Mathematical Programming manuscript No.
(will be inserted by the editor)

Deterministic Approximation Algorithms for Sphere Constrained Homogeneous Polynomial Optimization Problems

Anthony Man–Cho So

Received: date / Accepted: date

Abstract Due to their fundamental nature and numerous applications, sphere constrained polynomial optimization problems have received a lot of attention lately. In this paper, we consider three such problems: (i) maximizing a homogeneous polynomial over the sphere; (ii) maximizing a multilinear form over a Cartesian product of spheres; and (iii) maximizing a multiquadratic form over a Cartesian product of spheres. Since these problems are generally intractable, our focus is on designing polynomial–time approximation algorithms for them. By reducing the above problems to that of determining the L_2–diameters of certain convex bodies, we show that they can all be approximated to within a factor of $\Omega((\log n/n)^{d/2-1})$ *deterministically*, where n is the number of variables and d is the degree of the polynomial. This improves upon the currently best known approximation bound of $\Omega((1/n)^{d/2-1})$ in the literature. We believe that our approach will find further applications in the design of approximation algorithms for polynomial optimization problems with provable guarantees.

Keywords Polynomial Optimization · Approximation Algorithms · Diameters of Convex Bodies · Convex Programming

Mathematics Subject Classification (2000) 52A41 · 52B55 · 68W25 · 68W40 · 90C26

This research was supported by the Hong Kong Research Grants Council (RGC) General Research Fund (GRF) Projects CUHK 416908 and CUHK 419409.

Anthony Man–Cho So
Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong,
Shatin, N. T., Hong Kong
E-mail: manchoso@se.cuhk.edu.hk

1 Introduction

One of the most fundamental problems in optimization is that of maximizing (or minimizing) a multivariate polynomial over the Euclidean ball or sphere. Such a problem is certainly not foreign, as it captures the problem of finding the spectral norm of a matrix, as well as the problem of deciding the non–negativity of a homogeneous multivariate polynomial over the sphere, as special cases (the latter is closely related to the classical Hilbert's 17th problem in real algebraic geometry; see, e.g., [25, 29, 19]). More recently, sphere constrained polynomial optimization has also found applications in many different areas of study. These include numerical multilinear algebra (see, e.g., [26, 9, 10]), solid mechanics (see, e.g., [27, 6]), signal processing (see, e.g., [15, 20, 31]) and combinatorics (see, e.g., [11]), just to name a few. Consequently, there has been much research on the problem lately. From a computational complexity perspective, the problem is already NP–hard in some very simple settings, such as that of maximizing a cubic polynomial over the sphere [23] (see [13, 8, 9, 21] for related results). Many algorithms have been proposed to solve certain special cases of the problem (see, e.g., [18, 30, 28, 10]). However, almost all known provable approximation guarantees are for problems that have a quadratic objective. In view of the above discussion, it is natural to ask whether one can design efficient algorithms for a large class of sphere constrained polynomial optimization problems with provable approximation guarantees.

As a first step towards answering that question, de Klerk et al. [14] considered the problem of optimizing a fixed degree *even form* (i.e., a homogeneous polynomial with only even exponents) over the sphere and designed a polynomial–time approximation scheme (PTAS) for it. Later, Barvinok [1] showed that the problem of optimizing a certain class of multivariate polynomials over the sphere also admits a randomized PTAS. It should be noted that the results of de Klerk et al. and Barvinok do not imply each other, and they apply only when the polynomial objective function has some special structure. The problem of optimizing a more general multivariate polynomial over the Euclidean ball or sphere was not addressed until later, when Luo and Zhang [22] designed the first polynomial–time approximation algorithm for optimizing a multivariate quartic polynomial over a region defined by quadratic inequalities (which includes the Euclidean ball as a special case). Soon afterwards, Ling et al. [21] designed polynomial–time approximation algorithms for optimizing a biquadratic function over spheres. These results have set off a flurry of research activities. In particular, He et al. [8, 7] improved and extended the results in [22] by designing polynomial–time approximation algorithms for optimizing a multivariate polynomial or multilinear form over a region defined by quadratic inequalities. On another front, Nie [24] obtained bounds on the gap between the optimal value of a general polynomial optimization problem and that of a certain SDP relaxation due to Lasserre [18]. However, Nie did not discuss how to extract from the SDP relaxation a feasible solution to the original problem and quantify the loss in objective value. Thus, the bounds obtained by Nie are not approximation bounds in the usual sense.

Although the aforementioned results do shed some light on the approximability of sphere constrained polynomial optimization problems, they are not entirely satisfactory. On one hand, the approximation results developed in [14, 1] do not apply to general sphere constrained polynomial optimization problems. On the other hand, while the approximation bounds obtained in [22, 21, 8] do apply to the general problem, the best among them is of order $(1/n)^{d/2-1}$, which is not known to be tight when $d \geq 3$ (here, n is the number of variables and d is the degree of the polynomial). In fact, only fully polynomial–time approximation schemes (FPTAS) have been ruled out for general sphere constrained, fixed degree polynomial optimization problems [13]. This raises the obvious question of whether the bounds in [22, 21, 8] can be improved or not.

In this paper, we address the above issues and design polynomial–time approximation algorithms for sphere constrained homogeneous degree–d polynomial optimization problems, where $d \geq 3$ is assumed to be fixed. Our algorithms are deterministic, which should be contrasted with the randomized algorithms developed in [22, 8, 7]. Moreover, the approximation bounds we proved have the form $\Omega((\log n/n)^{d/2-1})$, which improves upon the bounds established in [22, 8]. Roughly speaking, our approach consists of two steps. First, we relate the objective value of our polynomial optimization problem to that of its multilinear relaxation. Such an idea is fairly standard; see, e.g., [12, 8]. Then, we reduce the problem of maximizing a multilinear form over spheres to that of maximizing a certain norm over the sphere, which, by standard duality arguments, is equivalent to determining the L_2–diameter of a certain convex body. The upshot of this reduction is that the latter problem can be tackled using powerful results from the algorithmic theory of convex bodies [5, 2]. We remark that a similar reduction has already appeared in a work by Khot and Naor [12], who used it to design a *randomized* polynomial–time approximation algorithm for maximizing a *trilinear* form over *binary* constraints. However, the approach we use for approximating the diameter is very different from that of Khot and Naor. Instead of adopting the random sampling approach described in [12], we follow the idea of Brieden et al. [2] and use certain polytopal approximation of the sphere to approximate the L_2–diameter. Under this approach, it is still relatively straightforward to obtain an approximation algorithm for maximizing a trilinear form over spheres. However, extending this result to higher degree multilinear forms becomes much more challenging. To achieve that goal, a natural idea is to apply our lower–degree results recursively. Specifically, given a degree–d multilinear form F and any $x^1 \in \mathbb{R}^{n_1}$, let $\hat{G}_{d-1}(x^1)$ be the value returned by our approximation algorithm when applied to the degree–$(d-1)$ multilinear optimization problem

$$G_{d-1}(x^1) = \max\left\{F(x^1, x^2, \ldots, x^d) : \|x^i\|_2 = 1 \text{ for } i = 2, 3, \ldots, d\right\}.$$

Clearly, if we can approximate the maximum value of $\hat{G}_{d-1}$ over the sphere, then we can approximate the maximum value of F over spheres. One of the main technical contributions of this paper is to show that under some mild conditions, the function $\hat{G}_{d-1}$ defines a norm on $\mathbb{R}^{n_1}$ for each $d \geq 3$ (see

Proposition 4). Consequently, the maximum value of the multilinear form F over spheres can be approximated by the maximum value of the norm $\hat{G}_{d-1}$ over the sphere, which by duality is equal to the L_2–diameter of the polar of the unit ball of $\hat{G}_{d-1}$. In particular, we can apply our argument for the base case to obtain an approximation algorithm for maximizing degree–d multilinear forms over spheres. We remark that although our approach leads to a more complicated construction than those in [12, 22, 8], it has the advantage that all the resulting algorithms are deterministic.

To further demonstrate the power of our approach, we study the problem of optimizing a multiquadratic form over spheres, which generalizes the sphere constrained biquadratic optimization problem considered in [21]. We show that our approach can be used to design a deterministic polynomial–time approximation algorithm for the problem with a relative approximation guarantee of $\Omega((\log n/n)^{d/2-1})$. It should be noted that even in the case of biquadratic optimization, our bound is sharper than that in [21].

The rest of the paper is organized as follows. In Section 2, we introduce the notation that will be used throughout the paper. Then, in Section 3, we study various sphere constrained polynomial optimization problems and develop approximation algorithms for them. Finally, we end with some closing remarks in Section 4.

2 Notation and Preliminaries

Our notation and definitions largely follow those in [16]. A *tensor* is a multidimensional array, and the *order* of a tensor is the number of dimensions. Thus, for instance, a matrix is a tensor of order two. Let $\mathcal{A} = (a_{i_1 i_2 \cdots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be a tensor of order d. We say that $\mathcal{A}$ is *non–zero* if at least one of its elements is non–zero, and is *cubical* if $n_1 = n_2 = \cdots = n_d$. A cubical tensor is said to be *super–symmetric* if every element $a_{i_1 i_2 \cdots i_d}$ is invariant under any permutation of the indices.

Now, let K and $j_1, \ldots, j_K$ be integers such that $1 \le K \le d$ and $1 \le j_1 < j_2 < \cdots < j_K \le d$. Furthermore, let $x^{j_k} \in \mathbb{R}^{n_{j_k}}$, where $k = 1, \ldots, K$, be given vectors. We use $\mathcal{A}(x^{j_1}, x^{j_2}, \ldots, x^{j_K})$ to denote the order–$(d-K)$ tensor that is obtained from the order–d tensor $\mathcal{A} = (a_{i_1 i_2 \cdots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ by "summing out" the indices $j_1, \ldots, j_K$. For instance, if $K = 2$, $j_1 = 2$ and $j_2 = 4$, then we have

$$\mathcal{A}\left(x^2, x^4\right)_{i_1 i_3 i_5 i_6 \cdots i_d} = \sum_{i_2=1}^{n_2} \sum_{i_4=1}^{n_4} a_{i_1 i_2 \cdots i_d} x^2_{i_2} x^4_{i_4}.$$

Given an order–d tensor $\mathcal{A} = (a_{i_1 i_2 \cdots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, we can associate with it a multilinear form $F_{\mathcal{A}} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_d} \to \mathbb{R}$ via

$$F_{\mathcal{A}}\left(x^1, x^2, \ldots, x^d\right) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} a_{i_1 i_2 \cdots i_d} x^1_{i_1} x^2_{i_2} \cdots x^d_{i_d}.$$

Furthermore, if $\mathcal{A}$ is super–symmetric with $n_1 = n_2 = \cdots = n_d = n$, then we can associate with it a homogeneous degree–d polynomial $f_{\mathcal{A}} : \mathbb{R}^n \to \mathbb{R}$ via

$$f_{\mathcal{A}}(x) = F_{\mathcal{A}}(x, x, \ldots, x) = \sum_{1 \le i_1, \ldots, i_d \le n} a_{i_1 i_2 \cdots i_d} x_{i_1} x_{i_2} \cdots x_{i_d}.$$

In fact, it is well known (see, e.g., [16]) that super–symmetric tensors are bijectively related to homogeneous polynomials. We shall drop the subscript $\mathcal{A}$ from $F_{\mathcal{A}}$ or $f_{\mathcal{A}}$ if the meaning is clear from the context.

