

Rounding on the Standard Simplex: Regular Grids for Global Optimization

Immanuel M. Bomze · Stefan Gollowitzer · E. Alper Yildirim

Received: date / Accepted: date

Abstract Given a point on the standard simplex, we calculate a proximal point on the regular grid which is closest with respect to any norm in a large class, including all ℓ^p -norms for $p \geq 1$. We show that the minimal ℓ^p -distance to the regular grid on the standard simplex can exceed one, even for very fine mesh sizes in high dimensions. Furthermore, for $p = 1$, the maximum minimal distance approaches the ℓ^1 -diameter of the standard simplex. We also put our results into perspective with respect to the literature on approximating global optimization problems over the standard simplex by means of the regular grid.

Keywords Rounding · regular grid · approximation · maximin distance · proximal point

1 Introduction

1.1 Motivations and Paper Organization

Challenging constrained global optimization problems occur already over relatively simple feasible sets, e.g., polytopes. The simplest polytope is the stan-

I. M. Bomze and S. Gollowitzer
Department of Statistics and Operations Research,
University of Vienna,
Austria
E-mail: {immanuel.bomze,stefan.gollowitzer}@univie.ac.at

E. A. Yildirim
Department of Industrial Engineering,
Koç University,
Turkey
E-mail: alperyildirim@ku.edu.tr

standard simplex given by

$$\Delta^n = \left\{ x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1 \right\} \subset \mathbb{R}^n,$$

where \mathbb{R}_+^n denotes the nonnegative orthant. Of course, neither convex nor concave objective functions pose a serious problem for this feasible set, but there are functions with indefinite curvature which constitute the hardest instances. For example, the maximum-clique problem [2] can be written as an indefinite quadratic optimization problem over Δ^n .

In a much more general context, because of their simple polyhedral structure, simplices play a central role in global optimization, see for example [11, 14, 15].

The standard simplex Δ^n can be approximated in a natural way by a sequence of increasingly finer regular grids Δ_r^n , $r = 1, 2, \dots$, where each grid Δ_r^n consists of $\mathcal{O}(n^r)$ rational points. Given an optimization problem over the standard simplex, the grid point with the best objective function value yields an approximate solution and the accuracy of this discretization can be improved as the grid gets finer. This approach was studied in detail by [13], see also [18, Chap. 8]; one important result there is that this algorithm has the best worst-case complexity in some situations. For various classes of objective functions, the regular grid serves as a tool to prove that these classes admit a polynomial-time approximation scheme, be it with respect to absolute or relative approximation ratio [10, 7, 8]. For a concise survey, see [6].

Motivated by these results, we study the properties of regular grids on the standard simplex. To begin with, let us mention the recent paper [4] where reduction strategies for branch-and-bound methods are proposed which avoid evaluation at all grid points. This paper also specifies application in blending and product design. Here, we follow a different approach. To be more precise, we give a simple characterization of the closest point on the grid with respect to a rather large class of norms, including all ℓ^p -norms for $p \geq 1$. We call this procedure a *rounding rule* since it may also be used in the reverse sense, when continuous relaxations of (mixed-)integer optimization problems are employed, to find a good feasible solution in a heuristic way, and at the same time keeping a balance constraint. Naturally, this avenue has been traveled before, one of the most influential papers being probably [5] dealing with a more general situation of integer optimization.

We then establish lower bounds on the maximum distance between any point on the standard simplex and the closest point on the grid, henceforth referred to as the maximin distance. For ℓ^p -norms, the maximin distance is formally defined as

$$A_p(n, r) := \sup_{\mathbf{x} \in \Delta^n} \min_{\mathbf{z} \in \Delta_r^n} \|\mathbf{x} - \mathbf{z}\|_p. \quad (1)$$

For $p \in \{1, 2, \infty\}$, we give an exact characterization of the maximin distance. We also study the asymptotic behavior of the maximin distance for fixed r

as $n \rightarrow \infty$, and we discuss some implications of our findings for the regular grid approximation of global optimization problems over the standard simplex, which also may have consequences for discrete rounding. For instance, due to the special structure of the simplex, we are able to reduce the estimate of [5, Thm.1] focusing on the case $p = \infty$ considerably, getting rid of a factor n .

To be more precise, this paper is organized as follows. In Section 2, we present the rounding rule, which yields a simple characterization of the proximal grid point with respect to a large class of norms. In this context, Section 3 establishes a lower bound for the maximin distance for all ℓ^p -norms for finite $p \geq 1$. In Section 4, we develop lower and upper bounds on the maximin distance as n tends to infinity. For $p \in \{1, 2, \infty\}$, exact maximin distances are presented in Section 5. Finally, Section 6 is devoted to the implications of our results in the context of approximating more general global optimization problems over the standard simplex.

1.2 Notation

The vector of all ones is denoted by $\mathbf{e} = [1, \dots, 1]^\top = \sum_{i=1}^n \mathbf{e}_i \in \mathbb{R}^n$, where \mathbf{e}_i denotes the i -th column of the $n \times n$ identity matrix. For a given real number $y \in \mathbb{R}$, we denote by $\lfloor y \rfloor$ the integer part of y , i.e., the largest integer not exceeding y , and by $s(y) = y - \lfloor y \rfloor$, the fractional part of y (so that we always have $0 \leq s(y) < 1$). Further, we abbreviate by $\mu_{rn} = \min\{n-1, r\}$, and for two integers m and k with $k \leq m$, we abbreviate by $[k : m] = \{k, k+1, \dots, m-1, m\}$, the set of all integers between k and m . Let us also denote by

$$\mathbb{R}_{>}^n = \{\mathbf{s} \in \mathbb{R}^n : s_1 \geq s_2 \geq \dots \geq s_n\},$$

the ordered vectors with n coordinates.

2 The Rounding Rule

Let n and r be given positive integers. Suppose $\mathbf{x} \in \Delta^n$, the standard simplex in \mathbb{R}^n , and fix the grid denominator at r . We define

$$\mathbb{N}_r^n := \left\{ \mathbf{m} \in \mathbb{N}^n : \sum_{i=1}^n m_i = \mathbf{e}^\top \mathbf{m} = r \right\}$$

and consider the regular grid $\Delta_r^n := \frac{1}{r} \mathbb{N}_r^n \subset \Delta^n$.

For $\mathbf{y} \in r\Delta^n$ we denote by $\mathbf{s}(\mathbf{y}) = [s(y_i)]_{i \in [1:n]}$ the fractional part vector of \mathbf{y} and observe that

$$k(\mathbf{y}) := \sum_{i=1}^n s(y_i) = r - \sum_{i=1}^n \lfloor y_i \rfloor \in [0 : n-1]$$

is a non-negative integer. The rounding rule is simple: round the $k(\mathbf{y})$ largest coordinates $s(y_i)$ up to one and all other coordinates $s(y_j)$ down to zero. We may and do re-order the coordinates $s(y_i)$ in non-increasing order so that $\mathbf{s}(\mathbf{y}) \in \mathbb{R}_{\geq}^n$. Thus, we obtain an integer grid vector

$$\mathbf{m}_r(\mathbf{y}) := [\lfloor y_1 \rfloor + 1, \dots, \lfloor y_{k(\mathbf{y})} \rfloor + 1, \lfloor y_{k(\mathbf{y})+1} \rfloor, \dots, \lfloor y_n \rfloor]^\top \in \mathbb{N}_r^n.$$

Indeed, $\mathbf{e}^\top \mathbf{m}_r(\mathbf{y}) = k(\mathbf{y}) + \sum_{i=1}^n \lfloor y_i \rfloor = r$. We now prove that this vector is proximal with respect to a large class of norms, including all ℓ^p -norms defined by $\|\mathbf{x}\|_p = [\sum_i |x_i|^p]^{1/p}$ for any finite $p \geq 1$. Note that if there are ties in the ordering of the coordinates of $\mathbf{s}(\mathbf{y})$, then $\mathbf{m}_r(\mathbf{y})$ need not be the unique proximal vector of \mathbf{y} in the grid \mathbb{N}_r^n .

