Moments Tensors, Hilbert's Identity, and k-wise Uncorrelated Random Variables

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

August 31, 2013

Abstract

In this paper we introduce a notion to be called k-wise uncorrelated random variables, which is similar but not identical to the so-called k-wise independent random variables in the literature. We show how to construct k-wise uncorrelated random variables by a simple procedure. The constructed random variables can be applied, e.g. to express the quartic polynomial $(x^{\mathrm{T}}Qx)^2$, where Q is an $n \times n$ positive semidefinite matrix, by a sum of fourth powered polynomial terms, known as Hilbert's identity. By virtue of the proposed construction, the number of required terms is no more than $2n^4 + n$. This implies that it is possible to find a $(2n^4 + n)$-point distribution whose fourth moments tensor is exactly the symmetrization of $Q \otimes Q$. Moreover, we prove that the number of required fourth powered polynomial terms to express $(x^{\mathrm{T}}Qx)^2$ is at least $n(n + 1)/2$. The result is applied to prove that computing the matrix $2 \mapsto 4$ norm is NP-hard. Extensions of the results to complex random variables are discussed as well.

Keywords: cone of moments, uncorrelated random variables, Hilbert's identity, matrix norm.

Mathematics Subject Classification: 78M05, 62H20, 15A60.

[*]Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, USA. Email: jiang373@umn.edu.

[†]Department of Management Sciences, City University of Hong Kong, Hong Kong. Email: simaihe@cityu.edu.hk.

[‡]Department of Mathematics, University of Portsmouth, Portsmouth PO1 3HF, 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

Given an n-dimensional random vector $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^{\mathrm{T}}$ with joint density function $p(\cdot)$, let us denote the n-dimensional d-th order tensor $\mathcal{F}$ to be the d-th order *moments* tensor associated with ξ as follows:

$$\mathcal{F}_{i_1 i_2 \ldots i_d} = \mathsf{E}\left[\prod_{k=1}^{d} \xi_{i_k}\right] = \int_{\mathbb{R}^n} \prod_{k=1}^{d} u_{i_k}\, p(u) du \quad \forall\, 1 \le i_1, i_2, \ldots, i_d \le n;$$

or equivalently,

$$\mathcal{F} = \int_{\mathbb{R}^n} \underbrace{u \otimes u \otimes \cdots \otimes u}_{d}\, p(u) du.$$

Since tensor $\mathcal{F}$ is in a finite dimensional space, by Carathéodory's theorem [6], it can be further rewritten as a sum of finite "rank one" terms, i.e., there exist t vectors $b^1, b^2, \ldots, b^t$ such that

$$\mathcal{F} = \sum_{i=1}^{t} \underbrace{b^i \otimes b^i \otimes \cdots \otimes b^i}_{d}. \tag{1}$$

An immediate consequence of the above construction is that $\mathcal{F}$ is super-symmetric, meaning that its component is invariant under permutation of the indices. For instance, the second order moments tensor can be easily derived from its covariance matrix, which is naturally symmetric and positive semidefinite. Indeed, thanks to the formulation (1), any $2d$-th order moments tensor is always positive semidefinite, in other words, the homogeneous polynomial function induced by this tensor is always nonnegative, i.e.,

$$f(x) = \mathcal{F}(\underbrace{x, x, \ldots, x}_{2d}) := \sum_{1 \le i_1, i_2, \ldots, i_{2d} \le n} \mathcal{F}_{i_1 i_2 \ldots i_{2d}} \prod_{k=1}^{2d} x_{i_k} = \sum_{i=1}^{t} \left((b^i)^{\mathrm{T}} x\right)^{2d} \ge 0.$$

However, the term 'nonnegativity' can be ambiguous in the case of higher order tensors. In our recent paper [11], this issue was particularly addressed. We shall only note here that the $2d$-th moments tensors form a specific nonnegative convex cone, whose membership query is a hard problem in general (see [11]). It is therefore interesting to know what kind of tensors are contained in this cone. For instance, one may wonder if the super-symmetric tensor associated with the polynomial $(x^{\mathrm{T}}x)^2$, which is clearly nonnegative, is a fourth order moments tensor or not. Interestingly, the answer is yes, due to a result of Hilbert [10], who showed that it is possible to express $(x^{\mathrm{T}}x)^d$ as $\sum_{i=1}^{t}(x^{\mathrm{T}}a^i)^{2d}$. As a consequence, the polynomial $(x^{\mathrm{T}}x)^2$ (the case $d = 2$) can be viewed as $\mathsf{E}[\xi^{\mathrm{T}}x]^4$ where ξ is a random vector, taking value $t^{1/4}a^i$ with probability $1/t$. Therefore, $\mathrm{sym}\,(I \otimes I)$ with I being the identity matrix is a fourth moments tensor, where the *symmetrization* mapping 'sym' turns a given tensor into a super-symmetric one by making the entries with the same set of indices all the same (taking the value of the average).

Apart from the above example, there are several other representations for general $2d$-th moments tensor other than (1). For example, with the help of Hilbert's identity [4], we can easily verify that $\mathrm{sym}\,(\underbrace{A \otimes A \otimes \cdots \otimes A}_{d})$ with $A \succeq 0$ also belongs to $2d$-th moments cone. Specifically, one can find vectors $a^1, a^2, \ldots, a^t$ such that

$$\mathrm{sym}\,(\underbrace{A \otimes A \otimes \cdots \otimes A}_{d}) = \sum_{i=1}^{t} \underbrace{a^i \otimes a^i \otimes \cdots \otimes a^i}_{2d}. \tag{2}$$

On the other hand, by letting the order of the tensor be $2d$ and $A^i = b^i \otimes b^i = b^i (b^i)^{\mathrm{T}}$ in (1), we have

$$\mathcal{F} = \sum_{i=1}^{t} \underbrace{b^i \otimes b^i \otimes \cdots \otimes b^i}_{2d} = \sum_{i=1}^{t} \mathrm{sym}\,(\underbrace{A^i \otimes A^i \otimes \cdots \otimes A^i}_{d}), \text{ with } A^i \succeq 0 \text{ and } \mathrm{rank}(A^i) = 1. \quad (3)$$

This implies that the rank-one constraint is redundant in terms of requiring $\mathcal{F}$ to be a $2d$-th moments tensor in (3).

In general, such decomposition of (2) is not unique. For example, one may verify that

$$(x_1^2 + x_2^2 + x_3^2)^2 = \frac{1}{3}\sum_{i=1}^{3} x_i^4 + \frac{1}{3} \sum_{1 \le i < j \le 3} \sum_{\beta_j = \pm 1} (x_i + \beta_j x_j)^4 = \frac{2}{3}\sum_{i=1}^{3} x_i^4 + \frac{1}{3} \sum_{\substack{\beta_2 = \pm 1 \\ \beta_3 = \pm 1}} (x_1 + \beta_2 x_2 + \beta_3 x_3)^4,$$

which leads to two different representations of the tensor $\mathrm{sym}\,(I_3 \otimes I_3)$. An interesting question is to find a succinct (preferably the shortest) representation among all the different representations, including the one from Hilbert's decomposition. However, from the original Hilbert's construction, the representation on the right hand side of (2) is exponential in n. By Carathéodory's theorem, there exists a decomposition such that the value of t in (2) is no more than $\binom{n+2d-1}{2d} + 1$. Unfortunately, Carathéodory's theorem is non-constructive. This motivates us to construct a *polynomial-size* representation, i.e., $t = O(n^k)$ for some constant k in (2).

One contribution of this paper is to give a 'short' (polynomial-size) representation for Hilbert's identity when $d = 2$. In fact, we also prove the number of terms for *any* representation can never be less than $n(n+1)/2$. An application of this polynomial-size representation will be discussed. Toward this end, let us first introduce the new notion of *k-wise uncorrelated random variables*, which may appear to be completely unrelated to the discussion of Hilbert's identity at first glance.

Definition 1.1 **(k-wise uncorrelation)** *A set of random variables $\{\xi_1, \xi_2, \ldots, \xi_n\}$ is called k-wise uncorrelated if*

$$\mathsf{E}\left[\prod_{j=1}^{n} \xi_j^{p_j}\right] = \prod_{j=1}^{n} \mathsf{E}\left[\xi_j^{p_j}\right] \quad \forall\, p_1, p_2, \ldots, p_n \in \mathbb{Z}_+ \text{ with } \sum_{i=1}^{n} p_i = k.$$

For instance, if $\xi_1, \xi_2, \ldots, \xi_n$ are i.i.d. random variables with finite supporting set $|\Delta| = q$, then they are k-wise uncorrelated. However the size of its corresponding sample space is q^n, which is exponential in n. It turns out that reducing the sample space while keeping the k-wise uncorrelation structure can be of great importance in many applications. For example, our result shows that the polynomial-size representation (2) can be obtained by finding k-wise uncorrelated random variables with polynomial-size sample space. Before addressing the issue of finding such random variables, below we shall first discuss a related notion known as the k-wise independence.