Now, let $d \ge 3$ be given. Let $\mathcal{A} = (a_{i_1 i_2 \cdots i_d}) \in \mathbb{R}^{n^d}$ be an arbitrary non–zero super–symmetric tensor of order d, and let $f : \mathbb{R}^n \to \mathbb{R}$ be the corresponding homogeneous polynomial. In this paper, we are mainly interested in the following sphere constrained homogeneous polynomial optimization problem:

$$(\mathsf{HP}) \qquad \begin{array}{rl} \bar{v} = \text{maximize} & f_{\mathcal{A}}(x) \equiv \displaystyle\sum_{1 \le i_1, \ldots, i_d \le n} a_{i_1 i_2 \cdots i_d} x_{i_1} x_{i_2} \cdots x_{i_d} \\ \text{subject to} & \|x\|_2 = 1. \end{array}$$

Since problem (HP) is generally NP–hard, our goal is to design polynomial–time approximation algorithms for it.

3 Sphere Constrained Homogeneous Polynomial Optimization

To begin, let us consider the following *multilinear relaxation* of (HP):

$$(\mathsf{ML}) \qquad \begin{array}{rl} \text{maximize} & F_{\mathcal{A}}\left(x^1, x^2, \ldots, x^d\right) \equiv \displaystyle\sum_{1 \le i_1, \ldots, i_d \le n} a_{i_1 i_2 \cdots i_d} x^1_{i_1} x^2_{i_2} \cdots x^d_{i_d} \\ \text{subject to} & \|x^i\|_2 = 1 \quad \text{for } i = 1, \ldots, d. \end{array}$$

As observed in [8], one can relate the objective value of (HP) to that of (ML) via the so–called polarization formula (see, e.g., [17]). Specifically, we have the following proposition, whose proof can be found in [8, Lemma 3.5]:

Proposition 1 *Let $x^1, x^2, \ldots, x^d \in \mathbb{R}^n$ be arbitrary, and let $\xi_1, \xi_2, \ldots, \xi_d$ be i.i.d. Bernoulli random variables (i.e.,* $\Pr(\xi_i = 1) = \Pr(\xi_i = -1) = 1/2$ *for $i = 1, \ldots, d$). Then, we have*

$$\mathbb{E}\left[\left(\prod_{i=1}^{d} \xi_i\right) f_{\mathcal{A}}\left(\sum_{j=1}^{d} \xi_j x^j\right)\right] = d! F_{\mathcal{A}}\left(x^1, x^2, \ldots, x^d\right).$$

Armed with Proposition 1, it can be shown that the problem of approximating (HP) reduces to that of approximating (ML). Specifically, we have the following

Theorem 1 *(He et al. [8]) Suppose there is a polynomial–time algorithm $\mathscr{A}_{ML}$ that, given any instance of* (ML)*, returns a feasible solution whose objective value is at least $\alpha \in (0,1]$ times the optimal value of* (ML) *(note that the optimal value of* (ML) *is always non–negative). Then, there is a polynomial–time algorithm $\mathscr{A}_{HP}$ that, given any instance of* (HP)*, returns a solution $\hat{x} \in \mathbb{R}^n$ with $\|\hat{x}\|_2 = 1$ and*

$$\begin{aligned} f_{\mathcal{A}}(\hat{x}) &\geq \alpha \cdot d! \cdot d^{-d} \cdot \bar{v} && \text{for odd } d \geq 3, && (1) \\ f_{\mathcal{A}}(\hat{x}) - \underline{v} &\geq \alpha \cdot d! \cdot d^{-d} \cdot (\bar{v} - \underline{v}) && \text{for even } d \geq 4, && (2) \end{aligned}$$

where $\bar{v} = \max_{\|x\|_2=1} f_{\mathcal{A}}(x)$ and $\underline{v} = \min_{\|x\|_2=1} f_{\mathcal{A}}(x)$. Furthermore, if $\mathscr{A}_{ML}$ is deterministic, then so is $\mathscr{A}_{HP}$. The factor α in (1) (resp. (2)) is known as the approximation ratio *(resp.* relative approximation ratio*).*

Thus, in the sequel, we may just focus on the problem (ML).

3.1 Sphere Constrained Multilinear Optimization

Let us first investigate the case where $d = 3$. Specifically, let $\mathcal{A} = (a_{ijk}) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an arbitrary non–zero order–3 tensor, and assume without loss of generality that $1 \leq n_1 \leq n_2 \leq n_3$. Consider the following optimization problem:

$$\begin{aligned} \bar{v}_{ml}(\mathcal{A}, 3) = \text{ maximize } \quad & \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} a_{ijk} x_i^1 x_j^2 x_k^3 \\ \text{subject to } \quad & \|x^1\|_2 = \|x^2\|_2 = \|x^3\|_2 = 1, \\ & x^1 \in \mathbb{R}^{n_1}, x^2 \in \mathbb{R}^{n_2}, x^3 \in \mathbb{R}^{n_3}. \end{aligned} \qquad (3)$$

As shown in [8], problem (3) is still NP–hard. However, we can approximate it in polynomial–time using an argument similar to that in Khot and Naor [12]. To begin, let

$$\left\| \mathcal{A}(x^1) \right\|_2 = \max_{u \in \mathbb{R}^{n_3} : \|u\|_2 \leq 1} \left\| \mathcal{A}(x^1) u \right\|_2 = \max_{\|u\|_2 \leq 1, \|v\|_2 \leq 1} v^T \mathcal{A}(x^1) u$$

be the largest singular value of the $n_2 \times n_3$ matrix $\mathcal{A}(x^1)$. Clearly, we have

$$\bar{v}_{ml}(\mathcal{A}, 3) = \max_{x^1 \in \mathbb{R}^{n_1} : \|x^1\|_2 \leq 1} \left\| \mathcal{A}(x^1) \right\|_2.$$

Observe that the function $x \mapsto \|x\|_{\mathcal{A}} \equiv \|\mathcal{A}(x)\|_2$ defines a semi–norm on $\mathbb{R}^{n_1}$. However, in the context of problem (3), we may assume without loss that it defines a *norm* on $\mathbb{R}^{n_1}$. To see this, consider the set $\mathcal{L} = \{x \in \mathbb{R}^{n_1} : \|x\|_{\mathcal{A}} = 0\}$. Note that $\mathcal{L}$ is the nullspace of the $(n_2 \times n_3) \times n_1$ matrix A, where

$$A_{(j,k),i} = a_{ijk} \quad \text{for } i = 1, \ldots, n_1; j = 1, \ldots, n_2; k = 1, \ldots, n_3. \qquad (4)$$

As such, it is a linear subspace in $\mathbb{R}^{n_1}$. Moreover, since $\mathcal{A}$ is non–zero, we have

$$\bar{v}_{ml}(\mathcal{A},3) \geq \max_{1\leq i\leq n_1, 1\leq j\leq n_2, 1\leq k\leq n_3} |a_{ijk}| > 0. \tag{5}$$

This implies that any optimal solution $(\bar{x}^1,\bar{x}^2,\bar{x}^3) \in \mathbb{R}^{n_1\times n_2\times n_3}$ to (3) must satisfy $\bar{x}^1 \in \mathcal{L}^\perp$, where $\mathcal{L}^\perp$ is the orthogonal complement of $\mathcal{L}$ in $\mathbb{R}^{n_1}$. Indeed, suppose that $\bar{x}^1 = y+z$, where $y \in \mathcal{L}^\perp$, $z \in \mathcal{L}$ and $y,z \neq \mathbf{0}$. Since $\bar{x}^1$ is feasible for (3), we have $\|y\|_2^2 + \|z\|_2^2 = 1$ by the Pythagoras theorem, which implies that $\|y\|_2 < 1$. Moreover, we have

$$0 < \bar{v}_{ml}(\mathcal{A},3) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} a_{ijk}\bar{x}_i^1\bar{x}_j^2\bar{x}_k^3 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} a_{ijk}y_i\bar{x}_j^2\bar{x}_k^3.$$

Now, consider the vector $\bar{y} = y/\|y\|_2 \in \mathbb{R}^{n_1}$. By construction, $(\bar{y},\bar{x}^2,\bar{x}^3)$ is a feasible solution to (3), and

$$\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} a_{ijk}\bar{y}_i\bar{x}_j^2\bar{x}_k^3 = \frac{1}{\|y\|_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} a_{ijk}y_i\bar{x}_j^2\bar{x}_k^3 = \frac{\bar{v}_{ml}(\mathcal{A},3)}{\|y\|_2} > \bar{v}_{ml}(\mathcal{A},3),$$

which contradicts the definition of $\bar{v}_{ml}(\mathcal{A},3)$. Hence, we have $\bar{x}^1 \in \mathcal{L}^\perp$, as desired. In particular, if $\mathcal{L} \neq \{\mathbf{0}\}$, then we can reduce the dimension of $\mathcal{A}$ and obtain a problem that is equivalent to (3).

Now, consider the unit ball of the norm $\|\cdot\|_{\mathcal{A}}$, which is given by

$$B_{\mathcal{A}} = \left\{x \in \mathbb{R}^{n_1} : \|x\|_{\mathcal{A}} \leq 1\right\}.$$

The set $B_{\mathcal{A}}$ is clearly centrally symmetric and convex. The following proposition shows that it is also bounded and has a non–empty interior, i.e., it is a centrally symmetric convex body. Consequently, we can use the ellipsoid algorithm to answer various queries about $B_{\mathcal{A}}$ and its polar efficiently. This will be crucial to the development of our approximation algorithms. In the sequel, we use $B_2^n(r)$ to denote the n–dimensional Euclidean ball centered at the origin with radius $r > 0$.

Proposition 2 *There exist rational numbers $0 < r \leq R < \infty$, whose encoding lengths are polynomially bounded by the input size of problem (3), such that $B_2^{n_1}(r) \subset B_{\mathcal{A}} \subset B_2^{n_1}(R)$.*

Proof Let $e^i \in \mathbb{R}^{n_1}$ be the i–th basis vector in $\mathbb{R}^{n_1}$, where $i = 1,\ldots,n_1$. Observe that for $i = 1,\ldots,n_1$, we have

$$0 < \|\mathcal{A}(e^i)\|_2 \leq r_i \equiv \sum_{j=1}^{n_2}\sum_{k=1}^{n_3} |a_{ijk}|.$$

Hence, we conclude that

$$\frac{1}{r'}\cdot \mathrm{conv}\left(\left\{\pm e^1,\ldots,\pm e^{n_1}\right\}\right) = \left\{x \in \mathbb{R}^{n_1} : \|x\|_1 \leq \frac{1}{r'}\right\} \subset B_{\mathcal{A}},$$

where $r' = \max_{1 \le i \le n_1} r_i$ and $\mathrm{conv}(S)$ is the convex hull of the points in $S \subset \mathbb{R}^{n_1}$. In particular, since $\|x\|_1 \le \sqrt{n_1} \cdot \|x\|_2$ for any $x \in \mathbb{R}^{n_1}$, we have $B_2^{n_1}(r) \subset B_{\mathcal{A}}$, where $r = 1/(\lceil \sqrt{n_1} \rceil \cdot r') > 0$ is a rational number, whose encoding length is polynomially bounded by the input size of problem (3).

