

Probability Bounds for Polynomial Functions in Random Variables

Simai HE ^{*} Bo JIANG [†] Zhening LI [‡] Shuzhong ZHANG [§]

October 23, 2013

Abstract

Random sampling is a simple but powerful method in statistics and in the design of randomized algorithms. In a typical application, random sampling can be applied to estimate an extreme value, say maximum, of a function f over a set $S \subseteq \mathbb{R}^n$. To do so, one may select a simpler (even finite) subset $S_0 \subseteq S$, randomly take some samples over S_0 for a number of times, and pick the best sample. The hope is to find a good approximate solution with reasonable chance. This paper sets out to present a number of scenarios for f , S and S_0 where certain probability bounds can be established, leading to a quality assurance of the procedure. In our setting, f is a multivariate polynomial function. We prove that if f is a d -th order homogeneous polynomial in n variables and F is its corresponding super-symmetric tensor, and ξ_i ($i = 1, 2, \dots, n$) are i.i.d. Bernoulli random variables taking 1 or -1 with equal probability, then $\text{Prob} \left\{ f(\xi_1, \xi_2, \dots, \xi_n) \geq \tau n^{-\frac{d}{2}} \|F\|_1 \right\} \geq \theta$, where $\tau, \theta > 0$ are two universal constants and $\|\cdot\|_1$ denotes the summation of the absolute values of all its entries. Several new inequalities concerning probabilities of the above nature are presented in this paper. Moreover, we show that the bounds are tight in most cases. Applications of our results in optimization are discussed as well.

Keywords: random sampling, probability bound, tensor form, polynomial function, polynomial optimization, approximation algorithm.

Mathematics Subject Classification: 60E15, 15A69, 90C10, 90C11, 90C59.

^{*}Department of Management Sciences, City University of Hong Kong, Hong Kong, Email: simaihe@cityu.edu.hk.

[†]Research Center for Management Science and Data Analytics, School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China. Email: isyebojiang@gmail.com.

[‡]Department of Mathematics, University of Portsmouth, Portsmouth PO3 1HF, United Kingdom. Email: zheningli@gmail.com.

[§]Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, USA. Email: zhangs@umn.edu.

1 Introduction.

Let $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function, and $S \subseteq \mathbb{R}^n$ be a given set, wherewith we consider: $\max_{x \in S} f(x)$. A possible generic approximation method for solving this problem would be randomization and sampling. In particular, we may proceed as follows: (i) choose a suitable and well-structured subset $S_0 \subseteq S$; (ii) design a suitable probability distribution ξ on S_0 ; (iii) take some random samples and pick the best solution. The quality of this approach, of course, depends on the chance of hitting some ‘good solutions’ by the random sampling. In other words, a bound in the following format is of crucial importance to us:

$$\text{Prob}_{\xi \sim S_0} \left\{ f(\xi) \geq \tau \max_{x \in S} f(x) \right\} \geq \theta, \quad (1)$$

where $\tau > 0$ and $0 < \theta < 1$ are certain constants.

In another situation, the original problem of interest is $\max_{x \in S_0} f(x)$. Replacing the constraint set to be $x \in S$ is a relaxation and it can help to create an easier problem to analyze. In this setting, a bound like (1) is useful in terms of deriving an approximate solution for solving the problem. A good example of this approach is the max-cut formulation of Goemans and Williamson [6], where S_0 is the set of rank-one positive semidefinite matrices with diagonal elements being all-ones, and S is S_0 dropping the rank-one restriction. In [19, 18, 10], this technique helped in the design of efficient randomized approximation algorithms for solving quadratically constrained quadratic programs by semidefinite programming (SDP) relaxation.

Motivated mainly due to its generic interest and importance, primarily in optimization, the current paper is devoted to the establishment of inequalities of type (1), under various assumptions. Of course such probability estimation cannot hold in general, unless some structures are in place. However, once (1) indeed holds, then with probability θ we will get a solution whose value is no worse than τ times the best possible value of $f(x)$ over S . In other words, with probability θ we will be able to generate a τ -approximate solution. In particular, if we independently draw m trials of ξ on S_0 and pick the one with the largest function value, then this process is a randomized approximation algorithm with approximation ratio τ , where the probability to this quality solution is at least $1 - (1 - \theta)^m$. If $m = \frac{\ln \frac{1}{\epsilon}}{\theta}$ then $1 - (1 - \theta)^m \geq 1 - \epsilon$, and this randomized algorithm indeed runs in polynomial-time in the problem dimensions.

In fact, the framework of our investigation, viz. the probability bound (1), is sufficiently rich to include some highly nontrivial results beyond optimization as well. As an example, let $f(x) = a^T x$ be a linear function, and $S = S_0 = \mathbb{B}^n := \{1, -1\}^n$ be a binary hypercube. Khot and Naor in [13] derived the following probability bound, which can be seen as a nontrivial instance of (1).

For every $\delta \in (0, \frac{1}{2})$, there is a constant $c_1(\delta) > 0$ with the following property: Fix $a = (a_1, a_2, \dots, a_n)^T \in \mathbb{R}^n$ and let $\xi_1, \xi_2, \dots, \xi_n$ be i.i.d. symmetric Bernoulli random variables (taking ± 1 with equal probability), then

$$\text{Prob} \left\{ \sum_{j=1}^n a_j \xi_j \geq \sqrt{\frac{\delta \ln n}{n}} \|a\|_1 \right\} \geq \frac{c_1(\delta)}{n^\delta}. \quad (2)$$

Since $\max_{x \in \mathbb{B}^n} a^T x = \|a\|_1$, (2) is of type (1), with $\tau = \sqrt{\frac{\delta \ln n}{n}}$ and $\theta = \frac{c_1(\delta)}{n^\delta}$. This bound indeed gives rise to an $\Theta\left(\sqrt{\frac{\ln n}{n}}\right)$ -approximation algorithm for the binary constrained trilinear

form maximization problem:

$$\begin{aligned} \max \quad & F(x, y, z) := \sum_{i,j,k=1}^n a_{ijk} x_i y_j z_k \\ \text{s.t.} \quad & x, y, z \in \mathbb{B}^n. \end{aligned}$$

To see why, let us denote its optimal solution to be $(x^*, y^*, z^*) = \arg \max_{x,y,z \in \mathbb{B}^n} F(x, y, z)$. By letting $a = F(\cdot, y^*, z^*) \in \mathbb{R}^n$ and $\xi_1, \xi_2, \dots, \xi_n$ be i.i.d. symmetric Bernoulli random variables, it follows from (2) that

$$\text{Prob} \left\{ F(\xi, y^*, z^*) \geq \sqrt{\frac{\delta \ln n}{n}} \|F(\cdot, y^*, z^*)\|_1 \right\} \geq \frac{c_1(\delta)}{n^\delta}. \quad (3)$$

Notice that by the optimality of (x^*, y^*, z^*) , we have $\|F(\cdot, y^*, z^*)\|_1 = F(x^*, y^*, z^*)$. Besides for any fixed ξ , the problem $\max_{y,z \in \mathbb{B}^n} F(\xi, y, z)$ is a binary constrained bilinear form maximization problem, which admits a deterministic approximation algorithm with approximation ratio 0.03 (see Alon and Naor [1]). Thus we are able to find two vectors $y_\xi, z_\xi \in \mathbb{B}^n$ in polynomial-time such that

$$F(\xi, y_\xi, z_\xi) \geq 0.03 \max_{y,z \in \mathbb{B}^n} F(\xi, y, z) \geq 0.03 F(\xi, y^*, z^*),$$

which by (3) implies

$$\text{Prob} \left\{ F(\xi, y_\xi, z_\xi) \geq 0.03 \sqrt{\frac{\delta \ln n}{n}} F(x^*, y^*, z^*) \right\} \geq \frac{c_1(\delta)}{n^\delta}.$$

Now we may independently draw $\xi_1, \xi_2, \dots, \xi_n$, followed by the algorithm proposed in [1] to solve $\max_{y,z \in \mathbb{B}^n} F(\xi, y, z)$. If we apply this procedure $\frac{n^\delta \ln \frac{1}{\epsilon}}{c_1(\delta)}$ times and pick the one with the largest objective value, then it is actually a polynomial-time randomized approximation algorithm with approximation ratio $0.03 \sqrt{\frac{\delta \ln n}{n}}$, whose chance of getting this quality bound is at least $1 - \epsilon$.

The scope of applications for results of type (1) is certainly beyond optimization per se; it is significant in the nature of probability theory itself. Recall that most classical results in probability theory is to upper bound the tail of a distribution (e.g. the Markov inequality and the Chebyshev inequality), say $\text{Prob} \{ \xi \geq a \} \leq b$. In other words, these are the *upper* bounds for the probability of a random variable beyond a threshold value. However, in some applications a *lower* bound for such probability can be relevant, in the form of

$$\text{Prob} \{ \xi \geq a \} \geq b. \quad (4)$$

One interesting example is a result due to Ben-Tal, Nemirovskii, and Roos [2], where they proved a lower bound of $1/8n^2$ for the probability that a homogeneous quadratic form of n i.i.d. symmetric Bernoulli random variables lies above its mean. More precisely, they proved the following:

If $F \in \mathbb{R}^{n \times n}$ is a symmetric matrix and $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ are i.i.d. symmetric Bernoulli random variables, then

$$\text{Prob} \{ \xi^T F \xi \geq \text{tr}(F) \} \geq \frac{1}{8n^2}.$$

As a matter of fact, the authors went on to conjecture in [2] that the lower bound can be as high as $\frac{1}{4}$, which was very recently disproved by Yuan [23]. However, the value of the tight bound remains unknown. A significant progress on this conjecture is due to He et al. [10], where the authors

improved the lower bound of $\frac{1}{8n^2}$ to 0.03. Note that the result of He et al. [10] also holds for any ξ_i 's being i.i.d. standard normal random variables. Luo and Zhang [17] provides a constant lower bound for the probability that a homogeneous quartic function of a zero mean multivariate normal distribution lies above its mean, which was a first attempt to extend such probability bound for functions of random variables beyond quadratic. For a univariate random variable, bounds of type (4) and its various extensions can be found in a recent paper by He, Zhang, and Zhang [11].

A well known result of Grünbaum [7] can also be put in the category of probability inequality (4). Grünbaum's theorem asserts:

If $S \subseteq \mathbb{R}^n$ is convex and ξ is uniformly distributed on S , then for any $c \in \mathbb{R}^n$,

$$\text{Prob} \{c^T \xi \geq c^T \mathbb{E} \xi\} \geq \frac{1}{e}.$$

The current paper aims at providing various new lower bounds for inequalities of type (1), when f is a multivariate polynomial function. To enable the presentation of our results, let us first briefly introduce the notations adopted in this paper. For any given d -th order tensor $F \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, we denote $F(x^1, x^2, \dots, x^d)$ to be the multilinear form induced by the tensor F , i.e.,

$$F(x^1, x^2, \dots, x^d) := \sum_{1 \leq i_1 \leq n_1, 1 \leq i_2 \leq n_2, \dots, 1 \leq i_d \leq n_d} F_{i_1 i_2 \dots i_d} x_{i_1}^1 x_{i_2}^2 \dots x_{i_d}^d = F \bullet (x^1 \otimes x^2 \otimes \dots \otimes x^d),$$

where $x^i \in \mathbb{R}^{n_i}$ for $i = 1, 2, \dots, d$. If $F \in \mathbb{R}^{n^d}$ is super-symmetric (the component is invariant under the permutation of the indices), we denote $f(x)$ to be the homogeneous polynomial function of $x \in \mathbb{R}^n$ induced by the super-symmetric tensor F , i.e.,

$$f(x) := F(\underbrace{x, x, \dots, x}_d) = \sum_{1 \leq i_1, i_2, \dots, i_d \leq n} F_{i_1 i_2 \dots i_d} x_{i_1} x_{i_2} \dots x_{i_d} = F \bullet (\underbrace{x \otimes x \otimes \dots \otimes x}_d).$$

For any given set $S \subseteq \mathbb{R}^n$, $\xi \sim S$ stands for that ξ is a multivariate *uniform* distribution on the support S . Two types of support sets are frequently used in this paper, namely

$$\mathbb{B}^n := \{1, -1\}^n \quad \text{and} \quad \mathbb{S}^n := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}.$$

It is easy to verify the following equivalent relationship:

1. $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T \sim \mathbb{B}^n$ is equivalent to $\xi_i \sim \mathbb{B}$ ($i = 1, 2, \dots, n$), and ξ_i 's are i.i.d. random variables;
2. $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T \sim \mathbb{S}^n$ is equivalent to $\eta / \|\eta\|_2$, with $\eta = (\eta_1, \eta_2, \dots, \eta_n)^T$ and η_i 's are i.i.d. standard normal random variables.

To simplify the presentation, the notion $\Theta(f(n))$ signifies the fact that there are positive universal constants α, β and n_0 such that $\alpha f(n) \leq \Theta(f(n)) \leq \beta f(n)$ for all $n \geq n_0$; i.e., it is of the same order as $f(n)$. To avoid confusion, the term *constant* sometimes also refers to a parameter depending only on the dimension of a polynomial function, which is a given number independent of the input data of the problem. In this paper, we use the L_1 norm (the sum of the absolute values of its entries) or L_2 norm (the square root of the sum of its squared entries) for vectors, matrices, and high order tensors.

