

# Calculus of the facial distance

Javier Peña <sup>\*</sup>      Elias Wirth <sup>†</sup>

August 31, 2026

## Abstract

We develop a few calculus rules to compute or lower bound the facial distance of a polytope. We illustrate our calculus rules on various popular polytopes. In particular, we provide a lower bound on the facial distance of the Birkhoff polytope.

## 1 Introduction

The *facial distance* of a polytope, or equivalently its *pyramidal width*, is a key condition measure of a polytope that determines the speed of convergence of variants of the Frank-Wolfe algorithm on a polytope. More precisely, consider a minimization problem of the form

$$\min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}) \tag{1.1}$$

where  $f$  is  $L$ -smooth and  $\mu$ -strongly convex, and  $\mathcal{C}$  is a polytope. The developments in [1, 2, 8–12, 14] show that several variants of the Frank-Wolfe algorithm applied to (1.1) generate a sequence of iterates  $\mathbf{x}^{(t)}$ ,  $t = 0, 1, \dots$  such that

$$f(\mathbf{x}^{(t)}) - \min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}) = \mathcal{O} \left( \exp \left( \frac{\mu}{L} \cdot \frac{\Phi(\mathcal{C})^2}{\text{diam}(\mathcal{C})^2} \cdot t \right) \right)$$

where  $\text{diam}(\mathcal{C})$  and  $\Phi(\mathcal{C})$  are respectively the diameter and the facial distance of  $\mathcal{C}$ .

The facial distance is also closely related to other interesting problems in convex geometry. In particular, it is closely related to the work of Deza et al. on *kissing polytopes* [3–5]. The articles [8, 9] compute the exact value of the facial distance for the following polytopes: the standard simplex, the  $\ell_\infty$  unit ball, and the  $\ell_1$  unit ball. Besides these special cases, there is very limited work on computing or even estimating the facial distance of general polytopes. Indeed, this computational task seems to pose a significant challenge.

The central goal of this paper is to provide some calculus rules that facilitate the computation of the facial distance. Our work is inspired by a recent development of Iommazzo et al. [7, Theorem 4.5] who derived an expression for the facial distance of the cartesian product of polytopes. We develop three main calculus rules: a product rule that generalizes [7, Theorem 4.5], a rule for obtaining lower bounds in terms of facet-exposing directions, and a rule for computing distances under linear transformations. We illustrate how these three rules yield exact expressions or lower bounds on the facial distance for new polytopes.

---

<sup>\*</sup>jfp@andrew.cmu.edu

<sup>†</sup>elias.s.wirth@gmail.com

A particularly interesting application of our results is a lower bound on the facial distance of the Birkhoff polytope, and a similar bound for more general *simplex-like* polytopes.

Our paper is organized as follows. Section 2 recalls the formal definition of facial distance and introduces some notation and terminology that we use throughout the paper. Section 3, Section 4, and Section 5 present our main developments, namely some calculus rules for products, facet-exposing directions, and linear transformations. Section 6 illustrates our calculus rules on several popular polytopes. Finally, Section 7 discusses a conjecture on the facial distance that is naturally suggested by the examples in Section 6.

## 2 Preliminaries

Let  $\mathbb{R}^n$  be endowed with a norm  $\|\cdot\|$  and let  $\mathcal{C} \subseteq \mathbb{R}^n$  be a polytope. Denote the sets of vertices and nonempty faces of  $\mathcal{C}$  by  $\text{vert}(\mathcal{C})$  and  $\text{faces}(\mathcal{C})$  respectively. For  $F, G \in \text{faces}(\mathcal{C})$  let

$$\text{dist}(F, G) = \min_{\mathbf{x} \in F, \mathbf{y} \in G} \|\mathbf{x} - \mathbf{y}\|.$$

To ease notation, we will write  $\mathcal{C}[G]$  as shorthand for  $\text{conv}(\text{vert}(\mathcal{C}) \setminus G)$  whenever  $\mathcal{C}$  is a polytope and  $G \in \text{faces}(\mathcal{C})$  with  $G \neq \mathcal{C}$ . For  $k \in \mathbb{N}$ , we will write  $[k]$  to denote the set  $\{1, \dots, k\}$ . We next recall the definition of our main object of discussion, namely the *facial distance* of a polytope. We also recall the definition of the much simpler and intuitive *diameter* of a polytope.

**Definition 2.1** (facial distance). Let  $\mathbb{R}^n$  be endowed with a norm  $\|\cdot\|$  and let  $\mathcal{C} \subseteq \mathbb{R}^n$  be a polytope. For  $F \in \text{faces}(\mathcal{C})$  the *facial distance* from  $F$  is defined as

$$\Phi(F, \mathcal{C}) := \min_{\substack{G \in \text{faces}(\mathcal{C}) \\ G \subsetneq \mathcal{C}}} \text{dist}(F, \mathcal{C}[G]).$$

We will write  $\Phi(\mathcal{C})$  as shorthand for  $\Phi(\mathcal{C}, \mathcal{C})$ . We note that  $\Phi(\mathcal{C}, \mathcal{C}) \leq \Phi(F, \mathcal{C})$  for every  $F \in \text{faces}(\mathcal{C})$ . Furthermore,  $\Phi(\mathcal{C}, \mathcal{C})$  is exactly the same as the *pyramidal width* defined in [8] as shown in [9].