Theorem 1 *Let $\mathbf{x} \in \Delta^n$ be an arbitrary point on the standard simplex. Suppose that the coordinates of \mathbf{x} are reordered such that $\mathbf{s}(r\mathbf{x}) \in \mathbb{R}_{\geq}^n$. Using the definition above, for any positive integer r , $\frac{1}{r}\mathbf{m}_r(r\mathbf{x})$ is one of the closest points on the regular grid Δ_r^n to \mathbf{x} with respect to any norm of the form $\|\mathbf{x}\| = \varphi(\sum_{i=1}^n \psi(|x_i|))$ with φ and ψ strictly increasing on \mathbb{R}_+ and ψ satisfying*

$$1 + \psi(t) \leq \psi(1 + t) \quad \text{and} \quad \psi(|1 - t|) < 1 + \psi(t) \quad \text{for all } t > 0. \quad (2)$$

Proof We show the analogue statement for $\mathbf{y} = r\mathbf{x}$ and $\mathbf{m}_r(\mathbf{y})$. So let $\mathbf{u} \in \mathbb{N}_r^n$ be a proximal point to \mathbf{y} with respect to the norm $\|\cdot\|$ of the supposed form. First, we claim that $u_i \in \{\lfloor y_i \rfloor, \lfloor y_i \rfloor + 1\}$ for all i . Indeed, assume $u_i < \lfloor y_i \rfloor$ for some i . Then there is a $j \neq i$ such that $u_j \geq \lfloor y_j \rfloor + 1 \geq 1$ (and also $u_j > y_j$), because otherwise we would get a contradiction to $\mathbf{e}^\top \mathbf{u} = r = \mathbf{e}^\top \mathbf{y}$. Now construct an integer point $\mathbf{v} \in \mathbb{N}_r^n$ with coordinates $v_i := u_i + 1 \leq \lfloor y_i \rfloor$, $v_j := u_j - 1 \geq 0$, and $v_k := u_k$ otherwise. We abbreviate by $z := y_i - u_i - 1 \geq 0$ and by $w := u_j - y_j > 0$. By construction and assumption, we get

$$\varphi^{-1}(\|\mathbf{y} - \mathbf{v}\|) - \varphi^{-1}(\|\mathbf{y} - \mathbf{u}\|) = \psi(z) - \psi(1 + z) + \psi(|1 - w|) - \psi(w) < 0,$$

contradicting proximality of \mathbf{u} . Similarly, we can show that $u_j \leq \lfloor y_j \rfloor + 1$ for all j . Hence we arrive at $u_i \in \{\lfloor y_i \rfloor, \lfloor y_i \rfloor + 1\}$ for all i . Finally, suppose that $\mathbf{u} \neq \mathbf{m}_r(\mathbf{y})$. Then, by the argument above, we know that for at least one $j \leq k(\mathbf{y})$ and at least one $i > k(\mathbf{y})$, we get $u_j = \lfloor y_j \rfloor$ while $u_i = \lfloor y_i \rfloor + 1$. Swapping $v_i := u_i - 1$, $v_j := u_j + 1$ and $v_k := u_k$ for all other k as before, we arrive again at $\mathbf{v} \in \mathbb{N}_r^n$ with

$$\begin{aligned} & \varphi^{-1}(\|\mathbf{y} - \mathbf{v}\|) - \varphi^{-1}(\|\mathbf{y} - \mathbf{u}\|) \\ &= \psi(s(y_i)) - \psi(1 - s(y_i)) + \psi(1 - s(y_j)) - \psi(s(y_j)) \leq 0, \end{aligned}$$

by monotonicity of ψ and $s(y_i) \leq s(y_j)$ as $i > k(\mathbf{y}) \geq j$. By proximality of \mathbf{u} , we conclude $\|\mathbf{y} - \mathbf{v}\| = \|\mathbf{y} - \mathbf{u}\|$. Continuing swaps if $\mathbf{v} \neq \mathbf{m}_r(\mathbf{y})$, we conclude that $\|\mathbf{y} - \mathbf{m}_r(\mathbf{y})\| = \|\mathbf{y} - \mathbf{u}\|$ and thus, $\mathbf{m}_r(\mathbf{y})$ is a proximal point of \mathbf{y} in the regular grid \mathbb{N}_r^n . \square

We remark that all ℓ^p -norms with finite $p \geq 1$ satisfy the hypotheses of Theorem 1. On the other hand, the ℓ^∞ -norm does not satisfy these hypotheses. However, the next result reveals that we can still extend the same proximality result to the ℓ^∞ -norm.

Theorem 2 *Let $\mathbf{x} \in \Delta^n$ be an arbitrary point on the standard simplex. Suppose that the coordinates of \mathbf{x} are reordered such that $\mathbf{s}(r\mathbf{x}) \in \mathbb{R}_>^n$. Using the definition above, for any positive integer r , $\frac{1}{r}\mathbf{m}_r(r\mathbf{x})$ is one of the closest points on the regular grid Δ_r^n to \mathbf{x} with respect to the ℓ^∞ -norm.*

Proof In a similar fashion, we establish an analogous result for $\mathbf{y} = r\mathbf{x}$ and $\mathbf{m}_r(\mathbf{y})$. We first establish that there exists a proximal point $\mathbf{u} \in \mathbb{N}_r^n$ such that $u_i \in \{\lfloor y_i \rfloor, \lfloor y_i \rfloor + 1\}$ for each $i = 1, \dots, n$. Let us define $\mathbf{s}(\mathbf{y})$ as before and let $k(\mathbf{y}) = \mathbf{e}^\top \mathbf{s}(\mathbf{y})$. If we round up any subset of $k(\mathbf{y})$ coordinates of \mathbf{y} and round down the remaining $n - k(\mathbf{y})$ coordinates, the resulting point \mathbf{v} satisfies $v_i \in \{\lfloor y_i \rfloor, \lfloor y_i \rfloor + 1\}$ for each $i = 1, \dots, n$. Furthermore, it is easy to verify that $\mathbf{v} \in \mathbb{N}_r^n$. Clearly, we have $\|\mathbf{y} - \mathbf{v}\|_\infty \leq 1$. On the other hand, if we round up fewer or more than $k(\mathbf{y})$ coordinates of \mathbf{y} , and round down the remaining ones, then the remaining point no longer belongs to \mathbb{N}_r^n .

For any point $\mathbf{w} \in \mathbb{N}_r^n$ such that $w_i < \lfloor y_i \rfloor$ for some $i \in \{1, \dots, n\}$, we obtain $\|\mathbf{y} - \mathbf{w}\|_\infty \geq |y_i - w_i| \geq 1$. Similarly, for any $\mathbf{z} \in \mathbb{N}_r^n$ such that $z_j > \lfloor y_j \rfloor + 1$ for some $j \in \{1, \dots, n\}$, we have $\|\mathbf{y} - \mathbf{z}\|_\infty \geq |y_j - z_j| \geq 1$. Therefore, it suffices to consider only the points $\mathbf{v} \in \mathbb{N}_r^n$ such that $v_i \in \{\lfloor y_i \rfloor, \lfloor y_i \rfloor + 1\}$ for each $i = 1, \dots, n$. By the previous discussion, we just need to determine the right subset of $k(\mathbf{y})$ coordinates of \mathbf{y} to be rounded up.

