Dispersion of Mass and the Complexity of Randomized Geometric Algorithms

Luis Rademacher [*,⋆]

Georgia Tech, Atlanta, GA 30332

Santosh Vempala [⋆⋆]

Georgia Tech, Atlanta, GA 30332

Abstract

How much can randomness help computation? Motivated by this general question and by volume computation, one of the few instances where randomness provably helps, we analyze a notion of dispersion and connect it to asymptotic convex geometry. We obtain a nearly quadratic lower bound on the complexity of randomized volume algorithms for convex bodies in $\mathbb{R}^n$ (the current best algorithm has complexity roughly n^4, conjectured to be n^3). Our main tools, dispersion of random determinants and dispersion of the length of a random point from a convex body, are of independent interest and applicable more generally; in particular, the latter is closely related to the variance hypothesis from convex geometry. This geometric dispersion also leads to lower bounds for matrix problems and property testing.

Key words: volume computation; lower bounds; randomized algorithms; variance hypothesis; dispersion of mass; determinant computation.

1 Introduction

Among the most intriguing questions raised by complexity theory is the following: how much can the use of randomness affect the computational complexity

[⋆] Supported by NSF ITR-0312354. At MIT when this research was performed.

[⋆⋆] Supported in part by a Guggenheim fellowship and by NSF CCF-0721503. At MIT when this research was performed.

[*] Corresponding author.

Email addresses: lrademac@cc.gatech.edu (Luis Rademacher), vempala@cc.gatech.edu (Santosh Vempala).

of algorithmic problems? At the present time, there are many problems for which randomized algorithms are simpler or faster than known deterministic algorithms but only a few known instances where randomness provably helps.

One problem for which randomness makes a dramatic difference is estimating the volume of a convex body in $\mathbb{R}^n$. The convex body can be accessed as follows: for any point $x \in \mathbb{R}^n$, we can determine whether x is in the body or not (a *membership* oracle). The complexity of an algorithm is measured by the number of such queries. The work of Elekes [12] and Bárány and Füredi [4] showed that any deterministic polynomial-time algorithm cannot estimate the volume to within an exponential (in n) factor. We quote their theorem below.

Theorem 1 ([4]) *For every deterministic algorithm that uses at most n^a membership queries and given a convex body K with $B_n \subseteq K \subseteq nB_n$ outputs two numbers A, B such that $A \leq \mathrm{vol}(K) \leq B$, there exists a body K' for which the ratio B/A is at least*

$$\left(\frac{cn}{a \log n}\right)^n$$

where c is an absolute constant.

In striking contrast, the celebrated paper of Dyer, Frieze and Kannan [10] gave a polynomial-time randomized algorithm to estimate the volume to arbitrary accuracy (the dependence on n was about n^{23}). This result has been much improved and generalized in subsequent work (n^{16}, [17]; n^{10}, [16,2]; n^8, [9]; n^7, [18]; n^5, [15]; n^4, [19]); the current fastest algorithm has complexity that grows as roughly $O(n^4/\epsilon^2)$ to estimate the volume to within relative error $1+\epsilon$ with high probability (for recent surveys, see [22,23]). Each improvement in the complexity has come with fundamental insights and lead to new isoperimetric inequalities, techniques for analyzing convergence of Markov chains, algorithmic tools for rounding and sampling logconcave functions, etc..

These developments lead to the question: what is the best possible complexity of any randomized volume algorithm? A lower bound of $\Omega(n)$ is straightforward. Here we prove a nearly quadratic lower bound: there is a constant $c > 0$ such that any randomized algorithm that approximates the volume to within a $(1 + c)$ factor needs $\Omega(n^2/\log n)$ queries. The formal statement appears in Theorem 2.

For the more restricted class of randomized nonadaptive algorithms (also called "oblivious"), an exponential lower bound is straightforward (Section 5.1). Thus, the use of full-fledged adaptive randomization is crucial in efficient volume estimation, but cannot improve the complexity below $n^2/\log n$.

In fact, the quadratic lower bound holds for a restricted class of convex bodies, namely parallelopipeds. A parallelopiped in $\mathbb{R}^n$ centered at the origin can be compactly represented using a matrix as $\{x : \|Ax\|_\infty \leq 1\}$, where A is an $n \times n$ nonsingular matrix; the volume is simply $2^n|\det(A)|^{-1}$. One way to interpret the lower bound theorem is that in order to estimate $|\det(A)|$ one needs almost as many bits of information as the number of entries of the matrix. The main ingredient of the proof is a dispersion lemma which shows that the determinant of a random matrix remains dispersed even after conditioning the distribution considerably. We discuss other consequences of the lemma in Section 8.

Our lower bound is nearly the best possible for this restricted class of convex bodies. Using $O(n^2 \log n)$ queries, we can find a close approximation to the entire matrix A and therefore any reasonable function of its entries. This naturally raises the question of what other parameters require a quadratic number of queries. We prove that estimating the product of the lengths of the rows of an unknown matrix A to within a factor of about $(1 + 1/\log n)$ also requires $\Omega(n^2/\log n)$ queries. The simplest version of this problem is the following: given a membership oracle for any unknown halfspace $a \cdot x \leq 1$, estimate $\|a\|$, the Euclidean length of the normal vector a (alternatively, estimate the distance of the hyperplane from the origin). This problem can be solved deterministically using $O(n \log n)$ oracle queries. We prove that any randomized algorithm that estimates $\|a\|$ to within an additive error of about $1/\sqrt{\log n}$ requires $\Omega(n)$ oracle queries.

Related earlier work includes [5,8], showing lower bounds for linear decision trees (i.e., every node of the tree tests whether an affine function of the input is nonnegative). [5] considers the problem of deciding whether given n real numbers, some k of them are equal, and they prove that it has complexity $\Theta(n \log(n/k))$. [8] proves that the n-dimensional knapsack problem has complexity at least $n^2/2$.

For these problems (length, product of lengths), the main tool in the analysis is a geometric dispersion lemma that is of independent interest in asymptotic convex geometry. Before stating the lemma, we give some background and motivation. There is an elegant body of work that studies the distribution of a random point X from a convex body K [3,6,7,21]. A convex body K is said to be in *isotropic* position if $\mathrm{vol}(K) = 1$ and for a random point X we have

$$\mathbb{E}(X) = 0, \qquad \text{and} \qquad \mathbb{E}(XX^T) = \alpha I \text{ for some } \alpha > 0.$$

We note that there is a slightly different definition of isotropy (more convenient for algorithmic purposes) which does not restrict $\mathrm{vol}(K)$ and replaces the second condition above by $\mathbb{E}(XX^T) = I$. Any convex body can be put in isotropic position by an affine transformation. A famous conjecture (*isotropic constant*) says that α is bounded by a universal constant for every convex body. It follows that $\mathbb{E}(\|X\|^2) = O(n)$. Motivated by the analysis of random walks, Lovász and Vempala made the following conjecture (under either definition). If true, then some natural random walks are significantly faster for isotropic convex bodies.

Conjecture 1 *For a random point X from an isotropic convex body,*

$$\mathrm{var}(\|X\|^2) = O(n).$$

The upper bound of $O(n)$ is achieved, for example, by the isotropic cube. The isotropic ball, on the other hand, has the smallest possible value, $\mathrm{var}(\|X\|^2) = O(1)$. The variance lower bound we prove in this paper (Theorem 6) directly implies the following: for an isotropic convex polytope P in $\mathbb{R}^n$ with at most $\mathrm{poly}(n)$ facets,

$$\mathrm{var}(\|X\|^2) = \Omega\left(\frac{n}{\log n}\right).$$

Thus, the conjecture is nearly tight for not just the cube, but *any* isotropic polytope with a small number of facets. Intuitively, our lower bound shows that the length of a random point from such a polytope is *not concentrated* as long as the volume is reasonably large. Roughly speaking, this says that in order to determine the length, one would have to localize the entire vector in a small region.

Returning to the analysis of algorithms, one can view the output of a randomized algorithm as a distribution. Proving a lower bound on the complexity is then equivalent to showing that the output distribution after some number of steps is *dispersed*. To this end, we define a simple parameter of a distribution:

Definition 1 *Let μ be a probability measure on $\mathbb{R}$. For any $0 < p < 1$, the p-*dispersion *of μ is*

$$\mathrm{disp}_\mu(p) = \inf\{|a - b| : a, b \in \mathbb{R}, \mu([a, b]) \geq 1 - p\}.$$

Thus, for any possible output z, and a random point X, with probability at least p, $|X - z| \geq \mathrm{disp}_\mu(p)/2$. We prove some useful properties about this parameter in Section 3.

2 Results

2.1 Complexity lower bounds

We begin with our lower bound for randomized volume algorithms. Besides the dimension n, the complexity also depends on the "roundness" of the input body. This is the ratio R/r where $rB_n \subseteq K \subseteq RB_n$. To avoid another parameter in our results, we ensure that R/r is bounded by a polynomial in n.

Theorem 2 (volume) *Let K be a convex body given by a membership oracle such that $B_n \subseteq K \subseteq O(n^8)B_n$. Then there exists a constant $c > 0$ such that any randomized algorithm that outputs a number V such that $(1-c)\,\mathrm{vol}(K) \leq V \leq (1+c)\,\mathrm{vol}(K)$ holds with probability at least $1 - 1/n$ has complexity $\Omega(n^2/\log n)$.*

We note that the lower bound can be easily extended to any algorithm with success probability $p > 1/2$ with a small overhead [14]. The theorem actually holds for parallelopipeds with the same roundness condition, i.e., convex bodies specified by an $n \times n$ real matrix A as $\{x \in \mathbb{R}^n : \forall\ 1 \leq i \leq n \quad |A_i \cdot x| \leq 1\}$ where A_i denotes the i'th row of A. In this case, the volume of K is simply $2^n|\det(A)|^{-1}$. We restate the theorem for this case.

Theorem 3 (determinant) *Let A be an matrix with entries in $[-1, 1]$ and smallest singular value at least $2^{-12}n^{-7}$ that can be accessed by the following oracle: for any x, the oracle determines whether $\|Ax\|_\infty \leq 1$ is true or false. Then there exists a constant $c > 0$ such that any randomized algorithm that outputs a number V such that*

$$(1-c)|\det(A)| \leq V \leq (1+c)|\det(A)|$$

holds with probability at least $1 - 1/n$, has complexity $\Omega(n^2/\log n)$.

A slightly weaker lower bound holds for estimating the product of the lengths of the rows of A. The proof is in Section 6.

Theorem 4 (product) *Let A be an unknown matrix that can be accessed by the following oracle: for any x, the oracle determines whether $||Ax||_\infty \leq 1$ is true or false. Then there exists a constant $c > 0$ such that any randomized algorithm that outputs a number L such that*

$$\left(1 - \frac{c}{\log n}\right) \prod_{i=1}^{n} \|A_i\| \leq L \leq \left(1 + \frac{c}{\log n}\right) \prod_{i=1}^{n} \|A_i\|$$

with probability at least $1 - 1/n$ has complexity $\Omega(n^2/\log n)$.

When A has only a single row, we get a stronger bound. In this case, the oracle is simply a membership oracle for a halfspace.

Theorem 5 (length) *Let a be a vector in $[-1,1]^n$ with $\|a\| \geq \sqrt{n} - 4\sqrt{\log n}$ and $a \cdot x \leq 1$ be the corresponding halfspace in $\mathbb{R}^n$ given by a membership oracle. Then there exists a constant $c > 0$ such that any randomized algorithm that outputs a number l such that*

$$\|a\| - \frac{c}{\sqrt{\log n}} \leq l \leq \|a\| + \frac{c}{\sqrt{\log n}}$$

with probability at least $1 - 1/n$ has complexity at least $n - 1$.

The restrictions on the input in all the above theorems ("roundness") only make them stronger. For example, the bound on the length of a above implies that it only varies in an interval of length $4\sqrt{\log n}$. To pin it down in an interval of length $c/\sqrt{\log n}$ (which is $O(\log \log n)$ bits of information) takes $\Omega(n)$ queries. This result is in the spirit of hardcore predicates [13].