**Definition 2.2** (diameter). Let  $\mathbb{R}^n$  be endowed with a norm  $\|\cdot\|$  and let  $\mathcal{C} \subseteq \mathbb{R}^n$  be a polytope. The *diameter* of  $\mathcal{C}$  is defined as

$$\text{diam}(\mathcal{C}) := \max_{\mathbf{x}, \mathbf{y} \in \mathcal{C}} \|\mathbf{x} - \mathbf{y}\|.$$

## 3 Products

Theorem 3.1 below gives a generic identity for the facial distance of the cartesian product of  $k$  polytopes in terms of the facial distances of the individual components. Throughout this section let  $n_1, \dots, n_k \in \mathbb{N}$  be fixed and assume that the norms on  $\mathbb{R}^{n_i}$  and on  $\mathbb{R}^{n_1 + \dots + n_k} = \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k}$  are related via

$$\|(\mathbf{x}_1, \dots, \mathbf{x}_k)\| = \|(\|\mathbf{x}_1\|, \dots, \|\mathbf{x}_k\|)\|$$

for some norm  $\|\cdot\|$  in  $\mathbb{R}^k$ , and some norms  $\|\cdot\|$  on each of the  $\mathbb{R}^{n_i}$ ,  $i = 1, \dots, k$ . Observe that  $\|\cdot\| = \|\cdot\|_2$  in the special case when each  $\mathbb{R}^{n_i}$  and  $\mathbb{R}^{n_1 + \dots + n_k}$  are endowed with the Euclidean norm.

We shall say that a norm  $\|\cdot\|$  on  $\mathbb{R}^k$  is *reasonable* if  $\|x\| \leq \|y\|$  whenever  $|x| \leq |y|$  componentwise. Notice that many popular norms on  $\mathbb{R}^k$ , including all  $\ell_p$  norms for  $p \in [1, \infty]$  are reasonable. Observe also that  $\|\cdot\|$  is reasonable if and only if  $\|\cdot\|^*$  is reasonable.

The following identity is inspired by and generalizes [7, Theorem 4.5].

**Theorem 3.1.** *Suppose that for  $i = 1, \dots, k$  the sets  $\mathcal{C}_i \subseteq \mathbb{R}^{n_i}$  are polytopes and  $F_i \in \text{faces}(\mathcal{C}_i)$ . If the norm  $\|\cdot\|$  on  $\mathbb{R}^k$  is reasonable then for  $F := F_1 \times \dots \times F_k$  and  $\mathcal{C} := \mathcal{C}_1 \times \dots \times \mathcal{C}_k$  it holds that*

$$\frac{1}{\Phi(F, \mathcal{C})} = \left\| \left( \frac{1}{\Phi(F_1, \mathcal{C}_1)}, \dots, \frac{1}{\Phi(F_k, \mathcal{C}_k)} \right) \right\|^*. \quad (3.1)$$

We note that the format of Theorem 3.1 is similar to that of the following simpler and intuitive identity for the diameter.

**Proposition 3.2.** *Suppose that for  $i = 1, \dots, k$  the sets  $\mathcal{C}_i \subseteq \mathbb{R}^{n_i}$  are polytopes and  $\mathcal{C} := \mathcal{C}_1 \times \dots \times \mathcal{C}_k$ . If the norm  $\|\cdot\|$  on  $\mathbb{R}^k$  is reasonable then*

$$\text{diam}(\mathcal{C}) = \|(\text{diam}(\mathcal{C}_1), \dots, \text{diam}(\mathcal{C}_k))\|.$$

*Proof.* This is a straightforward calculation. Since  $\|\cdot\|$  is reasonable, it follows that

$$\begin{aligned} \text{diam}(\mathcal{C}) &= \max_{\mathbf{x}, \mathbf{y} \in \mathcal{C}} \|\mathbf{x} - \mathbf{y}\| \\ &= \max_{\mathbf{x}, \mathbf{y} \in \mathcal{C}} \|(\|\mathbf{x}_1 - \mathbf{y}_1\|, \dots, \|\mathbf{x}_k - \mathbf{y}_k\|)\| \\ &= \left\| \left( \max_{\mathbf{x}_1, \mathbf{y}_1 \in \mathcal{C}_1} \|\mathbf{x}_1 - \mathbf{y}_1\|, \dots, \max_{\mathbf{x}_k, \mathbf{y}_k \in \mathcal{C}_k} \|\mathbf{x}_k - \mathbf{y}_k\| \right) \right\| \\ &= \|(\text{diam}(\mathcal{C}_1), \dots, \text{diam}(\mathcal{C}_k))\|. \end{aligned}$$

□

In contrast to Proposition 3.2, the proof of Theorem 3.1 quite a bit more elaborate. It relies on the following lemma, which is a slight modification of [7, Lemma 4.3].

**Lemma 3.3.** *Suppose  $\mathcal{C} \subseteq \mathbb{R}^n$ ,  $F \in \text{faces}(\mathcal{C})$ . Let*

$$G := \underset{\substack{G' \in \text{faces}(F) \\ G' \subsetneq \mathcal{C}}}{\text{argmin}} \text{dist}(\text{aff}(G'), \mathcal{C}[G'])$$

*Then the point  $\mathbf{p} \in \text{aff}(G)$  attaining  $\text{dist}(\text{aff}(G), \mathcal{C}[G])$  belongs to  $G$ . In particular,*

$$\Phi(F, \mathcal{C}) = \text{dist}(G, \mathcal{C}[G]) = \text{dist}(\text{aff}(G), \mathcal{C}[G]) = \min_{\substack{G' \in \text{faces}(F) \\ G' \subsetneq \mathcal{C}}} \text{dist}(\text{aff}(G'), \mathcal{C}[G']).$$

*Proof.* The proof is nearly identical to the proof of [7, Lemma 4.3]. We include it here since the above statement is a slight modification of this previous result.