The paper is organized as follows. In Section 2, we present probability inequalities of type (1) where f is a multilinear form, and ξ is either a random vector with i.i.d. symmetric Bernoulli random

variables, or a uniform distribution over hypersphere. Then in Section 3, we present another set of probability bounds of homogeneous polynomial function over a general class of independent random variables, including symmetric Bernoulli random variables and uniform distribution over hypersphere. We discuss some polynomial optimization problems where these probability bounds can be directly applied in Section 4. Finally, we summarize and discuss the main results presented in the paper in Section 5, with concluding remarks.

2 Multilinear tensor function in random variables.

In this section we present the following result, which provides tight probability bounds for multilinear form in two different sets of random variables.

Theorem 2.1 *Let $\xi^i \sim \mathbb{B}^{n_i}$ ($i = 1, 2, \dots, d$) be independent of each other, and $\eta^i \sim \mathbb{S}^{n_i}$ ($i = 1, 2, \dots, d$) be independent of each other. For any d -th order tensor $F \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ with $n_1 \leq n_2 \leq \dots \leq n_d$, and constant $\delta \in (0, \frac{1}{2})$, $\gamma \in (0, \frac{n_d}{\ln n_d})$, it follows that*

$$\text{Prob} \left\{ F(\xi^1, \xi^2, \dots, \xi^d) \geq c_3^{d-1} \sqrt{\frac{\delta \ln n_d}{\prod_{i=1}^d n_i}} \|F\|_1 \right\} \geq \frac{c_1(\delta) c_3^{2d-2}}{n_d^\delta \prod_{i=2}^d n_i^{i-1}}, \quad (5)$$

$$\text{Prob} \left\{ F(\eta^1, \eta^2, \dots, \eta^d) \geq \frac{1}{2^{\frac{d-1}{2}}} \sqrt{\frac{\gamma \ln n_d}{\prod_{i=1}^d n_i}} \|F\|_2 \right\} \geq \frac{c_2(\gamma)}{4^{d-1} n_d^{2\gamma} \sqrt{\ln n_d} \prod_{i=1}^{d-1} n_i}, \quad (6)$$

where $c_1(\delta)$ is a constant depended only on δ , $c_2(\gamma)$ is a constant depended only on γ , and $c_3 := \frac{8}{25\sqrt{5}} \approx 0.1431$. Moreover, the order of magnitude ‘ $\sqrt{\frac{\ln n_d}{\prod_{i=1}^d n_i}}$ ’ inside ‘Prob’ in (5) and (6) cannot be improved, if the probability bound on the right-hand-side is at least the reciprocal of a polynomial function in n_d .

We remark here that the degree d is deemed a fixed constant in our discussion. If we let $S = \mathbb{B}^{n_1 \times n_2 \times \dots \times n_d}$ and $S_0 = \{X \in \mathbb{B}^{n_1 \times n_2 \times \dots \times n_d} \mid \text{rank}(X) = 1\}$, then (5) is in the form of (1). Similarly, if we let $S = \mathbb{S}^{n_1 \times n_2 \times \dots \times n_d}$ and $S_0 = \{X \in \mathbb{S}^{n_1 \times n_2 \times \dots \times n_d} \mid \text{rank}(X) = 1\}$, then (6) is in the form of (1). For clarity, we shall prove (5) and (6) separately in the following two subsections.

Before doing this, let us first comment on the tightness of the bound $\tau_d := \Theta\left(\sqrt{\frac{\ln n_d}{\prod_{i=1}^d n_i}}\right) = \Theta\left(\sqrt{\frac{\ln \prod_{i=1}^d n_i}{\prod_{i=1}^d n_i}}\right)$, where the last equality holds because d is a fixed constant and $n_i \leq n_d$ for $i = 1, 2, \dots, d-1$. The tightness of the bounds is due to the inapproximability of computing the diameters of convex bodies, as shown below.

Lemma 2.2 (Khot and Naor [13]) *Let $K \in \mathbb{R}^n$ be a convex body with a weak optimization oracle. Then there is no randomized oracle-polynomial time algorithm that can compute the L_1 diameter of K with accuracy $\Theta\left(\sqrt{\frac{\ln n}{n}}\right)$.*

Lemma 2.3 (Brieden et al. [3, 4]) *Let $K \in \mathbb{R}^n$ be a convex body with a weak optimization oracle. Then there is no randomized oracle-polynomial time algorithm that can compute the L_2 diameter of K with accuracy $\Theta\left(\sqrt{\frac{\ln n}{n}}\right)$.*

These results in fact lead to the tightness of $\tau_1 = \Theta\left(\sqrt{\frac{\ln n_1}{n_1}}\right)$ in the case $d = 1$ (when the tensor F in (5) and (6) is a vector), for, if τ_1 could be improved, then applying the same argument as in the proof of Theorem 3.1 in [13]: drawing enough (polynomial number of) samples of $\xi \in \mathbb{B}^n$ for the L_1 case (respective $\eta \in \mathbb{S}^n$ for the L_2 case) followed by the oracle-polynomial time algorithm, would then improve the approximation bound τ_1 for the L_1 (respective L_2) diameter.

In fact, τ_1 is a tight bound not only for $\xi \sim \mathbb{B}^n$ but also for other structural distributions on the support set \mathbb{B}^n , also due to the inapproximability of computing the L_1 diameters of convex bodies (Lemma 2.2). Now, for any given degree d , if we denote $n = \prod_{i=1}^d n_i$, then (5) is essentially

$$\text{Prob} \left\{ F \bullet (\xi^1 \otimes \xi^2 \otimes \cdots \otimes \xi^d) \geq \Theta \left(\sqrt{\frac{\ln n}{n}} \|F\|_1 \right) \right\} \geq \Theta \left(\frac{1}{n_d^\alpha} \right) \quad (7)$$

for some constant α . Denote $\xi = \xi^1 \otimes \xi^2 \otimes \cdots \otimes \xi^d$, and clearly it is an implementable distribution on the support \mathbb{B}^n . Thus (7) can be regarded as in the form of (5) for $d = 1$. Due to the tightness of τ_1 , the bound $\tau_d = \Theta\left(\sqrt{\frac{\ln n_d}{\prod_{i=1}^d n_i}}\right) = \Theta\left(\sqrt{\frac{\ln n}{n}}\right)$ for general d in (5), once established, is tight too. The same argument of the structural distribution on the support set \mathbb{S}^n with $n = \prod_{i=1}^d n_i$ can be applied to prove the tightness of τ_1 in (6), using Lemma 2.3. It is interesting to note that the difference between a completely free ξ and the more restrictive $\xi = \xi^1 \otimes \xi^2 \otimes \cdots \otimes \xi^d$ lies in the fact that the latter is rank-one. Hence, the establishment of (5) and (6) actually implies that as far as the randomized solution is concerned, the rank-one restriction is immaterial.

2.1 Multilinear tensor function in Bernoulli random variables.

This subsection is dedicated to the proof of the first part of Theorem 2.1, namely (5). Let us start with some technical preparations. First, we have the following immediate probability estimation.

Lemma 2.4 *If $\xi \sim \mathbb{B}^n$, then for any vector $a \in \mathbb{R}^n$,*

$$\mathbb{E}|a^\top \xi| \geq 2c_3 \|a\|_2.$$

Proof. Denote $z = |a^\top \xi|$, and observe

$$\mathbb{E}z^2 = \mathbb{E} \left[\sum_{i=1}^n \xi_i a_i \right]^2 = \mathbb{E} \left[\sum_{i=1}^n a_i^2 + 2 \sum_{1 \leq i < j \leq n} \xi_i \xi_j a_i a_j \right] = \sum_{i=1}^n a_i^2 = \|a\|_2^2.$$

Direct computation shows that $\mathbb{E}z^4 \leq 9(\mathbb{E}z^2)^2$. By the Paley-Zygmund inequality [20], for every $\alpha \in (0, 1)$,

$$\text{Prob} \left\{ z \geq \sqrt{\alpha \mathbb{E}z^2} \right\} = \text{Prob} \left\{ z^2 \geq \alpha \mathbb{E}z^2 \right\} \geq (1 - \alpha)^2 (\mathbb{E}z^2)^2 / \mathbb{E}z^4 \geq (1 - \alpha)^2 / 9.$$

Since $z \geq 0$, we have

$$\mathbb{E}z \geq \text{Prob} \left\{ z \geq \sqrt{\alpha \mathbb{E}z^2} \right\} \sqrt{\alpha \mathbb{E}z^2} \geq \frac{(1 - \alpha)^2}{9} \sqrt{\alpha \mathbb{E}z^2} = \frac{(1 - \alpha)^2 \sqrt{\alpha}}{9} \|a\|_2.$$

By maximizing $\frac{(1 - \alpha)^2 \sqrt{\alpha}}{9}$ over $\alpha \in (0, 1)$, we have $\mathbb{E}z \geq \frac{16}{25\sqrt{5}} \|a\|_2 = 2c_3 \|a\|_2$. \square

We shall establish (5) by induction on the degree d . The first inductive step from $d = 1$ to $d = 2$ relies on the next lemma.

Lemma 2.5 *If $\xi \sim \mathbb{B}^n$, then for any matrix $A \in \mathbb{R}^{m \times n}$,*

$$\text{Prob} \left\{ \|A\xi\|_1 \geq \frac{c_3}{\sqrt{n}} \|A\|_1 \right\} \geq \frac{c_3^2}{m}.$$

Proof. Denote $a^i \in \mathbb{R}^n$ ($i = 1, 2, \dots, m$) to be the i -th row vector of the matrix A . By Lemma 2.4 we have for each $i = 1, 2, \dots, m$,

$$\mathbb{E} |\xi^T a^i| \geq 2c_3 \|a^i\|_2 \geq \frac{2c_3}{\sqrt{n}} \|a^i\|_1.$$

Summing over all $i = 1, 2, \dots, m$, we have

$$\mathbb{E} \|A\xi\|_1 = \sum_{i=1}^m \mathbb{E} |\xi^T a^i| \geq \frac{2c_3}{\sqrt{n}} \|A\|_1.$$

On the other hand,

$$(\mathbb{E} \|A\xi\|_1)^2 = \left(\sum_{i=1}^m \mathbb{E} |\xi^T a^i| \right)^2 \geq \sum_{i=1}^m (\mathbb{E} |\xi^T a^i|)^2 \geq \sum_{i=1}^m 4c_3^2 \|a^i\|_2^2 = 4c_3^2 \|A\|_2^2,$$

and

$$\mathbb{E} \|A\xi\|_1^2 = \mathbb{E} \left[\sum_{i=1}^m |\xi^T a^i| \right]^2 \leq \mathbb{E} \left[m \sum_{i=1}^m |\xi^T a^i|^2 \right] = m \sum_{i=1}^m \mathbb{E} [\xi^T a^i]^2 = m \sum_{i=1}^m \|a^i\|_2^2 = m \|A\|_2^2.$$

Thus by the Paley-Zygmund inequality we conclude that for any $\alpha \in (0, 1)$,

$$\text{Prob} \left\{ \|A\xi\|_1 \geq \frac{2\alpha c_3}{\sqrt{n}} \|A\|_1 \right\} \geq \text{Prob} \left\{ \|A\xi\|_1 \geq \alpha \mathbb{E} \|A\xi\|_1 \right\} \geq (1-\alpha)^2 \frac{(\mathbb{E} \|A\xi\|_1)^2}{\mathbb{E} \|A\xi\|_1^2} \geq (1-\alpha)^2 \frac{4c_3^2 \|A\|_2^2}{m \|A\|_2^2}.$$

Finally, letting $\alpha = \frac{1}{2}$ proves the lemma. \square

We remark that in the above inequality, the coefficient $\frac{c_3}{\sqrt{n}}$ in front of $\|A\|_1$ is independent of the number of rows (m) for matrix A . Towards proving (5) by induction for general d , for ease of referencing we state the following simple fact regarding joint conditional probability.

Proposition 2.6 *Suppose ξ and η are two random variables with support sets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ respectively. For $V' \subseteq V$, $W' \subseteq U \times V$ and $\delta > 0$, if*

$$\text{Prob}_{\xi} \{(\xi, y) \in W'\} \geq \delta \quad \forall y \in V$$

and

$$\text{Prob}_{\eta} \{\eta \in V'\} > 0,$$

then the joint conditional probability

$$\text{Prob}_{(\xi, \eta)} \left\{ (\xi, \eta) \in W' \mid \eta \in V' \right\} \geq \delta.$$

Proof. Notice that the first assumption is equivalent to

$$\text{Prob}_{(\xi, \eta)} \left\{ (\xi, \eta) \in W' \mid \eta = y \right\} \geq \delta \quad \forall y \in V. \quad (8)$$

Suppose that η has a density g in V , then

$$\begin{aligned} \text{Prob}_{(\xi, \eta)} \left\{ (\xi, \eta) \in W' \mid \eta \in V' \right\} &= \text{Prob}_{(\xi, \eta)} \left\{ (\xi, \eta) \in W', \eta \in V' \right\} / \text{Prob}_{\eta} \left\{ \eta \in V' \right\} \\ &= \int_{V'} \text{Prob}_{(\xi, \eta)} \left\{ (\xi, \eta) \in W', \eta = y \right\} g(y) dy / \text{Prob}_{\eta} \left\{ \eta \in V' \right\} \\ &\geq \int_{V'} \delta g(y) dy / \text{Prob}_{\eta} \left\{ \eta \in V' \right\} = \delta. \end{aligned}$$

The case where η is a discrete random variable can be handled similarly. \square

We are now ready to prove (5).

