

APPROXIMATING NONNEGATIVE POLYNOMIALS VIA SPECTRAL SPARSIFICATION*

ALPEREN A. ERGÜR†

Dedicated to Nuriye and Semih, and to thousands of souls longing for justice.

Abstract. We study polyhedral approximations to the cone of nonnegative polynomials. We show that any constant ratio polyhedral approximation to the cone of nonnegative degree $2d$ forms in n variables has to have exponentially many facets in terms of n . We also show that for any fixed $m \geq 3$, all linear m -dimensional sections of the nonnegative cone that include $(x_1^2 + x_2^2 + \cdots + x_n^2)^d$ have a constant ratio polyhedral approximation with $O(n^{m-2})$ many facets. Our approach is convex geometric, and parts of the argument rely on the recent solution of the Kadison–Singer problem. We also discuss a randomized polyhedral approximation which might be of independent interest.

Key words. polyhedral approximation, spectral sparsification, spectrahedron, polynomial optimization

AMS subject classifications. 90C22, 52B11, 52B55, 14P99

DOI. 10.1137/17M1121743

1. Introduction. Let $P_{n,2d}$ be the vector space of real homogenous degree $2d$ polynomials in n variables. The elements of $P_{n,2d}$ that are nonnegative on the sphere form a full-dimensional cone in $\mathbb{R}^{\dim(P_{n,2d})}$. The membership problem of this cone is algorithmically equivalent to global optimization of polynomials. In the case of quadratics ($d = 1$), membership of the cone can be checked efficiently. Starting with the $d = 2$ case, the membership problem is NP-Hard [1].

In general, it also seems hard to provide upper bounds for the complexity of the nonnegative cone membership problem. However, a very interesting result of Pebay, Rojas, and Thompson shows that for any fixed $\delta > 0$, deciding if the supremum of a polynomial with n variables and $n + n^\delta$ monomials exceeds a certain given number is $\text{NP}_{\mathbb{R}}$ -complete [27]. Indeed, it is now clear that the complexity of the membership problem for the cone of nonnegative polynomials is quite different in the case of sparse polynomials than in the case of arbitrary degree $2d$ homogenous forms (dense polynomials) [18, 25].

In this note, we are interested in polyhedral approximations to the cone of nonnegative polynomials. The discussion in the previous paragraph suggests that the computational complexity of the approximation will be quite different in the dense and the sparse cases. Our results confirm this intuition as follows: We show that any constant ratio polyhedral approximation in the dense case has to have exponentially many facets. We also show that for any subspace $E \subset P_{n,2d}$ with $\dim(E) = m$ and $(x_1^2 + x_2^2 + \cdots + x_n^2)^d \in E$, there exists a polyhedral cone with $O(n^{m-2})$ many facets which provides a constant ratio approximation to the nonnegative elements of E .

*Received by the editors March 20, 2017; accepted for publication (in revised form) January 8, 2019; published electronically March 21, 2019.

<http://www.siam.org/journals/siopt/29-1/M112174.html>

Funding: The author was partially supported by the Einstein Foundation, Berlin.

†Technische Universität Berlin, Institut für Mathematik, Sekretariat MA 3-2, Straße des 17. Juni 136, 10623, Berlin, Germany (erguer@math.tu-berlin.de).

To proceed we need to introduce some notation. We denote the cone of nonnegative degree $2d$ forms in n variables by $Pos_{n,2d}$:

$$Pos_{n,2d} := \{f \in P_{n,2d} : f(x) \geq 0 \text{ for all } x \in S^{n-1}\}.$$

It is more convenient to work with a compact base of $Pos_{n,2d}$ instead of the unbounded cone itself. To this end we define the following convex body:

$$\widetilde{Pos}_{n,2d} := \left\{ f \in Pos_{n,2d} : \int_{S^{n-1}} f(x) \sigma(x) = 1 \right\},$$

where σ is the uniform measure on the sphere with $\sigma(S^{n-1}) = 1$. Approximating $\widetilde{Pos}_{n,2d}$ with polytopes is equivalent to approximating $Pos_{n,2d}$ with polyhedral cones. Hence, in the rest of this note we are concerned with the polytope approximations to $\widetilde{Pos}_{n,2d}$. We begin with the familiar example of the cone of positive semidefinite matrices (the PSD cone).

Example 1.1 (the cone of nonnegative quadratic forms). For the $d = 1$ case, $Pos_{n,2}$ is the cone of nonnegative quadratics, or equivalently the PSD cone. In this case the base $\widetilde{Pos}_{n,2}$ can be expressed as follows:

$$\widetilde{Pos}_{n,2} := \{Q \in Pos_{n,2} : \text{Tr}(Q) = n\},$$

where Tr denotes the trace of a matrix. Now suppose that K is a polytope with the following property:

$$\widetilde{Pos}_{n,2} - \mathbb{I}_n \subseteq K \subseteq c(\widetilde{Pos}_{n,2} - \mathbb{I}_n),$$

where $c > 1$ is an absolute constant. Then, it follows from the work of Pokutta et al. [12, 13] that K has exponentially many facets in n .

A spectrahedron is the intersection of the PSD cone with an affine linear space. Semidefinite programming methods optimize a linear objective function over a spectrahedron (see, for instance, Chapter 2 of [9]). The well-established method of linear programming efficiently optimizes a linear objective function over a polyhedron. The example above provides a comparison of the expressive power of semidefinite programming versus linear programming. This was one of motivations for Pokutta and his collaborators in their work on the approximation limits of linear programming [12, 13].

Our first main theorem provides an inapproximability result for arbitrary degree d . Our proof is direct and simple, and it is based on some basic Gaussian concentration inequalities.

THEOREM 1.2. *Let $c > 1$ be a constant, let $r = (x_1^2 + x_2^2 + \cdots + x_n^2)^d$, and suppose K is a polytope with the following property:*

$$\widetilde{Pos}_{n,2d} - r \subseteq K \subseteq c^d (\widetilde{Pos}_{n,2d} - r).$$

Then, K has at least

$$a_0 e^{a_1 \frac{n}{cd}}$$

many facets, where a_0 and a_1 are absolute constants.

In Theorem 1.2 we assume d is any fixed degree, and one is at liberty to vary c to be any real number greater than 1. The constants a_0 and a_1 are some fixed numbers independent of n , d , and c . For instance, if one sets $c = 5$, the conclusion of

Theorem 1.2 is that any polyhedral approximation to $\widetilde{Pos}_{n,2d}$ with accuracy 5^d has to have $\Omega(\exp(\frac{n}{5d}))$ many facets.

After the inapproximability result for the dense case, we consider polyhedral approximations for structured subspaces of polynomials. For a subspace $E \subset P_{n,2d}$, we respectively define the cone of nonnegative elements and its compact base as

$$Pos_E := \{f \in E : f(x) \geq 0 \text{ for all } x \in S^{n-1}\},$$

$$\widetilde{Pos}_E := \left\{f \in Pos_E : \int_{S^{n-1}} f(x) \sigma(x) = 1\right\}.$$

We will explain in the third section that for the \widetilde{Pos}_E definition to be meaningful we need to assume $r = (x_1^2 + x_2^2 + \cdots + x_n^2)^d \in E$. We would like to present a natural example of such a subspace which is due to Choi, Lam, and Reznick [17].

Example 1.3 (even symmetric sextics). Let $M_k(x) = \sum_{i=1}^n x_i^k$, and consider the following vector space E :

$$E := \text{span}\{M_6, M_2M_4, M_2^3\},$$

$E \subset P_{n,6}$, and it is the subspace formed by even symmetric forms. Observe that E satisfies our assumption that $r \in E$ since $M_2^3 = (x_1^2 + x_2^2 + \cdots + x_n^2)^3 = r$. It follows from the results of Choi, Lam, and Reznick [17] that \widetilde{Pos}_E is precisely a regular n -gon.

In the beautiful paper [17] the authors exploit algebraic properties of even symmetric sextics, and it is not clear how to generalize their techniques to less structured families of sparse polynomials. In this paper, we consider arbitrary subspaces of forms with only the assumption that the element r is included the subspace. A corollary of our main theorem below is that for any fixed $m \geq 3$ and $n \rightarrow \infty$, all m -dimensional sections of $Pos_{n,2d}$ that include the element r are roughly polyhedral cones with $O(n^{m-2})$ many facets.

THEOREM 1.4. *Let $E \subset P_{n,2d}$ be a linear subspace of fixed dimension m , where $3 \leq m \leq \frac{n}{e}$. Also assume that $r \in E$. Then, there exists a polytope K with $O(n^{m-2})$ many facets which satisfies the following:*

$$\widetilde{Pos}_E - r \subseteq K \subseteq \left(1 + \frac{n}{m}\right)^{\frac{3m}{n}} (\widetilde{Pos}_E - r).$$

For the special case of quadratics, Theorem 1.4 yields the following corollary.