On the other hand, we have

$$\|x\|_{\mathcal{A}} \ge \max_{1 \le j \le n_2, 1 \le k \le n_3} \left| \sum_{i=1}^{n_1} a_{ijk} x_i \right| = \|Ax\|_\infty \ge \sqrt{\frac{\lambda_{min}(A^T A)}{n_2 n_3}} \cdot \|x\|_2,$$

where A is the $(n_2 \times n_3) \times n_1$ matrix given by (4), and $\lambda_{min}(A^T A)$ is the smallest eigenvalue of $A^T A$. Since $\mathcal{L} = \{\mathbf{0}\}$ and $1 \le n_1 \le n_2 \le n_3$ by assumption, the matrix A has full column rank. In particular, we have $\lambda_{min}(A^T A) > 0$. Thus, it follows that $B_{\mathcal{A}} \subset B_2^{n_1}(R)$, where $R = \left\lceil \sqrt{n_2 n_3 / \lambda_{min}(A^T A)} \right\rceil$ is a finite rational number. Moreover, the encoding length of R can be polynomially bounded by the input size of problem (3); see [5]. This completes the proof. $\square$

By Proposition 2, we conclude that the polar of $B_{\mathcal{A}}$, which is given by

$$B_{\mathcal{A}}^\circ = \left\{ y \in \mathbb{R}^{n_1} : x^T y \le 1 \text{ for all } x \in B_{\mathcal{A}} \right\},$$

is also a centrally symmetric convex body with $B_2^{n_1}(1/R) \subset B_{\mathcal{A}}^\circ \subset B_2^{n_1}(1/r)$. Moreover, we have

$$\begin{aligned} \bar{v}_{ml}(\mathcal{A}, 3) &= \max_{x \in \mathbb{R}^{n_1} : \|x\|_2 \le 1} \|x\|_{\mathcal{A}} = \max_{x \in \mathbb{R}^{n_1} : \|x\|_2 \le 1} \left(\max_{y \in B_{\mathcal{A}}^\circ} x^T y \right) \\ &= \max_{y \in B_{\mathcal{A}}^\circ} \left(\max_{x \in \mathbb{R}^{n_1} : \|x\|_2 \le 1} x^T y \right) = \max_{y \in B_{\mathcal{A}}^\circ} \|y\|_2 \\ &= \frac{1}{2} \mathrm{diam}_2(B_{\mathcal{A}}^\circ), \end{aligned} \tag{6}$$

where $\mathrm{diam}_2(B_{\mathcal{A}}^\circ)$ is the L_2–diameter[1] of $B_{\mathcal{A}}^\circ$. Thus, we have reduced the problem of approximating the optimal value of problem (3) in polynomial time to that of approximating $\mathrm{diam}_2(B_{\mathcal{A}}^\circ)$ in polynomial time. The latter can be achieved using the results in [5,2] if the so–called weak membership problem associated with $B_{\mathcal{A}}^\circ$ can be solved in polynomial time. Before we formally state our results, let us recall some definitions from the algorithmic theory of convex bodies (see [5] for further details).

Let K be a centrally symmetric convex body in $\mathbb{R}^n$. For any $\epsilon \ge 0$, the *outer parallel body* and *inner parallel body* of K are given by

$$K(\epsilon) = K + B_2^n(\epsilon) \quad \text{and} \quad K(-\epsilon) = \{x \in \mathbb{R}^n : x + B_2^n(\epsilon) \subset K\},$$

respectively. We say that K is *well–bounded* if there exist rational numbers $0 < r \le R < \infty$ such that $B_2^n(r) \subset K \subset B_2^n(R)$. The *weak membership problem* associated with K is defined as follows:

[1] Recall that the L_2–diameter of a set $S \subset \mathbb{R}^n$ is given by $\mathrm{diam}_2(S) = \sup_{x,y \in S} \|x - y\|_2$.

Weak Membership Problem. Given a vector $y \in \mathbb{Q}^n$ and a rational number $\epsilon > 0$, either (i) assert that $y \in K(\epsilon)$, or (ii) assert that $y \notin K(-\epsilon)$.

A *weak membership oracle* for K is a black box that solves the weak membership problem associated with K. By combining the results in [5] and [2], we have the following theorem:

Theorem 2 *(cf. Brieden et al. [2]) Given an integer $n \geq 1$, one can construct in deterministic polynomial time a centrally symmetric polytope P in $\mathbb{R}^n$ such that (i) $B_2^n(1) \subset P \subset B_2^n(O(\sqrt{n/\log n}))$, and (ii) for any well–bounded centrally symmetric convex body K in $\mathbb{R}^n$, one has*

$$\Omega\left(\sqrt{\frac{\log n}{n}}\right) \cdot diam_2(K) \leq diam_P(K) \leq diam_2(K),$$

where $diam_P(K)$ is the diameter of K with respect to the polytopal norm $\|\cdot\|_P$ induced by P (i.e., for any $x \in \mathbb{R}^n$, one has $\|x\|_P = \min\{\lambda \geq 0 : x \in \lambda P\}$, and P is the unit ball of the induced norm). Moreover, if K is equipped with a weak membership oracle, then for any given rational number $\epsilon > 0$, the quantity $diam_P(K)$ can be computed to an accuracy of ϵ in deterministic oracle–polynomial time[2], and a vector $x \in K(\epsilon)$ is delivered with $\|x\|_P \geq (1/2) \cdot diam_P(K) - \epsilon$.

Remarks.

1. Theorem 2 implies that if the weak membership oracle associated with K can carry out its computation in deterministic polynomial time, then the diameter of K can be approximated to arbitrary accuracy in deterministic polynomial time.
2. The constant behind the O–notation depends on the number of facets in P. Naturally, the more facets in P, the more accurate it can approximate $B_2^n(1)$, and hence the smaller the constant behind the O–notation. Following the derivation in [2], a rough calculation shows that when P has at most n^{c+1} facets (where $c \geq 2$ is a constant), the constant behind the O–notation is bounded above by $\sqrt{8 + 3/c} \leq 5/2$.

Let us now return to the problem of approximating $\mathrm{diam}_2(B_{\mathcal{A}}^\circ)$ in polynomial time. By Proposition 2, both $B_{\mathcal{A}}$ and $B_{\mathcal{A}}^\circ$ are well–bounded. Thus, in view of Theorem 2, it suffices to prove the following

Proposition 3 *The weak membership problem associated with $B_{\mathcal{A}}^\circ$ can be solved in deterministic polynomial time.*

Proof By the well–boundedness of $B_{\mathcal{A}}$ and the results in [5, Chapter 4], it suffices to show that the weak membership problem associated with $B_{\mathcal{A}}$ can be solved in deterministic polynomial time. This can be achieved as follows. First, given a vector $y \in \mathbb{Q}^{n_1} \backslash \{\mathbf{0}\}$ and a rational number $\epsilon > 0$, we can

[2] An algorithm has oracle–polynomial time complexity if its runtime is polynomial in both the input size and the number of calls to the oracle [5].

compute in deterministic polynomial time a number $\beta(y)$ such that $\beta(y) \le \|y\|_{\mathcal{A}} \le \beta(y) + \epsilon/\|y\|_2$ [5,3]. Now, suppose that $\beta(y) \le 1$. Then, we have $\|y\|_{\mathcal{A}} \le 1 + \epsilon/\|y\|_2$. We claim that $y \in B_{\mathcal{A}}(\epsilon)$. Indeed, let $\delta \in \mathbb{R}$ be such that

$$0 < \frac{\epsilon\|y\|_2}{\|y\|_2 + \epsilon} \le \delta < \min\{\epsilon, \|y\|_2\},$$

and consider the point $u = (1 - \delta/\|y\|_2)y$. We have

$$\|u\|_{\mathcal{A}} \le \left(1 - \frac{\delta}{\|y\|_2}\right)\left(1 + \frac{\epsilon}{\|y\|_2}\right) \le \left(1 - \frac{\epsilon}{\|y\|_2 + \epsilon}\right)\left(1 + \frac{\epsilon}{\|y\|_2}\right) = 1.$$

This shows that $u \in B_{\mathcal{A}}$, whence $y = u + \delta y/\|y\|_2 \in B_{\mathcal{A}}(\epsilon)$, as desired.

On the other hand, if $\beta(y) > 1$, then we have $\|y\|_{\mathcal{A}} > 1$, which trivially implies that $y \notin B_{\mathcal{A}}(-\epsilon)$. This completes the proof. □

From equation (6), Theorem 2 and Proposition 3, we see that the optimal *value* of problem (3) can be approximated to within a factor of $\Omega(\sqrt{\log n_1/n_1})$ in deterministic polynomial time. We now show how to construct a *feasible solution* $(\bar{x}^1, \bar{x}^2, \bar{x}^3) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ to problem (3) that attains the stated approximation ratio. To begin, let $\epsilon > 0$ be such that

$$\epsilon < \frac{\gamma}{4} \cdot \sqrt{\frac{\log n_1}{n_1}} \cdot \max_{1 \le i \le n_1, 1 \le j \le n_2, 1 \le k \le n_3} |a_{ijk}|,$$

where $\gamma \ge 2/5$ is the constant behind the Ω–notation in Theorem 2 (see the remarks after Theorem 2). Note that $\epsilon < (\gamma/4) \cdot \sqrt{\log n_1/n_1} \cdot \bar{v}_{ml}(\mathcal{A}, 3)$ by (5). Using the algorithm stated in Theorem 2 and Proposition 3, we compute a vector $\bar{y} \in B^\circ_{\mathcal{A}}(\epsilon)$ in deterministic polynomial time such that $\|\bar{y}\|_P \ge (1/2) \cdot \mathrm{diam}_P(B^\circ_{\mathcal{A}}) - \epsilon$. Furthermore, let $\bar{y}' \in B^\circ_{\mathcal{A}}$ be such that $\|\bar{y}'\|_2 = (1/2) \cdot \mathrm{diam}_2(B^\circ_{\mathcal{A}})$. We remark that such a $\bar{y}'$ exists, since $B^\circ_{\mathcal{A}}$ is centrally symmetric and compact. Now, by definition, we have

$$\|\bar{y}'\|_P \le \frac{1}{2}\mathrm{diam}_P(B^\circ_{\mathcal{A}}) \le \|\bar{y}\|_P + \epsilon. \tag{7}$$

Moreover, Theorem 2 implies that

$$\|x\|_P \le \|x\|_2 \le \frac{1}{\gamma} \cdot \sqrt{\frac{n_1}{\log n_1}} \cdot \|x\|_P \quad \text{for any } x \in \mathbb{R}^{n_1}. \tag{8}$$

Hence, we conclude that

$$\|\bar{y}\|_2 \ge \|\bar{y}'\|_P - \epsilon \ge \gamma \cdot \sqrt{\frac{\log n_1}{n_1}} \cdot \|\bar{y}'\|_2 - \epsilon \tag{9}$$

$$= \gamma \cdot \sqrt{\frac{\log n_1}{n_1}} \cdot \bar{v}_{ml}(\mathcal{A}, 3) - \epsilon, \tag{10}$$

where (9) follows from (7) and (8), and (10) follows from the definition of $\bar{y}'$ and (6). Upon setting

$$\bar{x}^1 = \frac{\bar{y}}{\|\bar{y}\|_2}, \quad (\bar{x}^2, \bar{x}^3) = \arg \max_{\|x^2\|_2 \le 1, \|x^3\|_2 \le 1} (x^2)^T \mathcal{A}(\bar{x}^1) x^3,$$

we see that $(\bar{x}^1, \bar{x}^2, \bar{x}^3) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is feasible for problem (3). Moreover, by (10) and our choice of ϵ, we have $\epsilon < \|\bar{y}\|_2$, which implies that $\bar{y} - \epsilon \bar{y}/\|\bar{y}\|_2 \in B^\circ_{\mathcal{A}}$. Hence, the solution $(\bar{x}^1, \bar{x}^2, \bar{x}^3)$ satisfies

$$\begin{aligned}\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} a_{ijk} \bar{x}^1_i \bar{x}^2_j \bar{x}^3_k = \|\bar{x}^1\|_{\mathcal{A}} &= \max_{y \in B^\circ_{\mathcal{A}}} (\bar{x}^1)^T y \\ &\ge (\bar{x}^1)^T \bar{y} - \epsilon \\ &\ge \frac{\gamma}{2} \cdot \sqrt{\frac{\log n_1}{n_1}} \cdot \bar{v}_{ml}(\mathcal{A}, 3),\end{aligned}$$

where the last inequality follows from (10) and the definition of ϵ. In other words, the objective value of the solution $(\bar{x}^1, \bar{x}^2, \bar{x}^3)$ is at least $\Omega(\sqrt{\log n_1/n_1})$ times the optimum. To summarize, we have proven the following