Proof of (5) in Theorem 2.1. *Proof.* The proof is based on induction on d . The case for $d = 1$ has been established by Khot and Naor [13]. Suppose the inequality holds for $d - 1$, by treating ξ^1 as a given parameter and taking $F(\xi^1, \cdot, \cdot, \dots, \cdot)$ as a tensor of order $d - 1$, one has

$$\text{Prob}_{(\xi^2, \xi^3, \dots, \xi^d)} \left\{ F(\xi^1, \xi^2, \dots, \xi^d) \geq c_3^{d-2} \sqrt{\frac{\delta \ln n_d}{\prod_{i=2}^d n_i}} \|F(\xi^1, \cdot, \cdot, \dots, \cdot)\|_1 \right\} \geq \frac{c_1(\delta) c_3^{2d-4}}{n_d^\delta \prod_{i=3}^d n_i^{i-2}}.$$

Define the event $E_1 = \left\{ \|F(\xi^1, \cdot, \cdot, \dots, \cdot)\|_1 \geq \frac{c_3}{\sqrt{n_1}} \|F\|_1 \right\}$. By applying Proposition 2.6 with $\xi = (\xi^2, \xi^3, \dots, \xi^d)$ and $\eta = \xi_1$, we have

$$\text{Prob}_{(\xi^1, \xi^2, \dots, \xi^d)} \left\{ F(\xi^1, \xi^2, \dots, \xi^d) \geq c_3^{d-2} \sqrt{\frac{\delta \ln n_d}{\prod_{i=2}^d n_i}} \|F(\xi^1, \cdot, \cdot, \dots, \cdot)\|_1 \mid E_1 \right\} \geq \frac{c_1(\delta) c_3^{2(d-2)}}{n_d^\delta \prod_{i=3}^d n_i^{i-2}}. \quad (9)$$

The desired probability can be lower bounded as follows:

$$\begin{aligned} &\text{Prob} \left\{ F(\xi^1, \xi^2, \dots, \xi^d) \geq c_3^{d-1} \sqrt{\frac{\delta \ln n_d}{\prod_{i=1}^d n_i}} \|F\|_1 \right\} \\ &\geq \text{Prob}_{(\xi^1, \xi^2, \dots, \xi^d)} \left\{ F(\xi^1, \xi^2, \dots, \xi^d) \geq c_3^{d-2} \sqrt{\frac{\delta \ln n_d}{\prod_{i=2}^d n_i}} \|F(\xi^1, \cdot, \cdot, \dots, \cdot)\|_1 \mid E_1 \right\} \cdot \text{Prob} \{E_1\} \\ &\geq \frac{c_1(\delta) c_3^{2d-4}}{n_d^\delta \prod_{i=3}^d n_i^{i-2}} \cdot \frac{c_3^2}{\prod_{i=2}^d n_i} = \frac{c_1(\delta) c_3^{2d-2}}{4^{d-1} n_d^\delta \prod_{i=2}^d n_i^{i-1}}, \end{aligned}$$

where the last inequality is due to (9) and Lemma 2.5. \square

2.2 Multilinear tensor function over hyperspheres.

In this subsection we shall prove the second part of Theorem 2.1, namely (6). The main construction is analogous to that of the proof for (5). First we shall establish a counterpart of inequality (2), i.e., we prove (6) for $d = 1$, which is essentially the following Lemma 2.7. Namely, if we uniformly and independently draw two vectors in \mathbb{S}^n , then there is non-trivial probability that their inner product is at least $\left(\sqrt{\frac{\gamma \ln n}{n}} \right)$ for certain positive γ .

Lemma 2.7 For every $\gamma > 0$, if $a, x \sim \mathbb{S}^n$ with $\gamma \ln n < n$ are drawn independently, then there is a constant $c_2(\gamma) > 0$, such that

$$\text{Prob} \left\{ a^\top x \geq \sqrt{\frac{\gamma \ln n}{n}} \right\} \geq \frac{c_2(\gamma)}{n^{2\gamma} \sqrt{\ln n}}.$$

Proof. By the symmetricity of \mathbb{S}^n , we may without loss of generality assume that a is a given vector in \mathbb{S}^n , e.g. $a = (1, 0, \dots, 0)^\top$. Let η_i ($i = 1, 2, \dots, n$) be i.i.d. standard normal random variables, then $x = \eta / \|\eta\|_2$ and $a^\top x = \eta_1 / \|\eta\|_2$.

First, we have for $n \geq 2$

$$\begin{aligned} \text{Prob} \left\{ \eta_1 \geq 2\sqrt{\gamma \ln n} \right\} &= \int_{2\sqrt{\gamma \ln n}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &\geq \int_{2\sqrt{\gamma \ln n}}^{4\sqrt{\gamma \ln n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &\geq \int_{2\sqrt{\gamma \ln n}}^{4\sqrt{\gamma \ln n}} \frac{1}{\sqrt{2\pi}} \frac{x}{4\sqrt{\gamma \ln n}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{32\pi\gamma \ln n}} \left(\frac{1}{n^{2\gamma}} - \frac{1}{n^{8\gamma}} \right). \end{aligned}$$

Secondly, we have

$$\text{Prob} \left\{ \|\eta\|_2 \geq 2\sqrt{n} \right\} \leq e^{-\frac{2n}{3}}. \quad (10)$$

To see why (10) holds, we may use a result on the χ^2 -distribution estimation by Laurent and Massart (Lemma 1 of [14]): For any vector $b = (b_1, b_2, \dots, b_n)^\top$ with $b_i \geq 0$ ($i = 1, 2, \dots, n$), denote $z = \sum_{i=1}^n b_i(\eta_i^2 - 1)$, then for any $t > 0$,

$$\text{Prob} \left\{ z \geq 2\|b\|_2\sqrt{t} + 2\|b\|_\infty t \right\} \leq e^{-t}.$$

Letting b to be the all-one vector and $t = \frac{2n}{3}$ leads to

$$\text{Prob} \left\{ \|\eta\|_2^2 \geq \frac{7n}{3} + \sqrt{\frac{8}{3}}n \right\} \leq e^{-\frac{2n}{3}},$$

which implies (10).

By these two inequalities, we conclude that

$$\begin{aligned} \text{Prob} \left\{ a^\top x \geq \sqrt{\frac{\gamma \ln n}{n}} \right\} &= \text{Prob} \left\{ \frac{\eta_1}{\|\eta\|_2} \geq \sqrt{\frac{\gamma \ln n}{n}} \right\} \\ &\geq \text{Prob} \left\{ \eta_1 \geq 2\sqrt{\gamma \ln n}, \|\eta\|_2 \leq 2\sqrt{n} \right\} \\ &\geq \text{Prob} \left\{ \eta_1 \geq 2\sqrt{\gamma \ln n} \right\} - \text{Prob} \left\{ \|\eta\|_2 \geq 2\sqrt{n} \right\} \\ &\geq \frac{1}{\sqrt{32\pi\gamma \ln n}} \left(\frac{1}{n^{2\gamma}} - \frac{1}{n^{8\gamma}} \right) - e^{-\frac{2n}{3}}. \end{aligned}$$

Therefore, there exists $n_0(\gamma) > 0$, depending only on γ , such that

$$\text{Prob} \left\{ a^T x \geq \sqrt{\frac{\gamma \ln n}{n}} \right\} \geq \frac{1}{\sqrt{32\gamma\pi \ln n}} \left(\frac{1}{n^{2\gamma}} - \frac{1}{n^{8\gamma}} \right) - e^{-\frac{2n}{3}} \geq \frac{1}{2n^{2\gamma} \sqrt{32\gamma\pi \ln n}} \quad \forall n \geq n_0(\gamma).$$

On the other hand, $0 < \gamma < \frac{n}{\ln n}$ implies that $\text{Prob} \left\{ a^T x \geq \sqrt{\frac{\gamma \ln n}{n}} \right\} > 0$. Therefore

$$\min_{n < n_0(\gamma), \gamma \ln n < n, n \in \mathbb{Z}} \text{Prob} \left\{ a^T x \geq \sqrt{\frac{\gamma \ln n}{n}} \right\} \cdot n^{2\gamma} \sqrt{\ln n} = t(\gamma) > 0,$$

where $t(\gamma)$ depends only on γ . Finally, letting $c_2(\gamma) = \min \left\{ t(\gamma), \frac{1}{2\sqrt{32\gamma\pi}} \right\}$ proves the lemma. \square

We remark that similar bound was proposed by Brieden et al. (Lemma 5.1 in [3], also in [4]), where the authors showed that

$$\text{Prob} \left\{ a^T x \geq \sqrt{\frac{\ln n}{n}} \right\} \geq \frac{1}{10\sqrt{\ln n}} \left(1 - \frac{\ln n}{n} \right)^{\frac{n-1}{2}},$$

for any $n \geq 2$. Lemma 2.7 gives a more flexible bound by incorporating the parameter γ , though the probability bound at $\gamma = 1$ is worse. Now, for any vector $a \in \mathbb{R}^n$, as $a/\|a\|_2 \in \mathbb{S}^n$, we have for $x \sim \mathbb{S}^n$

$$\text{Prob} \left\{ a^T x \geq \sqrt{\frac{\gamma \ln n}{n}} \|a\|_2 \right\} = \text{Prob} \left\{ \left(\frac{a}{\|a\|_2} \right)^T x \geq \sqrt{\frac{\gamma \ln n}{n}} \right\} \geq \frac{c_2(\gamma)}{n^{2\gamma} \sqrt{\ln n}}, \quad (11)$$

which implies (6) holds when $d = 1$. To proceed to the high order case, let us introduce the following intermediate result, which is analogous to Lemma 2.5 in previous subsection.

Lemma 2.8 *If $x \sim \mathbb{S}^n$, then for any matrix $A \in \mathbb{R}^{m \times n}$,*

$$\text{Prob} \left\{ \|Ax\|_2 \geq \frac{1}{\sqrt{2n}} \|A\|_2 \right\} \geq \frac{1}{4n}.$$

Proof. Let $A^T A = P^T \Lambda P$, where P is orthonormal and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ (since $A^T A$ is positive semidefinite). Denote $y = Px$. Since P is orthonormal and $x \sim \mathbb{S}^n$, we have $y \sim \mathbb{S}^n$. Notice that $\|Ax\|_2^2 = x^T A^T A x = x^T P^T \Lambda P x = y^T \Lambda y = \sum_{i=1}^n \lambda_i y_i^2$ and $\|A\|_2^2 = \text{tr}(A^T A) = \sum_{i=1}^n \lambda_i$, and the target probability is then

$$\text{Prob} \left\{ \|Ax\|_2 \geq \frac{1}{\sqrt{2n}} \|A\|_2 \right\} = \text{Prob} \left\{ \|Ax\|_2^2 \geq \frac{1}{2n} \|A\|_2^2 \right\} = \text{Prob} \left\{ \sum_{i=1}^n \lambda_i y_i^2 \geq \frac{1}{2n} \sum_{i=1}^n \lambda_i \right\},$$

where $y \sim \mathbb{S}^n$.

