Error bounds for monomial convexification in polynomial optimization

Error bounds for monomial convexification in polynomial optimization

Warren Adams, Akshay Gupte, and Yibo Xu

November 23, 2017

Abstract

Convex hulls of monomials have been widely studied in the literature, and monomial convexifications are implemented in global optimization software for relaxing polynomials. However, there has been no study of the error in the global optimum from such approaches. We give bounds on the worst-case error for convexifying a monomial over subsets of $[0,1]^n$. This implies additive error bounds for relaxing a polynomial optimization problem by convexifying each monomial separately. Our main error bounds depend primarily on the degree of the monomial, making them easy to compute. Since monomial convexification studies depend on the bounds on the associated variables, in the second part, we conduct an error analysis for a multilinear monomial over two different types of box constraints. As part of this analysis, we also derive the convex hull of a multilinear monomial over $[-1,1]^n$.

Keywords. Polynomial optimization, Monomial, Multilinear, Convex hull, Error analysis, Means inequality

AMS subject classification. 90C26, 65G99, 52A27

1 Introduction

A polynomial $p \in \mathbb{R}[x]$, where $\mathbb{R}[x] = \mathbb{R}[x_1, \ldots, x_n]$ is the ring of n-variate polynomials, is a linear combination of monomials and is expressed as $p(x) = \sum_\alpha c_\alpha x^\alpha$ where the sum is finite, $x^\alpha := \prod_{j=1}^n x_j^{\alpha_j}$ is a monomial, and every α_j is a nonnegative integer. A polynomial optimization problem is

$$z_S^* = \min \{p(x) \mid x \in S\}$$

for a compact convex set S and $p \in \mathbb{R}[x]$. It is common to assume that the degree of the polynomial is bounded by some constant m and this is denoted by $p \in \mathbb{R}[x]_m$. Polynomials, in general, are nonconvex functions, thereby necessitating the use of global optimization algorithms for optimizing them. Strong and efficiently computable convex relaxations are a major component of these algorithms, making them a subject of ongoing research. One approach for devising good relaxations is based on taking the convex envelope of each polynomial $p(x)$ over S. However, since this computation is NP-hard even in the most basic cases having $m = 2$ and $S = [0,1]^n$ or S being a standard simplex, a main emphasis of the envelope studies has been on finding the envelope either under structural assumptions on S or by considering only a subset of all the monomials appearing in $p(x)$.

Department of Mathematical Sciences, Clemson University
Email address: {wadams,agupte,yibox}@clemson.edu

Also, one is interested in obtaining polyhedral relaxations of the envelope so that lower bounds can be computed cheaply by solving linear programs (LPs) iteratively [LS14; MF05; SDL12; TRX13]. If $p(x)$ is a multilinear polynomial (i.e. $\alpha_j \in \{0,1\}$ for all j) and S is a box, then the envelopes are polyhedral and we know exponential sized extended formulations [Rik97; She97], as well as valid inequalities [CRH17; DPK16] and efficient cutting planes [Bao+15; MSF15] in projected spaces. A second method for obtaining lower bounds on the polynomial optimization problem has been to use the moments approach and Lasserre [Las01] hierarchy of semidefinite relaxations (SDPs) that converges to the global optimum [Las15; Lau09]. All of these techniques can of course also be used for relaxing a optimization problem that has polynomials in both the objective and constraints.

For a general polynomial $p(x) = \sum_\alpha c_\alpha x^\alpha$, given that it is hard to find the envelope explicitly and that computability of the SDP bounds does not scale well, a common relaxation technique, motivated by the classical work of McCormick [McC76], has been to replace each monomial x^α with a continuous variable, say w, and then add inequalities to convexify the graph of x^α over S, which is the set $\{(x,w) \in S \times \mathbb{R} \mid w = x^\alpha\}$. This is referred to as monomial convexification, and it typically yields a weaker relaxation than the envelope of the polynomial due to the fact that the envelope operator does not distribute over sums in general. However, because they may be cheaper and easier to generate than convexification of the entire polynomial, convex hulls of monomials have received significant attention [Bao+15; Bel+09; BD17; LP03] and are also routinely implemented in leading global optimization software [DS16; MF14; TS05]. We still do not know an explicit form for the convex hull of a general monomial, but a number of results are available for bivariate monomials [Loc16] and n-variate multilinear monomials [AKF83; BMN10; Ben04; Cra93; LNL12; MF04; RS01]. Moreover, there also exist challenging applications [BMW10] where the constraints can be formulated as having only monomial terms, thereby making monomial convexifications necessary for obtaining strong relaxations.

To quantify the strength of a relaxation of $p(x)$, one is interested in bounding the error produced with respect to the global optimum z_S^* by optimizing over this relaxation. Error bounds for converging solutions of iterative optimization algorithms have been the subject of study before [Pan97], but since these are not suited for studying relaxation strengths, different error measures have been proposed. Luedtke, Namazifar, and Linderoth [LNL12] studied a relative error measure for the relaxation of a bilinear polynomial $p \in \mathbb{R}[x]_2$ over $S = [0,1]^n$ obtained by convexifying each monomial with its McCormick envelopes. They showed that for every $x \in [0,1]^n$, the ratio of the difference between the McCormick overestimator and underestimator values at x and the difference between the concave and convex envelope values at x can be bounded by a constant that is solely in terms of the chromatic number of the co-occurrence graph of the bilinear polynomial. Recently, Boland et al. [Bol+17] showed that this same ratio cannot be bounded by a constant independent of n. Another, and somewhat natural, way of measuring the error from a relaxation is to bound the absolute gap $z_S^* - \tilde{z}_S$, where $\tilde{z}_S$ is a lower bound on z_S^* due to some convex relaxation of $\{(x,w) \in S \times \mathbb{R} \mid w = p(x)\}$. Such a bound helps determine how close one is to optimality in a global optimization algorithm. Also, there are examples (cf. $\prod_{j=1}^n x_j$ over $[1,r]^n$ in [LNL12, pp. 332]) where the relative error gap of McCormick relaxation goes to ∞, while this can never happen with the absolute gap. The only result that we know of on bounding absolute gaps for general polynomials is due to De Klerk and Laurent [DKL10] who used Bernstein approximation of polynomials for a hierarchy of LP and SDP relaxations. (On the contrary, [DKLS15; DKLS16] bound the absolute error from upper bounds on z_S^*.). We mention that the absolute errors arising from piecewise linear relaxations of bilinear monomials appearing in a specific application were studied by Dey and Gupte [DG15]. Finally, a third error measure is based on comparing the volume of a convex relaxation to the volume of the convex hull. This has been done for McCormick relaxations of a trilinear monomial over a box by Speakman and Lee [SL17].

Our contribution. In this paper, we bound the absolute gap to z_S^* from monomial convexification and thereby add to the small number of explicit error bounds for polynomial optimization. To bound this gap, we analyze the error in relaxing a monomial with its convex hull. This error analysis not only implies a bound on the absolute gap to z_S^* but it also can be used for bounding the error in relaxing any optimization problem with polynomials in both the objective and constraints. Our error measure is the maximum absolute deviation between the actual value and the approximate value of the monomial. Thus for any set X in the (x, w)-space, we denote the error of X with respect to x^α by $\mu(X)$, which is defined as

$$\mu(X) := \max_{(x,w)\in X} |w - x^\alpha|. \tag{1}$$

We will mostly be interested in the error $\mu(\cdot)$ for the convex hull of the graph of x^α and for the convex and concave envelopes of x^α. As mentioned earlier, monomial convexification errors have gone largely unnoticed in the literature, the only results being for the bilinear monomial x_1x_2. The folklore result [cf. AKF83] for x_1x_2 over a rectangle $[l_1, u_1] \times [l_2, u_2]$ states that the convex hull and envelope errors are attained at $(x_1, x_2) = (\frac{u_1+l_1}{2}, \frac{u_2+l_2}{2})$, which is the midpoint of the two diagonals of the box. Linderoth [Lin05] derived error formulae for x_1x_2 over triangles created by the two diagonals of $[l_1, u_1] \times [l_2, u_2]$. Since convex hull and envelope results for a bilinear polynomial are invariant to affine transformations, it is equivalent to consider x_1x_2 over $[0, 1]^2$. Substituting $n = 2$ and $\alpha_1 = \alpha_2 = 1$ in our forthcoming error bounds recover these known errors.

Notation. The vector of ones is $\mathbb{1}$, the i^{th} unit coordinate vector is $\mathbf{e}_i$, and the vector of zeros is $\mathbf{0}$; the dimensions will be apparent from the context in which these vectors are used. The convex hull of a set X is $\operatorname{conv} X$ and the relative interior of $\operatorname{conv} X$ is $\operatorname{rel.int} X$. A nonempty box in $\mathbb{R}^n$ is $[l, u] := [l_1, u_1] \times \cdots \times [l_n, u_n]$. The standard boxes that we focus on in this paper are $[0, 1]^n$, $[-1, 1]^n$, and $[1, r]^n$, for arbitrary scalar $r > 1$. Another compact convex set of interest to us is the standard n-simplex $\Delta_n := \operatorname{conv}\{\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_n\} = \left\{x \geq \mathbf{0} \mid \sum_{j=1}^n x_j \leq 1\right\}$. For convenience, we write $f(x) := x^\alpha$, $f_S^{\min} := \min_{x\in S} f(x)$, $f_S^{\max} := \max_{x\in S} f(x)$. The convex envelope of x^α over S, which is defined as the pointwise supremum of all convex underestimators of x^α over S, is denoted by $\operatorname{vex}_S[f]$. The concave envelope, which is analogously defined, is $\operatorname{cav}_S[f]$. The graph of a function $g(x)$ with domain S is denoted by $\mathcal{G}_S(g) := \{(x, w) \in S \times \mathbb{R} \mid w = g(x)\}$. The graphs of the monomial and its envelopes are $\mathcal{G}_S(f)$, $\mathcal{G}(\operatorname{vex}_S[f])$ and $\mathcal{G}(\operatorname{cav}_S[f])$. Two special types of monomials are the symmetric monomial and the multilinear monomial. The former has $\alpha = \alpha_0\mathbb{1}$ for some $\alpha_0 \in \mathbb{Z}_{\geq 1}$, and the latter, denoted by $m(x) := \prod_{j=1}^n x_j$, is a special case of the former with $\alpha = \mathbb{1}$. For $\beta \in \mathbb{R}_+^n$, we denote $|\beta| := \sum_{j=1}^n \beta_j$.

1.1 Main results

We obtain strong and explicit upper bounds on $\mu(\cdot)$ for different types of monomials. In the polynomial optimization literature, it is common to assume, upto scaling and translation, that the domain S of the problem is a subset of $[0, 1]^n$. When analyzing a single monomial, this assumption is not without loss of generality since the monomial basis of $\mathbb{R}[x]$ is not closed upto translating and scaling the variables. Hence we divide our analysis into two parts. First, we consider a general monomial $f(x) = x^\alpha$ over a compact convex set $S \subseteq [0, 1]^n$, and bound the errors without using explicit analytic forms of the envelopes, which are hard to compute and unknown in closed form for arbitrary S. The concave error is bounded by computing the error from a specific concave overestimator that is precisely the concave envelope of x^α over $[0, 1]^n$. On the convex side, we bound the error for any convex underestimator given as the pointwise supremum of (possibly uncountably

many) linear functions, each of which underestimates x^α over S. Thus our error analysis has a distinctly polyhedral flavor.

In the second part, we limit our attention to a multilinear monomial $m(x) = \prod_{j=1}^n x_j$, but the domain S is either a box with constant ratio or a symmetric box. By a box with constant ratio, we mean any box $[l, u]$ for which there exists a scalar $r > 1$ such that $u_i/l_i = r$ for all i with $l_i > 0$, and $l_i/u_i = r$ for all i with $l_i < 0$. By a symmetric box, we mean any box $[l, u]$ that has $u_i = -l_i$ for all i. Since these boxes are simple scalings of $[1, r]^n$ and $[-1, 1]^n$, respectively, and our error measure $\mu(\cdot)$ scales, we restrict our attention to only $[1, r]^n$ and $[-1, 1]^n$. Contrary to the first part, here we first derive explicit polyhedral characterizations of the envelopes and convex hulls over $[1, r]^n$ and $[-1, 1]^n$ and use them to perform a tight error analysis. The polyhedral representations for the $[1, r]^n$ case follow from the literature, whereas those over $[-1, 1]^n$ are established in this paper.

1.1.1 General monomial

Consider a monomial x^α with $\alpha_j \in \mathbb{Z}_{\geq 1}$ for all j. The degree of this monomial is $d := |\alpha| = \sum_{j=1}^n \alpha_j$. The following constants will be useful throughout the paper:

$$\mathfrak{C}_d^1 := \left(1 - \frac{1}{d}\right) d^{\frac{1}{1-d}}, \qquad \mathfrak{C}_d^2 := \left(1 - \frac{1}{d}\right)^d. \tag{2}$$

Theorem 1.1. *For the monomial $f(x) = x^\alpha$ over $S \subseteq [0, 1]^n$, we have*

$$\mu\left(\mathcal{G}(\mathrm{vex}_S[f])\right) \leq \left(1 - \frac{1}{|\gamma|}\right)^{|\gamma|} \leq \mathfrak{C}_d^2, \quad \mu\left(\mathcal{G}(\mathrm{cav}_S[f])\right) \leq \mu\left(\mathrm{conv}\, \mathcal{G}_S(f)\right) \leq \mathfrak{C}_d^1,$$

where for $\sigma_j := 1 - \max\{x_j \mid x \in S\}$, we define

$$\gamma_j := \begin{cases} \dfrac{1 - (1 - \sigma_j)^{\alpha_j}}{\sigma_j}, & \text{if } \sigma_j > 0, \\ \alpha_j, & \text{if } \sigma_j = 0, \end{cases} \qquad j = 1, \ldots, n.$$

If $\mathbf{0}, \mathbb{1} \in S$, then $\mu\left(\mathrm{conv}\, \mathcal{G}_S(f)\right) = \mu\left(\mathcal{G}(\mathrm{cav}_S[f])\right) = \mathfrak{C}_d^1$.

The monotonicity of $\mathfrak{C}_d^1$ and $\mathfrak{C}_d^2$ with respect to d suggests the intuitive result that convexifying higher degree monomials will likely produce greater errors. As $d \to \infty$, we have $\mathfrak{C}_d^1 \to 1$ and $\mathfrak{C}_d^2 \to 1/e$.

The bounds $\mathfrak{C}_d^1$ and $\mathfrak{C}_d^2$ depend only on the degree of the monomial. They are a consequence of some general error bounds, established in Theorem 3.1 for the concave error and in Theorem 3.2 for the convex error, that depend on how the monomial behaves over the domain S. The arguments used in proving Theorem 1.1 also imply that a family of convex relaxations of $\mathcal{G}_S(f)$ has error equal to $\mathfrak{C}_d^1$. We show this in Proposition 3.6. We also guarantee in Corollary 3.4 that the convex envelope error bound $\mathfrak{C}_d^2$ is tight for $m(x)$ over $S = [0, 1]^n$.

Theorem 1.1 has two immediate implications. First, we obtain the error in convexifying a monomial over $[0, 1]^n$.

Corollary 1.1. $\mu\left(\mathrm{conv}\, \mathcal{G}_{[0,1]^n}(f)\right) = \mathfrak{C}_d^1$.

Second, we obtain an additive error bound on polynomial optimization over subsets of $[0, 1]^n$. For a polynomial $p = \sum_\alpha c_\alpha x^\alpha \in \mathbb{R}[x]$, denote

$$L'(p) = \max\left\{ \max_{\alpha:\, c_\alpha > 0} c_\alpha \mathfrak{C}_d^2, \max_{\alpha:\, c_\alpha < 0} -c_\alpha \mathfrak{C}_d^1 \right\}. \tag{3}$$

Let $z_S^{mono} := \min\{\sum_\alpha c_\alpha w_\alpha \mid (x, w_\alpha) \in \operatorname{conv} \mathcal{G}_S(f) \ \forall \alpha\}$ be the lower bound[1] from monomial convexification on the global optimum $z_S^* = \min_{x \in S} p(x)$.

Corollary 1.2. *For any $p \in \mathbb{R}[x]_m$ and compact convex $S \subseteq [0,1]^n$,*

$$z_S^* - z_S^{mono} \leq L'(p) \binom{n+m}{n}.$$

Proof. We have $z_S^{mono} = \sum_{\alpha:\, c_\alpha > 0} c_\alpha \operatorname{vex}_S[f](x) + \sum_{\alpha:\, c_\alpha < 0} c_\alpha \operatorname{cav}_S[f](x)$. Therefore,

$$z_S^* - z_S^{mono} = \sum_{\alpha:\, c_\alpha > 0} c_\alpha (x^\alpha - \operatorname{vex}_S[f](x)) + \sum_{\alpha:\, c_\alpha < 0} (-c_\alpha)(\operatorname{cav}_S[f](x) - x^\alpha).$$

Applying Theorem 1.1 and the construction of $L'(p)$ gives us $z_S^* - z_S^{mono} \leq L'(p) \sum_\alpha 1$. Since $p \in \mathbb{R}[x]_m$, there are at most $\binom{n+m}{n}$ monomials in $p(x)$, leading to the claimed error bound. □

Computing $L'(p)$ may get tedious if $p(x)$ has a large number of monomials. A cheaper bound is possible by considering only the largest coefficient in $p(x)$.

Corollary 1.3. *For any $p \in \mathbb{R}[x]_m$ and compact convex $S \subseteq [0,1]^n$,*

$$z_S^* - z_S^{mono} \leq \max_\alpha |c_\alpha| \left(1 - \frac{1}{m}\right) m^{\frac{1}{1-m}} \binom{n+m}{n}.$$

Proof. Follows from Corollary 1.2 after using $d \leq m$ and $\mathcal{C}_d^1$ being monotone in d. □

The bounds from Theorem 1.1, although applicable to arbitrary $S \subseteq [0,1]^n$, can be weak if $\mathbf{0} \in S$ and $\mathbb{1} \notin S$. To emphasize this, we consider a monomial over the standard simplex Δ_n and obtain error bounds that depend on not just the degree of the monomial but also the exponent of each variable. These bounds are stronger than the bounds $\mathcal{C}_d^1$ and $\mathcal{C}_d^2$.

Theorem 1.2.

$$\mu\left(\mathcal{G}(\operatorname{cav}_{\Delta_n}[f])\right) \leq \mu\left(\operatorname{conv} \mathcal{G}_{\Delta_n}(f)\right) \leq \frac{(\alpha^\alpha)^{1/d}}{d} - \frac{\alpha^\alpha}{d^d}, \quad \mu\left(\mathcal{G}(\operatorname{vex}_{\Delta_n}[f])\right) = \frac{\alpha^\alpha}{d^d}.$$

All of the above bounds are tight for a symmetric monomial.

1.1.2 Multilinear monomial

Consider the multilinear monomial $m(x) = \prod_{j=1}^n x_j$.

Theorem 1.3. *Denote*

$$\mathcal{D}_{r,n} := \max_{i=1,\ldots,n-1} \left\{ \left(1 + \frac{i}{n}(r-1)\right)^n - r^i \right\}, \quad \mathcal{E}_{r,n} := 1 + \frac{r^n - 1}{(r-1)} \left[\frac{n-1}{n} \left(\frac{r^n - 1}{n(r-1)} \right)^{\frac{1}{n-1}} - 1 \right].$$

For $m(x)$ over $[1,r]^n$,

$$\mu\left(\mathcal{G}(\operatorname{cav}_{[1,r]^n}[m])\right) = \mathcal{E}_{r,n}, \quad \mu\left(\mathcal{G}(\operatorname{vex}_{[1,r]^n}[m])\right) = \mathcal{D}_{r,n}, \quad \mu\left(\operatorname{conv} \mathcal{G}_{[1,r]^n}(m)\right) = \max\{\mathcal{D}_{r,n}, \mathcal{E}_{r,n}\}.$$

All bounds are attained only on $\operatorname{rel.int}\{\mathbb{1}, r\mathbb{1}\}$.

