbb{R}^n} \phi(S).$$

Theorem 14 can be generalized to arbitrary logconcave measures.

▶ **Theorem 15.** *Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a logconcave function whose support has diameter D and let π_f be the induced measure. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\pi_f(S_1), \pi_f(S_2)\}.$$

The conclusion can be equivalently stated as $\Phi_f \geq 2/D$. For Euclidean distance this latter statement, applied to the subset S_1 essentially recovers the above lower bound on $\pi_f(S_3)$.

Next we ask if logconcave functions are the limit of such isoperimetry. Theorem 15 can be extended to s-concave functions in $\mathbb{R}^n$ for $s \geq -1/(n-1)$ with only a loss of a factor of 2 [10]. A nonnegative function f is s-concave if $f(x)^s$ is a concave function of s. Logconcave functions correspond to $s \to 0$. The only change in the conclusion is by a factor of 2 on the RHS.

▶ **Theorem 16.** *[10] Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an s-concave function with $s \geq -1/(n-1)$ and with support of diameter D and let π_f be the induced measure. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{d(S_1, S_2)}{D} \min\{\pi_f(S_1), \pi_f(S_2)\}.$$

This theorem is tight, in that there exist s-concave functions for $s \leq -1/(n-1-\varepsilon)$ for any $\varepsilon > 0$ whose isoperimetric ratio is exponentially small in εn.

Logconcave and s-concave functions are natural extensions of convex bodies. It is natural to ask what classes of nonconvex sets/functions might have good isoperimetry. In recent work [9], the isoperimetry of star-shaped bodies was studied. A star-shaped set S is one that contains at least one point x such that for any $y \in S$, the segment $[x, y]$ is also in S. The set of all such points x for which lines through x have convex intersections with S is called its kernel K_S. The kernel is always a convex set.

▶ **Theorem 17.** *[9] Let S_1, S_2, S_3 be a partition into measurable sets of a star-shaped body S of diameter D. Let $\eta = \mathrm{vol}(K_S)/\mathrm{vol}(S)$. Then,*

$$\mathrm{vol}(S_3) \geq \frac{d(S_1, S_2)}{4\eta D} \min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}.$$

In terms of diameter, Theorem 14 is the best possible, as shown by a cylinder. A more refined inequality is obtained in [27, 48] using the average distance of a point to the center of gravity (in place of diameter). It is possible for a convex body to have much larger diameter than average distance to its centroid (e.g., a cone). In such cases, the next theorem provides a better bound.

▶ **Theorem 18.** *[27] Let f be a logconcave density in $\mathbb{R}^n$ and π_f be the corresponding measure. Let z_f be the centroid of f and define $M(f) = \mathsf{E}_f(|x - z_f|)$. Then, for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{\ln 2}{M(f)} d(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

For an isotropic density, $M(f)^2 \leq \mathsf{E}_f(|x - z_f|^2) = n$ and so $M(f) \leq \sqrt{n}$. The diameter could be unbounded (e.g., an isotropic Gaussian). Thus, if A_f is the covariance matrix of the logconcave density f, the inequality can be re-stated as:

$$\pi_f(S_3) \geq \frac{c}{\sqrt{\mathsf{Tr}(A_f)}} d(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

A further refinement is called the KLS hyperplane conjecture [27]. Let $\lambda_1(A)$ be the largest eigenvalue of a symmetric matrix A.

▶ Conjecture 19 (KLS). Let f be a logconcave density in $\mathbb{R}^n$. There is an absolute constant c such that

$$\Phi_f \geq \frac{c}{\sqrt{\lambda_1(A_f)}}.$$

The KLS conjecture implies the thin shell conjecture (Conj. 12). Even the thin shell conjecture would improve the known isoperimetry of isotropic logconcave densities due to the following result of Bobkov [7].

▶ **Theorem 20.** *[7] For any logconcave density function f in $\mathbb{R}^n$,*

$$\Phi_f \geq \frac{c}{\mathsf{Var}(\|X\|^2)^{1/4}}.$$

Combining this with Fleury's bound 13, we get that for an isotropic logconcave function in $\mathbb{R}^n$,

$$\Phi_f \geq cn^{-7/16}. \tag{5}$$

This is slightly better than the bound of $cn^{-1/2}$ implied by Theorem 18 since $\mathsf{E}(\|X\|^2) = n$ for an isotropic density.

The next conjecture is perhaps the most well-known in convex geometry, and is called the slicing problem, the isotropic constant or simply the hyperplane conjecture.

▶ Conjecture 21 (Slicing). There exists a constant C, such that for an isotropic logconcave density f in $\mathbb{R}^n$,

$$f(0) \leq C^n.$$

In fact, a result of Ball [4] shows that it suffices to prove such a bound for the case of f being the uniform density over an isotropic convex body. In this case, the conjecture says the volume of an isotropic convex body in $\mathbb{R}^n$ is at least c^n for some constant c. The current best bound on $L_n = \sup_f f(0)^{1/n}$ is $O(n^{1/4})$ [8, 35].