Proceed by contradiction. Suppose  $\mathbf{p} \in \text{aff}(G)$  and  $\mathbf{q} \in \mathcal{C}[G]$  are such that  $\|\mathbf{p} - \mathbf{q}\| = \text{dist}(\text{aff}(G), \mathcal{C}[G])$  and  $\mathbf{p} \notin G$ . Let  $\tilde{G}$  be a facet of  $G$  such that  $\text{aff}(\tilde{G})$  separates  $\mathbf{p}$  from  $G$  and let  $\mathbf{v} \in \text{vert}(G) \setminus \text{aff}(\tilde{G})$ . Since  $\text{aff}(\tilde{G})$  has codimension one in  $\text{aff}(G)$  and the points  $\mathbf{p}, \mathbf{v} \in \text{aff}(G)$  are in opposite sides of this hyperplane, it follows that the segment  $[p, v]$  must

intersect  $\text{aff}(\tilde{G})$ . That is,  $\tilde{\mathbf{p}} := \lambda \mathbf{p} + (1 - \lambda) \mathbf{v} \in \text{aff}(\tilde{G})$  for some  $\lambda \in (0, 1)$ . Therefore  $\tilde{\mathbf{q}} := \lambda \mathbf{q} + (1 - \lambda) \mathbf{v} \in \mathcal{C}[\tilde{G}]$  and  $\|\tilde{\mathbf{p}} - \tilde{\mathbf{q}}\| = \lambda \|\mathbf{p} - \mathbf{q}\|$  and consequently

$$\text{dist}(\tilde{G}, \mathcal{C}[\tilde{G}]) \leq \|\tilde{\mathbf{p}} - \tilde{\mathbf{q}}\| = \lambda \|\mathbf{p} - \mathbf{q}\| < \text{dist}(\text{aff}(G), \mathcal{C}[G]).$$

This contradicts the choice of  $G$ . Thus it must be the case that  $\mathbf{p} \in G$ .

For the second statement, observe that since  $\mathbf{p} \in G$  it follows that

$$\Phi(F, \mathcal{C}) \leq \text{dist}(G, \mathcal{C}[G]) = \text{dist}(\text{aff}(G), \mathcal{C}[G]) = \min_{\substack{G' \in \text{faces}(F) \\ G' \subseteq \mathcal{C}}} \text{dist}(\text{aff}(G'), \mathcal{C}[G']).$$

Since the above right-most expression is evidently less than or equal to  $\Phi(F, \mathcal{C})$ , it follows that all of the above expressions are the same.  $\square$

*Proof of Theorem 3.1.* For  $i = 1, \dots, k$  let  $G_i \in \text{faces}(F_i)$  be such that

$$\delta_i := \Phi(F_i, \mathcal{C}_i) = \text{dist}(G_i, \mathcal{C}_i[G_i]).$$

Then  $G := G_1 \times \dots \times G_k \in \text{faces}(F)$ . For  $i \in [k]$  let  $\mathbf{p}_i \in G_i, \mathbf{q}_i \in \mathcal{C}_i[G_i]$  be such that  $\|\mathbf{p}_i - \mathbf{q}_i\| = \delta_i$ . First we show that the left-hand-side in (3.1) is larger than or equal than the right-hand-side. To that end, it suffices to construct  $\mathbf{p} \in G, \mathbf{q} \in \mathcal{C}[G]$  such that

$$\|\mathbf{p} - \mathbf{q}\| = \frac{1}{\|1/\delta\|^*} \quad (3.2)$$

for  $1/\delta := (1/\delta_1, \dots, 1/\delta_k) \in \mathbb{R}^k$ . Indeed, if (3.2) holds for some  $\mathbf{p} \in G, \mathbf{q} \in \mathcal{C}[G]$  then

$$\frac{1}{\Phi(F, \mathcal{C})} \geq \frac{1}{\text{dist}(G, \mathcal{C}[G])} \geq \|1/\delta\|^* = \left\| \left( \frac{1}{\Phi(F_1, \mathcal{C}_1)}, \dots, \frac{1}{\Phi(F_k, \mathcal{C}_k)} \right) \right\|^*. \quad (3.3)$$

The construction of  $\mathbf{p} \in G$  and  $\mathbf{q} \in \mathcal{C}[G]$  is as follows. Let

$$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k), \mathbf{q} = \lambda_1 \mathbf{q}^1 + \dots + \lambda_k \mathbf{q}^k$$

for  $\mathbf{q}^1, \dots, \mathbf{q}^k \in \mathcal{C}[G]$  and  $\lambda \in \Delta_{k-1}$  that we next describe. For  $i \in [k]$  let  $\mathbf{q}^i \in \mathcal{C}[G]$  be the vector obtained by changing the  $i$ -th block of  $\mathbf{p} \in G$  from  $\mathbf{p}_i \in G_i$  to  $\mathbf{q}_i \in \mathcal{C}_i[G_i]$ . In other words,

$$\mathbf{q}^1 = (\mathbf{q}_1, \mathbf{p}_2, \dots, \mathbf{p}_k), \mathbf{q}^2 = (\mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_3, \dots, \mathbf{p}_k), \dots, \mathbf{q}^k = (\mathbf{p}_1, \dots, \mathbf{p}_{k-1}, \mathbf{q}_k).$$

By definition of the dual norm,

$$\|1/\delta\|^* = \max_{\|z\| \leq 1} \langle 1/\delta, z \rangle.$$

Thus, after rescaling, there exists a  $\mathbf{z} \in \mathbb{R}^k$  such that  $\langle 1/\delta, \mathbf{z} \rangle = 1$  and  $\|\mathbf{z}\| = 1/\|1/\delta\|^*$ . Since the norm  $\|\cdot\|$  is reasonable and  $1/\delta > 0$  it follows that  $\mathbf{z} \geq 0$ . For  $i \in [k]$ , let  $\lambda_i := z_i/\delta_i$ . Since  $\mathbf{z} \geq 0$  and  $1/\delta > 0$  it follows that  $\lambda \geq 0$ . Also  $\mathbf{z} = (\lambda_1 \delta_1, \dots, \lambda_k \delta_k)$ . Finally,  $\lambda \in \Delta_{k-1}$  because

$$1 = \langle 1/\delta, \mathbf{z} \rangle = \sum_{i=1}^k z_i/\delta_i = \sum_{i=1}^k \lambda_i.$$

We next show that  $\mathbf{p}, \mathbf{q}$  satisfy (3.2). Indeed,

$$\|\mathbf{p} - \mathbf{q}\| = \|(\lambda_1(\mathbf{p}_1 - \mathbf{q}_1), \dots, \lambda_k(\mathbf{p}_k - \mathbf{q}_k))\| = \|(\lambda_1\delta_1, \dots, \lambda_k\delta_k)\| = \|\mathbf{z}\| = \frac{1}{\|1/\delta\|^*}.$$

Thus (3.2) holds and consequently so does (3.3).