COROLLARY 1.5. *Let E be an m -dimensional affine linear space of real $n \times n$ matrices with $m \leq \frac{n}{e}$ and $\mathbb{I}_n \in E$. We denote the spectrahedron defined by the intersection of the PSD cone with E by Pos_E . We define the following base \widetilde{Pos}_E for the spectrahedron Pos_E :*

$$\widetilde{Pos}_E := \{Q \in Pos_E : \text{Tr}(Q) = n\}.$$

Then, there exists a polytope K with $O(n^{m-2})$ facets which satisfies the following inclusions:

$$\widetilde{Pos}_E - \mathbb{I}_n \subseteq K \subseteq \left(1 + \frac{n}{m}\right)^{\frac{3m}{n}} (\widetilde{Pos}_E - \mathbb{I}_n).$$

Even though we stated Theorem 1.4 as there exists a polytope K satisfying the desired approximation, we actually prove existence of a polytope K where all the

facets are defined by single pointwise evaluations. In this respect, our construction relates to a basic question: Given an $\varepsilon > 0$ and $f \in E$ with $\int_{S^{n-1}} f(x) \sigma(x) > 0$, how many pointwise evaluations are needed to certify that $f(x) \geq -\varepsilon(\int_{S^{n-1}} f(x) \sigma(x))$ for all $x \in S^{n-1}$? Theorem 1.4 gives an estimate for $\varepsilon = (1 + \frac{n}{m})^{\frac{3m}{n}} - 1$.

The proof of Theorem 1.4 exploits convex geometric properties of Pos_E through spectral sparsification. In particular, we use results of Friedland and Youssef [15] which are based on the recent solution of the Kadison–Singer problem. The solution of the Kadison–Singer problem and results based on it (including Theorem 1.4) are not constructive. In that respect, we also study the random construction of a polyhedral approximation. To state our random approximation result we need a little more terminology. For all $p \geq 1$, we define the p -norm of a homogenous polynomial f as follows:

$$\|f\|_p := \left(\int_{S^{n-1}} |f(v)|^p \sigma(v) \right)^{\frac{1}{p}},$$

where σ is the uniform measure on S^{n-1} . We also define the following quantity for subspaces E of $P_{n,2d}$:

$$M(E) := \max_{f \in E} \frac{\|f\|_2^2}{\|f\|_1}.$$

Now we are ready to state the random polyhedral approximation theorem.

THEOREM 1.6. *Let $E \subset P_{n,2d}$ be a subspace with $r \in E$ and $\dim(E) = m$. Let μ be the measure defined in the last section of this paper. For a given α with $0 < \alpha < 1$, we set $t = O(m^2(\frac{M(E)}{1-\alpha})^m \ln(\frac{M(E)}{1-\alpha}))$. Let x_1, x_2, \dots, x_t be independent random vectors in E distributed according to μ . We define a polytope K_α as follows:*

$$K_\alpha = \{y \in E : \langle y, r \rangle = 0 \text{ and } \langle y, x_i \rangle \geq -1 \text{ for all } 1 \leq i \leq t\}.$$

Then K_α satisfies the inclusions

$$\widetilde{Pos_E} - r \subseteq K_\alpha \subseteq \frac{1}{\alpha} (\widetilde{Pos_E} - r)$$

with probability at least $1 - 4(\frac{1-\alpha}{M(E)})^{m^2}$.

Bounding $M(E)$ from above seems to require more information than just the dimension of the subspace. However, it is easy to prove $M(E) \leq 2^{2d}$ for any subspace E of $P_{n,2d}$ (see Corollary 6.6).

As a special case of Theorem 1.6, consider $2d = 4$ and $\alpha = \frac{n-1}{n}$. Then, $K_{\frac{n-1}{n}}$ has $O(m^2(16n)^m \ln(n))$ many facets and it satisfies

$$\widetilde{Pos_E} - r \subseteq K_{\frac{n-1}{n}} \subseteq \frac{n}{n-1} (\widetilde{Pos_E} - r)$$

with probability greater than $1 - 4n^{-m^2}$.

The rest of the paper is structured as follows: In the second section, we review the background material coming from geometric functional analysis. In the third section, we discuss convex geometric properties of the cone of nonnegative polynomials. In the fourth section, we prove Theorem 1.2 using Gaussian concentration inequalities and convex geometric duality introduced in the third section. In the fifth section, we prove Theorem 1.4 using the tools introduced in the second and the third sections. Finally, we prove Theorem 1.6 in the last section using a tool coming from computational geometry, namely the ε -net theorem.

2. John's theorem and spectral sparsification. We begin by recalling a fundamental theorem in convex geometry due to Fritz John [19].

THEOREM 2.1 (see John's theorem). *Every convex body $K \subset \mathbb{R}^n$ is contained in a unique ellipsoid of the minimal volume E_{\min} . Moreover,*

$$\frac{1}{n}E_{\min} \subset K \subset E_{\min}.$$

The minimal volume ellipsoid E_{\min} is the Euclidean unit ball B_2^n if and only if the following conditions are satisfied: $K \subset B_2^n$, there are unit vectors $(u_i)_{i=1}^m$ on the boundary of K and positive real numbers c_i such that

$$\sum_{i=1}^m c_i u_i = 0,$$

and for all $x \in \mathbb{R}^n$ we have

$$\sum_i c_i \langle u_i, x \rangle^2 = \|x\|_2^2.$$

If the minimal volume ellipsoid of a convex body K is the unit ball, we say K is in John's position. Henceforth, we call the conditions in Theorem 2.1 characterizing John's position as John's decomposition. One way to view John's decomposition is to observe that the family of unit vectors $\{u_i\}_{i=1}^m$ work like an orthogonal basis in \mathbb{R}^n . This phenomenon is studied in depth in frame theory. The following lemma is included here for completeness, and it can be found in any frame theory textbook.

LEMMA 2.2. *We denote the map that sends x to $\langle x, y \rangle z$ by $y \otimes z$. Then the following are equivalent.*

1.

$$\mathbb{I}_n = \sum_i c_i u_i \otimes u_i.$$

2. *For every $x \in \mathbb{R}^n$,*

$$x = \sum_i c_i \langle x, u_i \rangle u_i.$$

3. *For every $x \in \mathbb{R}^n$,*

$$\sum_i c_i \langle u_i, x \rangle^2 = \|x\|_2^2.$$

Another perspective on John's decomposition is to view the decomposition as a discrete measure supported on the vectors u_i with weights c_i , and the identity as being the covariance matrix of the measure. This measure theoretic interpretation is formalized in the notion of isotropic measures, which we present below.

DEFINITION 2.3. *A finite Borel measure Z on the sphere S^{n-1} of an n -dimensional real vector space V is said to be isotropic if*

$$\|x\|_2^2 = \int_{S^{n-1}} \langle x, u \rangle^2 dZ(u)$$

for all $x \in V$. Moreover, we define the centroid of a measure Z supported on the sphere S^{n-1} as

$$\frac{1}{Z(S^{n-1})} \int_{S^{n-1}} u Z(u).$$

We say the measure is centered at 0 if the centroid is the origin.

An isotropic measure supported on the sphere with centroid 0 is the continuous analog of John's decomposition. It is known that a convex body is in John's position if and only if the touching points of the convex body to the unit ball are the support of an isotropic measure with centroid 0 [14]. The advantage of this continuous point of view is that interesting convex bodies such as the convex hull and Minkowski sum of compact group orbits, or the dual of the cone of nonnegative polynomials, are easily shown to support an isotropic measure with their (possibly) infinitely many touching points to the unit ball.

In general, convex bodies with fewer than $\frac{n(n+3)}{2}$ touching points to their minimal volume ellipsoid form a dense family in the space of convex bodies [16]. The $O(n^2)$ many touching points in this dense family bring the problem of sparsification in John's decomposition. The main goal of this approach is to find $O(n)$ many vectors among the initial decomposition which form an approximate decomposition of identity. This line of reasoning is closely related to the recent solution of the Kadison–Singer problem; we begin our discussion with a remarkable theorem of Batson, Spielman, and Srivastava, which was a precursor to the solution of the Kadison–Singer problem (see Theorem 1.6 of [31]).

THEOREM 2.4. *Fix $\varepsilon \in (0, 1)$ and $n, m \in \mathbb{N}$. For every $x_1, x_2, \dots, x_m \in \mathbb{R}^n$ there exist $s_1, s_2, \dots, s_m \in [0, \infty)$ such that*

$$\#\{s_i : s_i \neq 0\} \leq \left\lceil \frac{n}{\varepsilon^2} \right\rceil$$

and for all $y \in \mathbb{R}^n$ we have

$$(1 - \varepsilon)^2 \sum_{i=1}^m \langle x_i, y \rangle^2 \leq \sum_{i=1}^m s_i \langle x_i, y \rangle^2 \leq (1 + \varepsilon)^2 \sum_{i=1}^m \langle x_i, y \rangle^2.$$

In particular, for the $\sum_{i=1}^m x_i \otimes x_i = \mathbb{I}_n$ case we have

$$(1 - \varepsilon)^2 \mathbb{I}_n \preceq \sum_{i=1}^m s_i x_i \otimes x_i \preceq (1 + \varepsilon)^2 \mathbb{I}_n.$$

Theorem 2.4 was recently refined in an article of Friedland and Youssef [15]. Friedland and Youssef's work uses the solution of the Kadison–Singer problem as an intermediate step and then provides refined estimates on a suite of problems including spectral sparsification, restricted invertibility, and the isomorphic Dvoretzky problem. The following result (Theorem 4.1 of [15]) will be used in the fifth section.

