

# Low-Degree Polynomial Approximation of the Cross-Polytope

Xiaoyu Li

University of New South Wales  
xiaoyu.li2@unsw.edu.au

## Abstract

How much approximation power does polynomial degree buy for a symmetric convex body? Barvinok’s lifted ellipsoid construction shows that degree  $2t$  always suffices for distortion  $O(\sqrt{d/t})$  when  $t \leq d$ , but the construction itself does not identify a body that forces this curve throughout the degree range. We show that the cross-polytope does. If  $\kappa_t$  denotes the optimum over positive sum-of-squares forms and  $\kappa_t^{\text{nn}}$  the optimum over all globally nonnegative forms, then

$$\frac{1}{\sqrt{2e}} \sqrt{\frac{d}{t}} \leq \kappa_t^{\text{nn}}(B_1^d) \leq \kappa_t(B_1^d) \leq \sqrt{2e} \sqrt{\frac{d}{t}}, \quad 1 \leq t \leq d.$$

Thus degree  $\Theta(d)$  is necessary and sufficient for constant distortion, even when the lower-bound adversary may use any nonnegative form rather than an SoS certificate.

The lower proof is representation-free. Signed-permutation averaging removes orientation but still leaves a large invariant algebra. Evaluating instead on the flat points of the  $\ell_1$ -sphere makes every invariant monomial forget its partition structure and remember only the support size  $k$ . Every candidate then has the profile  $V(k) = k^{-2t}Q(k)$ , where  $\deg Q \leq t$  and  $Q(0) = 0$ ; after the substitution  $u = 1/k$ , Lagrange interpolation shows that this degree budget cannot keep the profile nearly level across  $d$  support scales. The matching SoS construction is dictated by the same symmetry: averaging the powers of the sign-vector facet normals gives the Rademacher form  $\mathbb{E}_\varepsilon(\varepsilon^\top x)^{2t}$ .

The mechanism also diagnoses what does and does not generalize. The polar cube has a constant certificate at logarithmic degree, and a two-dimensional quartic shows that its familiar coordinate certificate need not be optimal. For a general symmetric polytope, weighted facet powers yield a concave boundary-floor program whose worst-direction oracle reduces to convex dual-norm problems; at degree two it recovers classical optimal design and the John ellipsoid. This is an oracle-model certificate optimization, not an end-to-end complexity result. Finally, a weighted reciprocal-grid lemma for  $k^{-\alpha}Q(k)$  yields the  $\ell_p$  consequences and identifies  $p = 2$  as the precise point where the interpolation mechanism stops.

## 1 Introduction

John’s theorem says that a quadratic form can approximate every origin-symmetric convex body in  $\mathbb{R}^d$  with loss at most  $\sqrt{d}$ . Allowing a homogeneous polynomial of degree  $2t$  gives a sequence of increasingly rich exact-degree algebraic models. The basic quantitative question is then: *what does the additional degree buy?* If degree is the available approximation budget, how quickly can the best worst-case distortion available through degree  $2t$  fall as  $t$  grows?

There is one point of interpretation. The  $2t$ -th root of a nonnegative form is homogeneous, but it need not be convex and hence need not be a norm. The quantity relevant here is only its multiplicative sandwich on the

boundary of the body. For a form  $p$  that is positive away from the origin, write

$$\rho_K(p^{1/(2t)}) = \frac{\max_{\|x\|_K=1} p(x)^{1/(2t)}}{\min_{\|x\|_K=1} p(x)^{1/(2t)}}.$$

We keep two polynomial cones separate:

$$\kappa_t(K) = \inf_{\substack{p \in \Sigma_{d,2t} \\ p(x) > 0 \text{ for } x \neq 0}} \rho_K(p^{1/(2t)}), \quad \kappa_t^{\text{nn}}(K) = \inf_{\substack{p \in \mathcal{N}_{d,2t} \\ p(x) > 0 \text{ for } x \neq 0}} \rho_K(p^{1/(2t)}), \quad (1.1)$$

where  $\Sigma_{d,2t}$  is the sum-of-squares cone and  $\mathcal{N}_{d,2t}$  is the cone of globally nonnegative forms. Thus  $\kappa_t^{\text{nn}}(K) \leq \kappa_t(K)$ . This distinction will matter in opposite directions: our upper constructions are SoS, while the principal lower bound holds against the larger nonnegative cone. Here  $t$  specifies an exact homogeneous degree; the feasible classes at successive values of  $t$  are not literally nested, and we do not assume that  $\kappa_t(K)$  is monotone. When degree is meant as a cumulative budget, one may instead take the best value over all  $s \leq t$ ; the preliminaries make this monotone envelope explicit. The cross-polytope law below has the same order under either convention.

**The universal curve and the missing worst case.** Barvinok's polynomial approximation theorem, in its all-degree SoS form, gives

$$\kappa_t(K) \leq \binom{d+t-1}{t}^{1/(2t)} = O(\sqrt{d/t}) \quad (1 \leq t \leq d) \quad (1.2)$$

for every symmetric body (Barvinok, 2003; Barvinok and Veomett, 2008). This tells us what every body can achieve, but not whether any body forces us to pay that much degree.

The cube and the cross-polytope make the gap in our understanding concrete. They are polar, share the same signed-permutation symmetry, and already behave very differently under the most immediate constructions. For the cube,

$$\sum_{i=1}^d x_i^{2t} \quad \text{gives} \quad \kappa_t(B_\infty^d) \leq d^{1/(2t)},$$

so logarithmic degree suffices for a constant certificate. For the cross-polytope, by contrast, Barvinok proved an order- $\sqrt{d}$  sextic lower bound in the published version of his paper, and Barvinok and Veomett recorded the quartic and sextic cases (Barvinok, 2003)(Barvinok and Veomett, 2008, Sec. 2.3). These fixed-degree obstructions suggest that the universal curve reflects a genuine worst case, rather than an artifact of the lifted construction. The question is whether the obstruction persists for every degree, and whether it depends on the SoS representation.

**The degree law.** Our main result answers both questions.

**Theorem 1.1** (Cross-polytope degree law; informal version of Theorem 3.1). *For every  $d \geq 1$  and  $1 \leq t \leq d$ ,*

$$\frac{1}{\sqrt{2e}} \sqrt{\frac{d}{t}} \leq \kappa_t^{\text{nn}}(B_1^d) \leq \kappa_t(B_1^d) \leq \sqrt{2e} \sqrt{\frac{d}{t}}.$$

Consequently, the cross-polytope matches Barvinok's universal SoS upper bound up to an absolute constant, and constant distortion requires and is achieved by degree of order  $d$ . There are two nontrivial

uniformities in the statement. The estimate holds throughout  $1 \leq t \leq d$ , rather than at a fixed degree, and its lower half does not inspect an SoS representation at all. It extends the previously recorded fixed-degree cases to the full degree range while retaining a representation-free lower bound over all globally nonnegative forms, and hence over the full SoS cone. To the best of our knowledge, this is the first degree-uniform cross-polytope lower bound over the full cone of globally nonnegative forms; the narrower sums-of-powers precedent and the classical flat-ray literature are separated explicitly in the related-work discussion below.

The theorem asks for two logically different arguments. The lower bound must simplify *every* nonnegative form without making it harder to approximate the body. The upper bound needs only one form, but it must explain the same  $\sqrt{d/t}$  scale. Both arguments come from the signed-permutation symmetry, used in opposite directions.

**How the obstruction is isolated.** A direct attack on the coefficients of an arbitrary nonnegative form retains far more information than the sandwich ratio can detect. Signed-permutation averaging is the first compression: it erases orientation and cannot increase the ratio. It does not, however, make the problem one-dimensional. An invariant degree- $2t$  form can still involve many products of even power sums.

The second compression is the decisive one. A lower bound does not require locating the true boundary extrema; it is enough to find a small set of directions on which every candidate already varies. We choose the flat points

$$x^{(k)} = \frac{1}{k}(\underbrace{1, \dots, 1}_k, 0, \dots, 0), \quad k = 1, \dots, d,$$

because on this family every even power sum depends only on the support size  $k$ . More importantly, a product of power sums forgets the underlying partition and remembers only its number of factors. The restriction of every invariant form therefore has the forced profile

$$V(k) = p(x^{(k)}) = k^{-2t}Q(k), \quad \deg Q \leq t, \quad Q(0) = 0.$$

At this point the high-dimensional geometry has disappeared. Dimension supplies  $d$  support scales at which the profile must remain positive, while degree gives the numerator only  $t$  degrees of freedom with which to compensate for the factor  $k^{-2t}$ .

The mechanism is already visible when  $t = 1$ : then  $V(k) = a/k$ , so the vertex and diagonal differ by a factor  $d$ , yielding the classical  $\sqrt{d}$  distortion. Higher degree can partially flatten this profile, but it cannot tune the first  $t$  support sizes and the distant scale  $k = d$  independently. The change of variables  $u = 1/k$  makes this precise:  $V(1/u) = u^t R(u)$  with  $\deg R \leq t - 1$ , so the values at  $1, 1/2, \dots, 1/t$  determine  $R(1/d)$ . Bounding that extrapolation functional by Lagrange interpolation gives the lower bound. The candidate's strict positivity away from the origin keeps the sampled profile positive; the argument never inspects how its global nonnegativity is certified.

**How the matching form emerges.** The upper construction is dictated by the same geometry rather than guessed as an unrelated probabilistic device. The oriented facet normals of the cross-polytope are the sign vectors  $\varepsilon \in \{\pm 1\}^d$ ; equivalently, opposite pairs define the slabs  $|\varepsilon^\top x| \leq 1$ . Averaging the  $2t$ -th powers of these functionals gives the canonical symmetric form

$$f(x) = \mathbb{E}_\varepsilon[(\varepsilon^\top x)^{2t}].$$

Schur convexity places its maximum on the  $\ell_1$ -sphere at a vertex and its minimum at the diagonal. Hence

$$\rho(f^{1/(2t)}) = \frac{d}{(\mathbb{E}[S_d^{2t}]^{1/(2t)}), \quad S_d = \varepsilon_1 + \dots + \varepsilon_d.$$

The collision-free pairings in this moment already supply the required scale, while Gaussian comparison gives the opposite moment estimate. In the regime  $t \rightarrow \infty$  with  $t = o(\sqrt{d})$ , the displayed ratio is  $\sqrt{e/2}\sqrt{d/t}(1 + o(1))$ . We also retain Barvinok’s upper bound because it is uniform beyond this moment-asymptotic regime.

This gives the central proof map:

$$\begin{array}{ccccc} \text{arbitrary form} & \xrightarrow{\text{average}} & \text{support-size profile} & \xrightarrow{\text{interpolate}} & \text{lower bound,} \\ \text{one facet orbit} & \xrightarrow{\text{average}} & \text{Rademacher form} & \xrightarrow{\text{moments}} & \text{upper bound.} \end{array}$$

The two rows are deliberately asymmetric: the first must compress every adversary without increasing its ratio, whereas the second needs to manufacture only one candidate with the matching order.

**Polarity as a stress test.** Is the lower-bound mechanism a generic consequence of polarity, or a feature of the  $\ell_1$  geometry? The cube certificate above and our cross-polytope lower bound settle the same-degree comparison without requiring the cube’s exact optimum. At  $t = \lceil \log_2 d \rceil$ , the cube has an SoS certificate of distortion at most  $\sqrt{2}$ , while every globally nonnegative certificate for the cross-polytope has distortion

$$\Omega\left(\sqrt{\frac{d}{\log d}}\right).$$

Thus polarity does not preserve the degree scale at which a prescribed distortion is attained.

This comparison also provides an early scope check. The coordinate form for the cube is sufficient for the separation, but sufficiency is not optimality. Already for  $B_\infty^2$  at  $t = 2$ , the SoS form

$$x^4 + y^4 - x^2y^2 = (x^2 - y^2)^2 + (xy)^2$$

has distortion  $(4/3)^{1/4} < 2^{1/4}$ . The negative mixed coefficient is unavailable to a nonnegative mixture of coordinate fourth powers, but it remains compatible with an SoS representation.