It is worth noting that a very simple algorithm can approximate the length as in the theorem with probability at least $3/4$ and $O(n \log^2 n)$ queries: the projection of a onto a given vector b can be computed up to an additive error of $1/\operatorname{poly}(n)$ in $O(\log n)$ queries (binary search along the line spanned by b). If b is random in S_{n-1}, then $\mathbb{E}((a \cdot b)^2) = \|a\|^2/n$. A Chernoff-type bound gives that the average of $O(n \log n)$ random projections allows the algorithm to localize $\|a\|$ in an interval of length $O(1/\sqrt{\log n})$ with probability at least $3/4$.

2.2 Variance of polytopes

The next theorem states that the length of a random point from a polytope with few facets has large variance. This is a key tool in our lower bounds. It also has a close connection to the variance hypothesis (which conjectures an upper bound for all isotropic convex bodies), suggesting that polytopes might be the limiting case of that conjecture.

Theorem 6 *Let $P \subseteq \mathbb{R}^n$ be a polytope with at most n^k facets and contained in the ball of radius n^q. For a random point X in P,*

$$\operatorname{var} \|X\|^2 \geq \operatorname{vol}(P)^{\frac{4}{n} + \frac{3c}{n \log n}} e^{-c(k+3q)} \frac{n}{\log n}$$

where c is a universal constant.

Thus, for a polytope of volume at least 1 contained in a ball of radius at most $\operatorname{poly}(n)$, with at most $\operatorname{poly}(n)$ facets, we have $\operatorname{var} \|X\|^2 = \Omega(n/\log n)$.

In particular this holds for any isotropic polytope with at most poly(n) facets. The proof of Theorem 6 is given in Section 7.

2.3 Dispersion of the determinant

In our proof of the volume lower bound, we begin with a distribution on matrices for which the determinant is dispersed. The main goal of the proof is to show that even after considerable conditioning, the determinant is still dispersed. The next definition will be useful in describing the structure of the distribution and how it changes with conditioning.

Definition 2 *Let $\mathcal{M}$ be a set of $n \times n$ matrices. We say that $\mathcal{M}$ is a* product set along rows *if there exist sets $\mathcal{M}_i \subseteq \mathbb{R}^n$, $1 \le i \le n$,*

$$\mathcal{M} = \{M : \forall 1 \le i \le n, M_i \in \mathcal{M}_i\}.$$

Let B_n denote the n-dimensional Euclidean unit ball centered at the origin.

Lemma 7 *There exists a constant $c > 0$ such that for any partition $\{\mathcal{A}^j\}_{j \in N}$ of $(\sqrt{n}B_n)^n$ into $|N| \le 2^{n^2-2}$ parts where each part is a product set along rows, there exists a subset $N' \subseteq N$ such that*

a. $\mathrm{vol}(\bigcup_{j \in N'} \mathcal{A}^j) \ge \frac{1}{2}\mathrm{vol}\Big((\sqrt{n}B_n)^n\Big)$ *and*
b. for any $u > 0$ and a random point X from $\mathcal{A}^j$ for any $j \in N'$, we have

$$\Pr\Big(|\det X| \notin [u, u(1+c)]\Big) \ge \frac{1}{2^7 n^6}.$$

3 Preliminaries

Throughout the paper, we assume that $n > 12$ to avoid trivial complications.

We define $\pi_V(u)$ to be the projection of a vector u to a subspace V. Given a matrix R, let R_i denote the i'th row of R, and let $\hat{R}$ be the matrix having the rows of R normalized to be unit vectors. Let $\tilde{R}_i$ be the projection of R_i to the subspace orthogonal to $R_1, \ldots, R_{i-1}$. For any row R_i of matrix R, let R_{-i} denote (the span of) all rows except R_i. So $\pi_{R_{-i}^{\perp}}(R_i)$ is the projection of R_i orthogonal to the subspace spanned by all the other rows of R.

The volume of the Euclidean unit ball is given by $\pi^{n/2}/\Gamma(n/2 + 1)$, and its surface area is $2\pi^{n/2}/\Gamma(n/2)$.

3.1 Dispersion

We begin with two simple cases in which large variance implies large dispersion.

Lemma 8 *Let X be a real random variable with finite variance σ^2.*

a. If the support of X is contained in an interval of length M then

$$\operatorname{disp}_X\left(\frac{3\sigma^2}{4M^2}\right) \geq \sigma.$$

b. If X has a logconcave density then $\operatorname{disp}_X(p) \geq (1-p)\sigma$.

Proof. Let $a, b \in \mathbb{R}$ be such that $b - a < \sigma$. Let $\alpha = \Pr(X \notin [a, b])$. Then

$$\operatorname{var} X \leq (1-\alpha)\left(\frac{b-a}{2}\right)^2 + \alpha M^2.$$

This implies

$$\alpha > \frac{3\sigma^2}{4M^2}.$$

For the second part, Lemma 5.5(a) from [20] implies that a logconcave density with variance σ^2 is never greater than $1/\sigma$. This implies that if $a, b \in \mathbb{R}$ are such that $\Pr(X \in [a, b]) \geq p$ then we must have $b - a \geq p\sigma$. □

Lemma 9 *Let X, Y be real-valued random variables and Z be a random variable that is generated by setting it equal to X with probability α and equal to Y with probability $1-\alpha$. Then,*

$$\operatorname{disp}_Z(\alpha p) \geq \operatorname{disp}_X(p).$$

Lemma 10 *Let $f : [0, M] \to \mathbb{R}_+$ be a density function with mean μ and variance σ^2. Suppose the distribution function of f is logconcave. Then f can be decomposed into a convex combination of densities g and h, i.e., $f(x) = \alpha g(x) + (1-\alpha)h(x)$, where g is uniform over an interval $[a, b]$, with $a \geq \mu$ and $\alpha(a-b)^2 = \Omega\Big(\sigma^2/\log(M/\sigma)\Big)$.*

This lemma is proved in Section 6.

3.2 Yao's lemma

We will need the following version of Yao's lemma. Informally, the probability of failure of a randomized algorithm ν on the worst input is at least the probability of failure of the best deterministic algorithm against some distribution μ.

Lemma 11 *Let μ be a probability measure on inputs I (a "distribution on inputs") and let ν be a probability measure on deterministic algorithms A (a "randomized algorithm"). Then*

$$\begin{aligned}&\inf_{a\in A} \Pr(\textit{algorithm } a \textit{ fails on measure } \mu)\\ &\qquad\qquad\qquad \leq \sup_{i \in I} \Pr(\textit{randomized algorithm } \nu \textit{ fails on input } i).\end{aligned}$$

Let I be a set (a subset of the inputs of a computational problem, for example the set of all well-rounded convex bodies in $\mathbb{R}^n$ for some n). Let O be another set (the set of possible outputs of a computational problem, for example, real numbers that are an approximation to the volume of a convex body). Let A be a set of functions from I to O (these functions represent deterministic algorithms that take elements in I as inputs and have outputs in O). Let $C : I \times A \to \mathbb{R}$ (for $a \in A$ and $i \in I$, $C(i, a)$ is a measure of the badness of the algorithm a on input i, such as the indicator of a giving a wrong answer on i).

Lemma 12 *Let μ and ν be probability measures over I and A, respectively. Let $C : I \times A \to \mathbb{R}$ be integrable with respect to $\mu \times \nu$. Then*

$$\inf_{a\in A} \mathbb{E}_{\mu(i)} C(i, a) \leq \sup_{i\in I} \mathbb{E}_{\nu(a)} C(i, a)$$

Proof. By means of Fubini's theorem and the integrability assumption we have

$$\mathbb{E}_{\nu(a)} \mathbb{E}_{\mu(i)} C(i, a) = \mathbb{E}_{\mu(i)} \mathbb{E}_{\nu(a)} C(i, a).$$

Also

$$\mathbb{E}_{\nu(a)} \mathbb{E}_{\mu(i)} C(i, a) \geq \inf_{a\in A} \mathbb{E}_{\mu(i)} C(i, a)$$

and

$$\mathbb{E}_{\mu(i)} \mathbb{E}_{\nu(a)} C(i, a) \leq \sup_{i\in I} \mathbb{E}_{\nu(a)} C(i, a).$$

□

Proof (of Lemma 11) Let $C : I \times A \to \mathbb{R}$, where for $i \in I$, $a \in A$ we have

$$C(i,a) = \begin{cases} 1 & \text{if } a \text{ fails on } i \\ 0 & \text{otherwise.} \end{cases}$$

Then the consequence of Lemma 12 for this C is precisely what we want to prove. $\square$

3.3 The query model and decision trees

We have already discussed the standard query model (let us call it Q): A membership oracle for a convex body K takes any $q \in \mathbb{R}^n$ and outputs YES if $q \in K$ and NO otherwise. When K is a parallelopiped specified by a matrix A, the oracle outputs YES if $\|Aq\|_\infty \leq 1$ and NO otherwise.

It is useful to view the computation of a deterministic algorithm as a decision tree representing the sequence of queries: the nodes (except the leaves) represent queries, the root is the first query made by the algorithm and there is one query subtree per answer. The leaves do not represent queries but instead the answers to the last query along every path. Any leaf l has a set P_l of inputs that are consistent with the corresponding path of queries and answers on the tree. Thus the set of inputs is partitioned by the leaves.

To prove our main lower bound results for parallelopipeds, it will be convenient to consider a modified query model Q' that can output more information: Given $q \in \mathbb{R}^n$, the modified oracle outputs YES as before if $\|Aq\|_\infty \leq 1$; otherwise it outputs a pair (i, s) where i is the "least index among violated constraints", $i = \min\{j : |A_j q| > 1\}$, and $s \in \{-1, 1\}$ is the "side", $s = \operatorname{sign}(A_i q)$. An answer from Q' gives at least as much information as the respective answer from Q, and this implies that a lower bound for algorithms with access to Q' is also a lower bound for algorithms with access to Q. The modified oracle Q' has the following useful property (see Definition 2):

Lemma 13 *If the set of inputs is a product set along rows, then the leaves of a decision tree in the modified query model Q' induce a partition of the input set where each part is itself a product set along rows.*

Proof. We start with $\mathcal{M}$, a product set along rows with components $\mathcal{M}_i$. Let us observe how this set is partitioned as we go down a decision tree. A YES answer imposes two additional constraints of the form $-1 \leq q \cdot x \leq 1$ on every set $\mathcal{M}_i$. For a NO answer with response (i, s), we get two constraints for all $\mathcal{M}_j$, $1 \leq j < i$, one constraint for the i'th set and no new constraints for the remaining sets. Given this information, a particular setting of any row (or subset of rows) gives no additional information about the other rows. Thus, the set of possible matrices at each child of the current query is a product set along rows. The lemma follows by applying this argument recursively. □

Apart from the product property given by the previous lemma, if one assumes additionally that the set of inputs is convex, then in the query model Q' each part of the partition is a convex set. This property is used in the proof of the product lower bound (Theorem 4), but is not used in the volume lower bound (Theorem 2). Thus, for the volume lower bound one could use an oracle like Q' that outputs the index i but not the sign s, and the product property would be preserved (Lemma 13) but not the convexity.

3.4 Distributions and concentration properties

We use two distributions on $n \times n$ matrices called D and D' for the lower bounds in this paper. A random matrix from D is obtained by selecting each row independently and uniformly from the ball of radius $\sqrt{n}$. A random matrix from D' is obtained by selecting each entry of the matrix independently and uniformly from the interval $[-1, 1]$. In the analysis, we will also encounter random matrices where each entry is selected independently from $N(0, 1)$. We use the following property.

Lemma 14 *Let σ be the minimum singular value of an $n \times n$ matrix G with independent entries from $N(0,1)$. For any $t > 0$,*

$$\Pr\left(\sigma\sqrt{n} \leq t\right) \leq t.$$

Proof. To bound σ, we will consider the formula for the density of $\lambda = \sigma^2$ given in [11, Theorem 3.1]:

$$f(\lambda) = \frac{n}{2^{n-1/2}} \frac{\Gamma(n)}{\Gamma(n/2)} \lambda^{-1/2} e^{-\lambda n/2} U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}\right)$$

where U is the Tricomi function, which satisfies for all $\lambda \geq 0$:

- $U(\frac{n-1}{2}, -\frac{1}{2}, 0) = \Gamma(3/2)/\Gamma((n+2)/2)$,
- $U(\frac{n-1}{2}, -\frac{1}{2}, \lambda) \geq 0$
- $\frac{d}{d\lambda} U(\frac{n-1}{2}, -\frac{1}{2}, \lambda) \leq 0$

(The first two properties are from [11, Theorem 3.1], the third from [1, 13.1.3 and 13.4.21].)