Definition 1.2 **(k-wise independence)** *A set of random variables $\Xi = \{\xi_1, \xi_2, \ldots, \xi_n\}$ with each taking values on the set $\Delta = \{\delta_1, \delta_2, \ldots, \delta_q\}$ is called k-wise independent, if any k different random variables $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k}$ of Ξ are independent, i.e.,*

$$\mathrm{Prob}\,\{\xi_{i_1} = \delta_{i_1}, \xi_{i_2} = \delta_{i_2}, \ldots, \xi_{i_k} = \delta_{i_k}\} = \prod_{j=1}^{k} \mathrm{Prob}\,\left\{\xi_{i_j} = \delta_{i_j}\right\} \quad \forall\, \delta_{i_j} \in \Delta,\ j = 1, 2, \ldots, k.$$

Note that when $k = 2$, k-wise independence is usually called pair-wise independence. Since 1980's, k-wise independence has been a popular topic in theoretical computer science. Essentially, working with k-wise independence (instead of the full independence) means that one can reduce the size of the sample space in question. In many cases, this feature is crucial. For instance, when $\Delta = \{0, 1\}$ and $\text{Prob}\,\{\xi_1 = 0\} = \text{Prob}\,\{\xi_1 = 1\} = \frac{1}{2}$, Alon, Babai, and Itai [1] constructed a sample space of size being approximately $n^{\frac{k}{2}}$. For the same Δ, when $\xi_1, \xi_2, \dots, \xi_n$ are independent but not identical, Karloff and Mansour [13] proved that the size of sample space can be upper bounded by $O(n^k)$. In the case of $\Delta = \{0, 1, \dots, q-1\}$ with q being a prime number, the total number of random variables being k-wise independent are quite restricted. For given $k < q$, Joffe [12] showed that there are up to $q + 1$ random variables form a k-wise independent set and the size of the sample space is q^k.

Clearly, k-wise independence implies k-wise uncorrelation. Therefore, we may apply the existing results of k-wise independence to get k-wise uncorrelated random variables. However, the aforementioned constructions of k-wise independent random variables heavily depend on the structure of Δ (e.g. it requires that $|\Delta| = 2$ or $k < |\Delta|$). Moreover, the construction of k-wise independent random variables is typically complicated and technically involved (see [13]). In fact, for certain problems (e.g. polynomial-size representation of Hilbert's identity in this case), we only need the random variables to be k-wise uncorrelated. Therefore, in this paper we propose a tailor-made simple construction which suits the structure of k-wise uncorrelated random variables. As we shall see later, our approach can handle the more general support set:

$$\Delta_q := \{1, \omega_q, \dots, \omega_q^{q-1}\}, \text{ with } \omega_q = e^{\boldsymbol{i}\frac{2\pi}{q}} = \cos\frac{2\pi}{q} + \boldsymbol{i}\sin\frac{2\pi}{q} \text{ and } q \text{ is prime}, \tag{4}$$

and k can be any parameter. Conceptually, our approach is rather generic: the k-wise uncorrelated random variables are constructed based only on the product of a small set of i.i.d. random variables with their powers; the sample space would be polynomial-size if the number of such i.i.d. random variables is $O(\log n)$. Consequently, we not only find polynomial-size representation for the fourth moments tensor in form of $\text{sym}\,(A \otimes A)$, but also for complex $2^d q$-th moments tensor. As an application, this construction can be used to prove that the matrix $2 \mapsto 4$ norm problem [5], whose complexity was previously unknown[1], is actually NP-hard.

The rest of this paper is organized as follows. In Section 2 we introduce Hilbert's identity and its connections to $2d$-th moments tensor. Then, in Section 3 we present a randomized algorithm, as well as a deterministic one, to construct k-wise uncorrelated random variables. As a result, we find polynomial-size representation of fourth moments tensor and complex $2^d q$-th moments tensor in Section 4. In Section 5, we discuss the shortest representation of Hilbert's identity and its related tensor rank problem, in particular providing a lower bound for the number of terms in the identity. Finally, we conclude this paper with an application of determining the complexity of matrix $2 \mapsto 4$ norm problem, to illustrate the usefulness of our approach.

Notation. Throughout we adopt the notation of the lower-case letters to denote vectors (e.g. $x \in \mathbb{R}^n$), the capital letters to denote matrices (e.g. $A \in \mathbb{R}^{n^2}$), and the capital calligraphy letters to denote higher (≥ 3) order tensors (e.g. $\mathcal{F} \in \mathbb{R}^{n^4}$), with subscriptions of indices being their entries (e.g. $x_1, A_{ij}, \mathcal{F}_{i_1 i_2 i_3 i_4} \in \mathbb{R}$). A tensor is said to be *super-symmetric* if its entries are invariant under all permutations of its indices. As mentioned earlier, the *symmetrization* mapping 'sym' makes a given tensor to be super-symmetric, which is $\mathcal{F} = \text{sym}\,(\mathcal{G})$ with

$$\mathcal{F}_{i_1 i_2 \dots i_d} = \frac{1}{|\Pi(i_1 i_2 \dots i_d)|} \sum_{\pi \in \Pi(i_1 i_2 \dots i_d)} \mathcal{G}_\pi \quad \forall\, 1 \leq i_1, i_2, \dots, i_d \leq n,$$

[1]During the review process of this paper, Barak et al. [3] independently proved that it is NP-hardness to compute the matrix $2 \mapsto 4$ norm.

where $\Pi(i_1 i_2 \ldots i_d)$ is the set of all distinct permutations of the indices $\{i_1, i_2, \ldots, i_d\}$. The symbol '$\otimes$' represents the outer product of vectors or matrices. In particular, if $\mathcal{F} = \underbrace{x \otimes x \otimes \cdots \otimes x}_{d}$ for some $x \in \mathbb{R}^n$, then $\mathcal{F}_{i_1 i_2 \ldots i_d} = \prod_{k=1}^{d} x_{i_k}$; and if $\mathcal{G} = \underbrace{X \otimes X \otimes \cdots \otimes X}_{d}$ for some $X \in \mathbb{R}^{n^2}$, then $\mathcal{G}_{i_1 i_2 \ldots i_{2d}} = \prod_{k=1}^{d} X_{i_{2k-1} i_{2k}}$. Besides, Δ denotes the supporting set of certain random variable, and $\Omega \subseteq \mathbb{R}^n$ is the sample space of a set of random variables $\{\xi_1, \xi_2, \ldots, \xi_n\}$, i.e., the space of all possible outcomes of $(\xi_1, \xi_2, \ldots, \xi_n)^{\mathrm{T}}$. Finally, the following two subsets of $\mathbb{Z}_+^n$ are frequently used in the discussion,

$$\mathbb{P}_k^n := \left\{ (p_1, p_2, \ldots, p_n)^{\mathrm{T}} \in \mathbb{Z}_+^n \mid p_1 + p_2 + \cdots + p_n = k \right\},$$

and for given prime number q,

$$\mathbb{P}_k^n(q) := \{ p \in \mathbb{P}_k^n \mid \exists\, i \,(1 \le i \le n) \text{ such that } q \nmid p_i \}.$$

It is easy to see that $|\mathbb{P}_k^n(q)| \le |\mathbb{P}_k^n| = \binom{n+k-1}{k}$.

2 Hilbert's Identity and $2d$-th Moments Tensor

Let us start our discussion with the famous Hilbert's identity, which states that for any fixed positive integers d and n, there always exist rational vectors $b^1, b^2, \ldots, b^t \in \mathbb{R}^n$ such that

$$\left(\sum_{i=1}^{n} x_i^2 \right)^d = \sum_{j=1}^{t} \left((b^j)^{\mathrm{T}} x \right)^{2d} \quad \forall\, x = (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n. \tag{5}$$

For instance, when $n = 4$ and $d = 2$, we have

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)^2 = \frac{1}{6} \sum_{1 \le i < j \le 4} (x_i + x_j)^4 + \frac{1}{6} \sum_{1 \le i < j \le 4} (x_i - x_j)^4, \tag{6}$$

which is called Liouville's identity. It is worth mentioning that Hilbert's identity is very well known and is a fundamental result in mathematics. For example, with the help of (5), Reznick [18] managed to prove the following result:

> Let $p(x)$ be $2d$-th degree homogeneous positive polynomial in $x \in \mathbb{R}^n$. Then there exists a positive integer r and vectors $b^1, b^2, \ldots, b^r \in \mathbb{R}^n$ such that
>
> $$\|x\|_2^{2r-2d} p(x) = \sum_{i=1}^{r} ((b^i)^{\mathrm{T}} x)^{2r}.$$

Reznick's result above solved Hilbert's seventeenth problem constructively (albeit only for the case $p(x)$ being positive definite). As another example, Hilbert [10] in 1909 solved Waring's problem:

> Can every positive integer be expressed as a sum of at most $g(k)$ k-th powers of positive integers, where $g(k)$ depends only on k, not on the number being represented?