Recently, Eldan and Klartag [18] have shown that the slicing conjecture is implied by the thin shell conjecture (and therefore by the KLS conjecture as well, a result that was earlier shown by Ball).

▶ **Theorem 22.** *[18] There exists a constant C such that*

$$L_n = \sup_f f(0)^{1/n} \leq C\sigma_n = C \sup_f \sqrt{\mathsf{E}_f((\|X\| - \sqrt{n})^2)}$$

where the suprema range over isotropic logconcave density functions.

We conclude this section with a discussion of another direction in which classical isoperimetry has been extended. The *cross-ratio distance* (also called Hilbert metric) between two points u, v in a convex body K is computed as follows: Let p, q be the endpoints of the chord in K through u and v such that the points occur in the order p, u, v, q. Then

$$d_K(u, v) = \frac{|u - v||p - q|}{|p - u||v - q|} = (p : v : u : q).$$

where $(p : v : u : q)$ denotes the classical cross-ratio. We can now define the cross-ratio distance between two sets S_1, S_2 as

$$d_K(S_1, S_2) = \min\{d_K(u, v) : u \in S_1, v \in S_2\}.$$

The next theorem was proved in [42] for convex bodies and extended to logconcave densities in [47].

▶ **Theorem 23.** *[42] Let f be a logconcave density in $\mathbb{R}^n$ whose support is a convex body K and let π_f be the induced measure. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq d_K(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

All the inequalities so far are based on defining the distance between S_1 and S_2 by the *minimum* over pairs of some notion of pairwise distance. It is reasonable to think that perhaps a much sharper inequality can be obtained by using some *average* distance between S_1 and S_2. Such an inequality was proved in [46], leading to a substantial improvement in the analysis of hit-and-run.

▶ **Theorem 24.** *[46] Let K be a convex body in $\mathbb{R}^n$. Let $f : K \to \mathbb{R}_+$ be a logconcave density with corresponding measure π_f and $h : K \to \mathbb{R}_+$, an arbitrary function. Let S_1, S_2, S_3 be any partition of K into measurable sets. Suppose that for any pair of points $u \in S_1$ and $v \in S_2$ and any point x on the chord of K through u and v,*

$$h(x) \leq \frac{1}{3}\min(1, d_K(u, v)).$$

Then

$$\pi_f(S_3) \geq \mathsf{E}_f(h(x)) \min\{\pi_f(S_1), \pi_f(S_2)\}.$$

The coefficient on the RHS has changed from a "minimum" to an "average". The weight $h(x)$ at a point x is restricted only by the minimum cross-ratio distance between pairs u, v from S_1, S_2 respectively, such that x lies on the line between them (previously it was the overall minimum). In general, it can be much higher than the minimum cross-ratio distance between S_1 and S_2.

3.4 Localization

Most of the known isoperimetry theorems can be proved using the localization lemma of Lovász and Simonovits [44, 27]. It reduces inequalities in high dimension to 1-dimensional inequalities by a sequence of bisections and a limit argument. To prove that an inequality is true, one assumes it is false and derives a contradiction in the 1-dimensional case.

▶ **Lemma 25.** *[44] Let $g, h : \mathbb{R}^n \to \mathbb{R}$ be lower semi-continuous integrable functions such that*

$$\int_{\mathbb{R}^n} g(x)\,dx > 0 \quad \text{and} \quad \int_{\mathbb{R}^n} h(x)\,dx > 0.$$

Then there exist two points $a, b \in \mathbb{R}^n$ and a linear function $\ell : [0,1] \to \mathbb{R}_+$ such that

$$\int_0^1 \ell(t)^{n-1} g((1-t)a + tb)\,dt > 0 \quad \text{and} \quad \int_0^1 \ell(t)^{n-1} h((1-t)a + tb)\,dt > 0.$$

The points a, b represent an interval A and one may think of $l(t)^{n-1}dA$ as the cross-sectional area of an infinitesimal cone with base area dA. The lemma says that over this cone truncated at a and b, the integrals of g and h are positive. Also, without loss of generality, we can assume that a, b are in the union of the supports of g and h.

A variant of localization was used in [9], where instead of proving an inequality for *every* needle, one instead averages over the set of needles produced by the localization procedure. Indeed, each step of localization is a bisection with a hyperplane and for an inequality to hold for the original set, it often suffices for it hold "on average" over the partition produced rather than for both parts separately. This approach can be used to prove isoperimetry for star-shaped sets (Theorem 17) or to recover Bobkov's isoperimetric inequality (Lemma 20).

We state here a strong version of a conjecture related to localization, which if true implies the KLS hyperplane conjecture.

▶ Conjecture 26. Let f be a logconcave density function and $\mathcal{P}$ be any partition of $\mathbb{R}^n$ such that each part is convex. Then there exists a constant C such that

$$\sum_{\mathcal{P}} \sigma_f(P)^2 \pi_f(P) \le C\sigma_f^2$$

where σ_f^2 is the largest variance of f along any direction and $\sigma_f(P)^2$ is the largest variance of f restricted to the set P.