We will now prove that for any n the density function of $t = \sqrt{n\lambda}$ is at most 1. To see this, the density of t is given by

$$g(t) = f\left(\frac{t^2}{n}\right)\frac{2t}{n} = 2f(\lambda)\sqrt{\frac{\lambda}{n}} = \frac{\sqrt{n}}{2^{n-3/2}}\frac{\Gamma(n)}{\Gamma(n/2)}e^{-\lambda n/2}U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}\right).$$

Now,

$$\frac{d}{dt}g(t) = \frac{\sqrt{n}}{2^{n-3/2}}\frac{\Gamma(n)}{\Gamma(n/2)}\times$$
$$\times\left[-\frac{n}{2}e^{-\lambda n/2}U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}\right) + e^{-\lambda n/2}\frac{d}{d\lambda}U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}\right)\right]\frac{2t}{n} \le 0.$$

Thus, the maximum of g is at $t = 0$, and

$$g(0) = \frac{\sqrt{n}}{2^{n-3/2}}\frac{\Gamma(n)}{\Gamma(n/2)}\frac{\Gamma(3/2)}{\Gamma(\frac{n+2}{2})} \le 1.$$

It follows that $\Pr(\sigma\sqrt{n} \le \alpha) \le \alpha$. □

Lemma 15 *Let X be a random n-dimensional vector with independent entries from $N(0,1)$. Then for $\epsilon > 0$*

$$\Pr\Big(\|X\|^2 \ge (1+\epsilon)n\Big) \le \Big((1+\epsilon)e^{-\epsilon}\Big)^{n/2}$$

and for $\epsilon \in (0,1)$

$$\Pr\Big(\|X\|^2 \le (1-\epsilon)n\Big) \le \Big((1-\epsilon)e^{\epsilon}\Big)^{n/2}.$$

For a proof, see [24, Lemma 1.3].

Lemma 16 *Let X be a uniform random vector in the n-dimensional ball of radius r. Let Y be an independent random n-dimensional unit vector. Then,*

$$\mathbb{E}(\|X\|^2) = \frac{nr^2}{n+2} \qquad \text{and} \qquad \mathbb{E}\Big((X\cdot Y)^2\Big) = \frac{r^2}{n+2}.$$

Proof. For the first part, we have

$$\mathbb{E}(\|X\|^2) = \frac{\int_0^r t^{n+1}dt}{\int_0^r t^{n-1}dt} = \frac{nr^2}{n+2}.$$

For the second part, because of the independence and the symmetry we can assume that Y is any fixed vector, say $(1,0,\ldots,0)$. Then $\mathbb{E}\Big((X\cdot Y)^2\Big) = \mathbb{E}(X_1^2)$.

But

$$\mathbb{E}(X_1^2) = \mathbb{E}(X_2^2) = \cdots = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2) = \frac{\mathbb{E}(\|X\|^2)}{n} = \frac{r^2}{n+2}.$$

□

Lemma 17 *There exists a constant $c > 0$ such that if $P \subseteq \mathbb{R}^n$ compact and X is a random point in P then*

$$\mathbb{E}\,\|X\|^2 \geq c(\operatorname{vol} P)^{2/n} n$$

Proof. For a given value of $\operatorname{vol} P$, the value $\mathbb{E}\,\|X\|^2$ is minimized when P is a ball centered at the origin. For some $c > 0$ we have that the volume of the ball of radius r is

$$\frac{\pi^{n/2} r^n}{\Gamma(n/2+1)} = \frac{2\pi^{n/2} r^n}{n\Gamma(n/2)} \geq \frac{2\pi^{n/2} r^n}{n(n/2)^{n/2}} \geq \frac{c^{n/2} r^n}{n^{n/2}}.$$

This implies that, for a given value of $\operatorname{vol} P$, the radius r of the ball of that volume satisfies

$$\frac{c^{n/2} r^n}{n^{n/2}} \geq \operatorname{vol} P. \tag{1}$$

On the other hand, Lemma 16 claims that for Y a random point in the ball of radius r, we have

$$\mathbb{E}\,\|Y\|^2 = \frac{nr^2}{n+2}. \tag{2}$$

Combining (1), (2) and the minimality of the ball, we get

$$\left(\frac{c\,\mathbb{E}\,\|X\|^2 (n+2)}{n^2}\right)^{n/2} \geq \operatorname{vol} P$$

and this implies the desired inequality. □

We conclude this section with two elementary properties of variance.

Lemma 18 *Let X, Y be independent real-valued random variables. Then*

$$\frac{\operatorname{var}(XY)}{(\mathbb{E}(XY))^2} = \left(1 + \frac{\operatorname{var} X}{(\mathbb{E}\,X)^2}\right)\left(1 + \frac{\operatorname{var} Y}{(\mathbb{E}\,Y)^2}\right) - 1 \geq \frac{\operatorname{var} X}{(\mathbb{E}\,X)^2} + \frac{\operatorname{var} Y}{(\mathbb{E}\,Y)^2}.$$

Lemma 19 *For real-valued random variables X, Y, $\operatorname{var} X = \mathbb{E}_Y \operatorname{var}(X \mid Y) + \operatorname{var}_Y \mathbb{E}(X \mid Y)$.*

4 Lower bound for length estimation

In this section, we prove Theorem 5. Let a be uniform random vector from $[-1,1]^n$. By Lemma 15, $\|a\| \geq \sqrt{n} - 4\sqrt{\log n}$ as required by the theorem with probability at least $1 - 1/n^2$. We will prove that there exists a constant $c > 0$ such that any deterministic algorithm that outputs a number l such that

$$\|a\| - \frac{c}{\sqrt{\log n}} \leq l \leq \|a\| + \frac{c}{\sqrt{\log n}}$$

with probability at least $1 - O(1/n \log n)$ makes at least $n-1$ halfspace queries. Along with Yao's lemma this proves the theorem.

Our access to a is via a membership oracle for the halfspace $a \cdot x \leq 1$. Consider the decision tree of height h for some deterministic algorithm. This will be a binary tree. The distribution at a leaf l is uniform over the intersection of $[-1,1]^n$ with the halfspaces given by the path (queries, responses) to the leaf l from the root r, i.e., uniform over a polytope P_l with at most $2n + h$ facets.

The volume of the initial set is 2^n. The volume of leaves with $\operatorname{vol}(P_l) < 1$ is less than $|L| = 2^h$ and so the total volume of leaves with $\operatorname{vol}(P_l) \geq 1$ is at least $2^n - 2^h$. Setting $h = n-1$, this is 2^{n-1} and so with probability at least $1/2$, $\operatorname{vol}(P_l) \geq 1$. For a random point X from any such P_l, Theorem 6 implies that $\operatorname{var} \|X\|^2 \geq cn/\log n$ for some absolute constant $c > 0$. Now by Lemma 8(a), and the fact that the support of $\|X\|^2$ is an interval of length n, we get that for any b,

$$\Pr\left(\left|\|X\|^2 - b\right| \geq \frac{1}{2}\sqrt{\frac{cn}{\log n}}\right) \geq \frac{3c}{4n \log n}.$$

It follows that $\|X\|$ is dispersed after $n-1$ queries. We note that the lower bound can be extended to any algorithm that succeeds with probability $1 - 1/n^\epsilon$ by a standard trick to boost the success probability: we repeat the algorithm $O(1/\epsilon)$ times and use the median of the results.

5 Complexity of randomized volume algorithms

We will use the distribution D on parallelopipeds (or matrices, equivalently). Recall that a random $n \times n$ matrix R is generated by choosing its rows $R_1, \ldots, R_n$ uniformly and independently from the ball of radius $\sqrt{n}$. The convex body corresponding to R is a parallelopiped having the rows of R as facets' normals:

$$\{x \in \mathbb{R}^n : (\forall i)|R_i \cdot x| \leq 1\}$$

Its volume is $V : \mathbb{R}^{n\times n} \to \mathbb{R}$ given (a.s.) by $V(R) = 2^n |\det R|^{-1}$.

At a very high level, the main idea of the lower bound is the following: after an algorithm makes all its queries, the set of inputs consistent with those queries is a product set along rows (in the oracle model Q'), while the level sets of the function that the algorithm is trying to approximate, $|\det(\cdot)|$, are far from being product sets. In the partition of the set of inputs induced by any decision tree of height $O(n^2/\log n)$, all parts are product sets along rows and most parts have large volume, and therefore V is dispersed in most of them. To make this idea more precise, we first examine the structure of a product set along rows all with *exactly* the same determinant. This abstract "hyperbola" has a rather sparse structure.

Theorem 20 *Let $R \subseteq \mathbb{R}^{n\times n}$ be such that $R = \prod_{i=1}^n R_i$, $R_i \subseteq \mathbb{R}^n$ convex and there exists $c > 0$ such that $|\det M| = c$ for all $M \in R$. Then, for some ordering of the R_i's, $R_i \subseteq S_i$, with S_i an $(i-1)$-dimensional affine subspace, $0 \notin S_i$ and satisfying: S_i is a translation of the linear hull of S_{i-1}.*

Proof. By induction on n. It is clearly true for $n = 1$. For arbitrary n, consider the dimension of the affine hull of each R_i, and let R_1 have minimum dimension. Let $a \in R_1$. There will be two cases:

If $R_1 = \{a\}$, then let A be the hyperplane orthogonal to a. If we denote T_i the projection of R_i onto A, then we have that $T = \prod_{i=1}^{n-1} T_i$ satisfies the hypotheses in $A \cong \mathbb{R}^{n-1}$ with constant $c/\|a\|$ and the inductive hypothesis implies that, for some ordering, the $T_2, \dots, T_n$ are contained in affine subspaces not containing 0 of dimensions $0, \dots, n-2$ in A, that is, $R_2, \dots, R_n$ are contained in affine subspaces not containing 0 of dimensions $1, \dots, n-1$.

If there are $a, b \in R_1$, $b \neq a$, then there is no zero-dimensional R_i. Also, because of the condition on the determinant, b is not parallel to a. Let $x_\lambda = \lambda a + (1-\lambda) b$ and consider the argument of the previous paragraph applied to x_λ and its orthogonal hyperplane. That is, for every λ there is some region T_i in A that is zero-dimensional. In other words, the corresponding R_i is contained in a line. Because there are only $n-1$ possible values of i but an infinite number of values of λ, we have that there exists one region R_i that is picked as the zero-dimensional for at least two different values of λ. That is, R_i is contained in the intersection of two non-parallel lines, and it must be zero-dimensional, which is a contradiction. $\square$

Now we need to extend this to an approximate hyperbola, i.e., a product set along rows with the property that for most of the matrices in the set, the determinant is restricted in a given interval. This extension is the heart of the proof and is captured in Lemma 7. We will need a bit of preparation for its proof.

We define two properties of a matrix $R \in \mathbb{R}^{n\times n}$:

- Property $P_1(R,t)$: $\prod_{i=1}^n \|\pi_{R_{-i}^\perp}(R_i)\| \le t$ ("short 1-D projections").
- Property $P_2(R,t)$: $|\det \hat{R}| \ge t$ ("angles not too small").

Lemma 21 *Let R be drawn from distribution D. Then for any $\alpha > 1$,*

a. $\Pr\Big(P_1(R,\alpha^n)\Big) \ge 1 - \frac{1}{\alpha^2}$,

b. there exists $\beta > 1$ (that depends on α) such that $\Pr\Big(P_2(R,1/\beta^n)\Big) \ge 1 - \frac{1}{n^\alpha}$.

Proof. For part (a), by the AM-GM inequality and Lemma 16 we have

$$\mathbb{E}\left(\left(\prod_i \|\pi_{R_{-i}^\perp}(R_i)\|^2\right)^{1/n}\right) \le \frac{1}{n}\sum_i \mathbb{E}\,\|\pi_{R_{-i}^\perp}(R_i)\|^2 = \frac{n}{n+2}.$$

Thus, by Markov's inequality,

$$\Pr\left(\prod_i \|\pi_{R_{-i}^\perp}(R_i)\| \ge c^n\right) = \Pr\left(\left(\prod_i \|\pi_{R_{-i}^\perp}(R_i)\|^2\right)^{1/n} \ge c^2\right) \le \frac{1}{c^2}.$$

For part (b), we can equivalently pick each entry of R independently as $N(0,1)$. In any case,

$$|\det \hat{R}| = \frac{|\det R|}{\prod_i \|R_i\|} = \frac{\prod_i \|\tilde{R}_i\|}{\prod_i \|R_i\|}.$$

We will find an upper bound for the denominator and a lower bound for the numerator.