Next we show that the left-hand-side in (3.1) is less than or equal than the right-hand-side. To that end, it suffices to show that for all  $G' \in \text{faces}(F)$ ,  $G' \subsetneq \mathcal{C}$  it holds that

$$\text{dist}(\text{aff}(G'), \mathcal{C}[G']) \geq \frac{1}{\|1/\delta\|^*}. \quad (3.4)$$

Indeed, if (3.4) holds for all  $G' \in \text{faces}(F)$ ,  $G' \subsetneq \mathcal{C}$  then Lemma 3.3 yields

$$\begin{aligned} \frac{1}{\Phi(F, \mathcal{C})} &= \frac{1}{\min_{\substack{G' \in \text{faces}(F) \\ G' \subsetneq \mathcal{C}}} \text{dist}(\text{aff}(G'), \mathcal{C}[G'])} \\ &\leq \|1/\delta\|^* \\ &= \left\| \left( \frac{1}{\Phi(F_1, \mathcal{C}_1)}, \dots, \frac{1}{\Phi(F_k, \mathcal{C}_k)} \right) \right\|^*. \end{aligned} \quad (3.5)$$

Suppose  $G' \in \text{faces}(F)$ ,  $G' \subsetneq \mathcal{C}$ . Then  $G' = G'_1 \times \dots \times G'_k$  for some  $G'_i \in \text{faces}(F_i)$ ,  $i = 1, \dots, k$ . Since  $G' \subsetneq \mathcal{C}$ , we must have  $G'_i \subsetneq \mathcal{C}_i$  for some  $i = 1, \dots, k$ . Without loss of generality assume  $G'_i \subsetneq \mathcal{C}_i$  for  $i = 1, \dots, j$  with  $1 \leq j \leq k$ . For each  $i = 1, \dots, j$  let  $\delta'_i = \text{dist}(\text{aff}(G'_i), \mathcal{C}_i[G'_i])$  and let  $\mathbf{a}_i \in \mathbb{R}^{n_i}$  be such that  $\|\mathbf{a}_i\|^* = 1$ ,  $\langle \mathbf{a}_i, \mathbf{x} \rangle$  is constant on  $\text{aff}(G'_i)$ , and  $\langle \mathbf{a}_i, \mathbf{x} - \mathbf{y} \rangle \geq \delta'_i$  for all  $\mathbf{x} \in \text{aff}(G'_i)$ ,  $\mathbf{y} \in \mathcal{C}_i[G'_i]$ . Let

$$\mathbf{a} := \frac{1}{\| (1/\delta'_1, \dots, 1/\delta'_j, 0, \dots, 0) \|} \cdot \left( \frac{\mathbf{a}_1}{\delta'_1}, \dots, \frac{\mathbf{a}_j}{\delta'_j}, \mathbf{0}, \dots, \mathbf{0} \right).$$

The construction of  $\mathbf{a}$  ensures that  $\|\mathbf{a}\|^* = 1$  and  $\langle \mathbf{a}, \mathbf{x} \rangle$  is constant on  $\text{aff}(G')$ . Furthermore, for any  $\mathbf{x} \in \text{aff}(G')$  the minimum of  $\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle$  over  $\mathbf{y} \in \mathcal{C} \setminus G'$  is attained at some  $\mathbf{v} \in \text{vert}(\mathcal{C} \setminus G')$ . Thus  $\mathbf{v}_i \in \mathcal{C}_i \setminus G'_i$  for some  $i = 1, \dots, j$  and consequently

$$\begin{aligned} \min\{\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle \mid \mathbf{y} \in \mathcal{C}[G']\} &\geq \frac{1}{\| (1/\delta'_1, \dots, 1/\delta'_j, 0, \dots, 0) \|} \cdot \frac{\langle \mathbf{a}_i, \mathbf{x}_i - \mathbf{v}_i \rangle}{\delta'_i} \\ &\geq \frac{1}{\| (1/\delta'_1, \dots, 1/\delta'_j, 0, \dots, 0) \|}. \end{aligned}$$

In particular, since  $\|\mathbf{a}\|^* = 1$ , it follows that for all  $\mathbf{x} \in \text{aff}(G')$ ,  $\mathbf{y} \in \mathcal{C}[G']$

$$\|\mathbf{x} - \mathbf{y}\| \geq \langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle \geq \frac{1}{\| (1/\delta'_1, \dots, 1/\delta'_j, 0, \dots, 0) \|}.$$

and consequently

$$\text{dist}(\text{aff}(G'), \mathcal{C}[G']) \geq \frac{1}{\| (1/\delta'_1, \dots, 1/\delta'_j, 0, \dots, 0) \|}.$$

Furthermore  $\delta'_i \geq \delta_i = \Phi(F_i, \mathcal{C}_i)$  for  $i = 1, \dots, j$  because  $\delta'_i = \text{dist}(G'_i, \mathcal{C}_i[G'_i])$  and  $G'_i \in \text{faces}(F_i)$  with  $G'_i \neq \mathcal{C}_i$ . Therefore, since  $\|\cdot\|^*$  is reasonable, it follows that

$$\begin{aligned} \text{dist}(\text{aff}(G'), \mathcal{C}[G']) &\geq \frac{1}{\|\langle 1/\delta'_1, \dots, 1/\delta'_j, 0, \dots, 0 \rangle\|^*} \\ &\geq \frac{1}{\|\langle 1/\delta_1, \dots, 1/\delta_j, 0, \dots, 0 \rangle\|^*} \\ &\geq \frac{1}{\|1/\boldsymbol{\delta}\|^*}. \end{aligned}$$

Thus (3.4) holds and consequently so does (3.5). □

## 4 Facet-exposing directions

Theorem 4.1 below gives bounds on the facial distance of a polytope in term of *facet-exposing directions* and *gaps* described next. In contrast to the facial distance, the latter quantities are often easily computable for polytopes with few facets or with a suitable polyhedral description as some of our examples in Section 6 illustrate.