We have already seen that the conjecture holds when f is a Gaussian with $C = 1$ (Theorem 11).

4 Algorithms

In this section, we describe the current best algorithms for the five focus problems and several open questions. These algorithms are related to each other, e.g., the current best algorithms for integration and optimization (in the membership oracle model) are based on sampling, the current best algorithms for sampling and rounding proceed in tandem etc.

In fact, all five problems have an intimate relationship with random sampling. Integration (volume estimation), optimization and rounding are solved by reductions to random sampling; in the case of integration, this is the only known efficient approach. For rounding, the sampling approach gives a better bound than the current best alternatives. As for the learning problem, its input is a random sample.

The Ellipsoid method has played a central role in the development of algorithmic convex geometry. Following Khachiyan's polynomial bound for linear programs [34], the generalization to convex optimization [21, 53, 32] has been a powerful theoretical tool. Besides optimization, the method also gave an efficient rounding algorithm [41] achieving a sandwiching ratio of $n^{1.5}$, and a polynomial-time algorithm for PAC-learning linear threshold functions. It is a major open problem to PAC-learn an intersection of *two* halfspaces.

Several polynomial-time algorithms have been proposed for linear programming and convex optimization, including: Karmarkar's algorithm [31] for linear programming, interior-point methods for convex optimization over sets with *self-concordant barriers* [52], polynomial implementation of the perceptron algorithm for linear [13] and conic program [5], simplex-with-rescaling for linear programs [33] and the random walk method [6] that requires only membership oracle access to the convex set. The field of convex optimization continues to be very active, both because of its many applications but also due to the search for more practical algorithms that work on larger inputs.

One common aspect of all the known polynomial-time algorithms is that they use use some type of repeated rescaling of space to effectively make the convex set "more round", leading to a complexity that depends on the accuracy to which the output is computed. In principle, such a dependence can be avoided for linear programming and it is a major open problem to find a strongly polynomial algorithm for solving linear programs.

We note here that although there have been many developments in continuous optimization algorithms in the decades since the ellipsoid method, and in applications to combinatorial problems, there has been little progress in the past two decades on general integer programming, with the current best complexity being $n^{O(n)}$ from Kannan's improvement [26] of Lenstra's algorithm [40].

4.1 Geometric random walks

Sampling is achieved by rapidly mixing geometric random walks. For an in-depth survey of this topic, up-to-date till 2005, the reader is referred to [62]. Here we outline the main results, including developments since then.

To sample from a density proportional to a function $f : \mathbb{R}^n \to \mathbb{R}_+$, we set up a Markov chain whose state space is $\mathbb{R}^n$ and stationary distribution has density proportional to f. At a point x, one step of the chain picks a *neighbor* y of x from some distribution P_x that depends only on x, where P_x is defined in some efficiently sampleable manner, and then we move to y with some probability or stay at x. For example, for uniformly sampling a convex body K, the *ball walk* defines the neighbors of x as all points within a fixed distance δ from x, and a random neighbor y is accepted as long as y is also in K. For sampling a general density, one could use the same neighbor set, and modify the probability of transition to $\min\{1, f(y)/f(x)\}$. This rejection mechanism is called the Metropolis filter and the Markov chain itself is the ball walk with a Metropolis filter.

Another Markov chain that has been successfully analyzed is called *hit-and-run*. At a point x, we pick a uniform random line l passing x, then a random point on the line l according to the density induced by the target distribution along l. In the case of uniformly sampling a convex body, this latter distribution is simply the uniform distribution on a chord; for a logconcave function it is the one-dimensional density proportional to the target density f along the chosen line l. This process has the advantage of not needing a step-size parameter δ.

For ease of analysis, we typically assume that the Markov chain is *lazy*, i.e., it stays put with probability 1/2 and attempts a move with probability 1/2. Then, it follows that

the distribution of the current point converges to a unique stationary distribution assuming the conductance of the Markov chain is nonzero. The rate of convergence is approximately determined by the conductance as given by the following theorem due to Lovász and Simonovits [44], extending earlier work by Diaconis and Stroock [11], Alon [2] and Jerrum and Sinclair [58].

▶ **Theorem 27.** *[44] Let Q be the stationary distribution of a lazy Markov chain with Q_0 being its starting distribution and Q_t the distribution of the current point after t steps.*

a. Let $M = \sup_A Q_0(A)/Q(A)$. Then,

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

b. Let $0 < s \leq \frac{1}{2}$ and $H_s = \sup\{|Q_0(A) - Q(A)| : Q(A) \leq s\}$. Then,

$$d_{tv}(Q_t, Q) \leq H_s + \frac{H_s}{s}\left(1 - \frac{\phi_s^2}{2}\right)^t.$$

c. For any $\varepsilon > 0$,

$$d_{tv}(Q_t, Q) \leq \varepsilon + \sqrt{\frac{\chi^2(Q_0, Q) + 1}{\varepsilon}}\left(1 - \frac{\phi^2}{2}\right)^t.$$

Thus the main quantity to analyze in order to bound the mixing time is the conductance.