THEOREM 2.5. *There exists a universal constant c such that the following holds. Let $\varepsilon > 0$ and let $\{c_i, x_i\}_{i=1}^m$ be a John decomposition of identity in \mathbb{R}^n (i.e., x_i are unit vectors and $c_i > 0$). Then there exists a multiset σ of indices from $[m]$ of size at most $\frac{n}{c\varepsilon^2}$ such that*

$$(1 - \varepsilon) \mathbb{I}_n \preceq \frac{n}{|\sigma|} \sum_{i \in \sigma} (x_i - u) \otimes (x_i - u) \preceq (1 + \varepsilon) \mathbb{I}_n,$$

where $u = \frac{1}{|\sigma|} \sum_{i \in \sigma} x_i$ satisfies $\|u\| \leq \frac{2\varepsilon}{3\sqrt{n}}$.

3. Convex geometry of nonnegative polynomials. In this section, we would like to introduce a modern convex geometry point of view on nonnegative polynomials. Our plan is to first introduce the concepts and results in the dense case, and then

write the implications for the sparse case in a separate section. Most of the results in this section have appeared in the literature [4, 8] with possibly different proofs.

We start by defining an inner product on $P_{n,2d}$. For $f, g \in P_{n,2d}$, the inner product $\langle f, g \rangle$ is defined as follows:

$$\langle f, g \rangle := \int_{S^{n-1}} f(x)g(x) \sigma(x),$$

where σ is the uniform measure on the sphere S^{n-1} . It must be clear that the norm introduced by this inner product is the 2-norm defined in the introduction. Throughout the paper this norm will be denoted by $\|\cdot\|_2$.

We consider the action of $SO(n)$ on $P_{n,2d}$. For $T \in SO(n)$ and $f \in P_{n,2d}$, we denote the result of the action of T on f by $T \circ f$, and the action is defined by pointwise evaluations on $x \in S^{n-1}$ as follows:

$$T \circ f(x) := f(T^{-1}x).$$

Since knowing all pointwise evaluations on the sphere uniquely defines the homogenous polynomial $T \circ f$, this action is well defined. For any $f, g \in P_{n,2d}$, we clearly have

$$\langle T \circ f, T \circ g \rangle = \langle f, g \rangle.$$

Hence, the inner product is $SO(n)$ invariant. Now, we consider pointwise evaluation maps on the vector space $P_{n,2d}$. Let $v \in S^{n-1}$ and consider the following map:

$$l_v : P_{n,2d} \rightarrow \mathbb{R}, \quad l_v(f) = f(v).$$

The operator norm of l_v is defined as follows:

$$\|l_v\| := \max_{\|f\|=1} |l_v(f)| = \max_{\|f\|=1} |f(v)|.$$

For any two arbitrary unit vectors $u, v \in S^{n-1}$, one can find $T \in SO(n)$ such that $T(u) = v$. Then for all $f \in P_{n,2d}$, one has $T \circ f(v) = f(u)$. Since $\|\cdot\|_2$ is invariant under the $SO(n)$ action, we immediately have $\|l_v\| = \|l_u\|$. Hence, the operator norm $\|l_v\|$ is fixed for all $v \in S^{n-1}$.

We introduced a Hilbert space structure on $P_{n,2d}$, so we have the Riesz representation theorem. That is, for all $v \in S^{n-1}$ there exists a corresponding unique $p_v \in P_{n,2d}$ such that for all $f \in P_{n,2d}$, we have

$$l_v(f) = f(v) = \langle f, p_v \rangle.$$

Since $\|l_v\| = \|p_v\|_2$, and since the norm of l_v is fixed over the sphere, we conclude that $\|p_v\|_2$ is fixed over the sphere as well.

The results of this section are basis independent; the polynomials p_v only depend on the inner product. In the lemma below, we write a concrete expansion of p_v for an arbitrary orthonormal basis and derive some basic properties.

LEMMA 3.1. *Let $u_1, u_2, \dots, u_N \in P_{n,2d}$ be an orthonormal basis for $P_{n,2d}$, where $N = \binom{n+d-1}{d}$. For every $v \in S^{n-1}$, we define the following polynomial p_v :*

$$p_v(x) := \sum_{i=1}^N u_i(v)u_i(x).$$

Then p_v possesses the following properties:

1. For all $q \in P_{n,2d}$, we have

$$\langle q, p_v \rangle = q(v)$$

and

$$\frac{\max_{v \in S^{n-1}} |q(v)|}{\|q\|_2} \leq \sqrt{N}.$$

2. For $v, w \in S^{n-1}$ and $T \in SO(n)$ we have

$$p_v(w) = p_w(v) = p_{T(v)}(T(w)).$$

3. For all $v \in S^{n-1}$ we have

$$N = p_v(v) = \|p_v\|^2.$$

Proof. Given $q \in P_{n,2d}$ we have

$$q(v) = \sum_{i=1}^N \langle q, u_i \rangle u_i(v) = \left\langle q, \sum_i u_i(v) u_i \right\rangle = \langle q, p_v \rangle.$$

The second part of the first claim will directly follow from the Cauchy–Schwarz inequality once we establish the third claim. We continue with the proof of second claim: For any $T \in SO(n)$ and any $f \in P_{n,2d}$, we have

$$\langle f, p_{T(w)} \rangle = f(T(w)) = \langle T^{-1} \circ f, p_w \rangle = \langle f, T \circ p_w \rangle.$$

Since f is arbitrary, this proves $p_{T(w)} = T \circ p_w$ and completes the proof of the second claim. Now we would like to show that $p_v(v) = N$ for all $v \in S^{n-1}$. Since $p_v(v) = p_w(w)$ for all $v, w \in S^{n-1}$, we have

$$p_v(v) = \int_{S^{n-1}} p_v(v) \sigma(v) = \int_{S^{n-1}} \langle p_v, p_v \rangle \sigma(v).$$

Expanding the rightmost equation, we have

$$p_v(v) = \int_{S^{n-1}} \left\langle \sum_i u_i(v) u_i, \sum_i u_i(v) u_i \right\rangle \sigma(v) = \sum_i \int_{S^{n-1}} u_i(v)^2 \sigma(v) = N. \quad \square$$

For any nonnegative polynomial $p \in Pos_{n,2d}$ we have

$$\int_{S^{n-1}} p(x) \sigma(x) = \langle p, r \rangle \geq 0,$$

where $r = (x_1^2 + x_2^2 + \cdots + x_n^2)^d$. We observe that

$$\widetilde{Pos}_{n,2d} = \{p \in Pos_{n,2d} : \langle p, r \rangle = 1\}.$$

Thus whenever $f \in \widetilde{Pos}_{n,2d}$ we have $r \perp (f - r)$. So we define the subspace of polynomials that are orthogonal to r as follows:

$$U := \left\{ f \in P_{n,2d} : \langle f, r \rangle = \int_{S^{n-1}} f(x) \sigma(x) = 0 \right\}.$$

In words, U is the subspace of mean zero polynomials. Now we define a map from S^{n-1} to $P_{n,2d}$ as follows:

$$\phi : S^{n-1} \rightarrow P_{n,2d}, \quad \phi(v) := p_v - r.$$

Observe that for all $v \in S^{n-1}$ we have $\|\phi(v)\|_2 = \sqrt{N-1}$. Moreover, we have

$$\langle \phi(v), r \rangle = \langle p_v - r, r \rangle = 0.$$

Hence $\phi(S^{n-1}) \subset U$. Now let σ be the uniform measure on S^{n-1} and let μ be the pushforward measure of σ under ϕ . For all $q \in U$ we have the following equality:

$$\|q\|^2 = \int_{S^{n-1}} q(v)^2 \sigma(v) = \int_{S^{n-1}} \langle q, p_v \rangle^2 \sigma(v) = \int_{S^{n-1}} \langle q, \phi(v) \rangle^2 \sigma(v).$$

By definition of the pushforward measure we have

$$\|q\|^2 = \int_{S^{n-1}} \langle q, \phi(v) \rangle^2 \sigma(v) = \int_U \langle q, u \rangle^2 \mu(u).$$

Therefore we observe that μ is an isotropic measure supported on the $\sqrt{N-1}$ scaled sphere of U . Hence μ creates a decomposition of identity!

In order to view the support of μ as a John decomposition we also need to compute its centroid:

$$q = \int u \mu(u) = \int_{S^{n-1}} (p_v - r) \sigma(v).$$

By construction, q is invariant under the action of $SO(n)$. Therefore we have $q = a(x_1^2 + x_2^2 + \cdots + x_n^2)^d$ for some $a \in \mathbb{R}$. Since $q \in U$, we deduce that $a = 0$. Hence the measure μ is centered at the origin. Now we define the body of pointwise evaluations:

$$B := \text{conv}(\text{Im}(\phi)).$$

We observed that $\frac{1}{\sqrt{N-1}}B$ is the convex hull of the support of an isotropic measure μ which has its centroid at 0. It immediately follows from the discussion in the previous section that $\frac{1}{\sqrt{N-1}}B$ is in John's position. Now we consider the dual convex body B° :

$$B^\circ = \{q \in U : \langle q, p \rangle \leq 1 \text{ for all } p \in B\} = \{q \in U : \langle q, p_v - r \rangle \leq 1 \text{ for all } v \in S^{n-1}\}.$$

By definition of U , we have $\langle q, p_v - r \rangle = \langle q, p_v \rangle = q(v)$. This gives the following equivalence:

$$-B^\circ + r = \widetilde{Pos}_{n,2d}.$$

This nice convex geometric duality allows us to approximate $\widetilde{Pos}_{n,2d}$ by approximating B .