Suppose  $\mathcal{C} \subseteq \mathbb{R}^n$  is a polytope and  $F \in \text{faces}(\mathcal{C})$ ,  $F \subsetneq \mathcal{C}$ . We say that  $\mathbf{a} \in \mathbb{R}^n$  is a *face-exposing direction* of  $F$  if  $\|\mathbf{a}\|^* = 1$  and

$$F = \text{argmin}\{\langle \mathbf{a}, \mathbf{x} \rangle \mid \mathbf{x} \in \mathcal{C}\}.$$

When this is the case, define the *gap* of  $\mathbf{a}$  as follows

$$\sigma(\mathbf{a}) := \min\{\langle \mathbf{a}, \mathbf{v} \rangle \mid \mathbf{v} \in \text{vert}(\mathcal{C}) \setminus F\} = \min\{\langle \mathbf{a}, \mathbf{x} \rangle \mid \mathbf{x} \in \mathcal{C}[F]\} > 0.$$

Observe that  $\text{dist}(\text{aff}(F), \mathcal{C}[F]) \geq \sigma(\mathbf{a})$  for any face exposing direction of  $F$  and the equality is attained for a judiciously chosen face-exposing direction. We will say *facet-exposing direction* in lieu of face-exposing direction in the special case when  $F$  is a facet.

Next, suppose  $F_1, \dots, F_m$  are the facets of  $\mathcal{C}$ . For  $F \in \text{faces}(\mathcal{C})$  let  $I(F) := \{j \mid F \subseteq F_j\}$ . Observe that  $I(F) \neq \emptyset$  whenever  $F \subsetneq \mathcal{C}$  and in that case

$$F = \bigcap_{j \in I(F)} F_j.$$

Theorem 4.1 below gives bounds on the facial distance of  $\mathcal{C}$  in terms of facet-exposing directions of the facets of  $\mathcal{C}$ .

**Theorem 4.1.** *Suppose  $\mathcal{C} \subseteq \mathbb{R}^n$  is a polytope and  $F_1, \dots, F_m$  are the facets of  $\mathcal{C}$  with facet-exposing directions  $\mathbf{a}_j$ ,  $j = 1, \dots, m$ . Let  $\sigma_j := \sigma(\mathbf{a}_j)$ ,  $j = 1, \dots, m$ . Then for  $F \in \text{faces}(\mathcal{C})$  with  $F \subsetneq \mathcal{C}$*

$$\text{dist}(\text{aff}(F), \mathcal{C}[F]) \geq \frac{1}{\|\sum_{j \in I(F)} \mathbf{a}_j / \sigma_j\|^*}.$$

Consequently for  $F \in \text{faces}(\mathcal{C})$  it holds that

$$\Phi(F, \mathcal{C}) \geq \min_{\substack{G \in \text{faces}(F) \\ G \subsetneq \mathcal{C}}} \frac{1}{\|\sum_{j \in I(G)} \mathbf{a}_j / \sigma_j\|^*}.$$

*Proof.* The choice of  $\mathbf{a}_j$  and  $\sigma_j$  implies that  $\|\mathbf{a}_j\|^* = 1$ ,  $\langle \mathbf{a}_j, \mathbf{x} \rangle$  is constant on  $\text{aff}(F_j)$ , and  $\langle \mathbf{a}_j, \mathbf{x} - \mathbf{y} \rangle \geq \sigma_j$  for all  $\mathbf{x} \in \text{aff}(F_j)$  and all  $\mathbf{y} \in \text{conv}(\text{vert}(\mathcal{C}) \setminus F_j)$ . Let

$$\mathbf{a} := \sum_{j \in I(F)} \frac{a_j}{\sigma_j}.$$

It follows that  $\langle \mathbf{a}, \mathbf{x} \rangle$  is constant on  $\text{aff}(F) = \bigcap_{j \in I(F)} \text{aff}(F_j)$ . Furthermore, for  $\mathbf{v} \in \text{vert}(\mathcal{C}) \setminus \text{aff}(F) = \text{vert}(\mathcal{C}) \setminus \bigcap_{j \in I(F)} \text{aff}(F_j)$  there must exist at least one  $j \in I(F)$  such that  $\mathbf{v} \notin F_j$  and thus for all  $\mathbf{x} \in \text{aff}(F) = \bigcap_{j \in I(F)} \text{aff}(F_j)$  it holds that

$$\langle \mathbf{a}, \mathbf{x} - \mathbf{v} \rangle \geq \frac{\langle \mathbf{a}_j, \mathbf{x} - \mathbf{v} \rangle}{\sigma_j} \geq 1.$$

Since this holds for all  $\mathbf{x} \in \text{aff}(F)$  and  $\mathbf{v} \in \text{vert}(\mathcal{C}) \setminus \text{aff}(F)$  it follows that

$$\text{dist}(\text{aff}(F), \mathcal{C}[F]) \geq \frac{1}{\|\mathbf{a}\|^*} \geq \frac{1}{\|\sum_{j \in I(F)} \mathbf{a}_j / \sigma_j\|^*}.$$

□

For  $F \in \text{faces}(\mathcal{C})$  let  $J(F) := \{j \mid F \cap F_j \neq \emptyset\}$ . Observe that  $J(F) \neq \emptyset$  for all  $F \in \text{faces}(\mathcal{C})$ . In particular,  $J(\mathcal{C}) = \{1, \dots, m\}$ . The following corollary readily follows from Theorem 4.1 and the triangle inequality.