Recall that $\mathbf{m}_r(\mathbf{y})$ is obtained by rounding up the $k(\mathbf{y})$ largest coordinates of $\mathbf{s}(\mathbf{y})$ and rounding down the remaining $n - k(\mathbf{y})$ coordinates. We now establish that $\mathbf{m}_r(\mathbf{y})$ is one of the closest points to \mathbf{y} . Suppose that there exists a proximal point $\mathbf{u} \in \mathbb{N}_r^n$ that is obtained by rounding up a different subset of $k(\mathbf{y})$ coordinates of \mathbf{y} . Since $\mathbf{s}(\mathbf{y}) \in \mathbb{R}_>^n$, there exist two coordinates $i > k(\mathbf{y}) \geq j$ such that $u_i = \lfloor y_i \rfloor + 1$ and $u_j = \lfloor y_j \rfloor$. If we define a new point $\mathbf{v} \in \mathbb{N}_r^n$ such that $v_i = \lfloor y_i \rfloor$, $v_j = \lfloor y_j \rfloor + 1$, and $v_k = u_k$ otherwise, then

$$\begin{aligned} \|\mathbf{y} - \mathbf{v}\|_\infty &= \max\{s(y_i), 1 - s(y_j), \max_{k \notin \{i,j\}} |y_k - v_k|\}, \\ &\leq \max\{s(y_j), 1 - s(y_i), \max_{k \notin \{i,j\}} |y_k - v_k|\}, \\ &= \|\mathbf{y} - \mathbf{u}\|_\infty, \end{aligned}$$

where we used $s(y_i) \leq s(y_j)$ to derive the inequality. If $\mathbf{v} \neq \mathbf{m}_r(\mathbf{y})$, we can similarly continue swaps until $\mathbf{v} = \mathbf{m}_r(\mathbf{y})$. It follows that $\mathbf{m}_r(\mathbf{y})$ is a proximal point of \mathbf{y} . \square

Since $\mathbf{s}(\mathbf{y}) = \mathbf{0}$ if and only if \mathbf{y} is integer with $k(\mathbf{y}) = 0$, i.e., if and only if $\frac{1}{r}\mathbf{y} \in \Delta_r^n$, we only need to consider the case $k(\mathbf{y}) \geq 1$ in the following sections.

3 The Maximin ℓ^p Distance to the Regular Grid

Next, we show that the maximin ℓ^p distance can be computed by solving a finite number of global optimization problems:

Theorem 3 *Let $A_p(n, r)$ denote the maximin ℓ^p distance over Δ^n with respect to the regular grid Δ_r^n . For any $p \in [1, \infty)$, we have*

$$A_p(n, r) = \frac{1}{r} \max_{k \in [1: \mu_{rn}]} \left\{ \sup_{\mathbf{s} \in S_k} [\psi_p(\mathbf{s}, k)]^{1/p} \right\} \quad (3)$$

with

$$\psi_p(\mathbf{s}, k) = \sum_{i \leq k} (1 - s_i)^p + \sum_{i > k} s_i^p,$$

and, for $k \in [1: \mu_{rn}]$,

$$S_k = \left\{ \mathbf{s} \in \mathbb{R}_{>}^n : \mathbf{e}^\top \mathbf{s} = k, 0 \leq s_i < 1 \text{ for all } i \right\}. \quad (4)$$

Further, we have

$$[A_p(n, r)]^p \geq \frac{n}{r^p} \max_{k \in [1: \mu_{rn}]} \varphi_p \left(\frac{k}{n} \right) \quad (5)$$

with

$$\varphi_p(t) = \chi_p(t) + \chi_p(1 - t), \quad \text{where } \chi_p(t) = t(1 - t)^p, \quad t \in [0, 1]. \quad (6)$$

Proof Given $\mathbf{x} \in \Delta^n$, let us define $\mathbf{y} = r\mathbf{x}$. Note that $\mathbf{s}(\mathbf{y}) \in S_k$ for some $k \in [1: \mu_{rn}]$, namely for $k = k(\mathbf{y})$, after possibly rearranging the coordinates in non-increasing order. Conversely, for each $\mathbf{s} \in S_k$, $k \in [1: \mu_{rn}]$, there exists $\mathbf{x} \in \Delta^n$ such that $\mathbf{y} = r\mathbf{x}$ satisfies $\mathbf{s}(\mathbf{y}) = \mathbf{s}$. Indeed, take for instance $y_1 := r - k + s_1 \geq 0$ and all other $y_i = s_i \geq 0$. We therefore have

$$\begin{aligned} A_p(n, r) &= \sup_{\mathbf{x} \in \Delta^n} \min_{\mathbf{z} \in \Delta_r^n} \|\mathbf{x} - \mathbf{z}\|_p \\ &= \left(\frac{1}{r} \right) \max_{k \in [1: \mu_{rn}]} \left\{ \sup_{\mathbf{y} \in \mathbb{N}_r^n : \mathbf{s}(\mathbf{y}) \in S_k} \|\mathbf{y} - \mathbf{m}_r(\mathbf{y})\|_p \right\}. \end{aligned}$$

Thus, (3) follows. Let us now fix k and consider the inner optimization problem, which can be equivalently stated as

$$\sup \left\{ [\psi_p(\mathbf{s}, k)]^{1/p} : \mathbf{s} \in S_k \right\}.$$

Clearly, $\mathbf{s} = \frac{k}{n} \mathbf{e} \in S_k$. It follows that

$$\begin{aligned} \sup_{\mathbf{y} \in \mathbb{N}_r^n : \mathbf{s}(\mathbf{y}) \in S_k} \|\mathbf{y} - \mathbf{m}_r(\mathbf{y})\|_p^p &\geq \sum_{i \leq k} \left(1 - \frac{k}{n} \right)^p + \sum_{j > k} \left(\frac{k}{n} \right)^p \\ &= k \left(1 - \frac{k}{n} \right)^p + (n - k) \left(\frac{k}{n} \right)^p. \end{aligned}$$

Therefore we also have shown (5). \square

4 Asymptotic Values for the Maximin Distance

First, let us derive an asymptotic lower bound for $A_p(n, r)$ for fixed r as $n \rightarrow \infty$:

Theorem 4 For $1 \leq p < \infty$ define, as usual, the conjugate q by $\frac{1}{p} + \frac{1}{q} = 1$; we define $\frac{1}{q} = 0$ if $p = 1$. Then, for any positive integer r , we have

$$\liminf_{n \rightarrow \infty} A_p(n, r) \geq \left(\frac{1}{r}\right)^{\frac{1}{q}}. \quad (7)$$

Proof First, we determine $\dot{\varphi}_p(0) = \dot{\chi}_p(0) - \dot{\chi}_p(1) = 1 > 0$, which means that φ_p is strictly increasing on some interval $[0, \bar{t}]$ with $\bar{t} > 0$. Therefore, if $n \geq (r/\bar{t})$, we have $\frac{k}{n} \in [0, \bar{t}]$ for all $k \in [1:r]$, and therefore, we get from (5)

$$\frac{n}{r} \max_{k \in [1:\mu_{rn}]} \varphi_p\left(\frac{k}{n}\right) = \frac{n}{r} \varphi_p\left(\frac{r}{n}\right) \rightarrow \dot{\varphi}_p(0) = 1 \quad \text{as } n \rightarrow \infty.$$

Hence, taking the p -th root in (5), we arrive at the asymptotic lower bound

$$\liminf_{n \rightarrow \infty} A_p(n, r) \geq \left[r^{1-p} \lim_{n \rightarrow \infty} \frac{n}{r} \max_{k \in [1:\mu_{rn}]} \varphi_p\left(\frac{k}{n}\right) \right]^{\frac{1}{p}} = r^{-\frac{1}{q}}.$$

□

A similar asymptotic result would hold if we allow r to vary with n in a way that $\frac{r}{n} \rightarrow 0$ as $n \rightarrow \infty$. Anyhow, we do not yet know how large $A_p(n, r)$ can actually get. To derive upper bounds, we need to investigate in more detail the optimization problem (3) for general p . For $k \in [1:\mu_{rn}]$, let us define S_k as in (4). Consider the following inner optimization problem in (3) over the closure \bar{S}_k of S_k , which we look at in a parameterized way:

$$\max \{ \psi_p(\mathbf{s}, k) : \mathbf{s} \in \bar{S}_k \} = \max_{\tau \in [0,1]} z_k(\tau), \quad k \in [1:\mu_{rn}], \quad (8)$$

where $z_k(\tau)$ is the optimal value of the problem $(P_k(\tau))$ defined by

$$z_k(\tau) = \max \{ \psi_p(\mathbf{s}, k) : \mathbf{s} \in \bar{S}_k, s_{k+1} \leq \tau \leq s_k \}, \quad k \in [1:\mu_{rn}].$$

Let us fix $\tau \in [0, 1]$. We will first establish some properties of an optimal solution of $(P_k(\tau))$.