[1] To avoid tediousness and with a slight abuse of notation, for each monomial we write $(x, w_\alpha) \in \operatorname{conv} \mathcal{G}_S(f)$ with the understanding that those x_j that appear in the monomial are included.

We conjecture that $\mathcal{D}_{r,n} \leq \mathcal{E}_{r,n}$ for all r, n and provide a strong empirical evidence in support of this claim. We prove this conjecture to be asymptotically true by showing that $\lim_{n\to\infty} \mathcal{D}_{r,n}/\mathcal{E}_{r,n} \leq 1/e$.

For $S = [-1,1]^n$, we characterize the convex hull in Theorem 4.1 and show that it has the following errors.

Theorem 1.4. *For $m(x)$ over $[-1,1]^n$,*

$$\mu\left(\mathcal{G}(\text{cav}_{[-1,1]^n}[m])\right) = \mu\left(\mathcal{G}(\text{vex}_{[-1,1]^n}[m])\right) = \mu\left(\text{conv}\,\mathcal{G}_{[-1,1]^n}(m)\right) = 1 + \left(\frac{n-2}{n}\right)^n.$$

This maximum error is attained at all the 2^n reflections of the point $(\frac{n-2}{n}\mathbb{1}, -1)$.

The exact description of the reflected points will be provided when we prove this theorem. Taking $n \to \infty$, this error approaches $1 + 1/e^2$ from below.

1.1.3 Outline

Our analysis begins with some preliminaries on the error measure. We observe that the error scales with the box and present a lower bound on the error, which we remark is also the proposed upper bound for the two cases $S = [0,1]^n$ and $S = [1,r]^n$. We also formally note the intuition that the convex hull error can be computed as the maximum of the two envelope errors, due to which our error analysis in the remainder of the paper involves analyzing the concave envelope and the convex envelope separately. §3.1 and §3.2 analyze these errors for a general monomial x^α over $S \subseteq [0,1]^n$. The main error bounds presented in §1.1.1 are proved in §3.3 and we compare them to those from literature in §3.4. The multilinear monomial over $[1,r]^n$ and $[-1,1]^n$ is analyzed in §4.1 and §4.2.

2 Preliminaries on $\mu(\cdot)$

The error defined in (1) is obviously monotone with respect to set inclusion: $\mu(X_1) \leq \mu(X_2)$ for any $X_1 \subseteq X_2$. This enables us to upper bound the convex hull error by using $\mu\left(\text{conv}\,\mathcal{G}_S(f)\right) \leq \mu(X)$ for any convex relaxation X of $\mathcal{G}_S(f)$, and also implies that the convex hull error over a smaller variable domain is upper bounded by the convex hull error over a larger domain. Another property we observe is that computing the convex hull error is equivalent to computing the error due to the convex envelope $\text{vex}_S[f]$ and that due to the concave envelope $\text{cav}_S[f]$. This intuitively seems correct given the well-known fact that $\text{conv}\,\mathcal{G}_S(f) = \{(x,w) \in S \times \mathbb{R} \mid \text{vex}_S[f](x) \leq w \leq \text{cav}_S[f](x)\}$, and the fact that the monomial convexification and envelope errors are

$$\mu\left(\text{conv}\,\mathcal{G}_S(f)\right) = \max_{(x,w)\in \text{conv}\,\mathcal{G}_S(f)} |w - x^\alpha|, \quad \mu\left(\mathcal{G}(\text{vex}_S[f])\right) = \max_{x\in S} x^\alpha - \text{vex}_S[f](x),$$
$$\mu\left(\mathcal{G}(\text{cav}_S[f])\right) = \max_{x\in S} \text{cav}_S[f](x) - x^\alpha.$$

Observation 2.1. *Let $X := \{(x,w) \in S \times \mathbb{R} \mid f_1(x) \leq w \leq f_2(x)\}$, where f_1 and f_2 are, respectively, convex and concave continuous functions with $f_1(x) \leq x^\alpha \leq f_2(x)$ for all $x \in S$. Then*

$$\mu\left(\text{conv}\,\mathcal{G}_S(f)\right) \leq \mu(X) = \max\left\{\mu\left(\mathcal{G}_S(f_1)\right), \mu\left(\mathcal{G}_S(f_2)\right)\right\},$$

and equality holds if $f_1 = \text{vex}_S[f]$ and $f_2 = \text{cav}_S[f]$.

The proof is straightforward and is left to the reader. Based on this observation, our error analysis in the rest of the paper involves analyzing the concave envelope and the convex envelope separately.

A third and final property we note is that the error scales with the box. For $c \in \mathbb{R}^n_{\neq 0}$,

$$[cl, cu] := \{x \in \mathbb{R}^n \mid c_j l_j \leq x_j \leq c_j u_j \ \forall j \text{ s.t. } c_j > 0,\ c_j l_j \geq x_j \geq c_j u_j \ \forall j \text{ s.t. } c_j < 0\}$$

is the coordinate-wise scaled version of $[l, u]$. The bijective linear map $\mathcal{D}(x) := (c_1 x_1, \ldots, c_n x_n)$ gives us the relation $[cl, cu] = \mathcal{D}([l, u])$. Denote $c^\alpha := \prod_{j=1}^n c_j^{\alpha_j}$.

Observation 2.2. *For any $c \in \mathbb{R}^n_{\neq 0}$, we have $\mu\left(\operatorname{conv} \mathcal{G}_{[cl,cu]}(f)\right) = |c^\alpha|\, \mu\left(\operatorname{conv} \mathcal{G}_{[l,u]}(f)\right)$, with (x, w) being optimal to $\mu\left(\operatorname{conv} \mathcal{G}_{[l,u]}(f)\right)$ if and only if $(\mathcal{D}(x), c^\alpha w)$ is optimal to $\mu\left(\operatorname{conv} \mathcal{G}_{[cl,cu]}(f)\right)$.*

Observation 2.2 allows us to focus on boxes with specific bounds l_j and u_j, and to then extend to slightly more general boxes via scalings. In particular, error results for

- $[0, 1]^n$ scale to any box having a vertex at $\mathbf{0}$,
- $[1, r]^n$ scale to any box for which the ratio between lower and upper bounds is the same positive scalar in each coordinate, and
- $[-1, 1]^n$ scale to any box that is symmetric with respect to $\mathbf{0}$.

Finally, we observe a lower bound on $\mu\left(\mathcal{G}(\operatorname{cav}_S[f])\right)$, and hence on $\mu\left(\operatorname{conv} \mathcal{G}_S(f)\right)$, when S contains two points on the ray $\{t\mathbb{1} \mid t \geq 0\}$, which happens for example when $S = [t_1, t_2]^n$ for some $t_1 < t_2$ with $t_2 > 0$.

Lemma 2.1. *Suppose $S \cap \{t\mathbb{1} \mid t \geq 0\} \neq \emptyset$ and let $t_1, t_2 \geq 0$ be the minimum and maximum values such that $t_1\mathbb{1}, t_2\mathbb{1} \in S$. Let $\tilde{f}$ be a concave overestimator of x^α on S and let X be a convex relaxation of $\mathcal{G}_S(f)$. Then, $\mu\left(\mathcal{G}_S(\tilde{f})\right) \geq \phi(\xi')$ and $\mu(X) \geq \phi(\xi')$, where $\phi\colon \xi \in [0, 1] \mapsto t_1^d + (t_2^d - t_1^d)\xi - (t_1 + (t_2 - t_1)\xi)^d$ and*

$$\xi' = \left(\frac{t_2^d - t_1^d}{d}\right)^{\frac{1}{d-1}} (t_2 - t_1)^{\frac{d}{1-d}} - \frac{t_1}{t_2 - t_1}.$$

Proof. The assumption $t_1\mathbb{1}, t_2\mathbb{1} \in S$ implies $(t_1\mathbb{1}, t_1^d), (t_2\mathbb{1}, t_2^d) \in \mathcal{G}_S(f)$. Convexity of X and $\mathcal{G}_S(f) \subset X$ lead to $(((1-\xi)t_1 + \xi t_2)\mathbb{1}, (1-\xi)t_1^d + \xi t_2^d) \in X$ for all $\xi \in [0, 1]$. Therefore

$$\begin{aligned}\mu(X) \ \geq\ \max_{0\leq\xi\leq 1} \left|(1-\xi)t_1^d + \xi t_2^d - ((1-\xi)t_1 + \xi t_2)^d\right| \ &=\ \max_{0\leq\xi\leq 1} (1-\xi)t_1^d + \xi t_2^d - ((1-\xi)t_1 + \xi t_2)^d \\ &=\ \max_{0\leq\xi\leq 1} \phi(\xi),\end{aligned}$$

where the equality is due to $t_1, t_2 \geq 0$ and convexity of the function $t \mapsto t^d$ on $\mathbb{R}_+$. Since $\phi(0) = \phi(1) = 0$, by Rolle's theorem, there exists a stationary point in $[0, 1]$ and this point is exactly ξ' stated above. Since ϕ is concave, ξ' must be a maxima. For a concave overestimator $\tilde{f}$, we have

$$\begin{aligned}\mu\left(\mathcal{G}_S(\tilde{f})\right) = \max_{x\in S} \tilde{f}(x) - x^\alpha \ &\geq\ \max_{0\leq\xi\leq 1} \tilde{f}((1-\xi)t_1 + \xi t_2) - ((1-\xi)t_1 + \xi t_2)^d \\ &\geq\ \max_{0\leq\xi\leq 1} (1-\xi)\tilde{f}(t_1) + \xi\tilde{f}(t_2) - ((1-\xi)t_1 + \xi t_2)^d \\ &\geq\ \max_{0\leq\xi\leq 1} (1-\xi)t_1^d + \xi t_2^d - ((1-\xi)t_1 + \xi t_2)^d \\ &=\ \phi(\xi').\end{aligned}$$

□

Remark 1. For lower bounding $\mu(X)$, the above proof really only requires $(t_1\mathbb{1}, t_1^d), (t_2\mathbb{1}, t_2^d) \in X$. The stronger assumption $t_1\mathbb{1}, t_2\mathbb{1} \in S$ is made for convenience.

Remark 2. The above method of lower bounding the error can also be utilized by considering arbitrary $l, u \in S$ with $\mathbf{0} \leq l \leq u$. This generalization is made possible by the observation that the function $\xi \mapsto \prod_{j=1}^n (l_j + (u_j - l_j)\xi)^{\alpha_j}$ is convex over $\mathbb{R}_+$. Since the derivation gets extremely tedious and does not yield new insight, we omit the general case here.

Substituting $t_1 = 0, t_2 = 1$ in Lemma 2.1 yields the critical point to be $\xi' = (1/d)^{1/(d-1)}$ so that $\phi(\xi') = \xi' - \xi'^d = \xi'(1 - \xi'^{d-1}) = (1/d)^{\frac{1}{d-1}}(1 - 1/d) = \mathcal{C}_d^1$, where the constant $\mathcal{C}_d^1$ was introduced in equation (2). Thus the significance of the lower bound from this lemma is that we prove in Theorem 1.1 that it is indeed equal to the maximum error of the convex hull when $\{\mathbf{0}, \mathbb{1}\} \subset S \subseteq [0,1]^n$. For a multilinear monomial over $S = [1, r]^n$ for some $r > 1$, or equivalently $S = [\frac{1}{r}, 1]^n$ using the scaling from Observation 2.2, the constant $\mathcal{E}_{r,n}$ defined in the statement of Theorem 1.3 is exactly the lower bound obtained from Lemma 2.1 by substituting $t_1 = 1, t_2 = r$ and we prove that this is the maximum concave envelope error and conjecture, with strong empirical evidence in support, that it is also the maximum convex hull error.

3 Monomial over $[0,1]^n$

This section considers a general multivariate monomial x^α, for some $\alpha \in \mathbb{Z}^n_{\geq 1}$, over a nonempty compact convex set $S \subseteq [0,1]^n$. It follows that $\mathcal{G}_S(f) \subseteq [0,1]^{n+1}$. Our main error bounds on $\mu\left(\operatorname{conv} \mathcal{G}_S(f)\right)$ depend only on the degree $d := \sum_{j=1}^n \alpha_j$ of the monomial and therefore are independent of how the monomial behaves on its domain S. However, en route to deriving these formulas, we establish tighter bounds that depend on the minimum and maximum value of x^α over S and thus are expensive to compute in general. The error formulas for the multilinear case will follow after substituting $\alpha = \mathbb{1}$. Motivated by Observation 2.1, we bound the convex hull error by bounding the envelope errors separately.

Before we begin, we recall that the envelopes of $m(x)$ were shown by Crama [Cra93] to be

$$\operatorname{vex}_{[0,1]^n}[m](x) = \max\left\{0, 1 + \sum_{j=1}^n (x_j - 1)\right\}, \quad \operatorname{cav}_{[0,1]^n}[m](x) = \min_{j=1,\dots,n} x_j. \tag{4a}$$

Remark 3. The envelopes of $m(x)$ over a box $[l, u]$ having one of its vertices at the origin, i.e., $l_j u_j = 0$ for all j, can be obtained by scaling the variables in (4a) as $x_j \leftarrow u_j x_j$ for $j \in J_1 := \{j \mid l_j = 0\}$, $x_j \leftarrow l_j x_j$ for $j \in J_2 := \{j \mid u_j = 0\}$, and $w_j \leftarrow w \prod_{j \in J_1} u_j \prod_{j \in J_2} l_j$.

The concave envelope in (4a) is also the concave envelope of x^α over $[0,1]^n$ for every $\alpha \geq \mathbb{1}$, i.e.

$$\operatorname{cav}_{[0,1]^n}[f](x) = \min_{j=1,\dots,n} x_j. \tag{4b}$$

This is because $f(x) = m(x)$ for $x \in \{0,1\}^n$ and a monomial x^α with $\alpha \geq \mathbb{1}$ is known to be concave-extendable from the vertices of $[0,1]^n$ (meaning that $\operatorname{cav}_{[0,1]^n}[f]$ can be obtained by looking at the values of $f(x)$ solely at $\{0,1\}^n$); see [TS02]. One can also establish this fact independently without using concave-extendability of $f(x)$.

For notational convenience throughout this section, we denote

$$E_0 := \{x \in [0,1]^n \mid x_i = 0 \text{ for some } i\}, \quad E_j := \operatorname{conv}\{\mathbb{1}, \mathbb{1} - \mathbf{e}_j\}, \ \ j = 1, \dots, n.$$

That is, E_0 is the union of all the coordinate plane facets of $[0,1]^n$ and E_j is the j^{th} edge of $[0,1]^n$ that is incident to the vertex $\mathbb{1}$.

3.1 Concave overestimator error

Throughout, we consider the piecewise linear concave function $f^{\text{conc}}(x) := \min_{j=1\dots n} x_j$, which we noted in (4b) to be the concave envelope of $f(x)$ over $[0,1]^n$. First, we treat the general case where S is any subset of $[0,1]^n$, and later we consider the case of S being a standard simplex.

3.1.1 General case

For arbitrary $S \subseteq [0,1]^n$, we have $\text{cav}_S[f](\cdot) \leq f^{\text{conc}}(\cdot)$ due to $\alpha \geq \mathbb{1}$ and $x \in [0,1]^n$ implying $0 \leq x^\alpha \leq x_j^{\alpha_j} \leq x_j$ for all j. We observe that this overestimator is exact only on E_0 or on edges E_i's along which the monomial is linear.

Proposition 3.1. *$f^{\text{conc}}(x) = x^\alpha$ if and only if $x = \mathbb{1}$ or $x \in E_0$ or $x \in E_i$ for some i with $\alpha_i = 1$.*

Proof. For $S \subseteq [0,1]^n$ and $\alpha \geq \mathbb{1}$, $f^{\text{conc}}(x) \geq x^\alpha$ follows from the facts and $x_i^{\alpha_i} x_j^{\alpha_j} \leq x_i^{\alpha_i}\ \forall i \neq j$. The equalities $f^{\text{conc}}(\mathbb{1}) = \mathbb{1}^\alpha = 1$ and $f^{\text{conc}}(x) = x^\alpha = 0$, for all $x \in E_0$, are obvious. For any $x \in \text{rel.int}\, E_i$, $x_i \in (0,1)$ and $x_j = 1\ \forall j \neq i$ give us $f^{\text{conc}}(x) = x_i$ and $f(x) = x_i^{\alpha_i}$. Thus it is obvious that for $x \in \text{rel.int}\, E_i$, $f^{\text{conc}}(x) = f(x)$ if and only if $\alpha_i = 1$. Now let x be any point in S that does not belong to a coordinate plane nor to any edge E_i. Then there exist distinct indices i, j with $x_i, x_j \in (0,1)$ and $x_i \leq x_j \leq x_k\ \forall k \neq i, j$. Therefore $0 < x^\alpha \leq x_i^{\alpha_i} x_j^{\alpha_j} < x_i = f^{\text{conc}}(x)$. □

Since $\text{cav}_S[f](\cdot) \leq f^{\text{conc}}(\cdot)$, the error due to f^{conc}, which is the maximum value of the difference $f^{\text{conc}}(x) - x^\alpha$ over S, provides an upper bound on the error from $\text{cav}_S[f]$. Proposition 3.1 tells us that this maximum difference occurs either in the interior of $[0,1]^n$ or in the relative interior of some face of $[0,1]^n$ passing through $\mathbb{1}$. In the following result, we give a tight upper bound on $f^{\text{conc}}(x) - x^\alpha$ that is attained at a specific point on the diagonal between $\mathbf{0}$ and $\mathbb{1}$. This is our main error bound for $\text{cav}_S[f]$.

Theorem 3.1. *$\mu\left(\mathcal{G}_S(f^{\text{conc}})\right) \leq \xi'^{1/d} - \xi'$, where $\xi' = \min\{\max\{f_S^{\min}, d^{\frac{d}{1-d}}\}, f_S^{\max}\}$. This bound can be attained only at the point $\xi'^{1/d}\mathbb{1} \in \text{rel.int}\,\{\mathbf{0}, \mathbb{1}\}$ and hence is tight if and only if $\xi'^{1/d}\mathbb{1} \in S$.*

Proof. Since $x \in [0,1]^n$ and $\alpha \geq \mathbb{1}$, we have $(\min_i x_i)^d \leq x^\alpha \leq \min_i x_i$. This implies $f^{\text{conc}}(x) \leq (x^\alpha)^{\frac{1}{d}}$ for $x \in [0,1]^n$, which leads to

$$\max_{x \in S} f^{\text{conc}}(x) - x^\alpha \ \leq\ \max_{x \in S} (x^\alpha)^{\frac{1}{d}} - x^\alpha. \tag{5}$$

Since $f(x)$ is a continuous function with minimum and maximum values $f_S^{\min}$ and $f_S^{\max}$ on the closed convex set S, the intermediate value theorem implies that

$$\max_{x \in S} (x^\alpha)^{\frac{1}{d}} - x^\alpha = \max\{\xi^{\frac{1}{d}} - \xi \mid f_S^{\min} \leq \xi \leq f_S^{\max}\}.$$

We have $0 \leq f_S^{\min} \leq f_S^{\max} \leq 1$ due to $S \subseteq [0,1]^n$. Elementary calculus tells us that the function $\xi^{1/d} - \xi$ is concave on $[0,1]$ with a unique stationary point at $\xi_0 = d^{\frac{d}{1-d}}$ and is increasing on $[0, \xi_0)$ and decreasing on $(\xi_0, 1]$. Hence the maximum value of this function on $[f_S^{\min}, f_S^{\max}]$ is $\xi'^{1/d} - \xi'$, where $\xi' = \min\{\max\{f_S^{\min}, d^{\frac{d}{1-d}}\}, f_S^{\max}\}$. Combining this with (5) gives us the desired upper bound.