**What the two proof mechanisms test next.** The sign design suggests weighting powers of facet normals for a general symmetric polytope. Maximizing the boundary floor gives a concave, one-sided certificate. In the oracle model, entropic mirror ascent gives an additive outer-iteration bound, while homogeneity represents each worst-direction oracle by finitely many convex dual-norm problems. At  $t = 1$  the program reduces to  $G$ -optimal design and the John ellipsoid (Kiefer and Wolfowitz, 1960). The quartic example then supplies the necessary scope check: even a fully symmetric optimized facet design need not be optimal over the full SoS cone.

The  $\ell_p$  family instead tests the weighted flat-profile obstruction. Its exponent is  $\alpha = 2t/p$ , so  $p = 2$  is exactly the point where  $\alpha$  meets the numerator degree. The resulting theorem gives the exact quadratic law, exact algebraic resonance points, and matching dimension exponents up to  $d^{o(1)}$  for fixed  $p < 2$  with  $\log t = o(\log d)$ . Beyond the threshold, the argument stops without predicting the answer. The formal statements of both extensions appear in Section 3.

**Related work and scope.** The polynomial-approximation framework originates in Barvinok (2003) and the survey of Barvinok and Veomett (2008). Polynomial norms and the geometry separating nonnegative forms, sums of squares, and sums of powers are studied in Ahmadi et al. (2019); Blekherman (2004). For the much narrower family of sums of  $2t$ -th powers of linear forms, the polynomial root is an  $L_{2t}$  subspace norm; the classical estimate  $T_2(L_{2t}) = \Theta(\sqrt{t})$  already gives an  $\Omega(\sqrt{d/t})$  obstruction by testing coordinate vectors

and their Rademacher sums (Dümbgen et al., 2010, Lem. 3.2). The scope gain here is that the lower proof is representation-free and applies to every globally nonnegative form.

The support-size rays themselves are classical in symmetric nonnegativity. For  $d \geq 3$ , Choi–Lam–Reznick proved that an even symmetric sextic is nonnegative if and only if it is nonnegative on the rays  $(\mathbf{1}_k, 0^{d-k})$ ,  $k = 1, \dots, d$  (Choi et al., 1987, Thm. 3.7). Timofte placed this criterion in a broader theory of test sets with few distinct coordinate levels, followed by half-degree and reflection-group developments (Acevedo and Velasco, 2016; Riener, 2012; Timofte, 2003). Recent work extends the flat-ray criterion to a hook-shaped subspace expressed in elementary symmetric polynomials, while showing that the corresponding finite criterion for the power-sum hook-shaped class fails in every degree  $r \geq 4$ —equivalently, from even degree  $2r \geq 8$  after the square substitution (Acevedo et al., 2025, Prop. 5.5, Rem. 5.7, and Thms. 5.6, 5.8). Our use is quantitative rather than a new test-set claim: we do not assert that these rays contain all extrema or test nonnegativity in every degree. The normalization also matters. The customary simplex point is  $y = k^{-1}\mathbf{1}_k$ , whose pullback under  $y_i = x_i^2$  is  $x = k^{-1/2}\mathbf{1}_k$ ; our  $\ell_1$ -boundary point is instead  $k^{-1}\mathbf{1}_k$ . Homogeneity gives  $p(k^{-1}\mathbf{1}_k) = k^{-t}p(k^{-1/2}\mathbf{1}_k)$ . This radial weight is part of the profile whose all-degree oscillation yields the distortion bound.

The reciprocal-grid step is adjacent to discrete norming and Remez-type estimates for polynomials bounded on finite sets (Coppersmith and Rivlin, 1992; Yomdin, 2011). Its nodes, prescribed weight, and objective are different: we control the oscillation of positive values of  $k^{-\alpha}Q(k)$  on a reciprocal grid and assume no positivity between grid points. We therefore use the estimate as a portable proof lemma, not as a general replacement for discrete Remez theory.

The quadratic endpoint of the facet program is classical optimal design (Kiefer and Wolfowitz, 1960); algorithms for the associated John ellipsoid include Cohen et al. (2019) and subsequent refinements (Cao et al., 2025; Li et al., 2026; Woodruff and Yasuda, 2024). We use entropic mirror ascent in its standard nonsmooth form (Bubeck, 2015).

**Organization.** Section 2 fixes the sandwich framework. Section 3 states the formal results in the order of the questions above. Section 4 isolates the support-size obstruction, and Section 5 develops the matching sign construction and then the general facet certificate, including its strict limitation. Section 6 digests what the argument proves, what it rules out, and which component must change in the remaining problems. Deferred assemblies and the  $\ell_p$  proofs appear in Section A.

## 2 Preliminaries

**Basic notation.** We write  $[m] = \{1, \dots, m\}$  and  $\Delta_n = \{w \in \mathbb{R}^n : w_i \geq 0, \sum_i w_i = 1\}$ . When the ambient body is fixed and clear, we suppress it from the subscript of the sandwich ratio  $\rho_K$ .

**Bodies and gauges.** A body  $K \subset \mathbb{R}^d$  is *symmetric* if it is convex, compact, full-dimensional, and satisfies  $K = -K$ . Its gauge is

$$\|x\|_K = \inf\{\lambda > 0 : x \in \lambda K\}.$$

For a symmetric polytope  $P = \{x \in \mathbb{R}^d : |a_i^\top x| \leq 1, i \in [n]\}$ , where each  $a_i$  represents one opposite pair of facet normals and the rows span  $\mathbb{R}^d$ , this becomes  $\|x\|_P = \max_i |a_i^\top x|$ .

**Two polynomial cones.** Let  $\mathcal{N}_{d,2t}$  be the cone of globally nonnegative forms of degree  $2t$ , and let

$$\Sigma_{d,2t} = \{p : p = \sum_r q_r^2, \text{ } q_r \text{ a form of degree } t\}$$

be the sum-of-squares (SoS) cone. In general  $\Sigma_{d,2t} \subsetneq \mathcal{N}_{d,2t}$ . This distinction matters: Barvinok's construction lies in the SoS cone, whereas our lower bound for the cross-polytope holds against the larger nonnegative cone.

For  $p \in \mathcal{N}_{d,2t}$  that is positive away from the origin, put  $g = p^{1/(2t)}$ . The function  $g$  is positively homogeneous, but it need not be convex and need not be a norm. Its sandwich ratio against  $K$  is

$$\rho_K(g) = \frac{\max_{\|x\|_K=1} g(x)}{\min_{\|x\|_K=1} g(x)}.$$

For a nonnegative form that vanishes at a nonzero point, we use the extended-value convention  $\rho_K(p^{1/(2t)}) = +\infty$ . Thus degenerate forms can be included harmlessly in later infima over closed weight simplices. Writing  $m = \min_{\partial K} g$  and  $M = \max_{\partial K} g$ , homogeneity gives  $m\{g \leq 1\} \subseteq K \subseteq M\{g \leq 1\}$ , so the ratio  $M/m$  is precisely the multiplicative loss. We therefore define

$$\kappa_t(K) = \inf_{\substack{p \in \Sigma_{d,2t} \\ p(x) > 0 \text{ for } x \neq 0}} \rho_K(p^{1/(2t)}), \quad \kappa_t^{\text{nn}}(K) = \inf_{\substack{p \in \mathcal{N}_{d,2t} \\ p(x) > 0 \text{ for } x \neq 0}} \rho_K(p^{1/(2t)}). \quad (2.1)$$

Thus  $\kappa_t^{\text{nn}}(K) \leq \kappa_t(K)$ . Unless a superscript is shown,  $\kappa_t$  denotes the SoS distortion. At  $t = 1$  the two cones coincide and  $\kappa_1(K)$  is the best ellipsoidal sandwich factor; John's theorem gives  $\kappa_1(K) \leq \sqrt{d}$ .

**Exact degree and cumulative degree budgets.** The index  $t$  in (2.1) specifies the exact homogeneous degree  $2t$ . Successive feasible classes are not literally nested, so no monotonicity in  $t$  is assumed. If a monotone degree budget is desired, define

$$\hat{\kappa}_t(K) = \min_{1 \leq s \leq t} \kappa_s(K), \quad \hat{\kappa}_t^{\text{nn}}(K) = \min_{1 \leq s \leq t} \kappa_s^{\text{nn}}(K).$$

For the cross-polytope, Theorem 3.1 gives the same  $\Theta(\sqrt{d/t})$  order for these envelopes when  $1 \leq t \leq d$ : use the theorem's lower bound at every  $s \leq t$  and its upper bound at  $s = t$ .

**Facet-power certificates.** If  $P$  is a symmetric polytope and  $w \in \Delta_n$ , then

$$f_{t,w}(x) = \sum_i w_i (a_i^\top x)^{2t} = \sum_i w_i ((a_i^\top x)^t)^2$$

is an SoS form. On  $\partial P$  it satisfies  $f_{t,w}(x) \leq \max_i (a_i^\top x)^{2t} = 1$ . We use these forms for upper bounds and for the oracle design problem of Section 5; they are a strict subclass of all SoS forms. The three levels used in the paper are therefore

$$\mathcal{N}_{d,2t} \supseteq \Sigma_{d,2t} \supseteq \{f_{t,w} : w \in \Delta_n\},$$

with the inclusion directions reversed for minimization: enlarging the candidate class can only lower the best distortion. The main lower bound controls the first and largest class; the facet program optimizes a certificate inside the last and smallest one.

**The Veronese lift.** Let  $v_t : \mathbb{R}^d \rightarrow \mathbb{R}^{D_t}$ ,  $D_t = \binom{d+t-1}{t}$ , be the symmetric degree- $t$  Veronese embedding, normalized so that  $\langle v_t(a), v_t(x) \rangle = (a^\top x)^t$  (the polynomial-kernel feature map). Then  $(a_i^\top x)^{2t} = \langle v_t(a_i), v_t(x) \rangle^2$  and

$$f_{t,w}(x) = v_t(x)^\top M_t(w) v_t(x), \quad M_t(w) = \sum_i w_i v_t(a_i) v_t(a_i)^\top,$$

so the sublevel set of a facet-power form is the inverse image, under  $v_t$ , of a possibly degenerate ellipsoid in the lifted space  $\mathbb{R}^{D_t}$ . For positive weights—or, more generally, spanning support—the form is positive away from the origin. This is the sense in which each degree level lifts the John ellipsoid: at  $t = 1$ ,  $v_1$  is the identity and  $M_t(w) = \sum_i w_i a_i a_i^\top$  is the usual information matrix. We use the lift for the algorithm of [Section 5](#); the lower bound of [Section 4](#) works directly in  $\mathbb{R}^d$ .

**Barvinok’s baseline.** Barvinok’s construction, in the all-degree formulation recorded by Barvinok and Veomett, gives for every symmetric body and every  $t \geq 1$  an SoS form satisfying ([Barvinok, 2003](#); [Barvinok and Veomett, 2008](#))

$$\kappa_t(K) \leq \binom{d+t-1}{t}^{1/(2t)}. \quad (2.2)$$

For  $1 \leq t \leq d$ , the right-hand side is at most  $\sqrt{2ed/t}$ . Barvinok and Veomett also recorded the following explicit construction for the cube ([Barvinok and Veomett, 2008](#), Sec. 2.3):

**Fact 2.1** (A logarithmic-degree cube certificate). *The SoS form  $p(x) = \sum_{i=1}^d x_i^{2t}$  satisfies*

$$\kappa_t(B_\infty^d) \leq d^{1/(2t)}.$$

*In particular,  $t = \lceil \log_2 d \rceil$  gives  $\kappa_t(B_\infty^d) \leq \sqrt{2}$ .*

The displayed quantity is an upper bound, not the exact full-cone distortion; [Proposition 3.9](#) gives a strict SoS improvement already for  $d = t = 2$ . By contrast, Barvinok and Veomett proved that the cross-polytope still has order- $\sqrt{d}$  distortion at degrees 4 and 6. Our main theorem extends that lower bound to every  $1 \leq t \leq d$ .