By the symmetricity of uniform distribution on the sphere, we have $\mathbb{E}[y_1^2] = \mathbb{E}[y_2^2] = \dots = \mathbb{E}[y_n^2]$. Combining with $\mathbb{E}[\sum_{i=1}^n y_i^2] = 1$ leads to $\mathbb{E}[y_i^2] = \frac{1}{n}$ for all $1 \leq i \leq n$. Therefore

$$\mathbb{E} \left[\sum_{i=1}^n \lambda_i y_i^2 \right] = \sum_{i=1}^n \lambda_i \mathbb{E}[y_i^2] = \frac{1}{n} \sum_{i=1}^n \lambda_i.$$

We are going to complete the proof by the Paley-Zygmund inequality. To this end, let us estimate $\mathbb{E} \left[\sum_{i=1}^n \lambda_i y_i^2 \right]^2$. Again by the symmetricity of uniform distribution on the sphere, we have $\mathbb{E}[y_i^4] = \alpha$ for all $1 \leq i \leq n$, and $\mathbb{E}[y_i^2 y_j^2] = \beta$ for any $1 \leq i < j \leq n$, where $\alpha, \beta > 0$ are constants to be determined. First

$$1 = \mathbb{E} \left[\sum_{i=1}^n y_i^2 \right]^2 \geq \mathbb{E} \left[\sum_{i=1}^n y_i^4 \right] = \alpha n \implies \alpha \leq \frac{1}{n}.$$

Next

$$0 \leq \mathbb{E}[y_1^2 - y_2^2]^2 = \mathbb{E}[y_1^4] + \mathbb{E}[y_2^4] - 2\mathbb{E}[y_1^2 y_2^2] = 2\alpha - 2\beta \implies \beta \leq \alpha \leq 1/n.$$

Noticing that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ leads to

$$\mathbb{E} \left[\sum_{i=1}^n \lambda_i y_i^2 \right]^2 = \alpha \sum_{i=1}^n \lambda_i^2 + 2\beta \sum_{1 \leq i < j \leq n} \lambda_i \lambda_j \leq \frac{1}{n} \left(\sum_{i=1}^n \lambda_i \right)^2 = n \left(\mathbb{E} \left[\sum_{i=1}^n \lambda_i y_i^2 \right] \right)^2.$$

Finally, by the Paley-Zygmund inequality, we have

$$\begin{aligned} \text{Prob} \left\{ \sum_{i=1}^n \lambda_i y_i^2 \geq \frac{1}{2n} \sum_{i=1}^n \lambda_i \right\} &= \text{Prob} \left\{ \sum_{i=1}^n \lambda_i y_i^2 \geq \frac{1}{2} \mathbb{E} \left[\sum_{i=1}^n \lambda_i y_i^2 \right] \right\} \\ &\geq \left(1 - \frac{1}{2} \right)^2 \frac{(\mathbb{E} [\sum_{i=1}^n \lambda_i y_i^2])^2}{\mathbb{E} [\sum_{i=1}^n \lambda_i y_i^2]^2} \geq \frac{1}{4n}. \end{aligned}$$

□

With the above preparations, the proof of (6) in Theorem 2.1 now follows from a similar induction argument as the proof of (5); the details are omitted here. Essentially, Lemma 2.7 helps with the basis case, and Lemma 2.8 helps to complete the inductive step.

3 Homogeneous polynomial function in random variables.

The previous section is concerned with tensor forms of independent entry vectors. One important aspect of the tensors is the connection to the polynomial functions. As is well known, a homogeneous d -th degree polynomial uniquely determines a super-symmetric tensor of d entry vectors. In this section we shall discuss the probability for polynomial function of random variables. In our discussion, the notion of *square-free* tensor plays an important role. Essentially, in the case of matrices, ‘square-free’ is equivalent to that the diagonal elements are all zero. For a general tensor $F = (a_{i_1 i_2 \dots i_d})$, ‘square-free’ means that $a_{i_1 i_2 \dots i_d} = 0$ whenever at least two indices are equal.

Theorem 3.1 *Let $F \in \mathbb{R}^{n^d}$ be a square-free super-symmetric tensor of order d , and let $f(x) = F(x, x, \dots, x)$ be a homogeneous polynomial function induced by F . If $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ are independent random variables with $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = 1$, $\mathbb{E}\xi_i^4 \leq \kappa$ for $i = 1, 2, \dots, n$, then*

$$\text{Prob} \left\{ f(\xi) \geq \sqrt{\frac{d!}{16\kappa}} \|F\|_2 \right\} \geq \frac{2\sqrt{3} - 3}{9d^2 (d!)^2 36^d \kappa^d}, \quad (12)$$

$$\text{Prob} \left\{ f(\xi) \geq \sqrt{\frac{d!}{16\kappa n^d}} \|F\|_1 \right\} \geq \frac{2\sqrt{3} - 3}{9d^2 (d!)^2 36^d \kappa^d}. \quad (13)$$

Compared to Theorem 2.1 in the previous section, Theorem 3.1 only requires the random variables to be independent from each other, and each with a bounded kurtosis, including the Bernoulli random variables and normal random variables as special cases. It is easy to verify that, under the square-free property of F , together with the assumptions $\mathbb{E}\xi_i = 0$ and ξ_i 's are independent from each other ($i = 1, 2, \dots, n$), we then have $\mathbb{E}[f(\xi)] = 0$. Since $\mathbb{E}\xi_i^2 = 1$ ($i = 1, 2, \dots, n$), we compute that $\text{Var}(f(\xi)) = \Theta(\|F\|_2^2)$. This means that the standard deviation of $f(\xi)$ is in the same order of $\|F\|_2$. Assertion (12) essentially states that given any set of independent random variables with bounded kurtosis, any square-free polynomial of these random variables will have a certain thickness of the tail at some point.

The proof for Theorem 3.1 is technically involved, and we shall delegate the details to the appendix. Although our main results in Theorem 3.1 are valid for arbitrary random variables, it is interesting to discuss its implications when the random variables are uniform distributions on \mathbb{B}^n and \mathbb{S}^n . In case of quadratic polynomial of Bernoulli random variables, we have the following:

Proposition 3.2 *If F is a diagonal-free symmetric matrix and $\xi \sim \mathbb{B}^n$, then*

$$\text{Prob} \left\{ \xi^T F \xi \geq \frac{\|F\|_2}{2\sqrt{30}} \right\} \geq \frac{2\sqrt{3} - 3}{135}.$$

The proof of this proposition will be discussed in appendix too. We remark that Proposition 3.2 is an extension to the result of Ben-Tal, Nemirovskii, and Roos [2] where it was shown that $\text{Prob} \{x^T F x \geq 0\} \geq \frac{1}{8n^2}$, and the result of He et al. [10] where it was shown that $\text{Prob} \{x^T F x \geq 0\} \geq 0.03$. Essentially, Proposition 3.2 is on the probability of a *strict tail* rather than the probability above the mean.

Proposition 3.3 *Let $F \in \mathbb{R}^{n^d}$ be a square-free super-symmetric tensor of order d , and let $f(x) = F(x, x, \dots, x)$ be a homogeneous polynomial function induced by F . If $\xi \sim \mathbb{B}^n$, then*

$$\text{Prob} \left\{ f(\xi) \geq \sqrt{\frac{d!}{16n^d}} \|F\|_1 \right\} \geq \frac{2\sqrt{3} - 3}{9d^2(d!)^2 36^d}.$$

Moreover, the order of magnitude $n^{-\frac{d}{2}}$ inside ‘Prob’ cannot be improved for $d = 2, 4$.

As a remark, Proposition 3.3 can be seen as an instance of (1) in the case $f(X) = F \bullet X$, $S = \{X \in \mathbb{B}^{n^d} : X \text{ is super-symmetric}\}$ and $S_0 = \{X \in S : \text{rank}(X) = 1\}$. The probability bound in Proposition 3.3 directly follows from (13), since $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = \mathbb{E}\xi_i^4 = 1$ for all $i = 1, 2, \dots, n$. It remains to show that even in this special case, the bounds are tight when $d = 2$ and $d = 4$, which are illustrated by the following examples.

Example 3.4 *For the case $d = 2$, define $F = I - E$, where I is the identity and E is the all-one matrix. In this case, for any $x \in \mathbb{B}^n$, $x^T F x = n - (e^T x)^2 \leq n$ and $\|F\|_1 = n^2 - n$. Therefore $x^T F x / \|F\|_1 \leq 1/(n-1)$ for any $x \in \mathbb{B}^n$, implying that the ratio cannot be better than $\Theta(n^{-1})$ for any positive probability.*

Example 3.5 For the case $d = 4$, define F to be the square-free tensor of order 4, with all non-square-free components being -1 . It is obvious that $\|F\|_1 = \Theta(n^4)$, and for any $x \in \mathbb{B}^n$

$$\begin{aligned}
F(x, x, x, x) &= \sum_{i=1}^n x_i^4 + 12 \sum_{i \neq j, j \neq k, i \neq k} x_i^2 x_j x_k + 6 \sum_{i \neq j} x_i^2 x_j^2 + 4 \sum_{i \neq j} x_i^3 x_j - \left(\sum_{i=1}^n x_i \right)^4 \\
&= n + 12(n-2) \sum_{j \neq k} x_j x_k + 3n(n-1) + 4 \sum_{i \neq j} x_i x_j - \left(\sum_{i=1}^n x_i \right)^4 \\
&= 3n^2 - 2n + (6n-10) \sum_{j \neq k} 2x_j x_k - \left(\sum_{i=1}^n x_i \right)^4 \\
&= 3n^2 - 2n + (6n-10) \left(\left(\sum_{i=1}^n x_i \right)^2 - n \right) - \left(\sum_{i=1}^n x_i \right)^4 \\
&= 3n^2 - 2n - n(6n-10) + (3n-5)^2 - \left(\left(\sum_{i=1}^n x_i \right)^2 - (3n-5) \right)^2 \\
&\leq 6n^2 - 22n + 25.
\end{aligned}$$

Thus we have $F(x, x, x, x)/\|F\|_1 \leq \Theta(n^{-2})$, implying that the ratio cannot be better than $\Theta(n^{-2})$ for any positive probability.

We believe that examples of the above type exist for any given $d \geq 4$; however, so far we are unable to explicitly construct a general example.

Let us now specialize the random variables to be uniformly distributed on the hypersphere. Since the components of the unit vector are not independent, we cannot directly apply Theorem 3.1. However, similar results can still be obtained.

Proposition 3.6 Let $F \in \mathbb{R}^{n^d}$ be a square-free super-symmetric tensor of order d , and let $f(x) = F(x, x, \dots, x)$ be a homogeneous polynomial function induced by F . If $\eta \sim \mathbb{S}^n$, then

$$\text{Prob} \left\{ f(\eta) \geq \sqrt{\frac{d!}{48(4n)^d}} \|F\|_2 \right\} \geq \frac{2\sqrt{3}-3}{9d^2(d!)^2 108^d} - e^{-\frac{2n}{3}}.$$

Proof. Let $\eta = \xi/\|\xi\|_2$ with $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ being i.i.d. standard normal random variables. Since $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = 1$, $\mathbb{E}\xi_i^4 = 3$ for all $1 \leq i \leq n$, by applying (12) in Theorem 3.1 with $\kappa = 3$, we have

$$\text{Prob} \left\{ f(\xi) \geq \sqrt{\frac{d!}{48}} \|F\|_2 \right\} \geq \frac{2\sqrt{3}-3}{9d^2(d!)^2 108^d}.$$

Together with (10), we have

$$\begin{aligned}
\text{Prob} \left\{ f(\eta) \geq \sqrt{\frac{d!}{48(4n)^d}} \|F\|_2 \right\} &= \text{Prob} \left\{ f \left(\frac{\xi}{\|\xi\|_2} \right) \geq \sqrt{\frac{d!}{48(4n)^d}} \|F\|_2 \right\} \\
&\geq \text{Prob} \left\{ f(\xi) \geq \sqrt{d!/48} \|F\|_2, \|\xi\|_2 \leq 2\sqrt{n} \right\} \\
&\geq \text{Prob} \left\{ f(\xi) \geq \sqrt{d!/48} \|F\|_2 \right\} - \text{Prob} \left\{ \|\xi\|_2 \geq 2\sqrt{n} \right\} \\
&\geq \frac{2\sqrt{3} - 3}{9d^2(d!)^2 108^d} - e^{-\frac{2n}{3}}.
\end{aligned}$$

□

Before concluding this section, we remark that Proposition 3.6 can still be categorized to the type of (1) with $f(X) = F \bullet X$, $S = \{X \in \mathbb{S}^{n^d} : X \text{ is super-symmetric}\}$ and $S_0 = \{X \in S : \text{rank}(X) = 1\}$. Luo and Zhang [17] offered a constant lower bound for the probability that a homogeneous quartic form of a zero mean multivariate normal distribution lies above its mean. In particular, by restricting the distributions to be i.i.d. standard normals and quartic form to be square-free, applying Theorem 3.1 in the case of $d = 4$, we obtain a constant bound for the probability that the quartic form above the mean *plus* some constant times the standard deviation. We may view this as a strengthening of the result in [17].

4 Applications of polynomial function optimization.

As discussed in the introduction, the probability bounds in the form of (1) have immediate applications in optimization. In particular, in this section we shall apply the bounds derived in Sections 2 and 3 to polynomial function optimization problems. We shall derive polynomial-time randomized approximation algorithms, with the approximation ratios improving the existing ones in the literature.

4.1 Polynomial optimization in binary variables.

The general unconstrained binary polynomial optimization model is $\max_{x \in \mathbb{B}^n} p(x)$, where $p(x)$ is a multivariate polynomial function. He, Li, and Zhang [9] proposed a polynomial-time randomized approximation algorithm with a relative performance ratio. When the polynomial $p(x)$ is homogeneous, this problem has many applications in graph theory; e.g. the max-cut problem [6] and the matrix cut-norm problem [1]. In particular we shall discuss two models in this subsection:

$$\begin{aligned}
(B_1) \quad &\max \quad F(x^1, x^2, \dots, x^d) \\
&\text{s.t.} \quad x^i \in \mathbb{B}^{n_i}, i = 1, 2, \dots, d; \\
(B_2) \quad &\max \quad f(x) = F(\underbrace{x, x, \dots, x}_d) \\
&\text{s.t.} \quad x \in \mathbb{B}^n.
\end{aligned}$$

When $d = 2$, (B_1) is to compute the matrix $\infty \mapsto 1$ norm, which is related to so called matrix cut-norm problem. The current best approximation ratio is 0.56, due to Alon and Naor [1]. When $d = 3$, (B_1) is a slight generalization of the model considered by Khot and Naor [13], where F was assumed to be super-symmetric and square-free. The approximation ratio estimated in [13] is $\Theta\left(\sqrt{\frac{\ln n_1}{n_1}}\right)$, which is the best bound till now. Recently, He, Li, and Zhang [9] proposed a

polynomial-time randomized approximation algorithm for (B_1) for any fixed degree d , with approximation performance ratio $\Theta\left(\prod_{i=1}^{d-2} \sqrt{\frac{1}{n_i}}\right)$. The results in this subsection will improve this approximation ratio for fixed d , thanks to Theorem 2.1.