We define some disjoint intervals for τ whose union covers the interval $(0, 1]$:

$$\text{For } u \in [0:n-1] \text{ define } I_k(u) := \begin{cases} \left(\frac{k-u-1}{n-u-1}, \frac{k-u}{n-u} \right], & \text{if } u \in [0:k-1], \\ \left(\frac{k}{u+1}, \frac{k}{u} \right], & \text{if } u \in [k:n-1]. \end{cases} \quad (9)$$

Lemma 1 Let $1 < p < \infty$ and $k \in [1:n-1]$ and fix $\tau \in [0, 1]$. Then the optimal solution $\mathbf{s}_k(\tau)$ of $(P_k(\tau))$ is unique, and is given as follows:

(a) for all $\tau \in I_k(u)$ as $u \in [0:k-1]$, we have

$$s_i(\tau) = \begin{cases} 1, & i \in [1:u], \\ k - u - (n - u - 1)\tau, & i = u + 1, \\ \tau, & i \in [u + 2:n]; \end{cases}$$

(b) for all $\tau \in I_k(v)$ as $v \in [k:n-1]$,

$$s_i(\tau) = \begin{cases} \tau, & i \in [1:v], \\ k - v\tau, & i = v + 1, \\ 0, & i \in [v + 2:n]. \end{cases}$$

Proof Let $\hat{\mathbf{s}} \in \mathbb{R}^n$ denote an optimal solution of $(P_k(\tau))$; one such solution must exist due to compactness of \bar{S}_k . We claim that there exists at most one index $i \in [1:k]$ such that $\tau < \hat{s}_i < 1$. Indeed, suppose, to the contrary, that there exist two indices $i_1 \in [1:k]$ and $i_2 \in [1:k]$ such that $i_1 \neq i_2$ and $\tau < \hat{s}_i < 1$ for $i \in \{i_1, i_2\}$. We can assume that $\hat{s}_{i_1} \geq \hat{s}_{i_2}$. Then, by the hypothesis, there exists $\delta > 0$ such that we can increase \hat{s}_{i_1} by δ and decrease \hat{s}_{i_2} by δ without changing the remaining coordinates of $\hat{\mathbf{s}}$, and the resulting solution, after possibly re-ordering the coordinates, is still feasible for $(P_k(\tau))$. By the definition of the objective function $\psi_p(\mathbf{s}, k)$ and our hypotheses that $i_1 \in [1:k]$ and $i_2 \in [1:k]$, we can assume, for simplicity, that the order of the coordinates of the modified solution, denoted by $\bar{\mathbf{s}}$, is the same as that of $\hat{\mathbf{s}}$. We claim that $\bar{\mathbf{s}}$ has a better objective function value than that of $\hat{\mathbf{s}}$, i.e., $\psi_p(\bar{\mathbf{s}}, k) > \psi_p(\hat{\mathbf{s}}, k)$: The function $(1-t)^p$ is strictly convex in $t \in (0, 1)$ as $1 < p < \infty$. Therefore, the linear approximation at any point \hat{t} lies strictly under the function, i.e.,

$$(1-t)^p > (1-\hat{t})^p - p(1-\hat{t})^{p-1}(t-\hat{t}), \quad \text{for all } t \in (0, 1), t \neq \hat{t}.$$

It follows that

$$\begin{aligned} (1-\bar{s}_{i_1})^p &= (1-(\hat{s}_{i_1} + \delta))^p > (1-\hat{s}_{i_1})^p - p\delta(1-\hat{s}_{i_1})^{p-1}, \text{ and} \\ (1-\bar{s}_{i_2})^p &= (1-(\hat{s}_{i_2} - \delta))^p > (1-\hat{s}_{i_2})^p + p\delta(1-\hat{s}_{i_2})^{p-1}. \end{aligned}$$

Adding the two inequalities, we obtain

$$(1-\bar{s}_{i_1})^p + (1-\bar{s}_{i_2})^p > (1-\hat{s}_{i_1})^p + (1-\hat{s}_{i_2})^p + p\delta [(1-\hat{s}_{i_2})^{p-1} - (1-\hat{s}_{i_1})^{p-1}].$$

Since the last term on the right-hand side is nonnegative, it follows that $\psi_p(\bar{\mathbf{s}}, k) > \psi_p(\hat{\mathbf{s}}, k)$, which contradicts the optimality of $\hat{\mathbf{s}}$.

We next claim that there exists at most one index $j \in [k+1:n]$ such that $0 < \hat{s}_j < \tau$. Relying on the strict convexity of the function t^p on $(0, 1)$, we can employ a similar contradiction argument to establish this claim.

Furthermore, we cannot have an index pair $(i, j) \in [1:k] \times [k+1:n]$ such that the inequalities $\tau < \hat{s}_i < 1$ and $0 < \hat{s}_j < \tau$ hold simultaneously. Otherwise, by decreasing \hat{s}_i by a sufficiently small number $\delta > 0$ and increasing \hat{s}_j by the same number δ , we obtain a strictly better feasible solution, which is again a contradiction.

Next, if $\tau = 0$ or $\tau = 1$, the only feasible solution is given by $s_1 = \dots = s_k = 1$

and $s_{k+1} = \dots = s_n = 0$, which implies that $z_k(\tau) = 0$.

Let us now assume that $0 < \tau \leq k/n < 1$. We claim that $\hat{s}_j = \tau$ for $j \in [k+1:n]$. Otherwise, there exists $j \in [k+1:n]$ such that $\hat{s}_j < \tau$, which implies that $\sum_{j=k+1}^n \hat{s}_j < (n-k)\tau$. Since $\mathbf{e}^\top \hat{\mathbf{s}} = k$, it follows that $\sum_{i=1}^k \hat{s}_i > k - (n-k)\tau$. Hence, there exists $i \in [1:k]$ such that

$$\hat{s}_i > \frac{k - (n-k)\tau}{k} = 1 + \tau - \frac{n\tau}{k} \geq \tau.$$

Therefore, we obtain an index pair $(i, j) \in [1:k] \times [k+1:n]$ such that $\tau < \hat{s}_i < 1$ and $0 < \hat{s}_j < \tau$, which is a contradiction.

Let $u(\tau)$ denote the number of entries of $\hat{\mathbf{s}}$ that are equal to one. Therefore, an optimal solution of $(P_k(\tau))$ is given by

$$\hat{s}_i = \begin{cases} 1, & \text{if } i \in [1:u(\tau)], \\ k - u(\tau) - \tau(n - u(\tau) - 1), & \text{if } i = u(\tau) + 1, \\ \tau, & \text{if } i \in [u(\tau) + 2:n]. \end{cases}$$

By definition of $u(\tau)$, we have $\hat{s}_i \in [\tau, 1)$ for $i = u(\tau) + 1$. Therefore, $u(\tau)$ is the unique nonnegative integer that satisfies

$$k - 1 - \frac{(n-k)\tau}{1-\tau} < u(\tau) \leq k - \frac{(n-k)\tau}{1-\tau},$$

i.e.,

$$u(\tau) = \left\lfloor k - \frac{(n-k)\tau}{1-\tau} \right\rfloor. \quad (10)$$

Note that the function $u(\tau)$ has the following closed form expression:

$$u(\tau) = \begin{cases} k-1, & \text{if } 0 < \tau \leq \frac{1}{n-k+1}, \\ k-2, & \text{if } \frac{1}{n-k+1} < \tau \leq \frac{2}{n-k+2}, \\ \vdots & \vdots \\ 1, & \text{if } \frac{k-2}{n-2} < \tau \leq \frac{k-1}{n-1}, \\ 0, & \text{if } \frac{k-1}{n-1} < \tau \leq \frac{k}{n}. \end{cases} \quad (11)$$

So, by definition of $I_k(u)$, we see that $u(\tau) = u$ if and only if $\tau \in I_k(u)$. Thus, assertion (a) follows.