**Corollary 4.2.** *Suppose  $\mathcal{C} \subseteq \mathbb{R}^n$  is a polytope with facets  $F_1, \dots, F_m$  and  $\delta_i := \text{dist}(\text{aff}(F_i), \mathcal{C}[F_i])$  for  $i = 1, \dots, m$ . Then for each  $F \in \text{faces}(\mathcal{C})$  it holds that*

$$\Phi(F, \mathcal{C}) \geq \frac{1}{\sum_{j \in J(F)} 1/\delta_j}.$$

In particular,

$$\frac{1}{\sum_{j=1}^m 1/\delta_j} \leq \Phi(\mathcal{C}) \leq \min\{\delta_j \mid j = 1, \dots, m\}.$$

## 5 Linear transformations

Throughout this section we shall assume that  $\mathbb{R}^m$  and  $\mathbb{R}^n$  be endowed with some norms,  $\mathcal{C} \subseteq \mathbb{R}^n$  be a polytope,  $A \in \mathbb{R}^{m \times n}$ , and  $\mathcal{D} = A(\mathcal{C}) \subseteq \mathbb{R}^m$ . A natural question is how the facial distances of  $\mathcal{D}$  and of  $\mathcal{C}$  are related. The following lemma provides some tools to tackle that question.

**Lemma 5.1.** *Let  $\mathcal{C} \subseteq \mathbb{R}^n$  be a polytope,  $A \in \mathbb{R}^{m \times n}$ ,  $\mathcal{D} = A(\mathcal{C})$ . Then the following hold:*

- (a)  $\text{vert}(\mathcal{D}) \subseteq A(\text{vert}(\mathcal{C}))$ .
- (b) If  $G \in \text{faces}(\mathcal{D})$  then  $A^{-1}(G) \cap \mathcal{C} \in \text{faces}(\mathcal{C})$ .
- (c) If  $G \in \text{faces}(\mathcal{D})$  and  $G \subsetneq \mathcal{D}$  then  $F := A^{-1}(G) \cap \mathcal{C} \subsetneq \mathcal{C}$  and

$$\text{dist}(G, \mathcal{D}[G]) = \min\{\|A(\mathbf{y} - \mathbf{x})\| \mid \mathbf{x} \in F, \mathbf{y} \in \mathcal{C}[F]\}. \quad (5.1)$$

*Proof.* (a) Observe that  $\mathcal{D} = A(\mathcal{C}) = A(\text{conv}(\text{vert}(\mathcal{C}))) = \text{conv}(A(\text{vert}(\mathcal{C})))$ . Therefore  $\text{vert}(\mathcal{D}) \subseteq A(\text{vert}(\mathcal{C}))$  since  $\text{vert}(\mathcal{D})$  is the smallest set of points whose convex hull is  $\mathcal{D}$ .

(b) Since  $G \in \text{faces}(\mathcal{D})$ , there exists  $\mathbf{a} \in \mathbb{R}^m$  such that

$$G = \underset{\mathbf{y} \in \mathcal{D}}{\text{argmin}} \langle \mathbf{a}, \mathbf{y} \rangle = A \left( \underset{\mathbf{x} \in \mathcal{C}}{\text{argmin}} \langle \mathbf{a}, A\mathbf{x} \rangle \right) = A \left( \underset{\mathbf{x} \in \mathcal{C}}{\text{argmin}} \langle A^\top \mathbf{a}, \mathbf{x} \rangle \right).$$

It thus follows that

$$F = A^{-1}(G) \cap \mathcal{C} = \underset{\mathbf{x} \in \mathcal{C}}{\text{argmin}} \langle A^\top \mathbf{a}, \mathbf{x} \rangle,$$

and consequently  $F \in \text{faces}(\mathcal{C})$ .

(c) Equation (5.1) is an immediate consequence of parts (a) and (b). Indeed, parts (a) and (b) imply that for any two  $\mathbf{z}, \mathbf{w} \in \mathbb{R}^m$  it holds that  $\mathbf{z} \in G$  and  $\mathbf{w} \in \mathcal{D}[G]$  if and only if  $\mathbf{z} = A\mathbf{x}$  and  $\mathbf{w} = A\mathbf{y}$  for some  $\mathbf{x} \in F$  and  $\mathbf{y} \in \text{conv}(\text{vert}(\mathcal{C}) \setminus F)$ . Equation (5.1) thus follows.  $\square$

**Corollary 5.2.** *Let  $\mathcal{C} \subseteq \mathbb{R}^n$  be a polytope,  $A \in \mathbb{R}^{n \times n}$  be non-singular, and  $G \in \text{faces}(\mathcal{D})$ . Then for  $F := A^{-1}(G) \cap \mathcal{C} \in \text{faces}(\mathcal{C})$  it holds that*

$$\frac{\Phi(F, \mathcal{C})}{\|A^{-1}\|} \leq \Phi(G, \mathcal{D}) \leq \|A\| \Phi(F, \mathcal{C}). \quad (5.2)$$

*Proof.* Equation (5.1) in Lemma 5.1 implies that for all  $G' \in \text{faces}(G)$  with  $G' \neq \mathcal{C}$  and  $F' := A^{-1}(G') \cap \mathcal{C}$

$$\begin{aligned} \text{dist}(G', \mathcal{D}[G']) &= \min\{\|A(\mathbf{y} - \mathbf{x})\| \mid \mathbf{x} \in F', \mathbf{y} \in \mathcal{C}[F']\} \\ &= \frac{1}{\|A^{-1}\|} \cdot \min\{\|A^{-1}\| \cdot \|A(\mathbf{y} - \mathbf{x})\| \mid \mathbf{x} \in F', \mathbf{y} \in \mathcal{C}[F']\} \\ &\geq \frac{1}{\|A^{-1}\|} \min\{\|\mathbf{y} - \mathbf{x}\| \mid \mathbf{x} \in F', \mathbf{y} \in \mathcal{C}[F']\} \\ &= \frac{\text{dist}(F', \mathcal{C}[F'])}{\|A^{-1}\|} \\ &\geq \frac{\Phi(F, \mathcal{C})}{\|A^{-1}\|}. \end{aligned}$$

Since this holds for all  $G' \in \text{faces}(G)$  with  $G' \neq \mathcal{C}$ , the lower bound of (5.2) follows.