Now we claim that this bound can be tight only on $\text{rel.int}\,\{\mathbf{0}, \mathbb{1}\}$. Suppose this is not true and there exists a $y \in S \setminus \text{rel.int}\,\{\mathbf{0}, \mathbb{1}\}$ such that $f^{\text{conc}}(y) - y^\alpha = \xi'^{1/d} - \xi'$. The fact that $\xi' > 0$ and $d \geq 2$ makes it obvious that $y \neq \mathbf{0}, \mathbb{1}$. Thus $y \notin \text{conv}\{\mathbf{0}, \mathbb{1}\}$. Since $f^{\text{conc}}(y) \leq (y^\alpha)^{1/d}$ and $\xi'^{1/d} - \xi'$ is the maximum value of the right hand side in (5), we have

$$\xi'^{1/d} - \xi' = f^{\text{conc}}(y) - y^\alpha \leq (y^\alpha)^{1/d} - y^\alpha \leq \xi'^{1/d} - \xi',$$

implying that equality holds throughout. Hence $f^{\text{conc}}(y) = (y^\alpha)^{1/d}$. However this is a contradiction to $y \in S \setminus \text{conv}\{\mathbf{0}, \mathbb{1}\}$ because observe that for any $x \geq \mathbf{0}$, $f^{\text{conc}}(x) = (x^\alpha)^{1/d}$ if and only if $x_1 = x_2 = \cdots = x_n$, which is equivalent to $x \in \text{conv}\{\mathbf{0}, \mathbb{1}\}$. Therefore $S \cap \text{rel. int}\{\mathbf{0}, \mathbb{1}\} \neq \emptyset$ is necessary for the proposed upper bound to be tight.

Suppose that $S \cap \text{conv}\{\mathbf{0}, \mathbb{1}\} = \text{conv}\{\xi_1\mathbb{1}, \xi_2\mathbb{1}\}$ for some $0 \leq \xi_1 \leq \xi_2 \leq 1$. On $\text{rel. int}\{\mathbf{0}, \mathbb{1}\}$, the function $f^{\text{conc}}(x) - x^\alpha$ transforms to the univariate concave function $\xi - \xi^d$ for $\xi \in (0, 1)$, which has a unique stationary point at $\tilde{\xi} = d^{1/(1-d)}$, giving us $\tilde{\xi} - \tilde{\xi}^d = d^{1/(1-d)} - d^{d/(1-d)} = \xi'^{1/d} - \xi'$, if $\xi' = d^{\frac{d}{1-d}}$. The function $\xi - \xi^d$ is increasing on $(0, \tilde{\xi})$ and decreasing on $(\tilde{\xi}, 1)$. By construction of $f_S^{\min}$ and $f_S^{\max}$, it follows that $0 \leq f_S^{\min} \leq \xi_1^d \leq \xi_2^d \leq f_S^{\max} \leq 1$. Therefore $f^{\text{conc}}(x) - x^\alpha = \xi'^{1/d} - \xi'$ for some $x \in S$ if and only if $x = \xi'^{1/d}\mathbb{1}$ and $\xi_1 \leq \xi'^{1/d} \leq \xi_2$. □

The upper bound presented in Theorem 3.1 depends on the minimum and maximum values of the monomial over S, which can be hard to compute for arbitrary S, and not just on the degree of the monomial. However, an immediate consequence is that the constant $\mathcal{C}_d^1$, defined as $\mathcal{C}_d^1 := (d-1)\, d^{\frac{d}{1-d}}$ in equation (2), is a degree-dependent bound on the error from $f^{\text{conc}}(x)$.

Corollary 3.1. *$\mu\left(\mathcal{G}_S(f^{\text{conc}})\right) \leq \mathcal{C}_d^1$, and this bound is tight if $d^{1/(1-d)}\mathbb{1} \in S$ and only if $f_S^{\min} \leq d^{d/(1-d)} \leq f_S^{\max}$.*

Proof. The function $\xi^{1/d} - \xi$ attains its maxima over $[0, 1]$ uniquely at $\xi_0 = d^{d/(1-d)}$. The definition of ξ' then gives us

$$\xi'^{1/d} - \xi' \leq \left(d^{\frac{d}{1-d}}\right)^{\frac{1}{d}} - d^{\frac{d}{1-d}} = d^{\frac{1}{1-d}} - d^{\frac{d}{1-d}} = (d-1)d^{\frac{d}{1-d}} = \mathcal{C}_d^1,$$

and subsequently, Theorem 3.1 leads to $\mathcal{C}_d^1$ being an upper bound on $f^{\text{conc}}(x) - x^\alpha$. The uniqueness of the maxima of $\xi^{1/d} - \xi$ also implies that for $\mathcal{C}_d^1$ to be a tight bound, we must have $\xi' = d^{d/(1-d)}$, which is equivalent to $f_S^{\min} \leq d^{d/(1-d)} \leq f_S^{\max}$. □

Notice that the necessity of $f_S^{\min} \leq d^{d/(1-d)} \leq f_S^{\max}$ in the above corollary is not immediate from the statement of Theorem 3.1. This can be explained as follows. Denote $S \cap \text{conv}\{\mathbf{0}, \mathbb{1}\} = \text{conv}\{\xi_1\mathbb{1}, \xi_2\mathbb{1}\}$ for some $0 \leq \xi_1 \leq \xi_2 \leq 1$. Since we showed that $\mathcal{C}_d^1$ is an upper bound on $\xi'^{1/d} - \xi'$, Theorem 3.1 implies that if $\mathcal{C}_d^1$ is a tight bound then $\xi_1 \leq \xi'^{1/d} \leq \xi_2$. By construction, $\xi' \in \{f_S^{\min}, f_S^{\max}, d^{d/(1-d)}\}$ and $(f_S^{\min})^{1/d} \leq \xi_1 \leq \xi_2 \leq (f_S^{\max})^{1/d}$. So, by Theorem 3.1, it is possible to have $d^{1/(1-d)} < (f_S^{\min})^{1/d}$ or $d^{1/(1-d)} > (f_S^{\max})^{1/d}$, if $\mathcal{C}_d^1$ is tight. However, Corollary 3.1 rules out this possibility. Furthermore, the condition $f_S^{\min} \leq d^{d/(1-d)} \leq f_S^{\max}$ is not sufficient to guarantee tightness of $\mathcal{C}_d^1$. The reason being that this condition does not enforce non-emptiness of $S \cap \text{rel. int}\{\mathbf{0}, \mathbb{1}\}$, which we know to be necessary from Theorem 3.1.

If the minimum and maximum values of x^α over S are low-enough and high-enough, respectively, as per Corollary 3.1, then we have a precise characterization of when $\mathcal{C}_d^1$ is a tight bound on $f^{\text{conc}}(x) - x^\alpha$.

Corollary 3.2. *For any $S \subseteq [0, 1]^n$ with $f_S^{\min} \leq d^{d/(1-d)} \leq f_S^{\max}$, the upper bound $\mathcal{C}_d^1$ on $f^{\text{conc}}(x) - x^\alpha$ is tight if and only if $d^{1/(1-d)}\mathbb{1} \in S$. In particular, $\mu\left(\mathcal{G}(\text{cav}_{[0,1]^n}[f])\right) = \mathcal{C}_d^1$.*

Proof. The assumptions of $f_S^{\min}$ and $f_S^{\max}$ imply $\xi' = d^{d/(1-d)}$ in Theorem 3.1, thereby leading to the first claim. Since $f_{[0,1]^n}^{\min} = 0$, $f_{[0,1]^n}^{\max} = 1$, and $d^{1/(1-d)}\mathbb{1} \in \text{rel. int}\{\mathbf{0}, \mathbb{1}\}$, the second claim follows from the first part and $f^{\text{conc}} = \text{cav}_{[0,1]^n}[f]$ from (4b). □

For the simplex $\Delta_n^{\mathbb{1}} := \operatorname{conv}\{\mathbb{1}, \mathbb{1}-\mathbf{e}_1, \ldots, \mathbb{1}-\mathbf{e}_n\}$, clearly, $f_S^{\min} = 0, f_S^{\max} = 1$ for any $S \supseteq \Delta_n^{\mathbb{1}}$. This simplex can be described as $\Delta_n^{\mathbb{1}} = \{x \mid \sum_{j=1}^n x_j \geq n-1,\ x \leq \mathbb{1}\}$. When $d = n$, i.e., multilinear monomial, it is easy to verify graphically that $n^{1/(1-n)} < 1 - 1/n$ so that the point $n^{1/(1-n)}\mathbb{1}$ does not belong to $\Delta_n^{\mathbb{1}}$. However, the function $d^{1/(1-d)}$ being monotone in d, for large enough values of d, we have $d^{1/(1-d)} \geq 1 - 1/n$, as can be verified numerically, and consequently, $d^{1/(1-d)}\mathbb{1} \in \Delta_n^{\mathbb{1}}$. Hence, the bound $\mathcal{C}_d^1$ from Corollary 3.2 is tight for arbitrary $S \supseteq \Delta_n^{\mathbb{1}}$ when the monomial degree is large.

3.1.2 Standard simplex

For monomials considered over the standard n-simplex Δ_n, we obtain a bound in Proposition 3.2 that is tight only for symmetric monomials. The proof of this result uses the following lemma which will be useful also in proving Theorem 1.1 later in §3.3.

Lemma 3.1. *$d^{(d-1)^2} > d^{d(d-2)} \geq (d-1)^{(d-1)^2}$ for all $d \geq 2$, and $d^{d(d-2)} = (d-1)^{(d-1)^2}$ if and only if $d = 2$.*

Proof. Obviously $d^{(d-1)^2} = d^{d^2-2d+1} > d^{d(d-2)}$. Since $d/(d-1) = 1 + 1/(d-1)$, binomial expansion gives us

$$\left(\frac{d}{d-1}\right)^{(d-1)^2} \geq 1 + \frac{(d-1)^2}{d-1} = d, \qquad d \geq 2.$$

This is equivalent to $d^{d(d-2)} \geq (d-1)^{(d-1)^2}$. Clearly, equality holds for $d = 2$. For $d \geq 3$, binomial expansion gives us

$$\left(\frac{d}{d-1}\right)^{(d-1)^2} \geq 1 + \frac{(d-1)^2}{d-1} + \binom{(d-1)^2}{2}\frac{1}{(d-1)^2} = d + \frac{d(d-2)}{2} > d, \qquad d \geq 3,$$

thereby leading to $d^{d(d-2)} > (d-1)^{(d-1)^2}$. □

Proposition 3.2.

$$\mu\left(\mathcal{G}_{\Delta_n}(f^{\text{conc}})\right) \leq \frac{(\alpha^\alpha)^{1/d}}{d} - \frac{\alpha^\alpha}{d^d},$$

and this bound is tight if and only if $\alpha_1 = \cdots = \alpha_n$.

Proof. $f_{\Delta_n}^{\min} = 0$ because $\mathbf{0} \in \Delta_n$. The maximum value of x^α over Δ_n is obviously attained in the relative interior of the face defined by the plane $\sum_{j=1}^n x_j = 1$. Solving the KKT system for $f_{\Delta_n}^{\max} = \max_x \{x^\alpha \mid \sum_{j=1}^n x_j = 1\}$ gives us $f_{\Delta_n}^{\max} = \frac{\alpha^\alpha}{d^d}$. For fixed integers $2 \leq n \leq d$, it is easy to argue that

$$\max_\alpha \left\{ \alpha^\alpha \mid \alpha \in \mathbb{Z}_{\geq 1}^n, \sum_{j=1}^n \alpha_j = d \right\} = (d-n+1)^{d-n+1},$$

using the convexity of $t \mapsto t\log t$ and the integrality of the polytope $\{\alpha \in \mathbb{R}_{\geq 1}^n \mid \sum_j \alpha_j = d\}$. Therefore for fixed d, the maximum value of α^α is achieved with $n = 2$ and is equal to $(d-1)^{d-1}$. Thus, $f_{\Delta_n}^{\max} = \alpha^\alpha/d^d \leq (d-1)^{d-1}/d^d$. By Lemma 3.1, we have $d^{d(d-2)/(d-1)} \geq (d-1)^{d-1}$ and so

$$f_{\Delta_n}^{\max} \leq \frac{d^{d(d-2)/(d-1)}}{d^d} = d^{d/(1-d)}.$$

This implies that $\xi' = f_{\Delta_n}^{\max}$ in Theorem 3.1, thereby giving us the proposed upper bound on $f^{\text{conc}}(x) - x^\alpha$. Theorem 3.1 also tells us that this bound is tight if and only if $\frac{(\alpha^\alpha)^{1/d}}{d}\mathbb{1} \in \Delta_n$, which is equivalent to showing $\alpha^\alpha \leq (d/n)^d$. Observe the following.

Claim 3.1. *$\alpha^\alpha \geq (d/n)^d$ for $\alpha \geq \mathbb{1}$, with equality holding if and only if $\alpha_1 = \cdots = \alpha_n$.*

Proof of Claim. This inequality is obtained by applying Jensen's inequality to the convex function $t \in (0,\infty) \mapsto t \log t$ with the n points being $t_i = \alpha_i \ \forall i$ and the convex combination weights being all equal to $1/n$. The equality condition is due to $t \log t$ being strictly convex. ◇

Therefore our bound is tight if and only if the monomial is symmetric. □

3.2 Convex underestimator error

We address the case of a simplex first because it is easy.

Proposition 3.3. *Suppose that S is a $0\backslash 1$ polytope with $\mathbb{1} \notin S$. Then $\mathrm{vex}_S[f](\cdot) = 0$. In particular, $\mathrm{vex}_{\Delta_n}[f](\cdot) = 0$, and the error due to this envelope is equal to α^α/d^d.*

Proof. Observe the following fact which is an immediate consequence of applying Jensen's inequality to the definition of convex envelope: for a continuous function $\phi\colon X \mapsto [\phi_0, \infty)$ for some finite ϕ_0 and bounded polyhedral domain X, if $\phi(v) = \phi_0$ for every vertex v of X, then $\mathrm{vex}_X[\phi](\cdot) = \phi_0$. Since $f(x) \geq 0$ for $x \geq \mathbf{0}$ and $f(x) = 0$ for $x \in \{0,1\}^n \setminus \{\mathbb{1}\}$, it follows from the assumption on S that $\mathrm{vex}_S[f](\cdot) = 0$. The standard n-simplex Δ_n satisfies the assumption on S and so the convex envelope over it is the zero function, thereby making the error equal to $f^{\max}_{\Delta_n}$. This value was argued in the proof of Proposition 3.2 to be equal to α^α/d^d. □

Hereafter, we let S be an arbitrary subset of $[0,1]^n$, with a special interest in $S = [0,1]^n$, or more generally $S \supseteq \Delta^{\mathbb{1}}_n(\lambda)$, where

$$\Delta^{\mathbb{1}}_n(\lambda) := \mathrm{conv}\{\mathbb{1}, \mathbb{1} - \lambda_1 \mathbf{e}_1, \ldots, \mathbb{1} - \lambda_n \mathbf{e}_n\} = \left\{ x \leq \mathbb{1} \mid \sum_{j=1}^n \frac{x_j}{\lambda_j} \geq \sum_{j=1}^n \frac{1}{\lambda_j} - 1 \right\}, \quad \mathbf{0} < \lambda \leq \mathbb{1}, \tag{6a}$$

is a n-simplex cornered at $\mathbb{1}$. For convenience, we write $\Delta^{\mathbb{1}}_n(\mathbb{1})$ simply as $\Delta^{\mathbb{1}}_n$. The motivation for studying the case $S \supseteq \Delta^{\mathbb{1}}_n(\lambda)$ is clear from Proposition 3.3 which highlights the significance of the vertex $\mathbb{1}$ belonging to S. Also note that the polytope $\Delta^{\mathbf{0}}_n$, the complement of $\Delta^{\mathbb{1}}_n$ defined as

$$\Delta^{\mathbf{0}}_n := \mathrm{conv}\left(\{0,1\}^n \setminus \{\mathbb{1}\}\right) = \left\{ x \in [0,1]^n \mid \sum_{j=1}^n x_j \leq n - 1 \right\}, \tag{6b}$$

is a $0\backslash 1$ polytope not containing $\mathbb{1}$. Note that $\Delta^{\mathbf{0}}_n$ is *not* the simplex cornered at $\mathbf{0}$, which was defined in §1 to be Δ_n. If $\Delta^{\mathbf{0}}_n \subseteq S \subseteq [0,1]^n$, $\mathrm{vex}_S[f](x) = 0$ for all $x \in \Delta^{\mathbf{0}}_n$, and therefore one would be interested in finding strong convex underestimators of x^α over $S \setminus \Delta^{\mathbf{0}}_n$. We will derive a piecewise linear convex underestimator later in Proposition 3.5.

We begin by establishing an error bound in Theorem 3.2. This bound does not have an explicit expression or formula, rather it is stated as the infimum of a certain function. However, it serves as a stepping stone towards deriving explicit error bounds in §3.2.2 that depend only on the degree of the polynomial, and hence towards proving our main result in §3.3.

3.2.1 Implicit bound

Unlike §3.1 where we calculate the error from a specific concave overestimator, here we consider a general convex underestimator defined as the pointwise supremum of a family of affine functions,

$$f_{\mathcal{B}}^{\text{cvx}}(x) := \max\left\{0, \sup_{\beta\in\mathcal{B}} \sigma(\beta) + \sum_{j=1}^{n} \beta_j(x_j - 1)\right\}, \tag{7a}$$

for some nonempty (possibly countably infinite) set $\mathcal{B} \subseteq \mathbb{1} + \mathbb{R}^n_+$, where

$$\sigma(\beta) := \min_{x\in S} x^\alpha - \sum_{j=1}^{n} \beta_j(x_j - 1) \tag{7b}$$

for each $\beta \in \mathcal{B}$ to ensure that the linear function $\sigma(\beta)+\sum_{j=1}^n \beta_j(x_j-1)$ underestimates and touches the graph of x^α. For finite $\mathcal{B}$, $f_{\mathcal{B}}^{\text{cvx}}$ is a piecewise linear convex underestimator, otherwise $f_{\mathcal{B}}^{\text{cvx}}$ could represent the convex envelope of x^α over S. The assumption of nonnegativity on β is due to the fact that the gradient of x^α at any point in $\mathbb{R}^n_+$ is a nonnegative vector. For convenience, we allow only positive β and scale it greater than equal to 1 by assuming $\mathcal{B} \subseteq \mathbb{1} + \mathbb{R}^n_+$. The multilinear monomial with $S = [0,1]^n$ would have $f_{\mathcal{B}}^{\text{cvx}}(x) = \max\{0, 1+\sum_{j=1}^n (x_j - 1)\}$ (cf. (4a)) with $\mathcal{B} = \{\mathbb{1}\}$ and $\sigma(\mathbb{1}) = 1$.

Denote $|\beta| := \sum_{j=1}^n \beta_j$. This gives us

$$\sigma(\beta) = |\beta| + \min_{x\in S} x^\alpha - \beta^\top x. \tag{7c}$$

Towards proving our main error bound in terms of only the degree of the monomial, we first obtain in Theorem 3.2 a error bound that depends on $\sigma(\beta)$'s. We make some remarks on $\sigma(\beta)$ here. An explicit formula for $\sigma(\beta)$ for arbitrary S seems hard and the function is expected to be nonconvex ($\sigma(\beta)$ is a translate of the negative of the Fenchel conjugate of x^α). However, it is possible to find bounds on it, which we state next.

Proposition 3.4. *We have the following for $\sigma(\beta)$ when $S \subseteq [0,1]^n$:*

1. $0 \le \sigma(\beta) < |\beta|$.
2. *If $S = [0,1]^n$, then $0 \le \sigma(\beta) \le 1$.*

Let $(\cdot)$ be the permutation that sorts β as $\beta_{(1)} \ge \beta_{(2)} \ge \cdots \ge \beta_{(n)}$.