**Rademacher moments.** Let  $S_d = \sum_{j=1}^d \varepsilon_j$  be a sum of independent Rademacher signs  $\varepsilon_j \in \{\pm 1\}$ , and  $(2t-1)!! = \frac{(2t)!}{2^t t!} = \prod_{i=1}^t (2i-1)$  the double factorial (the  $2t$ -th Gaussian moment). We write  $B_d$  for the hyperoctahedral group of signed permutations of  $\mathbb{R}^d$ , the symmetry group of both  $B_\infty^d$  and  $B_1^d$ . The  $k$ -flat points are  $x^{(k)} = \frac{1}{k}(\underbrace{1, \dots, 1}_k, 0, \dots, 0)$ ; each lies on the  $\ell_1$ -sphere,  $\|x^{(k)}\|_1 = 1$ .

### 3 Main results

The formal results follow the same diagnostic chain as the introduction. What forces the universal rate? Which explicit form matches it? Does the obstruction survive polarity? How much of the upper construction extends to general facet systems, and where does that restricted model fail? Finally, which part of the mechanism remains visible for  $\ell_p$  balls? Proofs appear in [Sections A, 4](#) and [5](#).

#### 3.1 The extremal cross-polytope

**Theorem 3.1** (Cross-polytope degree law). *For all  $d \geq 1$  and  $1 \leq t \leq d$ ,*

$$\frac{1}{\sqrt{2e}} \sqrt{\frac{d}{t}} \leq \kappa_t^{\text{nn}}(B_1^d) \leq \kappa_t(B_1^d) \leq \min \left\{ \binom{d+t-1}{t}^{1/(2t)}, \frac{d}{(\mathbb{E}[S_d^{2t}])^{1/(2t)}} \right\} \leq \sqrt{2e} \sqrt{\frac{d}{t}}. \quad (3.1)$$

In particular, both the SoS and the nonnegative-form distortions are  $\Theta(\sqrt{d/t})$ . Moreover, the sign design has the sharper asymptotic value

$$\frac{d}{(\mathbb{E}[S_d^{2t}])^{1/(2t)}} = \sqrt{\frac{e}{2}} \sqrt{\frac{d}{t}} (1 + o(1)) \quad \text{when } t \rightarrow \infty \text{ and } t = o(\sqrt{d}).$$

Consequently:

1. the cross-polytope matches Barvinok's universal upper bound up to an absolute constant, and is therefore extremal among symmetric bodies up to constants;
2. for every fixed  $C > 1$ , degree  $2t = \Theta_C(d)$  is necessary and sufficient for either distortion to be at most  $C$ .

*Proof idea.* Average an arbitrary nonnegative form over signed permutations and restrict it to the flat points; invariant theory forces the support-size profile  $V(k) = k^{-2t}Q(k)$  with  $\deg Q \leq t$  and  $Q(0) = 0$ , and interpolation forces its variation. For the other direction, uniformly averaging the sign-vector facets gives the Rademacher design of [Lemma 3.3](#), while Barvinok's theorem supplies the remaining uniform upper bound.  $\square$

**Why the rate has this form.** At  $t = 1$ , an invariant quadratic is a multiple of  $\sum_i x_i^2$ , whose value at  $x^{(k)}$  is proportional to  $1/k$ ; the vertex and diagonal already differ by a factor  $d$ , producing the classical  $\sqrt{d}$  distortion. Higher degree does not remove this tension. It gives the numerator degree  $t$  with which to flatten a profile whose geometric normalization is  $k^{2t}$ . Interpolation proves that this extra budget buys the factor  $\sqrt{t}$ , but cannot buy more.

Barvinok proved the cross-polytope lower bound for  $t = 3$ , and Barvinok and Veomett recorded the cases  $t = 2, 3$  ([Barvinok, 2003](#)) ([Barvinok and Veomett, 2008](#), Sec. 2.3). The lower half of [\(3.1\)](#), proved in [Section 4](#), extends these fixed-degree precedents to every  $1 \leq t \leq d$  while retaining a comparison against the full cone of nonnegative forms.

*Remark 3.2* (The nontrivial regime). The displayed lower bound exceeds the trivial bound 1 when  $t < d/(2e)$ . For larger  $t \leq d$ , both  $\sqrt{d/t}$  and the distortion are of constant order. The necessity statement in part 2 of [Theorem 3.1](#) uses the lower bound at  $t = \Theta_C(d)$ ; sufficiency follows from Barvinok's bound for all  $t$ , including  $t > d$  when  $C$  is close to 1.

The upper construction is forced by the facet geometry. The oriented facet normals of  $B_1^d$  are sign vectors, so the most symmetric facet-power form is their uniform average.

**Lemma 3.3** (The sign design). *For  $B_1^d$  and  $1 \leq t \leq d$ , the uniform design over the sign vectors  $\{\pm 1\}^d$  gives the form  $f(x) = \mathbb{E}_{\varepsilon \sim \{\pm 1\}^d}[(\varepsilon^\top x)^{2t}]$  with  $\rho(f^{1/(2t)}) = d/(\mathbb{E}[S_d^{2t}])^{1/(2t)}$ , and*

$$\frac{\sqrt{d}}{((2t-1)!!)^{1/(2t)}} \leq \rho(f^{1/(2t)}) \leq e \sqrt{\frac{d}{t}}.$$

If  $\binom{t}{2} < d$ , the upper bound can be sharpened to

$$\rho(f^{1/(2t)}) \leq \frac{\sqrt{d}}{((2t-1)!!)^{1/(2t)} (1 - \binom{t}{2}/d)^{1/(2t)}}.$$

Thus  $\kappa_t(B_1^d) \leq \rho(f^{1/(2t)})$ , and the ratio is  $\sqrt{e/2} \sqrt{d/t} (1 + o(1))$  when  $t \rightarrow \infty$  and  $t = o(\sqrt{d})$ .

*Proof idea.* The form  $f$  is invariant under signed permutations. Equalizing two coordinates while preserving their  $\ell_1$  mass can only lower  $f$ ; equivalently, Schur convexity pins the maximum, 1, at a vertex  $e_1$  and the minimum,  $d^{-2t}\mathbb{E}[S_d^{2t}]$ , at the diagonal  $x^{(d)}$ . Geometry is thereby reduced to one Rademacher moment. Gaussian comparison bounds this moment from above, while the collision-free perfect-matching terms in its expansion already give the correct lower scale. See [Section 5](#).  $\square$

### 3.2 The lower bound, in one variable

The main new ingredient is the lower bound. It uses two distinct compressions. Symmetrization replaces an arbitrary form by an invariant one without improving the candidate's ratio. Flat-point evaluation then replaces the remaining invariant algebra by a support-size curve. Only the second step makes the problem one-dimensional.

**Lemma 3.4** (Symmetrization and flat-point reduction). *Let  $p$  be any nonnegative form of degree  $2t$ , positive on the  $\ell_1$ -sphere. Then:*

1. *averaging  $p$  over  $B_d$  preserves nonnegativity and does not increase its sandwich ratio; if  $p$  is SoS, its average is SoS as well. Consequently both infima in (2.1) may be restricted to  $B_d$ -invariant forms;*
2. *a  $B_d$ -invariant form is a linear combination of products of even power sums, and its values on the flat points are*

$$V(k) := p(x^{(k)}) = \sum_{\ell=1}^t b_\ell k^{\ell-2t} = k^{-2t} Q(k), \quad Q(k) = \sum_{\ell=1}^t b_\ell k^\ell,$$

*with  $Q$  a real polynomial of degree  $\leq t$  and  $Q(0) = 0$ ;*

3. *consequently  $\rho(p^{1/(2t)})^{2t} \geq \max_{k \in [d]} V(k) / \min_{k \in [d]} V(k)$ .*

After this reduction, all geometry used by the proof is contained in the positive numbers  $V(1), \dots, V(d)$ . The form away from the flat points, and any certificate representing its nonnegativity, have disappeared.

The remaining estimate is cleaner when the prescribed power of  $k$  is left as a parameter. This separates the geometric input, which determines  $\alpha$ , from the interpolation step, which sees only the gap between  $\alpha$  and the available polynomial degree.

**Lemma 3.5** (Weighted reciprocal-grid profile lemma). *Let  $d, t$  be positive integers with  $1 \leq t \leq d$ , let  $\alpha \in \mathbb{R}$  satisfy  $\alpha \geq t$ , and let  $Q$  be a real polynomial with  $\deg Q \leq t$  and  $Q(0) = 0$ . If  $V(k) = k^{-\alpha}Q(k) > 0$  for every  $k \in [d]$ , put  $\beta = \alpha - t$  and let  $\ell_i$  be the Lagrange basis polynomial for the reciprocal nodes  $1, 1/2, \dots, 1/t$ . Then*

$$\frac{\max_{k \in [d]} V(k)}{\min_{k \in [d]} V(k)} \geq \max \left\{ 1, \frac{d^\beta}{\sum_{i=1}^t i^\beta |\ell_i(1/d)|} \right\} \geq \max \left\{ 1, 2(2e)^{-t} \left( \frac{d}{t} \right)^\beta \right\}. \quad (3.2)$$

*For  $t = 1$ , the exact ratio is  $d^{\alpha-1}$ .*

The exponent gap in (3.2) is structural. At  $\alpha = t$ , the choice  $Q(k) = k^t$  gives a constant profile, so no dimension growth is possible without additional information. When  $\alpha > t$ , the known weight creates the possibility of dimension growth, but the conclusion becomes nontrivial only when the displayed weighted interpolation term exceeds one. On a short grid an exactly level profile may still exist; this is why the maximum with 1 in (3.2) is essential.

**Lemma 3.6** (A discrete univariate bound). *There is an absolute constant  $c_1 = 1/\sqrt{2e}$  such that for every  $d \geq 1$ , every  $1 \leq t \leq d$ , and every real polynomial  $Q$  of degree  $\leq t$  with  $Q(0) = 0$  and  $Q(k) > 0$  for all  $k \in [d]$ , the sequence  $V(k) = k^{-2t}Q(k)$  satisfies*

$$\frac{\max_{k \in [d]} V(k)}{\min_{k \in [d]} V(k)} \geq 2 \left( \frac{d}{2et} \right)^t \geq (c_1 \sqrt{d/t})^{2t}.$$

For  $t = 1$ , the proof recovers the exact ratio  $d$  before the final factorial relaxation. The displayed lower bound is at least 2 whenever  $t \leq d/(2e)$ .

Together, Lemmas 3.4 and 3.6 give the stronger bound  $\kappa_t^{\text{nn}}(B_1^d) \geq c_1 \sqrt{d/t}$ . The weighted version also removes a repeated interpolation calculation from the  $\ell_p$  proof. We sketch its argument here and give the full proof in Section 4.

*Proof idea for Lemma 3.5.* Put  $\beta = \alpha - t \geq 0$  and write  $V(1/u) = u^\beta R(u)$ , where  $R$  has degree at most  $t - 1$ . Its values at  $u = 1, 1/2, \dots, 1/t$  determine  $R(1/d)$  exactly. The data satisfy  $R(1/i) = i^\beta V(i)$ , while  $V(d) = d^{-\beta} R(1/d)$ ; if all  $V(k)$  were nearly equal, the factor  $(t/d)^\beta$  would force the last value to be too small. The reciprocal substitution thus exposes both the effective degree and the exponent gap. Bounding the Lagrange weights and using  $t! \geq (t/e)^t$  gives (3.2); setting  $\alpha = 2t$  gives Lemma 3.6.  $\square$

### 3.3 A same-degree contrast with the cube

Polarity is a stress test for the obstruction. The cube construction of Fact 2.1 and the cross-polytope lower bound can be compared at the same degree without knowing the cube's exact optimum.

**Corollary 3.7** (Polar bodies at the same degree). *Let  $t_C^*(K) = \min\{t \geq 1 : \kappa_t(K) \leq C\}$ . For every fixed  $C > 1$  and every  $d \geq 2$ ,*

$$t_C^*(B_\infty^d) \leq \left\lceil \frac{\log d}{2 \log C} \right\rceil, \quad t_C^*(B_1^d) = \Theta_C(d).$$

*In particular, for  $d \geq 2$  and  $t = \lceil \log_2 d \rceil$ ,*

$$\kappa_t(B_\infty^d) \leq \sqrt{2}, \quad \kappa_t(B_1^d) \geq \frac{1}{\sqrt{2e}} \sqrt{\frac{d}{\lceil \log_2 d \rceil}}.$$

*Thus an explicit logarithmic-degree SoS form already gives constant distortion for the cube, while the polar cross-polytope still has diverging distortion and requires degree of order  $d$  for a constant factor.*

### 3.4 An oracle max–min design

Facet-power forms lead to a useful certificate problem. Its logic has four steps: the boundary floor is concave in the facet weights; a worst boundary point is a supergradient; homogeneity converts the nonconvex-looking boundary search into dual-norm optimization; and at  $t = 1$  the entire picture collapses to classical optimal design. The result below states exactly what this optimization certifies.