For the denominator, Markov's inequality and the fact that $\mathbb{E}\prod \|R_i\|^2 = n^n$ give

$$\Pr\left(\prod_{i=1}^n \|R_i\|^2 \ge tn^n\right) \le 1/t. \tag{3}$$

For the numerator, let $\mu_i = \mathbb{E}\,\|\tilde{R}_i\|^2 = n - i + 1$, let $\mu = \mathbb{E}\prod_{i=1}^n \|\tilde{R}_i\|^2 = \prod_{i=1}^n \mu_i = n!$.

Now, concentration of a Gaussian vector (Lemma 15) gives

$$\Pr\Big(\|\tilde{R}_i\|^2 \ge \mu_i/2\Big) \ge 1 - 2^{-(n-i+1)/8} \tag{4}$$

Alternatively, for $t \in (0,1)$ the fact that the density of $N(0,1)$ is less 1 gives

$$\Pr\Big(\|\tilde{R}_i\|^2 \geq t\mu_i\Big) \geq 1 - \sqrt{t}(n-i+1). \tag{5}$$

Let $c > 0$ be such that $2^{-(n-i+1)/8} \leq 1/(2n^{\alpha+1})$ for $i \leq n - c\log n$. Using inequality (4) for $i \leq n - c\log n$ and (5) for the rest with

$$t = \frac{1}{2n^{2\alpha}(c\log n)^{5/2}}$$

we get

$$\begin{aligned} &\Pr\Bigg(\prod_{i=1}^{n}\|\tilde{R}_i\|^2 \geq \frac{\mu}{2^{n-c\log n}t^{c\log n}}\Bigg) \\ &\qquad \geq \prod_{i=1}^{n-c\log n}\Pr\Big(\|\tilde{R}_i\|^2 \geq \frac{\mu_i}{2}\Big)\prod_{i=n-c\log n}^{n}\Pr\Big(\|\tilde{R}_i\|^2 \geq t\mu_i\Big) \\ &\qquad \geq 1 - \frac{1}{n^\alpha} \end{aligned} \tag{6}$$

where, for some $\gamma > 1$ we have $2^{n-c\log n}t^{c\log n} \leq \gamma^n$. The result follows from equations (6) and (3). □

Proof (of Lemma 7) The idea of the proof is the following: If we assume that $|\det(\cdot)|$ of most matrices in a part fits in an interval $[u, u(1+\epsilon)]$, then for most choices R_{-n} of the first $n-1$ rows in that part we have that most choices Y of the last row in that part have $|\det(R_{-n}, Y)|$ in that interval. Thus, in view of the formula[1] $|\det(R_{-n}, Y)| = \|\tilde{Y}\| \prod_{i=1}^{n-1}\|\tilde{R}_i\|$ we have that, for most values of Y,

$$\|\tilde{Y}\| \in \Big[u, u(1+\epsilon)\Big]\prod_{i=1}^{n-1}\|\tilde{R}_i\|^{-1}$$

where $\tilde{Y}$ is the projection of Y to the line orthogonal to $R_1, \ldots, R_{n-1}$. In other words, most choices of the last row are forced to be contained in a set of the form $\{x : b \leq |a \cdot x| \leq c\}$, that we call a double band, and the same argument works for the other rows. In a similar way, we get a pair of double bands of "complementary" widths for every pair of rows. These constraints on the part imply that it has small volume, giving a contradiction. This argument only works for parts containing mostly "matrices that are not too singular" —matrices that satisfy P_1 and P_2—, and we choose the parameters of these properties so that at least half of $(\sqrt{n}B_n)^n$ satisfies them.

We will firstly choose N' as the family of large parts that satisfy properties P_1 and P_2 for suitable parameters so that (a) is satisfied. We will say "probability

[1] Recall that $\tilde{R}_i$ is the projection of R_i to the subspace orthogonal to $R_1, \ldots, R_{i-1}$.

of a subset of $(\sqrt{n}B_n)^n$" to mean its probability with respect to the uniform probability measure on $(\sqrt{n}B_n)^n$. The total probability of the parts having probability at most α is at most $\alpha|N|$. Thus, setting $\alpha = 1/(4|N|)$, the parts having probability at least $1/4|N| \geq 1/2^{n^2}$ have total probability at least $3/4$. Since $\operatorname{vol} \cup_{j \in N} \mathcal{A}^j \geq 2^{n^2}$, each of those parts has volume at least 1. Let these parts be indexed by $N'' \subseteq N$. We choose parameters in Lemma 21 (say, $\alpha = 4$ for part (a), $\alpha = 2$ for part (b), giving the existence of some β) so that at least $7/8$ of $(\sqrt{n}B_n)^n$ satisfy $P_1(\cdot, 4^n)$ and $P_2(\cdot, 1/\beta^n)$, and then at least $3/4$ of the parts in probability satisfy $P_1(\cdot, 4^n)$ and $P_2(\cdot, 1/\beta^n)$ for at least half of the part in probability. Let $N''' \subseteq N$ be the set of indices of these parts. Let $N' = N'' \cap N'''$. We have that $\cup_{j \in N'} \mathcal{A}^j$ has probability at least $1/2$.

We will now prove (b). Let $A = \prod_{i=1}^n A_i$ be one of the parts indexed by N'. Let X be random in A. Let ϵ be a constant and $p_1(n)$ be a function of n both to be fixed later. Assume for a contradiction that there exists u such that

$$\Pr\Big(|\det X| \notin [u, u(1+\epsilon)]\Big) < p_1(n). \tag{7}$$

Let $G \subseteq A$ be the set of $M \in A$ such that $|\det M| \in [u, u(1+\epsilon)]$. Let $p_2(n)$, $p_3(n)$ be functions of n to be chosen later. Consider the subset of points $R \in G$ satisfying:

I. $P_1(R, 4^n)$ and $P_2(R, 1/\beta^n)$,

II. for any $i \in \{1, \dots, n\}$, for at most a $p_2(n)$ fraction of $Y \in A_i$ we have $(Y, R_{-i}) \notin G$, and

III. for any $i, j \in \{1, \dots, n\}$, $i \neq j$, for at most a $p_3(n)$ fraction of $(Y, Z) \in A_i \times A_j$ we have $(Y, Z, R_{-ij}) \notin G$.

Because of the constraints, such a subset is a

$$1 - \Pr(X \notin G) - \Pr(X \in G \text{ and not as I, II and III}) \geq$$
$$\geq 1 - p_1(n) - \frac{1}{2} - n\frac{p_1(n)}{p_2(n)} - n^2 \frac{p_1(n)}{p_3(n)} \tag{8}$$

fraction of A. The function $p_1(n)$ will be chosen at the end so that the right hand side is positive. Fix a matrix $R = (R_1, \dots, R_n)$ in that subset.

The constraints described in the first paragraph of the proof are formalized in Lemma 22, which, for all i, j, gives sets B_{ij} (double bands, of the form $\{x : b \leq |a \cdot x| \leq c\}$), such that most of A_i is contained in $\cap_{j=1}^n B_{ij}$. Lemma 22 is invoked in the following way: For each pair i, j with $i < j$, let E be the two-dimensional subspace orthogonal to all the rows of R except i, j. We set X_1 (respectively X_2) distributed as the marginal in E of the uniform probability measure on A_i (respectively A_j). We also set $a_1 = \pi_E(R_i)$, $a_2 = \pi_E(R_j)$, $p = p_3(n)$, $q = p_2(n)$ and u and ϵ as here, while γ will be chosen later.

Let l_{ij} be the width of (each component of) the double band B_{ij}. Then, according to Lemma 22, the following relations hold:

$$\begin{aligned}
&l_{ii} \le \epsilon \|\pi_{R_{-i}^{\perp}}(R_i)\| && \text{for any } i,\\
&l_{ij} \le 4\epsilon \|\pi_{R_{-i}^{\perp}}(R_i)\| \|\pi_{R_{-j}^{\perp}}(R_j)\| / l_{ji} && \text{for } i > j.
\end{aligned}$$

Since each double band has two components, the intersection of all the n bands associated to a particular region A_i, namely $\cap_{j=1}^{n} B_{ij}$, is the union of 2^n congruent parallelopipeds. Thus, using properties P_1 and P_2 of R and fixing ϵ as a sufficiently small constant, the "feasible region" defined by the double bands, $B = \prod_{i=1}^{n} \cap_{j=1}^{n} B_{ij}$, satisfies:

$$\begin{aligned}
\operatorname{vol} B &\le 2^{n^2} \frac{\prod_{i,j=1}^{n} l_{ij}}{|\det \hat{R}|^n}\\
&\le 2^{n^2} \frac{\prod_{i=1}^{n} \left(\epsilon \|\pi_{R_{-i}^{\perp}}(R_i)\| \prod_{j=2}^{i} 4\epsilon \|\pi_{R_{-i}^{\perp}}(R_i)\| \|\pi_{R_{-j}^{\perp}}(R_j)\| \right)}{|\det \hat{R}|^n}\\
&= 2^{n^2} \frac{\epsilon^{\binom{n}{2}} 4^{\binom{n-1}{2}} \prod_i \|\pi_{R_{-i}^{\perp}}(R_i)\|^n}{|\det \hat{R}|^n}\\
&\le 1/4^n.
\end{aligned}$$

Each region A_i is not much bigger than the intersection of the corresponding double bands $B_i = \cap_{j=1}^{n} B_{ij}$ as follows: restricting to the double band B_{ii} removes at most a $p_2(n)$ fraction of A_i, each double band B_{ij} for $j < i$ removes at most a γ fraction of A_i, and each double band B_{ij} for $j > i$ removes a $p_2(n) + (p_3(n)/\gamma)$ fraction of A_i. We set $\gamma = 1/4n^2$, $p_2(n) = 1/(4n^2)$ and $p_3(n) = 1/(16n^4)$ so that, as a fraction of $\operatorname{vol} A_i$, $\operatorname{vol} B_i$ is no less than

$$1 - np_2(n) - \binom{n}{2}\gamma - \binom{n}{2}\left(p_2(n) + \frac{p_3(n)}{\gamma}\right) \ge 1/2.$$

Thus, $\operatorname{vol} A \le 2^n \operatorname{vol} B \le 1/2^n$, which is a contradiction. The condition on $p_1(n)$ given by Equation (8) is satisfied for $p_1(n) = 1/(2^7 n^6)$. □

Lemma 22 (2-D lemma) *Let X_1, X_2 be two independent random vectors in $\mathbb{R}^2$ with bounded support (not necessarily with the same distribution). Let X be a random matrix with rows X_1, X_2. Assume that there exist $u > 0$, $0 < \epsilon \le 1$ such that*

$$\Pr\Big(|\det X| \notin [u, u(1+\epsilon)]\Big) < p.$$

Let $G = \{M \in \mathbb{R}^{2\times 2} : |\det M| \in [u, u(1+\epsilon)]\}$. Let $a_1, a_2 \in \mathbb{R}^2$ be such that $(a_1, a_2) \in G$ and

$$\Pr(X_1 : (X_1, a_2) \notin G) \le q, \qquad \Pr(X_2 : (X_2, a_1) \notin G) \le q.$$

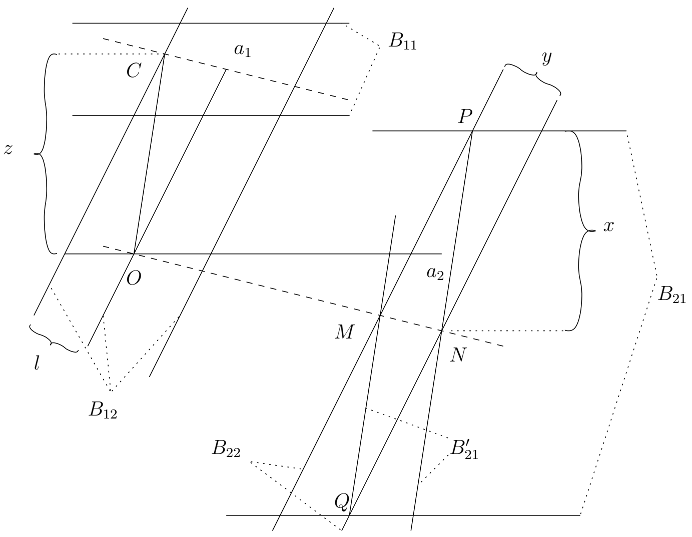

Fig. 1. The 2-D argument.