3. *If $\Delta_n^{\mathbf{0}} \cap \Delta_n^{\mathbb{1}} \subseteq S \subseteq \Delta_n^{\mathbf{0}}$, then $\sigma(\beta) = \beta_{(n)}$.*

The proof is moved to Appendix A. The case $S = \Delta_n$ is not covered in the above proposition since the error over Δ_n was already dealt with in Proposition 3.3 and hence we would have no use of the bounds on $\sigma(\beta)$ in this case.

To establish an upper bound on $x^\alpha - f_{\mathcal{B}}^{\text{cvx}}(x)$, we define the following constants for every linear underestimator $\sigma(\beta) + \sum_j \beta_j(x_j - 1)$:

$$r(\beta,\kappa) := \max_j \frac{\beta_j}{\kappa_j}, \quad \mathcal{C}(\beta,\kappa) := \left(1 - \frac{\sigma(\beta)}{|\beta|}\right)^{\frac{|\beta|}{r(\beta,\kappa)}}, \quad \text{for } \kappa \ge \mathbb{1}. \tag{8}$$

It is clear that $r(\beta,\kappa) < |\beta|$ and so $0 < r(\beta,\kappa)/|\beta| < 1$. Since $0 \le \sigma(\beta) < |\beta|$ by Proposition 3.4, we have $0 < \mathcal{C}(\beta,\kappa) \le 1$. For any $\beta \in \mathcal{B}$, $r(\beta,\cdot)$ is a nonincreasing function and so $\mathcal{C}(\beta,\cdot)$ is also a nonincreasing function:

$$0 < \mathcal{C}(\beta,\kappa') \le \mathcal{C}(\beta,\kappa) \le 1, \quad \mathbb{1} \le \kappa \le \kappa'. \tag{9}$$

We do not know how $\mathcal{C}(\cdot,\kappa)$ behaves. The significance of the scalar $\mathcal{C}(\beta,\kappa)$ is as follows.

Lemma 3.2. *Define* $\varphi_{\beta,\kappa}\colon t \mapsto |\beta| - \sigma(\beta) + t - |\beta|\, t^{\frac{r(\beta,\kappa)}{|\beta|}}$. *For* κ, β *with* $\kappa \not\geq \beta$,

$$\mathcal{C}(\beta,\kappa) = \max_{t\in[0,1]} \min\{t, \varphi_{\beta,\kappa}(t)\}.$$

Proof. Since $r(\beta,\kappa)/|\beta| \in (0,1)$, $\varphi_{\beta,\kappa}$ is convex over $[0,\infty)$. It is decreasing only over $[0,t_0]$, where $t_0 := r(\beta,\kappa)^{|\beta|/(|\beta|-r(\beta,\kappa))}$ is the unique stationary point of $\varphi_{\beta,\kappa}$. Note that $\varphi_{\beta,\kappa}(0) = |\beta| - \sigma(\beta) > 0$ and observe that $\mathcal{C}(\beta,\kappa)$, which lies in $(0,1]$, is the unique fixed point of $\varphi_{\beta,\kappa}$ on $\mathbb{R}$. Hence $\min\{t, \varphi_{\beta,\kappa}(t)\} = t$ if and only if $t \in [0, \mathcal{C}(\beta)]$. The assumption $\kappa \not\geq \beta$ is equivalent to $r(\beta,\kappa) \geq 1$. Therefore $|\beta|/r(\beta,\kappa) \leq |\beta|$. We claim that

$$\left(1 - \frac{\sigma(\beta)}{|\beta|}\right)^{\frac{|\beta|}{r(\beta,\kappa)}} \geq \left(1 - \frac{\sigma(\beta)}{|\beta|}\right)^{|\beta|} \geq 1 - \sigma(\beta).$$

The first inequality is obvious whereas the second is due to the monotonicity of the function $\sigma \mapsto (1 - \frac{\sigma}{|\beta|})^{|\beta|} + \sigma - 1$ on $[0,|\beta|]$. Thus we have argued that $\varphi_{\beta,\kappa}(1) \leq \varphi_{\beta,\kappa}(\mathcal{C}(\beta,\kappa))$. Now the monotone behavior of $\varphi_{\beta,\kappa}$ on $[t_0,\infty)$ means that $t_0 > \mathcal{C}(\beta,\kappa)$ because otherwise we would have the contradiction $\varphi_{\beta,\kappa}(1) > \varphi_{\beta,\kappa}(\mathcal{C}(\beta,\kappa))$. This implies that the maximum value of $\min\{t, \varphi_{\beta,\kappa}(t)\}$ on the $[0,1]$ interval occurs at $t = \mathcal{C}(\beta,\kappa)$ and, since this is a fixed point, it is equal to $\mathcal{C}(\beta,\kappa)$. □

Since we need $\kappa \not\geq \beta$ in the above lemma and forthcoming results, define

$$\begin{aligned}\mathcal{K}(\mathcal{B}) &:= \{\kappa \in \mathbb{R}^n_+ \mid \mathbb{1} \leq \kappa \leq \alpha,\ \kappa \not\geq \beta\ \ \forall \beta \in \mathcal{B}\} \\ &= \{\kappa \in \mathbb{R}^n_+ \mid \mathbb{1} \leq \kappa \leq \alpha,\ \kappa_j \leq \min_{\beta\in\mathcal{B}} \beta_j \ \text{ for some } j\}.\end{aligned} \tag{10}$$

The assumption $\mathcal{B} \subseteq \mathbb{1} + \mathbb{R}^n_+$ makes it obvious that $\mathbb{1} \in \mathcal{K}(\mathcal{B})$. The structure of $\varphi_{\beta,\kappa}$ discussed in the proof of Lemma 3.2 implies the following claim.

Lemma 3.3. *For every* $\kappa \in \mathcal{K}(\mathcal{B})$,

$$\max_{t\in[0,1]} \inf_{\beta\in\mathcal{B}} \min\{t, \varphi_{\beta,\kappa}(t)\} \;=\; \inf_{\beta\in\mathcal{B}} \max_{t\in[0,1]} \min\{t, \varphi_{\beta,\kappa}(t)\} \;=\; \inf_{\beta\in\mathcal{B}} \mathcal{C}(\beta,\kappa).$$

We are now ready to state our upper bound on error from the convex underestimator $f^{\text{cvx}}_{\mathcal{B}}$.

Theorem 3.2.

$$\mu\left(\mathcal{G}_S(f^{\text{cvx}}_{\mathcal{B}})\right) \;\leq\; \inf_{\beta\in\mathcal{B}} \mathcal{C}(\beta,\kappa^*),$$

where κ^* *is a maximal element of* $\mathcal{K}(\mathcal{B})$ *under the partial order* $\leq$. *In particular, if there exists some* j *such that* $\alpha_j \leq \beta_j$ *for all* $\beta \in \mathcal{B}$, *then*

$$\mu\left(\mathcal{G}_S(f^{\text{cvx}}_{\mathcal{B}})\right) \;\leq\; \mathcal{C}(\beta^*,\alpha) \;:=\; \inf_{\beta\in\mathcal{B}} \mathcal{C}(\beta,\alpha),$$

and this bound is tight only if $\beta^* = \alpha$ *and is attained only at the point* $\mathcal{C}(\alpha,\alpha)^{1/d}\mathbb{1} \in \text{rel. int}\,\{\mathbf{0},\mathbb{1}\}$.

Proof. Choose some $\kappa \in \mathcal{K}(\mathcal{B})$. For every $x \in [0,1]^n$ and j, $\beta_j/\kappa_j \leq r(\beta,\kappa)$ gives us $x_j^{\beta_j/\kappa_j} \geq x_j^{r(\beta,\kappa)} \geq 0$ and $\kappa_j \leq \alpha_j$ gives us $x_j^{\kappa_j} \geq x_j^{\alpha_j} \geq 0$. Thus

$$x^\beta = \prod_{j=1}^n \left(x_j^{\frac{\beta_j}{\kappa_j}}\right)^{\kappa_j} \geq \prod_{j=1}^n \left(x_j^{r(\beta,\kappa)}\right)^{\kappa_j} = \prod_{j=1}^n \left(x_j^{\kappa_j}\right)^{r(\beta,\kappa)} \geq \prod_{j=1}^n \left(x_j^{\alpha_j}\right)^{r(\beta,\kappa)} = (x^\alpha)^{r(\beta,\kappa)}. \tag{11a}$$

The generalized arithmetic-geometric means inequality tells us that $\sum_{j=1}^n \beta_j x_j \geq |\beta| \, (x^\beta)^{1/|\beta|}$, which combined with (11a) leads to $\sum_{j=1}^n \beta_j x_j \geq |\beta| \, (x^\alpha)^{r(\beta,\kappa)/|\beta|}$. Therefore

$$\begin{aligned} x^\alpha - f_{\mathcal{B}}^{\text{cvx}}(x) &= \min\left\{ x^\alpha, \inf_{\beta\in\mathcal{B}} x^\alpha - \sigma(\beta) + |\beta| - \sum_{j=1}^n \beta_j x_j \right\} \\ &\leq \min\left\{ x^\alpha, \inf_{\beta\in\mathcal{B}} x^\alpha - \sigma(\beta) + |\beta| - |\beta| \, (x^\alpha)^{\frac{r(\beta,\kappa)}{|\beta|}} \right\} \\ &= \inf_{\beta\in\mathcal{B}} \min\left\{ x^\alpha, x^\alpha - \sigma(\beta) + |\beta| - |\beta| \, (x^\alpha)^{\frac{r(\beta,\kappa)}{|\beta|}} \right\}, \end{aligned}$$

which leads to

$$\max_{x\in S} x^\alpha - f_{\mathcal{B}}^{\text{cvx}}(x) \leq \max_{x\in S} \inf_{\beta\in\mathcal{B}} \min\left\{ x^\alpha, x^\alpha - \sigma(\beta) + |\beta| - |\beta| \, (x^\alpha)^{\frac{r(\beta,\kappa)}{|\beta|}} \right\}. \tag{11b}$$

Since $f(x) = x^\alpha$ is a continuous function with minimum and maximum values $f_S^{\min}, f_S^{\max} \in [0,1]$ on S, the intermediate value theorem implies that (11b) transforms to

$$\max_{x\in S} x^\alpha - f_{\mathcal{B}}^{\text{cvx}}(x) \leq \max_{t\in[0,1]} \inf_{\beta\in\mathcal{B}} \min\{t, \varphi_{\beta,\kappa}(t)\}, \tag{11c}$$

where $\varphi_{\beta,\kappa}(t) = |\beta| - \sigma(\beta) + t - |\beta| \, t^{\frac{r(\beta,\kappa)}{|\beta|}}$ as in Lemma 3.2. Lemma 3.3 leads to $\max_{x\in S} x^\alpha - f_{\mathcal{B}}^{\text{cvx}}(x) \leq \inf_\beta \mathcal{C}(\beta,\kappa)$. Since κ was arbitrarily chosen in $\mathcal{K}(\mathcal{B})$ and we know from (9) that $\mathcal{C}(\beta,\cdot)$ is a nonincreasing function for every β, we may set κ equal to a maximal κ^* to obtain $\max_{x\in S} x^\alpha - f_{\mathcal{B}}^{\text{cvx}}(x) \leq \inf_{\beta\in\mathcal{B}} \mathcal{C}(\beta,\kappa^*)$. If $\alpha_j \leq \min_{\beta\in\mathcal{B}} \beta_j$ for some j, then α is the unique maximal element in $\mathcal{K}(\mathcal{B})$ and setting $\kappa^* = \alpha$ yields the upper bound $\inf_{\beta\in\mathcal{B}} \mathcal{C}(\beta,\alpha)$.

The bound $\mathcal{C}(\beta^*,\alpha)$ is tight if and only if there is equality throughout in (11a) with $\kappa = \alpha, \beta = \beta^*$, and in the means inequality $\sum_{i=1}^n \beta_i^* x_i \geq |\beta^*| \, (x^{\beta^*})^{1/|\beta^*|}$. Equation (11a) is an equality if and only if $\kappa = \alpha = \beta^*$, implying that $\beta^* = \alpha$ is a necessary condition for tightness. The means inequality is an equality if and only if $x_1 = x_2 = \cdots = x_n$ and hence the bound can be attained only at $\mathcal{C}(\alpha,\alpha)^{1/d}\mathbb{1}$. □

Remark 4. We will show in Proposition 3.5 that $\sigma(\alpha) = 1$, implying that $\mathcal{C}(\alpha,\alpha) = (1-1/d)^d$, which is exactly the constant $\mathcal{C}_d^2$ defined in (2), and therefore the above bound is attained only at $(1-1/d)\mathbb{1}$.

Any polyhedral relaxation of the epigraph of x^α can be encoded by the set $\mathcal{B}$ in equation (7a). Hence Theorem 3.2 yields an upper bound on the error from any polyhedral relaxation that is chosen apriori. Since we do not know the behavior of $\mathcal{C}(\cdot,\kappa)$, a analytic expression for the infimum in Theorem 3.2 does not seem possible in general. Even if $\mathcal{B}$ is finite, $\mathcal{C}(\beta,\kappa)$ requires the computation of $\sigma(\beta)$, which we know to be hard in general. However, one may derive upper bounds on the error using the lower bounds on $\sigma(\beta)$ from Proposition 3.4. Note though that this does not help for $S = [0,1]^n$ because the lower bound of 0 on $\sigma(\beta)$ gives a trivial upper bound of 1 on the error.

We use the bound in Theorem 3.2 to derive a degree-dependent bound on the convex envelope error. To do so, let us view this upper bound from a different perspective. By construction of $\mathcal{C}(\beta,\kappa)$, in order to obtain a smaller error bound, we would intuitively want to pick $\mathcal{B}$ such that it contains only those β that make $\sigma(\beta)$ to be as high as possible. For $S = [0,1]^n$, or more generally S containing $\mathbb{1}$, we know the highest that $\sigma(\beta)$ can be is 1. Hence we could do the following reverse construction — instead of choosing a set $\mathcal{B}$ and then computing $\sigma(\beta)$ for each $\beta \in \mathcal{B}$ as done before,

we could fix $\sigma(\beta) = 1$ and find the values of $\beta \geq \mathbb{1}$ that enable $1 + \sum_{j=1}^{n} \beta_j(x_j - 1)$ to be a valid linear underestimator (cf. equation (7a)) to x^α over S. This would alleviate the issue of having to compute $\sigma(\beta)$ for $\mathcal{C}(\beta, \kappa)$ and could possibly lead to simpler and explicit error bounds that depend only on exponent α and degree d. We follow this path for the rest of this section. Note also that the convex envelope of the multilinear monomial $\prod_{j=1}^{n} x_j$ over $[0,1]^n$ is $\max\{0, 1 + \sum_j (x_j - 1)\}$, meaning that there is only one β, the vector $\mathbb{1}$, with $\sigma(\mathbb{1}) = 1$. Thus our forthcoming derivation implies the error from the convex envelope of a multilinear monomial over $[0,1]^n$.

3.2.2 Explicit bounds

Denote

$$\ell_\beta(x) := 1 + \sum_{j=1}^{n} \beta_j(x_j - 1), \quad \beta \geq \mathbb{1}.$$

This linear function is exact at $x = \mathbb{1}$: $\ell_\beta(\mathbb{1}) = 1 = f(\mathbb{1})$. The convex underestimator on x^α is

$$g^{\text{cvx}}_{\mathcal{B}_1}(x) = \max\left\{0, \sup_{\beta \in \mathcal{B}_1} \ell_\beta(x)\right\}, \quad \text{where } \mathcal{B}_1 := \{\beta \geq \mathbb{1} \mid \ell_\beta(x) \leq x^\alpha \ \forall x \in S\}. \tag{12}$$

$\mathcal{B}_1$ is a closed convex set[2], due to linearity of $\ell_\beta(x)$ in β for fixed x, and it represents all the linear functions that are exact at $x = \mathbb{1}$ and underestimate x^α everywhere on $S \subseteq [0,1]^n$. Clearly, $\beta \leq \beta'$ implies $\ell_\beta(x) \geq \ell_{\beta'}(x)$ for all $x \in [0,1]^n$, and so $\beta \in \mathcal{B}_1$ implies $\beta' \in \mathcal{B}_1$. But then we could simply delete such a β' from $\mathcal{B}_1$ without affecting the supremum in $g^{\text{cvx}}_{\mathcal{B}_1}$. Hence we define the nondominated subset of $\mathcal{B}_1$ to be the following:

$$\mathcal{ND}(\mathcal{B}_1) := \{\beta \in \mathcal{B}_1 \mid \nexists\, \mathbb{1} \leq \beta' \lneqq \beta \text{ s.t. } \ell_{\beta'}(x) \leq x^\alpha \ \forall x \in S\}, \tag{13}$$

so that

$$g^{\text{cvx}}_{\mathcal{B}_1}(x) = \max\left\{0, \sup_{\beta \in \mathcal{ND}(\mathcal{B}_1)} \ell_\beta(x)\right\}. \tag{14}$$

A strong error bound from $g^{\text{cvx}}_{\mathcal{B}_1}$ would obviously depend on the elements in $\mathcal{ND}(\mathcal{B}_1)$ (cf. Theorem 3.2), making it important to obtain a (partial) characterization of $\mathcal{B}_1$ and $\mathcal{ND}(\mathcal{B}_1)$ based on the structure of S. We mention two cases where $\mathcal{ND}(\mathcal{B}_1)$ is easily seen to be equal to $\{\mathbb{1}\}$, the most trivial value.

Multilinear over $[0,1]^n$. Here $\alpha = \mathbb{1}, S = [0,1]^n$ and equation (4a) tells us $\mathbb{1} \in \mathcal{B}_1$, and therefore $\mathcal{ND}(\mathcal{B}_1) = \{\mathbb{1}\}$.

We will generalize this in Proposition 3.5 by showing that $\mathcal{ND}(\mathcal{B}_1) = \{\alpha\}$ when $S \supseteq \Delta^{\mathbb{1}}_n(\lambda)$.

Subsets of $\Delta^{\mathbf{0}}_n$. Here α is arbitrary and $S \subseteq \Delta^{\mathbf{0}}_n = \text{conv}(\{0,1\}^n \setminus \{\mathbb{1}\})$. We know that ℓ_β is valid to S if and only if $\sigma(\beta) \geq 1$, where $\sigma(\beta) = \min_{x \in S} x^\alpha - \sum_j \beta_j(x_j - 1)$. Clearly ℓ_β is valid to S if it is valid to $\Delta^{\mathbf{0}}_n$. We argued in Proposition 3.4 that $\sigma(\beta) = \beta_{(n)}$ for $\Delta^{\mathbf{0}}_n$ and since $\beta \geq \mathbb{1}$ by assumption, it follows that ℓ_β is valid to S for all $\beta \geq \mathbb{1}$. Therefore $\mathcal{ND}(\mathcal{B}_1) = \{\mathbb{1}\}$.

[2] It does not seem that $\mathcal{B}_1$ will be a polyhedron even for $S = [0,1]^n$. Since general monomials are not vertex-extendable over $[0,1]^n$, it is not clear whether the validity of ℓ_β over the entire box can be certified by checking at only a finite number of points.

For an arbitrary integer exponent α and $S \not\subseteq \Delta_n^{\mathbf{0}}$, it is not at all obvious what the set $\mathcal{B}_1$ should be. Note that this includes the case of a monomial over $S = [0,1]^n$. As a generalization of the multilinear case, is it true that $\alpha \in \mathcal{B}_1$? The function $\ell_\alpha(\cdot)$ is Taylor's first-order approximation of x^α at the point $x = \mathbb{1}$. Having $\alpha \in \mathcal{B}_1$ would mean that the gradient inequality at $x = \mathbb{1}$ holds true, which is not at all obvious since x^α is a nonconvex function. We show in Proposition 3.5 that $\alpha \in \mathcal{B}_1$ is always true, regardless of S, and in fact construct a $\beta \leq \alpha$ with $\beta \in \mathcal{B}_1$, so that $\alpha \notin \mathcal{ND}(\mathcal{B}_1)$ in general. This β depends on S and is constructed by taking projections of S onto each coordinate. We also present some conditions under which $\mathcal{ND}(\mathcal{B}_1)$ can be (partially) characterized.