in the affirmative for all k. The key underpinning tool in the proof is also Hilbert's identity (5); see e.g. [7, 16] for more stories on Warning's problem and Hilbert's identity. In fact, Hilbert's identity can be readily extended to a more general setting. For any given $A \succeq 0$, by letting $y = A^{\frac{1}{2}}x$ and applying (5), one has

$$(x^{\mathrm{T}}Ax)^2 = (y^{\mathrm{T}}y)^2 = \sum_{j=1}^{t}\left((b^j)^{\mathrm{T}}y\right)^{2d} = \sum_{j=1}^{t}\left((b^j)^{\mathrm{T}}A^{\frac{1}{2}}x\right)^{2d},$$

which guarantees the existence of vectors $a^1, a^2, \dots, a^t \in \mathbb{R}^n$ with $a^j = A^{\frac{1}{2}}b^j$ for $j = 1, 2, \dots, t$ such that

$$(x^{\mathrm{T}}Ax)^d = \sum_{j=1}^{t}\left((a^j)^{\mathrm{T}}x\right)^{2d}. \tag{7}$$

The discussion so far appears to be only concerned about decomposing a specific polynomial function. Let us now relate Hilbert's identity to the moments tensor. Observe that super-symmetric tensors are bijectively related to homogenous polynomial functions. In particular, if

$$f(x) = \sum_{1\le i_1\le i_2\le\cdots\le i_d\le n} \mathcal{G}_{i_1 i_2 \dots i_d}\prod_{k=1}^{d} x_{i_k}$$

is a d-th degree homogenous polynomial, then its associated super-symmetric tensor $\mathcal{F}$ with $\mathcal{F}_{i_1 i_2 \dots i_d} = \mathcal{G}_{i_1 i_2 \dots i_d}/|\Pi(i_1 i_2 \dots i_d)|$ is uniquely determined by $f(x) = \mathcal{F}(\underbrace{x, x, \dots, x}_{d})$, and vice versa. This is the same as the one-to-one correspondence between symmetric matrices and quadratic forms. Therefore, the tensor $\mathrm{sym}\,(\underbrace{A\otimes A\otimes\cdots\otimes A}_{d})$ is associated with the polynomial $\left(x^{\mathrm{T}}Ax\right)^d$, and the following relationship holds immediately.

Proposition 2.1 *For any $A \succeq 0$, there exit vectors $a^1, a^2, \dots, a^t \in \mathbb{R}^n$ such that $(x^{\mathrm{T}}Ax)^2 = \sum_{j=1}^{t}\left((a^j)^{\mathrm{T}}x\right)^{2d}$, i.e., $\mathrm{sym}\,(\underbrace{A\otimes A\otimes\cdots\otimes A}_{d}) = \sum_{i=1}^{t}\underbrace{a^i\otimes a^i\otimes\cdots\otimes a^i}_{2d}$. This implies that tensor $\mathrm{sym}\,(\underbrace{A\otimes A\otimes\cdots\otimes A}_{d})$ is a $2d$-th moments tensor if $A \succeq 0$.*

As we mentioned earlier, the size of such representation from Hilbert's identity is exponential in n. To see this, let us recall the claim of Hilbert (see [14]):

> Given fixed positive integers d and n, there exist $2d+1$ real numbers $\beta_1, \beta_2, \dots, \beta_{2d+1}$, $2d+1$ positive real numbers $\rho_1, \rho_2, \dots, \rho_{2d+1}$, and a positive real number α_d, such that
>
> $$(x^{\mathrm{T}}x)^d = \frac{1}{\alpha_d}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_{2d+1}=1}^{n}\rho_{i_1}\rho_{i_2}\dots\rho_{i_{2d+1}}(\beta_{i_1}x_1+\beta_{i_2}x_2+\cdots+\beta_{i_{2d+1}}x_{i_{2d+1}})^{2d}. \tag{8}$$

It is obvious that the number of $2d$-powered linear terms on the right hand side of (8) is $(2d+1)^n$, which is too lengthy for practical purposes. In the following, let us focus on how to get a polynomial-size decomposition of Hilbert's identity, or essentially the tensor $\mathrm{sym}\,(\underbrace{A\otimes A\otimes\cdots\otimes A}_{d})$ with $A \succeq 0$.

In light of the above discussion, it suffices to find a polynomial-size representation of (5). Toward this end, let us first rewrite $(x^{\mathrm{T}}x)^d$ in terms of the expectation of a polynomial function. In particular, by defining i.i.d. random variables $\xi_1, \xi_2, \dots, \xi_n$ with supporting set $\Delta = \{\beta_1, \beta_2, \dots, \beta_{2d+1}\}$

and $\text{Prob}\,(\xi_k = \beta_i) = \frac{\rho_i}{\gamma_d}$ for all $1 \le i \le 2d+1$ and $1 \le k \le n$, where $\gamma_d = \sum_{i=1}^{2d+1} \beta_i$, identity (8) is equivalent to

$$(x^{\mathrm{T}}x)^d = \frac{\gamma_d^d}{\alpha_d} \mathsf{E}\left[\left(\sum_{j=1}^n \xi_j x_j\right)^{2d}\right] = \frac{\gamma_d^d}{\alpha_d} \sum_{p \in \mathbb{P}_{2d}^n} \mathsf{E}\left[\prod_{j=1}^n \xi_j^{p_j}\right] \prod_{j=1}^n x_j^{p_j} = \frac{\gamma_d^d}{\alpha_d} \sum_{p \in \mathbb{P}_{2d}^n} \prod_{j=1}^n \mathsf{E}\left[\xi_j^{p_j}\right] \prod_{j=1}^n x_j^{p_j}. \quad (9)$$

As a consequence, if for any n random variables $\eta_1, \eta_2, \ldots, \eta_n$ satisfying

$$\mathsf{E}\left[\prod_{j=1}^n \eta_j^{p_j}\right] = \prod_{j=1}^n \mathsf{E}\left[\eta_j^{p_j}\right] \quad \forall\, p \in \mathbb{P}_{2d}^n, \quad (10)$$

and $\mathsf{E}\left[\eta_j^p\right] = \mathsf{E}\left[\xi_1^p\right]$ for all $0 < p \le 2d$ and $1 \le j \le n$, then it is straightforward to verify that $(x^{\mathrm{T}}x)^d = \frac{\gamma_d^d}{\alpha_d} \mathsf{E}\left[\left(\sum_{j=1}^n \eta_j x_j\right)^{2d}\right]$. Notice that (10) is actually equivalent to $\eta_1, \eta_2, \ldots, \eta_n$ being $2d$-wise uncorrelated, and we have the next result following (9) and (10).

Proposition 2.2 *If $\xi_1, \xi_2, \ldots, \xi_n$ are i.i.d. random variables, and $\eta_1, \eta_2, \ldots, \eta_n$ are $2d$-wise uncorrelated, satisfying the moments constraints $\mathsf{E}\left[\eta_j^p\right] = \mathsf{E}\left[\xi_1^p\right]$ for all $0 < p \le 2d$ and $1 \le j \le n$, then $\mathsf{E}\left[\left(\sum_{j=1}^n \xi_j x_j\right)^{2d}\right] = \mathsf{E}\left[\left(\sum_{j=1}^n \eta_j x_j\right)^{2d}\right]$.*

We end this section with the conclusion that the key to reducing the length of representation in (5) is to construct $2d$-wise uncorrelated random variables satisfying certain moments conditions, such that the sample space is as small as possible, which will be the subject of our subsequent discussions. As we will see later, the construction makes use of the structure of the support set (4). For general support sets, the techniques considered in [13] may be useful, and it is a topic for future research.

3 Construction of k-wise Uncorrelated Random Variables

In this section, we shall construct k-wise uncorrelated random variables, which are identical and uniformly distributed on Δ_q defined by (4). The rough idea is as follows. We first generate m i.i.d. random variables $\xi_1, \xi_2, \ldots, \xi_m$, based on which we can define new random variables $\eta_1, \eta_2, \ldots, \eta_n$ such that $\eta_i := \prod_{1 \le j \le m} \xi_j^{c_{ij}}$ for $i = 1, 2, \ldots, n$. Therefore, the size of sample space of $\{\eta_1, \eta_2, \ldots, \eta_n\}$ is bounded above by q^m, which yields a polynomial-size space if we let $m = O(\log_q n)$. The remaining part of this section is devoted to the discussion of the property for the power indices c_{ij}'s, in order to guarantee $\eta_1, \eta_2, \ldots, \eta_n$ to be k-wise uncorrelated, and how to find those power indices.

3.1 k-wise Regular Sequence

Let us start with some notations and definitions for the preparation. Suppose c is a number with m digits and $c[\ell]$ is the value of its ℓ-th bit. We call c to be endowed with the base q, if $c[\ell] \in \{0, 1, \ldots, q-1\}$ for all $1 \le \ell \le m$. In other words, $c = \sum_{\ell=1}^m c[\ell] q^{\ell-1}$. Now we can define the concept of k-wise regular sequence as follows.

Definition 3.1 *A sequence of m digits numbers $\{c_1, c_2, \ldots, c_n\}$ of base q is called k-wise regular if for any $p \in \mathbb{P}_k^n(q)$, there exists $\ell\,(1 \le \ell \le m)$ such that*

$$\sum_{j=1}^{n} p_j \cdot c_j[\ell] \neq 0 \ mod\ q.$$

Why are we interested in such regular sequences? The answer lies in the following proposition.