The proof for (b) follows the same lines, only that now we can show that all larger coordinates $\hat{s}_i = \tau$ for all $i \leq k$ whereas there can be at most one smaller coordinate $\hat{s}_{v+1} \in (0, \tau)$, where $v = v(\tau) \geq k$ now denotes the number of coordinates of $\hat{\mathbf{s}}$ that are equal to τ . Here, $v = v(\tau)$ if and only if $\tau \in \left(\frac{k}{v+1}, \frac{k}{v}\right] = I_k(v)$. \square

Lemma 2 *Let $1 < p < \infty$. For any $k \in [1:n-1]$, we can calculate the optimal value $z_k(\tau)$ of the problem $(P_k(\tau))$ as follows:*

$$z_k(\tau) = \begin{cases} [1 - k + u + (n - u - 1)\tau]^p + (k - u - 1)[1 - \tau]^p + (n - k)\tau^p, & \text{if } \tau \in I_k(u) \text{ and } u \in [0:k-1], \\ k[1 - \tau]^p + (v - k)\tau^p + [k - v\tau]^p, & \text{if } \tau \in I_k(v) \text{ and } v \in [k:n-1]. \end{cases} \quad (12)$$

Furthermore, the function $z_k(\tau)$ is convex in τ on every subinterval $I_k(u)$, for all $u \in [0:n-1]$; and continuous over the whole interval $[0, 1]$.

Proof The formulae (12) follow immediately from Lemma 1. Convexity is immediate by $1 < p < \infty$, and continuity follows by direct inspection. \square

Optimizing $z_k(\tau)$ is in itself a demanding global optimization problem in one variable, see Figure 1 (where $p = 5$, $k = 10$, and $n = 30$). This serves also as a counterexample that $\tau = \frac{k}{n} \in [0, 1]$ is not always the maximizer of z_k .

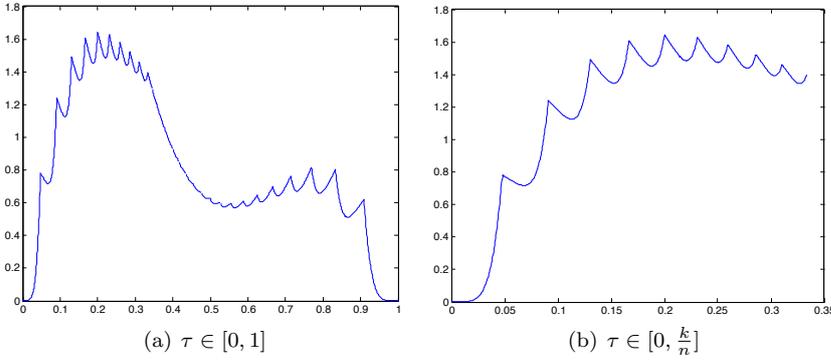


Fig. 1 The function $z_k(\tau)$ for $p = 5$, $k = 10$, and $n = 30$

Lemma 3 *For $1 < p < \infty$ and any positive integer r , we have*

$$\limsup_{n \rightarrow \infty} A_p(n, r) = \left(\frac{1}{r}\right)^{\frac{1}{q}} \limsup_{n \rightarrow \infty} [\max\{1\} \cup \left\{ \frac{v}{r} \varphi_p\left(\frac{k}{v}\right) : k \in [1:r], v \in [k:n-1] \right\}]^{\frac{1}{p}}.$$

Proof Lemma 2 implies that $\max\{z_k(\tau) : \tau \in [0, 1]\}$ can be written, for every fixed $k \in [1:r]$, as a finite maximization problem, namely as

$$\max_{\tau \in [0, 1]} z_k(\tau) = \max \left\{ \max_{u \in [0:k-1]} z_k\left(\frac{k-u}{n-u}\right), \max_{v \in [k:n-1]} z_k\left(\frac{k}{v}\right) \right\}.$$

Now, let \bar{t} be defined as in the proof of Theorem 4. For $u \in [0:k-1]$, we have $\frac{k-u}{n-u} \leq \frac{r}{n} \leq \bar{t}$, so $\varphi_p\left(\frac{k-u}{n-u}\right) \leq \varphi_p\left(\frac{r}{n}\right)$ as φ_p is increasing on the interval $[0, \bar{t}]$.

We arrive at

$$\frac{1}{r}z_k \left(\frac{k-u}{n-u} \right) = \frac{n-u}{r} \varphi_p \left(\frac{k-u}{n-u} \right) \leq \frac{n}{r} \varphi_p \left(\frac{r}{n} \right) \rightarrow \dot{\varphi}_p(0) = 1 \quad \text{as } n \rightarrow \infty.$$

Hence, we only have to deal with $\frac{1}{r}z_k \left(\frac{k}{v} \right) = \frac{v}{r} \varphi_p \left(\frac{k}{v} \right)$ for $v \in [k:n-1]$ and $k \in [1:r]$. \square

Theorem 5 *For $2 \leq p < \infty$ and any positive integer r , we have*

$$\lim_{n \rightarrow \infty} A_p(n, r) = \left(\frac{1}{r} \right)^{\frac{1}{q}}.$$

Proof We study $\frac{v}{r} \varphi_p \left(\frac{k}{v} \right)$ where $v \in [k:n-1]$ and $k \in [1:r]$, and consider the function $\beta_p(t) = \frac{1}{t} \varphi_p(t) = (1-t)\sigma_p(t)$ with $\sigma_p(t) = t^{p-1} + (1-t)^{p-1}$. Now, it is easily seen that for $2 \leq p < \infty$, the function σ_p is convex and takes values between 2^{-p+2} and 1 on $[0, 1]$, so that

$$\frac{v}{r} \varphi_p \left(\frac{k}{v} \right) = \frac{k}{r} \left(1 - \frac{k}{v} \right) \sigma_p \left(\frac{k}{v} \right) \leq 1 \quad \text{for all } v \in [k:n-1], k \in [1:r].$$

Therefore Lemma 3 yields $\limsup_{n \rightarrow \infty} A_p(n, r) = \left(\frac{1}{r} \right)^{\frac{1}{q}}$, and the assertion follows by Theorem 4. \square

Interestingly, even though the case $p = \infty$ does not match the assumptions of Theorem 1, the exact answer is known even for finite n (see Theorem 7 below), and fits the result above when $q = 1$. On the other hand, unfortunately, for $1 \leq p < 2$, we cannot transfer above arguments, but we can find in a similar way an explicit asymptotic upper bound:

Theorem 6 *For $1 \leq p < 2$ and fixed positive integer r , we have*

$$\limsup_{n \rightarrow \infty} A_p(n, r) \leq \sqrt[p]{2} \left(\frac{1}{r} \right)^{\frac{1}{q}}.$$

Proof We discuss the function β_p from the proof of Theorem 5 directly. From

$$\dot{\beta}_p(t) = t^{p-2} [p(1-t) - 1] - p(1-t)^{p-1},$$

we see that all critical points of β_p must be less than $\frac{1}{2}$; indeed, $p < 2$ and $t \geq \frac{1}{2}$ imply $p(1-t) < 1$ and hence, $\dot{\beta}_p(t) < 0$. As $\beta_p(0) = 1$ for $p > 1$ while $\beta_1(0) = 2$ and $\beta_p(1) = 0$ always hold, we only need to show $\beta_p(t) \leq 2$ for all $t \leq \frac{1}{2}$. However, this is immediate from

$$\beta_p(t) = (1-t) [t^{p-1} + (1-t)^{p-1}] \leq 2(1-t)^p \leq 2.$$

The result follows as in the proof of Theorem 5. \square

For $p = 1$ we can get an exact answer for finite n in the following section, which shows that the upper bound in the previous theorem is tight. Strangely, this is also true for $p = \infty$, the only case where both bounds from Theorems 5 and 6 coincide.