Remark 3.1. Readers who incline more toward algebraic geometry can think of B as the convex hull of the d th Veronese embedding.

3.1. Structured polynomials. Let $E \subset P_{n,2d}$ be a linear subspace and let $\dim(E) = m$ and $r = (x_1^2 + x_2^2 + \cdots + x_n^2)^d \in E$. Recall the definition of Pos_E :

$$Pos_E := \{f \in E : f(x) \geq 0 \text{ for all } x \in S^{n-1}\}.$$

We use the restriction of the inner product on $P_{n,2d}$ to E . For all $f \in \widetilde{Pos_E}$, we trivially have $\langle f, r \rangle \geq 0$. Now we recall the definition of the compact base $\widetilde{Pos_E}$:

$$\widetilde{Pos_E} := \{f \in Pos_E : \langle f, r \rangle = 1\}.$$

Let Π_E denote the orthogonal projection map on E . Following the notation in Lemma 3.1, for $f \in E$ we have

$$f(v) = \langle f, p_v \rangle = \langle f, \Pi_E(p_v) \rangle.$$

For instance, $r \in E$ and this implies $1 = r(v) = \langle r, p_v \rangle = \langle r, \Pi_E(p_v) \rangle$. Using this observation we define a map ϕ_E as follows:

$$\phi_E : S^{n-1} \rightarrow E, \quad \phi_E(v) := \Pi_E(p_v - r) = \Pi_E(p_v) - r.$$

A first observation is that $\langle \phi_E(v), r \rangle = 0$. So we define $U(E)$, the subspace of polynomials orthogonal to r , as follows:

$$U(E) := \{f \in E : \langle f, r \rangle = 0\}.$$

Hence, $\dim(U(E)) = m - 1$ and $\phi_E(S^{n-1}) \subset U(E)$. The second observation is that $\|\phi_E(v)\|_2 = (\|\Pi_E(p_v)\|_2^2 - 1)^{\frac{1}{2}}$. This observation shows that $\|\phi_E(v)\|_2$ can change at every point $v \in S^{n-1}$ in contrast to the situation in the previous section. This change in the norm can occur because E is not necessarily closed under the $SO(n)$ action, and certain directions on the sphere are preferred over others by the structure of the subspace E .

We would like to continue with the isotropic measure observation of the previous section. We define μ_E to be the pushforward measure of σ (the uniform measure on S^{n-1}) to E under the map ϕ_E . Now, for all $q \in E$ we have

$$\|q\|^2 = \int_{S^{n-1}} \langle q, \phi_E(v) \rangle^2 \sigma(v) = \int_E \langle q, u \rangle^2 \mu_E(u).$$

Therefore, μ_E is an isotropic measure. Proving that μ_E has centroid at 0 is also easy. Even though we lost the control on the norms of $\|\phi_E(v)\|_2$, μ_E being an isotropic measure has the following immediate consequence (which can be seen by taking the trace of the covariance matrix of μ_E):

$$\int_{S^{n-1}} \|\phi_E(v)\|_2^2 \sigma(v) = m - 1.$$

Now, we define the body of pointwise evaluations:

$$B(E) := \text{conv}\{\phi_E(v) : v \in S^{n-1}\}.$$

For any $f \in E$ with $\langle f, r \rangle = 0$ and for all $v \in S^{n-1}$ we have the following relation:

$$f(v) \geq -1 \Leftrightarrow \langle f, p_v \rangle \geq -1 \Leftrightarrow \langle f, \phi_E(v) \rangle \geq -1.$$

Also note that $q \in \widetilde{Pos_E}$ if and only if $\langle q - r, r \rangle = 0$ and $(q - r)(v) \geq -1$ for all $v \in S^{n-1}$. Hence, we conclude that

$$\widetilde{Pos_E} = -B(E)^\circ + r.$$

Remark 3.2. One of the reasons that isotropic measures are appealing is a theorem of Rudelson. Let $x \in \mathbb{R}^m$ be an isotropic random vector. Let x_1, x_2, \dots, x_M be independent copies of x . Then, Rudelson's remarkably general theorem [29] states the following:

$$\mathbb{E} \left\| \frac{1}{M} \sum_{i=1}^M x_i \otimes x_i - \mathbb{I}_n \right\| \leq C \sqrt{\frac{\log(m)}{M}} (\mathbb{E} \|x_i\|_2^{\log M})^{\frac{1}{\log M}}.$$

Hence, if one has any control on $\max_{v \in S^{n-1}} \|\phi_E(v)\|_2$, Rudelson's theorem provides a randomized way to obtain an approximate decomposition of identity, and that is all that is needed for the construction of a polyhedral approximation to $B(E)$.

4. Limits of approximation with few facets. We have established the convex geometric duality between $\widetilde{Pos}_{n,2d}$ (the section of the cone of nonnegative polynomials) and B (the convex body of pointwise evaluations). Thanks to this duality, searching for a polytope Q with few facets that is sandwiched between $\frac{1}{c}\widetilde{Pos}_{n,2d}$ and $\widetilde{Pos}_{n,2d}$ (for some constant $c > 1$) is equivalent to searching for a polytope P with few vertices that is sandwiched between B and cB . In this section, we show that for any constant $c > 1$, a polytope P satisfying

$$B \subseteq P \subseteq c^d B$$

has to have exponentially many vertices in terms of n .

Our result in this section is a direct application of basic properties of the Gaussian measure. Similar inapproximability results for the special case of quadratics were obtained by Pokutta et al. with a completely different approach [12, 13].

We start by presenting two facts about the Gaussian measure that are going to be used in our proof. The first fact is a tail bound for polynomial maps on Gaussian random variables.

LEMMA 4.1. *Let f be a homogenous degree $2d$ polynomial with $\int_{S^{n-1}} f(x) \sigma(x) = 0$, and let $\gamma_n \sim \mathcal{N}(0, \mathbb{I})$ be the standard Gaussian measure on \mathbb{R}^n . Then, for all $t \geq \sqrt{n+2d}$ we have*

$$\gamma_n(\{x : |f(x)| \geq t^{2d} \|f\|_2\}) \leq a_0 \exp\left(-a_1 \frac{t^2}{n+2d}\right),$$

where a_0 and a_1 are positive absolute constants.

Proof. We start by presenting a standard tail estimate for polynomials with normal random variables (see, for instance, Corollary 5.5.7 in [10]). We denote the standard Gaussian measure on \mathbb{R}^n by γ_n . Then, for a polynomial f with $\int_{S^{n-1}} f(x) \sigma(x) = 0$ the standard tail bound reads as follows:

$$\gamma_n\left(\left\{|f(x)| \geq s^{2d} \left(\int_{\mathbb{R}^n} f(x)^2 \gamma_n(x)\right)^{\frac{1}{2}}\right\}\right) \leq a_0 e^{-a_1 s^2},$$

where a_0, a_1 are absolute constants. Now we just need to rewrite this estimate with the norm $\|\cdot\|_2$ of this paper. Let us recall a basic integral identity:

$$\int_{S^{n-1}} f(x)^2 \sigma(x) = \frac{\Gamma(\frac{n}{2})}{2^{2d} \Gamma(\frac{n}{2} + 2d)} \int_{\mathbb{R}^n} f(x)^2 \gamma_n(x).$$

Therefore, we have

$$\left(\int_{\mathbb{R}^n} f(x)^2 \gamma_n(x) \right)^{\frac{1}{2}} = 2^d \|f\|_2 \left(\frac{\Gamma(\frac{n}{2} + 2d)}{\Gamma(\frac{n}{2})} \right)^{\frac{1}{2}} \leq \|f\|_2 (n + 2d)^d.$$

Hence, for all $s \geq 1$ we have

$$\begin{aligned} \gamma_n \left(\{ |f(x)| \geq s^{2d} (n + 2d)^d \|f\|_2 \} \right) &\leq \gamma_n \left(\left\{ |f(x)| \geq s^{2d} \left(\int_{\mathbb{R}^n} f(x)^2 \gamma_n(x) \right)^{\frac{1}{2}} \right\} \right) \\ &\leq a_0 e^{-a_1 s^2}. \end{aligned}$$

Setting $t^2 = s^2(n + 2d)$ completes the proof. \square

The second fact we are going to use is a standard tail bound for the norm of a Gaussian vector (see, for instance, [3]).

LEMMA 4.2. *Let v be a random vector distributed according to the standard Gaussian measure $\gamma_n \sim \mathcal{N}(0, 1)$ on \mathbb{R}^n . Then,*

$$\gamma_n \{ \|v\|_2^2 \leq (1 - \varepsilon)n \} \leq e^{-\frac{\varepsilon^2 n}{4}}.$$

Now we are ready to present the main result of this section.

THEOREM 4.3. *Let B be the body of pointwise evaluations in $P_{n,2d}$ as defined in the previous section, let $c > 1$ be a constant, and assume there exists a polytope $P = \text{conv}\{f_i : 1 \leq i \leq N\}$ with the following property:*

$$B \subseteq P \subset c^d B.$$

Then, we have

$$|N| \geq a_0 e^{a_1 \frac{n}{cd}},$$

where a_0 and a_1 are absolute constants.