Proposition 3.2 *Suppose m digits numbers $\{c_1, c_2, \ldots, c_n\}$ of base q are k-wise regular, where q is a prime number, and $\xi_1, \xi_2, \ldots, \xi_m$ are i.i.d. random variables uniformly distributed on Δ_q. Then $\eta_1, \eta_2, \ldots, \eta_n$ with*

$$\eta_i := \prod_{1 \le \ell \le m} \xi_\ell^{c_i[\ell]},\ i = 1, 2, \ldots, n \tag{11}$$

are k-wise uncorrelated.

Proof. Let $\eta_1, \eta_2, \ldots, \eta_n$ be defined as in (11). As ξ_i is uniformly distributed on Δ_q for $1 \le i \le m$ and q is prime, we have

$$\mathsf{E}[\xi_i^p] = \mathsf{E}\left[\eta_j^p\right] = \begin{cases} 1 & \text{if } q \mid p, \\ 0 & \text{otherwise} \end{cases}$$

for any i and any j with $c_j \neq (0, 0, \ldots, 0)$. Otherwise if $c_j = (0, 0, \ldots, 0)$ for some j, then $\mathsf{E}\left[\eta_j^p\right] = 1$.

For any given $p \in \mathbb{P}_k^n$, if $q \mid p_i$ for all $1 \le i \le n$, then

$$\begin{aligned}
\mathsf{E}\left[\prod_{j=1}^{n} \eta_j^{p_j}\right] &= \mathsf{E}\left[\left(\prod_{1 \le \ell \le m} \xi_\ell^{p_1 \cdot c_1[\ell]}\right)\left(\prod_{1 \le \ell \le m} \xi_\ell^{p_2 \cdot c_2[\ell]}\right) \cdots \left(\prod_{1 \le \ell \le m} \xi_\ell^{p_n \cdot c_n[\ell]}\right)\right] \\
&= \prod_{1 \le \ell \le m} \mathsf{E}\left[\xi_\ell^{\sum_{j=1}^{n} p_j \cdot c_j[\ell]}\right] = 1 = \prod_{j=1}^{n} \mathsf{E}\left[\eta_j^{p_j}\right].
\end{aligned}$$

Otherwise, there exists some i_0 such that $q \nmid p_{i_0}$, implying that $p \in \mathbb{P}_k^n(q)$. By k-wise regularity, we can find some ℓ_0 satisfying $\sum_{j=1}^{n} p_j \cdot c_j[\ell_0] \neq 0$ mod q, implies that $\mathsf{E}\left[\xi_{\ell_0}^{\sum_{j=1}^{n} p_j \cdot c_j[\ell_0]}\right] = 0$. Moreover, there exists some j_0 such that $p_{j_0} \cdot c_{j_0}[\ell_0] \neq 0$ mod q, i.e., $q \nmid p_{j_0}$ and $c_{j_0}[\ell_0] \neq 0$. This leads to $\mathsf{E}\left[\eta_{j_0}^{p_{j_0}}\right] = 0$, and we have

$$\mathsf{E}\left[\prod_{j=1}^{n} \eta_j^{p_j}\right] = \prod_{1 \le \ell \le m} \mathsf{E}\left[\xi_\ell^{\sum_{j=1}^{n} p_j \cdot c_j[\ell]}\right] = 0 = \prod_{j=1}^{n} \mathsf{E}\left[\eta_j^{p_j}\right],$$

and the conclusion follows. □

3.2 A Randomized Algorithm

We shall now focus on how to find such k-wise regular sequence $\{c_1, c_2, \ldots, c_n\}$ of base q. First, we present a randomized process, in which $c_i[\ell]$ is randomly and uniformly chosen from $\{0, 1, \ldots, q-1\}$ for all $1 \le i \le n$ and $1 \le \ell \le m$. The algorithm is as follows.

Algorithm RAN

Input:	Dimension n and $m := \lceil k \log_q n \rceil$.
Output:	A sequence $\{c_1, c_2, \ldots, c_n\}$ in m digits of base q.
Step 0:	Construct $S = \{(\underbrace{0, \ldots, 0, 0}_{m}), (\underbrace{0, \ldots, 0, 1}_{m}), \ldots, (\underbrace{q-1, \ldots, q-1, q-1}_{m})\}$ of base q.
Step 1:	Independently and uniformly take $c_i \in S$ for $i = 1, 2, \ldots, n$.
Step 2:	Assemble the sequence $\{c_1, c_2, \ldots, c_n\}$ and exit.

Theorem 3.3 *If $1 < k < n$ and q is a prime number, then* **Algorithm RAN** *returns a k-wise m-digit regular sequence $\{c_1, c_2, \ldots, c_n\}$ of base q with probability at least $1 - \frac{(1.5)^{k-1}}{k!}$, which is independent of n and q.*

Proof. Since $\{c_1, c_2, \ldots, c_n\}$ is a sequence of m-digit numbers of base q, if it is not regular, then there exist $p \in \mathbb{P}_k^n$, such that

$$\sum_{j=1}^{n} p_j \cdot c_j[\ell] = 0 \bmod q \quad \forall\, 1 \le \ell \le m.$$

Therefore, we have

$$\text{Prob}\,\{\{c_1, c_2, \ldots, c_n\} \text{ is not } k\text{-wise regular}\} \le \sum_{p \in \mathbb{P}_k^n(q)} \text{Prob}\left\{\sum_{j=1}^{n} p_j \cdot c_j[\ell] = 0 \bmod q,\, \forall\, 1 \le \ell \le m\right\}.$$

For any given $p \in \mathbb{P}_k^n(q)$, we may without loss of generality assume that $q \nmid p_n$. If we fix $c_1, c_2, \ldots, c_{n-1}$, as q is prime, then there is only one solution for c_n such that $\sum_{j=1}^{n} p_j \cdot c_j[\ell] = 0 \bmod q, \forall\, 1 \le \ell \le m$. Combining the fact that $c_1, c_2, \ldots, c_n$ are independently and uniformly generated, we have

$$\begin{aligned}
& \text{Prob}\left\{\sum_{j=1}^{n} p_j \cdot c_j[\ell] = 0 \bmod q,\, \forall\, 1 \le \ell \le m\right\} \\
= \; & \text{Prob}\left\{\sum_{j=1}^{n} p_j \cdot c_j[\ell] = 0 \bmod q,\, \forall\, 1 \le \ell \le m \,\middle|\, c_1 = d_1, c_2 = d_2, \ldots, c_{n-1} = d_{n-1}\right\} \cdot \\
& \sum_{d_1, d_2, \ldots, d_{n-1} \in S} \text{Prob}\,\{c_1 = d_1, c_2 = d_2, \ldots, c_{n-1} = d_{n-1}\} \\
= \; & \frac{1}{q^m} \sum_{d_1, d_2, \ldots, d_{n-1} \in S} \text{Prob}\,\{c_1 = d_1, c_2 = d_2, \ldots, c_{n-1} = d_{n-1}\} \\
\le \; & \frac{1}{n^k}.
\end{aligned} \tag{12}$$

Finally,

$$\begin{aligned}
& \text{Prob}\,\{\{c_1, c_2, \ldots, c_n\} \text{ is } k\text{-wise regular}\} \\
= \; & 1 - \text{Prob}\,\{\{c_1, c_2, \ldots, c_n\} \text{ is not } k\text{-wise regular}\} \\
\ge \; & 1 - |\mathbb{P}_k^n(q)| \cdot \frac{1}{n^k} \ge 1 - |\mathbb{P}_k^n| \cdot \frac{1}{n^k} = 1 - \binom{n+k-1}{k} \cdot \frac{1}{n^k} \ge 1 - \frac{(1.5)^{k-1}}{k!}.
\end{aligned}$$

$\square$

For some special q and k, in particular relating to the the simplest case of Hilbert's identity (4-wise regular sequence of base 2), the lower bound of the probability in Theorem 3.3 can be improved.

Proposition 3.4 *If $k = 4$ and $q = 2$, then* **Algorithm RAN** *returns a 4-wise regular sequence $\{c_1, c_2, \ldots, c_n\}$ of base 2 with probability at least $1 - \frac{1}{2n^2} - \frac{1}{4!}$.*

The proof is similar to that of Theorem 3.3, and thus is omitted.

3.3 Derandomization

Although k-wise regular sequence always exists and can be found with high probability, one may however wish to construct such regular sequence *deterministically*. In fact, this is possible if we apply Theorem 3.3 in a slightly different manner, which is shown in the following algorithm. Basically, we start with a short regular sequence C, and enumerate all the remaining numbers in order to find c such that $C \cup \{c\}$ is also regular. Updating C with $C \cup \{c\}$, we repeat this procedure until the cardinality of C reaches n. Moreover, thanks to the polynomial-size sample space, this 'brute force' approach still runs in polynomial-time.

Algorithm DET

Input: Dimension n and $m := \lceil k \log_q n \rceil$.

Output: A sequence $\{c_1, c_2, \ldots, c_n\}$ in m digits of base q.

Step 0: Construct $S = \{(\underbrace{0, \ldots, 0, 0}_{m}), (\underbrace{0, \ldots, 0, 1}_{m}), \ldots, (\underbrace{q-1, \ldots, q-1, q-1}_{m})\}$ of base q, and a sequence $C := \{c_1, c_2, \ldots, c_k\}$ in m digits, where $c_i := (0, \ldots, 0, 0, 1, \underbrace{0, \ldots, 0, 0}_{k-1})$ for $i = 1, 2, \ldots, k$. Let the index count be $\tau := k$.