5 Exact Values for the Maximin Distance

Let us start with $p = \infty$, the maximum norm. The next result reveals that we can get an explicit expression for the maximin distance even for finite n and any positive integer r ; note that the general result in [5, Thm.1], adapted to our case, would give an upper bound of $\frac{n}{r}$ which is considerably larger, but this is understandable due to the special structure of the standard simplex.

Theorem 7 *For any $n \geq 2$ and any positive integer r , we have*

$$A_\infty(n, r) = \frac{1}{r} \left(1 - \frac{1}{n} \right). \quad (13)$$

Proof Let $\mathbf{x} \in \Delta^n$ and let $\mathbf{y} = r\mathbf{x}$. Suppose that the coordinates are reordered so that $\mathbf{s}(\mathbf{y}) \in \mathbb{R}_>^n$. By Theorem 2, $\mathbf{m}_r(\mathbf{y}) \in \mathbb{N}_r^n$ is one of the closest points to \mathbf{y} with respect to the ℓ^∞ -norm. Note that

$$\|\mathbf{y} - \mathbf{m}_r(\mathbf{y})\|_\infty = \max\{1 - s(y_k), s(y_{k+1})\},$$

where $k := k(\mathbf{y})$. Let us fix $k \in [1: \mu_{rn}]$ and consider the following optimization problem parameterized by k :

$$\zeta_k := \sup\{\max\{1 - s_k, s_{k+1}\} : \mathbf{s} \in S_k\},$$

where S_k is defined as in (4). We first claim that

$$\zeta_k = \max\left\{1 - \frac{1}{n-k+1}, \frac{k}{k+1}\right\}. \quad (14)$$

To this end, we first show that $s_k \geq \frac{1}{n-k+1}$ for each $\mathbf{s} \in S_k$. Indeed, if $s_k < \frac{1}{n-k+1}$, then

$$\sum_{i=1}^{k-1} s_i = k - \sum_{j \geq k} s_j > k - 1,$$

which contradicts $s_i < 1$ for each i . Furthermore, for sufficiently small $\epsilon > 0$, the solution $\hat{\mathbf{s}} \in \mathbb{R}^n$ given by $\hat{s}_i = 1 - \epsilon$ for $i \in [1:k-1]$, $\hat{s}_k = \frac{1}{n-k+1} + (k-1)\epsilon$, and $\hat{s}_j = \frac{1}{n-k+1}$ for $j \in [k+1:n]$ belongs to S_k . Letting $\epsilon \searrow 0$, it follows that $\inf\{s_k : \mathbf{s} \in S_k\} = \frac{1}{n-k+1}$. Note that these arguments hold also in case $k = 1$. In a similar fashion, we claim that $s_{k+1} \leq \frac{k}{k+1}$ for each $\mathbf{s} \in S_k$. Indeed, if $s_{k+1} > \frac{k}{k+1}$, then

$$k \geq \sum_{i=1}^{k+1} s_i > k,$$

which is a contradiction. Clearly, the solution $\tilde{\mathbf{s}} \in \mathbb{R}^n$ given by $\tilde{s}_i = \frac{k}{k+1}$ for $i \in [1:k+1]$ and $\tilde{s}_j = 0$ otherwise also belongs to S_k . So $\max\{s_{k+1} : \mathbf{s} \in S_k\} = \frac{k}{k+1}$, and (14) is established. It follows that

$$\zeta_k = \begin{cases} 1 - \frac{1}{n-k+1}, & \text{if } 1 \leq k \leq \frac{n}{2}, \\ \frac{k}{k+1}, & \text{if } \frac{n}{2} < k \leq \mu_{rn}. \end{cases}$$

For $1 \leq k \leq \frac{n}{2}$, since $1 - \frac{1}{n-k+1}$ is a decreasing function of k , the largest value of $1 - \frac{1}{n}$ is achieved when $k = 1$. On the other hand, if $\frac{n}{2} < k \leq \mu_{rn}$, then $k \leq n - 1$. Since $\frac{k}{k+1}$ is an increasing function of k , the largest value never exceeds $\frac{n-1}{n} = 1 - \frac{1}{n}$ in this case. It follows that

$$A_\infty(n, r) = \frac{1}{r} \max_{k \in [0: \mu_{rn}]} \zeta_k = \frac{1}{r} \left(1 - \frac{1}{n} \right),$$

which establishes our assertion. \square

It is worth noticing that the maximin bound $A_\infty(n, r)$ is actually attained. Indeed, let $\hat{\mathbf{y}} \in \mathbb{R}^n$ be given by $\hat{y}_1 = r - 1 + \frac{1}{n}$ and $\hat{y}_j = \frac{1}{n}$ for each $j \in [2: n]$. Clearly, $\hat{\mathbf{y}} \in r\Delta^n$ and $\mathbf{s}(\hat{\mathbf{y}}) = \frac{1}{n}\mathbf{e}$. It is easy to verify that $\|\hat{\mathbf{y}} - \mathbf{m}_r(\hat{\mathbf{y}})\|_\infty = 1 - \frac{1}{n}$ for each $n \geq 2$.

Unfortunately, this favourable behaviour cannot be exploited too well; see Section 6 for a discussion.

Next we deal with $p = 1$. Beforehand, we note, for any positive integer r and any $\mathbf{y} \in \mathbb{N}_r^n$, that

$$\|\mathbf{y} - \mathbf{m}_r(\mathbf{y})\|_1 = \sum_{i \leq k(\mathbf{y})} (1 - s_i(\mathbf{y})) + \sum_{i > k(\mathbf{y})} s_i(\mathbf{y}) = 2 \sum_{i > k(\mathbf{y})} s_i(\mathbf{y}). \quad (15)$$

Theorem 8 *Let r be a positive integer and let $A_1(n, r)$ denote the maximin ℓ^1 distance over Δ^n with respect to the regular grid Δ_r^n . For any $n \geq 2$, we have*

$$A_1(n, r) = \begin{cases} \frac{n}{2r}, & \text{if } 2 \leq n \leq 2r \text{ and } n \text{ is even,} \\ \frac{(n-1)(n+1)}{2n}, & \text{if } 2 \leq n \leq 2r \text{ and } n \text{ is odd,} \\ 2 \left(1 - \frac{r}{n} \right), & \text{if } n > 2r. \end{cases} \quad (16)$$

Proof Let us fix $k \in [1: \mu_{rn}]$ and consider the corresponding optimization problem given by (8). We claim that $\mathbf{s}_k^* = (k/n)\mathbf{e} \in S_k$ is a maximizer of $\psi_1(\mathbf{s}, k)$ over \bar{S}_k . Note that $\psi_1(\mathbf{s}_k^*, k) = 2k(n-k)/n$. Suppose, for a contradiction, that there exists $\mathbf{v} \in \mathbb{N}_r^n$ such that $\hat{\mathbf{s}} := \mathbf{s}(\mathbf{v}) \in \bar{S}_k$ and $\psi_1(\hat{\mathbf{s}}, k) = \|\mathbf{v} - \mathbf{m}_r(\mathbf{v})\|_1 > 2k(n-k)/n$ when $p = 1$. By (15), it follows that

$$2 \sum_{j=k+1}^n \hat{s}_j = \|\mathbf{v} - \mathbf{m}_r(\mathbf{v})\|_1 > \frac{2k(n-k)}{n},$$

which implies that $\hat{s}_{k+1} > k/n$. By definition of \bar{S}_k , it follows that $\hat{s}_1 \geq \hat{s}_2 \geq \dots \geq \hat{s}_k \geq \hat{s}_{k+1} > k/n$. Therefore,

$$\sum_{i=1}^k \hat{s}_i + \sum_{j=k+1}^n \hat{s}_j > k \left(\frac{k}{n} \right) + \frac{k(n-k)}{n} = k,$$

which contradicts the assumption that $\hat{\mathbf{s}} \in \bar{S}_k$. Therefore,

$$\begin{aligned} A_1(n, r) &= \left(\frac{2}{r} \right) \max_{k \in [1: \mu_{rn}]} \frac{k(n-k)}{n} \\ &= \left(\frac{2n}{r} \right) \max_{k \in [1: \mu_{rn}]} \left(\frac{k}{n} \right) \left(1 - \frac{k}{n} \right). \end{aligned} \quad (17)$$