Algorithm B_1 (Randomized Algorithm for (B_1))

1. Sort and rename the dimensions if necessary, so as to satisfy $n_1 \leq n_2 \leq \dots \leq n_d$.
2. Randomly and independently generate $\xi^i \sim \mathbb{B}^{n_i}$ for $i = 1, 2, \dots, d-2$.
3. Solve the following bilinear form optimization problem

$$\begin{aligned} \max \quad & F(\xi^1, \xi^2, \dots, \xi^{d-2}, x^{d-1}, x^d) \\ \text{s.t.} \quad & x^{d-1} \in \mathbb{B}^{n_{d-1}}, x^d \in \mathbb{B}^{n_d} \end{aligned}$$

using the deterministic algorithm of Alon and Naor [1], and get its approximate solution (ξ^{d-1}, ξ^d) .

4. Compute the objective value $F(\xi^1, \xi^2, \dots, \xi^d)$.
5. Repeat the above procedures $\frac{\prod_{i=1}^{d-2} n_i^\delta}{0.03 (c_1(\delta))^{d-2}} \ln \frac{1}{\epsilon}$ times for any constant $\delta \in (0, \frac{1}{2})$ and choose a solution whose objective function is the largest.

We remark that Algorithm B_1 was already mentioned in [13] for odd d , where a similar order of approximation bound as in Theorem 4.1 was suggested; however, no explicit polynomial-time algorithm and the detailed proofs of approximation guarantee were provided. The approximation ratio for Algorithm B_1 and its proof are in the following theorem.

Theorem 4.1 *Algorithm B_1 solves (B_1) in polynomial-time with probability at least $1 - \epsilon$, and its approximation performance ratio is $\delta^{\frac{d-2}{2}} \prod_{i=1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}}$.*

The proof is based on mathematical induction. Essentially, if an algorithm solves (B_1) of order $d-1$ approximately with an approximation ratio τ , then there is an algorithm solves (B_1) of order d approximately with an approximation ratio $\tau \sqrt{\frac{\delta \ln n}{n}}$, where n is the dimension of the additional order. *Proof.* For given problem degree d , the proof is based on induction on $t = 2, 3, \dots, d$. Suppose $(\xi^1, \xi^2, \dots, \xi^d)$ is the approximate solution generated by Algorithm B_1 . For $t = 2, 3, \dots, d$, we treat $(\xi^1, \xi^2, \dots, \xi^{d-t})$ as given parameters and define the following problems

$$\begin{aligned} (D_t) \quad \max \quad & F(\xi^1, \xi^2, \dots, \xi^{d-t}, x^{d-t+1}, x^{d-t+2}, \dots, x^d) \\ \text{s.t.} \quad & x^i \in \mathbb{B}^{n_i}, i = d-t+1, d-t+2, \dots, d, \end{aligned}$$

whose optimal value is denoted by $v(D_t)$. By applying Algorithm B_1 to (D_t) , $(\xi^{d-t+1}, \xi^{d-t}, \dots, \xi^d)$ can be viewed as an approximate solution generated. In the remaining, we shall prove by induction that for each $t = 2, 3, \dots, d$,

$$\text{Prob}_{(\xi^{d-t+1}, \xi^{d-t+2}, \dots, \xi^d)} \left\{ F(\xi^1, \xi^2, \dots, \xi^d) \geq \delta^{\frac{t-2}{2}} \prod_{i=d-t+1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}} v(D_t) \right\} \geq \frac{0.03 (c_1(\delta))^{t-2}}{\prod_{i=d-t+1}^{d-2} n_i^\delta}. \quad (14)$$

In other words, $(\xi^{d-t+1}, \xi^{d-t+2}, \dots, \xi^d)$ has a non-trivial probability to be a $\delta^{\frac{t-2}{2}} \prod_{i=d-t+1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}}$ -approximate solution of (D_t) .

For the initial case $t = 2$, the deterministic algorithm by Alon and Naor [1] (Step 3 of Algorithm B_1) guarantees a constant ratio, i.e., $F(\xi^1, \xi^2, \dots, \xi^d) \geq 0.03 v(D_2)$, implying (14). Suppose now (14) holds for $t - 1$. To prove that (14) holds for t , we notice that $(\xi^1, \xi^2, \dots, \xi^{d-t})$ are given fixed parameters. Denote $(z^{d-t+1}, z^{d-t+2}, \dots, z^d)$ to be an optimal solution of (D_t) , and define the following events

$$\begin{aligned} E_3 &= \left\{ z \in \mathbb{B}^{n_{d-t+1}} \mid F(\xi^1, \dots, \xi^{d-t}, z, z^{d-t+2}, \dots, z^d) \geq \sqrt{\frac{\delta \ln n_{d-t+1}}{n_{d-t+1}}} v(D_t) \right\}; \\ E_4 &= \left\{ \xi^{d-t+1} \in E_3, \xi^{d-t+2} \in \mathbb{B}^{n_{d-t+2}}, \dots, \xi^d \in \mathbb{B}^{n_d} \mid \right. \\ &\quad \left. F(\xi^1, \dots, \xi^d) \geq \delta^{\frac{t-3}{2}} \prod_{i=d-t+2}^{d-2} \sqrt{\frac{\ln n_i}{n_i}} F(\xi^1, \dots, \xi^{d-t}, \xi^{d-t+1}, z^{d-t+2}, \dots, z^d) \right\}. \end{aligned}$$

Then we have

$$\begin{aligned} &\text{Prob}_{(\xi^{d-t+1}, \dots, \xi^d)} \left\{ F(\xi^1, \dots, \xi^d) \geq \delta^{\frac{t-2}{2}} \prod_{i=d-t+1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}} v(D_t) \right\} \\ &\geq \text{Prob}_{(\xi^{d-t+1}, \dots, \xi^d)} \left\{ (\xi^{d-t+1}, \dots, \xi^d) \in E_4 \mid \xi^{d-t+1} \in E_3 \right\} \cdot \text{Prob}_{\xi^{d-t+1}} \left\{ \xi^{d-t+1} \in E_3 \right\}. \end{aligned} \quad (15)$$

To lower bound (15), first note that (z^{d-t+2}, \dots, z^d) is a feasible solution of (D_{t-1}) , and so we have

$$\begin{aligned} &\text{Prob}_{(\xi^{d-t+1}, \dots, \xi^d)} \left\{ (\xi^{d-t+1}, \dots, \xi^d) \in E_4 \mid \xi^{d-t+1} \in E_3 \right\} \\ &\geq \text{Prob}_{(\xi^{d-t+1}, \dots, \xi^d)} \left\{ F(\xi^1, \dots, \xi^d) \geq \delta^{\frac{t-3}{2}} \prod_{i=d-t+2}^{d-2} \sqrt{\frac{\ln n_i}{n_i}} v(D_{t-1}) \mid \xi^{d-t+1} \in E_3 \right\} \\ &\geq \frac{0.03 (c_1(\delta))^{t-3}}{\prod_{i=d-t+2}^{d-2} n_i^\delta}, \end{aligned}$$

where the last inequality is due to the induction assumption on $t - 1$, and Proposition 2.6 for joint conditional probability with $\xi = (\xi^{d-t+2}, \dots, \xi^d)$ and $\eta = \xi^{d-t+1}$. Secondly, we have

$$\begin{aligned} &\text{Prob}_{\xi^{d-t+1}} \left\{ \xi^{d-t+1} \in E_3 \right\} \\ &= \text{Prob}_{\xi^{d-t+1}} \left\{ F(\xi^1, \dots, \xi^{d-t+1}, z^{d-t+2}, \dots, z^d) \geq \sqrt{\frac{\delta \ln n_{d-t+1}}{n_{d-t+1}}} F(\xi^1, \dots, \xi^{d-t}, z^{d-t+1}, \dots, z^d) \right\} \\ &= \text{Prob}_{\xi^{d-t+1}} \left\{ F(\xi^1, \dots, \xi^{d-t+1}, z^{d-t+2}, \dots, z^d) \geq \sqrt{\frac{\delta \ln n_{d-t+1}}{n_{d-t+1}}} \|F(\xi^1, \dots, \xi^{d-t}, \cdot, z^{d-t+2}, \dots, z^d)\|_1 \right\} \\ &\geq \frac{c_1(\delta)}{n_{d-t+1}^\delta}, \end{aligned}$$

where the last inequality is due to Theorem 2.1 for the case $d = 1$. With the above two facts, we can lower bound the right hand side of (15), and conclude

$$\text{Prob}_{(\xi^{d-t+1}, \dots, \xi^d)} \left\{ F(\xi^1, \dots, \xi^d) \geq \delta^{\frac{t-2}{2}} \prod_{i=d-t+1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}} v(D_t) \right\} \geq \frac{0.03 (c_1(\delta))^{t-3}}{\prod_{i=d-t+2}^{d-2} n_i^\delta} \cdot \frac{c_1(\delta)}{n_{d-t+1}^\delta} = \frac{0.03 (c_1(\delta))^{t-2}}{\prod_{i=d-t+1}^{d-2} n_i^\delta}.$$

As (D_d) is exactly (B_1) , Algorithm B_1 solves (B_1) approximately with probability at least $1 - \epsilon$. \square

We remark that theoretically we may get a better approximate solution, using the 0.56-randomized algorithm in [1] to replace the subroutine in Step 3 of Algorithm B_1 , though that algorithm is quite complicated. In a similar vein, we obtain approximation results for (B_2) .

Algorithm B_2 (Randomized Algorithm for (B_2))

1. Randomly generate $\xi \sim \mathbb{B}^n$ and compute $f(\xi)$.
2. Repeat this procedure $\frac{9d^2(d!)^2 36^d}{2\sqrt{3}-3} \ln \frac{1}{\epsilon}$ times and choose a solution whose objective function is the largest.

The model (B_2) has been studied extensively in the quadratic cases, i.e., $d = 2$. Goemans and Williamson [6] gave a 0.878-approximation ratio for the case F being the Laplacian of a given graph. Later, Nesterov [19] gave a 0.63-approximation ratio for the case F being positive semidefinite. For diagonal-free matrix, the best possible approximation ratio is $\Theta(1/\ln n)$, due to Charikar and Wirth [5], which is also known to be tight. For $d = 3$ and F is square-free, Khot and Naor [13] gave an $\Theta\left(\sqrt{\frac{\ln n}{n}}\right)$ -approximation bound. They also pointed out an iterative procedure to get an $\Theta\left(\frac{\ln^{d/2-1} n}{n^{d/2-1}}\right)$ -approximation bound for odd d , which requires a linkage between multilinear tensor function and homogeneous polynomial of any degree (see Lemma 1 of [8]). For general d , He, Li, and Zhang [9] proposed polynomial-time randomized approximation algorithms with approximation ratio $\Theta\left(n^{-\frac{d-2}{2}}\right)$ when F is square-free for odd d ; however for even d , they can only propose a relative approximation ratio $\Theta\left(n^{-\frac{d-2}{2}}\right)$. Now, by virtue of Theorem 3.1 (more precisely Proposition 3.3), since $\|F\|_1$ is an upper bound for the optimal value of (B_2) , *absolute* approximation ratios are also established when d is even, as shown below.

Theorem 4.2 *When d is even and F is square-free, Algorithm B_2 solves (B_2) in polynomial-time (in terms of $\ln \frac{1}{\epsilon}$) with probability at least $1 - \epsilon$, and approximation performance ratio $\sqrt{\frac{d!}{16n^d}}$.*

4.2 Polynomial optimization over hyperspheres.

Polynomial function optimization over hyperspheres have much applications in biomedical engineering, material sciences, numerical linear algebra, among many others. Readers are referred to [8, 15, 16] and references therein for more information. Let us consider:

$$\begin{aligned}
 (S_1) \quad & \max F(x^1, x^2, \dots, x^d) \\
 & \text{s.t. } x^i \in \mathbb{S}^{n_i}, i = 1, 2, \dots, d; \\
 (S_2) \quad & \max f(x) = F(\underbrace{x, x, \dots, x}_d) \\
 & \text{s.t. } x \in \mathbb{S}^n.
 \end{aligned}$$

When $d = 2$, (S_1) and (S_2) reduce to computing matrix spectrum norms and can be solved in polynomial-time. However they are NP-hard when $d \geq 3$. For general d , (S_2) is to compute the largest eigenvalue of the tensor F . As far as approximation algorithms are concerned, He, Li, and Zhang [8] proposed polynomial-time approximation algorithms for (S_1) with approximation ratio $\Theta\left(\prod_{i=1}^{d-2} \sqrt{\frac{1}{n_i}}\right)$. In [8], a generic linkage relating (S_2) and (S_1) is established. This linkage

enables one to get a solution with the same approximation ratio (relative ratio for even d though) for (S_2) whenever a solution with an approximation ratio for (S_1) is available. Therefore, let us now focus on (S_1) . For (S_1) , recently So [21] improved the result of [8] from $\Theta\left(\prod_{i=1}^{d-2}\sqrt{\frac{1}{n_i}}\right)$ to $\Theta\left(\prod_{i=1}^{d-2}\sqrt{\frac{\ln n_i}{n_i}}\right)$. Unfortunately, the method in [21] relies on the equivalence (polynomial-time reduction) between convex optimization and membership oracle queries using the ellipsoid method, and it is computationally impractical. On the other hand, the algorithm that we present below is straightforward, while retaining the same quality of approximation as the result in [21].