Theorem 3 *There is a deterministic polynomial–time approximation algorithm for problem (3) with approximation ratio $\Omega(\sqrt{\log n_1/n_1})$. In particular, there is a deterministic polynomial–time approximation algorithm for* (HP) *when $d = 3$, with approximation ratio $\Omega(\sqrt{\log n/n})$.*

To obtain a polynomial–time approximation algorithm for (ML) (and hence for (HP)) when $d \ge 4$, a natural idea is to apply the above argument recursively. Specifically, given a degree–d multilinear form F and any $x^1 \in \mathbb{R}^{n_1}$, let $\hat{G}_{d-1}(x^1)$ be the value returned by our approximation algorithm when applied to the degree–$(d-1)$ multilinear optimization problem

$$G_{d-1}(x^1) = \max \left\{ F(x^1, x^2, \ldots, x^d) : \|x^i\|_2 = 1 \text{ for } i = 2, 3, \ldots, d \right\}.$$

If $\hat{G}_{d-1}$ defines a norm on $\mathbb{R}^{n_1}$, then we can repeat our earlier argument and obtain an approximation algorithm for the problem $\max_{\|x^1\|_2=1} \hat{G}_{d-1}(x^1)$. This would then yield an approximation algorithm for the original problem

$$\max_{\|x^1\|_2=1} G_{d-1}(x^1) = \max \left\{ F(x^1, x^2, \ldots, x^d) : \|x^i\|_2 = 1 \text{ for } i = 1, 2, \ldots, d \right\}.$$

To implement the above idea, we need the following proposition, which forms the heart of our construction.

Proposition 4 *Let $d \ge 3$ be given, and let $\mathcal{A} = \{a_{i_1 i_2 \cdots i_d}\} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be an arbitrary non–zero order–d tensor. Let $\|\cdot\|_{\nu_i} : \mathbb{R}^{n_{i+1}} \to \mathbb{R}_+$ be an* *arbitrary norm on $\mathbb{R}^{n_i}$, where $i = 1, \ldots, d-3$. Define the functions $\{\Lambda_i^{\mathcal{A},d}\}_{i=1}^{d-2}$ inductively as follows:*

$$\Lambda_{d-2}^{\mathcal{A},d}\left(x^1, x^2, \ldots, x^{d-2}\right) = \left\|\mathcal{A}\left(x^1, x^2, \ldots, x^{d-2}\right)\right\|_2,$$

$$\Lambda_i^{\mathcal{A},d}\left(x^1, x^2, \ldots, x^i\right) = diam_{\nu_i}\left[\left\{y \in \mathbb{R}^{n_{i+1}} : \Lambda_{i+1}^{\mathcal{A},d}\left(x^1, x^2, \ldots, x^i, y\right) \le 1\right\}^\circ\right]$$

for $i = d-3, d-4, \ldots, 1$. Then, the following hold:

(a) For $j = 1, 2, \ldots, d-2$ and for any $\bar{x}^1, \ldots, \bar{x}^{k-1}, \bar{x}^{k+1}, \ldots, \bar{x}^j$, where $\bar{x}^i \in \mathbb{R}^{n_i}$, the function $\bar{\Lambda}_{j,k}^{\mathcal{A},d} : \mathbb{R}^{n_k} \to \mathbb{R}_+$ given by

$$\bar{\Lambda}_{j,k}^{\mathcal{A},d}(x) = \Lambda_j^{\mathcal{A},d}\left(\bar{x}^1, \ldots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \ldots, \bar{x}^j\right)$$

is a semi–norm on $\mathbb{R}^{n_k}$ for any $k \in \{1, \ldots, j\}$.

(b) Suppose that the $(n_2 \times \cdots \times n_d) \times n_1$ matrix A, where

$$A_{(i_2,\ldots,i_d),i_1} = a_{i_1 i_2 \cdots i_d}, \tag{11}$$

has full column rank. Then, the function $\Lambda_1^{\mathcal{A},d}$ defines a norm on $\mathbb{R}^{n_1}$.

Proof Let $d \ge 3$ be given. We prove (a) by backward induction on j. The base case (i.e., $j = d-2$) is clear, since we have

$$\bar{\Lambda}_{d-2,k}^{\mathcal{A},d}(x) = \left\|\mathcal{A}\left(\bar{x}^1, \ldots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \ldots, \bar{x}^{d-2}\right)\right\|_2$$

for $k = 1, \ldots, d-2$. Now, suppose that $j < d-2$. Let $k \in \{1, \ldots, j\}$ and $\bar{x}^1, \ldots, \bar{x}^{k-1}, \bar{x}^{k+1}, \ldots, \bar{x}^j$ be arbitrary, where $\bar{x}^i \in \mathbb{R}^{n_i}$. Consider

$$\begin{aligned}\bar{\Lambda}_{j,k}^{\mathcal{A},d}(x) &= \Lambda_j^{\mathcal{A},d}\left(\bar{x}^1, \ldots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \ldots, \bar{x}^j\right) \\ &= \mathrm{diam}_{\nu_j}\left[\left\{y \in \mathbb{R}^{n_{j+1}} : \Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \ldots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \ldots, \bar{x}^j, y\right) \le 1\right\}^\circ\right].\end{aligned}$$

By the inductive hypothesis, for any given $x \in \mathbb{R}^{n_k}$, the function

$$y \mapsto \Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \ldots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \ldots, \bar{x}^j, y\right)$$

defines a semi–norm on $\mathbb{R}^{n_{j+1}}$. Hence, the set

$$\bar{B}_{j,k}(x) = \left\{y \in \mathbb{R}^{n_{j+1}} : \Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \ldots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \ldots, \bar{x}^j, y\right) \le 1\right\}$$

is centrally symmetric and convex. In particular, we have

$$\bar{\Lambda}_{j,k}^{\mathcal{A},d}(x) = \frac{2}{r_{\nu_j}(\bar{B}_{j,k}(x))},$$

where $r_{\nu_j}(\bar{B}_{j,k}(x))$ is the radius of the largest ball $\mathscr{B}$ with respect to the norm $\|\cdot\|_{\nu_j}$ that is contained in $\bar{B}_{j,k}(x)$ (see [4]). We may assume that the center of $\mathscr{B}$ coincides with the origin, since $\bar{B}_{j,k}(x)$ is centrally symmetric. We now claim that $x \mapsto \bar{\Lambda}_{j,k}^{\mathcal{A},d}(x)$ defines a semi–norm on $\mathbb{R}^{n_k}$. Observe that the only non–trivial task is to establish the convexity of $\bar{\Lambda}_{j,k}^{\mathcal{A},d}$. Towards that end, let $x, x' \in \mathbb{R}^{n_k}$ be arbitrary, and set

$$\bar{x} = \frac{x + x'}{2}, \quad r = r_{\nu_j}(\bar{B}_{j,k}(x)), \quad r' = r_{\nu_j}(\bar{B}_{j,k}(x')).$$

Consider the following cases:

Case 1: $r = r' = \infty$.

Then, we have

$$\begin{aligned} &\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \dots, \bar{x}^j, y\right) \\ = &\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, x', \bar{x}^{k+1}, \dots, \bar{x}^j, y\right) = 0 \end{aligned}$$

for *any* $y \in \mathbb{R}^{n_{j+1}}$. By the inductive hypothesis, the function

$$u \mapsto \Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, u, \bar{x}^{k+1}, \dots, \bar{x}^j, y\right)$$

defines a semi–norm on $\mathbb{R}^{n_k}$. Hence, we have

$$\begin{aligned} 0 \le\ &\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, \bar{x}, \bar{x}^{k+1}, \dots, \bar{x}^j, y\right) \\ \le\ &\frac{1}{2}\Big[\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \dots, \bar{x}^j, y\right) \\ &\quad + \Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, x', \bar{x}^{k+1}, \dots, \bar{x}^j, y\right)\Big] = 0, \end{aligned}$$

or equivalently,

$$\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, \bar{x}, \bar{x}^{k+1}, \dots, \bar{x}^j, y\right) = 0. \tag{12}$$

Since (12) holds for all $y \in \mathbb{R}^{n_{j+1}}$, we conclude that $\bar{\Lambda}_{j,k}^{\mathcal{A},d}(\bar{x}) = 0$, as required.

Case 2: $r' \ge r > 0$, with $r < \infty$.

Let $\bar{y} \in \bar{B}_{j,k}(x)$ be such that $\|\bar{y}\|_{\nu_j} = r$. By the inductive hypothesis, we have

$$\begin{aligned} &\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, \bar{x}, \bar{x}^{k+1}, \dots, \bar{x}^j, \bar{y}\right) \\ \le\ &\frac{1}{2}\Big[\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \dots, \bar{x}^j, \bar{y}\right) \\ &\quad + \Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, x', \bar{x}^{k+1}, \dots, \bar{x}^j, \bar{y}\right)\Big]. \end{aligned} \tag{13}$$

Since $\bar{y} \in \bar{B}_{j,k}(x)$, it follows from the definition of $\bar{B}_{j,k}(x)$ that

$$\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1, \dots, \bar{x}^{k-1}, x, \bar{x}^{k+1}, \dots, \bar{x}^j, \bar{y}\right) \le 1. \tag{14}$$

Consider now the following sub–cases:

Case 2a: $r' = \infty$.

Then, we have $\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1,\ldots,\bar{x}^{k-1},x',\bar{x}^{k+1},\ldots,\bar{x}^j,\bar{y}\right)=0$. It follows from (13) and (14) that

$$\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1,\ldots,\bar{x}^{k-1},\bar{x},\bar{x}^{k+1},\ldots,\bar{x}^j,\bar{y}\right)\le\frac{1}{2}.$$

In particular, we have $2\bar{y}\in\bar{B}_{j,k}(\bar{x})$. Upon noting that $\|2\bar{y}\|_{\nu_j}=2r$, we obtain $r_{\nu_j}(\bar{B}_{j,k}(\bar{x}))\ge 2r$, which implies that

$$\bar{\Lambda}_{j,k}^{\mathcal{A},d}(\bar{x})=\frac{2}{r_{\nu_j}(\bar{B}_{j,k}(\bar{x}))}\le\frac{1}{r}=\frac{1}{2}\left[\bar{\Lambda}_{j,k}^{\mathcal{A},d}(x)+\bar{\Lambda}_{j,k}^{\mathcal{A},d}(x')\right],$$

as required.

<u>**Case 2b**</u>: $r'<\infty$.