Proof. We use an idea of Barvinok, which is to study the maxima of a linear functional on the polytope to bound the number of its vertices [2]. Our linear functional will be the pointwise evaluation map l_v at a random Gaussian vector $v \in \mathbb{R}^n$. We aim to use the known properties of the convex body B to arrive at a lower bound on the number of vertices of P . We define a symmetric polytope \tilde{P} for convenience:

$$\tilde{P} := \text{conv}\{f_i, -f_i : 1 \leq i \leq N\}.$$

We observe that

$$B \subseteq \tilde{P} \subseteq c^d \binom{n + 2d - 1}{2d}^{\frac{1}{2}} B_2^{m-1}.$$

Note that here the unit ball is defined with respect to the $\|\cdot\|_2$ norm introduced by the inner product. Using the fact that the maximum of a linear functional is attained at the vertices of a polytope and performing a basic union bound, we have the following inequality:

$$\mathbb{P} \left\{ v : \max_{f \in \tilde{P}} f(v) \geq \tau \right\} = \mathbb{P} \left\{ v : \max_{f \in P} |f(v)| \geq \tau \right\} \leq |N| \max_{1 \leq i \leq N} \mathbb{P} \{ v : |f_i(v)| \geq \tau \}.$$

For any vector $v \in \mathbb{R}^n$, we define $\tilde{v} = \frac{v}{\|v\|}$. Now using the properties of $p_{\tilde{v}}$ from Lemma 3.1, we have the following estimate from below:

$$\max_{f \in B} f(v) \geq p_v(v) = \|v\|^{2d} p_{\tilde{v}}(\tilde{v}) \geq \binom{n+2d-1}{2d} \|v\|^{2d}.$$

Therefore, we have the following lower bound:

$$\begin{aligned} \mathbb{P}\left\{v : \max_{f \in P} |f(v)| \geq \tau\right\} &\geq \mathbb{P}\left\{v : \|v\|^{2d} \geq \tau \binom{n+2d-1}{2d}^{-1}\right\} \\ &= \mathbb{P}\left\{\|v\|_2^2 \geq \tau^{\frac{1}{d}} \binom{n+2d-1}{2d}^{-\frac{1}{d}}\right\}. \end{aligned}$$

For all $f_i \in P$, we have $\|f\|_2 \leq c^d \binom{n+2d-1}{2d}^{\frac{1}{2}}$. Then, Lemma 4.1 gives the following estimate:

$$\mathbb{P}\left\{v : |f_i(v)| \geq t^{2d} \binom{n+2d-1}{2d}^{\frac{1}{2}}\right\} \leq a_0 \exp\left(-a_1 \frac{t^2}{c(n+2d)}\right).$$

We set $\tau = t^{2d} \binom{n+2d-1}{2d}^{\frac{1}{2}}$. Then, the very first inequality in this proof gives us the following lower bound on $|N|$:

$$|N| \geq \frac{\mathbb{P}\{\|v\|_2^2 \geq t^2 \binom{n+2d-1}{2d}^{-\frac{1}{2d}}\}}{\max_{1 \leq i \leq N} \mathbb{P}\{v : |f_i(v)| \geq t^{2d} \binom{n+2d-1}{2d}^{\frac{1}{2}}\}}.$$

We choose $t^2 = \frac{n(n+2d)}{4ed}$, which ensures

$$t^2 \binom{n+2d-1}{2d}^{-\frac{1}{2d}} \leq \frac{n}{2}$$

and also

$$\frac{t^2}{c(n+2d)} = \frac{n}{4ced}.$$

We obtained an inequality for bounding $\mathbb{P}\{v : |f_i(v)| \geq t^{2d} \binom{n+2d-1}{2d}^{\frac{1}{2}}\}$ from above.

We also have Lemma 4.2 to bound $\mathbb{P}\{\|v\|_2^2 \geq t^2 \binom{n+2d-1}{2d}^{-\frac{1}{2d}}\}$ from below. Putting these two inequalities together we obtain the following estimate with $b_0 = a_0^{-1}$ and $b_1 = \frac{a_1}{4e}$:

$$|N| \geq b_0 e^{b_1 \frac{n}{cd}} (1 - e^{-\frac{n}{16}}). \quad \square$$

5. Approximation polytopes. We start this section with a simple observation: Suppose that a convex body K and a polytope P satisfy $P \subset K \subset \alpha P$. Then, for any invertible linear map T we have the following inclusions:

$$T(P) \subset T(K) \subset \alpha T(P).$$

Therefore, to approximate K with polytopes, we can select a suitable linear map T and approximate the “easier” convex body $T(K)$ instead. In this section, we will

prove existence of approximation polytopes to $B(E)$ with few facets. Based on the simple observation above, we assume without loss of generality that $B(E)$ is in John's position. Recall that $B(E)$ lies inside the subspace $U(E) := \{f \in E : \langle f, r \rangle = 0\}$ by construction. Therefore, being in John's position means that we have

$$\frac{1}{m-1}B_2^{m-1} \subseteq B(E) \subseteq B_2^{m-1}$$

and the touching points of $B(E)$ to the unit ball of $U(E)$ form a John decomposition of identity.

The following theorem is a direct corollary of the work of Friedland and Youssef (see Theorem 2.5).

THEOREM 5.1. *There exists a universal constant c so that the following holds. For every $\varepsilon > 0$, there exists a multiset of vectors $S \subseteq S^{n-1}$ with $|S| \leq \frac{m-1}{c\varepsilon^2}$ and corresponding polynomials $q_v \in \partial B(E) \cap B_2^{m-1}$ for all $v \in S$, so that*

$$(1 - \varepsilon)\mathbb{I}_{m-1} \preceq \frac{m-1}{|S|} \sum_{v \in S} q_v \otimes q_v \preceq (1 + \varepsilon)\mathbb{I}_{m-1}$$

and

$$\left\| \frac{1}{|S|} \sum_{v \in S} q_v \right\| \leq \frac{2\varepsilon}{3\sqrt{m}},$$

where \mathbb{I}_{m-1} is the identity map on the subspace $U(E)$ of E .

Proof. If T is the map which puts $B(E)$ in John's position, then one can consider $T(\phi_E(v))$ as the parametrization of $TB(E) \subset \mathbb{R}^{m-1}$ and apply Theorem 2.5 to the John decomposition given by the touching points of $TB(E)$ to the unit ball. \square

Remark 5.1. It is important to notice that the existence of a John decomposition of identity with $\frac{(m-1)(m+2)}{2}$ many vectors is already guaranteed by Fritz John's classical article [19]. The goal of the spectral sparsification is to reduce the order of number of the vectors in the decomposition from quadratic to linear. Hence, all the results in the spectral sparsification literature are effective only for $\frac{1}{\sqrt{m}} \leq \varepsilon \leq 1$.

Remark 5.2. Note that S being a set or a multiset where some elements are counted with multiplicity does not affect the rest of our arguments. Therefore, from this point on we consider S as a set for simplicity.

Using Theorem 5.1, we will first construct a polytope Q with $O(m)$ facets which gives the following rough approximation:

$$\widetilde{Pos}_E - r \subseteq Q \subseteq m^{\frac{3}{2}} \left(\widetilde{Pos}_E - r \right).$$

Then, in the next subsection we will improve the accuracy of approximation by using a "tensorization" trick. We start with a set of unit vectors S with $|S| \leq \frac{m-1}{c\varepsilon^2}$ which satisfies the following:

$$\left(1 - \frac{\varepsilon}{4}\right)\mathbb{I}_{m-1} \prec T = \frac{1}{|S|} \sum_{v \in S} f_v \otimes f_v \prec \left(1 + \frac{\varepsilon}{4}\right)\mathbb{I}_{m-1}.$$

Existence of such a set S is guaranteed by taking $f_v = \sqrt{m-1}q_v$ in Theorem 5.1.

Since $\|T - \mathbb{I}_{m-1}\| \leq \frac{\varepsilon}{4}$, for any $p \in E$ we have the following upper and lower bounds:

$$(1 - \varepsilon) \|p\|^2 \leq \langle Tp, p \rangle \leq (1 + \varepsilon) \|p\|^2.$$

Since $\langle Tp, p \rangle = \frac{1}{|S|} \sum_{v \in S} \langle p, f_v \rangle^2$, the pigeon hole principle implies that there exists a $v \in S$ such that $|\langle f_v, p \rangle| \geq (1 - \varepsilon) \|p\|$. Now, we define our approximation polytope P as follows:

$$P := \{f_v, -f_v : v \in S\}.$$

By the above observation, for all $f \in E$ we have $\max_{q \in P} \langle f, q \rangle \geq (1 - \varepsilon) \|f\|$. This implies $(1 - \varepsilon) B_2^{m-1} \subseteq P$. Since $f_v \in \sqrt{m-1} B(E)$ and $B(E)$ is in John's position, $\|f_v\| \leq \sqrt{m-1}$. In summary, we have

$$(1 - \varepsilon) B_2^{m-1} \subseteq P \subseteq \sqrt{m-1} B_2^{m-1}.$$

This implies the following inclusion by John's theorem:

$$(1 - \varepsilon) B(E) \subseteq P \subseteq m^{\frac{3}{2}} B(E)$$

Taking the duals of all sides gives the following:

$$m^{-\frac{3}{2}} B(E)^\circ \subseteq P^\circ \subseteq (1 - \varepsilon)^{-1} B(E)^\circ.$$

Setting $Q = -m^{\frac{3}{2}} P^\circ$ and using $-B(E)^\circ = \widetilde{Pos}_E - r$, we have the following inclusions:

$$\widetilde{Pos}_E - r \subseteq Q \subseteq (1 + 2\varepsilon) m^{\frac{3}{2}} (\widetilde{Pos}_E - r).$$

Note that, by construction, P has $O(m)$ vertices and Q has $O(m)$ facets.