Ball walk. The following bound holds on the conductance of the ball walk [28].

▶ **Theorem 28.** *For any $0 \leq s \leq 1$, we can choose the step-size δ for the ball walk in a convex body K of diameter D so that*

$$\phi_s \geq \frac{s}{200nD}.$$

The proof of this inequality is the heart of understanding the convergence. Examining the definitions of the conductance ϕ and the isoperimetric ratio Φ_Q of the stationary distribution, one sees that they are quite similar. In fact, the main difference is the following: in defining conductance, the weight between two points x and y depends on the probability density of stepping from x to y in one step; for the isoperimetric ratio, the distance is a geometric notion such as Euclidean distance. Connecting these two notions — geometric distance and probabilistic distance — along with isoperimetry bounds leads to the conductance bound above. This is the generic line of proof with each of these components chosen based on the Markov chain being analyzed. For a detailed description, we refer to [62].

Using Theorem 27(b), we conclude that from an M-warm start, the variation distance of Q_t and Q is smaller than ε after

$$t \geq C\frac{M^2}{\varepsilon^2}n^2D^2\ln\left(\frac{2M}{\varepsilon}\right) \tag{6}$$

steps, for some absolute constant C.

The bound of $O(n^2D^2)$ on the mixing rate is the best possible in terms of the diameter, as shown by a cylinder. Here D^2 can be replaced by $\mathsf{E}(\|X - \mathsf{E}X\|^2)$ using the isoperimetric inequality given by Theorem 18. For an isotropic convex body this gives a mixing rate of $O(n^3)$. The current best isoperimetry for convex bodies (5) gives a bound of $O(n^{2.875})$. The KLS conjecture (19) implies a mixing rate of $O(n^2)$, which matches the best possible for the ball walk, as shown for example by an isotropic cube.

Hit-and-run. For hit-and-run, one obtains a qualitatively better bound in the following sense. In the case of the ball walk, the starting distribution heaving affects the mixing rate. It is possible to start at points close to the boundary with very small local conductance, necessitating many attempts to make a single step. Hit-and-run does not have this problem and manages to exponentially accelerate out of any corner. This is captured in the next theorem, using the isoperimetric inequality given by Theorem 24.

▶ **Theorem 29.** *[46] The conductance of hit-and-run in a convex body of diameter D is $\Omega(1/nD)$.*

▶ **Theorem 30.** *[46] Let K be a convex body that contains a unit ball and has centroid z_K. Suppose that $\mathsf{E}_K(|x - z_K|^2) \leq R^2$ and $\chi^2(Q_0, Q) \leq M$. Then after*

$$t \geq Cn^2R^2 \ln^3 \frac{M}{\varepsilon},$$

steps, where C is an absolute constant, we have $d(Q_t, Q)_{tv} \leq \varepsilon$.

The theorem improves on the bound for the ball walk (6) by reducing the dependence on M and ε from polynomial (which is unavoidable for the ball walk) to logarithmic, while maintaining the (optimal) dependence on R and n. For a body in near-isotropic position, $R = O(\sqrt{n})$ and so the mixing time is $O^*(n^3)$. One also gets a polynomial bound starting from *any single* interior point. If x is at distance d from the boundary, then the distribution obtained after one step from x has $\chi^2(Q_1, Q) \leq (n/d)^n$ and so applying the above theorem, the mixing time is $O(n^4 \ln^3(n/d\varepsilon))$.

Theorems 29 and 30 have been extended in [45] to arbitrary logconcave functions.

▶ **Theorem 31.** *[45] Let f be a logconcave density function with support of diameter D and assume that the level set of measure $1/8$ contains a unit ball. Then,*

$$\phi_s \geq \frac{c}{nD \ln(nD/s)}$$

where c is a constant.

This implies a mixing rate that nearly matches the bound for convex bodies.

Affine-invariant walks. In both cases above, the reader will notice the dependence on rounding parameters and the improvement achieved by assuming isotropic position. As we will see in the next section, efficient rounding can be achieved by interlacing with sampling. Here we mention two random walks which achieve the rounding "implicitly" by being affine invariant. The first is a multi-point variant of hit-and-run that can be applied to sampling any density function. It maintains m points $x_1, \ldots, x_m$. For each x_j, it picks a random combination of the current points,

$$y = \sum_{i=1}^{m} \alpha_i (x_i - \overline{x})$$

where the α_i are drawn from $N(0, 1)$; the chosen point x_j is replaced by a random point along this line through x_j in the direction of y. This process in affine invariant and hence one can effectively assume that the underlying distribution is isotropic. The walk was analyzed in [6] assuming a warm start. It is open to analyze the rate of convergence from a general starting distribution.