Considering the function $\varphi_1(t) = 2t(1-t)$ over $t \in [0, 1]$, which increases on $[0, \frac{1}{2}]$ and decreases on $[\frac{1}{2}, 1]$, it follows that the maximum is attained at $k = r$ if $r < n/2$. If $r \geq n/2$, then the maximum is attained at $k = \frac{n}{2}$ if n is even and at $k = \frac{n+1}{2}$ (or $k = \frac{n-1}{2}$) if n is odd. The assertion follows. \square

We remark that asymptotically, if $n \rightarrow \infty$ and $\frac{r}{n} \rightarrow 0$, the maximin distance $A_1(n, r)$ approaches the ℓ^1 -diameter of Δ (given by $\max_{x, y \in \Delta} \|x - y\|_1 = 2$). Note also that this illustrates that Theorem 5 does not hold for $p = 1$. Rather, the upper bound given by Theorem 6 is tight.

Finally we turn to the most frequently used norm, the Euclidean norm where $p = 2$.

Theorem 9 *For $p = 2$, $n \geq 2$, and any positive integer r , we get the following explicit expression:*

$$A_2(n, r) = \begin{cases} \frac{\sqrt{n}}{2r}, & \text{if } n \leq 2r, n \text{ even} \\ \frac{1}{2r} \sqrt{\frac{n^2-1}{n}}, & \text{if } n \leq 2r, n \text{ odd} \\ \sqrt{\frac{n-r}{rn}}, & \text{if } n > 2r. \end{cases} \quad (18)$$

Proof If $p = 2$, then the function $z_k(\tau)$ is non-increasing in τ over the interval $\tau \in (\frac{k}{n}, 1] = \bigcup_{v=k}^{n-1} I_k(v)$. This follows from specializing (12) to $p = 2$, which gives $z_k(\tau) = (v+1)[v\tau^2 - 2k\tau] + k(k+1)$, with derivative $\dot{z}_k(\tau) = 2(v+1)[v\tau - k] \leq 0$ on $\tau \in I_k(v)$. So, arguing as in Section 4 above, we obtain

$$\max_{\tau \in [0, 1]} z_k(\tau) = \max_{u \in [0: k-1]} z_k\left(\frac{k-u}{n-u}\right) = (n-k) \max_{u \in [0: k-1]} \frac{k-u}{n-u} = \frac{k(n-k)}{n}.$$

The claim follows by maximizing the latter expression over $k \in [1: \mu_{rn}]$. \square

6 Approximation Considerations

First, let us establish the dependence of a Lipschitz constant $L_p(f)$ on the dimension n .

Lemma 4 *For a smooth function f with a bound*

$$M_f = \max \left\{ \left| \frac{\partial f}{\partial x_i}(\mathbf{z}) \right| : \mathbf{z} \in \Delta^n, i \in [1: n] \right\}$$

on the partial derivatives, we get, for any p with $1 < p \leq \infty$, a Lipschitz constant

$$L_p(f) \leq \alpha(f, q) := \max_{\mathbf{z} \in \Delta} \|\nabla f(\mathbf{z})\|_q \leq M_f n^{1/q}.$$

Proof By Hölder's inequality and the mean value theorem, for some $\mathbf{z} \in \text{conv}\{\mathbf{x}, \mathbf{y}\} \subset \Delta^n$,

$$\begin{aligned} |f(\mathbf{y}) - f(\mathbf{x})| &= |(\mathbf{y} - \mathbf{x})^\top \nabla f(\mathbf{z})| \\ &\leq \|\mathbf{y} - \mathbf{x}\|_p \|\nabla f(\mathbf{z})\|_q \\ &\leq \alpha(f, q) \|\mathbf{y} - \mathbf{x}\|_p. \end{aligned}$$

Now $\|\nabla f(\mathbf{z})\|_q^q = \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(\mathbf{z}) \right|^q \leq n M_f^q$, so that $\alpha(f, q) \leq M_f n^{1/q}$ for all $p > 1$. \square

Looking at small departures from linear functions, we see that the estimates of Lemma 4 are tight.

Suppose now that we want to approximate the optimal value in a quite natural way:

$$f^* := \min \{f(\mathbf{x}) : \mathbf{x} \in \Delta^n\} \approx f_r^* := \min \{f(\mathbf{m}) : \mathbf{m} \in \Delta_r^n\} \geq f^*,$$

i.e., by the minimal value of f over the regular grid. This approach has been studied extensively by many authors, yielding (in-)approximability results for various classes of objective functions f , and with respect to various measures for departure from optimality in f .

Let us start with d -degree polynomials $f(\mathbf{x})$ over Δ^n . By multiplying some terms with suitable powers of $\mathbf{e}^\top \mathbf{x} = 1$, and adding a suitable constant $K(\mathbf{e}^\top \mathbf{x})^d$, we may and do assume that f is homogeneous and strictly positive over Δ^n . If we now introduce squared variables $y_i^2 = x_i$ to get rid of the sign constraints $x_i \geq 0$ over Δ^n , we end up with the problem to minimize a strictly positive homogeneous polynomial $p(\mathbf{y})$ of degree $2d$ over the Euclidean sphere S^n . If f is not linear, then $d > 1$ so $2d \geq 4$. The papers [10,8], see also [6, Cor.4.2], use the regular grid to show the existence of a PTAS (polynomial-time approximation scheme) for such class of polynomials. By contrast, [12, Prop.3.2.2] says that unless $\mathbf{P}=\mathbf{NP}$, there cannot exist any PTAS for minimizing a homogeneous polynomial $p(\mathbf{y})$ of *odd* degree $2d + 1 > 3$ over the Euclidean sphere S^n .

Next, let us discuss more general Lipschitz functions f with, e.g., $L_2(f) \leq L$ for L fixed. If, in addition, we know that f is *convex*, then the ellipsoid algorithm [13] can find a regular grid point $\tilde{\mathbf{x}} \in \Delta_r^n$ after $2n(n+1)(\ln \frac{L}{\varepsilon} + \ln n) + 2$ iterations (with polynomial effort) such that $f(\tilde{\mathbf{x}}) - f^* \leq \varepsilon$. For non-convex f , [7] obtained the following approximation result:

$$f_r^* - f^* \leq \frac{2L}{r^2}. \quad (19)$$

Of course, the field of global Lipschitz optimization is vast, for optimality principles and similar approximation ideas see, e.g. [9,17], and a recent survey [16]. However, in view of above approach, we know that $L_2(f) \approx M_f \sqrt{n}$, and it may be more natural to fix M_f rather than $L_2(f)$ with increasing n ; the right-hand side of (19) then should be replaced by the bound $2M_f \frac{\sqrt{n}}{r^2}$ which is

still dominating the general result we get from the maximin proximal distance: if $f_r^* = f(\mathbf{x}^*)$ where $\mathbf{x}^* \in \Delta^n$ is an optimal solution and $\bar{\mathbf{x}} = \frac{1}{r}\mathbf{m}_r(r\mathbf{x}^*) \in \Delta_r^n$ is a proximal point to \mathbf{x}^* on the grid, then

$$f_r^* - f^* \leq f(\bar{\mathbf{x}}) - f(\mathbf{x}^*) \leq L_p(f) \|\bar{\mathbf{x}} - \mathbf{x}^*\|_p \leq L_p(f) A_p(n, r) \leq L_p(f) \sqrt[p]{2} r^{-\frac{1}{q}} \quad (20)$$

if n is large. This estimate is not as tight as the one in (19), regardless of the choice of p . Similarly, for $p = q = 2$, we get also for an M_f -based estimate an upper bound of $M_f \sqrt{1 - \frac{r}{n}} > 2M_f \frac{\sqrt{n}}{r^2}$. Furthermore, note that only for $p = 1$, the estimate $L_1(f) \leq M_f$ from Lemma 4 does not increase with n but this does not help in view of Theorem 8.