5.1. Improved accuracy with more facets. In this section we will use a standard construction from multilinear algebra, namely the tensor power of a vector space. Tensor powers will help us produce approximation polytopes with more facets and improved accuracy. Now, let V be a vector space equipped with inner product $\langle \cdot, \cdot \rangle$. For an integer $k \geq 1$, the k th tensor power of V is defined as follows:

$$V^{\otimes k} = \underbrace{V \otimes V \otimes V \otimes \cdots \otimes V}_{k \text{ times}}.$$

It is natural to consider $V^{\otimes k}$ with the following inner product:

$$\langle x_1 \otimes x_2 \otimes \cdots \otimes x_k, y_1 \otimes y_2 \otimes \cdots \otimes y_k \rangle = \prod_{i=1}^k \langle x_i, y_i \rangle.$$

The symmetric part of $V^{\otimes k}$, $Sym(V^{\otimes k})$ is the subspace spanned by tensors $x^{\otimes k} = x \otimes x \otimes \cdots \otimes x$. We consider $Sym(E^k)$, the symmetric part of the k th tensor power of the polynomial space E . Recall that we defined a subspace $U(E) := \{f \in E : \langle f, r \rangle = 0\}$. Note that $\dim(U(E)) = m - 1$, and $\dim(Sym(U(E)^{\otimes k})) = \binom{m+k-2}{k}$. After this line, we write U for $U(E)$, hoping that no confusion will arise. We define the following object in $Sym(U^{\otimes k})$:

$$B(E)^{\otimes k} := \text{conv}\{\phi_E(v)^{\otimes k} : v \in S^{n-1}\}.$$

We assume $B(E)^{\otimes k}$ is in John's position and repeat the reasoning presented in the previous section. This proves existence of a set $S_k \subseteq S^{n-1}$ with $|S_k| \leq \binom{m+k-2}{k} c\varepsilon^{-2}$ such that the following inequalities are satisfied for all $q^{\otimes k} \in U^{\otimes k}$:

$$(1 - \varepsilon) \max_{g^{\otimes k} \in B(E)^{\otimes k}} \langle g^{\otimes k}, q^{\otimes k} \rangle \leq \max_{v \in S_k} |\langle f_v^{\otimes k}, q^{\otimes k} \rangle| \\ \leq \binom{m+k-2}{k}^{\frac{3}{2}} \max_{g^{\otimes k} \in B(E)^{\otimes k}} \langle g^{\otimes k}, q^{\otimes k} \rangle.$$

We can also write these inequalities as follows:

$$(1 - \varepsilon) \max_{g^{\otimes k} \in B(E)^{\otimes k}} \langle g, q \rangle^k \leq \max_{v \in S_k} |\langle f_v, q \rangle|^k \leq \binom{m+k-2}{k}^{\frac{3}{2}} \max_{g^{\otimes k} \in B(E)^{\otimes k}} \langle g, q \rangle^k.$$

We fix $\varepsilon = \frac{1}{2e}$ and define $P_k := \text{conv}\{f_v, -f_v : v \in S_k\}$. Then, we have the following inequalities:

$$\left(1 - \frac{1}{2e}\right)^{\frac{1}{k}} \max_{g \in B(E)} \langle g, q \rangle \leq \max_{g \in P_k} \langle g, q \rangle \leq \binom{m+k-2}{k}^{\frac{3}{2k}} \max_{g \in B(E)} \langle g, q \rangle.$$

The standard Stirling estimate for binomial coefficients gives

$$\binom{m+k-2}{k} \leq e^{m-2} \left(1 + \frac{k}{m-2}\right)^{m-2}.$$

If we set $k = s(m-2)$, then we have

$$\binom{m+k-2}{k}^{\frac{3}{2k}} \leq \left(e \left(1 + \frac{k}{m-2}\right)\right)^{\frac{3(m-2)}{2k}} \leq (e(1+s))^{\frac{3}{2s}}.$$

Note also that $(1 - \frac{1}{2e})^{-1} \leq 1 + \frac{1}{e}$. Since the above inequalities on $\max_{P_k} \langle g, q \rangle$ hold for arbitrary $q \in U$, we have the following inclusions:

$$\left(1 + \frac{1}{e}\right)^{-\frac{1}{s(m-2)}} B(E) \subseteq P_k \subseteq (e(1+s))^{\frac{3}{2s}} B(E).$$

We set $Q_k = -(e(1+s))^{\frac{3}{2s}} P_k^\circ$ and we assume $s \geq e$. Note that $(1 + \frac{1}{e})^{\frac{1}{s(m-2)}} \leq (\frac{s+1}{e})^{\frac{1}{s}}$. Therefore,

$$\left(1 + \frac{1}{e}\right)^{\frac{1}{s(m-2)}} (e(1+s))^{\frac{3}{2s}} \leq (s+1)^{\frac{3}{s}}.$$

Now using the relation $-B(E)^\circ = \widetilde{Pos_E} - r$, we conclude that

$$\widetilde{Pos_E} - r \subseteq Q_k \subseteq \left(1 + \frac{k}{m-2}\right)^{\frac{3(m-2)}{k}} (\widetilde{Pos_E} - r).$$

Let us summarize the result of this section in a theorem statement.

THEOREM 5.2. *Let $E \subset P_{n,2d}$ be a linear subspace, let $r = (x_1^2 + x_2^2 + \cdots + x_n^2)^d$ and assume $r \in E$. Let $m = \dim(E)$, assume $m \geq 3$, and let k be an integer with $k \geq e(m-2)$. Then, there exists a polytope Q_k with $O(k^{m-2})$ facets which satisfies*

$$\widetilde{Pos_E} - r \subseteq Q_k \subseteq \left(1 + \frac{k}{m-2}\right)^{\frac{3(m-2)}{k}} (\widetilde{Pos_E} - r).$$

Remark 5.3. Alexander Barvinok recently pointed out to me his article [6]. The arguments in [6] are well optimized for approximating arbitrary convex bodies with arbitrary polytopes. The optimization of the approximation in [6] is done by using Chebyshev nodes. In our particular case, the order of approximation given by [6] is not better than results of this article due to the large coefficient of symmetry of the cone of nonnegative polynomials [8]. The main difference in our argument is that we do not aim to approximate with arbitrary polytopes, and instead approximate with polytopes created out of pointwise evaluations in E .

6. Random approximation polytopes. The deterministic construction of the approximation polytopes in previous section includes two main steps: Computation of John's ellipsoid and computation of an approximate decomposition of identity, as stated in Theorem 5.1. The computation of John's ellipsoid is known to be hard [32]. Therefore, we will take a different approach in this section and construct approximation polytopes by random sampling.

If one is equipped with the knowledge that $\|P_E(p_v)\| \leq a\sqrt{\dim(E)}$ for some universal constant a , then one can bypass the John ellipsoid computation and directly use the theorem of Rudelson as explained in Remark 3.2. However, as we have seen in the third section one does not always have such an upper bound. Hence, different tools are needed.

We will use a tool coming from computational geometry, namely the ε -net theorem. It is hard to do any justice to the beautiful mathematics behind the ε -net theorem in a very limited space. Therefore, we just state a special case of a version of the ε -net theorem due to Komlós, Pach, and Woeginger [20], and refer the reader to [23] and references therein.

THEOREM 6.1 (the ε -net theorem; a special case for half-spaces). *Let \mathcal{F} be a family of half-spaces in \mathbb{R}^{m-1} . Let μ be a probability measure on \mathbb{R}^{m-1} . Assume that for all half-spaces $H_+ \in \mathcal{F}$, we have $\mu(H_+) \geq \varepsilon$, where $0 < \varepsilon \leq (3e)^{-2}$. Let $t = \lceil \frac{3m}{\varepsilon} \ln(\frac{1}{\varepsilon}) \rceil$, and let x_1, x_2, \dots, x_t be independent random variables distributed with respect to μ . Then, the set $X = \{x_1, x_2, \dots, x_t\}$ has a nonempty intersection with all members of \mathcal{F} with probability greater than $1 - 4(9e^2\varepsilon)^m$.*

Naszódi's recent article [24] was a source of inspiration for the ideas developed in this section. One can obtain the above theorem as a special case of Lemma 3.2 in Naszodi's article by setting $C = 3$, $\delta = 4(9e^2\varepsilon)^m$, and using the fact that a collection of half-spaces in \mathbb{R}^{m-1} has VC-dimension at most m .