Since $r_{\nu_j}(\bar{B}_{j,k}(x'))=r'$, we have $r^{-1}r'\bar{y}\in\bar{B}_{j,k}(x')$, or equivalently,

$$\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1,\ldots,\bar{x}^{k-1},x',\bar{x}^{k+1},\ldots,\bar{x}^j,\bar{y}\right)\le\frac{r}{r'}.\tag{15}$$

It then follows from (13), (14) and (15) that

$$\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1,\ldots,\bar{x}^{k-1},\bar{x},\bar{x}^{k+1},\ldots,\bar{x}^j,\bar{y}\right)\le\frac{1}{2}\left(1+\frac{r}{r'}\right)=\frac{r+r'}{2r'}.$$

Now, set $y'=2(r+r')^{-1}r'\bar{y}\in\mathbb{R}^{n_{j+1}}$. Then, we have

$$\Lambda_{j+1}^{\mathcal{A},d}\left(\bar{x}^1,\ldots,\bar{x}^{k-1},\bar{x},\bar{x}^{k+1},\ldots,\bar{x}^j,y'\right)\le 1,$$

i.e., $y'\in\bar{B}_{j,k}(\bar{x})$. Moreover, we have $\|y'\|_{\nu_j}=2rr'/(r+r')$, whence

$$r_{\nu_j}(\bar{B}_{j,k}(\bar{x}))\ge\frac{2rr'}{r+r'}.$$

However, this implies that

$$\begin{aligned}\bar{\Lambda}_{j,k}^{\mathcal{A},d}(\bar{x})&=\frac{2}{r_{\nu_j}(\bar{B}_{j,k}(\bar{x}))}\le\frac{1}{2}\cdot\frac{2(r+r')}{rr'}\\&=\frac{1}{2}\left(\frac{2}{r}+\frac{2}{r'}\right)=\frac{1}{2}\left[\bar{\Lambda}_{j,k}^{\mathcal{A},d}(x)+\bar{\Lambda}_{j,k}^{\mathcal{A},d}(x')\right],\end{aligned}$$

which completes the proof of (a).

We now proceed to prove (b). Note that $\Lambda_1^{\mathcal{A},d}(x)=\bar{\Lambda}_{1,1}^{\mathcal{A},d}(x)$ for all $x\in\mathbb{R}^{n_1}$. Thus, in view of (a), it suffices to show that $\Lambda_1^{\mathcal{A},d}(x)>0$ for all $x\in\mathbb{R}^{n_1}\backslash\{\mathbf{0}\}$. Suppose that this is not the case, i.e., there exists an $\bar{x}\in\mathbb{R}^{n_1}\backslash\{\mathbf{0}\}$ such that $\Lambda_1^{\mathcal{A},d}(\bar{x})=0$. Since

$$\Lambda_1^{\mathcal{A},d}(\bar{x})=\mathrm{diam}_{\nu_1}\left[\left\{x^2\in\mathbb{R}^{n_2}:\Lambda_2^{\mathcal{A},d}\left(\bar{x},x^2\right)\le 1\right\}^\circ\right]$$

and $x^2 \mapsto \Lambda_2^{\mathcal{A},d}(\bar{x}, x^2)$ defines a semi–norm on $\mathbb{R}^{n_2}$, it follows that $\Lambda_2^{\mathcal{A},d}(\bar{x}, x^2) = 0$ for all $x^2 \in \mathbb{R}^{n_2}$. By using the definitions of $\Lambda_2^{\mathcal{A},d}, \ldots, \Lambda_{d-2}^{\mathcal{A},d}$ and iterating, we see that

$$\Lambda_{d-2}^{\mathcal{A},d}\left(\bar{x}, x^2, \ldots, x^{d-2}\right) = \left\|\mathcal{A}\left(\bar{x}, x^2, \ldots, x^{d-2}\right)\right\|_2 = 0$$

for any $x^i \in \mathbb{R}^{n_i}$, $i = 2, 3, \ldots, d-2$. However, this implies that $\mathcal{A}(\bar{x}) = \mathbf{0}$, or equivalently, $\bar{x}$ belongs to the nullspace of the matrix A. This contradicts the fact that A has full column rank. Hence, we have $\Lambda_1^{\mathcal{A},d}(x) > 0$ for all $x \in \mathbb{R}^{n_1} \backslash \{\mathbf{0}\}$, as desired. $\square$

Armed with Proposition 4, we can now design a polynomial–time approximation algorithm for the following multilinear optimization problem:

$$\begin{aligned} \bar{v}_{ml}(\mathcal{A}, d) = \text{ maximize } \quad & \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} a_{i_1 i_2 \cdots i_d} x_{i_1}^1 x_{i_2}^2 \cdots x_{i_d}^d \\ \text{subject to } \quad & \|x^i\|_2 = 1 \quad \text{for } i = 1, \ldots, d, \\ & x^i \in \mathbb{R}^{n_i} \quad \text{for } i = 1, \ldots, d, \end{aligned} \tag{16}$$

where $d \geq 3$ is a given integer. Specifically, we prove the following

Theorem 4 *For any given $d \geq 3$, let $1 \leq n_1 \leq n_2 \leq \ldots \leq n_d$ be arbitrary integers, and let $\|\cdot\|_{P_i} : \mathbb{R}^{n_{d-1-i}} \to \mathbb{R}_+$, where $i = d-2, d-3, \ldots, 1$, be polytopal norms possessing the properties guaranteed in Theorem 2. Furthermore, let $\mathcal{A} = (a_{i_1 i_2 \cdots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be an arbitrary non–zero order–d tensor, and let A be the $(n_2 \times \cdots \times n_d) \times n_1$ matrix given by (11). Suppose that A has full column rank. Then, the functions $\{\Lambda_j^{\mathcal{A},d}\}_{j=1}^{d-2}$, where $\Lambda_j^{\mathcal{A},d} : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j} \to \mathbb{R}_+$ is given by*

$$\begin{aligned} & \Lambda_j^{\mathcal{A},d}\left(x^1, \ldots, x^j\right) \\ & = \begin{cases} \mathit{diam}_{P_{d-2-j}}\left(\left\{x \in \mathbb{R}^{n_{j+1}} : \Lambda_{j+1}^{\mathcal{A},d}\left(x^1, \ldots, x^j, x\right) \leq 1\right\}^\circ\right) \\ \qquad \textit{for } j = 1, \ldots, d-3, \\ \left\|\mathcal{A}\left(x^1, \ldots, x^{d-2}\right)\right\|_2 \qquad \textit{for } j = d-2, \end{cases} \end{aligned}$$

satisfy

$$\begin{aligned} \Omega\left(\prod_{i=1}^{d-2}\sqrt{\frac{\log n_i}{n_i}}\right) \cdot \bar{v}_{ml}(\mathcal{A}, d) & \leq \frac{1}{2}\mathit{diam}_{P_{d-2}}\left(\left\{x \in \mathbb{R}^{n_1} : \Lambda_1^{\mathcal{A},d}(x) \leq 1\right\}^\circ\right) \\ & \leq \bar{v}_{ml}(\mathcal{A}, d), \end{aligned}$$

and $\Lambda_1^{\mathcal{A},d}$ is an efficiently computable norm on $\mathbb{R}^{n_1}$. Moreover, there exist rational numbers $0 < r \leq R < \infty$, whose encoding lengths are polynomially bounded by the input size of problem (16), such that

$$B_2^{n_1}(r) \subset \left\{x \in \mathbb{R}^{n_1} : \Lambda_1^{\mathcal{A},d}(x) \leq 1\right\} \subset B_2^{n_1}(R).$$

Consequently, the quantity $\mathrm{diam}_{P_{d-2}}\left(\left\{x \in \mathbb{R}^{n_1} : \Lambda_1^{\mathcal{A},d}(x) \le 1\right\}^\circ\right)$ *can also be efficiently computed.*

Remarks.

1. The assumption that A has full column rank can be made without loss of generality, since vectors in the nullspace of A can never be part of an optimal solution to problem (16) (unless $\mathcal{A} = \mathbf{0}$).
2. When we say that a quantity is *efficiently computable*, we mean that it can be computed to any desired accuracy by a deterministic algorithm whose runtime is polynomial in the input size of (16) and the level of accuracy, provided that $d \ge 3$ is *fixed*.
3. Theorem 4 states that the optimal value of problem (16) can be approximated to within a factor of $\Omega(\prod_{i=1}^{d-2}\sqrt{\log n_i/n_i})$. Using an argument similar to that in the paragraph preceding Theorem 3, one can find a feasible solution to problem (16) that attains the stated approximation ratio.

Proof We proceed by induction on $d \ge 3$. The base case follows from the results of Theorem 3. Now, suppose that $d > 3$. Let $x^1 \in \mathbb{R}^{n_1}\backslash\{\mathbf{0}\}$ be arbitrary, and consider the order–$(d-1)$ tensor $\mathcal{A}(x^1) \in \mathbb{R}^{n_2 \times n_3 \times \cdots \times n_d}$. Without loss of generality, we may assume that the $(n_3 \times \cdots \times n_d) \times n_2$ matrix $A(x^1)$, where $\left[A(x^1)\right]_{(i_3,\ldots,i_d),i_2} = \left[\mathcal{A}(x^1)\right]_{i_2 i_3 \cdots i_d}$, has full column rank. Then, by the inductive hypothesis, the functions $\{\Lambda_j^{\mathcal{A}(x^1),d-1}\}_{j=1}^{d-3}$, where $\Lambda_j^{\mathcal{A}(x^1),d-1} : \mathbb{R}^{n_2 \times \cdots \times n_{j+1}} \to \mathbb{R}_+$ is given by

$$\begin{aligned}
&\Lambda_j^{\mathcal{A}(x^1),d-1}\left(x^2,\ldots,x^{j+1}\right) \\
&= \begin{cases} \mathrm{diam}_{P_{d-3-j}}\left(\left\{x \in \mathbb{R}^{n_{j+2}} : \Lambda_{j+1}^{\mathcal{A}(x^1),d-1}\left(x^2,\ldots,x^{j+1},x\right) \le 1\right\}^\circ\right) \\ \qquad \text{for } j = 1,\ldots,d-4, \\ \left\|\left[\mathcal{A}\left(x^1\right)\right]\left(x^2,x^3,\ldots,x^{d-2}\right)\right\|_2 \qquad \text{for } j = d-3, \end{cases}
\end{aligned}$$

satisfy

$$\begin{aligned}
&\Omega\left(\prod_{i=2}^{d-2}\sqrt{\frac{\log n_i}{n_i}}\right) \cdot \bar{v}_{ml}(\mathcal{A}(x^1),d-1) \\
&\le \frac{1}{2}\mathrm{diam}_{P_{d-3}}\left(\left\{x \in \mathbb{R}^{n_2} : \Lambda_1^{\mathcal{A}(x^1),d-1}(x) \le 1\right\}^\circ\right) \\
&\le \bar{v}_{ml}(\mathcal{A}(x^1),d-1). \qquad (17)
\end{aligned}$$

Now, define the functions $\{\Lambda_j^{\mathcal{A},d}\}_{j=1}^{d-2}$ by

$$\Lambda_1^{\mathcal{A},d}\left(x^1\right) = \mathrm{diam}_{P_{d-3}}\left(\left\{x \in \mathbb{R}^{n_2} : \Lambda_2^{\mathcal{A},d}(x^1,x) \le 1\right\}^\circ\right),$$

$$\Lambda_j^{\mathcal{A},d}\left(x^1,x^2,\ldots,x^j\right) = \Lambda_{j-1}^{\mathcal{A}(x^1),d-1}\left(x^2,\ldots,x^j\right) \quad \text{for } j = 2,\ldots,d-2.$$