Let $\gamma > p/(1-q)$. *Then there exist double bands* $B_{ij} \subseteq \mathbb{R}^2$, $b_{ij} \geq 0$, $i, j \in \{1, 2\}$, $l \geq 0$,

$$\begin{aligned}
B_{11} &= \Big\{x : |a_2^\perp \cdot x| \in \Big[b_{11}, b_{11} + \epsilon \|\pi_{a_2^\perp}(a_1)\|\Big]\Big\} \\
B_{22} &= \Big\{x : |a_1^\perp \cdot x| \in \Big[b_{22}, b_{22} + \epsilon \|\pi_{a_1^\perp}(a_2)\|\Big]\Big\} \\
B_{12} &= \Big\{x : |a_1^\perp \cdot x| \in \Big[b_{12}, b_{12} + l\Big]\Big\} \\
B_{21} &= \Big\{x : |a_2^\perp \cdot x| \in \Big[b_{21}, b_{21} + 4\epsilon \|\pi_{a_2^\perp}(a_1)\| \|\pi_{a_1^\perp}(a_2)\|/l\Big]\Big\}
\end{aligned}$$

such that

$$\begin{aligned}
&\Pr(X_1 \notin B_{11}) \leq q && \Pr(X_1 \notin B_{12}) \leq q + (p/\gamma) \\
&\Pr(X_2 \notin B_{21}) \leq \gamma && \Pr(X_2 \notin B_{22}) \leq q.
\end{aligned}$$

Proof. The proof refers to Figure 1 which depicts the bands under consideration.

A double band of the form $\{x : |a \cdot x| \in [u, v]\}$ has (additive or absolute) width $v - u$ and relative (or multiplicative) width v/u. Consider the expansion $|\det X| = \|X_2\| \|\pi_{X_2^\perp}(X_1)\|$ and the definition of a_2 to get

$$\Pr\Big(\|\pi_{a_2^\perp}(X_1)\| \notin \|a_2\|^{-1}[u, u(1+\epsilon)]\Big) \leq q.$$

That is, with probability at most q we have X_1 outside of a double band of relative width $1 + \epsilon$:

$$B_{11} = \Big\{x : \|\pi_{a_2^\perp}(x)\| \in \|a_2\|^{-1}[u, u(1+\epsilon)]\Big\}.$$

Because $a_1 \in B_{11}$, the absolute width is at most $\epsilon\|\pi_{a_2^\perp}(a_1)\|$. If we exchange the roles of a_1 and a_2 in the previous argument, we get a double band B_{22}.

Let $\mathcal{A}$ be the set of $a \in \mathbb{R}^2$ satisfying: $(a, a_2) \in G$ and with probability at most γ over X_2 we have $(a, X_2) \notin G$. We have that

$$\Pr(X_1 \in \mathcal{A}) \geq 1 - q - \frac{p}{\gamma}.$$

Consider a point $C \in \mathcal{A}$ that maximizes the distance to the span of a_1. Similarly to the construction of B_{11}, by definition of $\mathcal{A}$ and with probability at most γ we have X_2 outside of a double band of relative width $1+\epsilon$. We denote it B'_{21}. In order to have better control of the angles between the bands, we want to consider a bigger double band parallel to B_{11}, the minimum such a band that contains the intersection of B_{22} and B'_{21}. Call this band B_{21}. Consider the line though the origin O parallel to $C - a_1$, and points M and N where the boundary of one component of the double band B_{22} intersects the line, M is the point closest to the origin, N, the farthest. The boundary of B'_{21} intersects the boundary of B_{11} precisely at $\pm M$ and $\pm N$, because for any vector $v \in \mathbb{R}^2$ parallel to $C - a_1$ we have $|\det(v, C)| = |\det(v, a_1)|$. Consider the components of B'_{21} and B_{22} containing M and N and let P be any of the other two points where the boundaries of those components meet. This implies that triangles Oa_1C and PMN are similar. The width of B_{21} is at most $2x$, where $x = \max\{\|\pi_{a_2^\perp}(P - M)\|, \|\pi_{a_2^\perp}(P - N)\|\}$. Then,

$$\frac{x}{z} = \frac{y}{l},$$

where $l = \|\pi_{a_1^\perp}(C)\|$ is the width of a band imposed on $\mathcal{A}$ by definition of C, y is the width of B_{22}, $y \leq \epsilon\|\pi_{a_1^\perp}(a_2)\|$, and z is the distance between C or a_1 and the span of a_2, whichever is larger, that is,

$$z = \max\{\|\pi_{a_2^\perp}(C)\|, \|\pi_{a_2^\perp}(a_1)\|\} \leq (1+\epsilon)\|\pi_{a_2^\perp}(a_1)\| \leq 2\|\pi_{a_2^\perp}(a_1)\|.$$

Thus, $x \leq 2\epsilon\|\pi_{a_2^\perp}(a_1)\|\|\pi_{a_1^\perp}(a_2)\|/l$. Let B_{12} be the band imposed on $\mathcal{A}$ by definition of C. $\square$

We are now ready to prove the complexity lower bounds.

Proof of Theorem 3. In view of Yao's lemma, it is enough to prove a lower bound on the complexity of deterministic algorithms against a distribution and then a lower bound on the minimum singular value of matrices according to that distribution. The deterministic lower bound is a consequence of the dispersion of the determinant proved in Lemma 7, the bound on the minimum singular value is an easy adaptation of a bound on the minimum singular value of a Gaussian matrix given by Lemma 14. These two claims are formalized below.

Claim 1: Let R be a random input according to distribution D. Then there exists a constant $c > 0$ such that any deterministic algorithm that outputs a number V such that

$$(1-c)|\det R| \leq V \leq (1+c)|\det R|$$

with probability at least $1 - 1/(2^8 n^6)$ makes more than

$$\frac{n^2-2}{\log_2(2n+1)}$$

queries in the oracle model Q'.

Claim 2: Let A be an $n \times n$ random matrix from distribution D. Let σ be the minimum singular value of A. Then for any $t \geq 0$

$$\Pr(\sigma\sqrt{n} \leq t) \leq 4t + \frac{n}{2^{n-1}}$$

(the choice of $t = 1/(2^{12} n^6)$ proves Theorem 3).

Proof of Claim 1: For a deterministic algorithm and a value of n, consider the corresponding decision tree. Let

$$h \leq \frac{n^2-2}{\log_2(2n+1)}$$

be the height and L be the set of leaves of this tree. Let $(P_l)_{l \in L}$ be the partition on the support of D induced by the tree.

Every query has at most $2n+1$ different answers, and every path has height at most h. Thus,

$$|L| \leq (2n+1)^h = 2^{n^2-2}.$$

The sets P_l are product sets along rows by Lemma 13, and hence by Lemma 7 we have that there exists a constant $c > 0$ such that with probability at least $1/(2^8 n^6)$ and for any $a > 0$ we have that $|\det R|$ is outside of $[a, (1+c)a]$. Claim 1 follows.

Proof of Claim 2: We will bound $\|A^{-1}\|_2 = 1/\sigma$. To achieve this, we will reduce the problem to the case where the entries of the matrix are $N(0,1)$ and independent. We write $A = GDE$, where G has its entries independently as $N(0,1)$, D is the diagonal matrix that normalizes the rows of G and E is another random diagonal matrix independent of (G, D) that scales the rows of GD to give them the length distribution of a random vector in $\sqrt{n}B_n$. We have

$$\|A^{-1}\|_2 \le \|D^{-1}\|_2 \|E^{-1}\|_2 \|G^{-1}\|_2. \tag{9}$$

Now, with probability at least $1 - n/2^n$ the diagonal entries of E are at least $\sqrt{n}/2$. Thus, except for an event that happens with probability $n/2^n$,

$$\|E^{-1}\|_2 \le 2/\sqrt{n}. \tag{10}$$

On the other hand, Lemma 15 (with $\epsilon = 3$) implies that with probability at least $1 - n/2^n$ the diagonal entries of D^{-1} are at most $2\sqrt{n}$. Thus, except for an event that happens with probability $n/2^n$,

$$\|D^{-1}\|_2 \le 2\sqrt{n}. \tag{11}$$

From (9), (10) and (11), we get $\|A^{-1}\|_2 \le 4\|E^{-1}\|$. Using Lemma 14 which bounds the singular values for a Gaussian matrix, Claim 2 follows. □

Finally, Theorem 2 is a simple consequence.

Proof of Theorem 2. It remains to prove that a parallelopiped given by a matrix A as in Theorem 3 contains $B_n/\sqrt{n}$ and is contained in $\frac{\sqrt{n}}{\sigma}B_n$ whenever $\sigma > 0$, where σ is the minimum singular value of A. The first inclusion is evident since the entries must be from $[-1, 1]$. It is sufficient to prove the second inclusion for the vertices of the parallelopiped, i.e., solutions to $Ax = b$ for any $b \in \{-1, 1\}^n$. That is, $x = A^{-1}b$ and therefore

$$\|x\| \le \|A^{-1}\|_2 \|b\| \le \sqrt{n}/\sigma.$$

□

5.1 Nonadaptive volume algorithms

An algorithm is *nonadaptive* if its queries are independent of the input.

Theorem 23 (nonadaptive lower bound) *Let K be a convex body given by a membership oracle such that $B_n \subseteq K \subseteq 2nB_n$. Then any nonadaptive randomized algorithm that outputs a number V such that $.9\,\mathrm{vol}(K) \le V \le 1.1\,\mathrm{vol}(K)$ holds with probability at least $3/4$ has complexity at least $\frac{1}{2e(n+2)}n^{n/2}$.*

Proof. Consider the distribution on parallelopipeds induced by the following procedure: first, with equal probability choose one of the following bodies:

- ("brick") $\Big\{x \in \mathbb{R}^n : (\forall i \in \{2, \dots, n\})\ |x_i| \leq 1\Big\} \cap nB_n$
- ("double brick") $\Big\{x \in \mathbb{R}^n : (\forall i \in \{2, \dots,\ n\})\ |x_i| \leq 1\Big\} \cap 2nB_n$

and then, independently of the first choice, apply a random rotation.

We will prove the following claim, from which the desired conclusion can be obtained by means of Yao's lemma.

Claim: Let K be a parallelopiped according to the previous distribution. Then any nonadaptive deterministic algorithm that outputs a number V such that

$$.9\,\mathrm{vol}(K) \leq V \leq 1.1\,\mathrm{vol}(K) \tag{12}$$

holds with probability more than $\frac{1}{2} + \frac{Qn}{2}(\frac{2}{n\pi})^{n/2}$ has complexity at least Q.

Proof of Claim: To satisfy Equation (12), the algorithm has to actually distinguish between the brick and the double brick. Let the *bad surface* be the intersection between the input and the sphere of radius n. In order to distinguish between the two bodies, the algorithm has to make at least one query whose ray hits the bad surface. We will prove that the probability of this event is no more than $2Q(2/e\pi n)^{n/2}$. To see this, observe that the probability of a query hitting the bad surface is at most the volume of the bad surface divided by the volume of the sphere of radius n. The former can be bounded in the following way: Let $x = (x_2, \dots, x_n)$ be the coordinates along the normals to the $n-1$ facets of the body. Parameterize one of the hemispheres determined by the hyperplane containing those normals as $F(x_2, \dots, x_n) = \sqrt{n^2 - x_2^2 - \cdots - x_n^2}$.