To use the ε -net theorem we need to specify a probability measure on our space. We will use a modified version of the probability measure defined in the third section. Let us recall how the construction works. We have $E \subseteq P_{n,2d}$ with $\dim(E) = m$ and $r = (x_1^2 + x_2^2 + \dots + x_n^2)^d \in E$. We define a subspace U as $U := \{f \in E : \langle f, r \rangle = 0\}$. We showed that, for every $v \in S^{n-1}$, there exists a polynomial that we denoted by $\Pi_E(p_v - r) \in U$ such that for all $f \in U$, we have $f(v) = \langle f, \Pi_E(p_v - r) \rangle$. Then, we defined a map ϕ_E as follows:

$$\begin{aligned}\phi_E : S^{n-1} &\rightarrow E, \\ \phi_E(v) &= \Pi_E(p_v - r).\end{aligned}$$

Out of this map, we created the following convex body:

$$B(E) := \text{conv}\{\phi_E(v) : v \in S^{n-1}\}.$$

We showed that $B(E) \subset U$ is dual to a base of the cone of nonnegative polynomials, and hence our objective is to approximate $B(E)$ with polytopes having as few vertices

as possible. Now we define the probability measure that is going to be used in this section. Let λ be the uniform measure on Δ^m , where Δ^m is defined as follows:

$$\Delta^m := \left\{ x \in \mathbb{R}^m : x_i \geq 0 \text{ and } \sum_i x_i = 1 \right\}.$$

Now, let $T^m = S^{n-1} \times S^{n-1} \times \cdots \times S^{n-1}$ be the Cartesian product of S^{n-1} with itself m times and let $\sigma_m = \sigma \times \sigma \times \cdots \times \sigma$ be the product measure on T^m , where σ is the uniform measure on S^{n-1} . Let Ψ be a map from $\Delta^m \times T^m$ to U defined as follows:

$$\Psi : \Delta^m \times T^m \rightarrow U,$$

$$\Psi(a_1, a_2, \dots, a_m, v_1, v_2, \dots, v_m) = \sum_{i=1}^m a_i \phi_E(v_i).$$

We define μ to be the pushforward measure of $\lambda \times \sigma_m$ under Ψ ; μ is clearly a probability measure, and it is supported on $B(E)$, i.e., $\mu(B(E)) = 1$. Now we need to pick a special family of half-spaces and show that their measure is bounded from below by a certain threshold. We consider the family of half-spaces defined by supporting hyperplanes of $\alpha B(E)$ for a fixed $0 < \alpha \leq 1$:

$$\mathcal{F}_\alpha := \{H^+ : H \text{ is a supporting hyperplane of } \alpha B(E)\}.$$

Suppose we have a set $V = \{v_1, v_2, \dots, v_t\} \subseteq B(E)$ where for all half-spaces $H^+ \in \mathcal{F}_\alpha$ there exists a $v_i \in H^+$. This would imply the following inclusions:

$$\alpha B(E) \subseteq \text{conv}\{v_1, v_2, \dots, v_t\} \subseteq B(E).$$

Our goal in the rest of this section is to use this observation to construct random approximation polytopes to $B(E)$. We first need some preparatory lemmas.

LEMMA 6.2. *Let $f \in P_{n,2d}$ be a polynomial with $\int_{S^{n-1}} f(x) \sigma(x) = 0$, where σ is the uniform measure on S^{n-1} . Then, we have*

$$\sigma(\{x : f(x) > 0\}) \geq \frac{\|f\|_1^2}{4\|f\|_2^2}.$$

Proof. We define a function f_+ as follows:

$$f_+(x) = \begin{cases} f(x) & \text{if } f(x) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Now consider $f_+(x)$ as a random variable, where x is distributed according to σ . Clearly $f_+(x) \geq 0$ for all $x \in S^{n-1}$, and we also have

$$\sigma(\{x : f(x) > 0\}) = \sigma(\{x : f_+(x) > 0\}).$$

We apply the Paley–Zygmund inequality to $f_+(x)$, which gives the following:

$$\sigma(\{x : f_+(x) > 0\}) \geq \frac{(\mathbb{E}f_+(x))^2}{\mathbb{E}f_+(x)^2}.$$

Observe that $\int_{S^{n-1}} f(x) \sigma(x) = 0$ implies $2 \int_{S^{n-1}} f_+(x) \sigma(x) = \|f\|_1$. Also, we trivially have $\mathbb{E}f_+(x)^2 \leq \|f\|_2^2$. \square

As a corollary of this lemma, we have the following lower bound on the measure of half-spaces from \mathcal{F}_α .

COROLLARY 6.3. *Let $H_+ \in \mathcal{F}_\alpha$ be a half-space defined by an $f \in U$ as follows:*

$$H_+ := \left\{ g \in U : \langle f, g \rangle \geq \alpha \max_{v \in S^{n-1}} f(v) \right\}.$$

Then, we have

$$\mu(H_+) \geq \left(\frac{(1-\alpha) \|f\|_1^2}{4 \|f\|_2^2} \right)^m.$$

Proof. First, let us note that

$$\max_{g \in \alpha B(E)} \langle f, g \rangle = \alpha \max_{v \in S^{n-1}} f(v).$$

Therefore, all the supporting hyperplanes of $\alpha B(E)$ will be in the form considered in the statement of the corollary. Now let H_+^1 be the translate of H_+ supporting $B(E)$, and let $\phi_E(v_0) \in \partial B(E) \cap H_+^1$. Also, let H_+^0 be the translate of H_+ passing through the origin. Then $\alpha \phi_E(v_0) + (1-\alpha)(B(E) \cap H_+^0)$ is included inside $B(E) \cap H_+$ due to convexity.

Now let $A_f := \{v \in S^{n-1} : f(v) \geq 0\}$, and let $A_f^m = A_f \times A_f \times \cdots \times A_f$ be the Cartesian product of A_f with itself m times. Then, the image of $(1-\alpha)\Delta^m \times A_f^m$ under the map Ψ is included in $(1-\alpha)(B(E) \cap H_+^0)$. By the definition of the measure μ , we then have the following lower bound:

$$\begin{aligned} \mu(H_+) &\geq \mu(\alpha \phi_E(v_0) + (1-\alpha)(B(E) \cap H_+^0)) \geq (1-\alpha)^{m-1} \sigma(A_f)^m \\ &\geq (1-\alpha)^m \sigma(A_f)^m. \end{aligned}$$

Using Lemma 6.2 completes the proof:

$$\mu(H_+) \geq \left(\frac{(1-\alpha) \|f\|_1^2}{4 \|f\|_2^2} \right)^m. \quad \square$$

Now we are ready to state and prove the main result of this section.

THEOREM 6.4. *Let $E \subset P_{n,2d}$ be a subspace with $r \in E$ and $\dim(E) = m$. Let $B(E)$ and the measure μ be as defined above. We set $M(E) = \max_{f \in E} \|f\|_2^2 / \|f\|_1^2$. Now let*

$$t = \left\lceil 27e^2 m^2 \left(\frac{M(E)}{1-\alpha} \right)^m \ln \left(\frac{M(E)}{1-\alpha} \right) \right\rceil,$$

and let x_1, x_2, \dots, x_t be independent random vectors in E distributed according to μ . Then,

$$\alpha B(E) \subseteq \text{conv}\{x_1, x_2, \dots, x_t\} \subseteq B(E)$$

with probability at least $1 - 4\left(\frac{1-\alpha}{M(E)}\right)^{m^2}$.

Proof. We will use the ε -net theorem with the measure μ and the family of half-spaces \mathcal{F}_α . From Corollary 6.3, we have that for all $H_+ \in \mathcal{F}_\alpha$, we have $\mu(H_+) \geq \left(\frac{1-\alpha}{4M(E)}\right)^m$. We set $\varepsilon = \frac{(1-\alpha)^m}{9e^2 M(E)^m}$, for which we clearly have $9e^2 \varepsilon \leq 1$, and also $\mu(H_+) \geq \varepsilon$ for all $H_+ \in \mathcal{F}_\alpha$. Now, let $t = \lceil \frac{3m}{\varepsilon} \ln(\frac{1}{\varepsilon}) \rceil$ and let x_1, x_2, \dots, x_t be independent random vectors distributed according to μ . Then, the ε -net theorem yields that $X = \{x_1, x_2, \dots, x_t\}$ is a transversal of \mathcal{F}_α with probability at least $1 - 4(9e^2 \varepsilon)^m$. \square

Below is the restatement of the theorem with dual convex bodies.

COROLLARY 6.5. *Let $E \subset P_{n,2d}$ be a subspace with $r \in E$ and $\dim(E) = m$. Let μ be the measure as defined above. We set $M(E) = \max_{f \in E} \|f\|_2^2 / \|f\|_1^2$. For a given α with $0 < \alpha \leq 1$, we set*

$$t = O\left(m^2 \left(\frac{M(E)}{1-\alpha}\right)^m \ln\left(\frac{M(E)}{1-\alpha}\right)\right).$$

Let x_1, x_2, \dots, x_t be independent random vectors distributed with μ . We define the polytope $K_\alpha = \{y \in U : \langle y, x_i \rangle \geq -1 \text{ for all } 1 \leq i \leq t\}$. Then K_α satisfies

$$\widetilde{\text{Pos}}_E - r \subseteq K_\alpha \subseteq \frac{1}{\alpha} \left(\widetilde{\text{Pos}}_E - r \right)$$

with probability at least $1 - 4\left(\frac{1-\alpha}{M(E)}\right)^{m^2}$.

We do not know any way to bound $M(E)$ from above using the low-dimensionality of E . What we can do is to bound $M(P_{n,2d})$ from above in terms of d only.