For polytopes whose facets are given explicitly, Kannan and Narayanan [29] analyzed an affine-invariant process called the *Dikin walk*. At each step, the Dikin walk computes an ellipsoid based on the current point (and the full polytope) and moves to a random point in this ellipsoid. The Dikin ellipsoid at a point x in a polytope $Ax \leq 1$ with m inequalities is defined as:

$$D_x = \{z \in \mathbb{R}^n \, : \, (x-z)^T \sum_{i=1}^{m} \frac{a_i a_i^T}{(1-a_i^T x)^2}(x-z) \leq 1\}.$$

At x, a new point y is sampled from a Dikin ellipsoid and the walk moves to y with prob

$$\min\left\{1, \frac{\mathrm{vol}(D_y)}{\mathrm{vol}(D_x)}\right\}.$$

For polytopes with few facets, this process uses fewer arithmetic operations than the current best bounds for the ball walk or hit-and-run.

Open problems. One open problem for both hit-and-run and the ball walk is to find a single starting point that is as good a start as a warm start. E.g., does one of these walks started at the centroid converge at the same rate as starting from a random point?

The random walks we have considered so far for general convex bodies and density functions rely only on membership oracles or function oracles. Can one do better using a separation oracle? When presented with a point x such an oracle either declares the point is in the convex set K or gives a hyperplane that separates x from K. At least for convex bodies, such an oracle is typically realizable in the same complexity as a membership oracle.

One attractive process based on a separation oracle is the following *reflection walk* with parameter δ: At a point x, we pick a random point y in the ball of radius δ. We move along the straight line from x to y either reaching y or encountering a hyperplane H that separates y from K; in the latter case, we reflect y about H and continue moving towards y.

It is possible that for this process, the number of oracle calls (not the number of Markov chain steps) is only $O^*(nD)$ rather than $O^*(n^2D^2)$, and even $O^*(n)$ for isotropic bodies. It is a very interesting open problem to analyze the reflection walk.

Our next open problem is to understand classes of distributions that can be efficiently sampled by random walks. A recent extension of the convex setting is to sampling star-shaped bodies [9], and this strongly suggests that the full picture is far from clear. One necessary property is good isoperimetry. Can one provide a sufficient condition that depends on isoperimetry and some local property of the density to be sampled? More concretely, for what manifolds with nonnegative curvature (for which isoperimetry is known) can an efficient sampling process be defined?

Our final question in this section is about sampling a discrete subset of a convex body, namely the set of lattice points that lie in the body. In [30], it is shown that this problem can be solved if the body contains a sufficiently large ball, by a reduction to the continuous sampling problem. The idea is that, under this condition, the volume of the body is roughly the same as the number of lattice points and thus sampling the body and rounding to a nearby lattice point is effective. Can this condition be improved substantially? E.g., can one sample lattice points of a near-isotropic convex body? Or lattice points of a body that contains at least half of a ball of radius $\mathrm{O}(\sqrt{n})$?

4.2 Annealing

Rounding, optimization and integration can all be achieved by variants of the same algorithmic technique that one might call *annealing*. The method starts at an "easy" distribution F_0 and goes through a sequence of distributions $F_1, \ldots, F_m$ where the F_i are chosen so that moving from F_{i-1} to F_i is efficient and F_m is a target distribution. This approach can be traced back to [43]. It was analyzed for linear optimization over convex sets in [25], for volume computation and rounding convex sets in [47] and extended to integration, rounding and maximization of logconcave functions in [45].

For example, in the case of convex optimization, the distribution F_m is chosen so that most of its mass on points whose objective value is at least $(1-\varepsilon)$ times the maximum. Sampling directly from this distribution could be difficult since one begins at an arbitrary point.

For integration of logconcave functions, we define a series of functions, with the final function being the one we wish to integrate and the initial one being a function that is easy to integrate. The distribution in each phase has density proportional to the corresponding function. We use samples from the current distribution to estimate the ratios of integrals of consecutive functions. Multiplying all these ratios and the integral of the initial function gives the estimate for the integral of the target function.

For rounding a logconcave density, as shown by Theorem 4, we need $O(n)$ random samples. Generating these directly from the target density could be expensive since sampling (using a random walk) takes time that depends heavily on how well-rounded the density function is. Instead, we consider again a sequence of distributions, where the first distribution in the sequence is chosen to be easy to sample and consecutive distributions have the property that if F_i is isotropic, then F_{i+1} is near-isotropic. We then sample F_{i+1} efficiently, apply an isotropic transformation and proceed to the next phase.

For both optimization and integration, this rounding procedure is incorporated into the main algorithm to keep the sampling efficient. All three cases are captured in the generic description below.

Annealing

1. For $i = 0, \ldots, m$, define
$$a_i = b\left(1 + \frac{1}{\sqrt{n}}\right)^i \quad \text{and} \quad f_i(x) = f(x)^{a_i}.$$
2. Let $X_0^1, \ldots, X_0^k$ be independent random points with density proportional to f_0.
3. For $i = 0, \ldots, m-1$: starting with $X_i^1, \ldots, X_i^k$, generate random points $X_{i+1} = \{X_{i+1}^1, \ldots, X_{i+1}^k\}$; update a running estimate g based on these samples; update the isotropy transformation using the samples.
4. Output the final estimate of g.