The following technical lemma will be useful. It is proved in Appendix A.

Lemma 3.4. *Let $\lambda_1 \in \mathbb{Z}_{\geq 1}, \lambda_2 \geq 1$. Consider the univariate polynomial $\phi(\sigma) := (1-\sigma)^{\lambda_1} + \lambda_2\sigma - 1$ which has a trivial root at 0.*

1. *If $\lambda_2 \geq \lambda_1$, $\phi(\sigma) > 0$ for all $\sigma \in (0,1]$.*

For $\lambda_2 < \lambda_1$,

2. *ϕ has exactly one root in $(0,1]$, denoted σ^*, and $\sigma^* > 1 - (\lambda_2/\lambda_1)^{\frac{1}{\lambda_1 - 1}}$.*
3. *$\phi(\sigma) < 0$ for all $\sigma \in (0, \sigma^*)$ and $\phi(\sigma) > 0$ for all $\sigma \in (\sigma^*, 1]$.*
4. *$(1-\sigma)^{\lambda_1} > 1 - \lambda\sigma$ for all $\lambda \in (\lambda_2, \infty), \sigma \in [\sigma^*, 1]$, and $(1-\sigma)^{\lambda_1} < 1 - \lambda\sigma$ for all $\lambda \in [1, \lambda_2), \sigma \in [0, \sigma^*)$*

Finally, there is a root in $(1, \infty)$ if and only if λ_1 is odd, and there is a root in $(-\infty, 0)$ if and only if $\lambda_2 > \lambda_1$.

Remark 5. Finding an analytic expression for the root σ^* seems difficult, and an algebraic root may not even exist, as can be verified using computational algebra software for the polynomial $\phi(\sigma) = (1-\sigma)^6 + 3\sigma - 1$, whose roots are in bijection to that of $\sigma^6 - 3\sigma + 2$ under the mapping $\sigma \mapsto 1 - \sigma$. However, our forthcoming analysis circumvents this issue since it does not depend on the exact value of σ^*.

We also need to introduce some notation. For every i, denote the projection of S onto the x_i-subspace by

$$\mathrm{Proj}_{x_i} S := [1 - \sigma_i^1, 1 - \sigma_i^2], \qquad \text{for some } 0 \leq \sigma_i^2 \leq \sigma_i^1 \leq 1,$$

and define

$$\gamma_i := \begin{cases} \dfrac{1 - (1-\sigma_i^2)^{\alpha_i}}{\sigma_i^2} & \text{if } \sigma_i^2 > 0, \\ \alpha_i & \text{if } \sigma_i^2 = 0, \end{cases} \qquad i = 1, \ldots, n. \tag{15}$$

This γ is exactly the γ from the statement of Theorem 1.1 in §1.1.1. Note that if $S \cap E_i \neq \emptyset$, $S \cap E_j \neq \emptyset$ for distinct i, j, then $\sigma^2 = \mathbf{0}$.

Lemma 3.5. *$1 \leq \gamma_i < \alpha_i$ for every i with $\sigma_i^2 > 0$. Hence $\gamma = \alpha$ if and only if $\sigma^2 = \mathbf{0}$.*

Proof. $\gamma_i \geq 1$ is obvious due to $\sigma_i^2 \in (0,1)$ and $\alpha_i \geq 1$. Since $\alpha_i \in \mathbb{Z}_{\geq 1}$, we have $\frac{1-\chi^{\alpha_i}}{1-\chi} = 1 + \chi + \chi^2 + \cdots + \chi^{\alpha_i - 1}$, making $\frac{1-\chi^{\alpha_i}}{1-\chi}$ an increasing function on $[0,1]$. Hence, by complementing to $\sigma = 1 - \chi$, $\frac{1-(1-\sigma)^{\alpha_i}}{\sigma}$ is a decreasing function on $[0,1]$. L'Hôpital's rule gives $\lim_{\sigma \to 0} \frac{1-(1-\sigma)^{\alpha_i}}{\sigma} = \alpha_i$. ∎

Proposition 3.5. *We have the following:*

1. $\gamma, \alpha \in \mathcal{B}_1$.

Consider any $\beta \geq \mathbb{1}$ *and suppose* $I := \{i \mid S \cap \operatorname{rel.int} E_i \neq \emptyset, \beta_i \leq \alpha_i\}$ *is nonempty. For* $i \in I$ *denote* $1 - \tau_i^2 = \max\{x_i \mid x \in S \cap E_i\}$.

2. $\beta \in \mathcal{B}_1$ *only if* $\alpha_i(1-\tau_i^2)^{\alpha_i - 1} \leq \beta_i \leq \alpha_i$ *for* $i \in I$ *with* $\tau_i^2 > \sigma_i^2$, *and* $\gamma_i \leq \beta_i \leq \alpha_i$ *for* $i \in I$ *with* $\tau_i^2 = \sigma_i^2$.

3. *Suppose* $\mathbb{1} \in S$. *Then* $\beta \in \mathcal{B}_1$ *only if* $\beta_i = \alpha_i$ *for all* $i \in I$.

Finally,

4. *If* $S \supseteq \Delta_n^{\mathbb{1}}(\lambda) := \operatorname{conv}(\cup_{i=1}^n \{\mathbb{1} - \lambda_i \mathbf{e}_i\})$ *for some* $\mathbf{0} < \lambda \leq \mathbb{1}$, *then* $\mathcal{ND}(\mathcal{B}_1) = \{\alpha\}$.

Proof. (1) Observe that showing $\ell_\beta(x) \leq x^\alpha$ for all $x \in S$ is equivalent to showing $\ell_\beta(x) \leq x^\alpha$ for all $x \in S$ such that $x > \mathbf{0}, x \neq \mathbb{1}$. Indeed, $\ell_\beta(x)$ is exact at $x = \mathbb{1}$ and for any $x \in E_0$, $x_i = 0$ implies that $\ell_\beta(x) = 1 - \beta_i + \sum_{j \neq i} \beta_j(x_j - 1)$ which is nonpositive due to $\beta \geq \mathbb{1}$ and $x \in [0,1]^n$. Therefore to show $\gamma \in \mathcal{B}_1$, we prove $\ell_\gamma(x) \leq x^\alpha$ for every $x \in S, x > \mathbf{0}, x \neq \mathbb{1}$.

Consider such an x and let $k = |\{i \mid 0 < x_i < 1\}|$. Assume wlog that $x_i = 1 - \sigma_i$ for $i = 1, \ldots, k$ with $\sigma_i \in [\sigma_i^2, \sigma_i^1], \sigma_i \in (0,1)$, and $x_i = 1$ for $i \geq k+1$. We must show that

$$\prod_{i=1}^k (1-\sigma_i)^{\alpha_i} \geq 1 - \sum_{i=1}^k \gamma_i \sigma_i.$$

We argue this inequality by induction on k. Take $k = 1$. We obtain $(1-\sigma_1)^{\alpha_1} \geq 1 - \gamma_1\sigma_1$ from the following claim.

Claim 3.2. *For any* i *and* $\sigma \in [\sigma_i^2, 1]$, *we have* $(1-\sigma)^{\alpha_i} \geq 1 - \beta_i\sigma$ *for all* $\beta_i \geq \gamma_i$.

Proof of Claim. If $\sigma_i^2 = 0$, then $\gamma_i = \alpha_i$ and applying the first item in Lemma 3.4 with $\lambda_1 = \alpha_i$ and $\lambda_2 = \beta_i$ tells us $(1-\sigma)^{\alpha_i} \geq 1 - \beta_i\sigma$ for all $\beta_i \geq \gamma_i$. Otherwise $\sigma_i^2 > 0$ and Lemma 3.5 allows us to apply Lemma 3.4 with $\lambda_1 = \alpha_i$ and $\lambda_2 = \gamma_i$. It is readily seen from the construction of γ in (15) that σ_i^2 is a root of $\phi(\omega) = (1-\omega)^{\alpha_i} + \gamma_i\omega - 1$ and by the second item of Lemma 3.4, it is the unique root in $(0,1]$. Now $\sigma \in [\sigma_i^2, 1]$ and the fourth item of Lemma 3.4 yield $(1-\sigma)^{\alpha_i} \geq 1 - \beta_i\sigma$ for all $\beta_i \geq \gamma_i$. ◇

Assume that the inequality is true for $k \geq 1$ and let us argue it for $k+1$. The induction hypothesis gives us

$$\prod_{i=1}^k (1-\sigma_i)^{\alpha_i} - \gamma_{k+1}\sigma_{k+1} \geq 1 - \sum_{i=1}^k \gamma_i\sigma_i - \gamma_{k+1}\sigma_{k+1} = 1 + \sum_{i=1}^n \gamma_i(x_i - 1).$$

Let $\prod_{i=1}^k (1-\sigma_i)^{\alpha_i} = 1 + \chi$ for some $\chi \in (-1, 0)$; such a χ exists because $\sigma_i \in (0,1)$ for $i = 1, \ldots, k$. Hence, the induction hypothesis becomes

$$1 + \chi - \gamma_{k+1}\sigma_{k+1} \geq 1 + \sum_{i=1}^n \gamma_i(x_i - 1).$$

Now,

$$x^\alpha = \prod_{i=1}^{k+1} (1-\sigma_i)^{\alpha_i} = (1+\chi)(1-\sigma_{k+1})^{\alpha_{k+1}} \geq (1+\chi)(1-\gamma_{k+1}\sigma_{k+1}),$$

where the inequality is by applying Claim 3.2 to $i = k+1$, and using $1+\chi > 0$. Since $(1+\chi)(1-\gamma_{k+1}\sigma_{k+1}) = 1+\chi-\gamma_{k+1}\sigma_{k+1}-\gamma_{k+1}\sigma_{k+1}\chi$ and $\chi < 0, \gamma_{k+1}, \sigma_{k+1} > 0$, we have

$$x^\alpha > 1+\chi-\gamma_{k+1}\sigma_{k+1} \geq 1+\sum_{i=1}^n \gamma_i(x_i-1), \tag{16}$$

where $\geq$ is from the induction hypothesis. This finishes our inductive proof for showing $\gamma \in \mathcal{B}_1$. Thus every $x \in S$ with $|\{i \mid x_i \in (0,1)\}| \geq 2$ has $\ell_\gamma(x) < x^\alpha$. The closedness of $\mathcal{B}_1$ under monotonicity and $\gamma \leq \alpha$ give us $\alpha \in \mathcal{B}_1$.

(2) Choose some $i \in I$. If $\beta_i = \alpha_i$, then there is nothing to prove because $\alpha_i \geq \gamma_i$ and $\tau_i^2 \in [0,1]$. So assume $\beta_i < \alpha_i$. Consider a point $\bar{x} \in S \cap \operatorname{rel.int} E_i$, which can be written as $\bar{x}_i = 1-\tau$ and $\bar{x}_j = 1\ \forall j \neq i$, where $\tau = \tau_i^2$ if $\tau_i^2 > 0$, otherwise τ is a small positive real. Note that $\bar{x}^\alpha = (1-\tau_i^2)^{\alpha_i}$ and $\ell_\beta(\bar{x}) = 1-\beta_i\tau$. The second and third items of Lemma 3.4 with $\lambda_1 = \alpha_i, \lambda_2 = \beta_i$ tell us that $(1-\tau)^{\alpha_i} < 1-\beta_i\tau$ if $\tau \leq 1-(\beta_i/\alpha_i)^{\frac{1}{\alpha_i-1}}$. This means that $\tau_i^2 \geq 1-(\beta_i/\alpha_i)^{\frac{1}{\alpha_i-1}}$, which rearranges to $\beta_i \geq \alpha_i(1-\tau_i^2)^{\alpha_i-1}$, is necessary for ℓ_β to be a valid linear underestimator.

(3) It is easy to see that the convexity of S makes $\mathbb{1} \in S$ equivalent to $\tau_i^2 = 0$ for all $i \in I$. We also have $\mathbb{1} \in S$ implying $\sigma^2 = \mathbf{0}$. Therefore $\gamma = \alpha$. Now (2) gives us $\beta_i = \alpha_i$ for $i \in I$.

(4) The assumption $S \supseteq \Delta_n^{\mathbb{1}}(\lambda)$ implies $S \cap \operatorname{rel.int} E_i \neq \emptyset$ for all i, $\mathbb{1} \in S$, $\tau^2 = \sigma^2 = \mathbf{0}$ and hence $\gamma = \alpha$. The claim then follows from (3). □

Remark 6. Due to the functions $h_1(t) = \alpha_i(1-t)^{\alpha_i-1}$ and $h_2(t) = h_1(t) - (1-(1-t)^{\alpha_i})/t$ being nonincreasing and nonpositive, respectively, over $[0,1]$, it follows that $\alpha_i(1-\tau_i^2)^{\alpha_i-1} \leq \gamma_i$ in Proposition 3.5, meaning that the lower bound on β_i with $\tau_i^2 > \sigma_i^2$ is weaker than the lower bound on β_i with $\tau_i^2 = \sigma_i^2$. This happens because while arguing this part, we used a lower bound on the root of $(1-\sigma)^{\lambda_1} + \lambda_2\sigma - 1$ in $(0,1]$ from Lemma 3.4, since finding a analytic expression for the root seems difficult (cf. Remark 5). Therefore if $I \setminus I' \neq \emptyset$, then there is no guarantee that γ is a nondominated point in $\mathcal{B}_1$.

Remark 7. The second item in Proposition 3.5 indicates that a tight lower bound on a valid β can get arbitrarily close to α.

The vector γ in (15) can be constructed only when projections of S are readily available or can be computed quickly. When these projections are difficult to compute, we could use the first claim of Proposition 3.5 telling us that ℓ_α is a underestimator of x^α. The last item in this proposition provides a clean and simple expression for $g_{\mathcal{B}_1}^{\text{cvx}}$ in (14).

The preceding results on $\mathcal{B}_1$ and $\mathcal{ND}(\mathcal{B}_1)$, combined with Theorem 3.2, imply explicit bounds on the error from the convex underestimator $g_{\mathcal{B}_1}^{\text{cvx}}$. Recall the constants from (8). Denoting $\mathcal{C}(\beta,\beta)$ simply as $\mathcal{C}(\beta)$, we have for ℓ_α and ℓ_γ, respectively,:

$$\mathcal{C}(\alpha) = \left(1-\frac{1}{d}\right)^d = \mathcal{C}_d^2, \qquad \mathcal{C}(\gamma) = \left(1-\frac{1}{|\gamma|}\right)^{|\gamma|},$$

where we recall that $\mathcal{C}_d^2$ was defined in (2) and $|\gamma| = \sum_{j=1}^n \gamma_j$.

Corollary 3.3. $\mu\left(\mathcal{G}_S(g_{\mathcal{B}_1}^{\text{cvx}})\right) \leq \mathcal{C}(\gamma) \leq \mathcal{C}_d^2$, *and equality holds throughout if* $\mathbf{0} \in S$ *and* $S \supset \Delta_n^{\mathbb{1}}(\lambda)$ *for some* $\mathbf{0} < \lambda \leq \mathbb{1}$.

Proof. We first observe that $\max_{x\in S} x^\alpha - \max\{0, \ell_\gamma(x)\} \leq \mathcal{C}(\gamma)$. This is obtained by applying Theorem 3.2 with $f_{\mathcal{B}_1}^{\text{cvx}}$ replaced by $\max\{0,\ell_\gamma\}$ and noting that γ is a maximal element of $\mathcal{K}(\{\gamma\})$. Since $\gamma \in \mathcal{B}_1$ by Proposition 3.5, $g_{\mathcal{B}_1}^{\text{cvx}}(\cdot) \geq \max\{0,\ell_\gamma(\cdot)\}$ and hence $\max_{x\in S} x^\alpha - g_{\mathcal{B}_1}^{\text{cvx}}(x) \leq \mathcal{C}(\gamma)$.

Since $t\ln(1-\frac{1}{t})$ is concave increasing over $[2,\infty)$ and $\gamma \le \alpha$ by construction, we get $\mathcal{C}(\gamma) \le \mathcal{C}(\alpha)$. If $S \supseteq \Delta_n^{\mathbb{1}}(\lambda)$, then $\gamma = \alpha$ and the last claim in Proposition 3.5 tells us $\mathcal{ND}(\mathcal{B}_1) = \{\alpha\}$ and $g_{\mathcal{B}_1}^{\text{cvx}}(x) = \max\{0, \ell_\alpha(x)\}$. Now recall Theorem 3.2. We have $\beta^* = \alpha$ due to $\mathcal{ND}(\mathcal{B}_1) = \{\alpha\}$. This theorem tells us that the bound on $\max_{x\in S} x^\alpha - g_{\mathcal{B}_1}^{\text{cvx}}(x)$ can be attained only at $\mathcal{C}(\alpha)\mathbb{1}$. The assumptions $\mathbf{0} \in S$ and $\Delta_n^{\mathbb{1}}(\lambda) \subset S$ lead to $\text{conv}\{\mathbf{0}, \mathbb{1}\} \subset S$ and therefore $\mathcal{C}(\alpha)\mathbb{1} \in S$. □

A direct implication is a tight bound on the error of the convex envelope of a multilinear monomial considered over $[0,1]^n$.

Corollary 3.4. *We have* $\max_{x\in[0,1]^n} x^\alpha - \max\{0, \ell_\alpha(x)\} = \mathcal{C}_d^2$. *In particular, for a multilinear monomial,* $\mu\left(\mathcal{G}(\text{vex}_{[0,1]^n}[m])\right) = (1-\frac{1}{n})^n$.

Proof. Since $S = [0,1]^n \supset \Delta_n^{\mathbb{1}}$, the last item in Proposition 3.5 tells us $g_{\mathcal{B}_1}^{\text{cvx}}(x) = \max\{0, \ell_\alpha(x)\}$ and then the first equality follows immediately from Corollary 3.3. For a multilinear monomial, equation (4a) gives us $\text{vex}_{[0,1]^n}[m](x) = \max\{0, \ell_{\mathbb{1}}(x)\}$. The claimed error follows by using $\alpha = \mathbb{1}$ in the expression for $\mathcal{C}_d^2$. □

3.3 Convex hull error

Proof of Theorem 1.1. Since $\text{cav}_S[f](x) \le f^{\text{conc}}(x)$ for $x \in [0,1]^n$, the upper bound of $\mathcal{C}_d^1$ on $\text{cav}_S[f](x) - x^\alpha$ is due to $\mu\left(\mathcal{G}_S(f^{\text{conc}})\right) \le \mathcal{C}_d^1$ from Corollary 3.1. Similarly the upper bounds on $x^\alpha - \text{vex}_S[f](x)$ are due to $g_{\mathcal{B}_1}^{\text{cvx}}(\cdot) \le \text{vex}_S[f](\cdot)$ and Corollary 3.3. By Observation 2.1, we then have that $\mu\left(\text{conv}\,\mathcal{G}_S(f)\right) \le \max\{\mathcal{C}_d^1, \mathcal{C}_d^2\}$. To show this error is upper bounded by $\mathcal{C}_d^1$, we argue the following.

Claim 3.3. $\mathcal{C}_d^2 \le \mathcal{C}_d^1$ *for* $d \ge 2$ *and equality holds if and only if* $d = 2$.

Proof of Claim. The two constants are $\mathcal{C}_d^2 = \mathcal{C}(\alpha) = (1-1/d)^d$ and $\mathcal{C}_d^1 = (1-1/d)d^{1/(1-d)}$. Therefore the following equivalence holds:

$$\mathcal{C}_d^1 \ge \mathcal{C}_d^2 \iff \left(\frac{1}{d}\right)^{\frac{1}{d-1}} \ge \left(1-\frac{1}{d}\right)^{d-1} \iff d \le \left(\frac{d}{d-1}\right)^{(d-1)^2} \iff (d-1)^{(d-1)^2} \le d^{d(d-2)}.$$

Lemma 3.1 proves the last inequality and that it holds at equality only when $d = 2$. ◇

Thus we have $\mu\left(\text{conv}\,\mathcal{G}_S(f)\right) \le \mathcal{C}_d^1$ for any $S \subseteq [0,1]^n$.