We have that

$$\frac{d}{dx_i} F(x) = \frac{x_i}{F(x)}.$$

In the domain of integration $[-1,1]^{n-1}$ we have $\|x\|^2 \leq n$ and this implies that in that domain

$$\|\nabla F(x)\|^2 = \frac{\|x\|^2}{n^2 - \|x\|^2} \leq \frac{1}{n-1}.$$

The volume of the bad surface is given by

$$2\int_{[-1,1]^{n-1}} \sqrt{1 + \|\nabla F(x)\|^2}\, dx \leq 2^n \sqrt{1 + \frac{1}{n-1}} \leq 2^{n+1}$$

The volume of the sphere of radius n is

$$\frac{2n^{n-1}\pi^{n/2}}{\Gamma(n/2)} \geq \frac{2n^{n-1}\pi^{n/2}}{(n/2)^{n/2}} = \frac{2}{n}(2n\pi)^{n/2}.$$

Thus, the probability that a particular query hits the bad surface is at most

$$n\left(\frac{2}{n\pi}\right)^{n/2}.$$

Therefore the algorithm gives the wrong answer with probability at least

$$\frac{1}{2}\left(1 - Qn\left(\frac{2}{n\pi}\right)^{n/2}\right).$$

□

6 Lower bound for the product

Proof. (of Lemma 10.) Let the distribution function be $F(t) = \Pr(X \leq t) = e^{g(t)}$ for some concave function g and the density is $f(t) = g'(t)e^{g(t)}$ where $g'(t)$ is nonincreasing. First, we observe that logconcavity implies that $F(\mu) \geq 1/4$. To see this, let $\mu - l$ be the point where $F(\mu - l) = F(\mu)/2$. Then, $F(\mu - il) \leq F(\mu)/2^i$ and

$$\begin{aligned}\int_0^\mu (\mu - x)f(x)\,dx &\leq \sum_{i\geq 1}\Big(F(\mu-(i-1)l) - F(\mu - il)\Big)(il)\\ &\leq F(\mu)l + \sum_{i>1} F(\mu - il)\Big((i+1) - i\Big)l\\ &\leq F(\mu)l\sum_{i\geq 0}\frac{1}{2^i} = 2lF(\mu).\end{aligned}$$

On the other hand (assuming $F(\mu) \leq 1/4$, otherwise, there is nothing to prove),

$$\int_\mu^\infty (x-\mu)f(x)\,dx \geq \sum_{i=1}^{\lfloor \log(1/F(\mu))\rfloor}(2^i - 2^{i-1})F(\mu)(i-1)l \geq \frac{\log\Big(1/F(\mu)\Big)}{2}l.$$

Therefore, we must have $2F(\mu) \geq \log(1/F(\mu))/2$ which implies $F(\mu) \geq 1/4$.

Next,

$$\int_0^\mu (\mu - x)f(x)\,dx \geq \int_0^{\mu - l}(\mu - x)f(x)\,dx \geq F(\mu - l)l \geq \frac{l}{8}.$$

Therefore, since μ is the mean,

$$\int_\mu^\infty (x - \mu)f(x)\,dx \geq \frac{l}{8}.$$

It follows that

$$\int_{\mu}^{\infty} (x-\mu)^2 f(x)\, dx \geq \frac{l^2}{64}. \tag{13}$$

Suppose $l < \sigma/4$. Then,

$$\begin{aligned}\int_0^{\mu} (x-\mu)^2 f(x)\, dx &\leq \sum_{i\geq 1} \Big(F(\mu-(i-1)l) - F(\mu - il)\Big)(il)^2 \\ &\leq F(\mu)l^2 + \sum_{i>1} F(\mu - il)\Big((i+1)^2 - i^2\Big)l^2 \\ &\leq F(\mu)l^2 \sum_{i\geq 1} \frac{2i+1}{2^i} = 5l^2 F(\mu) \leq \sigma^2/2.\end{aligned}$$

Since

$$\sigma^2 = \int_0^{\infty} (x-\mu)^2 f(x)\, dx = \int_0^{\mu} (x-\mu)^2 f(x)\, dx + \int_{\mu}^{\infty} (x-\mu)^2 f(x)\, dx,$$

we must have

$$\int_{\mu}^{\infty} (x-\mu)^2 f(x)\, dx \geq \frac{\sigma^2}{2}.$$

Using this and (13), we have (regardless of the magnitude of l),

$$\int_{\mu}^{\infty} (x-\mu)^2 f(x) \geq \frac{\sigma^2}{2^{10}}. \tag{14}$$

Now we consider intervals to the right of μ. Let $J_0 = (\mu, x_0]$ where x_0 is the smallest point to the right of μ for which $f(x_0) \leq 1/\sigma$ (J_0 could be empty). Let J_i, for $i = 1, 2, \ldots, m = 3\log(M/\sigma) + 14$ be $[x_{i-1}, x_i]$ where x_i is the smallest point for which $f(x_i) \leq 1/(\sigma 2^i)$. For any $t \geq t' \geq \mu$, $f(t') \geq f(t)F(t')/F(t) \geq f(t)F(\mu) \geq f(t)/4$. Therefore, the function f is approximately constant in any interval J_i for $i \geq 1$. If $x_0 > \mu + \sigma/64$, then the interval $[\mu, \mu + \sigma/64]$ satisfies the desired property (as $f(x) \geq f(x_0)$ for x in this interval, we can take $\alpha = f(x_0)\sigma/64 = 1/64$). Otherwise,

$$\int_{J_0} (x-\mu)^2 f(x)\, dx \leq \sigma^2/2^{12}.$$

Also,

$$\int_{x_m}^{\infty} (x-\mu)^2 f(x)\, dx \leq 4M^3 f(x_m) \leq \sigma^2/2^{12}.$$

Therefore, from (14), for some $i^* \geq 1$ we have

$$\int_{J_{i^*}} (x-\mu)^2 f(x)\, dx \geq \frac{\sigma^2}{2^{12} m}.$$

The interval $[\mu, x_{i^*}]$ then completes the proof: For this interval we can take $\alpha = f(x_{i^*})(x_{i^*} - \mu)$, and we have

$$\begin{aligned}\int_{J_{i^*}} (x-\mu)^2 f(x)\,dx &\leq 8(x_{i^*}-\mu)^2(x_{i^*}-x_{i^*-1})f(x_{i^*})\\ &\leq 8\alpha(x_{i^*}-\mu)^2.\end{aligned}$$

□

Proof of Theorem 4 For this lower bound, we use the distribution D' on matrices. Let R be an $n \times n$ random matrix having each entry uniformly and independently in $[-1, 1]$. On input R from distribution D' having rows $(R_1, \ldots, R_n)$ and with probability at least $1/2$ over the inputs, we consider algorithms that output an approximation to $f(R) = \prod_i \|R_i\|$. The next claim for deterministic algorithms, along with Yao's lemma, proves Theorem 4.

Claim: Suppose that a deterministic algorithm makes at most

$$h := \frac{\frac{n^2}{2} - 1}{\log_2(2n+1)}$$

queries on any input R and outputs V. Then there exists a constant $c > 0$ such that the probability of the event

$$\left(1 - \frac{c}{\log n}\right) f(R) \leq V \leq \left(1 + \frac{c}{\log n}\right) f(R)$$

is at most $1 - O(1/n)$.

To prove the claim, we consider a decision tree corresponding to a deterministic algorithm. Let P_l be the set of matrices associated with a leaf l. By Lemma 13, we have that the set P_l is a product set along rows, that is $P_l = \prod_i \mathcal{R}_i$, where $\mathcal{R}_i \subseteq \mathbb{R}^n$ is the set of possible choices of the row R_i consistent with l. The conditional distribution of R at a leaf l consists of *independent*, uniform choices of the rows from their corresponding sets. Moreover, the sets $\mathcal{R}_i$ are polytopes with at most $f = 2n + 2h$ facets. Every query has at most $2n + 1$ different answers, and every path has height at most h. Thus, $|L| \leq (2n+1)^h = 2^{\frac{n^2}{2}-1}$. The total probability of the leaves having probability at most α is at most $\alpha|L|$. Thus, setting $\alpha = 1/(2|L|)$, the leaves having probability at least

$$\frac{1}{2|L|} \geq \frac{1}{2^{n^2/2}}$$

have total probability at least $1/2$. Because $\operatorname{vol} \cup_{l \in L} P_l = 2^{n^2}$, we have that those leaves have volume at least $2^{n^2/2}$. Further, since $P_l = \prod_i \mathcal{R}_i$, we have that for such P_l at least $n/2$ of the $\mathcal{R}_i$'s have volume at least 1. Theorem

6 implies that for those $\operatorname{var}\|R_i\|^2 \geq \Omega(n/\log n)$. Along with the fact that $\|R_i\| \leq \sqrt{n}$ and Lemma 18, for a random matrix R from such a P_l, we get

$$\frac{\operatorname{var}\Big(f(R)^2\Big)}{\Big(\mathbb{E}(f(R)^2)\Big)^2} \geq \sum_i \frac{\operatorname{var}(\|R_i\|^2)}{\Big(\mathbb{E}(\|R_i\|^2)\Big)^2} = \Omega\left(\frac{1}{\log n}\right).$$

Thus, the variance of $f(R)$ is large. However, this does not directly imply that $f(R)$ is dispersed since the support of $f(R)$ could be of exponential length and its distribution is not logconcave.

Let $X = \prod_{i=1}^n X_i$ where $X_i = \|R_i\|^2$. To prove the lower bound, we need to show that $\operatorname{disp}_X(p)$ is large for p at least inverse polynomial in n. For i such that $\operatorname{vol}(\mathcal{R}_i) \geq 1$, we have $\operatorname{var} X_i = \Omega(n/\log n)$ by Theorem 6. As remarked earlier at least $n/2$ sets satisfy the volume condition and we will henceforth focus our attention on them. We also get

$$\mathbb{E}(X_i) \geq n/16 \tag{15}$$

from this. The distribution function of each X_i is logconcave (although not its density) and its support is contained in $[0, n]$. So by Lemma 10, we can decompose the density f_i of each X_i as $f_i(x) = p_i g_i(x) + (1 - p_i) g_i'(x)$. where g_i is the uniform distribution over an interval $[a_i, b_i]$ of length L_i and

$$p_i L_i^2 = \Omega\left(\frac{n}{\log^2 n}\right) \quad \text{and} \quad p_i = \Omega\left(\frac{1}{n\log^2 n}\right).$$

We will assume that $p_i L_i^2 = cn/\log^2 n$ and $p_i = \Omega(1/n^2)$. This can be achieved by noting that L_i is originally at most n and truncating the interval suitably. Let X_i' be a random variable drawn uniformly from the interval $[a_i, b_i]$. Let $Y_i = \log X_i'$, I be a subset of $\{1, 2, \ldots, n\}$ and $Y_I = \sum_{i\in I} \log X_i'$. The density of Y_i is $h_i(t) = e^t/L_i$ for $\log a_i \leq t \leq \log b_i$ and zero outside this range. Thus Y_i has a logconcave density and so does Y_I (the sum of random variables with logconcave density also has a logconcave density). Also, $\operatorname{var}(Y_I) = \sum_{i\in I} \operatorname{var}(Y_i)$. To bound the variance of Y_i, we note that since $a_i \geq \mathbb{E}(X_i) \geq n/16$ by Lemma 10 and Equation (15), we have $b_i \leq 16a_i$ and so $h_i(t)$ varies by a factor of at most 16. Thus, we can decompose h_i further into h_i' and h_i'' where h_i' is uniform over $[\log a_i, \log b_i]$ and

$$h_i(x) = \frac{1}{16} h_i'(x) + \frac{15}{16} h_i''(x).$$

Let Y_i' have density h_i'. Then

$$\operatorname{var}(Y_i) \geq \frac{1}{16} \operatorname{var}(Y_i') = \frac{(\log b_i - \log a_i)^2}{192}.$$

Therefore

$$\mathrm{var}(Y_I) \geq \frac{1}{192} \sum_{i \in I} (\log b_i - \log a_i)^2$$

From this we get a bound on the dispersion of Y_I using the logconcavity of Y_I and Lemma 8(b). The bound depends on the set I of indices that are chosen. This set is itself a random variable defined by the decompositions of the X_i's. We have

$$\mathbb{E}_I\Big(\mathrm{var}(Y_I)\Big) \geq \frac{1}{192} \sum_{i=1}^{n} p_i (\log b_i - \log a_i)^2 \geq \frac{1}{192} \sum_{i=1}^{n} p_i \frac{L_i^2}{(8a_i)^2} \geq \frac{c_1}{\log^2 n}$$

On the other hand,

$$\begin{aligned}
\mathrm{var}_I\Big(\mathrm{var}(Y_I)\Big) &\leq \sum_{i=1}^{n} p_i (\log b_i - \log a_i)^4 \\
&\leq \sum_{i=1}^{n} p_i \frac{L_i^4}{a_i^4} \\
&\leq \frac{16^4}{n^4} \sum_{i=1}^{n} \frac{p_i^2 L_i^4}{p_i} \\
&= \frac{16^4}{n^4} \frac{c^2 n^2}{\log^4 n} \sum_{i=1}^{n} \frac{1}{p_i}.
\end{aligned}$$