Then, by construction, we have

$$\begin{aligned}\varLambda_{d-2}^{\mathcal{A},d}\left(x^1,x^2,\dots,x^{d-2}\right) &= \left\|\left[\mathcal{A}\left(x^1\right)\right]\left(x^2,x^3,\dots,x^{d-2}\right)\right\|_2 \\ &= \left\|\mathcal{A}\left(x^1,x^2,\dots,x^{d-2}\right)\right\|_2\end{aligned}$$

and

$$\begin{aligned}&\varLambda_{j}^{\mathcal{A},d}\left(x^1,x^2,\dots,x^{j}\right) \\ &= \mathrm{diam}_{P_{d-2-j}}\left(\left\{x\in\mathbb{R}^{n_{j+1}} : \varLambda_{j}^{\mathcal{A}(x^1),d-1}\left(x^2,\dots,x^j,x\right)\le 1\right\}^\circ\right) \\ &= \mathrm{diam}_{P_{d-2-j}}\left(\left\{x\in\mathbb{R}^{n_{j+1}} : \varLambda_{j+1}^{\mathcal{A},d}\left(x^1,\dots,x^j,x\right)\le 1\right\}^\circ\right)\end{aligned}$$

for $j=d-3,d-4,\dots,1$. Thus, by Proposition 4 and the fact that A has full column rank, the function $\varLambda_1^{\mathcal{A},d}:\mathbb{R}^{n_1}\to\mathbb{R}_+$ defines a norm on $\mathbb{R}^{n_1}$. Moreover, by the inductive hypothesis and the fact that $d\ge 4$ is fixed, the function $x\mapsto \varLambda_2^{\mathcal{A},d}(x^1,x)=\varLambda_1^{\mathcal{A}(x^1),d-1}(x)$ is an efficiently computable norm on $\mathbb{R}^{n_2}$, and there exist rational numbers $0<r^1\le R^1<\infty$, whose encoding lengths can be polynomially bounded by the input size of the multilinear optimization problem associated with $\mathcal{A}(x^1)$, such that

$$B_2^{n_2}(r^1)\subset\left\{x\in\mathbb{R}^{n_2} : \varLambda_2^{\mathcal{A},d}(x^1,x)\le 1\right\}\subset B_2^{n_2}(R^1).$$

Hence, by arguing as in the proof of Proposition 3 and applying Theorem 2, we conclude that $\varLambda_1^{\mathcal{A},d}$ is efficiently computable.

Now, let $B_{\varLambda_1^{\mathcal{A},d}}=\left\{x\in\mathbb{R}^{n_1} : \varLambda_1^{\mathcal{A},d}(x)\le 1\right\}$ be the unit ball of the norm $\varLambda_1^{\mathcal{A},d}$. To show that $B_{\varLambda_1^{\mathcal{A},d}}$ is well–bounded, we first recall from (17) and the definition of $\varLambda_1^{\mathcal{A},d}$ that

$$\varOmega\left(\prod_{i=2}^{d-2}\sqrt{\frac{\log n_i}{n_i}}\right)\cdot\bar{v}_{ml}(\mathcal{A}(x),d-1)\le\frac{1}{2}\varLambda_1^{\mathcal{A},d}(x)\le\bar{v}_{ml}(\mathcal{A}(x),d-1)$$

for any $x\in\mathbb{R}^{n_1}$. Let $e^i\in\mathbb{R}^{n_1}$ be the i–th basis vector in $\mathbb{R}^{n_1}$, where $i=1,\dots,n_1$. Then, we have

$$\varLambda_1^{\mathcal{A},d}(e^i)\le 2\bar{v}_{ml}(\mathcal{A}(e^i),d-1)\le r_i\equiv 2\sum_{i_2=1}^{n_2}\cdots\sum_{i_d=1}^{n_d}|a_{i,i_2\cdots i_d}|.$$

Upon setting $r'=\max_{1\le i\le n_1} r_i$, we see that

$$\frac{1}{r'}\cdot\mathrm{conv}\left(\left\{\pm e^1,\dots,\pm e^{n_1}\right\}\right)=\left\{x\in\mathbb{R}^{n_1} : \|x\|_1\le\frac{1}{r'}\right\}\subset B_{\varLambda_1^{\mathcal{A},d}}.$$

In particular, we have $B_2^{n_1}(r)\subset B_{\varLambda_1^{\mathcal{A},d}}$, where $r=1/(\lceil\sqrt{n_1}\rceil\cdot r')>0$ is a rational number, whose encoding length is polynomially bounded by the input size of problem (16).

On the other hand, for any $x \in \mathbb{R}^{n_1}$, we compute

$$\bar{v}_{ml}(\mathcal{A}(x), d-1) \geq \max_{1 \leq i_2 \leq n_2, \ldots, 1 \leq i_d \leq n_d} \left| \sum_{i_1=1}^{n_1} a_{i_1 i_2 \cdots i_d} x_{i_1} \right| = \|Ax\|_\infty$$

$$\geq \sqrt{\frac{\lambda_{min}(A^T A)}{n_2 \cdots n_d}} \cdot \|x\|_2.$$

Since A has full column rank, we conclude that $B_{\Lambda_1^{\mathcal{A},d}} \subset B_2^{n_1}(R)$, where

$$R = O\left(\sqrt{\frac{n_2 \cdots n_d}{\lambda_{min}(A^T A)} \cdot \prod_{i=2}^{d-2} \frac{n_i}{\log n_i}} \right)$$

can be chosen as a finite rational number. Moreover, the encoding length of R can be polynomially bounded by the input size of problem (16). Now, by arguing as in the proof of Proposition 3 and applying Theorem 2, we conclude that the quantity $\mathrm{diam}_{P_{d-2}}\left(\left\{ x \in \mathbb{R}^{n_1} : \Lambda_1^{\mathcal{A},d}(x) \leq 1 \right\}^\circ \right)$ is efficiently computable.

Finally, by maximizing the terms in (17) over x^1, we have

$$\Omega\left(\prod_{i=2}^{d-2} \sqrt{\frac{\log n_i}{n_i}} \right) \cdot \bar{v}_{ml}(\mathcal{A}, d) \leq \frac{1}{2} \max_{x \in \mathbb{R}^{n_1} : \|x\|_2 = 1} \Lambda_1^{\mathcal{A},d}(x) \leq \bar{v}_{ml}(\mathcal{A}, d). \quad (18)$$

In particular, we have reduced the problem of approximating $\bar{v}_{ml}(\mathcal{A}, d)$ to that of approximating $\max_{x \in \mathbb{R}^{n_1} : \|x\|_2 = 1} \Lambda_1^{\mathcal{A},d}(x)$. Now, by mimicking the derivation of (6), it can be shown that

$$\max_{x \in \mathbb{R}^{n_1} : \|x\|_2 = 1} \Lambda_1^{\mathcal{A},d}(x) = \frac{1}{2} \mathrm{diam}_2\left(B^\circ_{\Lambda_1^{\mathcal{A},d}} \right). \quad (19)$$

Moreover, by Theorem 2 and the definition of $\|\cdot\|_{P_{d-2}}$, we have

$$\Omega\left(\sqrt{\frac{\log n_1}{n_1}} \right) \cdot \mathrm{diam}_2\left(B^\circ_{\Lambda_1^{\mathcal{A},d}} \right) \leq \mathrm{diam}_{P_{d-2}}\left(B^\circ_{\Lambda_1^{\mathcal{A},d}} \right) \leq \mathrm{diam}_2\left(B^\circ_{\Lambda_1^{\mathcal{A},d}} \right). \quad (20)$$

It then follows from (18), (19) and (20) that

$$\Omega\left(\prod_{i=1}^{d-2} \sqrt{\frac{\log n_i}{n_i}} \right) \cdot \bar{v}_{ml}(\mathcal{A}, d) \leq \frac{1}{2} \mathrm{diam}_{P_{d-2}}\left(\left\{ x \in \mathbb{R}^{n_1} : \Lambda_1^{\mathcal{A},d}(x) \leq 1 \right\}^\circ \right)$$

$$\leq \bar{v}_{ml}(\mathcal{A}, d).$$

This completes the inductive step and also the proof of Theorem 4. $\square$

Upon combining the results of Theorem 4 with that of Theorem 1, we have the following

Corollary 1 *Let $d \geq 3$ be given. Then, there is a deterministic polynomial–time approximation algorithm for* (HP) *with approximation ratio (resp. relative approximation ratio)* $\alpha = \Omega\left(\log^{\frac{d}{2}-1} n \middle/ n^{\frac{d}{2}-1}\right)$ *when d is odd (resp. even).*

The bound on α in Corollary 1 improves upon the $\Omega((1/n)^{d/2-1})$ bound established in [8].

3.2 Sphere Constrained Multiquadratic Optimization

The techniques introduced in the previous section can also be used to design a polynomial–time approximation algorithm for sphere constrained multiquadratic optimization problems. Specifically, let $d \geq 2$ be a given integer. Let $\mathcal{A} = (a_{i_1 i_2 \cdots i_{2d-1} i_{2d}}) \in \mathbb{R}^{n_1^2 \times \cdots \times n_d^2}$ be a non–zero order–$2d$ tensor that is *partially symmetric*, i.e.,

$$a_{i_1 i_2 \cdots i_{2j-1} i_{2j} \cdots i_{2d-1} i_{2d}} = a_{i_1 i_2 \cdots i_{2j} i_{2j-1} \cdots i_{2d-1} i_{2d}}$$

for $j = 1, \ldots, d$ and $i_{2k-1}, i_{2k} = 1, \ldots, n_k$, where $k = 1, \ldots, d$. Without loss of generality, we assume that $1 \leq n_1 \leq \cdots \leq n_d$. Let

$$f\left(x^1, \ldots, x^d\right) \equiv \sum_{1 \leq i_1, i_2 \leq n_1} \cdots \sum_{1 \leq i_{2d-1}, i_{2d} \leq n_d} a_{i_1 i_2 \cdots i_{2d-1} i_{2d}} x^1_{i_1} x^1_{i_2} \cdots x^d_{i_{2d-1}} x^d_{i_{2d}},$$

and consider the following multiquadratic optimization problem:

$$(\mathsf{MQ}) \qquad \begin{array}{rll} \bar{v}_{mq} = \text{maximize} & f\left(x^1, \ldots, x^d\right) & \\ \text{subject to} & \|x^i\|_2 = 1 & \text{for } i = 1, \ldots, d, \\ & x^i \in \mathbb{R}^{n_i} & \text{for } i = 1, \ldots, d. \end{array}$$

Note that when $d = 2$, problem (MQ) is simply the biquadratic optimization problem introduced in [21]. Similar to the approach we used for the homogeneous polynomial optimization problem (HP), we begin by studying the following multilinear relaxation of (MQ):

$$(\mathsf{MQL}) \qquad \begin{array}{rll} \text{maximize} & F\left(x^1, \ldots, x^{2d}\right) & \\ \text{subject to} & \|x^i\|_2 = 1 & \text{for } i = 1, \ldots, 2d, \\ & x^{2i-1}, x^{2i} \in \mathbb{R}^{n_i} & \text{for } i = 1, \ldots, d, \end{array}$$

where

$$F\left(x^1, \ldots, x^{2d}\right) \equiv \sum_{1 \leq i_1, i_2 \leq n_1} \cdots \sum_{1 \leq i_{2d-1}, i_{2d} \leq n_d} a_{i_1 i_2 \cdots i_{2d-1} i_{2d}} x^1_{i_1} x^2_{i_2} \cdots x^{2d-1}_{i_{2d-1}} x^{2d}_{i_{2d}}.$$

Clearly, we have $f\left(x^1, \ldots, x^d\right) = F\left(x^1, x^1, \ldots, x^d, x^d\right)$ for any $x^i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, d$. Similar to Proposition 1, the following polarization–type formula establishes the relationship between the objective values of (MQ) and (MQL). We relegate its proof to the appendix (see Appendix A).