If $\mathbf{0}, \mathbb{1} \in S$, then setting $t_1 = 0, t_2 = 1$ in Lemma 2.1 yields the critical point to be $\xi' = (1/d)^{1/(d-1)}$ so that

$$\phi(\xi') = \xi' - \xi'^d = \xi'(1-\xi'^{d-1}) = (1/d)^{\frac{1}{d-1}}(1-1/d) = \mathcal{C}_d^1.$$

Therefore the convex hull error and the concave envelope error are lower bounded by $\mathcal{C}_d^1$, making each of them equal to $\mathcal{C}_d^1$. □

The arguments used in proving Theorem 1.1 also imply that a family of convex relaxations of $\mathcal{G}_S(f)$ has error equal to $\mathcal{C}_d^1$. Recall the convex underestimator $f_{\mathcal{B}}^{\text{cvx}}$ from (7a) for any $\mathcal{B} \subseteq \mathbb{1} + \mathbb{R}_+^n$ and consider the convex relaxation

$$P_{\mathcal{B}} := \{(x,w) \in [0,1]^n \times \mathbb{R} \mid f_{\mathcal{B}}^{\text{cvx}}(x) \le w \le f^{\text{conc}}(x)\}.$$

Note that x is not restricted to be in S here. Assume $\alpha \in \mathcal{B}$. Also assume $\mathbb{1} \in S$ so that $\sigma(\beta) \le 1$ for every $\beta \in \mathcal{B}$, as per Proposition 3.4. We claim that

Proposition 3.6. $\mu\left(P_{\mathcal{B}}\right) = \mathcal{C}_d^1$.

Proof. The proof of $\mu\left(P_{\mathcal{B}}\right) \leq \mathcal{C}_d^1$ is the same as that in Theorem 1.1, along with using the assumption $\alpha \in \mathcal{B}$ to get $\max_{x\in S} x^\alpha - f_{\mathcal{B}}^{\text{cvx}}(x) \leq \max_{x\in S} x^\alpha - \max\{0, \ell_\alpha(x)\} = \mathcal{C}_d^2$. Tightness of this bound is obtained by applying Lemma 2.1 and Remark 1 after noting that $(\mathbf{0}, 0), (\mathbb{1}, 1) \in P_{\mathcal{B}}$. The point $(\mathbf{0}, 0)$ belongs to $P_{\mathcal{B}}$ because $f^{\text{conc}}(\mathbf{0}) = 0$, and $f_{\mathcal{B}}^{\text{cvx}}(\mathbf{0}) = \max\{0, \sup_{\beta\in\mathcal{B}} \sigma(\beta) - |\beta|\}$, which is equal to 0 since Proposition 3.4 states that $\sigma(\beta) < |\beta|$. The point $(\mathbb{1}, 1)$ belongs to $P_{\mathcal{B}}$ because $f^{\text{conc}}(\mathbb{1}) = 1$, and $f_{\mathcal{B}}^{\text{cvx}}(\mathbb{1}) = \max\{0, \sup_{\beta\in\mathcal{B}} \sigma(\beta)\}$, which is less than equal to 1 due to $\mathbb{1} \in S$. □

The next proof is that of the error bounds over a simplex.

Proof of Theorem 1.2. The concave envelope error bound is from Proposition 3.2 and the fact that $\text{cav}_{\Delta_n}[f] \leq f^{\text{conc}}$. The convex envelope error bound was observed in Proposition 3.3. To upper bound $\mu\left(\text{conv}\, \mathcal{G}_{\Delta_n}(f)\right)$, we note that

$$\frac{(\alpha^\alpha)^{1/d}}{d} - \frac{\alpha^\alpha}{d^d} \geq \frac{\alpha^\alpha}{d^d} \iff \left(\frac{\alpha^\alpha}{d^d}\right)^{\frac{d-1}{d}} \leq \frac{1}{2} \iff 2^{\frac{1}{d-1}} \leq \frac{d}{\alpha^{\alpha/d}}.$$

Denoting $\alpha_{(n)} = \max_i \alpha_i$, we have $\alpha^{\alpha/d} \leq \alpha_{(n)}^{\sum_i \alpha_i/d} = \alpha_{(n)}$. Thus it suffices to show that $d/\alpha_{(n)} \geq 2^{\frac{1}{d-1}}$, equivalently, $(d/\alpha_{(n)})^{d-1} \geq 2$. Since $\alpha_n \leq d-1$ due to $n \geq 2$,

$$\left(\frac{d}{\alpha_{(n)}}\right)^{d-1} \geq \left(\frac{d}{d-1}\right)^{d-1} = \left(1 + \frac{1}{d-1}\right)^{d-1} \geq 1 + \frac{d-1}{d-1} = 2,$$

where the last inequality is from binomial expansion. □

We end by mentioning that for $S = [\frac{1}{r}, 1]^n$, or equivalently for $S = [1, r]^n$ upto scaling, our upper bounds on the convex hull error are the same as those in Theorem 1.1 whereas a lower bound can be obtained by setting $t_1 = 1/r, t_2 = 1$ in Lemma 2.1. However these bounds are not tight, which is not all that surprising since we do not know the exact form of the envelopes of a general monomial over $[\frac{1}{r}, 1]^n$. In §4, we consider a multilinear monomial over $[1, r]^n$ and use the explicit characterization of its envelopes to derive tight error bounds. It so happens that in the multilinear case, the lower bound from Lemma 2.1 with $t_1 = 1, t_2 = r$ seems to be the convex hull error, a claim that is verified empirically for random r, n and shown to be true for every $r > 1$ as $n \to \infty$.

3.4 Comparison with another error bound

For the problem of optimizing $p \in \mathbb{R}[x]_m$ over $S = [0, 1]^n$: $z^*_{[0,1]^n} = \min\{p(x) \mid x \in [0, 1]^n\}$, De Klerk and Laurent [DKL10] present a LP and a SDP relaxation of $z^*_{[0,1]^n}$ based on two different positivstellensatz and also give a common error bound for these relaxations. Their bound is [DKL10, Theorem 1.4]:

$$z^*_{[0,1]^n} - \tilde{z}^\delta_{[0,1]^n} \leq \frac{L(p)}{\delta} \binom{m+1}{3} n^m,$$

where $\tilde{z}^\delta_{[0,1]^n}$ is either of their two relaxations, $\delta \geq m$ is an integer with $n\delta$ being a degree bound on polynomials in the positivstellensatz, and

$$L(p) = \max_\alpha |c_\alpha| \frac{\prod_j \alpha_j!}{|\alpha|!}.$$

As $\delta \to \infty$, the two relaxations converge to $z^*_{[0,1]^n}$ (the SDP relaxation has finite convergence). Corollary 1.2 states that the monomial convexification approach would yield a error bound, as per our analysis, of $z^*_{[0,1]^n} - z^{mono}_{[0,1]^n} \leq L'(p)\binom{n+m}{n}$ for $L'(p)$ defined in (3). This bound was weakened subsequently in Corollary 1.3 for ease of computation.

We note that for the LP and SDP relaxations to provide a better worst case guarantee, the degrees of the polynomials considered in the respective positivstellensatz must grow cubic in the degree of $p(x)$.

Proposition 3.7. *For $p \in \mathbb{R}[x]_m$ with $c_\alpha = 0, \pm 1$, and fixed n, the worst case error bound from $\tilde{z}^\delta_{[0,1]^n}$ is better than the worst case error bound from $z^{mono}_{[0,1]^n}$ only if $\delta = \Omega(m^3)$.*

Proof. The assumption $c_\alpha = 0, \pm 1$ implies $L(p) = \max_{\alpha:\,|c_\alpha|=1} \frac{\prod_j \alpha_j!}{|\alpha|!}$, which is lower bounded by $\frac{1}{m!}$. We have $L'(p) \leq \mathcal{C}^1_m = (1 - \frac{1}{m})m^{\frac{1}{(1-m)}}$ from Corollary 1.3. Therefore, the LP and SDP relaxations of [DKL10] give better error bounds than monomial convexification only if

$$\delta \geq \hat{\delta} := \frac{\binom{m+1}{3} n^m}{m! \mathcal{C}^1_m \binom{n+m}{n}} = \frac{m^2(m+1)}{6m^{\frac{1}{1-m}}(1+\frac{m}{n})\cdots(1+\frac{1}{n})} = \Omega(m^3) \text{ for fixed } n. \qquad \square$$

4 Multilinear monomial

Here we consider a multilinear monomial $m(x) = \prod_{j=1}^n x_j$ over either a box with constant ratio or a symmetric box. Since these boxes are simple scalings of $[1, r]^n$ and $[-1, 1]^n$, respectively, and our error measure $\mu(\cdot)$ scales as noted in Observation 2.2, we henceforth restrict our attention to only $[1, r]^n$ and $[-1, 1]^n$. As in §3, the convex hull error is computed by bounding the convex and concave envelope errors separately.

4.1 Box with constant ratio

Proposition 4.1 ([Ben04; TRX13])**.**

$$\mathrm{cav}_{[1,r]^n}[m](x) \;=\; \left[\min_{\sigma\in\Sigma_n} \sum_{j=1}^n r^{j-1} x_{\sigma(j)}\right] \;-\; \sum_{j=1}^{n-1} r^j,$$

where Σ_n is the set of all permutations of $\{1, \ldots, n\}$, and

$$\mathrm{vex}_{[1,r]^n}[m](x) \;=\; \max_{i=1,\ldots,n} r^{i-1}\left(\sum_{j=1}^n x_j \;-\; (n-i) \;-\; r(i-1)\right).$$

Proof. To obtain $\mathrm{cav}_{[1,r]^n}[m]$, we simply substitute $l_j = 1, u_j = r$ in [Ben04, Theorem 1] which states $\mathrm{cav}_{[l,u]}[m]$ for arbitrary l, u with $l \geq \mathbf{0}$. The convex envelope can be derived from [TRX13, Theorem 4.6]. This theorem gives a piecewise linear function with n pieces as the convex envelope of a function $g(y)\colon [0,1]^n \mapsto \mathbb{R}$ when g is convex-extendable from $\{0,1\}^n$ and there exists a convex function $\rho\colon \mathbb{R}_+ \mapsto \mathbb{R}$ such that $g(y) = \rho(\sum_{j=1}^n y_j)$ for every $y \in \{0,1\}^n$. Consider $\prod_{j=1}^n x_j$. Writing $x_j = 1 + (r-1)y_j$, the multilinear term becomes $g(y) = \prod_{j=1}^n (1 + (r-1)y_j)$ for $y \in [0,1]^n$. Since this $g(y)$ is a multilinear function of y, it is convex-extendable from $\{0,1\}^n$. Furthermore, for $y \in \{0,1\}^n$, $g(y) = r^{\sum_{j=1}^n y_j}$, and obviously $r^{(\cdot)}$ is convex over $\mathbb{R}_+$. Therefore, the convex envelope formula follows from [TRX13, Theorem 4.6]. $\square$

Applying a straightforward scaling argument, similar to the one used for the $[0,1]^n$ box at the beginning of §3, gives us the convex hull of $\mathcal{G}(m)$ when $l_i u_i > 0$ for all i and for some $r > 0$, $u_i/l_i = r$ for all i with $l_i > 0$ and $l_i/u_i = r$ for all i with $u_i < 0$. We omit the details.

Before proving Theorem 1.3 which claims that $\mathcal{D}_{r,n}$ and $\mathcal{E}_{r,n}$ are the maximum envelope errors for $\prod_{j=1}^n x_j$ over $[1,r]^n$, we provide some background on these two constants. The value $\mathcal{E}_{r,n}$ is obtained by applying Lemma 2.1: set $t_1 = 1, t_2 = r$ to get $\xi' = (\frac{r^n-1}{n})^{\frac{1}{(n-1)}}(r-1)^{\frac{n}{(1-n)}} - \frac{1}{(r-1)}$ and $\phi(\xi')$, upon simplification, becomes equal to $\mathcal{E}_{r,n}$. There is no simple explicit closed form formula for $\mathcal{D}_{r,n}$. However, $\mathcal{D}_{r,n}$ can be bounded as follows. After replacing $t = i/n$, the formula for $\mathcal{D}_{r,n}$ requires solving an integer program:

$$\mathcal{D}_{r,n} = \max\{\psi(t) \mid t \in \{1/n, 2/n, \ldots, 1\}\}, \qquad \text{where } \psi(t) = (1+(r-1)t)^n - r^{nt}.$$

Note that ψ is a difference of two convex increasing functions ψ_1 and ψ_2. After separating the maximizations over ψ_1 and ψ_2, we obtain the trivial upper bound $\mathcal{D}_{r,n} \le r^n - r$. But this bound can be very weak. A tighter bound can be derived by considering the continuous relaxation of the problem:

$$\mathcal{D}_{r,n} \le \max\{\psi(t) \mid t \in [0,1]\}.$$

Since ψ is differentiable with $\psi(0) = \psi(1) = 0$, by Rolle's theorem, there exists at least one stationary point of ψ in $[0,1]$. Based on these stationary points, we can say the following.

Proposition 4.2. *Let $t^* = \min\{t \in [0,1] \mid \psi'(t) = 0\}$ be the smallest stationary point of ψ on $[0,1]$, and t^{**} be the global maxima of ψ on $[0,1]$. If $t^* \ge \frac{n-1}{n}$, then $\mathcal{D}_{r,n} = (1+\frac{n-1}{n}(r-1))^n - r^{n-1}$, otherwise if $t^{**} \le \frac{n-1}{n}$, then $\mathcal{D}_{r,n} \le r^n \left(\frac{\ln r}{r-1}\right)^{\frac{n}{n-1}} - r^{n-1}$, otherwise $\mathcal{D}_{r,n} \le r^{\frac{n^2}{n-1}} \left(\frac{\ln r}{r-1}\right)^{\frac{n}{n-1}} - r^n$.*

The proof is in Appendix A. Obviously, $t^* \le t^{**}$. We conjecture that $t^* = t^{**}$.

We now prove our main result in this section.

4.1.1 Proof of Theorem 1.3

Proof. We only prove the maximum errors for the envelopes, the formula for $\mu\left(\operatorname{conv} \mathcal{G}_{[1,r]^n}(m)\right)$ follows subsequently from Observation 2.1. Consider the concave envelope first. We noted earlier that the value $\mathcal{E}_{r,n}$ comes from applying Lemma 2.1 with $t_1 = 1, t_2 = r$. Hence to prove that the maximum concave envelope error is equal to $\mathcal{E}_{r,n}$, it suffices to argue that there exists a point in $\operatorname{rel.int}\{\mathbb{1}, r\mathbb{1}\}$ which maximizes this error. Suppose, for sake of contradiction, that this is not the case. Since $\operatorname{cav}_{[1,r]^n}[m](\mathbb{1}) = m(\mathbb{1})$ and $\operatorname{cav}_{[1,r]^n}[m](r\mathbb{1}) = m(r\mathbb{1})$, we know that these two points do not maximize the error. Then our assumption means that for every maximizer x^* there exists some index i such that $x^*_{(i)} < x^*_{(i+1)}$, where $(\cdot)$ is the permutation that permutes variables as $x^*_{(1)} \le x^*_{(2)} \le \cdots \le x^*_{(n)}$. Since $r > 1$, for every $x \in [1,r]^n$, the minimum over Σ_n in the expression for $\operatorname{cav}_{[1,r]^n}[m]$, which is given in Proposition 4.1, occurs at a permutation σ such that $x_{\sigma(1)} \ge x_{\sigma(2)} \ge \cdots \ge x_{\sigma(n)}$. Therefore, $\operatorname{cav}_{[1,r]^n}[m](x) = \sum_{j=1}^n r^{n-j} x_{(j)} - \sum_{j=1}^{n-1} r^j$. In particular, $\operatorname{cav}_{[1,r]^n}[m](x^*) = \sum_{j=1}^n r^{n-j} x^*_{(j)} - \sum_{j=1}^{n-1} r^j$, and the maximum error is $z^* = \sum_{j=1}^n r^{n-j} x^*_{(j)} - \prod_{j=1}^n x^*_j - \sum_{j=1}^{n-1} r^j$. Now consider two points $\hat{x}$ and $\tilde{x}$ obtained from x^* by setting, respectively, $\hat{x}_{(i)} = x^*_{(i+1)}$ and $\tilde{x}_{(i+1)} = x^*_{(i)}$. Since the error at these points cannot be larger than z^*, we have $r^{n-i}(x^*_{(i+1)} - x^*_{(i)}) \le (x^*_{(i+1)} - x^*_{(i)}) \prod_{j \ne i} x_{(j)}$ and $r^{(n-i-1)}(x^*_{(i)} - x^*_{(i+1)}) \le (x^*_{(i)} - x^*_{(i+1)}) \prod_{j \ne i+1} x_{(j)}$, and consequently, $r^{n-i} - \prod_{j \ne i} x^*_{(j)} \le 0$ and $r^{n-i-1} - \prod_{j \ne i+1} x^*_{(j)} \ge 0$. Hence

$$r^{n-i} \;\le\; r^{n-i} x^*_{(i)} \;\le\; \prod_{j=1}^n x^*_j \;\le\; r^{n-i-1} x^*_{(i+1)} \;\le\; r^{n-i}.$$

Equality holds in above if and only if $x^*_{(i)} = 1$ and $x^*_{(i+1)} = r$. Therefore $x^*_{(1)} = \cdots = x^*_{(i)} = 1$, $x^*_{(i+1)} = \cdots = x^*_{(n)} = r$, but at such a point, the error is zero due to

$$\operatorname{cav}_{[1,r]^n}[m](x^*) \;=\; \sum_{j=1}^{i} r^{n-j} + \sum_{j=i+1}^{n} r^{n+1-j} - \sum_{j=1}^{n-1} r^j \;=\; r^{n-i} \;=\; m(x^*).$$

Thus we have reached a contradiction to x^* being a maximizer. Hence it must be that the error is maximized on $\operatorname{rel.int}\{\mathbb{1}, r\mathbb{1}\}$.