LEMMA 6.6. *The following holds for all $f \in P_{n,2d}$:*

$$\frac{\|f\|_2^2}{\|f\|_1^2} \leq 2^{2d}.$$

Proof. We have

$$\begin{aligned} \int_{\mathbb{R}^n} |f(x)| \gamma_n(x) &= \frac{|S^{n-1}|}{(2\pi)^{\frac{n}{2}}} \int_0^\infty r^{n+2d-1} e^{-\frac{r^2}{2}} dr \int_{S^{n-1}} |f(x)| \sigma(x), \\ \int_{\mathbb{R}^n} |f(x)| \gamma_n(x) &= \frac{|S^{n-1}|}{(2\pi)^{\frac{n}{2}}} 2^{\frac{n}{2}+d-1} \Gamma\left(\frac{n}{2} + d\right) \int_{S^{n-1}} |f(x)| \sigma(x). \end{aligned}$$

Following the same pattern we also derive the following:

$$\int_{\mathbb{R}^n} f(x)^2 \gamma_n(x) = \frac{|S^{n-1}|}{(2\pi)^{\frac{n}{2}}} 2^{\frac{n}{2}+2d-1} \Gamma\left(\frac{n}{2} + 2d\right) \int_{S^{n-1}} f(x)^2 \sigma(x).$$

Also note that $\frac{|S^{n-1}|}{(2\pi)^{\frac{n}{2}}} 2^{\frac{n}{2}} = \frac{2}{\Gamma(\frac{n}{2})}$. This gives us the following ratio:

$$\frac{\int_{\mathbb{R}^n} f(x)^2 \gamma_n(x)}{(\int_{\mathbb{R}^n} |f(x)| \gamma_n(x))^2} = \frac{\|f\|_2^2}{\|f\|_1^2} \frac{\Gamma(\frac{n}{2})\Gamma(\frac{n}{2} + 2d)}{\Gamma(\frac{n}{2} + d)^2}.$$

The reverse Hölder inequalities for polynomials with normal random variables [11, 26] tell us that the left-hand side of the equality is less than 2^{2d} . Therefore, we have

$$\frac{\|f\|_2^2}{\|f\|_1^2} \leq 2^{2d} \frac{\Gamma(\frac{n}{2} + d)^2}{\Gamma(\frac{n}{2})\Gamma(\frac{n}{2} + 2d)} \leq 2^{2d}. \quad \square$$

7. Acknowledgments. I had the idea of Theorem 4.3 after having a short discussion with Alexander Barvinok in December 2016 at ICERM, Brown University. I also had a chance to meet with him at Physikzentrum Bad Honnef on late November 2017. I would like to thank Alexander Barvinok for enjoyable discussions, and to ICERM and Physikzentrum Bad Honnef for their hospitality. I also would like to present my thanks to Cynthia Vinzant and Seth Sullivant for discussions at Raleigh during fall 2016, and to anonymous referees whose remarks helped to clarify the presentation.

REFERENCES

- [1] A. A. AHMADI, *Lecture notes for Princeton ORFE course*, Lecture 15; available at <https://www.princeton.edu/~amirali/Public/Teaching/ORF523>.
- [2] A. BARVINOK, *A bound for the number of vertices of a polytope with applications*, *Combinatorica*, 33 (2013), pp. 1–10.
- [3] A. BARVINOK, *A Course in Convexity*, Grad. Stud. Math. 54, American Mathematical Society, Providence, RI, 2002.
- [4] A. BARVINOK AND G. BLEKHERMAN, *Convex geometry of orbits*, in *Combinatorial and Computational Geometry*, Math. Sci. Res. Inst. Publ. 52, Cambridge University Press, Cambridge, 2005, pp. 51–77.
- [5] A. BARVINOK, *Approximating L^∞ norms by L^{2k} norms for functions on orbits*, *Found. Comput. Math.*, 2 (2002), pp. 393–412.
- [6] A. BARVINOK, *Thrifty approximations of convex bodies by polytopes*, *Int. Math. Res. Not.*, 2014 (2014), pp. 4341–4356.
- [7] A. BHARDWAJ, P. ROSTALSKI, AND R. SANYAL, *Deciding polyhedrality of spectrahedra*, *SIAM J. Optim.*, 25 (2015), pp. 1873–1884.
- [8] G. BLEKHERMAN, *Convexity properties of the cone of nonnegative polynomials*, *Discrete Comput. Geom.*, 32 (2004), pp. 345–371.
- [9] G. BLEKHERMAN, P. A. PARRILO, AND R. THOMAS, EDS., *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Ser. Optim. 13, SIAM, Philadelphia, PA, 2013.
- [10] V. I. BOGACHEV, *Gaussian Measures*, Math. Surveys Monogr. 62, American Mathematical Society, Providence, RI, 1998.
- [11] S. G. BOBKOV, *Remarks on the growth of L_p -norms of polynomials*, *Geometric Aspects of Functional Analysis*, Lecture Notes in Math. 1745, Springer, Berlin, 2000.
- [12] G. BRAUN AND S. POKUTTA, *The matching polytope does not admit fully-polynomial size relaxation schemes*, in *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM, Philadelphia, PA, 2015, pp. 837–846.
- [13] G. BRAUN, S. FIORINI, S. POKUTTA, AND D. STEURER, *Approximation limits of linear programs (beyond hierarchies)*, *Math. Oper. Res.*, 40 (2015), pp. 756–772.
- [14] S. BRAZITIKOS, A. GIANNOPOULOS, P. VALETTAS, AND B. VRITSIOU, *Geometry of Isotropic Convex Bodies*, Math. Surveys Monogr. 196, American Mathematical Society, Providence, RI, 2014.
- [15] O. FRIEDLAND AND P. YOUSEF, *Approximating convex bodies and matrices through Kadison–Singer*, *Int. Math. Res. Not.* (2017), <https://doi.org/10.1093/imrn/rnx206>.
- [16] P. M. GRUBER, *Minimal ellipsoids and their duals*, *Rend. Circ. Mat. Palermo* (2), 37 (1988), pp. 35–64.
- [17] M. D. CHOI, T. Y. LAM, AND B. REZNICK, *Even symmetric sextics*, *Math. Z.*, 195 (1987), pp. 559–580.
- [18] S. ILIMAN AND T. DE WOLFF, *Lower bounds for polynomials with simplex Newton polytopes based on geometric programming*, *SIAM J. Optim.*, 26 (2016), pp. 1128–1146.
- [19] F. JOHN, *Extremum problems with inequalities as subsidiary conditions*, in *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, Interscience Publishers, New York, NY, 1948, pp. 187–204.
- [20] J. KOMLÓS, J. PACH, AND G. WOEGINGER, *Almost tight bounds for ε -nets*, *Discrete Comput. Geom.*, 7 (1992), pp. 163–173.
- [21] A. MARCUS, D. SPIELMAN, AND N. SRIVASTAVA, *Interlacing families I: Bipartite Ramanujan graphs of all degrees*, *Ann. Math.*, 182 (2015), pp. 307–325.
- [22] A. MARCUS, D. SPIELMAN, AND N. SRIVASTAVA, *Interlacing families II: Mixed characteristic polynomials and the Kadison–Singer problem*, *Ann. Math.*, 182 (2015), pp. 327–350.
- [23] J. MATOUSEK, *Lectures on Discrete Geometry*, Grad. Texts in Math. 212, Springer, New York, 2002.

- [24] M. NASZÓDI, *Approximating a Convex Body with a Polytope Using the Epsilon-Net Theorem*, preprint, <https://arxiv.org/pdf/1705.07754.pdf>, 2017.
- [25] P. A. PARRILO, *Exploiting Algebraic Structure in Sum of Squares Programs*, in Positive Polynomials in Control, D. Henrion and A. Garulli, eds., Lect. Notes Control Inf. Sci. 312, Springer, Berlin, 2005, pp. 181–194.
- [26] Y. V. PROKHOROV, *On polynomials in normally distributed random variables*, Probab. Theory Appl., 37 (1992), pp. 692–694.
- [27] P. PEBAY, J. M. ROJAS, AND D. THOMPSON, *Optimizing n -variate $(n + k)$ -nomials for small k* , Theoret. Comput. Sci., 412 (2011), pp. 1457–1469.
- [28] M. V. RAMANA, *Polyhedra, spectrahedra, and semidefinite programming*, in Topics in Semidefinite and Interior-Point Methods, Fields Inst. Commun. 18, American Mathematical Society, Providence, RI, 1998, pp. 27–38.
- [29] M. RUDELSON, *Random vectors in the isotropic position*, J. Funct. Anal., 164 (1999), pp. 60–72.
- [30] M. RUDELSON, *Contact points of convex bodies*, Israel J. Math., 101 (1997), pp. 93–124.
- [31] N. SRIVASTAVA, *Spectral Sparsification and Restricted Invertibility*, Ph.D. Dissertation, Yale University, 2010; available at <https://math.berkeley.edu/~nikhil/dissertation.pdf>.
- [32] M. J. TODD, *Minimum-Volume Ellipsoids: Theory and Algorithms*, MOS-SIAM Ser. Optim. 23, SIAM, Philadelphia, PA, 2016.
- [33] R. VERSHYNIN, *High-Dimensional Probability: An Introduction with Applications in Data Sciences*, Camb. Ser. Stat. Probab. Math. 47, Cambridge University Press, Cambridge, 2018.
- [34] R. VERSHYNIN, *Introduction to the non-asymptotic analysis of random matrices*, in Compressed Sensing: Theory and Applications, Cambridge University Press, Cambridge, 2012, pp. 210–268.