Proposition 5 *Let $x^{2i-1}, x^{2i} \in \mathbb{R}^{n_i}$, where $i = 1, \ldots, d$, be arbitrary vectors. Let $\xi_1, \ldots, \xi_{2d}$ be i.i.d. Bernoulli random variables. Then, we have*

$$\mathbb{E}\left[\left(\prod_{l=1}^{2d}\xi_l\right) f\left(\sum_{i=1}^{2}\xi_i x^i, \sum_{i=3}^{4}\xi_i x^i, \ldots, \sum_{i=2d-1}^{2d}\xi_i x^i\right)\right] = 2^d F\left(x^1, \ldots, x^{2d}\right).$$

Proposition 5 allows us to focus on the multilinear optimization problem (MQL), for which a deterministic polynomial–time approximation algorithm is available by the results in the previous section. Now, by adapting an argument of He et al. [8] and using Proposition 5, we can prove the following

Theorem 5 *Let $d \geq 2$ be given. Then, there is a deterministic polynomial–time approximation algorithm for* (MQ) *with relative approximation ratio $2^{-d} \cdot \Omega\left(\prod_{i=1}^{d-1}\left(\log n_i / n_i\right)\right)$.*

Remark. When $d = 2$, the approximation bound we obtain is $\Omega(\log n_1/n_1)$, which improves upon the $\Omega(1/n_2^2)$ bound established in [21] (recall that $n_1 \leq n_2$ by assumption).

Proof Define the functions $H : \mathbb{R}^{n_1^2 \times \cdots \times n_d^2} \to \mathbb{R}$ and $h : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}_+$ by

$$H\left(x^1, x^2, \ldots, x^{2d-1}, x^{2d}\right) = \prod_{i=1}^{d}\left(x^{2i-1}\right)^T\left(x^{2i}\right),$$

$$h\left(x^1, \ldots, x^d\right) = H\left(x^1, x^1, \ldots, x^d, x^d\right) = \prod_{i=1}^{d}\|x^i\|_2^2.$$

Note that H (resp. h) is a multilinear (resp. multiquadratic) form. Now, let $\bar{x}^i \in \mathbb{R}^{n_i}$ be such that $\|\bar{x}^i\|_2 = 1$ for $i = 1, \ldots, d$, and define the function $G : \mathbb{R}^{n_1^2 \times \cdots \times n_d^2} \to \mathbb{R}$ by

$$\begin{aligned} &G\left(x^1, x^2, \ldots, x^{2d-1}, x^{2d}\right) \\ &= F\left(x^1, x^2, \ldots, x^{2d-1}, x^{2d}\right) - f\left(\bar{x}^1, \ldots, \bar{x}^d\right) H\left(x^1, x^2, \ldots, x^{2d-1}, x^{2d}\right). \end{aligned}$$

Consider the following auxiliary optimization problem:

$$(\mathsf{AUX}) \qquad \begin{array}{rll} \bar{v}_{aux} = \text{maximize} & G\left(x^1, x^2, \ldots, x^{2d-1}, x^{2d}\right) & \\ \text{subject to} & \|x^i\|_2 = 1 & \text{for } i = 1, \ldots, 2d, \\ & x^{2i-1}, x^{2i} \in \mathbb{R}^{n_i} & \text{for } i = 1, \ldots, d. \end{array}$$

Since (AUX) is a sphere constrained multilinear optimization problem, by Theorem 4, we can find a feasible solution $\{\hat{x}^i\}_{i=1}^{2d}$ to it with

$$G\left(\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^{2d-1}, \hat{x}^{2d}\right) \geq \alpha \cdot \bar{v}_{aux}, \quad \text{where } \alpha = \Omega\left(\prod_{i=1}^{d-1}\frac{\log n_i}{n_i}\right).$$

Now, let $\underline{v}_{mq}$ be the optimal value of the problem obtained from (MQ) by changing the word "maximize" to "minimize". Consider first the case where

$$f\left(\bar{x}^1,\ldots,\bar{x}^d\right) - \underline{v}_{mq} \le \frac{\alpha}{4}\left(\bar{v}_{mq} - \underline{v}_{mq}\right). \tag{21}$$

Since $\left|H\left(\hat{x}^1,\hat{x}^2,\ldots,\hat{x}^{2d-1},\hat{x}^{2d}\right)\right| \le 1$ by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&F\left(\hat{x}^1,\hat{x}^2,\ldots,\hat{x}^{2d-1},\hat{x}^{2d}\right) - \underline{v}_{mq}\cdot H\left(\hat{x}^1,\hat{x}^2,\ldots,\hat{x}^{2d-1},\hat{x}^{2d}\right)\\
&= G\left(\hat{x}^1,\hat{x}^2,\ldots,\hat{x}^{2d-1},\hat{x}^{2d}\right) + \left[f\left(\bar{x}^1,\ldots,\bar{x}^d\right) - \underline{v}_{mq}\right] H\left(\hat{x}^1,\ldots,\hat{x}^{2d}\right)\\
&\ge \alpha\cdot \bar{v}_{aux} - \left[f\left(\bar{x}^1,\ldots,\bar{x}^d\right) - \underline{v}_{mq}\right]\\
&\ge \alpha\cdot\left[\bar{v}_{mq} - f\left(\bar{x}^1,\ldots,\bar{x}^d\right)\right] - \frac{\alpha}{4}\left(\bar{v}_{mq} - \underline{v}_{mq}\right) && (22)\\
&\ge \frac{\alpha}{2}\left(\bar{v}_{mq} - \underline{v}_{mq}\right), && (23)
\end{aligned}$$

where (22) follows from (21) and the fact that the optimal solution to (MQ) is feasible for (AUX), and (23) follows since

$$f\left(\bar{x}^1,\ldots,\bar{x}^d\right) \le \frac{\alpha}{4}\bar{v}_{mq} + \left(1 - \frac{\alpha}{4}\right)\underline{v}_{mq}$$

by (21). Upon letting $u^i = \sum_{j=2i-1}^{2i}\xi_j\hat{x}^j$ for $i = 1,\ldots,d$ and using Proposition 5, we have

$$\begin{aligned}
&2^d\left[F\left(\hat{x}^1,\hat{x}^2,\ldots,\hat{x}^{2d-1},\hat{x}^{2d}\right) - \underline{v}_{mq}\cdot H\left(\hat{x}^1,\hat{x}^2,\ldots,\hat{x}^{2d-1},\hat{x}^{2d}\right)\right]\\
&= \mathbb{E}\left[\left(\prod_{l=1}^{2d}\xi_l\right)\left[f\left(u^1,u^2,\ldots,u^d\right) - \underline{v}_{mq}\cdot h\left(u^1,u^2,\ldots,u^d\right)\right]\right]\\
&= \Pr\left(\prod_{l=1}^{2d}\xi_l = 1\right)\cdot\mathbb{E}\left[f\left(u^1,u^2,\ldots,u^d\right) - \underline{v}_{mq}\cdot\prod_{i=1}^d\|u^i\|_2^2 \;\middle|\; \prod_{l=1}^{2d}\xi_l = 1\right]\\
&\quad - \Pr\left(\prod_{l=1}^{2d}\xi_l = -1\right)\cdot\mathbb{E}\left[f\left(u^1,u^2,\ldots,u^d\right) - \underline{v}_{mq}\cdot\prod_{i=1}^d\|u^i\|_2^2 \;\middle|\; \prod_{l=1}^{2d}\xi_l = -1\right]\\
&\le \frac{1}{2}\mathbb{E}\left[f\left(u^1,u^2,\ldots,u^d\right) - \underline{v}_{mq}\cdot\prod_{i=1}^d\|u^i\|_2^2 \;\middle|\; \prod_{l=1}^{2d}\xi_l = 1\right],
\end{aligned}$$

where the last inequality follows from the fact that the solution $\left\{u^i/\|u^i\|_2\right\}_{i=1}^d$ satisfies the constraints in (MQ), and hence

$$f\left(u^1,u^2,\ldots,u^d\right) - \underline{v}_{mq}\cdot\prod_{i=1}^d\|u^i\|_2^2 \ge 0.$$

In particular, we see that there exists a vector $\beta \in \{-1,1\}^{2d}$ such that $\prod_{l=1}^{2d} \beta_l = 1$ and

$$\frac{1}{2}\left[f\left(\sum_{i=1}^{2} \beta_i \hat{x}^i, \sum_{i=3}^{4} \beta_i \hat{x}^i, \ldots, \sum_{i=2d-1}^{2d} \beta_i \hat{x}^i \right) - \underline{v}_{mq} \cdot \prod_{i=1}^{d} \left\| \sum_{j=2i-1}^{2i} \beta_j \hat{x}^j \right\|_2^2 \right]$$
$$\geq \alpha \cdot 2^{d-1} \cdot \left(\bar{v}_{mq} - \underline{v}_{mq} \right). \tag{24}$$

Moreover, since $d \geq 2$ is fixed, such a vector can be found in constant time. Upon setting

$$\tilde{x}^j = \sum_{i=2j-1}^{2j} \beta_i \hat{x}^i \Bigg/ \left\| \sum_{i=2j-1}^{2j} \beta_i \hat{x}^i \right\|_2 \quad \text{for } j = 1, \ldots, d$$

and noting that $\left\| \sum_{i=2j-1}^{2j} \beta_i \hat{x}^i \right\|_2 \leq 2$ for $j = 1, \ldots, d$, we conclude from (24) that

$$f\left(\tilde{x}^1, \ldots, \tilde{x}^d \right) - \underline{v}_{mq} \geq \alpha \cdot 2^{-d} \cdot \left(\bar{v}_{mq} - \underline{v}_{mq} \right),$$

as desired. Now, suppose that the condition in (21) does not hold. Then, we have

$$f\left(\bar{x}^1, \ldots, \bar{x}^d \right) - \underline{v}_{mq} > \frac{\alpha}{4} \left(\bar{v}_{mq} - \underline{v}_{mq} \right) \geq \alpha \cdot 2^{-d} \cdot \left(\bar{v}_{mq} - \underline{v}_{mq} \right),$$

and the claim in the theorem statement is trivially satisfied. This completes the proof. □

4 Concluding Remarks

It has been known that the approximability of various polynomial optimization problems is closely related to the approximability of their multilinear relaxations. By reducing the problem of optimizing a multilinear form over spheres to that of determining the L_2–diameter of a certain convex body, we were able to utilize powerful results from the algorithmic theory of convex bodies to develop deterministic polynomial–time approximation algorithms for a host of sphere constrained polynomial optimization problems. Moreover, our algorithms have the best known approximation guarantees to date. We believe that our approach will find further applications in the design of approximation algorithms for other norm constrained polynomial optimization problems.

Acknowledgements The author would like to thank the reviewers for their comments and suggestions.