Now consider the convex envelope. We follow similar steps as in the proof of Theorem 3.2.

$$\begin{aligned}
\prod_{j=1}^{n} x_j \;-\; \operatorname{vex}_{[1,r]^n}[m](x) \;&=\; \prod_{j=1}^{n} x_j \;-\; \max_{i=1,\ldots,n} r^{i-1}\left(\sum_{j=1}^{n} x_j - (n-i) - r(i-1)\right) \\
&=\; \min_{i=1,\ldots,n}\left\{\prod_{j=1}^{n} x_j - r^{i-1}\left(\sum_{j=1}^{n} x_j - (n-i) - r(i-1)\right)\right\} \\
&\le\; \min_{i=1,\ldots,n}\left\{\prod_{j=1}^{n} x_j - r^{i-1}\left(n\sqrt[n]{\prod_{j=1}^{n} x_j} - (n-i) - r(i-1)\right)\right\},
\end{aligned}$$

where we employ the arithmetic-geometric means inequality. By regarding $\sqrt[n]{\prod_{j=1}^n x_j}$ as a scalar variable t, we get

$$\max_{x\in[1,r]^n} \prod_{j=1}^{n} x_j \;-\; \operatorname{vex}_{[1,r]^n}[m](x) \;\le\; \max_{t\in[1,r]} \min_{i=1,\ldots,n} \widehat{\varphi}_i(t),$$

where $\widehat{\varphi}_i \colon t \mapsto t^n - nr^{i-1}t + r^{i-1}[n - r + i(r-1)]$ is a convex function on $[1, r]$. Therefore we have to find the maximum value of the pointwise minimum function $\min_i \widehat{\varphi}_i(t)$ on the interval $[1, r]$ and it is apparent that this maximum value is attained at a breakpoint of the function, i.e., at a t^* such that $\widehat{\varphi}_i(t^*) = \widehat{\varphi}_{i+1}(t^*)$ for some $1 \le i \le n-1$. For any $1 \le i \le n-1$, solving for t in $\widehat{\varphi}_i(t) = \widehat{\varphi}_{i+1}(t)$ means that we must find t satisfying $t^n - r^{i-1}nt + r^{i-1}[n-r+i(r-1)] = t^n - r^i nt + r^i[n-i-1+ir]$, which upon canceling and rearranging terms leads to $r^{i-1}(r-1)nt = -r^{i-1}(n-i) - 2ir^i + nr^i + ir^{i+1}$. Therefore $(r-1)nt = -(n-i) - 2ir + nr + ir^2 = n(r-1) + i(r-1)^2$ and hence, $t = 1 + i(r-1)/n$. Substituting this breakpoint t into $\widehat{\varphi}_i$ yields

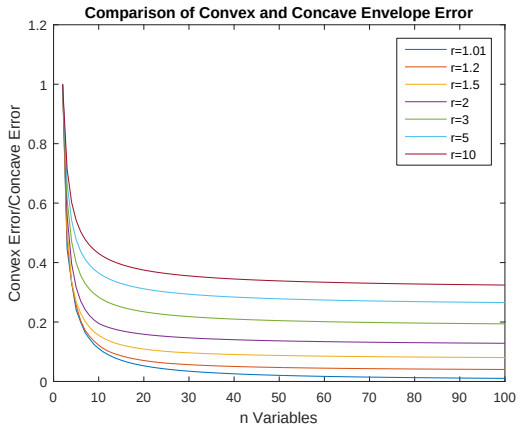
$$\widehat{\varphi}_i(t) = \left(1 + \frac{i}{n}(r-1)\right)^n - nr^{i-1}\left(1 + \frac{i}{n}(r-1)\right) + r^{i-1}[n - r + i(r-1)] = \left(1 + \frac{i}{n}(r-1)\right)^n - r^i.$$

The maximum, with respect to $i = 1, \ldots, n-1$, over all such values is the maximum of $\min_i \widehat{\varphi}_i(t)$ and notice that this maximum over i is exactly the constant $\mathcal{D}_{r,n}$. Hence $\mathcal{D}_{r,n}$ is a upper bound on the convex envelope error. This bound is tight because the means inequality is an equality when all the x_j's are equal to each other, and hence this error is attained on $\operatorname{rel.int}\{\mathbb{1}, r\mathbb{1}\}$. □

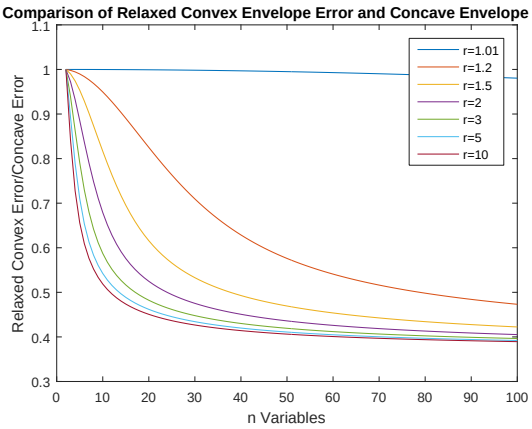
4.1.2 Comparing $\mathcal{D}_{r,n}$ and $\mathcal{E}_{r,n}$

We conjecture that $\mathcal{D}_{r,n} \le \mathcal{E}_{r,n}$ for every r, n, which would imply that the convex hull error is equal to $\mathcal{E}_{r,n}$. Although we were unable to prove this in general due to the extremely complicated forms for $\mathcal{E}_{r,n}$, and more specifically, for $\mathcal{D}_{r,n}$, we ran some simulations, graphed in Figure 1, to

support our claim. For every $n = 2, \ldots, 100$ and $r \in \{1.01, 1.2, 1.5, 2, 3, 5, 10\}$, we computed the ratio $\mathcal{D}_{r,n}/\mathcal{E}_{r,n}$ and plotted it in Figure 1(a). We also plotted the ratio between the error of the relaxed convex envelope (the relaxation is obtained by taking the maximum over $i \in \{1, n\}$ in the expression for $\mathrm{vex}_{[1,r]^n}[m]$) and $\mathcal{E}_{r,n}$, see Figure 1(b). As can be seen in these figures, the ratios are never larger than 1, thereby establishing a strong empirical basis in support of our conjecture that the error from the concave envelope dominates that from the convex envelope, and possibly even from the relaxed convex envelope.



(a) Concave and Convex envelopes



(b) Concave and Relaxed Convex envelopes

Figure 1: Error comparisons for $\prod_{j=1}^{n} x_j$ over $[1, r]^n$.

Asymptotically, $\mathcal{E}_{r,n}$ dominates $\mathcal{D}_{r,n}$ in the following sense. Recall t^* and t^{**} defined in Proposition 4.2.

Proposition 4.3. $\lim_{n\to\infty} \frac{\mathcal{E}_{r,n}}{r^n - 1} = 1$, *and* $\lim_{n\to\infty} \frac{\mathcal{D}_{r,n}}{r^n - 1} \leq \frac{1}{e}$ *if* $t^* \geq (n-1)/n$ *or* $t^{**} \leq (n-1)/n$.

Proof. We have

$$\lim_{n\to\infty} \frac{\mathcal{E}_{r,n}}{r^n - 1} = \frac{1}{r-1}\left[\lim_{n\to\infty} \frac{n-1}{n} \lim_{n\to\infty} \left(\frac{r^n - 1}{n(r-1)}\right)^{\frac{1}{n-1}} - 1\right] = \frac{1}{r-1}\left[1 \cdot r - 1\right] = 1.$$

Proposition 4.2 gives two bounds on $\mathcal{D}_{r,n}$. If $\mathcal{D}_{r,n} = (1 + \frac{n-1}{n}(r-1))^n - r^{n-1}$, then

$$\begin{aligned}
\lim_{n\to\infty} \frac{\mathcal{D}_{r,n}}{r^n - 1} = \lim_{n\to\infty} \frac{\mathcal{D}_{r,n}}{r^n} \lim_{n\to\infty} \frac{r^n}{r^n - 1} &= \lim_{n\to\infty} \left(\frac{1}{r} + \left(1 - \frac{1}{n}\right)\left(1 - \frac{1}{r}\right)\right)^n - \frac{1}{r} \\
&= \lim_{n\to\infty} \left(1 - \frac{1}{n} + \frac{1}{nr}\right)^n - \frac{1}{r} \\
&= e^{\frac{1}{r}-1} - \frac{1}{r}.
\end{aligned}$$

The above function of r is increasing over $[1, \infty)$ and converges to $1/e \approx 0.37$ as $r \to \infty$. The limit on the other value of $\mathcal{D}_{r,n}$ is $\frac{\ln r}{r-1} - \frac{1}{r}$ as $n \to \infty$, and the value of this function of r never exceeds 0.22. ∎

Thus $\mathcal{E}_{r,n}$ seems to grow much more rapidly than $\mathcal{D}_{r,n}$ in some cases.

4.2 Symmetric box

4.2.1 Convex hull

Luedtke, Namazifar, and Linderoth [LNL12] showed that the recursive McCormick relaxation, which Ryoo and Sahinidis [RS01] had used to obtain an extended formulation of $\operatorname{conv} \mathcal{G}_{[0,1]^n}(m)$, yields a compact extended formulation of $\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m)$. However, to the best of our knowledge, there is no known characterization of this convex hull in the (x, w)-space. We provide this next. A different proof based on constructive arguments is presented in a companion paper [AGX17].

Theorem 4.1. *Partition subsets of* $\{1, \dots, n\}$ *into* $\mathcal{N}^{even} := \{I \subseteq \{1, \dots, n\} \mid |I| \text{ is even}\}$ *and* $\mathcal{N}^{odd} := \{I \subseteq \{1, \dots, n\} \mid |I| \text{ is odd}\}$. *If* n *is odd, then*

$$\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m) = \Big\{(x, w) \in [-1, 1]^{n+1} \mid -(n-1) \leq \sum_{i \in I} x_i - \sum_{i \notin I} x_i + w \leq n - 1, \;\; I \in \mathcal{N}^{even}\Big\}.$$

If n *is even, then*

$$\begin{aligned}\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m) = \Big\{(x, w) \in [-1, 1]^{n+1} \mid & \sum_{i \in I} x_i - \sum_{i \notin I} x_i + w \leq n - 1, \;\; I \in \mathcal{N}^{odd} \\ & \sum_{i \in I} x_i - \sum_{i \notin I} x_i + w \geq -(n-1), \;\; I \in \mathcal{N}^{even}\Big\}.\end{aligned}$$

Before presenting our proof, we provide an intuition behind the proposed convex hull description. Denote $x_{n+1} = w$ to get $\mathcal{G}_{[-1,1]^n}(m) = \{x \in [-1, 1]^{n+1} \mid x_{n+1} = \prod_{j=1}^{n} x_j\}$. It is well-known [Rik97; She97] that for any box $[l, u]$, the extreme points of $\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m)$ are in bijection with the extreme points of $[l, u]$ (this is also true for a multilinear polynomial). Hence the set of extreme points of $\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m)$ is equal to $\mathcal{G}_{[-1,1]^n}(m) \cap \{-1, 1\}^{n+1}$. A point in $\{-1, 1\}^{n+1}$ violates $x_{n+1} = \prod_{j=1}^{n} x_j$ if and only if the set $\{i \in \{1, \dots, n+1\} \mid x_i = -1\}$ has odd cardinality. Every such inadmissible point in $\{-1, 1\}^{n+1}$ can be cut off using the "no-good" inequality

$$\sum_{i \in I} (x_i - (-1)) + \sum_{i \in \{1,\dots,n+1\} \setminus I} (1 - x_i) \geq 2$$

for some odd subset $I \subseteq \{1, \dots, n+1\}$. The no-good cut for subset I is valid to every point in $\{-1, 1\}^{n+1}$, except that point which takes the value -1 at exactly those elements indexed by I. This cut rearranges to

$$\sum_{i \in I} x_i - \sum_{\{1,\dots,n+1\} \setminus I} x_i \geq -(n-1). \tag{17}$$

Hence $\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m) = \operatorname{conv}\{x \in \{-1, 1\}^{n+1} \mid \text{eq. (17)} \;\; \forall I \subseteq \{1, \dots, n+1\}, \text{ odd } |I|\}$. Consider the polytope

$$P^{-1,1} := \{x \in [-1, 1]^{n+1} \mid \text{eq. (17)} \;\; \forall I \subseteq \{1, \dots, n+1\}, \text{ odd } |I|\}, \tag{18}$$

which is the LP relaxation of $\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m)$. By construction, this polytope has the property that $P^{-1,1} \cap \{-1, 1\}^{n+1} \subseteq \mathcal{G}_{[-1,1]^n}(m)$. We will show in the proof of Theorem 4.1 that the extreme points of $P^{-1,1}$ are in $\{-1, 1\}^{n+1}$, thereby implying that $\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m) = P^{-1,1}$. This equality, along with the following claim that is straightforward to verify, gives us the statement of Theorem 4.1.

Observation 4.1. *After denoting* $x_{n+1} = w$, *each of the convex hull descriptions in Theorem 4.1 becomes equal to the polytope* $P^{-1,1}$.

Proof of Theorem 4.1. We show that for any $c \in \mathbb{R}^{n+1}$, the linear program $z^{LP} = \max\{c^\top x \colon x \in P^{-1,1}\}$ has an optimal solution in $\mathcal{G}_{[-1,1]^n}(m) \cap \{-1,1\}^{n+1}$. We proceed by considering cases that are defined using $A = \{i : c_i = 0\}, B = \{i : c_i < 0\}, C = \{i : c_i > 0\}$. Note two things: (1) $z^{LP} \le \sum_{i \in B \cup C} |c_i|$ due to $x \in [-1,1]^{n+1}$ for every feasible x, (2) any $x \in \{-1,1\}^{n+1}$ belongs to $P^{-1,1}$ if and only if $\{i \in \{1,\dots,n+1\} \mid x_i = -1\}$ has even cardinality.

Case 1: $|B|$ is even. Since B has even cardinality, the point x^* with $x^*_i = -1$ for $i \in B$ and $x^*_i = 1$ for $i \in A \cup C$ belongs to $P^{-1,1}$. This x^* is optimal to z^{LP} because $c^\top x^* = \sum_{i \in B \cup C} |c_i|$.

Case 2: $|B|$ is odd and $|A| \ge 1$. Choose an arbitrary $j_0 \in A$ and set $x^*_i = -1$ for $i \in B \cup \{j_0\}$ and $x^*_i = 1$ for $i \in (A \setminus \{j_0\}) \cup C$. This x^* belongs to $P^{-1,1}$ because $B \cup \{j_0\}$ is even and is optimal to z^{LP} because $c^\top x^* = \sum_{i \in B \cup C} |c_i|$.

Case 3: $|B|$ is odd and $|A| = 0$. Let $j_1 \in \arg\min_{1 \le i \le n+1} |c_i|$. There are two subcases. When $j_1 \in B$, i.e. $c_{j_1} < 0$, the point $x^*_i = -1$ for $i \in B \setminus \{j_1\}$ and $x^*_i = 1$ for $i \in C \cup \{j_1\}$ is optimal with value $\sum_{i \in B \setminus \{j_1\}} (-c_i) + \sum_{i \in C \cup \{j_1\}} c_i$, because in this subcase

$$\begin{aligned} c^\top x &= \left(-c_{j_1}\right) \left(\sum_{i \in C} x_i - \sum_{i \in B} x_i \right) + \sum_{i \in C} \left(c_i + c_{j_1}\right) x_i + \sum_{i \in B} \left(c_i - c_{j_1}\right) x_i \\ &\le \left(-c_{j_1}\right)(n-1) + \sum_{i \in C} \left(c_i + c_{j_1}\right) + \sum_{i \in B} \left(c_{j_1} - c_i\right) \\ &= \sum_{i \in C} c_i + \sum_{i \in B} (-c_i) + 2c_{j_1} \\ &= \sum_{i \in B \setminus \{j_1\}} (-c_i) + \sum_{i \in C \cup \{j_1\}} c_i, \end{aligned}$$

where the $\le$ inequality is obtained by applying (17) with $I = B$. When $j_1 \in C$, i.e. $c_{j_1} > 0$, then the point $x^*_i = -1$ for $i \in B \cup \{j_1\}$ and $x^*_i = 1$ for $i \in C \setminus \{j_1\}$ is optimal with value $\sum_{i \in B \cup \{j_1\}} (-c_i) + \sum_{i \in C \setminus \{j_1\}} c_i$, because in this subcase

$$\begin{aligned} c^\top x &= c_{j_1} \left(\sum_{i \in C} x_i - \sum_{i \in B} x_i \right) + \sum_{i \in C} \left(c_i - c_{j_1}\right) x_i + \sum_{i \in B} \left(c_i + c_{j_1}\right) x_i \\ &\le c_{j_1}(n-1) + \sum_{i \in C} \left(c_i - c_{j_1}\right) + \sum_{i \in B} \left(-c_i - c_{j_1}\right) \\ &= \sum_{i \in C} c_i + \sum_{i \in B} (-c_i) - 2c_{j_1} \\ &= \sum_{i \in B \cup \{j_1\}} (-c_i) + \sum_{i \in C \setminus \{j_1\}} c_i. \end{aligned}$$

This completes our proof for showing that $P^{-1,1}$ has extreme points in $\{-1,1\}^{n+1}$. □

A scaling argument yields $\operatorname{conv} \mathcal{G}(m)$ when $l_i = -u_i$ for all i.

4.2.2 Errors

In order to prove Theorem 1.4, we make use of the reflection symmetry in the sets $\mathcal{G}_{[-1,1]^n}(m)$ and $\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)$, as described next. Let $\operatorname{sgn}(\cdot)$ denote the sign of a scalar, with $\operatorname{sgn}(0)$ considered positive. A point $(x,w)\in[-1,1]^{n+1}$ is said to have compatible signs if $\operatorname{sgn}(w)=\operatorname{sgn}(\prod_{j=1}^n x_j)$, i.e., $\operatorname{sgn}(w)$ is negative if and only if x has no zero entries and has an odd number of negative entries. Define the following binary relation on $[-1,1]^{n+1}$: $(x,w)\sim(x',w')$ if (i) $|w'|=|w|$ and $|x_j|=|x'_j|$ for all j, and (ii) both (x,w) and (x',w') have compatible signs or both (x,w) and (x',w') do not have compatible signs. Thus $(x,w)\sim(x',w')$ if and only if x' is obtained from x by reversing signs on odd (even) many entries of x and setting $w'=-w$ ($w'=w$). This binary relation has two important properties.

1. It preserves the error measure $h(x,w):=\left|w-\prod_{j=1}^n x_j\right|$. Indeed, one can easily argue that $h(x,w)=h(x',w')$ if $(x,w)\sim(x',w')$.
2. It is an equivalence relation, i.e., a reflexive symmetric transitive relation. This is obvious by construction of $\sim$.

Now consider $[(x,w)]:=\{(x',w')\in[-1,1]^n \mid (x,w)\sim(x',w')\}$, the equivalence class of (x,w) induced by $\sim$. Since $\sim$ is an equivalence relation on $[-1,1]^{n+1}$ and $\mathcal{G}_{[-1,1]^n}(m)$ and $\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)$ are subsets of $[-1,1]^{n+1}$, each of these sets is partitioned by $\sim$. Observe that the definition of $\sim$ means that for every $(x',w')\in[-1,1]^{n+1}$ having compatible (incompatible) signs, there exists $(x,w)\in[-1,1]^{n+1}$ such that $(x',w')\sim(x,w)$ and $(x,w)\geq\mathbf{0}$ ($x\geq\mathbf{0}, w<0$). Now, because every point in $\mathcal{G}_{[-1,1]^n}(m)$ has compatible signs and $(x,w)\in\mathcal{G}_{[-1,1]^n}(m)$ trivially implies $[(x,w)]\subset\mathcal{G}_{[-1,1]^n}(m)$, we have

$$\mathcal{G}_{[-1,1]^n}(m)=\bigcup\left\{[(x,w)]\colon (x,w)\in\mathcal{G}_{[-1,1]^n}(m),(x,w)\geq\mathbf{0}\right\}. \tag{19a}$$

To make a similar statement for $\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)$, we need a small modification because the convex hull contains points with both compatible and incompatible signs. In particular, we must drop the nonnegativity requirement on w. Also, if $(x,w)\in\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)$, then using the fact that (x,w) is a convex combination of points in $\mathcal{G}_{[-1,1]^n}(m)$, all of which have compatible signs, we get that $[(x,w)]\subset\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)$. Thus we have the following:

$$\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)=\bigcup\left\{[(x,w)]\colon (x,w)\in\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m),x\geq\mathbf{0}\right\} \tag{19b}$$

Now, the fact that $\sim$ is error-preserving leads to

$$\mu\left(\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)\right)=\max\left\{\left|w-\prod_{j=1}^n x_j\right|\colon (x,w)\in\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m),x\geq\mathbf{0}\right\}, \tag{19c}$$

meaning that we only need to consider nonnegative values of x when computing the convex hull error.