Similarly,

$$\begin{aligned} \text{dist}(G', \mathcal{D}[G']) &= \min\{\|A(\mathbf{y} - \mathbf{x})\| \mid \mathbf{x} \in F', \mathbf{y} \in \mathcal{C}[F']\} \\ &\leq \|A\| \cdot \min\{\|\mathbf{y} - \mathbf{x}\| \mid \mathbf{x} \in F', \mathbf{y} \in \mathcal{C}[F']\} \\ &= \|A\| \text{dist}(F', \mathcal{C}[F']) \end{aligned}$$

Since  $A$  is nonsingular, the map  $G' \mapsto F' := A^{-1}(G') \cap \mathcal{C}$  is a bijection between  $\text{faces}(G)$  and  $\text{faces}(F)$ . Thus,

$$\begin{aligned} \min_{\substack{G' \in \text{faces}(G) \\ G' \neq G}} \text{dist}(G', \mathcal{D}[G']) &\leq \min_{\substack{G' \in \text{faces}(G) \\ G' \neq G}} \|A\| \text{dist}(A^{-1}(G') \cap \mathcal{C}, \mathcal{C}[A^{-1}(G') \cap \mathcal{C}]) \\ &= \|A\| \min_{\substack{F' \in \text{faces}(F) \\ F' \neq F}} \text{dist}(F', \mathcal{C}[F']) \\ &= \|A\| \Phi(F, \mathcal{C}). \end{aligned}$$

Thus, the upper bound of (5.2) follows. □

## 6 Examples

The examples below illustrate how the above calculus rules can be used to bound or compute the facial distance. For the sake of intuition, throughout this section we assume that the norm in the underlying space is the canonical Euclidean norm  $\|\cdot\|_2$ .

**Example 6.1** (standard simplex). Suppose  $n > 1$ , and

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq 0, \langle \mathbf{1}, \mathbf{x} \rangle = 1\}.$$

The vertices of  $\mathcal{C}$  are the unitary vectors  $\mathbf{e}_i$ ,  $i = 1, \dots, n$ . The facets of  $\mathcal{C}$  are

$$F_i = \{\mathbf{x} \in \mathcal{C} \mid x_i = 0\} \text{ for } i = 1, \dots, n.$$

Thus the directions  $\mathbf{e}_i$ ,  $i = 1, \dots, n$  are face-exposing directions of the facets of  $\mathcal{C}$  and  $\sigma(\mathbf{e}_i) = 1$  for  $i = 1, \dots, n$ . For any  $F \in \text{faces}(\mathcal{C})$  with  $F \neq \mathcal{C}$  we have

$$I(F) = \{i \in [n] \mid x_i = 0 \text{ for all } \mathbf{x} \in F\}$$

and so  $|I(F)| \leq n - 1$ . Therefore Theorem 4.1 implies that for all  $F \in \text{faces}(\mathcal{C})$  with  $F \subsetneq \mathcal{C}$

$$\text{dist}(\text{aff}(F), \mathcal{C}[F]) \geq \frac{1}{\|\sum_{j \in I(F)} \mathbf{e}_j\|^*} = \frac{1}{\sqrt{|I(F)|}} \geq \frac{1}{\sqrt{n-1}},$$

and consequently Lemma 3.3 implies that

$$\Phi(\mathcal{C}) \geq \frac{1}{\sqrt{n-1}}.$$

We note that this lower bound is within a small constant of the exact value of  $\Phi(\mathcal{C})$  previously computed in [8, 9]:

$$\Phi(\mathcal{C}) = \begin{cases} \frac{2}{\sqrt{n}} & \text{if } n \text{ is even} \\ \frac{2}{\sqrt{n-1/n}} & \text{if } n \text{ is odd.} \end{cases}$$

**Example 6.2** ( $\ell_1$  unit ball). Suppose  $n \geq 1$  and

$$\mathcal{D} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_1 \leq 1\} = A(\mathcal{C})$$

where  $A = \begin{bmatrix} I & -I \end{bmatrix} \in \mathbb{R}^{n \times 2n}$  and  $\mathcal{C} \subseteq \mathbb{R}^{2n}$  is the standard simplex in  $\mathbb{R}^{2n}$ , that is,

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{2n} \mid \mathbf{x} \geq 0, \langle \mathbf{1}, \mathbf{x} \rangle = 1\}.$$

The set of vertices of  $\mathcal{D}$  is  $\{\pm \mathbf{e}_j \mid j \in [n]\}$ .

Let  $G \in \text{faces}(\mathcal{D})$ . We next compute  $\text{dist}(\text{aff}(F), \mathcal{D}[G])$ . Since  $\mathcal{D}$  and the Euclidean norm are invariant under changes of signs and permutations of any set of coordinates, we can assume without loss of generality that  $G$  is of the form  $\text{conv}\{\mathbf{e}_1, \dots, \mathbf{e}_k\} = \{\mathbf{z} \in \mathcal{D} : \mathbf{z}_{k+1:n} = \mathbf{0}\}$  for some  $k \leq n$ . It thus follows that  $F = A^{-1}(G) = \{\mathbf{x} \in \mathcal{C} \mid \mathbf{x}_{k+1:2n} = \mathbf{0}\}$  and hence  $\mathcal{C}[F] = \{\mathbf{y} \in \mathcal{C} \mid \mathbf{y}_{1:k} = \mathbf{0}\}$ . Lemma 5.1 implies that

$$\begin{aligned} \text{dist}(G, \mathcal{D}[G]) &= \min\{\|A(\mathbf{x} - \mathbf{y})\| \mid \mathbf{x} \in F, \mathbf{y} \in \mathcal{C}[F]\} \\ &= \min\{\| [\mathbf{x}_{1:n} - \mathbf{x}_{n+1:2n}] - [\mathbf{y}_{1:n} - \mathbf{y}_{n+1:2n}] \| \mid \mathbf{x} \in F, \mathbf{y} \in \mathcal{C}[F]\} \\ &= \min\left\{\left\| \begin{bmatrix} \mathbf{x}_{1:k} \\ \mathbf{0}_{k+1:n} \end{bmatrix} - \begin{bmatrix} -\mathbf{y}_{n+1:n+k} \\ \mathbf{y}_{k+1:n} - \mathbf{y}_{n+k+1:2n} \end{bmatrix} \right\| \mid \mathbf{x} \in F, \mathbf{y} \in \mathcal{C}[F]\right\} \\ &= \min\left\{\left\| \begin{bmatrix} \mathbf{x}_{1:k} + \mathbf{y}_{n+k+1:2n} \\ \mathbf{y}_{k+1:n} - \mathbf{y}_{n+k+1:2n} \end{bmatrix} \right\| \mid \mathbf{x} \in F, \mathbf{y} \in \mathcal{C}[F]\right\}. \end{aligned}$$