Step 1: If $\tau = n$, then go to Step 2; Otherwise enumerate $S \setminus C$ to find a $c \in S \setminus C$ such that $C \cup \{c\}$ is k-wise regular. Let $c_{\tau+1} := c$, $C := C \cup \{c_{\tau+1}\}$ and $\tau := \tau + 1$, and return to Step 1.

Step 2: Assemble the sequence $\{c_1, c_2, \ldots, c_n\}$ and exit.

It is obvious that the initial sequence $\{c_1, c_2, \ldots, c_k\}$ is k-wise regular. In order for **Algorithm DET** to exit successfully, it remains to argue that it is always possible to expand the k-wise regular sequence by one in Step 1, as long as $\tau < n$.

Theorem 3.5 *Suppose that $3 \le k \le \tau < n$, q is a prime number, and C with $|C| = \tau$ is k-wise regular. If we* uniformly *pick $c_{\tau+1}$ from S, then*

$$\text{Prob}\,\{C \cup \{c_{\tau+1}\} \text{ is } k\text{-wise regular}\} \ge 1 - \frac{(1.5)^k}{k!}\left(\frac{\tau+1}{n}\right)^k,$$

ensuring that $\{c_{\tau+1} \in S \mid C \cup \{c_{\tau+1}\} \text{ is } k\text{-wise regular}\} \ne \emptyset$.

Proof. Like in the proof of Theorem 3.3, we have

$$\text{Prob}\,\{C \cup \{c_{\tau+1}\} \text{ is not } k\text{-wise regular}\} \le \sum_{p \in \mathbb{P}_k^{\tau+1}(q)} \text{Prob}\left\{\sum_{j=1}^{\tau+1} p_j \cdot c_j[\ell] = 0 \bmod q,\ \forall\, 1 \le \ell \le m\right\}.$$

For any $p \in \mathbb{P}_k^{\tau+1}(q)$, since q is prime, by using a similar argument as of (12), we can get

$$\mathrm{Prob}\left\{\sum_{j=1}^{\tau+1} p_j \cdot c_j[\ell] = 0 \bmod q,\, \forall\, 1 \le \ell \le m\right\} \le \frac{1}{n^k}.$$

Essentially, the argument in (12) works by conditioning on the elements in C, the selection ordering in C during the previous steps is not important. Therefore,

$$\mathrm{Prob}\left\{C \cup \{c_{\tau+1}\} \text{ is } k\text{-wise regular}\right\} \ge 1 - \left|\mathbb{P}_k^{\tau+1}(q)\right| \frac{1}{n^k} \ge 1 - \binom{\tau+k}{k}\frac{1}{n^k} \ge 1 - \frac{(1.5)^k}{k!}\left(\frac{\tau+1}{n}\right)^k > 0.$$

□

By the above theorem, Step 1 of **Algorithm DET** guarantees to expand the k-wise regular sequence of base q before reaching the desired cardinality $\tau = n$. A straightforward computation shows that **Algorithm DET** requires an overall complexity of $O(n^{2k-1}\log_q n)$.

4 Polynomial-Size Representation of Moments Tensor

4.1 Polynomial-Size Representation of the Fourth Moments Tensor

With the help of k-wise uncorrelated random variables, we are able to construct polynomial-size representation of the fourth moments tensor. In Hilbert's construction (9), the support set Δ is too general to apply the result in Section 3. However as we mentioned earlier, such decomposition of (9) is not unique. In fact, when $d = 2$, we observe that

$$(x^{\mathrm{T}}x)^2 = \left(\sum_{i=1}^n x_i^2\right)^2 = \frac{2}{3}\sum_{i=1}^n x_i^4 + \frac{1}{3}\,\mathsf{E}\left[\left(\sum_{j=1}^n \xi_j x_j\right)^4\right], \tag{13}$$

where $\xi_1, \xi_2, \dots, \xi_n$ are i.i.d. symmetric Bernoulli random variables. Applying either **Algorithm RAN** or **Algorithm DET** leads to a 4-wise regular sequence of base 2, based on which we can define random variables $\eta_1, \eta_2, \dots, \eta_n$ as we did in (11). Proposition 3.2 guarantees that $\eta_1, \eta_2, \dots, \eta_n$ are 4-wise uncorrelated, and it is easy to check that

$$\mathsf{E}[\eta_j] = \mathsf{E}[\eta_j^3] = \mathsf{E}[\xi_1] = \mathsf{E}[\xi_1^3] = 0,\ \mathsf{E}[\eta_j^2] = \mathsf{E}[\eta_j^4] = \mathsf{E}[\xi_1^2] = \mathsf{E}[\xi_1^4] = 1 \quad \forall\, 1 \le j \le n.$$

Thus, by Proposition 2.2, we have $\mathsf{E}\left[\left(\sum_{j=1}^n \eta_j x_j\right)^4\right] = \mathsf{E}\left[\left(\sum_{j=1}^n \xi_j x_j\right)^4\right]$. Moreover, the size of the sample space of $\{\eta_1, \eta_2, \dots, \eta_n\}$ is at most $2^{\lceil k \log_q n\rceil} \le 2n^4$, which means the new representation has at most $n + 2n^4$ fourth powered terms. Combining with Proposition 2.1, we have the following main result.

Theorem 4.1 *Given a positive integer n, we can find τ $(\le 2n^4)$ vectors $b^1, b^2, \dots, b^\tau \in \mathbb{R}^n$ in polynomial time, such that*

$$(x^{\mathrm{T}}x)^2 = \frac{2}{3}\sum_{i=1}^n x_i^4 + \sum_{j=1}^{\tau}\left((b^j)^{\mathrm{T}}x\right)^4 \quad \forall\, x \in \mathbb{R}^n,$$

or equivalently,

$$\mathrm{sym}\,(I \otimes I) = \frac{2}{3}\sum_{i=1}^{n} e^i \otimes e^i \otimes e^i \otimes e^i + \sum_{j=1}^{\tau} b^j \otimes b^j \otimes b^j \otimes b^j,$$

where $e^i \in \mathbb{R}^n$ is the i-th unit vector (with the i-th entry 1 and other entries zeros).

The result can be extended to a more general setting as follows.

Corollary 4.2 *Given a positive semidefinite matrix $A \in \mathbb{R}^{n\times n}$, we can find $\tau\,(\le 2n^4+n)$ vectors $a^1, a^2, \ldots, a^\tau \in \mathbb{R}^n$ in polynomial time, such that*

$$(x^{\mathrm{T}}Ax)^2 = \sum_{i=1}^{\tau}\left((a^i)^{\mathrm{T}}x\right)^4 \quad \forall\, x \in \mathbb{R}^n,$$

or equivalently,

$$\mathrm{sym}\,(A \otimes A) = \sum_{i=1}^{\tau} a^i \otimes a^i \otimes a^i \otimes a^i.$$

Proof. Due to the one to one correspondence between super-symmetric tensors and homogeneous polynomials, we only need to prove the first identity. By letting $y = A^{\frac{1}{2}}x$ and applying Theorem 4.1, we can find $b^1, b^2, \ldots, b^\tau$ in polynomial time with $\tau \le 2n^4$, such that

$$(x^{\mathrm{T}}Ax)^2 = (y^{\mathrm{T}}y)^2 = \frac{2}{3}\sum_{i=1}^{n} y_i^4 + \sum_{j=1}^{\tau}\left((b^j)^{\mathrm{T}}y\right)^4 = \sum_{i=1}^{n}\left(\left(\frac{2}{3}\right)^{\frac{1}{4}}(e^i)^{\mathrm{T}}A^{\frac{1}{2}}x\right)^4 + \sum_{j=1}^{\tau}\left((b^j)^{\mathrm{T}}A^{\frac{1}{2}}x\right)^4.$$

The conclusion follows by letting $a^i = \left(\frac{2}{3}\right)^{\frac{1}{4}} A^{\frac{1}{2}}e^i$ for $i = 1, 2, \ldots, n$, and $a^{i+n} = A^{\frac{1}{2}}b^i$ for $i = 1, 2, \ldots, \tau$. □

4.2 Polynomial-Size Representation of Complex qd-th Moments Tensor

In this subsection we are going to generalize the result in Section 4.1 to qd-th moments tensor. Denote $\mathcal{I}^q$ to be the q-th order identity tensor, whose entry is 1 when all its indices are identical, and is zero otherwise. We are interested in whether $\underbrace{\mathcal{I}^q \otimes \mathcal{I}^q \otimes \cdots \otimes \mathcal{I}^q}_{d}$ is a qd-th moments tensor or not. If it is true, then for any given positive integers q, d and n, there exist vectors $a^1, a^2, \ldots, a^t \in \mathbb{R}^n$, such that

$$\mathrm{sym}\,(\underbrace{\mathcal{I}^q \otimes \mathcal{I}^q \otimes \cdots \otimes \mathcal{I}^q}_{d}) = \sum_{i=1}^{t} \underbrace{a^i \otimes a^i \otimes \cdots \otimes a^i}_{qd}, \tag{14}$$

or equivalently

$$\left(\sum_{i=1}^{n} x_i^q\right)^d = \sum_{j=1}^{t}\left((a^j)^{\mathrm{T}}x\right)^{qd} \quad \forall\, x \in \mathbb{R}^n. \tag{15}$$

Unfortunately, the above does not hold in general, as the following counter example shows.