References

1. Barvinok, A.: Integration and Optimization of Multivariate Polynomials by Restriction onto a Random Subspace. Foundations of Computational Mathematics **7**(2), 229–244 (2007)
2. Brieden, A., Gritzmann, P., Kannan, R., Klee, V., Lovász, L., Simonovits, M.: Deterministic and Randomized Polynomial–Time Approximation of Radii. Mathematika **48**, 63–105 (2001)
3. Golub, G.H., Van Loan, C.F.: Matrix Computations, third edn. The Johns Hopkins University Press, Baltimore, Maryland (1996)
4. Gritzmann, P., Klee, V.: Inner and Outer j–Radii of Convex Bodies in Finite–Dimensional Normed Spaces. Discrete and Computational Geometry **7**(1), 255–280 (1992)
5. Grötschel, M., Lovász, L., Schrijver, A.: Geometric Algorithms and Combinatorial Optimization, *Algorithms and Combinatorics*, vol. 2, second corrected edn. Springer–Verlag, Berlin Heidelberg (1993)
6. Han, D., Dai, H.H., Qi, L.: Conditions for Strong Ellipticity of Anisotropic Elastic Materials. Journal of Elasticity **97**(1), 1–13 (2009)
7. He, S., Li, Z., Zhang, S.: General Constrained Polynomial Optimization: an Approximation Approach. Tech. Rep. SEEM2009–06, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong (2009)
8. He, S., Li, Z., Zhang, S.: Approximation Algorithms for Homogeneous Polynomial Optimization with Quadratic Constraints. Mathematical Programming, Series B **125**(2), 353–383 (2010)
9. Hillar, C.J., Lim, L.H.: Most Tensor Problems are NP Hard (2009). Preprint
10. Ishteva, M., Absil, P.A., van Huffel, S., de Lathauwer, L.: On the Best Low Multilinear Rank Approximation of Higher–Order Tensors. In: M. Diehl, F. Glineur, E. Jarlebring, W. Michiels (eds.) Recent Advances in Optimization and its Applications in Engineering, pp. 145–164. Springer–Verlag, Berlin Heidelberg (2010)
11. Kannan, R.: Spectral Methods for Matrices and Tensors. In: Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC 2010), pp. 1–12 (2010)
12. Khot, S., Naor, A.: Linear Equations Modulo 2 and the L_1 Diameter of Convex Bodies. SIAM Journal on Computing **38**(4), 1448–1463 (2008)
13. de Klerk, E.: The Complexity of Optimizing over a Simplex, Hypercube or Sphere: a Short Survey. Central European Journal of Operations Research **16**(2), 111–125 (2008)
14. de Klerk, E., Laurent, M., Parrilo, P.A.: A PTAS for the Minimization of Polynomials of Fixed Degree over the Simplex. Theoretical Computer Science **361**(2–3), 210–225 (2006)
15. Kofidis, E., Regalia, P.A.: Tensor Approximation and Signal Processing Applications. In: V. Olshevsky (ed.) Structured Matrices in Mathematics, Computer Science and Engineering I: Proceedings of an AMS–IMS–SIAM Joint Summer Research Conference, *Contemporary Mathematics*, vol. 280, pp. 103–133. American Mathematical Society, Providence, Rhode Island (2001)
16. Kolda, T.G., Bader, B.W.: Tensor Decompositions and Applications. SIAM Review **51**(3), 455–500 (2009)
17. Kwapien, S.: Decoupling Inequalities for Polynomial Chaos. The Annals of Probability **15**(3), 1062–1071 (1987)
18. Lasserre, J.B.: Global Optimization with Polynomials and the Problem of Moments. SIAM Journal on Optimization **11**(3), 796–817 (2001)
19. Laurent, M.: Sums of Squares, Moment Matrices and Optimization over Polynomials. In: M. Putinar, S. Sullivant (eds.) Emerging Applications of Algebraic Geometry, *The IMA Volumes in Mathematics and Its Applications*, vol. 149, pp. 157–270. Springer Science+Business Media, LLC, New York (2009)
20. Lim, L.H., Comon, P.: Multiarray Signal Processing: Tensor Decomposition Meets Compressed Sensing. Comptes Rendus Mécanique **338**(6), 311–320 (2010)
21. Ling, C., Nie, J., Qi, L., Ye, Y.: Biquadratic Optimization over Unit Spheres and Semidefinite Programming Relaxations. SIAM Journal on Optimization **20**(3), 1286–1310 (2009)

22. Luo, Z.Q., Zhang, S.: A Semidefinite Relaxation Scheme for Multivariate Quartic Polynomial Optimization with Quadratic Constraints. SIAM Journal on Optimization **20**(4), 1716–1736 (2010)
23. Nesterov, Yu.: Random Walk in a Simplex and Quadratic Optimization over Convex Polytopes. CORE Discussion Paper 2003071, Université Catholique de Louvain, Belgium (2003)
24. Nie, J.: An Approximation Bound Analysis for Lasserre's Relaxation in Multivariate Polynomial Optimization (2009). Preprint
25. Parrilo, P.A.: Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization. Ph.D. thesis, California Institute of Technology, Pasadena, CA 91125 (2000)
26. Qi, L.: Eigenvalues of a Real Supersymmetric Tensor. Journal of Symbolic Computation **40**(6), 1302–1324 (2005)
27. Qi, L., Dai, H.H., Han, D.: Conditions for Strong Ellipticity and M–Eigenvalues. Frontiers of Mathematics in China **4**(2), 349–364 (2009)
28. Qi, L., Wang, F., Wang, Y.: Z–Eigenvalue Methods for a Global Polynomial Optimization Problem. Mathematical Programming, Series A **118**(2), 301–316 (2009)
29. Reznick, B.: Some Concrete Aspects of Hilbert's 17th Problem. In: C.N. Delzell, J.J. Madden (eds.) Real Algebraic Geometry and Ordered Structures, *Contemporary Mathematics*, vol. 253, pp. 251–272. American Mathematical Society, Providence, Rhode Island (2000)
30. So, A.M.C., Ye, Y., Zhang, J.: A Unified Theorem on SDP Rank Reduction. Mathematics of Operations Research **33**(4), 910–920 (2008)
31. Weiland, S., van Belzen, F.: Singular Value Decompositions and Low Rank Approximations of Tensors. IEEE Transactions on Signal Processing **58**(3), 1171–1182 (2010)

Appendix

A Polarization–Type Formula for Multiquadratic Optimization

In this section, we prove Proposition 5. For the reader's convenience, we reproduce the statement here:

Proposition 5 *Let $x^{2i-1}, x^{2i} \in \mathbb{R}^{n_i}$, where $i = 1, \ldots, d$, be arbitrary vectors. Let $\xi_1, \ldots, \xi_{2d}$ be i.i.d. Bernoulli random variables. Then, we have*

$$\mathbb{E}\left[\left(\prod_{u=1}^{2d}\xi_u\right) f\left(\sum_{i=1}^{2}\xi_i x^i, \sum_{i=3}^{4}\xi_i x^i, \ldots, \sum_{i=2d-1}^{2d}\xi_i x^i\right)\right] = 2^d F\left(x^1, \ldots, x^{2d}\right).$$

Proof Let $\mathcal{S} = \left\{(i_1, i_2, \ldots, i_{2d-1}, i_{2d}) \in \mathbb{Z}^{2d} : 1 \le i_1, i_2 \le n_1; \ldots; 1 \le i_{2d-1}, i_{2d} \le n_d\right\}$. By definition, we have

$$\begin{aligned} & f\left(\sum_{i=1}^{2}\xi_i x^i, \sum_{i=3}^{4}\xi_i x^i, \ldots, \sum_{i=2d-1}^{2d}\xi_i x^i\right) \\ &= \sum_{(i_1,i_2,\ldots,i_{2d-1},i_{2d})\in\mathcal{S}} a_{i_1\cdots i_{2d}} \prod_{j=1}^{d}\left(\xi_{2j-1}x^{2j-1}_{i_{2j-1}} + \xi_{2j}x^{2j}_{i_{2j-1}}\right)\left(\xi_{2j-1}x^{2j-1}_{i_{2j}} + \xi_{2j}x^{2j}_{i_{2j}}\right). \end{aligned}$$

Since

$$\begin{aligned} & \left(\xi_{2j-1}x^{2j-1}_{i_{2j-1}} + \xi_{2j}x^{2j}_{i_{2j-1}}\right)\left(\xi_{2j-1}x^{2j-1}_{i_{2j}} + \xi_{2j}x^{2j}_{i_{2j}}\right) \\ &= \sum_{\beta_{2j-1},\beta_{2j}=2j-1}^{2j} \xi_{\beta_{2j-1}}\xi_{\beta_{2j}} x^{\beta_{2j-1}}_{i_{2j-1}} x^{\beta_{2j}}_{i_{2j}}, \end{aligned}$$

it follows that

$$
\begin{aligned}
& f\left(\sum_{i=1}^{2}\xi_i x^i, \sum_{i=3}^{4}\xi_i x^i, \dots, \sum_{i=2d-1}^{2d}\xi_i x^i\right) \\
= & \sum_{(i_1,i_2,\dots,i_{2d-1},i_{2d})\in\mathcal{S}} a_{i_1 i_2\cdots i_{2d-1} i_{2d}} \prod_{j=1}^{d}\left(\sum_{\beta_{2j-1},\beta_{2j}=2j-1}^{2j} \xi_{\beta_{2j-1}}\xi_{\beta_{2j}} x_{i_{2j-1}}^{\beta_{2j-1}} x_{i_{2j}}^{\beta_{2j}}\right) \\
= & \sum_{1\le\beta_1,\beta_2\le 2}\cdots\sum_{2d-1\le\beta_{2d-1},\beta_{2d}\le 2d}\left[\left(\prod_{v=1}^{2d}\xi_{\beta_v}\right)\times\right. \\
& \qquad\qquad \left.\sum_{(i_1,i_2,\dots,i_{2d-1},i_{2d})\in\mathcal{S}} a_{i_1 i_2\cdots i_{2d-1} i_{2d}} x_{i_1}^{\beta_1} x_{i_2}^{\beta_2}\cdots x_{i_{2d-1}}^{\beta_{2d-1}} x_{i_{2d}}^{\beta_{2d}}\right] \\
= & \sum_{1\le\beta_1,\beta_2\le 2}\cdots\sum_{2d-1\le\beta_{2d-1},\beta_{2d}\le 2d}\left(\prod_{v=1}^{2d}\xi_{\beta_v}\right) F\left(x^{\beta_1}, x^{\beta_2}, \dots, x^{\beta_{2d}}\right).
\end{aligned}
$$

In particular, we obtain

$$
\begin{aligned}
& \mathbb{E}\left[\left(\prod_{u=1}^{2d}\xi_u\right) f\left(\sum_{i=1}^{2}\xi_i x^i, \sum_{i=3}^{4}\xi_i x^i, \dots, \sum_{i=2d-1}^{2d}\xi_i x^i\right)\right] \\
= & \sum_{1\le\beta_1,\beta_2\le 2}\cdots\sum_{2d-1\le\beta_{2d-1},\beta_{2d}\le 2d} F\left(x^{\beta_1}, x^{\beta_2}, \dots, x^{\beta_{2d}}\right)\mathbb{E}\left[\prod_{u=1}^{2d}\xi_u\cdot\prod_{v=1}^{2d}\xi_{\beta_v}\right] \\
= & \sum_{1\le\beta_1\ne\beta_2\le 2}\cdots\sum_{2d-1\le\beta_{2d-1}\ne\beta_{2d}\le 2d} F\left(x^{\beta_1}, x^{\beta_2}, \dots, x^{\beta_{2d}}\right) \qquad (25) \\
= & \, 2^d F\left(x^1, x^2, \dots, x^{2d}\right), \qquad (26)
\end{aligned}
$$

where (25) follows from the fact that

$$
\mathbb{E}\left[\prod_{u=1}^{2d}\xi_u\cdot\prod_{v=1}^{2d}\xi_{\beta_v}\right] = \begin{cases} 0 & \text{if } \beta_{2j-1}=\beta_{2j} \text{ for some } j=1,\dots,d, \\ 1 & \text{otherwise,} \end{cases}
$$

and (26) follows from the partial symmetry of the tensor $\mathcal{A}$. This completes the proof. □