**Theorem 3.8** (Optimization of a facet-design bound). *For a symmetric polytope  $P = \{x : |a_i^\top x| \leq 1, i \in [n]\}$ , whose facet normals span  $\mathbb{R}^d$ , let  $\Phi_t(w) = \min_{\|x\|_P=1} f_{t,w}(x)$  and define the facet-design certificate*

$$\beta_t(P) = \left( \max_{w \in \Delta_n} \Phi_t(w) \right)^{-1/(2t)}.$$

*The maximum is attained and is strictly positive, so this certificate is finite. Then:*

1.  $\Phi_t$  is concave and takes values in  $[0, 1]$ . Moreover,

$$\kappa_t^{\text{nn}}(P) \leq \kappa_t(P) \leq \inf_{w \in \Delta_n} \rho_P(f_{t,w}^{1/(2t)}) \leq \beta_t(P).$$

2. For  $n \geq 2$ , let  $T \geq 1$  and  $\delta \geq 0$ . Suppose an oracle, given  $w$ , returns  $x_w \in \partial P$  with  $f_{t,w}(x_w) \leq \Phi_t(w) + \delta$ . Entropic mirror ascent with

$$w_i^+ = \frac{w_i \exp(\eta(a_i^\top x_w)^{2t})}{\sum_j w_j \exp(\eta(a_j^\top x_w)^{2t})}$$

initialized at the uniform distribution, with  $\eta = \sqrt{2 \log n / T}$ , has an averaged iterate  $\bar{w}_T$  satisfying

$$\max_{w \in \Delta_n} \Phi_t(w) - \Phi_t(\bar{w}_T) \leq \sqrt{\frac{2 \log n}{T}} + \delta.$$

Thus, for every  $\varepsilon > 0$ , an exact oracle gives additive  $\varepsilon$ -accuracy in  $T = O(\max\{1, \log n\}/\varepsilon^2)$  calls. The case  $n = 1$  is immediate.

3. For  $w$  in the relative interior of the simplex, put  $h_w = f_{t,w}^{1/(2t)}$ . Then  $h_w$  is a norm and

$$\Phi_t(w)^{-1/(2t)} = \max_{i \in [n]} h_w^*(a_i), \quad h_w^*(a_i) = \max\{a_i^\top x : f_{t,w}(x) \leq 1\}.$$

Thus the worst-direction oracle reduces exactly to  $n$  convex norm-dual problems.

4. At  $t = 1$ , for  $M(w) = \sum_i w_i a_i a_i^\top \succ 0$ ,

$$\Phi_1(w) = \frac{1}{\max_i a_i^\top M(w)^{-1} a_i}.$$

Hence  $\max_w \Phi_1(w) = 1/d$  and its maximizers are the  $D$ -optimal designs by the Kiefer–Wolfowitz equivalence theorem. For any maximizer  $w^*$ , the ellipsoid  $\{x : d x^\top M(w^*) x \leq 1\}$  is the symmetric John ellipsoid.

**Proof mechanism.** The ceiling  $\max_{\partial P} f_{t,w} \leq 1$  turns certificate design into the problem of raising a concave boundary floor. A minimizing boundary point exposes a supergradient, so entropic mirror ascent repeatedly reweights the facet directions that see the least-protected point. Homogeneity then replaces the boundary equality constraint by dual norms on the unit ball of  $h_w$ . At  $t = 1$ , these dual norms are leverage scores and the construction reduces to classical  $D/G$ -optimal design.

**What the theorem does not optimize.** The theorem optimizes the certificate  $\beta_t$ ; it neither minimizes the true ratio within the facet-power family nor characterizes the full SoS optimum. The norm-dual identity removes the apparent nonconvexity of the boundary oracle, but its numerical cost and conditioning remain separate from the outer iteration bound.

Before treating the optimized certificate as a proxy for the SoS optimum, one should test the smallest symmetric nonquadratic case. A mixed quartic term already separates the two notions.

**Proposition 3.9** (A symmetric design gap). *For the square and quartic forms,*

$$\kappa_2(B_\infty^2) \leq \left(\frac{4}{3}\right)^{1/4} < 2^{1/4} = \beta_2(B_\infty^2).$$

The same strict gap holds for  $B_1^2$ , which is linearly equivalent to  $B_\infty^2$ . Thus signed-permutation symmetry alone does not make the facet-power certificate optimal over the full SoS cone.

### 3.5 Extensions to $\ell_p$ balls

The  $\ell_p$  family stress-tests the flat-point mechanism by changing one exponent while preserving the symmetry. It reveals a methodological transition at  $p = 2$ . We state only the ranges supported by the proof; determining the interior regime  $p > 2$  remains open.

**Theorem 3.10** (Three consequences for  $\ell_p$  balls). *For  $1 \leq t \leq d$  and  $p \in [1, \infty]$ , with the convention  $1/\infty = 0$ :*

1. *If  $p = 2j$  for a divisor  $j \mid t$ , then  $\kappa_t^{\text{nn}}(B_p^d) = \kappa_t(B_p^d) = 1$ .*
2. *At degree two,  $\kappa_1(B_p^d) = d^{1/2-1/p}$ .*
3. *If  $1 \leq p \leq 2$ , then*

$$\frac{1}{\sqrt{2e}} \left( \frac{d}{t} \right)^{1/p-1/2} \leq \kappa_t^{\text{nn}}(B_p^d) \leq \kappa_t(B_p^d) \leq d^{1/p-1/2}.$$

*Consequently, for fixed  $p < 2$  and  $\log t = o(\log d)$ , both distortions equal  $d^{1/p-1/2+o(1)}$ .*

The resonance statement in [Theorem 3.10](#) occurs because  $\|x\|_{2j}^{2t} = (\sum_i x_i^{2j})^{t/j}$  is itself an SoS form of degree  $2t$ . For  $1 \leq p \leq 2$ , evaluating an invariant form at  $k^{-1/p}(1, \dots, 1, 0, \dots, 0)$  replaces the exponent  $2t$  in [Lemma 3.4](#) by  $2t/p$ , after which [Lemma 3.5](#) applies directly. For  $p > 2$  the exponent crosses the interpolation degree, and this proof no longer determines the distortion. This is a boundary of the argument, not evidence for any particular law beyond it.

## 4 The obstruction: from a boundary problem to a grid

This section proves the stronger inequality  $\kappa_t^{\text{nn}}(B_1^d) \geq c_1 \sqrt{d/t}$ , where  $c_1 = 1/\sqrt{2e}$ . The proof performs three reductions, each discarding information in a direction safe for a lower bound. We first remove orientation by averaging, then choose an orbit family on which the invariant algebra remembers only support size, and finally invert that support-size grid so that ordinary polynomial interpolation applies. The decisive quantitative step is to use the classical flat-ray family as a distortion witness. Classifying invariant nonnegative forms directly would still leave many partition-indexed terms; on these rays all of those terms collapse onto support size. Once the profile  $k^{-2t}Q(k)$  appears, the remaining obstruction is one-dimensional.

### 4.1 Choosing the witness profile

**Lemma 3.4** (Symmetrization and flat-point reduction). *Let  $p$  be any nonnegative form of degree  $2t$ , positive on the  $\ell_1$ -sphere. Then:*

1. *averaging  $p$  over  $B_d$  preserves nonnegativity and does not increase its sandwich ratio; if  $p$  is SoS, its average is SoS as well. Consequently both infima in (2.1) may be restricted to  $B_d$ -invariant forms;*
2. *a  $B_d$ -invariant form is a linear combination of products of even power sums, and its values on the flat points are*

$$V(k) := p(x^{(k)}) = \sum_{\ell=1}^t b_\ell k^{\ell-2t} = k^{-2t} Q(k), \quad Q(k) = \sum_{\ell=1}^t b_\ell k^\ell,$$

*with  $Q$  a real polynomial of degree  $\leq t$  and  $Q(0) = 0$ ;*

3. *consequently  $\rho(p^{1/(2t)})^{2t} \geq \max_{k \in [d]} V(k) / \min_{k \in [d]} V(k)$ .*

*Proof.* Write  $S = \{x : \|x\|_1 = 1\}$  for the  $\ell_1$ -sphere and  $G = B_d$  for the group of signed permutations. Since signed permutations preserve  $\ell_1$ , the set  $S$  is  $G$ -invariant.

(1) *Symmetrization.* Let  $p$  be nonnegative of degree  $2t$ , positive on  $S$ . Define the orbit average  $\bar{p}(x) = \frac{1}{|G|} \sum_{U \in G} p(U^{-1}x)$ ; it is nonnegative, of degree  $2t$ , and  $G$ -invariant. For each  $x \in S$  every point  $U^{-1}x$  lies in  $S$ , so  $\min_S p \leq p(U^{-1}x) \leq \max_S p$ , and averaging preserves the bracket:  $\min_S p \leq \bar{p}(x) \leq \max_S p$  for all  $x \in S$ . Taking extrema over  $x$  gives  $\max_S \bar{p} \leq \max_S p$  and  $\min_S \bar{p} \geq \min_S p$ , hence  $\rho(\bar{p}^{1/(2t)}) \leq \rho(p^{1/(2t)})$ . If  $p = \sum_r q_r^2$  is SoS, then  $\bar{p} = |G|^{-1} \sum_{U,r} (q_r \circ U^{-1})^2$  is SoS as well. It follows that the infimum over either polynomial cone is unchanged when restricted to  $G$ -invariant forms.

This bracket argument is the entire reason averaging is safe: orbit averaging may pull a maximum down and push a minimum up, but it cannot increase the ratio. This is exactly the direction needed when minimizing over candidates. Notice also that no positivity representation has been used.

(2) *Power-sum coordinates.* A  $G$ -invariant polynomial is invariant under sign flips — so each variable appears to even powers, i.e. it is a polynomial in  $y_i = x_i^2$  — and under permutations, so it is a symmetric polynomial in  $y_1, \dots, y_d$ . By the fundamental theorem of symmetric polynomials and Newton identities, these are generated by the power sums  $\sum_i y_i^k = \sum_i x_i^{2k} = p_{2k}(x)$ . A degree- $2t$  form is therefore a combination of monomials  $\prod_k p_{2k}^{m_k}$  with  $\sum_k k m_k = t$ . On a flat point,  $p_{2k}(x^{(j)}) = j \cdot (1/j)^{2k} = j^{1-2k}$ , so a monomial with  $\ell = \sum_k m_k$  factors evaluates to  $j^{\ell-2t}$  (using  $\sum_k k m_k = t$ ), where  $1 \leq \ell \leq t$  ( $\ell = 1$  for the single factor  $p_{2t}$ ,  $\ell = t$  for  $p_2^t$ ). Aggregating by  $\ell$ ,

$$V(k) = p(x^{(k)}) = \sum_{\ell=1}^t b_\ell k^{\ell-2t} = k^{-2t} Q(k), \quad Q(k) = \sum_{\ell=1}^t b_\ell k^\ell,$$

so  $\deg Q \leq t$  and  $Q(0) = 0$ .

This is the information collapse on which the proof turns. At  $x^{(j)}$ , each power-sum factor contributes one copy of  $j$ , whereas the weighted degree contributes the common factor  $j^{-2t}$ . The shape of the partition indexing the invariant monomial disappears; only its number  $\ell$  of factors remains. A large invariant coefficient space has therefore become one polynomial of degree at most  $t$ .

(3) *Flat points bound the ratio.* The  $d$  flat points lie on  $S$ , so  $\max_S p \geq \max_{k \in [d]} V(k)$  and  $\min_S p \leq \min_{k \in [d]} V(k)$ ; dividing,  $\rho(p^{1/(2t)})^{2t} = \max_S p / \min_S p \geq \max_k V(k) / \min_k V(k)$ .  $\square$

*Remark 4.1.* Part (3) is a *lower* bound on  $\rho$ : it does not claim the worst direction of a general invariant form is a flat point. This is precisely why the witness strategy is economical. Locating the true extrema could require understanding the form on the entire boundary; the proof needs only  $d$  directions that already force substantial variation.