For optimization, the function f_m is set to be a sufficiently high power of f, the function to be maximized while g is simply the maximum objective value so far. For integration and rounding $f_m = f$, the target function to be integrated or rounded. For integration, the function g starts out as the integral of f_0 and is multiplied by the ratio of $\int f_{i+1} / \int f_i$ in

each step. For rounding, g is simply the estimate of the isotropic transformation for the current function.

We now state the known guarantees for this algorithm [45]. The complexity improves on the original $O^*(n^{10})$ algorithm of Applegate and Kannan [3].

▶ **Theorem 32.** *[45] Let f be a well-rounded logconcave function. Given $\varepsilon, \delta > 0$, we can compute a number A such that with probability at least $1 - \delta$,*

$$(1-\varepsilon)\int_{\mathbb{R}^n} f(x)\,dx \leq A \leq (1+\varepsilon)\int_{\mathbb{R}^n} f(x)\,dx$$

and the number of function calls is

$$O\left(n^4 \log^c \frac{n}{\varepsilon\delta}\right) = O^*(n^4)$$

where c is an absolute constant.

Any logconcave function can be put in near-isotropic position in time $O^*(n^4)$ steps, given a point that maximizes f. Near-isotropic position guarantees that f is well-rounded as remarked earlier. When the maximum of f is not known in advance or computable quickly, the complexity is a bit higher and is the same as that of maximization.

▶ **Theorem 33.** *[45] For any well-rounded logconcave function f, given $\varepsilon, \delta > 0$ and a point x_0 with $f(x_0) \geq \beta \max f$, the annealing algorithm finds a point x in $O^*(n^{4.5})$ oracle calls such that with probability at least $1-\delta$, $f(x) \geq (1-\varepsilon)\max f$ and the dependence on ε, δ and β is bounded by a polynomial in $\ln(1/\varepsilon\delta\beta)$.*

This improves significantly on the complexity of the Ellipsoid method in the memebership oracle model (the latter being $\Omega(n^{10})$). We observe here that in this general oracle model, the upper bound on the complexity of optimization is higher than that of integration. The gap gets considerably higher when we move to star-shaped sets. Integration remains $O^*(n^4/\eta^2)$ where η is the relative measure of the kernel of the star-shaped set, while linear optimization, even approximately is NP-hard. The hardness holds for star-shaped sets with η being any constant [9], i.e., the convex kernel takes up any constant fraction of the set. At one level, this is not so surprising since finding a maximum could require zooming in to a small hidden portion of the set while integration is a more global property. On the other hand, historically, efficient integration algorithms came later than optimization algorithms even for convex bodies and were considerably more challenging to analyze.

The known analysis of sampling-based methods for optimization rely on the current point being nearly random from a suitable distribution, with the distribution modified appropriately at each step. It is conceivable that a random walk type method can be substantially faster if its goal is optimization and its analysis directly measures progress on some distance function, rather than relying on the intermediate step of producing a random sample.

The modern era of volume algorithms began when Lovász asked the question of whether the volume of convex body can be estimated from a random sample (unlike annealing in which the sampling distribution is modified several times). Eldan [17] has shown that this could require a superpolynomial number of points for general convex bodies. On the other hand, it remains open to efficiently compute the volume from a random sample for polytopes with a bounded number of facets.

Annealing has been used to speed up the reduction from counting to sampling for discrete sets as well [59]. Here the traditional reduction with overhead linear in the underlying dimension [24] is improved to one that is roughly the square root of the dimension for

estimating a wide class of partition functions. The annealing algorithm as described above has to be extended to be nonadaptive, i.e., the exponents used change in an adaptive manner rather than according to a schedule fixed in advance.

4.3 PCA

The method we discuss in this section is classical, namely Principal Component Analysis (PCA). For a given set of points or a distribution in $\mathbb{R}^n$, PCA identifies a sequence of orthonormal vectors such that for any $k \leq n$, the span of the first k vectors in this sequence is the subspace that minimizes the expected squared distance to the given point set or distribution (among all k-dimensional subsets). The vectors can be found using the Singular Value Decomposition (SVD), which can be viewed as a greedy algorithm that finds one vector at a time. More precisely, given a distribution D in $\mathbb{R}^n$, with $\mathsf{E}_D(X) = 0$, the top principal component or singular vector is a unit vector that maximizes $\mathsf{E}((v^T x)^2)$. For $2 \leq i \leq n$, the i'th principal component is a unit vector that maximizes the same function among all unit vectors that are orthogonal to the first $i-1$ principal components. For a general Gaussian, the principal components are exactly the directions along which the component 1-dimensional Gaussians are generated. Besides the regression property for subspaces, PCA is attractive because it can be computed efficiently by simple, iterative algorithms. Replacing the expected squared distance by a different power leads to an intractable problem. We have already seen one application of PCA, namely to rounding via the isotropic position. Here we take the principal components of the covariance matrix and apply a transformation to make their corresponding singular values all equal to one (and therefore the variance in any direction is the same). The principal components are the unit vectors along the axes of the inertial ellipsoid corresponding to the covariance matrix of a distribution.