Example 4.3 *The function $f(x) = (x_1^3 + x_2^3)^2 = x_1^6 + 2x_1^3x_2^3 + x_2^6$ cannot be decomposed in the form of (15) with $q = 3$ and $d = 2$, i.e., a sum of sixth powered linear terms.*

This can be easily proven by contradiction. Suppose we can find $a^1, a^2, \ldots, a^t \in \mathbb{R}^n$, such that

$$x_1^6 + 2x_1^3x_2^3 + x_2^6 = \sum_{i=1}^{t} (a_i x_1 + b_i x_2)^6 \,. \tag{16}$$

There must exist some (a_j, b_j) with $a_j b_j \neq 0$, since otherwise there is no monomial $x_1^3x_2^3$ in the right hand side of (16). As a consequence, the coefficient of monomial $x_1^2x_2^4$ in the right hand side of (16) is at least $\binom{6}{2} a_j^2 b_j^4 > 0$, which is null on the left side of the equation, leading to a contradiction.

In the same vein one can actually show that (15) cannot hold for any $q \geq 3$. Therefore, we turn to qd-th moments tensor in the complex domain, i.e., both entries of the tensor and vector a^i's in (14) and (15) are now allowed to take complex values. Similar to (13), we have the following identity:

$$\left(\sum_{j=1}^{n} x_j^q\right)^2 = \left(1 - \frac{2}{\binom{2q}{q}}\right) \sum_{j=1}^{n} x_j^{2q} + \frac{2}{\binom{2q}{q}} \mathsf{E}\left[\left(\sum_{i=1}^{n} \xi_i x_i\right)^{2q}\right], \tag{17}$$

where $\xi_1, \xi_2, \ldots, \xi_n$ are i.i.d. random variables uniformly distributed on Δ_q. Moreover, we can further prove (15) for the more general complex case.

Proposition 4.4 *For any given positive integers q, d and n, there exist $a^1, a^2, \ldots, a^\tau \in \mathbb{C}^n$ such that*

$$\left(\sum_{i=1}^{n} x_i^q\right)^{2^d} = \sum_{j=1}^{\tau} \left((a^j)^{\mathrm{T}} x\right)^{2^d q} \quad \forall\, x \in \mathbb{C}^n, \tag{18}$$

or equivalently,

$$\mathrm{sym}\,(\underbrace{\mathcal{I}^q \otimes \mathcal{I}^q \otimes \cdots \otimes \mathcal{I}^q}_{2^d}) = \sum_{i=1}^{t} \underbrace{a^i \otimes a^i \otimes \cdots \otimes a^i}_{2^d q}.$$

Proof. Due to the one to one correspondence between super-symmetric tensors and homogeneous polynomials, we only need to prove the first identity, whose proof is based on mathematical induction. The case $d = 1$ is already guaranteed by (17). Suppose that (18) is true for $d - 1$, then there exist $b^1, b^2, \ldots, b^t \in \mathbb{C}^n$ such that

$$\left(\sum_{i=1}^{n} x_i^q\right)^{2^d} = \left(\left(\sum_{i=1}^{n} x_i^q\right)^{2^{d-1}}\right)^2 = \left(\sum_{j=1}^{t} \left((b^j)^{\mathrm{T}} x\right)^{2^{d-1} q}\right)^2.$$

By applying (17) to the above identity, there exist $c^1, c^2, \ldots, c^\tau \in \mathbb{C}^t$, such that

$$\left(\sum_{i=1}^{n} x_i^q\right)^{2^d} = \left(\sum_{j=1}^{t} \left((b^j)^{\mathrm{T}} x\right)^{2^{d-1} q}\right)^2 = \sum_{i=1}^{\tau} \left(\sum_{j=1}^{t} (c^i)_j \cdot (b^j)^{\mathrm{T}} x\right)^{2^d q} = \sum_{i=1}^{\tau} \left((c^i)^{\mathrm{T}} B^{\mathrm{T}} x\right)^{2^d q},$$

where $B = (b^1, b^2, \ldots, b^t) \in \mathbb{C}^{n \times t}$. Letting $a^i = Bc^i \; (1 \leq i \leq \tau)$ completes the inductive step. □

The next step is to reduce the number τ in (18). Under the condition that q is prime, we can get a k-wise regular sequence of base q using either **Algorithm RAN** or **Algorithm DET**. With the help of Theorem 2.2, we can further get a polynomial-size representation of complex Hilbert's identity and complex $2^d q$-th moments tensor, by applying a similar argument as in Theorem 4.1.

Theorem 4.5 *For any given positive integers q, d and n with q being prime, we can find $\tau \le O\left(n^{(2q)^{2^{d-1}}}\right)$ vectors $a^1, a^2, \ldots, a^\tau \in \mathbb{C}^n$ in polynomial time, such that*

$$\left(\sum_{i=1}^n x_i^q\right)^{2^d} = \sum_{i=1}^{\tau} \left((a^i)^{\mathrm{T}} x\right)^{(2^d q)} \quad \forall\, x \in \mathbb{C}^n,$$

or equivalently,

$$\mathrm{sym}\,(\underbrace{\mathcal{I}^q \otimes \mathcal{I}^q \otimes \cdots \otimes \mathcal{I}^q}_{2^d}) = \sum_{i=1}^{\tau} \underbrace{a^i \otimes a^i \otimes \cdots \otimes a^i}_{2^d q}.$$

5 Shortest Representation of Hilbert's Identity

In Section 4.1, we constructed polynomial-size representation of Hilbert's identity, in particular, the fourth moments tensor $\mathrm{sym}\,(I \times I)$. The number of fourth powered linear functions required (in Theorem 4.1) is $n + 2n^4$. As we shall see later, this size is in general not smallest possible. This raises the issue of how to find the shortest representation of the fourth moments tensor. In general, we are interested in the following quantity:

$$\tau_{2d}(n) := \min_{m \in \mathbb{Z}_+} \left\{ \exists\, b^1, b^2, \ldots, b^m \in \mathbb{R}^n, \text{ such that } \left(x^{\mathrm{T}} x\right)^d = \sum_{i=1}^m \left((b^i)^{\mathrm{T}} x\right)^{2d} \forall\, x \in \mathbb{R}^n \right\}.$$

If fact, $\tau_{2d}(n)$ is closely related to the rank of the super-symmetric tensor $\mathrm{sym}\,(\underbrace{I \otimes I \otimes \cdots \otimes I}_{d})$, which is the following:

$$\rho_{2d}(n) := \min_{r \in \mathbb{Z}_+} \left\{ \exists\, b^1, b^2, \ldots, b^r \in \mathbb{R}^n, \lambda \in \mathbb{R}^r, \text{ such that } \left(x^{\mathrm{T}} x\right)^d = \sum_{i=1}^r \lambda_i \left((b^i)^{\mathrm{T}} x\right)^{2d} \forall\, x \in \mathbb{R}^n \right\},$$

or in the language of tensors, the smallest r such that

$$\mathrm{sym}\,(\underbrace{I \otimes I \otimes \cdots \otimes I}_{d}) = \sum_{i=1}^{r} \lambda_i \underbrace{b^i \otimes b^i \otimes \cdots \otimes b^i}_{d}.$$

The difference between $\tau_{2d}(n)$ and $\rho_{2d}(n)$ lies in the fact that the latter one allows negative rank-one tensors. Therefore we have $\tau_{2d}(n) \ge \rho_{2d}(n)$. Computing the exact values for $\tau_{2d}(n)$ and $\rho_{2d}(n)$ is not easy for general n and d, and the only clear case is for $d = 1$ whereas $\tau_2(n) = \rho_2(n) = n$. In this section we focus on the case $d = 2$, i.e., $\tau_4(n)$ and $\rho_4(n)$. In fact, the lower bound for $\tau_{2d}(n)$ was already studied by Reznick [17]. Below we first summarize the result of Reznick [17].

Theorem 5.1 (Theorem 8.15 of [17]) *For any given positive integers d and n, the number of d-th powered linear terms in Hilberts identity (5) is at least $\binom{n+d-1}{n-1}$, i.e., $\tau_{2d}(n) \ge \binom{n+d-1}{n-1}$.*

Furthermore when $d = 2$, the exact values $\tau_{2d}(n)$ for some specific n's are known in the literature.

Proposition 5.2 (Proposition 9.26 of [17]) $\tau_4(n) = \binom{n+2-1}{n-1} + 1 = \frac{n(n+1)}{2} + 1$ *when $n = 4, 5, 6$.*

We remark that when $d=2$, $n(n+1)/2$ is also a lower bound for the number of rank-one terms to represent $\mathrm{sym}\,(A \otimes A)$ with $A \succ 0$. Besides, if $\xi_1, \xi_2, \dots, \xi_n$ are symmetric Bernoulli random variables, and they are 4-wise uncorrelated, then Theorem 5.1 also indicates that $n(n+1)/2$ is a lower bound for the size of sample space generated by $\{\xi_1, \xi_2, \dots, \xi_n\}$. In fact, $n(n+1)/2$ is also a lower bound for the rank of $\mathrm{sym}\,(I \otimes I)$, as the following theorem stipulates.