## 4.2 Weighted grid rigidity by interpolation

*Proof of Lemma 3.5.* Fix  $d, t, \alpha$  and a valid  $Q$ . Put  $\beta = \alpha - t \geq 0$ , and write  $s = \max_{k \in [d]} V(k)$  and  $m = \min_{k \in [d]} V(k)$ , both positive.

The goal is to use the first  $t$  support sizes to control  $V(d)$ . In its original variable,  $V(k)$  is a weighted polynomial profile. Passing to reciprocals separates the known power  $u^\beta$  from the low-degree part that is free to move. This is also why  $\beta = \alpha - t$  is the relevant parameter rather than  $\alpha$  alone.

*Change of variable.* Since  $Q(0) = 0$  and  $\deg Q \leq t$ , write  $Q(k) = \sum_{j=1}^t a_j k^j$ , so  $V(k) = \sum_{j=1}^t a_j k^{j-\alpha}$ . With  $u = 1/k$ ,

$$V(1/u) = u^\beta R(u), \quad R(u) = \sum_{j=1}^t a_j u^{t-j}, \quad \deg R \leq t-1.$$

Thus  $R(1/i) = i^\beta V(i)$  and  $V(d) = d^{-\beta} R(1/d)$ .

*Interpolate at the  $t$  largest grid points.* Let  $u_i = 1/i$  for  $i \in [t]$  and put  $u_0 = 1/d$ . Since  $\deg R \leq t-1$ , Lagrange interpolation through  $u_1, \dots, u_t$  is exact at  $u_0$ :

$$R(u_0) = \sum_{i=1}^t R(u_i) \ell_i(u_0), \quad \ell_i(u_0) = \prod_{\substack{1 \leq j \leq t \\ j \neq i}} \frac{u_0 - u_j}{u_i - u_j}.$$

Put  $L_i = |\ell_i(u_0)|$ . Since  $R(u_i) = i^\beta V(i)$  and  $V(i) \leq s$ ,

$$V(d) = d^{-\beta} |R(u_0)| \leq d^{-\beta} s \sum_{i=1}^t i^\beta L_i. \quad (4.1)$$

When  $t = d$ ,  $u_0 = u_t$ , so  $L_t = 1$  and the other weights vanish; the formula simply reads  $V(d) \leq s$ .

*Bound the interpolation weights.* With  $u_0 = 1/d$  and  $u_j = 1/j$  ( $1 \leq j \leq t \leq d$ ),

$$|u_0 - u_j| = \frac{d-j}{dj} \leq \frac{1}{j}, \quad |u_i - u_j| = \frac{|j-i|}{ij}.$$

Hence  $\prod_{j \neq i} |u_0 - u_j| \leq \prod_{j \neq i} \frac{1}{j} = \frac{i}{t!}$  and  $\prod_{j \neq i} |u_i - u_j| = \frac{(i-1)!(t-i)!}{i^{t-2}t!}$ , so

$$L_i \leq \frac{i/t!}{(i-1)!(t-i)!/(i^{t-2}t!)} = \frac{i^{t-1}}{(i-1)!(t-i)!} = \frac{i^{t-1}}{(t-1)!} \binom{t-1}{i-1}.$$

Summing, with  $i^{t-1} \leq t^{t-1}$  and  $\sum_{i=1}^t \binom{t-1}{i-1} = 2^{t-1}$ ,

$$\sum_{i=1}^t L_i \leq \frac{t^{t-1}}{(t-1)!} \cdot 2^{t-1}. \quad (4.2)$$

The weighted evaluation norm  $\sum_i i^\beta L_i$  gives the first bound in (3.2). The simpler constant uses only its  $\ell_1$  relaxation; the signs of the Lagrange weights and the exact extremizing polynomial are irrelevant for that estimate.

*Assemble.* Since  $m \leq V(d)$  and  $s/m \geq 1$ , (4.1) first gives

$$\frac{s}{m} \geq \max \left\{ 1, \frac{d^\beta}{\sum_i i^\beta L_i} \right\}.$$

Because  $i^\beta \leq t^\beta$ , (4.2) and  $t! \geq (t/e)^t$  (from  $e^t \geq t^t/t!$ ) give  $(t-1)! \geq t^{t-1}/e^t$  and therefore

$$\frac{s}{m} \geq \max \left\{ 1, \frac{(t-1)!}{2^{t-1}t^{t-1}} \left( \frac{d}{t} \right)^\beta \right\} \geq \max \left\{ 1, 2(2e)^{-t} \left( \frac{d}{t} \right)^\beta \right\}.$$

At  $t = 1$ , positivity forces  $Q(k) = a_1 k$  with  $a_1 > 0$ , so  $V(k) = a_1 k^{1-\alpha}$  and the exact ratio is  $d^{\alpha-1}$ .

The estimate should be read backwards as follows. If all of the first  $t$  profile values stay below  $s$ , they determine the low-degree part  $R$  and force the diagonal value  $V(d)$  to pay the factor  $(t/d)^\beta$ , up to the weighted interpolation norm. Once that weighted term exceeds one, some sampled values must separate. The factorials quantify the cost of extrapolating from the first  $t$  reciprocal scales to the last one. At  $\beta = 0$ , the example  $Q(k) = k^t$  shows why the method correctly produces no dimension growth; for  $\beta > 0$  on a short grid, the maximum with one still cannot be discarded.  $\square$

*Proof of Lemma 3.6.* Specialize the preceding calculation to  $\alpha = 2t$ , hence  $\beta = t$ . Dropping the maximum with 1 gives

$$\frac{\max_k V(k)}{\min_k V(k)} \geq 2(2e)^{-t} \left(\frac{d}{t}\right)^t = 2 \left(\frac{d}{2et}\right)^t.$$

Dropping the leading factor 2 gives  $((2e)^{-1/2} \sqrt{d/t})^{2t}$ . At  $t = 1$ , the exact statement in Lemma 3.5 gives ratio  $d$ ; and the displayed coarse bound is at least 2 whenever  $t \leq d/(2e)$ .  $\square$

*Proof of the lower bound in Theorem 3.1.* Take any nonnegative degree- $2t$  form that is positive away from the origin. By Lemma 3.4, after averaging we obtain a polynomial  $Q$  with degree at most  $t$ , with  $Q(0) = 0$  and  $Q(k) > 0$  on  $[d]$ , such that the form's ratio to the power  $2t$  is at least  $\max_k k^{-2t} Q(k) / \min_k k^{-2t} Q(k)$ . By Lemma 3.6, this is at least  $(c_1 \sqrt{d/t})^{2t}$ . Taking the infimum over all nonnegative forms gives  $\kappa_t^{\text{nn}}(B_1^d) \geq c_1 \sqrt{d/t}$ ; the SoS lower bound follows from  $\kappa_t^{\text{nn}} \leq \kappa_t$ .

The scope of the conclusion is now transparent. The candidate's strict positivity away from the origin ensures that the sampled values  $V(k)$  are positive and that their ratio is meaningful. The argument never asks how global nonnegativity is certified, which is why the lower bound survives the enlargement from SoS forms to all nonnegative forms.  $\square$

## 5 The construction and its facet generalization

### 5.1 Averaging the facet orbit

The oriented facet inequalities of the cross-polytope are  $\varepsilon^\top x \leq 1$  for  $\varepsilon \in \{\pm 1\}^d$ . Equivalently, modulo  $\varepsilon \sim -\varepsilon$ , opposite pairs form the slabs  $|\varepsilon^\top x| \leq 1$ . Because the power is even, averaging over all oriented normals merely duplicates each opposite pair and is the canonical construction preserving every symmetry of the body. The Rademacher variable below records the value of this orbit at the diagonal.

**Lemma 3.3** (The sign design). *For  $B_1^d$  and  $1 \leq t \leq d$ , the uniform design over the sign vectors  $\{\pm 1\}^d$  gives the form  $f(x) = \mathbb{E}_{\varepsilon \sim \{\pm 1\}^d} [(\varepsilon^\top x)^{2t}]$  with  $\rho(f^{1/(2t)}) = d/(\mathbb{E}[S_d^{2t}])^{1/(2t)}$ , and*

$$\frac{\sqrt{d}}{((2t-1)!!)^{1/(2t)}} \leq \rho(f^{1/(2t)}) \leq e \sqrt{\frac{d}{t}}.$$

If  $\binom{t}{2} < d$ , the upper bound can be sharpened to

$$\rho(f^{1/(2t)}) \leq \frac{\sqrt{d}}{((2t-1)!!)^{1/(2t)} (1 - \binom{t}{2}/d)^{1/(2t)}}.$$

Thus  $\kappa_t(B_1^d) \leq \rho(f^{1/(2t)})$ , and the ratio is  $\sqrt{e/2} \sqrt{d/t} (1 + o(1))$  when  $t \rightarrow \infty$  and  $t = o(\sqrt{d})$ .

*Proof.* Let

$$f(x) = \mathbb{E}_\varepsilon [(\varepsilon^\top x)^{2t}], \quad S = \{x : \|x\|_1 = 1\}.$$

The form is an average of squares of degree- $t$  forms and is therefore SoS. It is also invariant under signed permutations.

*Extrema on the  $\ell_1$ -sphere.* It is enough to work on the simplex  $\Delta = \{x \geq 0 : \sum_i x_i = 1\}$ . Fix two coordinates, write them as  $(m + \delta, m - \delta)$ , and freeze the others. If  $c_\varepsilon = \sum_{\ell \neq i,j} \varepsilon_\ell x_\ell$ , then

$$g(\delta) = \mathbb{E}_\varepsilon \left[ (c_\varepsilon + m(\varepsilon_i + \varepsilon_j) + \delta(\varepsilon_i - \varepsilon_j))^{2t} \right]$$

is convex in  $\delta$ . Swapping  $\varepsilon_i$  and  $\varepsilon_j$  shows that it is even, hence nondecreasing in  $|\delta|$ . Equalizing two coordinates can only decrease  $f$ , so  $f$  is Schur-convex on  $\Delta$ . The uniform point  $d^{-1}\mathbf{1}$  is majorized by every point in  $\Delta$ , while  $e_1$  majorizes every such point. Consequently,

$$\max_{x \in S} f(x) = f(e_1) = 1, \quad \min_{x \in S} f(x) = f(d^{-1}\mathbf{1}) = d^{-2t} \mathbb{E}[S_d^{2t}].$$

This proves the exact ratio

$$\rho(f^{1/(2t)}) = \frac{d}{(\mathbb{E}[S_d^{2t}])^{1/(2t)}}.$$

The two-coordinate calculation has a simple geometric reading. Moving  $\ell_1$  mass from a larger coordinate to a smaller one lowers the even signed moment. Repeating these elementary balancing transfers—often called Robin Hood transfers—leads to the diagonal; reversing them leads to a vertex. This is why no other boundary directions enter the exact ratio.

*Moment estimates.* The geometry has now been reduced to determining the scale of one moment. To obtain distortion  $\sqrt{d/t}$ , this moment should have scale  $(dt)^t$ . Gaussian replacement gives its ceiling, while the collision-free pairing terms already give its floor. The Rademacher moment satisfies

$$(2t-1)!! d^{\downarrow t} \leq \mathbb{E}[S_d^{2t}] \leq (2t-1)!! d^t, \quad (5.1)$$

where  $d^{\downarrow t} = d(d-1)\cdots(d-t+1)$ . For the upper bound, replace the signs one at a time by standard Gaussians. In the binomial expansion, the two variables have the same moments in degrees zero and two, while every higher even Gaussian moment dominates its Rademacher counterpart. Iteration gives

$$\mathbb{E}[S_d^{2t}] \leq \mathbb{E}[N(0, d)^{2t}] = (2t-1)!! d^t.$$

For the lower bound, expand the moment and retain only those monomials in which  $t$  distinct indices each occur twice. There are  $(2t-1)!! d^{\downarrow t}$  such monomials, each with expectation one; all other surviving monomials are nonnegative. This proves (5.1).