We next discuss the application of PCA to the problem of learning an intersection of k halfpsaces in $\mathbb{R}^n$. As remarked earlier, this is an open problem even for $k = 2$. We make two assumptions. First, k is small compared to n and so algorithms that are exponential in k but not n might be tolerable. Second, instead of learning such an intersection from an arbitrary distribution on examples, we assume the distribution is an unknown Gaussian. This setting was considered by Vempala [61, 64] and by Klivans et al [38]. The first algorithm learns an intersection of k halfspaces from any logconcave input distribution using an intersection of $O(k)$ halfspaces and has complexity $(n/\varepsilon)^{O(k)}$. The second learns an intersection of k halfspaces from any Gaussian distribution using a polynomial threshold function of degree $O(\log k/\varepsilon^4)$ and therefore has complexity $n^{O(\log k/\varepsilon^4)}$. Neither of these algorithms is polynomial when k grows with n.

In recent work [63], PCA is used to give an algorithm whose complexity is $\mathrm{poly}(n, k, 1/\varepsilon) + C(k, 1/\varepsilon)$ where $C(.)$ is the complexity of learning an intersection of k halfspaces in $\mathbb{R}^k$. Thus one can bound this using prior results as at most

$$\min\left\{k^{O(\log k/\varepsilon^4)}, (k/\varepsilon)^{O(k)}\right\}.$$

For fixed ε, the algorithm is polynomial for k up to $2^{O(\sqrt{\log n})}$. The algorithm is straightforward: first put the full distribution in isotropic position (using a sample), effectively making the input distribution a standard Gaussian; then compute the *smallest* k principal components of the examples that lie in the intersection of the unknown halfspaces. The main claim in the analysis is that this latter subspace must be close to the span of the normals to the unknown halfspaces. This is essentially due to Lemma 11, which guarantees that the smallest k principal components are in the subspace spanned by the normals, while

orthogonal to this subspace, all variances are equal to that of the standard Gaussian. The algorithm and its analysis can be extended to arbitrary convex sets whose normals lie in an unknown k-dimensional subspace. The complexity remains a fixed polynomial in n times an exponential in k.

For general convex bodies, given positive and negative examples from an unknown Gaussian distribution, it is shown in [38] that the complexity is $2^{\tilde{O}(\sqrt{n})}$ (ignoring the dependence on ε, along with a nearly matching lower bound. A similar lower bound of $2^{\Omega(\sqrt{n})}$ holds for learning a convex body given only random points from the body [20]. These constructions are similar to Eldan's [17] and use polytopes with an exponential number of facets. Thus an interesting open problem is to learn a polytope in $\mathbb{R}^n$ with m facets given either uniform random points from it or from a Gaussian. It is conceivable that the complexity is $\text{poly}(m, n, 1/\varepsilon)$. The general theory of VC-dimension already tells us that $O(mn)$ samples suffice. Such an algorithm would of course also give us an algorithm for estimating the volume of a polytope from a set of random points and not require samples from a sequence of distributions.

Acknowledgements. The author is grateful to Daniel Dadush, Elena Grigorescu, Ravi Kannan, Jinwoo Shin and Ying Xiao for helpful comments.

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 N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.

3 D. Applegate and R. Kannan. Sampling and integration of near log-concave functions. In *STOC '91: Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 156–163, New York, NY, USA, 1991. ACM.

4 K. M. Ball. Logarithmically concave functions and sections of convex sets in rn. *Studia Mathematica*, 88:69–84, 1988.

5 A. Belloni, R. M. Freund, and S. Vempala. An efficient rescaled perceptron algorithm for conic systems. *Math. Oper. Res.*, 34(3):621–641, 2009.

6 D. Bertsimas and S. Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556, 2004.

7 S. Bobkov. On isoperimetric constants for log-concave probability distributions. *Geometric aspects of functional analysis, Lect. notes in Math.*, 1910:81–88, 2007.

8 J. Bourgain. Random points in isotropic convex sets. *Convex geometric analysis*, 34:53–58, 1996.

9 K. Chandrasekaran, D. Dadush, and S. Vempala. Thin partitions: Isoperimetric inequalities and a sampling algorithm for star shaped bodies. In *SODA*, pages 1630–1645, 2010.

10 K. Chandrasekaran, A. Deshpande, and S. Vempala. Sampling s-concave functions: The limit of convexity based isoperimetry. In *APPROX-RANDOM*, pages 420–433, 2009.

11 P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

12 A. Dinghas. Uber eine klasse superadditiver mengenfunktionale vonbrunn-minkowski-lusternik-schem typus. *Math. Zeitschr.*, 68:111–125, 1957.

13 J. Dunagan and S. Vempala. A simple polynomial-time rescaling algorithm for solving linear programs. *Math. Prog.*, 114(1):101–114, 2008.

14 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.

15 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.

16 M. E. Dyer, A. M. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, 1991.

17 R. Eldan. A polynomial number of random points does not determine the volume of a convex body. *http://arxiv.org/abs/0903.2634*, 2009.