Suppose $p_i \geq c_2/n$ for all i. Then we get,

$$\mathrm{var}_I\Big(\mathrm{var}(Y_I)\Big) \leq \frac{c_2'}{\log^4 n}$$

and for c_2 large enough, $\mathrm{var}_I\Big(\mathrm{var}(Y_I)\Big) \leq \Big(\mathbb{E}_I \, \mathrm{var}(Y_I)\Big)^2/4$. Hence, using Chebychev's inequality, with probability at least $1/4$, $\mathrm{var}(Y_I) \geq c_1/(4 \log^2 n)$. By Lemma 8(b), with probability at least $1/4$, we have $\mathrm{disp}_{Y_I}(1/2) \geq \frac{\sqrt{c_1}}{4 \log n}$. This implies that for any u,

$$\Pr\left(X \in \left[u, u\left(1 + \frac{\sqrt{c_1}}{4 \log n}\right)\right]\right) \leq \frac{7}{8}.$$

Finally, if for some i, $p_i < c_2/n$, then for that Y_i, $L_i^2 = \Omega(n^2/\log^2 n)$ and using just that i, we get $\mathrm{disp}_{Y_i}(p_i/2) \geq \sqrt{L_i^2/a_i^2} = \Omega(1/\log^2 n)$ and once again X is dispersed as well (recall that $p_i = \Omega(1/n^2)$). $\square$

7 Variance of polytopes

Let $X \in K$ be a random point in a convex body K. Consider the parameter σ_K of K defined as

$$\sigma_K^2 = \frac{n \operatorname{var} \|X\|^2}{\left(\mathbb{E} \|X\|^2\right)^2}.$$

It has been conjectured that if K is isotropic, then $\sigma_K^2 \leq c$ for some universal constant c independent of K and n (the *variance hypothesis*). Together with the isotropic constant conjecture, it implies Conjecture 1. Our lower bound (Theorem 6) shows that the conjecture is nearly tight for isotropic polytopes with at most poly(n) facets and they might be the limiting case.

We now give the main ideas of the proof of Theorem 6. It is well-known that polytopes with few facets are quite different from the ball. Our theorem is another manifestation of this phenomenon: the width of an annulus that captures most of a polytope is much larger than one that captures most of a ball. The idea of the proof is the following: if $0 \in P$, then we bound the variance in terms of the variance of the cone induced by each facet. This gives us a constant plus the variance of the facet, which is a lower-dimensional version of the original problem. This is the recurrence in our Lemma 24. If $0 \notin P$ (which can happen either at the beginning or during the recursion), we would like to translate the polytope so that it contains the origin without increasing $\operatorname{var} \|X\|^2$ too much. This is possible if certain technical conditions hold (case 3 of Lemma 24). If not, the remaining situation can be handled directly or reduced to the known cases by partitioning the polytope. It is worth noting that the first case ($0 \in P$) is not generic: translating a convex body that does not contain the origin to a position where the body contains the origin may increase $\operatorname{var} \|X\|^2$ substantially. The next lemma states the basic recurrence used in the proof.

Lemma 24 (recurrence) *Let $T(n, f, V)$ be the infimum of $\operatorname{var} \|X\|^2$ among all polytopes in $\mathbb{R}^n$ with volume at least V, with at most f facets and contained in the ball of radius $R > 0$. Then there exist constants $c_1, c_2, c_3 > 0$ such that*

$$\begin{aligned} T(n, f, V) \geq \left(1 - \frac{c_1}{n}\right) T\left(n-1, f+2, \frac{c_2}{nR^2}\left(\frac{V}{Rf}\right)^{1+\frac{2}{n-1}}\right) \\ + \frac{c_3}{R^{8/(n-1)}}\left(\frac{V}{Rf}\right)^{\frac{4}{n-1}+\frac{8}{(n-1)^2}}. \end{aligned}$$

(Of course, T depends on R, but we omit that dependence to simplify the notation, given that, in contrast with the other parameters, R is the same for all appearances of T.)

Proof. Let P be a polytope as in the statement (not necessarily minimal). Let U be the nearest point to the origin in P. We will use more than one argument, depending on the case:

Case 1: (origin) $0 \in P$.

For every facet F of P, consider the cone C_F obtained by taking the convex hull of the facet and the origin. Consider the affine hyperplane H_F determined by F. Let U be the nearest point to the origin in H_F. Let Y_F be a random point in C_F, and decompose it into a random point $X_F + U$ in F and a scaling factor $t \in [0,1]$ with a density proportional to t^{n-1}. That is, $Y_F = t(X_F + U)$. We will express $\operatorname{var} \|Y_F\|^2$ as a function of $\operatorname{var} \|X_F\|^2$.

We have that $\|Y_F\|^2 = t^2(\|U\|^2 + \|X_F\|^2)$. Then,

$$\begin{aligned}\operatorname{var} \|Y_F\|^2 =&(\mathbb{E}\, t^4) \operatorname{var} \|X_F\|^2 \\ &+ (\operatorname{var} t^2)\Big(\|U\|^4 + (\mathbb{E}\, \|X_F\|^2)^2 + 2\|U\|^2\, \mathbb{E}\, \|X_F\|^2\Big)\end{aligned} \tag{16}$$

Now, for $k \geq 0$

$$\mathbb{E}\, t^k = \frac{n}{n+k}.$$

and

$$\operatorname{var} t^2 = \frac{4n}{(n+4)(n+2)^2} \geq \frac{c_1}{n^2}$$

for $c_1 = 1/2$ and $n \geq 3$. This in (16) gives

$$\begin{aligned}\operatorname{var} \|Y_F\|^2 &\geq \frac{n}{n+4} \operatorname{var} \|X_F\|^2 + \frac{c_1}{n^2}\left(\|U\|^4 + (\mathbb{E}\, \|X_F\|^2)^2 + 2\|U\|^2\, \mathbb{E}\, \|X_F\|^2\right) \\ &\geq \frac{n}{n+4} \operatorname{var} \|X_F\|^2 + \frac{c_1}{n^2}\Big(\mathbb{E}\, \|X_F\|^2\Big)^2.\end{aligned} \tag{17}$$

Now, by means of Lemma 17, we have that

$$\mathbb{E}\, \|X_F\|^2 \geq c_2 V_{n-1}(F)^{2/(n-1)}(n-1)$$

and this in (17) implies for some constant $c_3 > 0$ that

$$\operatorname{var} \|Y_F\|^2 \geq \frac{n}{n+4} \operatorname{var} \|X_F\|^2 + c_3 V_{n-1}(F)^{4/(n-1)}.$$

Using this for all cones induced by facets we get

$$\begin{aligned}\operatorname{var} \|X\|^2 &\geq \frac{1}{\operatorname{vol} P} \sum_{F \text{ facet}} \operatorname{vol} C_F \operatorname{var} \|Y_F\|^2 \\ &\geq \frac{1}{\operatorname{vol} P} \sum_{F \text{ facet}} \operatorname{vol} C_F \left(\frac{n}{n+4} \operatorname{var} \|X_F\|^2 + c_3 V_{n-1}(F)^{4/(n-1)}\right)\end{aligned} \tag{18}$$

Now we will argue that $\operatorname{var}\|X_F\|^2$ is at least $T(n-1, f, \frac{V}{Rf})$ for most facets. Because the height of the cones is at most R, we have that the volume of the cones associated to facets having $V_{n-1}(F) \leq \operatorname{vol} P/\alpha$ is at most

$$f\frac{1}{n}R\frac{\operatorname{vol} P}{\alpha}$$

That is, the cones associated to facets having $V_{n-1}(F) > \operatorname{vol} P/\alpha$ are at least a

$$1 - \frac{Rf}{\alpha n}$$

fraction of P. For $\alpha = Rf$ we have that a $1 - 1/n$ fraction of P is composed of cones having facets with $V_{n-1}(F) > \operatorname{vol} P/(Rf)$. Let $\mathcal{F}$ be the set of these facets. The number of facets of any facet F of P is at most f, which implies that for $F \in \mathcal{F}$ we have

$$\operatorname{var}\|X_F\|^2 \geq T(n-1, f, \frac{V}{Rf}).$$

Then (18) becomes

$$\begin{aligned}
\operatorname{var}\|X\|^2 &\geq \frac{1}{\operatorname{vol} P}\sum_{F\in\mathcal{F}} \operatorname{vol} C_F \left(\frac{n}{n+4}\operatorname{var}\|X_F\|^2 + c_3 V_{n-1}(F)^{4/(n-1)}\right) \\
&\geq \frac{1}{\operatorname{vol} P}\sum_{F\in\mathcal{F}} \operatorname{vol} C_F \left(\frac{n}{n+4}T\left(n-1, f, \frac{V}{Rf}\right) + c_3\left(\frac{V}{Rf}\right)^{4/(n-1)}\right) \\
&\geq \left(1 - \frac{1}{n}\right)\left(\frac{n}{n+4}T\left(n-1, f, \frac{V}{Rf}\right) + c_3\left(\frac{V}{Rf}\right)^{4/(n-1)}\right) \\
&\geq \left(1 - \frac{c_5}{n}\right)T\left(n-1, f, \frac{V}{Rf}\right) + c_4\left(\frac{V}{Rf}\right)^{4/(n-1)}
\end{aligned}$$

for some constants $c_5, c_4 > 0$.

Case 2: (slicing)

$$\operatorname{var}\mathbb{E}\Big(\|X\|^2 \mid X\cdot U\Big) \geq \beta = \frac{c_4}{16}\left(\frac{V}{Rf}\right)^{4/(n-1)}.$$

In this case, using Lemma 19,

$$\begin{aligned}
\operatorname{var}\|X\|^2 &= \mathbb{E}\operatorname{var}\Big(\|X\|^2 \mid X\cdot U\Big) + \operatorname{var}\mathbb{E}\Big(\|X\|^2 \mid X\cdot U\Big) \\
&\geq \mathbb{E}\operatorname{var}\Big(\|X\|^2 \mid X\cdot U\Big) + \beta
\end{aligned} \tag{19}$$

Call the set of points $X \in P$ with some prescribed value of $X \cdot U$ a slice. Now we will argue that the variance of a slice is at least $T\Big(n-1, f, \frac{V}{2nR}\Big)$ for most slices. Because the width of P is at most $2R$, we have that the volume of the slices S having $V_{n-1}(S) \leq V/\alpha$ is at most $2RV/\alpha$. That is, the slices having $V_{n-1}(S) > V/\alpha$ are at least a $1 - 2R/\alpha$ fraction of P. For $\alpha = 2nR$, we have that a $1 - 1/n$ fraction of P are slices with $V_{n-1}(S) > V/(2nR)$. Let $\mathcal{S}$ be the set of these slices. The number of facets of a slice is at most f, which implies that for $S \in \mathcal{S}$ we have $\operatorname{var}\Big(\|X\|^2 \mid X \in S\Big) \geq T\Big(n-1, f, \frac{V}{2nR}\Big)$. Then (19) becomes

$$\operatorname{var}\|X\|^2 \geq \left(1 - \frac{1}{n}\right) T\left(n-1, f, \frac{V}{2nR}\right) + \frac{c_4}{16}\left(\frac{V}{Rf}\right)^{4/(n-1)}.$$

Case 3: (translation) $\operatorname{var}(X \cdot U) \leq \beta$ and $\operatorname{var}\mathbb{E}\Big(\|X\|^2 \mid X \cdot U\Big) < \beta$.