Algorithm S_1 (Randomized Algorithm for (S_1))

1. Sort and rename the dimensions if necessary, so as to satisfy $n_1 \leq n_2 \leq \dots \leq n_d$.
2. Randomly and independently generate $\eta^i \sim \mathbb{S}^{n_i}$ for $i = 1, 2, \dots, d-2$.
3. Solve the largest singular value problem

$$\begin{aligned} \max \quad & F(\eta^1, \eta^2, \dots, \eta^{d-2}, x^{d-1}, x^d) \\ \text{s.t.} \quad & x^{d-1} \in \mathbb{S}^{n_{d-1}}, x^d \in \mathbb{S}^{n_d}, \end{aligned}$$

and get its optimal solution (η^{d-1}, η^d) .

4. Compute the objective value $F(\eta^1, \eta^2, \dots, \eta^d)$.
5. Repeat the above procedures $\frac{\prod_{i=1}^{d-2} n_i^{2\gamma} \sqrt{\ln n_i}}{(c_2(\gamma))^{d-2}} \ln \frac{1}{\epsilon}$ times for any constant $\gamma \in (0, \frac{n_1}{\ln n_1})$ and choose a solution whose objective function is the largest.

Theorem 4.3 *Algorithm S_1 solves (S_1) in polynomial-time with probability at least $1 - \epsilon$, and approximation performance ratio $\gamma^{\frac{d-2}{2}} \prod_{i=1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}}$.*

The proof is similar to that for Theorem 4.1, and is omitted here.

4.3 Polynomial function mixed integer programming.

This last subsection deals with optimization of polynomial functions under binary variables and variables with spherical constraints mixed up. Such problems have applications in matrix combinatorial problem, vector-valued maximum cut problem; see e.g. [9]. In [9], the authors considered

$$\begin{aligned} (M_1) \quad & \max \quad F(x^1, x^2, \dots, x^d, y^1, y^2, \dots, y^{d'}) \\ & \text{s.t.} \quad x^i \in \mathbb{B}^{n_i}, i = 1, 2, \dots, d; y^j \in \mathbb{S}^{m_j}, j = 1, 2, \dots, d'; \\ (M_2) \quad & \max \quad F(\underbrace{x, x, \dots, x}_d, \underbrace{y, y, \dots, y}_{d'}) \\ & \text{s.t.} \quad x \in \mathbb{B}^n, y \in \mathbb{S}^m; \\ (M_3) \quad & \max \quad F(\underbrace{x^1, x^1, \dots, x^1}_{d_1}, \dots, \underbrace{x^s, x^s, \dots, x^s}_{d_s}, \underbrace{y^1, y^1, \dots, y^1}_{d'_1}, \dots, \underbrace{y^t, y^t, \dots, y^t}_{d'_t}) \\ & \text{s.t.} \quad x^i \in \mathbb{B}^{n_i}, i = 1, 2, \dots, s; y^j \in \mathbb{S}^{m_j}, j = 1, 2, \dots, t; \end{aligned}$$

and proposed polynomial-time randomized approximation algorithms when the tensor F is square-free in x (the binary part). In fact, (M_3) is a generalization of (M_1) and (M_2) , and it can also be regarded as a generalization of (B_1) , (B_2) , (S_1) and (S_2) as well. Essentially the approximative

results can be applied by using the linkage we mentioned earlier (see [8]) if approximation result for (M_1) can be established. In fact, (M_1) plays the role as a cornerstone for the whole construction. The approximation ratio for (M_1) derived in [9] is $\Theta\left(\prod_{i=1}^{d-1} \sqrt{\frac{1}{n_i}} \prod_{j=1}^{d'-1} \sqrt{\frac{1}{m_j}}\right)$. The results in Section 2 lead to the following improvements:

Theorem 4.4 *Denote N to be the set of the $d+d'-2$ smallest numbers in $\{n_1, \dots, n_d, m_1, \dots, m_{d'}\}$. (M_1) admits a polynomial-time randomized approximation algorithm with approximation performance ratio $\Theta\left(\prod_{n \in N} \sqrt{\frac{\ln n}{n}}\right)$.*

The method for solving (M_1) is similar to that for solving (B_1) and (S_1) , and we shall not repeat the detailed discussions. Basically we shall sample multiple times to get a solution with high probability. For the $d+d'-2$ numbers in N , if it is the dimension of binary constraints, the algorithm uniformly picks a vector in the discrete hypercube; and if it is the dimension of spherical constraints, the algorithms uniformly pick a vector in the hypersphere. All the randomized procedures will be done independent from each other. As the first inductive step, we will then come across a bilinear function optimization problem in either of the three possible cases:

- $\max_{x \in \mathbb{B}^n, y \in \mathbb{B}^m} x^T F y$, which can be solved by the algorithm proposed in Alon and Naor [1] to get a solution with the guaranteed constant approximation ratio;
- $\max_{x \in \mathbb{B}^n, y \in \mathbb{S}^m} x^T F y$, which can be solved by the algorithm proposed in He, Li, and Zhang [9] to get a solution with the guaranteed constant approximation ratio;
- $\max_{x \in \mathbb{S}^n, y \in \mathbb{S}^m} x^T F y$, which can be solved by computing the largest singular value of matrix F .

5 Summary and concluding remark.

To put the results presented in the paper in perspective, in this section let us highlight and briefly summarize our new findings.

We set out to explore the probability bound in the form of $\text{Prob}_{\xi \sim S_0} \{f(\xi) \geq \tau \max_{x \in S} f(x)\} \geq \theta$ with $S_0 \subseteq S$, denoted by (1) in this paper. The function F in our discussion is either a multilinear tensor form or a homogeneous polynomial function. To enable probability bounds in the form of (1), we will need some structure in place. In particular, we consider the choice of the structural sets S_0 and S respectively as follows:

1. Consider $S = \mathbb{B}^{n_1 \times n_2 \times \dots \times n_d}$ and $S_0 = \{X \in S \mid \text{rank}(X) = 1\}$, and $F \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$. If we draw ξ uniformly over S_0 , then

$$\text{Prob} \left\{ F \bullet \xi \geq c_3^{d-1} \sqrt{\frac{\delta \ln n_d}{\prod_{i=1}^d n_i}} \max_{X \in S} F \bullet X = c_3^{d-1} \sqrt{\frac{\delta \ln n_d}{\prod_{i=1}^d n_i}} \|F\|_1 \right\} \geq \frac{c_1(\delta) c_3^{2d-2}}{n_d^\delta \prod_{i=2}^d n_i^{i-1}},$$

where $c_1(\delta)$ is a constant depending only on constant $\delta \in (0, \frac{1}{2})$ and $c_3 = \frac{8}{25\sqrt{5}}$. Moreover, the order of $\sqrt{\frac{\ln n_d}{\prod_{i=1}^d n_i}}$ cannot be improved if the bound is required to be at least a polynomial function of $\frac{1}{n_d}$.

2. Consider $S = \{X \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} \mid X \bullet X = 1\}$ and $S_0 = \{X \in S \mid \text{rank}(X) = 1\}$, and $F \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$. If we draw ξ uniformly over S_0 , then

$$\text{Prob} \left\{ F \bullet \xi \geq \frac{1}{2^{\frac{d-1}{2}}} \sqrt{\frac{\gamma \ln n_d}{\prod_{i=1}^d n_i}} \max_{X \in S} F \bullet X = \frac{1}{2^{\frac{d-1}{2}}} \sqrt{\frac{\gamma \ln n_d}{\prod_{i=1}^d n_i}} \|F\|_2 \right\} \geq \frac{c_2(\gamma)}{4^{d-1} n_d^{2\gamma} \sqrt{\ln n_d} \prod_{i=1}^{d-1} n_i},$$

where $c_2(\gamma)$ is a constant depended only on constant $\gamma \in (0, \frac{n_d}{\ln n_d})$. Moreover, the order of $\sqrt{\frac{\ln n_d}{\prod_{i=1}^d n_i}}$ cannot be improved if the bound is required to be at least a polynomial function of $\frac{1}{n_d}$.

3. Consider $S = \{X \in \mathbb{B}^{n^d} \mid X \text{ is super-symmetric}\}$ and $S_0 = \{X \in S \mid \text{rank}(X) = 1\}$, and a square-free super-symmetric tensor $F \in \mathbb{R}^{n^d}$. If we draw ξ uniformly over S_0 , then there exists a universal constant $c > 0$, such that

$$\text{Prob} \left\{ F \bullet \xi \geq \sqrt{\frac{d!}{16n^d}} \max_{X \in S} F \bullet X = \sqrt{\frac{d!}{16n^d}} \|F\|_1 \right\} \geq c.$$

Moreover, when $d = 2$ or $d = 4$, the order of $n^{-\frac{d}{2}}$ cannot be improved for any positive bound.

4. Consider $S = \{X \in \mathbb{R}^{n^d} : X \bullet X = 1, X \text{ is super-symmetric}\}$ and $S_0 = \{X \in S \mid \text{rank}(X) = 1\}$, and a square-free super-symmetric tensor $F \in \mathbb{R}^{n^d}$. If we draw ξ uniformly over S_0 , then there exists a universal constant $c > 0$, such that

$$\text{Prob} \left\{ F \bullet \xi \geq \sqrt{\frac{d!}{48(4n)^d}} \max_{X \in S} F \bullet X = \sqrt{\frac{d!}{48(4n)^d}} \|F\|_2 \right\} \geq c.$$

Applying the results straightforwardly, we obtain polynomial-time randomized approximation algorithms for solving various polynomial optimization models with high probability. Specifically, our results include:

1. $\Theta\left(\prod_{i=1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}}\right)$ -approximation ratio for

$$\begin{aligned} & \max F(x^1, x^2, \dots, x^d) \\ & \text{s.t. } x^i \in \mathbb{B}^{n_i}, i = 1, 2, \dots, d. \end{aligned}$$

This ratio improves that of $\Theta\left(\prod_{i=1}^{d-2} \sqrt{\frac{1}{n_i}}\right)$ proposed by He, Li, and Zhang [9].

2. $\Theta\left(n^{-\frac{d}{2}}\right)$ -approximation ratio for

$$\begin{aligned} & \max f(x) := F(x, \underbrace{x, \dots, x}_d) \\ & \text{s.t. } x \in \mathbb{B}^n, \end{aligned}$$

where $f(x)$ is a homogeneous polynomial function with the tensor F being square-free. This ratio is new. In the literature, when $d \geq 4$ and is even, the only previous approximation ratio for this model was in He, Li, and Zhang [9]; however, the ratio there is a relative one.

3. $\Theta\left(\prod_{i=1}^{d-2} \sqrt{\frac{\ln n_i}{n_i}}\right)$ -approximation ratio for

$$\begin{aligned} & \max F(x^1, x^2, \dots, x^d) \\ & \text{s.t. } x^i \in \mathbb{S}^{n_i}, i = 1, 2, \dots, d. \end{aligned}$$

This improves the $\prod_{i=1}^{d-2} \sqrt{\frac{1}{n_i}}$ approximation ratio in [8], and achieves the same theoretical bound as in So [21]. However, the algorithm proposed here is straightforward to implement, while the one in [21] is very involved.

4. $\Theta \left(\prod_{n \in N} \sqrt{\frac{\ln n}{n}} \right)$ -approximation ratio for

$$\begin{aligned} \max \quad & F(x^1, x^2, \dots, x^d, y^1, y^2, \dots, y^{d'}) \\ \text{s.t.} \quad & x^i \in \mathbb{B}^{n_i}, i = 1, 2, \dots, d, \\ & y^j \in \mathbb{S}^{m_j}, j = 1, 2, \dots, d', \end{aligned}$$

where N is the set of the $d + d' - 2$ smallest numbers in $\{n_1, \dots, n_d, m_1, \dots, m_{d'}\}$. This ratio improves that of $\Theta \left(\prod_{i=1}^{d-1} \sqrt{\frac{1}{n_i}} \prod_{j=1}^{d'-1} \sqrt{\frac{1}{m_j}} \right)$ proposed in [9].