18 R. Eldan and B. Klartag. Approximately gaussian marginals and the hyperplane conjecture. *http://arxiv.org/abs/1001.0875*, 2010.

19 B. Fleury. Concentration in a thin euclidean shell for log-concave measures. *J. Funct. Anal.*, 259(4):832–841, 2010.

20 N. Goyal and L. Rademacher. Learning convex bodies is hard. In *COLT*, 2009.

21 M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, 1988.

22 B. Grunbaum. Partitions of mass-distributions and convex bodies by hyperplanes. *Pacific J. Math.*, 10:1257–1261, 1960.

23 M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004.

24 M. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.*, 43:169–188, 1986.

25 A. T. Kalai and S. Vempala. Simulated annealing for convex optimization. *Math. Oper. Res.*, 31(2):253–266, 2006.

26 R. Kannan. Minkowski's convex body theorem and integer programming. *Math. Oper. Res.*, 12(3):415–440, 1987.

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

28 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.

29 R. Kannan and H. Narayanan. Random walks on polytopes and an affine interior point method for linear programming. In *STOC*, pages 561–570, 2009.

30 R. Kannan and S. Vempala. Sampling lattice points. In *STOC*, pages 696–700, 1997.

31 N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–396, 1984.

32 R. M. Karp and C. H. Papadimitriou. On linear characterization of combinatorial optimization problems. *SIAM J. Comp.*, 11:620–632, 1982.

33 J. A. Kelner and D. A. Spielman. A randomized polynomial-time simplex algorithm for linear programming. In *STOC*, pages 51–60, 2006.

34 L. G. Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20:53–72, 1980.

35 B. Klartag. On convex perturbations with a bounded isotropic constant. *Geom. and Funct. Anal.*, 16(6):1274–1290, 2006.

36 B. Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168:91–131, 2007.

37 B. Klartag. Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal.*, 245:284–310, 2007.

38 A. R. Klivans, Ryan O'Donnell, and R. A. Servedio. Learning geometric concepts via gaussian surface area. In *FOCS*, pages 541–550, 2008.

39 L. Leindler. On a certain converse of Hölder's inequality ii. *Acta Sci. Math. Szeged*, 33:217–223, 1972.

40 H. W. Lenstra. Integer programming with a fixed number of variables. *Math. of Oper. Res.*, 8(4):538–548, 1983.

41 L. Lovász. *An Algorithmic Theory of Numbers, Graphs and Convexity*, volume 50 of *CBMS-NSF Conference Series*. SIAM, 1986.

42 L. Lovász. Hit-and-run mixes fast. *Math. Prog.*, 86:443–461, 1998.

43 L. Lovász and M. Simonovits. On the randomized complexity of volume and diameter. In *Proc. 33rd IEEE Annual Symp. on Found. of Comp. Sci.*, pages 482–491, 1992.

44 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.

45 L. Lovász and S. Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *FOCS*, pages 57–68, 2006.

46 L. Lovász and S. Vempala. Hit-and-run from a corner. *SIAM J. Computing*, 35:985–1005, 2006.

47 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.

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

49 J. Luedtke, S. Ahmed, and G. L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Math. Program.*, 122(2):247–272, 2010.

50 E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.*, 177(1):1–43, 2009.

51 V. Milman and A. Pajor. Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. *Geometric aspects of Functional Analysis, Lect. notes in Math.*, pages 64–104, 1989.

52 Yu. Nesterov and A. Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. SIAM, 1994.

53 M. W. Padberg and M. R. Rao. The russian method for linear programming iii: Bounded integer programming. *NYU Research Report*, 1981.

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

55 A. Prekopa. Logarithmic concave measures and functions. *Acta Sci. Math. Szeged*, 34:335–343, 1973.

56 A. Prekopa. On logarithmic concave measures with applications to stochastic programming. *Acta Sci. Math. Szeged*, 32:301–316, 1973.

57 M. Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164:60–72, 1999.

58 A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82:93–133, 1989.

59 D. Stefankovic, S. Vempala, and E. Vigoda. Adaptive simulated annealing: A near-optimal connection between sampling and counting. In *FOCS*, pages 183–193, Washington, DC, USA, 2007. IEEE Computer Society.

60 P. M. Vaidya. A new algorithm for minimizing convex functions over convex sets. *Math. Prog.*, 73:291–341, 1996.

61 S. Vempala. A random sampling based algorithm for learning the intersection of half-spaces. In *FOCS*, pages 508–513, 1997.

62 S. Vempala. Geometric random walks: A survey. *MSRI Combinatorial and Computational Geometry*, 52:573–612, 2005.

63 S. Vempala. Learning convex concepts from gaussian distributions with PCA. In *FOCS*, 2010.

64 S. Vempala. A random sampling based algorithm for learning the intersection of half-spaces. *JACM, to appear*, 2010.

65 D. B. Yudin and A. S. Nemirovski. Evaluation of the information complexity of mathematical programming problems (in russian). *Ekonomika i Matematicheskie Metody*, 13(2):3–45, 1976.