Proof of Theorem 1.4. To upper bound the convex hull error. We only present arguments for when n is odd, since the even case is almost exactly the same due to similar characterizations of the convex hulls in Theorem 4.1. By equation (19c), we consider only $(x,w)\in\operatorname{conv}\mathcal{G}_{[-1,1]^n}(m)$

with $x \geq \mathbf{0}$. Thus, $\mu\left(\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m)\right)$ is equal to the maximum of the maximum errors of $\operatorname{vex}_{[-1,1]^n}[m](x)$ and $\operatorname{cav}_{[-1,1]^n}[m](x)$ calculated over $[0,1]^n$.

$$\begin{aligned}
\prod_{j=1}^{n} x_j \;-\; \operatorname{vex}_{[-1,1]^n}[m](x) \;&=\; \prod_{j=1}^{n} x_j \;-\; \max\left\{-(n-1) \;+\; \max_{I \in \mathcal{N}^{even}} \sum_{j \notin I} x_j - \sum_{j \in I} x_j,\; -1\right\} \\
&=\; \prod_{j=1}^{n} x_j \;-\; \max\left\{\sum_{j=1}^{n} x_j - (n-1),\; -1\right\} \\
&=\; \min\left\{\prod_{j=1}^{n} x_j - \sum_{j=1}^{n} x_j + (n-1),\; \prod_{j=1}^{n} x_j + 1\right\} \\
&\leq\; \min\left\{\prod_{j=1}^{n} x_j - n\sqrt[n]{\prod_{j=1}^{n} x_j} + (n-1),\; \prod_{j=1}^{n} x_j + 1\right\},
\end{aligned}$$

where $x \geq \mathbf{0}$ has given us the second equality, and the inequality in the last step from applying the arithmetic-geometric means inequality. Therefore, after regarding $\sqrt[n]{\prod_{j=1}^{n} x_j}$ as a scalar variable t, we get $\max_{t \in [0,1]} \min \varphi(t)$ to be an upper bound on the convex envelope error, where $\varphi(t) = \min\{t^n - nt + n - 1, t^n + 1\}$. The function $t^n - nt + n - 1$ is convex decreasing on $[0,1]$ whereas $t^n + 1$ is convex increasing on $[0,1]$, and hence the maximum value of φ on $[0,1]$ occurs at a breakpoint where the two functions have equal value. Solving for $t^n - nt + n - 1 = t^n + 1$ yields $t = 1 - 2/n$, and so the upper bound is $1 + (1-2/n)^n$. This bound is tight since it is attained at $x = (1-2/n)\mathbb{1}$ where $\operatorname{vex}_{[-1,1]^n}[m]((1-2/n)\mathbb{1}) = -1$. On the concave side, we have $\operatorname{cav}_{[-1,1]^n}[m](x) \leq 1$ and since $\prod_{j=1}^{n} x_j \geq 0$ for $x \geq \mathbf{0}$, the concave envelope error is upper bounded by 1. Thus, $\mu\left(\operatorname{conv} \mathcal{G}_{[-1,1]^n}(m)\right) = \max\{1 + (1-2/n)^n, 1\} = 1 + (1-2/n)^n$.

To find the points where this bound is attained, we already observed the point $(x, w) = ((1 - 2/n)\mathbb{1}, -1)$. Since our relation $\sim$ is error-preserving, all points in the equivalence class of $((1 - 2/n)\mathbb{1}, -1)$ have the same error, and there are 2^n many such points. Finally, note that for any point $(x', 1) \in [((1 - 2/n)\mathbb{1}, -1)]$, the above bounds on the envelopes would be reversed so that both the envelopes have the same maximum error over the entire $[-1,1]^n$ box. □

Acknowledgements

The first author was supported in part by ONR grant N00014-16-1-2168. The second author was supported in part by ONR grant N00014-16-1-2725. We thank two referees whose meticulous reading helped us clarify some of the technical details.

Appendix A Missing Proofs

Proof of Proposition 3.4. Since $S \subseteq [0,1]^n$ and $\beta \geq \mathbb{1}$ make $x^\alpha \leq \min_j x_j < \beta^\top x$ for all $x \in S$, we have $\sigma(\beta) < |\beta|$. The lower bound of 0 comes from

$$\sigma(\beta) \;\geq\; |\beta| + \min_{x \in S} x^\alpha - \max_{x \in S} \beta^\top x \;\geq\; |\beta| + 0 - |\beta| = 0.$$

If $S = [0,1]^n$, then $\mathbb{1} \in S$ implies that $\min_{x\in S} x^\alpha - \beta^\top x \le 1 - |\beta|$ and so by (7c), we have $\sigma(\beta) \le 1$. For the fourth claim we have $\Delta_n^{\mathbf{0}} \cap \Delta_n^{\mathbb{1}} = \{x \in [0,1]^n \mid \sum_i x_i = n-1\} = \operatorname{conv}\{\mathbb{1} - \mathbf{e}_1, \ldots, \mathbb{1} - \mathbf{e}_n\}$. Denote this simplex by $\Delta_{n-1}^{\mathbb{1}}$. The assumption $\Delta_{n-1}^{\mathbb{1}} \subseteq S$ means that $\mathbb{1} - \mathbf{e}_i \in S$ for all i. Substituting this point into (7c) gives us $\sigma(\beta) \le \sum_{j=1}^n \beta_j + 0 - \sum_{j\neq i} \beta_j = \beta_i$ for all i. This leads to $\sigma(\beta) \le \min_j \beta_j$. Since $\Delta_{n-1}^{\mathbb{1}} \subseteq S \subseteq \Delta_n^{\mathbf{0}}$, $\max_{x\in\Delta_{n-1}^{\mathbb{1}}} \beta^\top x \le \max_{x\in S} \beta^\top x \le \max_{x\in\Delta_n^{\mathbf{0}}} \beta^\top x$. Note that $\Delta_n^{\mathbf{0}} = \operatorname{conv}(\{x \in \{0,1\}^n \mid \sum_i x_i \le n-2\} \cup \Delta_{n-1}^{\mathbb{1}})$. The positivity of β then makes it clear that $\max_{x\in\Delta_n^{\mathbf{0}}} \beta^\top x = \max_{x\in\Delta_{n-1}^{\mathbb{1}}} \beta^\top x$. Hence $\max_{x\in S} \beta^\top x = \max_{x\in\Delta_{n-1}^{\mathbb{1}}} \beta^\top x = \sum_{j=1}^{n-1} \beta_{(j)}$, where $\beta_{(1)} \ge \beta_{(2)} \ge \cdots \ge \beta_{(n)}$. Now,

$$\sigma(\beta) \;\ge\; \sum_{j=1}^n \beta_j \;+\; \min_{x\in S} x^\alpha \;-\; \max_{x\in S} \beta^\top x \;\ge\; \sum_{j=1}^n \beta_j \;+\; 0 \;-\; \sum_{j=1}^{n-1} \beta_{(j)} \;=\; \beta_{(n)} \;=\; \min_j \beta_j.$$

Since we have already argued $\sigma(\beta) \le \min_j \beta_j$, it follows that $\sigma(\beta) = \min_j \beta_j$. □

Proof of Lemma 3.4. For nontriviality, assume $\lambda_1 > 1$.

(1) The first derivative is $\phi'(\sigma) = -\lambda_1(1-\sigma)^{\lambda_1 - 1} + \lambda_2$. If $\lambda_2 \ge \lambda_1$, then $\phi'(0) \ge 0$ and $\phi'(\sigma) > 0$ for all $\sigma \in (0,1]$ and hence ϕ is strictly increasing over $(0,1)$ and $\phi(\sigma) > \phi(0) = 0$ for all $\sigma \in (0,1]$.

(2 & 3) Now assume $1 \le \lambda_2 < \lambda_1$. Set $\tilde\sigma = 1 - (\lambda_2/\lambda_1)^{\frac{1}{\lambda_1 - 1}}$ and realize that $\phi'(\tilde\sigma) = 0$ and $\tilde\sigma \in (0,1)$. Then we have $\phi'(\sigma) < 0$ for $\sigma \in (0,\tilde\sigma)$. Therefore ϕ is decreasing on $(0,\tilde\sigma]$, which implies $\phi(\sigma) < \phi(0) = 0$ for $\sigma \in (0,\tilde\sigma]$. Hence $\phi(\tilde\sigma) < 0$. The construction of $\tilde\sigma$ also implies $\phi'(\sigma) > 0$, and hence ϕ is increasing, for $\sigma \in (\tilde\sigma, 1]$. Since $\phi(1) \ge 0$, it follows that there is a unique real number σ^* in $(\tilde\sigma, 1]$ such that $\phi(\sigma^*) = 0$. Thus we have $\phi(\sigma) \le 0$ for $\sigma \in [0,\sigma^*]$ and $\phi(\sigma) > 0$ for $\sigma \in (\sigma^*, 1]$. If λ_1 is odd, the other root is obtained by applying Descartes' rule of signs as in the first claim.

(4) Take $\lambda \in (\lambda_2, \infty)$ and define $g(\sigma) := (1-\sigma)^{\lambda_1} + \lambda\sigma - 1$. If $\lambda \ge \lambda_1$, then the first claim in this lemma, with λ_2 replaced by λ, gives us $g(\sigma) \ge 0$. Now assume $\lambda < \lambda_1$. Applying the second claim in this lemma, after replacing λ_2 with λ, tells us there is a unique real σ^{**} that is a root of g in $(0,1]$. Now $g(\sigma^*) = \phi(\sigma^*) + (\lambda - \lambda_2)\sigma^* > 0$ because $\phi(\sigma^*) = 0, \lambda > \lambda_2, \sigma^* > 0$. Then the third claim in this lemma, with λ_2 replaced by λ, gives us $\sigma^* > \sigma^{**}$ and consequently, the proposed fourth claim.

For the final part, note that the roots of ϕ and its complemented polynomial $\phi'(t) := t^{\lambda_1} - \lambda_2 t + \lambda_2 - 1$ are in bijection under the relation $\sigma = 1 - t$. Descartes' rule of signs tells us that ϕ' has exactly one positive root besides $t = 1$. When $\lambda_2 > \lambda_1$, this root must be in $(1,\infty)$ because otherwise we would get a contradiction to ϕ not having any roots in $(0,1]$. Descartes' rule also tells us there is exactly one negative root when λ_1 is odd. This translates to ϕ having a root in $(1,\infty)$ if and only if λ_1 is odd. □

Proof of Proposition 4.2. Note that $\psi(0) = \psi(1) = 0$. We first claim that ψ is strictly increasing on $(0,t^*)$. In fact, we argue the stronger claim that $\psi(t) > 0$ for all $t \in (0,1)$. This claim is equivalent to showing that $(\frac{1+(r-1)t}{r^t})^n > 1$, which is equivalent to $r^t - (r-1)t - 1 < 0$. The function $t \mapsto r^t - (r-1)t - 1$ is convex and is zero-valued at $t = 0$ and $t = 1$. Therefore, by convexity, $r^t - (r-1)t - 1 < 0$ for all $t \in (0,1)$, and hence, we have $\psi(t) > 0$ for all $t \in (0,1)$.

Since $\mathcal{D}_{r,n} \le \max_{t\in[0,1]} \psi(t)$, ψ is strictly increasing on $(0,t^*)$, and $\psi(1) = 0$, the condition $t^* \ge (n-1)/n$ implies that $i = (n-1)$ yields the maximum value in the formula for $\mathcal{D}_{r,n}$. Now

suppose $t^{**} \leq (n-1)/n$. Since t^{**} is a stationary point, $(1+t^{**}(r-1))^{n-1} = r^{nt^{**}\frac{\ln r}{r-1}}$. Now,

$$0 \;<\; \mathcal{D}_{r,n} \;\leq\; (1+t^{**}(r-1))^n - r^{nt^{**}} \;=\; r^{\frac{n^2 t^{**}}{n-1}} \left(\frac{\ln r}{r-1}\right)^{\frac{n}{n-1}} - r^{nt^{**}} \;=\; r^{nt^{**}} \left(r^{\frac{nt^{**}}{n-1}} \left(\frac{\ln r}{r-1}\right)^{\frac{n}{n-1}} - 1 \right)$$
$$\leq r^{n-1} \left(r \left(\frac{\ln r}{r-1}\right)^{\frac{n}{n-1}} - 1 \right),$$

where the last inequality uses $nt^{**} \leq n-1$ and $r > 1$. Finally, if $t^* < (n-1)/n < t^{**}$, since t^{**} can be arbitrarily close to 1, we can only bound $r^{nt^{**}}$ and $r^{\frac{nt^{**}}{n-1}}$ in above by r^n and $r^{\frac{n}{n-1}}$, respectively, to obtain the last proposed bound on $\mathcal{D}_{r,n}$. □

References

[AGX17] Warren Adams, Akshay Gupte, and Yibo Xu. "An RLT approach for convexifying symmetric multilinear polynomials". *working paper*. 2017.

[AKF83] F.A. Al-Khayyal and J.E. Falk. "Jointly constrained biconvex programming". In: *Mathematics of Operations Research* 8.2 (1983), pp. 273–286.

[Bao+15] Xiaowei Bao, Aida Khajavirad, Nikolaos V Sahinidis, and Mohit Tawarmalani. "Global optimization of nonconvex problems with multilinear intermediates". In: *Mathematical Programming Computation* 7.1 (2015), pp. 1–37.

[Bel+09] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter. "Branching and bounds tightening techniques for non-convex MINLP". In: *Optimization Methods and Software* 24.4 (2009), pp. 597–634.

[BMN10] Pietro Belotti, Andrew J Miller, and Mahdi Namazifar. "Valid inequalities and convex hulls for multilinear functions". In: *Electronic Notes in Discrete Mathematics* 36 (2010), pp. 805–812.

[Ben04] Harold P Benson. "Concave envelopes of monomial functions over rectangles". In: *Naval Research Logistics (NRL)* 51.4 (2004), pp. 467–476.

[Bol+17] Natashia Boland, Santanu S Dey, Thomas Kalinowski, Marco Molinaro, and Fabian Rigterink. "Bounding the gap between the McCormick relaxation and the convex hull for bilinear functions". In: *Mathematical Programming* 162 (2017), pp. 523–535.

[BD17] Christoph Buchheim and Claudia D'Ambrosio. "Monomial-wise optimal separable underestimators for mixed-integer polynomial optimization". In: *Journal of Global Optimization* 67.4 (2017), pp. 759–786.

[BMW10] Christoph Buchheim, Dennis Michaels, and Robert Weismantel. "Integer programming subject to monomial constraints". In: *SIAM Journal on Optimization* 20.6 (2010), pp. 3297–3311.

[Cra93] Yves Crama. "Concave extensions for nonlinear 0–1 maximization problems". In: *Mathematical Programming* 61.1-3 (1993), pp. 53–60.

[CRH17] Yves Crama and Elisabeth Rodríguez-Heck. "A class of valid inequalities for multilinear 0–1 optimization problems". In: *Discrete Optimization* (2017).

[DS16] Evrim Dalkiran and Hanif D Sherali. "RLT-POS: Reformulation-Linearization Technique-based optimization software for solving polynomial programming problems". In: *Mathematical Programming Computation* (2016), pp. 1–39.

[DKL10] Etienne De Klerk and Monique Laurent. "Error bounds for some semidefinite programming approaches to polynomial minimization on the hypercube". In: *SIAM Journal on Optimization* 20.6 (2010), pp. 3104–3120.

[DKLS15] Etienne De Klerk, Monique Laurent, and Zhao Sun. "An error analysis for polynomial optimization over the simplex based on the multivariate hypergeometric distribution". In: *SIAM Journal on Optimization* 25.3 (2015), pp. 1498–1514.

[DKLS16] Etienne De Klerk, Monique Laurent, and Zhao Sun. "Convergence analysis for Lasserres measure-based hierarchy of upper bounds for polynomial optimization". In: *Mathematical Programming* (2016), pp. 1–30.

[DPK16] Alberto Del Pia and Aida Khajavirad. "A polyhedral study of binary polynomial programs". In: *Mathematics of Operations Research* (2016).

[DG15] Santanu S. Dey and Akshay Gupte. "Analysis of MILP techniques for the pooling problem". In: *Operations Research* 63.2 (2015), pp. 412–427.

[Las01] Jean B Lasserre. "Global optimization with polynomials and the problem of moments". In: *SIAM Journal on Optimization* 11.3 (2001), pp. 796–817.

[Las15] Jean Bernard Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Vol. 52. Cambridge University Press, 2015.

[Lau09] Monique Laurent. "Sums of squares, moment matrices and optimization over polynomials". In: *Emerging applications of algebraic geometry*. Springer, 2009, pp. 157–270.

[LP03] Leo Liberti and Constantinos C Pantelides. "Convex envelopes of monomials of odd degree". In: *Journal of Global Optimization* 25.2 (2003), pp. 157–168.

[Lin05] Jeff Linderoth. "A simplicial branch-and-bound algorithm for solving quadratically constrained quadratic programs". In: *Mathematical Programming* 103.2 (2005), pp. 251–282.

[Loc16] Marco Locatelli. "Polyhedral subdivisions and functional forms for the convex envelopes of bilinear, fractional and other bivariate functions over general polytopes". In: *Journal of Global Optimization* Online First (2016). DOI: 10.1007/s10898-016-0418-4.

[LS14] Marco Locatelli and Fabio Schoen. "On convex envelopes for bivariate functions over polytopes". In: *Mathematical Programming* 144.1-2 (2014), pp. 65–91.

[LNL12] J. Luedtke, M. Namazifar, and J. Linderoth. "Some Results on the Strength of Relaxations of Multilinear Functions". In: *Mathematical Programming* 136.2 (2012), pp. 325–351.

[McC76] G.P. McCormick. "Computability of global solutions to factorable nonconvex programs: Part I. Convex underestimating problems". In: *Mathematical Programming* 10.1 (1976), pp. 147–175.

[MF04] C.A. Meyer and C.A. Floudas. "Trilinear monomials with mixed sign domains: Facets of the convex and concave envelopes". In: *Journal of Global Optimization* 29.2 (2004), pp. 125–155.

[MF05] C.A. Meyer and C.A. Floudas. "Convex envelopes for edge-concave functions". In: *Mathematical Programming* 103.2 (2005), pp. 207–224.

[MF14] Ruth Misener and Christodoulos A Floudas. "ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations". In: *Journal of Global Optimization* 59.2-3 (2014), pp. 503–526.

[MSF15] Ruth Misener, James B Smadbeck, and Christodoulos A Floudas. "Dynamically generated cutting planes for mixed-integer quadratically constrained quadratic programs and their incorporation into GloMIQO 2". In: *Optimization Methods and Software* 30.1 (2015), pp. 215–249.

[Pan97] Jong-Shi Pang. "Error bounds in mathematical programming". In: *Mathematical Programming* 79.1-3 (1997), pp. 299–332.

[Rik97] A.D. Rikun. "A convex envelope formula for multilinear functions". In: *Journal of Global Optimization* 10.4 (1997), pp. 425–437.

[RS01] Hong Seo Ryoo and Nikolaos V Sahinidis. "Analysis of bounds for multilinear functions". In: *Journal of Global Optimization* 19.4 (2001), pp. 403–424.

[SDL12] Hanif D Sherali, Evrim Dalkiran, and Leo Liberti. "Reduced RLT representations for nonconvex polynomial programming problems". In: *Journal of Global Optimization* 52.3 (2012), pp. 447–469.

[She97] H.D. Sherali. "Convex envelopes of multilinear functions over a unit hypercube and over special discrete sets". In: *Acta Mathematica Vietnamica* 22.1 (1997), pp. 245–270.

[SL17] Emily Speakman and Jon Lee. "Quantifying Double McCormick". In: *Mathematics of Operations Research* 42.4 (2017), pp. 1230–1253.

[TS02] M. Tawarmalani and N.V. Sahinidis. "Convex extensions and envelopes of lower semi-continuous functions". In: *Mathematical Programming* 93.2 (2002), pp. 247–263.

[TS05] M. Tawarmalani and N.V. Sahinidis. "A polyhedral branch-and-cut approach to global optimization". In: *Mathematical Programming* 103.2 (2005), pp. 225–249.

[TRX13] Mohit Tawarmalani, Jean-Philippe P Richard, and Chuanhui Xiong. "Explicit convex and concave envelopes through polyhedral subdivisions". In: *Mathematical Programming* 138.1-2 (2013), pp. 531–577.