The upper half of (5.1) gives

$$\rho(f^{1/(2t)}) \geq \frac{\sqrt{d}}{((2t-1)!!)^{1/(2t)}}.$$

For a uniform upper bound, use  $(2t-1)!! \geq t! \geq (t/e)^t$  and  $d^{\downarrow t} = t! \binom{d}{t} \geq (d/e)^t$ . The lower half of (5.1) then gives

$$\rho(f^{1/(2t)}) \leq e\sqrt{d/t}.$$

If  $\binom{t}{2} < d$ , the elementary inequality  $d^{\downarrow t} \geq d^t(1 - \binom{t}{2}/d)$  gives the sharper estimate in Lemma 3.3. Finally, when  $t = o(\sqrt{d})$ ,  $d^{\downarrow t} = d^t(1 + o(1))$ ; together with  $((2t-1)!!)^{1/(2t)} = \sqrt{2t/e}(1 + o(1))$ , this yields the stated asymptotic value.

Thus the factor  $\sqrt{t}$  gained over the quadratic approximation is already visible combinatorially: it is the  $2t$ -th root of the number of pairings that survive in the moment expansion.  $\square$

## 5.2 Optimizing a restricted construction

The sign design suggests allowing nonuniform weights on the facet normals of a general polytope. Every term  $(a_i^\top x)^{2t}$  is at most one on the boundary, so the normalized construction has a built-in ceiling. The quantity that remains to protect is its floor: maximize the smallest boundary value over the weights. This is the origin of  $\Phi_t$  and of the one-sided certificate  $\beta_t$ .

The associated matrix

$$M_t(w) = \sum_i w_i v_t(a_i) v_t(a_i)^\top$$

is the information matrix for degree- $t$  polynomial features, which makes classical optimal-design tools available. The objective here is nevertheless different from the full SoS problem: it optimizes the boundary floor of one restricted family of facet-power forms.

**Theorem 3.8** (Optimization of a facet-design bound). *For a symmetric polytope  $P = \{x : |a_i^\top x| \leq 1, i \in [n]\}$ , whose facet normals span  $\mathbb{R}^d$ , let  $\Phi_t(w) = \min_{\|x\|_P=1} f_{t,w}(x)$  and define the facet-design certificate*

$$\beta_t(P) = \left( \max_{w \in \Delta_n} \Phi_t(w) \right)^{-1/(2t)}.$$

*The maximum is attained and is strictly positive, so this certificate is finite. Then:*

1.  $\Phi_t$  is concave and takes values in  $[0, 1]$ . Moreover,

$$\kappa_t^{\text{nn}}(P) \leq \kappa_t(P) \leq \inf_{w \in \Delta_n} \rho_P(f_{t,w}^{1/(2t)}) \leq \beta_t(P).$$

2. For  $n \geq 2$ , let  $T \geq 1$  and  $\delta \geq 0$ . Suppose an oracle, given  $w$ , returns  $x_w \in \partial P$  with  $f_{t,w}(x_w) \leq \Phi_t(w) + \delta$ . Entropic mirror ascent with

$$w_i^+ = \frac{w_i \exp(\eta(a_i^\top x_w)^{2t})}{\sum_j w_j \exp(\eta(a_j^\top x_w)^{2t})}$$

*initialized at the uniform distribution, with  $\eta = \sqrt{2 \log n / T}$ , has an averaged iterate  $\bar{w}_T$  satisfying*

$$\max_{w \in \Delta_n} \Phi_t(w) - \Phi_t(\bar{w}_T) \leq \sqrt{\frac{2 \log n}{T}} + \delta.$$

*Thus, for every  $\varepsilon > 0$ , an exact oracle gives additive  $\varepsilon$ -accuracy in  $T = O(\max\{1, \log n\}/\varepsilon^2)$  calls. The case  $n = 1$  is immediate.*

3. For  $w$  in the relative interior of the simplex, put  $h_w = f_{t,w}^{1/(2t)}$ . Then  $h_w$  is a norm and

$$\Phi_t(w)^{-1/(2t)} = \max_{i \in [n]} h_w^*(a_i), \quad h_w^*(a_i) = \max\{a_i^\top x : f_{t,w}(x) \leq 1\}.$$

*Thus the worst-direction oracle reduces exactly to  $n$  convex norm-dual problems.*

4. At  $t = 1$ , for  $M(w) = \sum_i w_i a_i a_i^\top \succ 0$ ,

$$\Phi_1(w) = \frac{1}{\max_i a_i^\top M(w)^{-1} a_i}.$$

*Hence  $\max_w \Phi_1(w) = 1/d$  and its maximizers are the  $D$ -optimal designs by the Kiefer–Wolfowitz equivalence theorem. For any maximizer  $w^*$ , the ellipsoid  $\{x : d x^\top M(w^*) x \leq 1\}$  is the symmetric John ellipsoid.*

*Proof.* Write  $c_i(x) = (a_i^\top x)^{2t}$ , so that  $f_{t,w}(x) = \langle c(x), w \rangle$ .

*1: concavity and soundness.* The function

$$\Phi_t(w) = \min_{x \in \partial P} \langle c(x), w \rangle$$

is a pointwise minimum of linear functions of  $w$ , hence concave. The boundary is compact, so the minimum is attained. Moreover  $0 \leq c_i(x) \leq 1$  on  $\partial P$ , which gives  $0 \leq \Phi_t(w) \leq 1$ . In fact,  $|\Phi_t(w) - \Phi_t(v)| \leq \|w - v\|_1$ , so  $\Phi_t$  is continuous on the compact simplex and its maximum is attained. At the uniform weight vector, spanning of the facet normals makes  $f_{t,w}$  positive away from the origin; compactness of  $\partial P$  then gives  $\Phi_t(w) > 0$ . Thus  $\beta_t(P)$  is finite.

Every  $f_{t,w}$  is SoS. If  $m(w) = \min_{\partial P} f_{t,w}$  and  $M(w) = \max_{\partial P} f_{t,w}$ , then  $M(w) \leq 1$  and

$$\rho_P(f_{t,w}^{1/(2t)}) = \left( \frac{M(w)}{m(w)} \right)^{1/(2t)} \leq \Phi_t(w)^{-1/(2t)}.$$

Taking infima proves

$$\kappa_t^{\text{nn}}(P) \leq \kappa_t(P) \leq \inf_w \rho_P(f_{t,w}^{1/(2t)}) \leq \beta_t(P).$$

The last step has two potential sources of slack: it replaces the actual maximum  $M(w)$  by the bound one, and it restricts the construction to facet-power forms rather than all SoS forms.

*2: mirror ascent with an approximate oracle.* Suppose  $x_s \in \partial P$  satisfies  $f_{t,w_s}(x_s) \leq \Phi_t(w_s) + \delta$ , and put  $g_{s,i} = c_i(x_s)$ . For every  $w \in \Delta_n$ ,

$$\Phi_t(w) \leq \langle g_s, w \rangle \leq \Phi_t(w_s) + \langle g_s, w - w_s \rangle + \delta.$$

Thus  $g_s$  is a  $\delta$ -supergradient, and  $\|g_s\|_\infty \leq 1$ . For the exponential update, any  $w \in \Delta_n$  satisfies

$$D_{\text{KL}}(w \| w_{s+1}) - D_{\text{KL}}(w \| w_s) = -\eta \langle g_s, w \rangle + \log \sum_i w_{s,i} e^{\eta g_{s,i}} \leq -\eta \langle g_s, w - w_s \rangle + \frac{\eta^2}{8},$$

For the last inequality, the second derivative of the log moment is the variance of a variable in  $[0, 1]$ , hence at most  $1/4$ ; integrating twice gives the remainder  $\eta^2/8$ . Telescoping from the uniform distribution and taking  $\eta = \sqrt{2 \log n / T}$  gives, for  $n \geq 2$ ,

$$\sum_{s=1}^T \langle g_s, w - w_s \rangle \leq \frac{\log n}{\eta} + \frac{\eta T}{8} \leq \sqrt{2T \log n} \quad \text{for every } w \in \Delta_n$$

for the normalized update in [Theorem 3.8](#); this is the negative-entropy specialization of the standard mirror-descent inequality ([Bubeck, 2015](#), Sec. 4.6, Eq. (4.10)). Averaging the approximate-supergradient inequalities and using concavity,

$$\max_w \Phi_t(w) - \Phi_t(\bar{w}_T) \leq \sqrt{\frac{2 \log n}{T}} + \delta.$$

If the oracle errors vary with  $s$ , the same argument replaces  $\delta$  by  $T^{-1} \sum_s \delta_s$ ; the case  $n = 1$  is immediate.

Geometrically, the minimizing boundary point exposes a direction that the current weights protect least well. Its coordinates  $c_i(x_s)$  measure how strongly each facet functional sees that direction, and the exponential update reweights according to this exposure. The regret inequality certifies the average effect; it does not assert that every individual update increases  $\Phi_t$ .

3: *a convex representation of the oracle.* The equality constraint  $\|x\|_P = 1$  makes the boundary search appear nonconvex. Homogeneity removes that fixed scale: the minimum of one norm on the boundary of another is the reciprocal of a maximum over a norm ball. Since the polytope gauge is itself the maximum of the facet functionals, the problem then separates into finitely many dual-norm evaluations. For positive weights,  $h_w = f_{t,w}^{1/(2t)}$  is a norm because the facet normals span  $\mathbb{R}^d$ . By homogeneity,

$$\min_{\|x\|_P=1} h_w(x) = \left( \max_{h_w(x) \leq 1} \|x\|_P \right)^{-1} = \left( \max_i h_w^*(a_i) \right)^{-1}.$$

Raising to the power  $2t$  proves the identity in [Theorem 3.8](#). Each dual norm value is the convex program

$$h_w^*(a_i) = \max_x \{a_i^\top x : \sum_j w_j (a_j^\top x)^{2t} \leq 1\}.$$

If  $i$  and  $x$  attain the two maxima, then  $h_w(x) = 1$  and dual Hölder gives  $|a_j^\top x| \leq h_w^*(a_j) \leq h_w^*(a_i) = a_i^\top x$  for every  $j$ . Hence  $\|x\|_P = a_i^\top x$ , and  $x/\|x\|_P$  is a minimizing boundary point. The reduction therefore gives both the oracle value and an oracle direction.

4: *the quadratic case.* At  $t = 1$ , every object in the construction should reduce to classical ellipsoidal design. Indeed, the dual norm values below become leverage scores, and Kiefer–Wolfowitz closes the loop. When  $t = 1$ ,  $h_w(x)^2 = x^\top M(w)x$  and  $h_w^*(a_i)^2 = a_i^\top M(w)^{-1}a_i$ . Hence

$$\Phi_1(w) = \frac{1}{\max_i a_i^\top M(w)^{-1}a_i}.$$

For every nonsingular design, the sensitivities  $\ell_i(w) = a_i^\top M(w)^{-1}a_i$  satisfy

$$\sum_i w_i \ell_i(w) = \text{tr}(M(w)^{-1}M(w)) = d,$$

so  $\max_i \ell_i(w) \geq d$ . Conversely, let  $w^*$  maximize  $\log \det M(w)$ . The directional derivative toward the  $i$ th simplex vertex is  $\ell_i(w^*) - d \leq 0$ , hence every sensitivity is at most  $d$ . Concavity of  $\log \det$  proves the converse as well: any design with  $\max_i \ell_i(w) \leq d$  is  $D$ -optimal. This is the finite-design proof of the Kiefer–Wolfowitz equivalence theorem ([Kiefer and Wolfowitz, 1960](#), p. 364). If  $M(w)$  is singular, a nonzero kernel vector can be scaled to  $\partial P$  and gives  $\Phi_1(w) = 0$ ; therefore every maximizer is nonsingular. We have proved

$$\min_{w \in \Delta_n} \max_i a_i^\top M(w)^{-1}a_i = d$$

and identified its minimizers with the  $D$ -optimal designs. If  $w^*$  is one of them, then

$$E_{w^*} = \{x : d x^\top M(w^*)x \leq 1\}$$

is the symmetric John ellipsoid of  $P$ . Indeed, writing an ellipsoid as  $E_X = \{x : x^\top X^{-1}x \leq 1\}$ , the primal–dual KKT relation is  $X^{-1} = dM(w^*)$ . The sensitivity inequalities give feasibility for  $X = (dM(w^*))^{-1}$ , and the identity above forces equality whenever  $w_i^* > 0$ ; the multipliers  $dw_i^*$  therefore give complementary slackness.