Before concluding the whole paper, let us finally remark the results in this paper are also connected to Khintchine's inequality [12], which asserts that:

If $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ are i.i.d. symmetric Bernoulli random variables, then for any $p > 0$, there exist constants b_p and c_p such that for any vector $a \in \mathbb{R}^n$,

$$b_p \|a\|_2 \leq \left(\mathbb{E} |\xi^T a|^p \right)^{\frac{1}{p}} \leq c_p \|a\|_2. \quad (16)$$

Since Khintchine's result in early 1923, much effort has been on determining the sharp values of the constants b_p and c_p or some sort of extensions of Khintchine's inequality. In particular, similar inequality in the matrix case (the vector a is replaced by a matrix) has been established. Recently, So [22] proved a conjecture proposed by Nemirovskii through matrix version Khintchine's inequality.

Observe that Lemma 2.4 is exactly the lower bound part of Khintchine's inequality with $p = 1$. Furthermore our new probability inequality (Theorem 3.1) also implies the lower bound part of Khintchine's inequality where the random variables are endowed with some dependent structures.

Corollary 5.1 *Suppose $F \in \mathbb{R}^{n^d}$ is a square-free super-symmetric tensor of order d , and $\Xi = \underbrace{\xi \otimes \xi \otimes \dots \otimes \xi}_d$, where $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ are independent random variables with $\mathbb{E} \xi_i = 0, \mathbb{E} \xi_i^2 = 1, \mathbb{E} \xi_i^4 \leq \kappa$. Then for any $p > 0$, there exists a constant $b_p = \left(\frac{2\sqrt{3}-3}{4^{p+d} 9^{1+d} d^2 (d!)^{2-\frac{p}{2}} \kappa^{(d+\frac{p}{2})}} \right)^{\frac{1}{p}}$, such that*

$$b_p \|F\|_2 \leq (\mathbb{E} |F \bullet \Xi|^p)^{\frac{1}{p}}. \quad (17)$$

To see why, by Theorem 3.1 we have $\text{Prob} \left\{ |F \bullet \Xi| \geq \frac{\sqrt{d!} \|F\|_2}{4\sqrt{\kappa}} \right\} \geq \frac{2\sqrt{3}-3}{9d^2 (d!)^2 36^d \kappa^d}$. Therefore

$$\mathbb{E} |F \bullet \Xi|^p \geq \text{Prob} \left\{ |F \bullet \Xi|^p \geq \left(\frac{\sqrt{d!} \|F\|_2}{4\sqrt{\kappa}} \right)^p \right\} \cdot \left(\frac{\sqrt{d!} \|F\|_2}{4\sqrt{\kappa}} \right)^p \geq \frac{2\sqrt{3}-3}{9d^2 (d!)^2 36^d \kappa^d} \left(\frac{\sqrt{d!} \|F\|_2}{4\sqrt{\kappa}} \right)^p,$$

which implies (17).

Acknowledgments

The first author's research was supported in part by Hong Kong GRF Grant [Grant CityU143711]; the second and fourth author's research was supported in part by the National Science Foundation [Grant CMMI-1161242]; the third author's research was supported in part by Natural Science Foundation of China [Grant 11371242], Natural Science Foundation of Shanghai [Grant 12ZR1410100], and Ph.D. Programs Foundation of Chinese Ministry of Education [Grant 20123108120002].

References

- [1] N. Alon and A. Naor, *Approximating the Cut-Norm via Grothendieck's Inequality*, SIAM Journal on Computing, 35, 787–803, 2006.
- [2] A. Ben-Tal, A. Nemirovskii, and C. Roos, *Robust Solutions of Uncertain Quadratic and Conic-Quadratic Problems*, SIAM Journal on Optimization, 13, 535–560, 2002.
- [3] A. Brieden, P. Gritzmann, R. Kannan, V. Klee, L. Lovász, and M. Simonovits, *Approximation of Diameters: Randomization Doesn't Help*, The 39th Annual IEEE Symposium on Foundations of Computer Science, 244–251, 1998.
- [4] A. Brieden, P. Gritzmann, R. Kannan, V. Klee, L. Lovász, and M. Simonovits, *Deterministic and Randomized Polynomial-Time Approximation of Radii*, Mathematika, 48, 63–105, 2001.
- [5] M. Charikar and A. Wirth, *Maximizing Quadratic Programs: Extending Grothendieck's Inequality*, The 45th Annual IEEE Symposium on Foundations of Computer Science, 54–60, 2004.
- [6] M.X. Goemans and D.P. Williamson, *Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems using Semidefinite Programming*, Journal of the ACM, 42, 1115–1145, 1995.
- [7] B. Gröbbaum, *Partitions of Mass-Distributions and of Convex Bodies by Hyperplanes*, Pacific Journal of Mathematics, 10, 1257–1271, 1960.
- [8] S. He, Z. Li, and S. Zhang, *Approximation Algorithms for Homogeneous Polynomial Optimization with Quadratic Constraints*, Mathematical Programming, 125, 353–383, 2010.
- [9] S. He, Z. Li, and S. Zhang, *Approximation Algorithms for Discrete Polynomial Optimization*, Journal of the Operations Research Society of China, 1, 3–36, 2013.
- [10] S. He, Z.-Q. Luo, J. Nie, and S. Zhang, *Semidefinite Relaxation Bounds for Indefinite Homogeneous Quadratic Optimization*, SIAM Journal on Optimization, 19, 503–523, 2008.
- [11] S. He, J. Zhang, and S. Zhang, *Bounding Probability of Small Deviation: A Fourth Moment Approach*, Mathematics of Operations Research, 35, 208–232, 2010.
- [12] A. Khintchine, *Über Dyadische Brüche*, Mathematische Zeitschrift, 18, 109–116, 1923.
- [13] S. Khot and A. Naor, *Linear Equations Modulo 2 and the L_1 Diameter of Convex Bodies*, SIAM Journal on Computing, 38, 1448–1463, 2008.
- [14] B. Laurent and P. Massart, *Adaptive Estimation of a Quadratic Functional by Model Selection*, The Annals of Statistics, 28, 1302–1338, 2000.
- [15] Z. Li, *Polynomial Optimization Problems—Approximation Algorithms and Applications*, Ph.D. Thesis, The Chinese Univesrity of Hong Kong, Shatin, Hong Kong, June 2011.
- [16] Z. Li, S. He, and S. Zhang, *Approximation Methods for Polynomial Optimization: Models, Algorithms, and Applications*, SpringerBriefs in Optimization, Springer, New York, 2012.

- [17] Z.-Q. Luo and S. Zhang, *A Semidefinite Relaxation Scheme for Multivariate Quartic Polynomial Optimization with Quadratic Constraints*, SIAM Journal on Optimization, 20, 1716–1736, 2010.
- [18] Z.-Q. Luo, N.D. Sidiropoulos, P. Tseng, and S. Zhang, *Approximation Bounds for Quadratic Optimization with Homogeneous Quadratic Constraints*, SIAM Journal on Optimization, 18, 1–28, 2007.
- [19] Yu. Nesterov, *Semidefinite Relaxation and Nonconvex Quadratic Optimization*, Optimization Methods and Softwares, 9, 141–160, 1998.
- [20] R. Paley and A. Zygmund, *A Note on Analytic Functions in the Unit Circle*, Mathematical Proceedings of the Cambridge Philosophical Society, 28, 266–272, 1932.
- [21] A.M.-C. So, *Deterministic Approximation Algorithms for Sphere Constrained Homogeneous Polynomial Optimization Problems*, Mathematical Programming, 129, 357–382, 2011.
- [22] A.M.-C. So, *Moment Inequalities for Sums of Random Matrices and Their Applications in Optimization*, Mathematical Programming, 130, 125–151, 2011.
- [23] Y.-X. Yuan, *A Counter-Example to a Conjecture of Ben-Tal, Nemirovski and Roos*, Journal of the Operations Research Society of China, 1, 155–157, 2013.

A Proofs of Theorem 3.1 and Proposition 3.2.

The whole appendix is devoted to the proof of Theorem 3.1, among which Proposition 3.2 is proved as a byproduct. First, we observe that $\|F\|_2 \geq n^{-\frac{d}{2}}\|F\|_1$ since $F \in \mathbb{R}^{n^d}$, and thus (13) can be immediately derived from (12). Hence we shall focus on (12).

Furthermore, we observe that Theorem 3.1 is almost equivalent to the fact that any homogeneous polynomial function of independent random variables with bounded kurtosis should also have a bounded kurtosis itself, as formulated as follows:

Theorem A.1 *Let $F \in \mathbb{R}^{n^d}$ be a square-free super-symmetric tensor of order d , and let $f(x) = F(x, x, \dots, x)$ be a homogeneous polynomial function induced by F . If $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ are independent random variables with $\mathbb{E}\xi_i = 0, \mathbb{E}\xi_i^2 = 1, \mathbb{E}\xi_i^4 \leq \kappa$ for all $i = 1, 2, \dots, n$, then $\mathbb{E}f^4(\xi) \leq d^2(d!)^2 36^d \kappa^d (\mathbb{E}f^2(\xi))^2$.*

Before proving the theorem, let us note another important fact required in the proof, namely if a random variable has a bounded kurtosis, then it has a constant probability above the mean plus some constant proportion of the standard deviation.

Lemma A.2 *For any random variable z with its kurtosis upper bounded by $\kappa > 0$, namely*

$$\mathbb{E}[z - \mathbb{E}z]^4 \leq \kappa (\mathbb{E}[z - \mathbb{E}z]^2)^2,$$

we have

$$\text{Prob} \left\{ z \geq \mathbb{E}z + \frac{\sqrt{\text{Var}(z)}}{4\sqrt{\kappa}} \right\} \geq \frac{2\sqrt{3} - 3}{9\kappa}.$$

Proof. By normalizing z , i.e., letting $y = (z - \mathbf{E}z)/\sqrt{\text{Var}(z)}$, we shall have $\mathbf{E}y = 0$, $\mathbf{E}y^2 = 1$ and $\mathbf{E}y^4 \leq \kappa$. Thus we only need to show $\text{Prob} \left\{ y \geq \frac{1}{4\sqrt{\kappa}} \right\} \geq \frac{2\sqrt{3}-3}{9\kappa}$.

Denote $x = t - y$, where the constant $t > 0$ will be decided later. We have

$$\begin{aligned} \mathbf{E}x &= t - \mathbf{E}y = t, \\ \mathbf{E}x^2 &= t^2 - 2t\mathbf{E}y + \mathbf{E}y^2 = t^2 + 1, \\ \mathbf{E}x^4 &= t^4 - 4t^3\mathbf{E}y + 6t^2\mathbf{E}y^2 - 4t\mathbf{E}y^3 + \mathbf{E}y^4 \leq t^4 + 6t^2 + 4t\sqrt{\kappa} + \kappa, \end{aligned}$$

where $(\mathbf{E}y^3)^2 \leq \mathbf{E}y^2\mathbf{E}y^4 \leq \kappa$ is applied in the last inequality.

By applying Theorem 2.3 of [11], for any constant $v > 0$

$$\begin{aligned} \text{Prob} \{y \geq t\} &= \text{Prob} \{x \leq 0\} \\ &\geq \frac{4(2\sqrt{3}-3)}{9} \left(-\frac{2\mathbf{E}x}{v} + \frac{3\mathbf{E}x^2}{v^2} - \frac{\mathbf{E}x^4}{v^4} \right) \\ &\geq \frac{4(2\sqrt{3}-3)}{9} \left(-\frac{2t}{v} + \frac{3t^2+3}{v^2} - \frac{t^4+6t^2+4t\sqrt{\kappa}+\kappa}{v^4} \right) \\ \left(\text{let } t = \frac{1}{4\sqrt{\kappa}} \text{ and } v = \sqrt{\kappa} \right) &= \frac{4(2\sqrt{3}-3)}{9} \left(-\frac{1}{2\kappa} + \frac{3}{16\kappa^2} + \frac{3}{\kappa} - \frac{1}{256\kappa^4} - \frac{6}{16\kappa^3} - \frac{1}{\kappa^2} - \frac{1}{\kappa} \right) \\ &= \frac{4(2\sqrt{3}-3)}{9} \left(\frac{24}{16\kappa} - \frac{13}{16\kappa^2} - \frac{6}{16\kappa^3} - \frac{1}{256\kappa^4} \right) \\ (\text{notice } \kappa \geq \mathbf{E}y^4 \geq (\mathbf{E}y^2)^2 = 1) &\geq \frac{4(2\sqrt{3}-3)}{9} \cdot \frac{4}{16\kappa} = \frac{2\sqrt{3}-3}{9\kappa}. \end{aligned}$$

□

Let us now prove Theorem A.1. We start with a special case when $d = 2$ and ξ are symmetric Bernoulli random variables, which helps to illustrate the ideas underlying the proof for the general case.

Proposition A.3 *Let $F \in \mathbb{R}^{n \times n}$ be a diagonal-free symmetric matrix, and let $f(x) = x^\top Fx$. If $\xi \sim \mathbb{B}^n$, then $\mathbf{E}f^4(\xi) \leq 15(\mathbf{E}f^2(\xi))^2$.*

Proof. Rewrite $y = f(\xi) = \sum_{\sigma} a_{\sigma} \xi^{\sigma}$, where $\sigma \in \Pi := \{(1, 2), (1, 3), \dots, (n-1, n)\}$ and $\xi^{(i,j)} := \xi_i \xi_j$. Since $\mathbf{E}\xi_i^d = 0$ for odd d and $\mathbf{E}\xi_i^d = 1$ for even d , the non-zero terms in $\mathbf{E}y^4$ are all in the forms of $a_{ij}a_{ij}a_{ij}a_{ij}$, $a_{ij}a_{ij}a_{ik}a_{ik}$, $a_{ij}a_{ij}a_{kl}a_{kl}$ and $a_{ij}a_{ik}a_{jl}a_{kl}$, where we assume i, j, k and ℓ are distinctive. Let us count the different types of terms.