In [7], further classes of objective functions which admit a PTAS are identified, along similar ideas, namely to fix upper bounds on the (higher-order) derivatives of f over Δ^n . However, as stressed in [6]: “It is still an open question to completely classify the classes of functions that allow a PTAS.”

Now, let us pass to the relative approximation ratio, i.e., to an accuracy measure which is positively homogeneous of degree zero in f , to avoid scaling effects. Many such measures are possible [1], e.g., the relative error $\frac{f_r^* - f^*}{f^*}$ or, alternatively, $\frac{f_r^* - f^*}{L_p^*(f)}$ if $L_p^*(f)$ is the minimal global p -norm Lipschitz constant of f over Δ^n . For smooth functions, we may replace $L_p^*(f)$ by M_f which does not depend on p , and consider the quotient $\frac{f_r^* - f^*}{M_f}$; also note that the p -norm diameter of Δ^n equals $\sqrt[p]{2}$ for all n .

Corollary 1 *Let $1 \leq p \leq \infty$ and the conjugate q defined by $\frac{1}{p} + \frac{1}{q} = 1$ with the usual extensions. Then, we have*

$$\limsup_{n \rightarrow \infty} \left(\frac{1}{n} \right)^{\frac{1}{q}} \frac{|f_r^* - f^*|}{M_f} \leq \sqrt[p]{2} \left(\frac{1}{r} \right)^{\frac{1}{q}}.$$

Proof Let $f^* = f(\mathbf{x}^*)$, where $\mathbf{x}^* \in \Delta^n$ is an optimal solution and $\bar{\mathbf{x}} = \frac{1}{r}\mathbf{m}_r(r\mathbf{x}^*) \in \Delta_r^n$ be a proximal point to \mathbf{x}^* on the grid. Then, as in (20),

$$f_r^* - f^* \leq L_p(f) A_p(n, r) \leq M_f n^{\frac{1}{q}} A_p(n, r)$$

by Lemma 4. Theorem 6 yields the result. \square

Therefore, for large n and fixed r , we can, by this naive method, bound above the relative error $\frac{|f_r^* - f^*|}{M_f}$ by $\text{diam}_p(\Delta^n) \left(\frac{n}{r}\right)^{\frac{1}{q}}$, which, for $p > 1$, is even worse than the straightforward upper bound, namely $\frac{|f_r^* - f^*|}{M_f} \leq \text{diam}_p(\Delta^n)$. As before, more precise formulations for finite n can be derived for $p \in \{1, 2, \infty\}$ from Theorems 7, 8 and 9, however, this will not improve the results.

Another option for a scale-invariant accuracy measure would be $\frac{f_r^* - f^*}{\text{span}(f)}$ where $\text{span}(f) := \max_{\mathbf{x} \in \Delta^n} f(\mathbf{x}) - f^* \leq \sqrt[p]{2} M_f$. In these relative terms, and for quadratic functions f , the simplest class where the optimization problem over

Δ^n is NP-hard, it has been proved by methods of copositive approximations in [3] that

$$\frac{f_r^* - f^*}{\text{span}(f)} \leq \frac{1}{r+2},$$

which implies $\frac{|f_r^* - f^*|}{M_f} \leq \frac{\sqrt[r]{2}}{r+2}$. These results were later extended to polynomials of fixed degree d [10, 8]; see also [12, Thm.3.2.3]. More general functions f can be treated, assuming bounds similar to M_f on the (higher-order) derivatives over Δ^n , see [7]. All these methods use advanced techniques like conic (copositive) optimization, or closely related linear or semidefinite optimization algorithms for Lagrange interpolation on the regular grid Δ_r^n .

Our preceding findings shed, as we hope, some light on the nature of global optimization problems on the standard simplex, and, in particular, illustrate the fact that the naive approximation via proximal points is not powerful enough for establishing PTAS results for larger classes of objective functions, corroborating the need of refined methods, e.g., reduction strategies like those in [4]. Our observations also demonstrate, therefore, the importance of recent developments like copositive optimization concepts for the progress on approximation results.

Acknowledgements The authors are grateful to Oliver Stein for valuable remarks. E. Alper Yıldırım was supported in part by Turkish Scientific and Technological Research Council (TÜBİTAK) Grant 112M870 and by TÜBA-GEBİP (Turkish Academy of Sciences Young Scientists Award Program).

References

1. Ausiello, G., P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi: *Complexity and approximation: Combinatorial optimization problems and their approximability properties*. Springer, Berlin 1999.
2. Bomze, I.M., M. Budinich, P. Pardalos, and M. Pelillo: The maximum clique problem, in: D.-Z. Du, P.M. Pardalos (eds.), *Handbook of Combinatorial Optimization (supp. Vol. A)*, 1–74. Kluwer, Dordrecht 1999.
3. Bomze, I.M., and E. de Klerk: Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *J. Global Optim.* **24**, 163–185 (2002).
4. Casado, L. G., I. García, B. G. Tóth, and E. M. Hendrix: On determining the cover of a simplex by spheres centered at its vertices. *J. Global Optim.* **50**, 645–655 (2011).
5. Cook, W., A.M.H. Gerards, A. Schrijver, É. Tardos: Sensitivity theorems in integer linear programming. *Math. Programming* **34**, 251–264 (1986).
6. de Klerk, E.: The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central Eur. J. Oper. Res.*, **16**, 111–125 (2008).
7. de Klerk, E., D. den Hertog, and G. Elabwabi: On the complexity of optimization over the standard simplex. *Eur. J. Oper. Res.* **191**, 773–785 (2008).
8. de Klerk, E., M. Laurent, and P. Parrilo: A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theor. Comp. Sci.* **361**, 210–225 (2006).
9. Evtushenko, Y.: Numerical methods for finding global extreme (case of a non-uniform mesh). *USSR Computational Mathematics and Mathematical Physics* **11**, 38–54 (1971).
10. Fabusovich, L.: Global optimization of homogeneous polynomials on the simplex and on the sphere. In: C. Floudas, P. Pardalos (eds.), *Frontiers in global optimization*, 109–121. Kluwer, Dordrecht 2003.

11. Horst, R. and H. Tuy: *Global Optimization: Deterministic Approaches*. Springer, Berlin 1996.
12. Li, Z., S. He, and S. Zhang: *Approximation methods for polynomial optimization: Models, algorithms and applications*. Springer Briefs in Optimization, New York 2012.
13. Nemirovsky, A.S. and D.B. Yudin: *Problem complexity and method efficiency in optimization*. Wiley-Interscience, New York 1983.
14. Paulavičius, R. and J. Žilinskas: *Simplicial Global Optimization*. Springer, New York 2013.
15. Raber, U.: A simplicial branch-and-bound method for solving nonconvex all-quadratic programs. *J. Global Optim.* **13** 417–432 (1998).
16. Sergeyev, Y.D. and D.E. Kvasov: Lipschitz global optimization. In: J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith (eds.), *Wiley Encyclopedia of Operations Research and Management Science*. Wiley, New York 2010.
17. Sukharev, A.G.: *Minimax algorithms in problems of numerical analysis (in Russian)*. Nauka, Moscow 1989.
18. Traub J.F., G.W. Wasilkowski and H. Wozniakowski: *Information-based complexity*. Academic Press, New York 1988.