In simplex normalization, the iteration of [Cohen et al. \(2019\)](#) is

$$w_i^+ = \frac{w_i \ell_i(w)}{d}, \quad \ell_i(w) = a_i^\top M(w)^{-1}a_i.$$

Their weights sum to  $d$ ; the displayed formula follows after the change of variables  $v = dw$ . It optimizes the same  $D$ -optimal design problem, but it uses the log-determinant sensitivity  $\ell_i(w)$  rather than the worst-direction supergradient  $c_i(x_s)$  above. The optimizer is shared; the update is not.  $\square$

### 5.3 What the facet model misses

The generalization above creates a tempting but false inference: perhaps optimizing the symmetric facet-power family already optimizes over all SoS forms. The smallest nonquadratic case,  $d = t = 2$ , is enough to disprove it.

**Proposition 3.9** (A symmetric design gap). *For the square and quartic forms,*

$$\kappa_2(B_\infty^2) \leq \left(\frac{4}{3}\right)^{1/4} < 2^{1/4} = \beta_2(B_\infty^2).$$

*The same strict gap holds for  $B_1^2$ , which is linearly equivalent to  $B_\infty^2$ . Thus signed-permutation symmetry alone does not make the facet-power certificate optimal over the full SoS cone.*

*Proof.* For the square, consider

$$p(x, y) = x^4 + y^4 - x^2y^2 = (x^2 - y^2)^2 + (xy)^2.$$

This is an SoS form. On the boundary of  $B_\infty^2$ , one coordinate has squared magnitude one; if the other has squared magnitude  $z \in [0, 1]$ , then

$$p = 1 - z + z^2 \in [3/4, 1].$$

Therefore  $\kappa_2(B_\infty^2) \leq (4/3)^{1/4}$ .

The two opposite facet pairs of the square are represented by  $e_1, e_2$ . For  $w = (w_1, w_2)$ ,

$$\Phi_2(w) = \min_{\max\{|x|, |y|\}=1} (w_1x^4 + w_2y^4) = \min\{w_1, w_2\}.$$

Its maximum over the simplex is  $1/2$ , so  $\beta_2(B_\infty^2) = 2^{1/4}$ . Moreover, the boundary point  $(1, 1)$  gives  $w_1 + w_2 = 1$  for every  $w$ , so the boundary maximum is exactly one. Thus  $2^{1/4}$  is also the best actual ratio inside this facet-power family; the gap below is not caused by the ceiling relaxation.

Finally, the linear map  $(x, y) \mapsto (x + y, x - y)$  sends the  $\ell_1$ -norm to the  $\ell_\infty$ -norm. Distortion and the corresponding facet certificate are invariant under invertible linear changes of variables. Indeed, if  $A$  is invertible and  $b_i = A^{-\top}a_i$  are the transformed facet normals, then

$$\sum_i w_i (b_i^\top Ax)^{2t} = \sum_i w_i (a_i^\top x)^{2t}.$$

The map  $x \mapsto Ax$  therefore preserves the boundary extrema and  $\beta_t$ , while polynomial pullback gives the same invariance for  $\kappa_t$ . This yields the same strict gap for  $B_1^2$ . Explicitly, the transformed SoS form is  $x^4 + 14x^2y^2 + y^4$ .

The conceptual point is the mixed term. A nonnegative mixture of coordinate facet powers cannot use a negative  $x^2y^2$  coefficient, whereas the full SoS cone can: the interaction is absorbed by the square decomposition. Symmetry controls which terms may appear, but it does not collapse these two cones.  $\square$

*Remark 5.1* (What symmetry still gives). For the cube, symmetry gives  $\beta_t(B_\infty^d) = d^{1/(2t)}$ , but [Proposition 3.9](#) shows that this need not equal  $\kappa_t(B_\infty^d)$ . For the cross-polytope, concavity and group averaging make the uniform sign design optimal for  $\beta_t$ ; hence, for  $1 \leq t \leq d$ ,

$$\beta_t(B_1^d) = \frac{d}{(\mathbb{E}[S_d^{2t}])^{1/(2t)}} = \Theta(\sqrt{d/t}).$$

In this same range, the main theorem makes this certificate order-optimal, not exactly optimal, over the full SoS cone.

## 6 Discussion

**What the proof teaches.** The cross-polytope obstruction is carried by orbit geometry, not by a particular representation of nonnegativity. After averaging, dimension appears as the number of support scales  $k = 1, \dots, d$  on which a candidate must remain well balanced; degree appears as the complexity of the profile available to balance them. The flat points are effective not because they are known to contain every extremum, but because they make invariant forms forget almost everything: each restriction is  $k^{-2t}Q(k)$  with  $\deg Q \leq t$ . Interpolation then converts the mismatch between  $d$  scales and a degree- $t$  numerator into the factor  $\sqrt{d/t}$ .

This view also explains the strongest scope feature of the theorem. Once the sampled values are positive, the proof no longer sees whether the form is a sum of squares, a sum of powers, or merely globally nonnegative. The lower bound is therefore representation-free. The construction side uses the same symmetry differently: the facets form one sign-vector orbit, whose uniform power average produces the Rademacher form. The lower and upper bounds are thus two uses of one group—simplify every adversary, then generate one canonical match.

**A portable diagnostic.** The weighted form of [Lemma 3.5](#) separates the interpolation argument from the geometry that produces its exponent. To test another symmetric body, one may look for three ingredients: a group average that does not increase the ratio; boundary representatives indexed by a support parameter  $k \in [d]$ ; and a collapse of every averaged degree- $2t$  form to

$$V(k) = k^{-\alpha}Q(k), \quad \deg Q \leq t, \quad Q(0) = 0.$$

If these ingredients hold and  $\alpha \geq t$ , positivity on the representatives alone forces

$$\frac{\max_k V(k)}{\min_k V(k)} \geq 2(2e)^{-t} \left(\frac{d}{t}\right)^{\alpha-t}.$$

This is a test rather than a universal theorem about symmetric bodies: the body-specific work is precisely to find an orbit family on which the invariant algebra collapses in this way. For  $B_1^d$ ,  $\alpha = 2t$ ; for the  $\ell_p$  witnesses,  $\alpha = 2t/p$ . Thus the same lemma both proves the cross-polytope rate and locates the  $p = 2$  stopping point.

**A correct-order certificate need not be an optimizer.** The cube and the facet program give the same warning in two forms. The coordinate certificate  $\sum_i x_i^{2t}$  proves the cube–cross-polytope separation, but the quartic example in [Proposition 3.9](#) shows that it is not always optimal. Likewise, depending on the body,  $\beta_t$  may be looser than the best ratio in its own facet-power family; independently, that restricted family may be looser than the optimum over the full SoS cone.

For a weight vector  $w$ , the actual facet-form ratio is

$$\left(\frac{M(w)}{m(w)}\right)^{1/(2t)}, \quad m(w) = \min_{\partial P} f_{t,w}, \quad M(w) = \max_{\partial P} f_{t,w},$$

whereas  $\beta_t$  optimizes  $m(w)^{-1/(2t)}$  after using  $M(w) \leq 1$ . The normalization is one possible source of slack, and restricting to facet powers creates a second; [Proposition 3.9](#) proves that the latter occurs already for a symmetric quartic example. Indeed, for the square  $M(w) = 1$  for every  $w$ , since the boundary point  $(1, 1)$  gives  $w_1 + w_2 = 1$ ; the strict gap in [Proposition 3.9](#) therefore comes from the restricted form family, not from the ceiling relaxation.

The normalization slack also occurs by itself. For the regular hexagon described by three unit normals at angles  $0, \pi/3, 2\pi/3$ , uniform weights give information matrix  $I/2$  and boundary floor  $1/2$ . By [Theorem 3.8](#),

this design is optimal for the boundary-floor program, so  $\beta_1 = \sqrt{2}$ . Its boundary maximum is  $2/3$ , however, and the same facet form has actual ratio  $2/\sqrt{3} < \sqrt{2}$ . Dihedral averaging cannot increase a quadratic sandwich ratio and sends every quadratic form to a scalar multiple of  $\|x\|_2^2$ , so this Euclidean form is also optimal over the full quadratic cone. The example therefore isolates the ceiling relaxation. The value of the program is that its guarantee, optimizer, oracle, and failure mode are all explicit; it should not be identified with  $\kappa_t$ .

The quadratic endpoint is a useful consistency check, with one similar qualification. Maximizing  $\Phi_1$  is the  $G$ -optimal problem and produces a  $D$ -optimal design; after the factor- $d$  scaling in [Theorem 3.8](#), its information matrix defines the John ellipsoid. But  $\beta_1(P) = \sqrt{d}$  universally, while a particular body may have a better surrounding ellipsoidal ratio. What survives exactly at  $t = 1$  is the optimizer and the inscribed ellipsoid, not equality among every approximation objective.

**Sharp constants: which side must move?** For  $t \rightarrow \infty$  with  $t = o(\sqrt{d})$ , the sign design has asymptotic constant  $\sqrt{e/2}$ , whereas the interpolation lower bound has constant  $1/\sqrt{2e}$ . The factor- $e$  gap has two plausible sources, and the present argument does not choose between them. On the lower side, we bound only the  $\ell_1$  norm of the Lagrange evaluation functional and then relax a factorial. On the upper side, the quartic design gap warns that the uniform sign form may not be optimal in the full SoS cone. A sharper grid extremal inequality would move the first side; a construction using mixed SoS interactions would move the second. Determining which happens is the sharp-constant problem exposed by the proof.

**The cube: determine the optimum, not another certificate.** For the cube we know  $\kappa_t(B_\infty^d) \leq d^{1/(2t)}$ , which is enough for logarithmic-degree constant approximation. The next question is the order of the full SoS optimum for fixed  $t > 1$ . The two-dimensional quartic improvement rules out a proof that simply declares the coordinate certificate optimal by symmetry. Any successful lower bound must control the mixed invariant terms that facet powers omit.

**Beyond  $p = 2$ : replace the exponent-sensitive step.** For  $1 \leq p \leq 2$ , the flat-point exponent lies on the side where the same interpolation inequality yields a lower bound. At  $p > 2$  it crosses the degree of the numerator, and the direction of that comparison no longer controls oscillation. This is a structural stopping point of the proof, not a prediction of the answer. A concrete first target is the interval  $2 < p < 4$  at quartic degree, where the resonance at  $p = 4$  and the exact quadratic law provide two anchors. Progress there requires a replacement for the exponent-sensitive interpolation step, not more bookkeeping within the present estimate.

**Computation: protect a relative objective.** The outer mirror-ascent guarantee is additive in  $\Phi_t$ . The inner problem has an exact convex dual-norm formulation, but solving its  $n$  convex programs may dominate the running time, and the map  $\Phi_t \mapsto \Phi_t^{-1/(2t)}$  is ill-conditioned when the optimum is small. A useful algorithmic target is a relative-error guarantee under a quantitative lower bound on  $\max_w \Phi_t(w)$ , together with an oracle whose complexity tracks the conditioning of the induced norm. These questions concern the certificate program; they do not affect the geometric lower bound.

The research map is therefore modular: better constants require a sharper profile inequality or a new SoS construction; the exact cube problem requires control beyond facet powers; the region  $p > 2$  requires a new witness mechanism; and faster computation requires conditioning-aware dual norms. The proof does not merely leave these problems open; it locates each one at the step where the present framework loses the information a sharper result would need.

**AI Disclosure.** The author acknowledges the use of GPT-5.6 Solar as an assistive tool in preparing this manuscript, including support in drafting, revising, and polishing the exposition. All AI-assisted material was carefully reviewed and edited by the author, who takes full responsibility for the final manuscript and its results.