Let $X_0 = X - U$. We have,

$$\operatorname{var}\|X\|^2 = \operatorname{var}\|X_0\|^2 + 4\operatorname{var} X \cdot U + 4\operatorname{cov}(X \cdot U, \|X_0\|^2). \tag{20}$$

Now, Cauchy-Schwartz inequality and the fact that $\operatorname{cov}(A, B) = \operatorname{cov}(A, \mathbb{E}(B \mid A))$ for random variables A, B, give

$$\begin{aligned}
\operatorname{cov}(X \cdot U, \|X_0\|^2) &= \operatorname{cov}(X \cdot U, \|X\|^2 - 2X \cdot U + \|U\|^2) \\
&= \operatorname{cov}(X \cdot U, \|X\|^2) - 2\operatorname{var} X \cdot U \\
&= \operatorname{cov}(X \cdot U, \mathbb{E}(\|X\|^2 \mid X \cdot U)) - 2\operatorname{var} X \cdot U \\
&\geq -\sqrt{\operatorname{var} X \cdot U}\sqrt{\operatorname{var}\mathbb{E}(\|X\|^2 \mid X \cdot U)} - 2\operatorname{var} X \cdot U.
\end{aligned}$$

This in (20) gives

$$\begin{aligned}
\operatorname{var}\|X\|^2 &\geq \operatorname{var}\|X_0\|^2 - 4\operatorname{var} X \cdot U - 4\sqrt{\operatorname{var} X \cdot U}\sqrt{\operatorname{var}\mathbb{E}\Big(\|X\|^2 \mid X \cdot U\Big)} \\
&\geq \operatorname{var}\|X_0\|^2 - 8\beta.
\end{aligned}$$

Now, X_0 is a random point in a translation of P containing the origin, and thus case 1 applies, giving

$$\operatorname{var}\|X\|^2 \geq \left(1 - \frac{c_5}{n}\right) T\left(n-1, f, \frac{V}{Rf}\right) + \frac{c_4}{2}\left(\frac{V}{Rf}\right)^{4/(n-1)}.$$

Case 4: (partition) otherwise:

We want to control $\operatorname{var} X \cdot U$ to be able to apply the third case. To this end, we will subdivide P into parts so that one of previous cases applies to each part. Let $P_1 = P$, let U_i be the nearest point to the origin in P_i (or, if P_i is empty, the sequence stops), let $\hat{U}_i$ denote $U_i/\|U_i\|$,

$$Q_i = P_i \cap \left\{x : \|U_i\| \leq \hat{U}_i \cdot x \leq \|U_i\| + \sqrt{\beta}/R\right\},$$

and $P_{i+1} = P_i \setminus Q_i$. Observe that $\|U_{i+1}\| \geq \|U_i\| + \sqrt{\beta}/R$ and $\|U_i\| \leq R$, this implies that $i \leq R^2/\sqrt{\beta}$ and the sequence is always finite.

For any i and by definition of Q_i we have $\operatorname{var}(X \cdot U_i \mid X \in Q_i) = \|U_i\|^2 \operatorname{var}(X \cdot \hat{U}_i \mid X \in Q_i) \leq \beta$.

The volume of the parts Q_i having $\operatorname{vol} Q_i \leq V/\alpha$ is at most $\frac{VR^2}{\alpha\sqrt{\beta}}$. That is, the parts having $\operatorname{vol} Q_i > V/\alpha$ are at least a $1 - \frac{R^2}{\alpha\sqrt{\beta}}$ fraction of P. For $\alpha = nR^2/\sqrt{\beta}$ we have that a $1 - 1/n$ fraction of P are parts with $\operatorname{vol}(Q_i) > V\sqrt{\beta}/(nR^2)$. Let $\mathcal{Q}$ be the set of these parts. The number of facets of a part is at most $f + 2$. Thus, applying one of the three previous cases to each part in $\mathcal{Q}$, and using that $f \geq n$,

$$\begin{aligned}
\operatorname{var} \|X\|^2 &\geq \frac{1}{\operatorname{vol} P} \sum_{Q \in \mathcal{Q}} \operatorname{vol} Q \operatorname{var}(\|X\|^2 \mid X \in Q) \\
&\geq \left(1 - \frac{1}{n}\right) \left(\left(1 - \frac{c_5}{n}\right) T\left(n-1, f+2, \frac{V\sqrt{\beta}}{nR^3 \max\{f, 2n\}}\right) + \frac{c_4}{16} \left(\frac{V\sqrt{\beta}}{nR^3 f}\right)^{4/(n-1)} \right) \\
&\geq \left(1 - \frac{1}{n}\right) \left(\left(1 - \frac{c_5}{n}\right) T\left(n-1, f+2, \frac{V\sqrt{\beta}}{2fnR^3}\right) + \frac{c_4}{16} \left(\frac{V\sqrt{\beta}}{nR^3 f}\right)^{4/(n-1)} \right).
\end{aligned}$$

In any of these cases,

$$\operatorname{var} \|X\|^2 \geq \left(1 - \frac{c_6}{n}\right) T\left(n-1, f+2, \frac{V}{2Rf} \min\Bigl(1, \frac{\sqrt{\beta}}{nR^2}\Bigr)\right) + c_7 \left(\frac{V}{Rf} \min\Bigl(1, \frac{\sqrt{\beta}}{nR^2}\Bigr)\right)^{4/(n-1)}. \tag{21}$$

Now, by assumption, $V \leq 2^n R^n$, and this implies by definition that

$$\frac{\sqrt{\beta}}{nR^2} \leq O\left(\frac{1}{n}\right).$$

That is,

$$\min\left(1, \frac{\sqrt{\beta}}{nR^2}\right) = O\left(\frac{\sqrt{\beta}}{nR^2}\right)$$

and the lemma follows, after replacing the value of β in Equation (21). □

Proof (of Theorem 6) The inequality claimed in the theorem is invariant under (uniform) scaling (which would change the volume as well as the radius of the circumscribed sphere), and thus for the proof we can assume that

$\operatorname{vol} P = 1$, without loss of generality. For $n \geq 13$, this implies that $R \geq 1$. We use the recurrence lemma in a nested way $t = n/\log n$ times[2]. The radius R stays fixed, and the number of facets involved is at most $f + 2t \leq 3f$. Each time, the volume is raised to the power of at most $1 + \frac{2}{n-t}$ and divided by at most

$$u := c'nR^2\Big(R(f+2t)\Big)^{1+\frac{2}{n-t}} > 1,$$

for $c' = \max(c_2^{-1}, 1)$. That is, after t times the volume is at least (using the fact that $(1+\frac{2}{n-t})^t = O(1)$ and denoting $v = 1 + \frac{2}{n-t}$)

$$u^{-\sum_{i=0}^{t-1} v^i} \geq u^{-tv^t} = \left(c'nR^2\Big(R(f+2t)\Big)^{1+\frac{2}{n-t}}\right)^{-t(1+\frac{2}{n-t})^t} \geq 1/(3c'nR^3f)^{O(t)}.$$

That means that from the recurrence inequality we get (we ignore the expression in "?", as we will discard that term):

$$\begin{aligned} T(n,f,1) \geq \left(1 - \frac{c_1}{n}\right)^t T(n-t, f+2t, ?) + \\ + c_3 t \left(1 - \frac{c_1}{n}\right)^{t-1} \frac{1}{R^{8/(n-t-1)}} \left(\frac{1}{3Rf} \frac{1}{(3c'nR^3f)^{O(t)}}\right)^{\frac{4}{n-1} + \frac{8}{(n-1)^2}}. \end{aligned}$$

We discard the first term and simplify to get,

$$T(n,f,1) \geq \frac{n}{\log n} \left(\frac{1}{R^3 f}\right)^{O(1/\log n)}$$

Thus, for a polytope of arbitrary volume we get by means of a scaling that there exists a universal constant $c > 0$ such that

$$\operatorname{var} \|X\|^2 \geq (\operatorname{vol} P)^{4/n} \left(\frac{(\operatorname{vol} P)^{3/n}}{R^3 f}\right)^{c/\log n} \frac{n}{\log n}.$$

The theorem follows. □

8 Discussion

The results for determinant/volume hold with the following stronger oracle: we can specify any $k \times k$ submatrix A' of A and a vector $x \in \mathbb{R}^k$ and ask whether $\|A'x\|_\infty \leq 1$. In particular, this allows us to query individual entries of the matrix. More specifically, consider the oracle that takes indices i, j and $a \in \mathbb{R}$ and returns whether $A_{ij} \leq a$. Using this oracle, our proof (Lemma 7) yields the following result: there is a constant $c > 0$ such that any randomized algorithm

[2] To force t to be an integer would only add irrelevant complications that we omit.

that approximates the determinant to within a $(1+c)$ factor has complexity $\Omega(n^2)$. In the property testing framework, this rules out sublinear (in the input size) methods for estimating the determinant, even with randomized (adaptive) access to arbitrary entries of the input matrix.

A posteriori, the way the volume lower bound is proved resembles an idea used in communication complexity: discrepancy lower bounds. In that idea, one gives an upper bound to the size of "almost monochromatic rectangles", which implies a lower bound on the number of rectangles and, thus, the communication complexity of the given function. In our case, we give an upper bound to the measure of product sets where the determinant does not change too much. Moreover, our results imply a lower bound for the following multiparty problem: There are n players, player i gets to know only the ith row of a given $n \times n$ real matrix A, and they want to approximate $|\det A|$ up to a multiplicative constant. Then in any protocol where each of them broadcasts bits, they must broadcast $\Omega(n^2/\log n)$ bits, even for randomized protocols succeeding with high probability and even if the matrix is restricted to be far from singular as in Theorem 3.

In our lower bounds for the product, the error bound is $1+c/\log n$, where the logarithmic factor comes from the variance lemma. It is an open problem as to whether this factor can be removed in the variance lower bound.

For the volume problem itself, the best known algorithm has complexity roughly $O(n^4)$ but the complexity of that algorithm is conjectured to be n^3. It is conceivable that our lower bound for membership oracle queries can be improved to n^3, although one would have to use bodies other than parallelepipeds. Also, it is an open problem to give a faster algorithm using a separation oracle.

Finally, we hope that the tools introduced here are useful for other problems.

Acknowledgements. We would like to thank Noga Alon, Laci Lovász, Mike Saks and Avi Wigderson for helpful discussions.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions.* Dover, New York, 1972. Tenth printing.

[2] D. Applegate and R. Kannan. Sampling and integration of near log-concave functions. *Proceedings of the twenty-third annual ACM symposium on theory of computing*, pages 156–163, 1991.

[3] K. Ball. Normed spaces with a weak Gordon-Lewis property. *Lecture Notes in Mathematics*, 1470:36–47, 1991.

[4] I. Bárány and Z. Füredi. Computing the volume is difficult. *Discrete and Computational Geometry*, 2:314–326, 1987.

[5] A. Björner, L. Lovász, and A. C.-C. Yao. Linear decision trees: Volume estimates and topological bounds. In *STOC*, pages 170–177. ACM, 1992.

[6] S. G. Bobkov and A. Koldobsky. On the central limit property of convex bodies. In *Geometric aspects of functional analysis*, volume 1807 of *Lecture Notes in Math.*, pages 44–52. Springer, Berlin, 2003.

[7] J. Bourgain. On the distribution of polynomials on high-dimensional convex sets. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 127–137. Springer, Berlin, 1991.

[8] D. Dobkin and R. J. Lipton. A lower bound of $\frac{1}{2}n^2$ on linear search programs for the knapsack problem. *J. Comput. System Sci.*, 16(3):413–417, 1978.

[9] M. Dyer and A. Frieze. Computing the volume of convex bodies: a case where randomness provably helps. In *Probabilistic combinatorics and its applications (San Francisco, CA, 1991)*, volume 44 of *Proc. Sympos. Appl. Math.*, pages 123–169. Amer. Math. Soc., Providence, RI, 1991.

[10] M. Dyer, A. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.*, 38(1):1–17, 1991.

[11] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[12] G. Elekes. A geometric inequality and the complexity of computing volume. *Discrete Comput. Geom.*, 1(4):289–292, 1986.

[13] O. Goldreich and L. A. Levin. A hard-core predicate for all one-way functions. In *STOC*, pages 25–32. ACM, 1989.

[14] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.*, 43(2-3):169–188, 1986.

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

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

[17] L. Lovász and M. Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990)*, pages 346–354. IEEE Comput. Soc. Press, Los Alamitos, CA, 1990.

[18] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures and Algorithms*, 4(4):359–412, 1993.

[19] L. Lovász and S. Vempala. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *Journal of Computer and System Sciences*, 72(2):392–417, 2006.

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

[21] V. Milman and A. Pajor. Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. *Lecture Notes in Mathematics*, 1376:64–104, 1989.

[22] M. Simonovits. How to compute the volume in high dimension? *Mathematical Programming*, 97(1):337–374, 2003.

[23] S. Vempala. Geometric Random Walks: A Survey. *MSRI volume on Combinatorial and Computational Geometry*, 2005.

[24] S. S. Vempala. *The Random Projection Method*. American Mathematical Society, 2004.