Theorem 5.3 *For any positive integer n, it holds that $n(n+1)/2 \le \rho_4(n) \le n^2$.*

Proof. Denote the shortest representation to be

$$\left(\sum_{j=1}^n x_j^2\right)^2 = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j\right)^4 - \sum_{i=1}^{\ell} \left(\sum_{j=1}^n b_{ij} x_j\right)^4,$$

where $m + \ell = \rho_4(n)$. By comparing the coefficient of each monomial, we have

$$\left\{\begin{array}{ll} \sum_{i=1}^m a_{ij}^4 - \sum_{i=1}^{\ell} b_{ij}^4 = 1 & \forall\, 1 \le j \le n \\ \sum_{i=1}^m a_{ij_1}^2 a_{ij_2}^2 - \sum_{i=1}^{\ell} b_{ij_1}^2 b_{ij_2}^2 = \frac{1}{3} & \forall\, 1 \le j_1 \ne j_2 \le n \\ \sum_{i=1}^m a_{ij_1}^3 a_{ij_2} - \sum_{i=1}^{\ell} b_{ij_1}^3 b_{ij_2} = 0 & \forall\, 1 \le j_1 \ne j_2 \le n \\ \sum_{i=1}^m a_{ij_1}^2 a_{ij_2} a_{ij_3} - \sum_{i=1}^{\ell} b_{ij_1}^2 b_{ij_2} b_{ij_3} = 0 & \forall\, 1 \le j_1, j_2, j_3 \le n \text{ with } j_k \ne j_t \text{ if } k \ne t \\ \sum_{i=1}^m a_{ij_1} a_{ij_2} a_{ij_3} a_{ij_4} - \sum_{i=1}^{\ell} b_{ij_1} b_{ij_2} b_{ij_3} b_{ij_4} = 0 & \forall\, 1 \le j_1, j_2, j_3, j_4 \le n \text{ with } j_k \ne j_t \text{ if } k \ne t \end{array}\right. . \tag{19}$$

Construct matrices $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times \frac{n(n-1)}{2}}$, $C \in \mathbb{R}^{\ell \times n}$ and $D \in \mathbb{R}^{\ell \times \frac{n(n-1)}{2}}$, where

$$A = \begin{bmatrix} a_{11}^2 & a_{12}^2 & \dots & a_{1n}^2 \\ a_{21}^2 & a_{22}^2 & \dots & a_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^2 & a_{m2}^2 & \dots & a_{mn}^2 \end{bmatrix}, \; C = \begin{bmatrix} b_{11}^2 & b_{12}^2 & \dots & b_{1n}^2 \\ b_{21}^2 & b_{22}^2 & \dots & b_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ b_{\ell 1}^2 & b_{\ell 2}^2 & \dots & b_{\ell n}^2 \end{bmatrix},$$

$$B = \begin{bmatrix} a_{11}a_{12} & a_{11}a_{13} & \dots & a_{11}a_{1n} & a_{12}a_{13} & a_{12}a_{14} & \dots & a_{12}a_{1n} & \dots & a_{1,n-1}a_{1n} \\ a_{21}a_{22} & a_{21}a_{23} & \dots & a_{21}a_{2n} & a_{22}a_{23} & a_{22}a_{24} & \dots & a_{22}a_{2n} & \dots & a_{2,n-1}a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1}a_{m2} & a_{m1}a_{m3} & \dots & a_{m1}a_{mn} & a_{m2}a_{m3} & a_{m2}a_{m4} & \dots & a_{m2}a_{mn} & \dots & a_{m,n-1}a_{mn} \end{bmatrix}$$

and

$$D = \begin{bmatrix} b_{11}b_{12} & b_{11}b_{13} & \dots & b_{11}b_{1n} & b_{12}b_{13} & b_{12}b_{14} & \dots & b_{12}b_{1n} & \dots & b_{1,n-1}b_{1n} \\ b_{21}b_{22} & b_{21}b_{23} & \dots & b_{21}b_{2n} & b_{22}b_{23} & b_{22}b_{24} & \dots & b_{22}b_{2n} & \dots & b_{2,n-1}b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{\ell 1}b_{\ell 2} & b_{\ell 1}b_{\ell 3} & \dots & b_{\ell 1}b_{\ell n} & b_{\ell 2}b_{\ell 3} & b_{\ell 2}b_{\ell 4} & \dots & b_{\ell 2}b_{\ell n} & \dots & b_{\ell,n-1}b_{\ell n} \end{bmatrix}.$$

By (19), it is straightforward to verify that

$$[A,B]^{\mathrm{T}}[A,B] - [C,D]^{\mathrm{T}}[C,D] = \begin{bmatrix} A^{\mathrm{T}}A - C^{\mathrm{T}}C & A^{\mathrm{T}}B - C^{\mathrm{T}}D \\ B^{\mathrm{T}}A - D^{\mathrm{T}}C & B^{\mathrm{T}}B - D^{\mathrm{T}}D \end{bmatrix} = \begin{bmatrix} \frac{1}{3}E + \frac{2}{3}I & O \\ O & \frac{1}{3}I \end{bmatrix} \succ 0.$$

Thus $[A,B]^{\mathrm{T}}[A,B]$ is also positive definite, hence full-rank. Finally,

$$\rho_4(n) \ge m \ge \mathrm{rank}\,([A,B]) \ge \mathrm{rank}\left([A,B]^{\mathrm{T}}[A,B]\right) = n(n+1)/2.$$

The upper bound follows from the following identity (formula (10.35) in [17]):

$$\left(\sum_{j=1}^n x_j^2\right)^2 = \frac{1}{6}\sum_{j<k}(x_j+x_k)^4 + \frac{1}{6}\sum_{j<k}(x_j-x_k)^4 + \frac{4-n}{3}\sum_{j=1}^n x_j^4.$$

When $n \geq 5$, the coefficient $\frac{4-n}{3}$ is negative, and so it is not a valid representation of Hilbert's identity, but it is still a rank-one decomposition for $\left(\sum_{j=1}^n x_j^2\right)^2$. Since there are no more than n^2 rank one terms in this expression, it yields an upper bound of n^2 for $\rho_4(n)$. □

Remark as $\rho_4(n) \leq \tau_4(n)$, Theorem 5.3 immediately implies Theorem 5.1 when $d = 2$. The following examples show that $n(n+1)/2$ is the exact value for $\rho_4(n)$ as well as $\tau_4(n)$ when $n \leq 3$ (note that the case $n = 1$ is trivial).

Example 5.4 $(x_1^2 + x_2^2)^2 = \frac{1}{2}\left(x_1 + \frac{1}{\sqrt{3}}x_2\right)^4 + \frac{1}{2}\left(x_1 - \frac{1}{\sqrt{3}}x_2\right)^4 + \frac{8}{9}x_2^4.$

Example 5.5 $(x_1^2 + x_2^2 + x_3^2)^2 = \frac{1}{2(a^4+1)}\sum_{\beta=\pm1}\left((x_1+\beta a x_2)^4 + (x_2+\beta a x_3)^4 + (x_3+\beta a x_1)^4\right)$, *where* $a^2 = \frac{3\pm\sqrt{5}}{2}$.

We remark that the above tight representations are not unique. One may find other representations, e.g. (8.29) and (8.30) of [17], which include Examples 5.4 and 5.5 as special cases. Moreover, in light of Proposition 5.2, Liouville's identity (6) which involving 12 terms, is not tight for both $\rho_4(4)$ and $\tau_4(4)$. The following tight example for $\tau_4(4)$ only includes 11 terms.

Example 5.6 ((9.27)(i) of [17])

$$\begin{aligned}(x_1^2+x_2^2+x_3^2+x_4^2)^2 &= \frac{1}{32}(x_1+x_2+x_3+x_4)^4 + \frac{1}{192}\sum_{i=1}^4\left(3x_i - \sum_{j\neq i}x_j\right)^4 \\ &\quad + \frac{1}{192}\sum_{1\leq i<j\leq 4}\left((1+\sqrt{2})(x_i+x_j) - (1-\sqrt{2})\sum_{k\neq i,k\neq j}x_k\right)^4.\end{aligned}$$

This example along with Theorem 5.3 implies that $10 \leq \rho_4(4) \leq 11$. It remains an open problem to locate the exact value of $\rho_4(4)$. In general, finding the exact values (or a tighter upper bound) of $\tau_4(n)$ and $\rho_4(n)$, as well as finding a succinct algorithm to construct a shorter (less than $2n^4+n$) representation of the fourth moments tensor $\mathrm{sym}\,(I \otimes I)$, are interesting future research questions.