## References

- José Acevedo and Mauricio Velasco. Test sets for nonnegativity of polynomials invariant under a finite reflection group. *Journal of Pure and Applied Algebra*, 220(8):2936–2947, 2016. doi: 10.1016/j.jpaa.2016.01.010.
- Jose Acevedo, Grigoriy Blekherman, Sebastian Debus, and Cordian Riener. The wonderful geometry of the Vandermonde map. *Foundations of Computational Mathematics*, 2025. doi: 10.1007/s10208-025-09718-6.
- Amir Ali Ahmadi, Etienne de Klerk, and Georgina Hall. Polynomial norms. *SIAM Journal on Optimization*, 29(1):399–422, 2019. doi: 10.1137/18M1172843.
- Alexander Barvinok. Approximating a norm by a polynomial. In Vitali D. Milman and Gideon Schechtman, editors, *Geometric Aspects of Functional Analysis*, volume 1807 of *Lecture Notes in Mathematics*, pages 20–26. Springer, Berlin, Heidelberg, 2003. doi: 10.1007/978-3-540-36428-3\_2.
- Alexander Barvinok and Ellen Veomett. The computational complexity of convex bodies. In Jacob E. Goodman, János Pach, and Richard Pollack, editors, *Surveys on Discrete and Computational Geometry: Twenty Years Later*, volume 453 of *Contemporary Mathematics*, pages 117–137. American Mathematical Society, 2008.
- Grigoriy Blekherman. Volumes of nonnegative polynomials, sums of squares and powers of linear forms, 2004.
- Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3–4):231–357, 2015. doi: 10.1561/22000000050.
- Yang Cao, Xiaoyu Li, Zhao Song, Xin Yang, and Tianyi Zhou. Faster algorithms for structured John ellipsoid computation. In *Advances in Neural Information Processing Systems*, volume 38, 2025.
- Man-Duen Choi, Tsit-Yuen Lam, and Bruce Reznick. Even symmetric sextics. *Mathematische Zeitschrift*, 195(4):559–580, 1987. doi: 10.1007/BF01166704.
- Michael B. Cohen, Ben Cousins, Yin Tat Lee, and Xin Yang. A near-optimal algorithm for approximating the John ellipsoid. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the 32nd Conference on Learning Theory (COLT)*, volume 99 of *Proceedings of Machine Learning Research*, pages 849–873. PMLR, 2019.
- Don Coppersmith and Theodore J. Rivlin. The growth of polynomials bounded at equally spaced points. *SIAM Journal on Mathematical Analysis*, 23(4):970–983, 1992. doi: 10.1137/0523054.
- Lutz Dübgen, Sara A. van de Geer, Mark C. Veraar, and Jon A. Wellner. Nemirovski’s inequalities revisited. *The American Mathematical Monthly*, 117(2):138–160, 2010. doi: 10.4169/000298910X476059.

- Jack Kiefer and Jacob Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960. doi: 10.4153/CJM-1960-030-4.
- Xiaoyu Li, Junwei Yu, Jiaojiao Jiang, Junbin Gao, and Andi Han. Beyond averaging in John ellipsoid approximation: High-accuracy algorithms in the leverage-score model, 2026.
- Cordian Riener. On the degree and half degree principle for symmetric polynomials. *Journal of Pure and Applied Algebra*, 216(4):850–856, 2012. doi: 10.1016/j.jpaa.2011.08.012.
- Vlad Timofte. On the positivity of symmetric polynomial functions. part I: General results. *Journal of Mathematical Analysis and Applications*, 284(1):174–190, 2003. doi: 10.1016/S0022-247X(03)00301-9.
- David P. Woodruff and Taisuke Yasuda. John ellipsoids via lazy updates. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-2179.
- Yosef Yomdin. Remez-type inequality for discrete sets. *Israel Journal of Mathematics*, 186(1):45–60, 2011. doi: 10.1007/s11856-011-0131-4.

## A Deferred proofs

### A.1 Assembly of the main theorem

**Theorem 3.1** (Cross-polytope degree law). *For all  $d \geq 1$  and  $1 \leq t \leq d$ ,*

$$\frac{1}{\sqrt{2e}} \sqrt{\frac{d}{t}} \leq \kappa_t^{\text{nn}}(B_1^d) \leq \kappa_t(B_1^d) \leq \min \left\{ \binom{d+t-1}{t}^{1/(2t)}, \frac{d}{(\mathbb{E}[S_d^{2t}])^{1/(2t)}} \right\} \leq \sqrt{2e} \sqrt{\frac{d}{t}}. \quad (3.1)$$

*In particular, both the SoS and the nonnegative-form distortions are  $\Theta(\sqrt{d/t})$ . Moreover, the sign design has the sharper asymptotic value*

$$\frac{d}{(\mathbb{E}[S_d^{2t}])^{1/(2t)}} = \sqrt{\frac{e}{2}} \sqrt{\frac{d}{t}} (1 + o(1)) \quad \text{when } t \rightarrow \infty \text{ and } t = o(\sqrt{d}).$$

*Consequently:*

1. *the cross-polytope matches Barvinok’s universal upper bound up to an absolute constant, and is therefore extremal among symmetric bodies up to constants;*
2. *for every fixed  $C > 1$ , degree  $2t = \Theta_C(d)$  is necessary and sufficient for either distortion to be at most  $C$ .*

*Proof.* The lower-bound argument at the end of [Section 4](#) applies to every globally nonnegative form, and therefore gives

$$\frac{1}{\sqrt{2e}} \sqrt{\frac{d}{t}} \leq \kappa_t^{\text{nn}}(B_1^d) \leq \kappa_t(B_1^d).$$

Barvinok’s theorem supplies the first SoS upper bound in [\(3.1\)](#), and [Lemma 3.3](#) supplies the second. Finally,

$$\binom{d+t-1}{t} \leq \left( \frac{e(d+t-1)}{t} \right)^t \leq \left( \frac{2ed}{t} \right)^t$$

for  $1 \leq t \leq d$ , which proves the last inequality in (3.1). The more precise asymptotic value of the sign design was established in Section 5.

For the first consequence, the universal SoS upper bound is  $O(\sqrt{d/t})$  for every symmetric body, while the cross-polytope has the matching lower bound even over the larger nonnegative cone.

For the second consequence, if  $t \leq d$  and either distortion is at most  $C$ , then  $t \geq d/(2eC^2)$ ; if  $t > d$ , this lower bound on  $t$  is automatic. Conversely, the all-degree Barvinok bound implies that, for every fixed  $C > 1$ , some  $t = O_C(d)$  gives SoS distortion at most  $C$ , and hence also nonnegative-form distortion at most  $C$ . To see the linear dependence directly, the case  $d = 1$  being immediate, take  $t = \lceil cd \rceil$  for  $d \geq 2$  and use

$$\binom{d+t-1}{t} = \binom{d+t-1}{d-1} \leq \left( \frac{e(d+t-1)}{d-1} \right)^{d-1};$$

the  $1/(2t)$  power tends to one as the constant  $c$  tends to infinity.  $\square$

## A.2 The polar-body comparison

**Corollary 3.7** (Polar bodies at the same degree). *Let  $t_C^*(K) = \min\{t \geq 1 : \kappa_t(K) \leq C\}$ . For every fixed  $C > 1$  and every  $d \geq 2$ ,*

$$t_C^*(B_\infty^d) \leq \left\lceil \frac{\log d}{2 \log C} \right\rceil, \quad t_C^*(B_1^d) = \Theta_C(d).$$

*In particular, for  $d \geq 2$  and  $t = \lceil \log_2 d \rceil$ ,*

$$\kappa_t(B_\infty^d) \leq \sqrt{2}, \quad \kappa_t(B_1^d) \geq \frac{1}{\sqrt{2e}} \sqrt{\frac{d}{\lceil \log_2 d \rceil}}.$$

*Thus an explicit logarithmic-degree SoS form already gives constant distortion for the cube, while the polar cross-polytope still has diverging distortion and requires degree of order  $d$  for a constant factor.*

*Proof.* The cube certificate in Fact 2.1 gives

$$t_C^*(B_\infty^d) \leq \left\lceil \frac{\log d}{2 \log C} \right\rceil.$$

For the cross-polytope, the lower bound in Theorem 3.1 gives  $t_C^*(B_1^d) \geq d/(2eC^2)$ , while the all-degree Barvinok upper bound used above gives  $t_C^*(B_1^d) = O_C(d)$ . Thus  $t_C^*(B_1^d) = \Theta_C(d)$ .

If  $t = \lceil \log_2 d \rceil$ , then  $d^{1/(2t)} \leq \sqrt{2}$ , so the cube bound follows from Fact 2.1; the cross-polytope bound is (3.1).  $\square$

## A.3 The $\ell_p$ family

**Theorem 3.10** (Three consequences for  $\ell_p$  balls). *For  $1 \leq t \leq d$  and  $p \in [1, \infty]$ , with the convention  $1/\infty = 0$ :*

1. *If  $p = 2j$  for a divisor  $j \mid t$ , then  $\kappa_t^{\text{nn}}(B_p^d) = \kappa_t(B_p^d) = 1$ .*
2. *At degree two,  $\kappa_1(B_p^d) = d^{1/2-1/p}$ .*
3. *If  $1 \leq p \leq 2$ , then*

$$\frac{1}{\sqrt{2e}} \left( \frac{d}{t} \right)^{1/p-1/2} \leq \kappa_t^{\text{nn}}(B_p^d) \leq \kappa_t(B_p^d) \leq d^{1/p-1/2}.$$

*Consequently, for fixed  $p < 2$  and  $\log t = o(\log d)$ , both distortions equal  $d^{1/p-1/2+o(1)}$ .*

*Proof.* **1: resonances.** Let  $p = 2j$  and suppose  $j \mid t$ . Then

$$q(x) = \left( \sum_i x_i^{2j} \right)^{t/j} = \|x\|_p^{2t}.$$

The form is SoS: expanding the integer power of  $\sum_i (x_i^j)^2$  expresses it as a sum of squares of degree- $t$  monomials. Its ratio is one, which is the smallest possible value for either polynomial cone.

**2: degree two.** At  $t = 1$ , a nonnegative quadratic form is SoS. Averaging over signed permutations does not increase the ratio and turns its matrix into a scalar multiple of the identity. The ratio of the Euclidean norm to the  $\ell_p$  norm on the  $\ell_p$ -sphere is  $d^{|1/2-1/p|}$ , proving the stated equality.

**3: the range  $1 \leq p \leq 2$ .** Average an arbitrary nonnegative form over signed permutations and evaluate it at

$$x^{(k)} = k^{-1/p} (\underbrace{1, \dots, 1}_k, 0, \dots, 0).$$

Since  $p_{2r}(x^{(k)}) = k^{1-2r/p}$ , the resulting flat-point values have the form

$$V(k) = k^{-\alpha} Q(k), \quad \alpha = \frac{2t}{p}, \quad \deg Q \leq t, \quad Q(0) = 0.$$

This is the same support-size profile as for the cross-polytope, with one controlled change: the geometric normalization replaces the exponent  $2t$  by  $\alpha = 2t/p$ . The inequality  $p \leq 2$  is exactly the condition  $\alpha \geq t$  required by [Lemma 3.5](#). Applying that lemma gives

$$\frac{\max_k V(k)}{\min_k V(k)} \geq 2(2e)^{-t} \left( \frac{d}{t} \right)^{\alpha-t}.$$

After taking the  $2t$ -th root and dropping the harmless factor  $2^{1/(2t)}$ ,

$$\kappa_t^{\text{nn}}(B_p^d) \geq \frac{1}{\sqrt{2e}} \left( \frac{d}{t} \right)^{1/p-1/2}.$$

The SoS form  $(\sum_i x_i^2)^t$  has  $2t$ -th root  $\|x\|_2$ , whose ratio on the  $\ell_p$ -sphere is  $d^{1/p-1/2}$ . This proves the two-sided chain. If  $p < 2$  is fixed and  $\log t = o(\log d)$ , the logarithms of the lower and upper bounds are both  $(1/p - 1/2 + o(1)) \log d$ , proving the final assertion.  $\square$

*Remark A.1.* The three parts of [Theorem 3.10](#) arise from three different mechanisms: exact algebraic representation at the resonance points, uniqueness of the invariant quadratic at  $t = 1$ , and profile interpolation for  $p \leq 2$ . They should not be read as samples from a single smooth formula.

The upper bound for  $1 < p < 2$  does not recover the  $t$ -dependence in the lower bound. A matching construction would require an analogue of the sign design adapted to  $\ell_p$ . For  $p > 2$ , even the correct power of  $d$  away from the resonance points remains open.