Type A: $a_{ij}a_{ij}a_{ij}a_{ij}$. The total number of such type of terms is $\binom{n}{2}$;

Type B: $a_{ij}a_{ij}a_{ik}a_{ik}$. The total number of such type of terms is $n \cdot \binom{n-1}{2} \cdot \binom{4}{2}$;

Type C: $a_{ij}a_{ij}a_{kl}a_{kl}$. The total number of such type of terms is $\binom{n}{4} \cdot 3 \cdot \binom{4}{2}$;

Type D: $a_{ij}a_{ik}a_{jl}a_{kl}$. The total number of such type of terms is $\binom{n}{4} \cdot 3 \cdot 4!$.

Notice that

$$(\mathbf{E}y^2)^2 = \left(\sum_{\sigma \in \Pi} a_{\sigma}^2 \right)^2 = \sum_{\sigma \in \Pi} a_{\sigma}^4 + 2 \sum_{\sigma_1 \neq \sigma_2} a_{\sigma_1}^2 a_{\sigma_2}^2 =: \text{'Part I'} + \text{'Part II'}.$$

Type A terms constitute exactly ‘Part I’ in $(Ey^2)^2$; each item of Types B and C will appear exactly once in ‘Part II’ of $(Ey^2)^2$; each term of Type D can be bounded by an average of two terms in ‘Part II’ of $(Ey^2)^2$ since $a_{ij}a_{ik}a_{j\ell}a_{k\ell} \leq (a_{ij}^2a_{k\ell}^2 + a_{ik}^2a_{j\ell}^2)/2$. The number of the terms of Types B, C and D is:

$$n \cdot \binom{n-1}{2} \binom{4}{2} + \binom{n}{4} \cdot 3 \cdot \binom{4}{2} + \binom{n}{4} \cdot 3 \cdot 4! = \frac{n(n-1)(n-2)(15n-33)}{4} =: N$$

and there are

$$\binom{n}{2} \cdot \left(\binom{n}{2} - 1 \right) = \frac{n(n-1)(n-2)(n+1)}{4} =: N'$$

terms in ‘Part II’ of $(Ey^2)^2$. Clearly $N \leq 15N'$, which leads to $Ey^4 \leq 15(Ey^2)^2$. \square

We are now in a position to prove Proposition 3.2, which follows from Proposition A.3 and Lemma A.2.

Proof of Proposition 3.2 *Proof.* Since F is diagonal-free and symmetric, it is easy to verify $E[\xi^T F \xi] = 0$ and

$$\text{Var}(\xi^T F \xi) = \sum_{\sigma \in \Pi} a_\sigma^2 = 4 \sum_{\sigma \in \Pi} (a_\sigma/2)^2 = 2\|F\|_2^2.$$

By Lemma A.2 we have $\text{Prob} \left\{ \xi^T F \xi \geq \frac{\sqrt{\text{Var}(\xi^T F \xi)}}{4\sqrt{15}} \right\} \geq \frac{2\sqrt{3}-3}{135}$, the desired inequality holds. \square

Let us now come to the proof of main theorem in the appendix.

Proof of Theorem A.1 *Proof.* Let $I := \{1, 2, \dots, n\}$ be the index set, and Π be the set containing all the combinations of d distinctive indices in I . Obviously $|\Pi| = \binom{n}{d}$. For any $\pi \in \Pi$, we denote $x^\pi := \prod_{i \in \pi} x_i$ and $x^{\pi_1 + \pi_2} := x^{\pi_1} x^{\pi_2}$ (e.g. $x^{\{1,2\}} = x_1 x_2$ and $x^{\{1,2\} + \{1,3\}} = x^{\{1,2\}} x^{\{1,3\}} = x_1 x_2 \cdot x_1 x_3 = x_1^2 x_2 x_3$).

Since F is square-free and super-symmetric, y can be written as $\sum_{\pi \in \Pi} a_\pi x^\pi$, or simply $\sum_{\pi} a_\pi x^\pi$ (whenever we write summation over π , it means the summation over all $\pi \in \Pi$). We thus have

$$Ey^2 = E \left[\sum_{\pi_1, \pi_2} a_{\pi_1} x^{\pi_1} a_{\pi_2} x^{\pi_2} \right] = \sum_{\pi_1, \pi_2} a_{\pi_1} a_{\pi_2} E x^{\pi_1 + \pi_2} = \sum_{\pi_1 = \pi_2} a_{\pi_1} a_{\pi_2} E x^{\pi_1 + \pi_2} = \sum_{\pi} a_\pi^2.$$

Our task is to bound

$$Ey^4 = E \left[\sum_{\pi_1, \pi_2, \pi_3, \pi_4} a_{\pi_1} x^{\pi_1} a_{\pi_2} x^{\pi_2} a_{\pi_3} x^{\pi_3} a_{\pi_4} x^{\pi_4} \right] = \sum_{\pi_1, \pi_2, \pi_3, \pi_4} a_{\pi_1} a_{\pi_2} a_{\pi_3} a_{\pi_4} E x^{\pi_1 + \pi_2 + \pi_3 + \pi_4}. \quad (18)$$

For any combination quadruple $\{\pi_1, \pi_2, \pi_3, \pi_4\}$, there are in total $4d$ indices, with each index appearing at most 4 times. Suppose there are a number of indices appearing 4 times, b number of indices appearing 3 times, c number of indices appearing twice, and g number of indices appearing once. Clearly $4a + 3b + 2c + g = 4d$. In order to compute the summation of all the terms $a_{\pi_1} a_{\pi_2} a_{\pi_3} a_{\pi_4} E x^{\pi_1 + \pi_2 + \pi_3 + \pi_4}$ over $\pi_1, \pi_2, \pi_3, \pi_4 \in \Pi$ in (18), we shall group them according to different $\{a, b, c, g\}$.

1. $g \geq 1$: as we know $E x_i = 0$ for all $i \in I$, all the terms in this group will vanish.
2. $b = c = g = 0$: the summation of all the terms in this group is

$$\sum_{\pi_1 = \pi_2 = \pi_3 = \pi_4} a_{\pi_1} a_{\pi_2} a_{\pi_3} a_{\pi_4} E x^{\pi_1 + \pi_2 + \pi_3 + \pi_4} = \sum_{\pi_1} a_{\pi_1}^4 E x^{4\pi_1} \leq \kappa^d \sum_{\pi} a_\pi^4.$$

3. $g = 0$ and $b + c \geq 1$: we shall classify all the terms in this group step by step. In the following, we assume $|\Pi| \geq 2$ and $n \geq d + 1$ to avoid triviality.

- It is clear that $4a + 3b + 2c = 4d$, $0 \leq a \leq d - 1$, $0 \leq b \leq (4d - 4a)/3$ and b must be even. In this group, the number of different $\{a, b, c\}$ is at most $\sum_{a=0}^{d-1} (1 + \lfloor \frac{4d-4a}{6} \rfloor) \leq d^2$.
- For any given triple $\{a, b, c\}$, there are total $\binom{n}{a} \binom{n-a}{b} \binom{n-a-b}{c}$ number of distinctive ways to assign indices. Clearly, we have $\binom{n}{a} \binom{n-a}{b} \binom{n-a-b}{c} \leq n!/(n-a-b-c)! \leq n!/(n-2d)_{+!}$.
- For any given a indices appearing 4 times, b indices appearing 3 times, and c indices appearing twice, we shall count how many distinctive ways they can form a particular combination quadruple $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ (note that orders do count). For the indices appearing 4 times, they do not have choice but to be located in $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ each once; for indices appearing 3 times, each has at most 4 choices; for indices appearing twice, each has at most 6 choices. Therefore, the total number of distinctive ways to formulate the combination of quadruples is upper bounded by $4^b 6^c \leq 6^{2d}$.
- For any given combination quadruple $\{\pi_1, \pi_2, \pi_3, \pi_4\}$, noticing that $(\text{Ex}_i^3)^2 \leq \text{Ex}_i^2 \text{Ex}_i^4 \leq \kappa$ for all $i \in I$, we have $|\text{Ex}^{\pi_1+\pi_2+\pi_3+\pi_4}| \leq \kappa^a \cdot (\sqrt{\kappa})^b \cdot 1^c = \kappa^{a+b/2} \leq \kappa^d$.
- For any given combination quadruple $\{\pi_1, \pi_2, \pi_3, \pi_4\}$, in this group each combination can appear at most twice. Specifically, if we assume $i \neq j$ (implying $\pi_i \neq \pi_j$), then the forms of $\{\pi_1, \pi_1, \pi_1, \pi_2\}$ and $\{\pi_1, \pi_1, \pi_1, \pi_1\}$ do not appear. The only possible forms are $\{\pi_1, \pi_2, \pi_3, \pi_4\}$, $\{\pi_1, \pi_1, \pi_2, \pi_3\}$ and $\{\pi_1, \pi_1, \pi_2, \pi_2\}$. We notice that

$$\begin{aligned} a_{\pi_1} a_{\pi_2} a_{\pi_3} a_{\pi_4} &\leq (a_{\pi_1}^2 a_{\pi_2}^2 + a_{\pi_1}^2 a_{\pi_3}^2 + a_{\pi_1}^2 a_{\pi_4}^2 + a_{\pi_2}^2 a_{\pi_3}^2 + a_{\pi_2}^2 a_{\pi_4}^2 + a_{\pi_3}^2 a_{\pi_4}^2)/6, \\ a_{\pi_1} a_{\pi_1} a_{\pi_2} a_{\pi_3} &\leq (a_{\pi_1}^2 a_{\pi_2}^2 + a_{\pi_1}^2 a_{\pi_3}^2)/2, \\ a_{\pi_1} a_{\pi_1} a_{\pi_2} a_{\pi_2} &= a_{\pi_1}^2 a_{\pi_2}^2. \end{aligned}$$

Therefore, in any possible form, each $a_{\pi_1} a_{\pi_2} a_{\pi_3} a_{\pi_4}$ can be *on average* upper bounded by one item $a_{\pi_1}^2 a_{\pi_2}^2 (\pi_1 \neq \pi_2)$ in $\sum_{\pi_1 \neq \pi_2} a_{\pi_1}^2 a_{\pi_2}^2$.

Overall, in this group, by noticing the symmetry of Π , the summation of all the terms is upper bounded by $d^2 \cdot \frac{n!}{(n-2d)_{+!}} \cdot 6^{2d} \cdot \kappa^d$ number of items in form of $a_{\pi_1}^2 a_{\pi_2}^2 (\pi_1 \neq \pi_2)$ in $\sum_{\pi_1 \neq \pi_2} a_{\pi_1}^2 a_{\pi_2}^2$. Notice that there are in total $|\Pi|(|\Pi| - 1)/2 = \frac{1}{2} \binom{n}{d} (\binom{n}{d} - 1)$ items in $\sum_{\pi_1 \neq \pi_2} a_{\pi_1}^2 a_{\pi_2}^2$, and each item is evenly distributed. By symmetry, the summation of all the terms in this group is upper bounded by

$$\frac{d^2 \cdot \frac{n!}{(n-2d)_{+!}} \cdot 6^{2d} \cdot \kappa^d}{\frac{1}{2} \binom{n}{d} (\binom{n}{d} - 1)} \sum_{\pi_1 \neq \pi_2} a_{\pi_1}^2 a_{\pi_2}^2 \leq d^2 (d!)^2 36^d \kappa^d \cdot 2 \sum_{\pi_1 \neq \pi_2} a_{\pi_1}^2 a_{\pi_2}^2.$$

Finally, we are able to bound Ey^4 by

$$\begin{aligned} \text{Ey}^4 &\leq \kappa^d \sum_{\pi} a_{\pi}^4 + d^2 (d!)^2 36^d \kappa^d \cdot 2 \sum_{\pi_1 \neq \pi_2} a_{\pi_1}^2 a_{\pi_2}^2 \\ &\leq d^2 (d!)^2 36^d \kappa^d \left(\sum_{\pi} a_{\pi}^4 + 2 \sum_{\pi_1 \neq \pi_2} a_{\pi_1}^2 a_{\pi_2}^2 \right) \\ &= d^2 (d!)^2 36^d \kappa^d \left(\sum_{\pi} a_{\pi}^2 \right)^2 = d^2 (d!)^2 36^d \kappa^d (\text{Ey}^2)^2. \end{aligned}$$

Putting the pieces together, the theorem follows. \square

Finally, combining Theorem A.1 and Lemma A.2, and noticing $\text{Var}(f(\xi)) = d!\|F\|_2^2$ in Theorem 3.1, lead us to the probability bound (12) in Theorem 3.1, which concludes the whole proof.