6 Matrix $q \mapsto p$ Norm Problem

In this section, we shall illustrate the power of polynomial-size representation of moments tensor by a specific example. In particular, we consider the problem of computing the so-called $q \mapsto p$ $(1 \leq p, q \leq \infty)$ norm of a matrix A, defined as follows:

$$\|A\|_{q\mapsto p} := \max_{\|x\|_q=1}\|Ax\|_p.$$

This problem can be viewed as a natural extension of several useful problems. For instance, the case $p = q = 2$ corresponds to the largest singular value of A. The case $(p, q) = (1, \infty)$ corresponds to the bilinear optimization problem in binary variables, which is related to the so-called matrix cut norm and Grothendieck's constant; see Alon and Naor [2]. In case $p = q$, the problem becomes the matrix p-norm problem, which has applications in scientific computing; cf. [9].

In terms of the computational complexity, three easy cases are well known: (1) $q = 1$ and $p \geq 1$ is a rational number; (2) $p = \infty$ and $q \geq 1$ is a rational number; (3) $p = q = 2$. Steinberg [19] showed that computing $\|A\|_{q \mapsto p}$ is NP-hard for general $1 \leq p < q \leq \infty$, and she further conjectured that the above mentioned three cases are the only exceptional easy cases where the matrix $q \mapsto p$ norm can be computed in polynomial time. Hendrickx and Olshevsky [8] made some progress along this line by figuring out the complexity status of the "diagonal" case of $p = q$. Moreover, very recently Bhaskara and Vijayaraghavan [5] proved that this problem is NP-hard to approximate to any constant factor when $2 < p \leq q$. However, the problem of determining the complexity status for the case $p > q$ still remains open. Here we shall show that the problem $\|A\|_{q \mapsto p}$ is NP-hard when $p = 4$ and $q = 2$. To this end, let us first present the following lemma.

Lemma 6.1 *Given positive integers n, i, j with $1 \leq i < j \leq n$, we can find $t\,(\leq 2n^4 + n + 2)$ vectors $a^1, a^2, \dots, a^t$ in polynomial time, such that*

$$2x_i^2 x_j^2 + (x^{\mathrm{T}} x)^2 = \sum_{k=1}^{t} \left((a^k)^{\mathrm{T}} x \right)^4 .$$

Proof. Recall in Theorem 4.1, we can find $\tau\,(\leq 2n^4)$ vectors $a^1, a^2, \dots, a^\tau \in \mathbb{R}^n$ in polynomial time, such that

$$\frac{2}{3} \sum_{\ell=1}^{n} x_\ell^4 + \sum_{\ell=1}^{\tau} \left((a^\ell)^{\mathrm{T}} x \right)^4 = (x^{\mathrm{T}} x)^2. \tag{20}$$

On the other hand, one verifies straightforwardly that for $1 \leq i \neq j \leq n$ we have

$$\frac{1}{2} \left((x_i + x_j)^4 + (x_i - x_j)^4 \right) + x_i^4 + x_j^4 + 2 \sum_{1 \leq \ell \leq n,\, \ell \neq i,j} x_\ell^4 = 6 x_i^2 x_j^2 + 2 \sum_{\ell=1}^{n} x_\ell^4. \tag{21}$$

Dividing by 3 on both sides of (21) and then summing up with $\sum_{\ell=1}^{\tau} \left((a^\ell)^{\mathrm{T}} x \right)^4$ yields

$$\begin{aligned}
& \sum_{\ell=1}^{\tau} \left((a^\ell)^{\mathrm{T}} x \right)^4 + \frac{1}{3} \left(\frac{1}{2} \left((x_i + x_j)^4 + (x_i - x_j)^4 \right) + x_i^4 + x_j^4 + 2 \sum_{1 \leq \ell \leq n,\, \ell \neq i,j} x_\ell^4 \right) \\
= \;& \sum_{\ell=1}^{\tau} \left((a^\ell)^{\mathrm{T}} x \right)^4 + 2 x_i^2 x_j^2 + \frac{2}{3} \sum_{\ell=1}^{n} x_\ell^4 \\
= \;& 2 x_i^2 x_j^2 + (x^{\mathrm{T}} x)^2,
\end{aligned}$$

where the last equality is due to (20). □

Now we are in a position to prove the main theorem of this section.

Theorem 6.2 *Computing $\|A\|_{2 \mapsto 4} = \max_{\|x\|_2 = 1} \|Ax\|_4$ is NP-hard.*

Proof. The reduction is made from computing the maximum (vertex) independence set of a graph. In particular, for a given graph $G = (V, E)$, Nesterov [15] showed that the following problem can be reduced from the maximum independence number problem:

$$\begin{array}{ll} \max & 2 \sum_{(i,j) \in E,\, i<j} x_i^2 x_j^2 \\ \text{s.t.} & \|x\|_2 = 1,\, x \in \mathbb{R}^n, \end{array}$$

hence is NP-hard. Moreover, the above is obviously equivalent to

$$\begin{array}{lll} (P) & \max & 2\sum_{(i,j)\in E,\, i<j} x_i^2 x_j^2 + |E|\cdot\|x\|_2^4 = \sum_{(i,j)\in E,\, i<j}\left(2x_i^2x_j^2 + (x^{\mathrm{T}}x)^2\right) \\ & \text{s.t.} & \|x\|_2 = 1,\, x\in\mathbb{R}^n. \end{array}$$

By Lemma 6.1, the objective in (P) can be expressed by no more than $|E|\cdot(2n^4+n+2)$ number of fourth powered linear terms, making (P) be an instance of $\|A\|_{2\mapsto 4}$ (polynomial-size). The polynomial reduction is thus complete. □

Suppose that p' and q' are the *conjugates* of p and q respectively, i.e., $\frac{1}{p}+\frac{1}{p'}=1$ and $\frac{1}{q}+\frac{1}{q'}=1$. By using the fact that $\|x\|_p = \max_{\|y\|_{p'}=1} y^{\mathrm{T}}x$, one can prove that $\|A\|_{q\mapsto p} = \|A^{\mathrm{T}}\|_{p'\mapsto q'}$. Therefore, Theorem 6.2 implies that computing $\|A\|_{\frac{4}{3}\mapsto 2}$ is also NP-hard. We remark that Theorem 6.2 was independently proved by Barak et al. [3] using a similar argument, after the initial version of this paper was submitted.

Acknowledgments

We would like to thank the anonymous referees and the associated editor for their insightful comments, which helped to significantly improve this paper from its original version. In this paper, 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, L. Babai, and A. Itai, *A Fast and Simple Randomized Algorithm for the Maximal Independent Set Problem*, Journal of Algorithms, 7, 567–583, 1986.

[2] N. Alon and A. Naor, *Approximating the Cut-Norm via Grothendieck's Inequality*, SIAM Journal on Computing, 35, 787–803, 2006.

[3] B. Barak, F. Brandao, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou, *Hypercontractivity, Sum-of-Squares Proofs, and Their Applications*, Proceedings of the 44th Annual ACM Symposium on Theory of Computing, 307–326, 2012.

[4] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics, Volume 54, American Mathematical Society, 2002.

[5] A. Bhaskara and A. Vijayaraghavan, *Approximating Matrix p-Norms.* Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms, 497–511, 2011.

[6] C. Carathéodory, *Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen*, Rendiconti del Circolo Matematico di Palermo, 32, 193–217, 1911.

[7] W. J. Ellison, *Waring's Problem*, The American Mathematical Monthly, 78, 10–36, 1971.

[8] J. M. Hendrickx and A. Olshevsky, *Matrix p-Norms Are NP-Hard to Approximate If $p \neq 1, 2, \infty$*, SIAM Journal on Matrix Analysis and Applications, 31, 2802–2812, 2010.

[9] N. J. Higham, *Estimating the Matrix p-Norm*, Numerische Mathematik, 62, 539–555, 1992.

[10] D. Hilbert, *Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl n^{ter} Potenzen (Waringsches problem)*, Mathematische Annalen, 67, 281–300, 1909.

[11] B. Jiang, Z. Li, and S. Zhang, *On Cones of Nonnegative Quartic Forms*, Technical Report, Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, 2011.

[12] A. Joffe, *On a Set of Almost Deterministic k-Wise Independent Random Variables*, Annals of Probability, 2, 161–162, 1974.

[13] H. Karloff and Y. Mansour, *On Construction of k-Wise Independent Random Variables*, Combinatorica, 17, 91–107, 1997.

[14] M. B. Nathanson, *Additive Number Thoery: The Classical Bases*, Graduate Texts in Mathematics, Volumn 164, Springer-Verlag, New York, 1996.

[15] Yu. Nesterov, *Random Walk in a Simplex and Quadratic Optimization over Convex Polytopes*, CORE Discussion Paper, UCL, Louvain-la-Neuve, Belgium, 2003.

[16] P. Pollack, *On Hilbert's Solution of Waring's Problem*, Central European Journal of Mathematics, 9, 294–301, 2011.

[17] B. Reznick, *Sums of Even Powers of Real Linear Forms*, Memoirs of the American Mathematical Society, Volume 96, Number 463, 1992.

[18] B. Reznick, *Uniform Denominators in Hilbert's Seventeenth Problem*, Mathematische Zeitschrift, 220, 75–97, 1995.

[19] D. Steinberg, *Computation of Matrix Norms with Applications to Robust Optimization*, Research Thesis, Technion – Israel University of Technology, Haifa, Israel, 2005.