A simple calculation then shows that

$$\text{dist}(G, \mathcal{D}[G]) = \begin{cases} \min\{\|\mathbf{x}_{1:k}\| \mid \mathbf{x} \in F\} = \frac{1}{\sqrt{k}} & \text{if } k < n \\ \min\{\|\mathbf{x}_{1:n} + \mathbf{y}_{n+1:2n}\| \mid \mathbf{x} \in F, \mathbf{y} \in \mathcal{C}[F]\} = \frac{2}{\sqrt{n}} & \text{if } k = n. \end{cases}$$

Consequently,

$$\Phi(\mathcal{D}) = \frac{1}{\sqrt{n-1}}.$$

**Example 6.3** ( $\ell_\infty$  unit ball). Suppose  $n \geq 1$  and

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_\infty \leq 1\} = \mathcal{C}_1 \times \dots \times \mathcal{C}_n$$

where  $\mathcal{C}_i = [-1, 1]$  for each  $i = 1, \dots, n$ . In this case  $\Phi(\mathcal{C}_i) = 2$  for  $i = 1, \dots, n$ . Thus Theorem 3.1 implies that

$$\Phi(\mathcal{C}) = \frac{1}{\| [1/2, \dots, 1/2] \|_2} = \frac{1}{\sqrt{n}/2} = \frac{2}{\sqrt{n}}.$$

**Example 6.4** (Birkhoff polytope). Suppose  $\mathbb{R}^{n \times n}$  is endowed with the Frobenius norm and  $\mathcal{C}$  is the set of  $n \times n$  doubly stochastic matrices, that is,

$$\mathcal{C} = \left\{ \mathbf{x} \in \mathbb{R}^{n \times n} \mid \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_{ij} = 1 \text{ and } \sum_{j=1}^n x_{ij} = 1 \text{ for } i, j \in [n] \right\}.$$

We will focus on the case  $n \geq 3$  since for the cases  $n = 1$  and  $n = 2$  the Birkhoff polytope is respectively a singleton and a segment and thus of limited interest.

It is known that the vertices of  $\mathcal{C}$  are the set of permutation matrices and the facets of  $\mathcal{C}$  are of the form

$$F_{ij} = \{\mathbf{x} \in \mathcal{C} \mid x_{ij} = 0\} \text{ for } (ij) \in [n] \times [n].$$

Thus the directions  $\mathbf{e}_i \mathbf{e}_j^\top$  for  $(ij) \in [n] \times [n]$  are face-exposing directions of the facets of  $\mathcal{C}$  and  $\sigma(\mathbf{e}_i \mathbf{e}_j^\top) = 1$  for  $(ij) \in [n] \times [n]$ . For any  $F \in \text{faces}(\mathcal{C})$  with  $F \neq \mathcal{C}$  we have

$$I(F) = \{(ij) \in [n] \times [n] \mid x_{ij} = 0 \text{ for all } \mathbf{x} \in F\}.$$

and so  $|I(F)| \leq n^2 - 1$ . Therefore Theorem 4.1 implies that for all  $F \in \text{faces}(\mathcal{C})$  with  $F \subsetneq \mathcal{C}$

$$\text{dist}(\text{aff}(F), \mathcal{C}[F]) \geq \frac{1}{\|\sum_{j \in I(F)} \mathbf{e}_j\|^*} = \frac{1}{\sqrt{|I(F)|}} \geq \frac{1}{\sqrt{n^2 - 1}},$$

and consequently Lemma 3.3 implies that

$$\Phi(\mathcal{C}) \geq \frac{1}{\sqrt{n^2 - 1}}.$$

The bounds in Example 6.1 and Example 6.4 generalize to the broader class of *simplex-like* polytopes introduced by Garber and Meshi [6].

**Definition 6.5.** A polytope  $\mathcal{C} \subseteq \mathbb{R}^n$  is *simplex-like* if it is of the form  $\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$  and  $\text{vert}(\mathcal{C}) \subseteq \{0, 1\}^n$ .

Example 6.7 relies on the following property of polytopes with a polyhedral description.

**Proposition 6.6.** Suppose  $\mathcal{C} \subseteq \mathbb{R}^n$  is a polytope of the form  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ . Then each facet of  $\mathcal{C}$  has a face-exposing direction of the form  $\mathbf{e}_j$  for some  $j \in [n]$ .

*Proof.* Suppose  $F \subseteq \mathcal{C}$  is a facet. Let

$$J(F) := \{i \in [n] \mid \langle \mathbf{e}_i, \mathbf{x} \rangle = 0 \text{ for all } \mathbf{x} \in F\}.$$

In other words,  $J(F) \subseteq [n]$  is the set of inequalities that are active at all points in  $F$ . Since  $F$  is a facet of  $\mathcal{C}$ , in particular  $F \neq \mathcal{C}$  and thus there exists  $j \in J(F)$  such that  $\langle \mathbf{e}_j, \mathbf{x} \rangle > 0$  for some  $\mathbf{x} \in \mathcal{C} \setminus F$ . Since  $\mathcal{C} \subseteq \{\mathbf{x} \in \mathcal{C} \mid \langle \mathbf{e}_j, \mathbf{x} \rangle \geq 0\}$ , to show that  $\mathbf{e}_j$  is a facet-exposing direction for  $F$  it suffices to show that

$$F = \{\mathbf{x} \in \mathcal{C} \mid \langle \mathbf{e}_j, \mathbf{x} \rangle = 0\}. \quad (6.1)$$

The construction of  $J(F)$  and the fact that  $j \in J(F)$  guarantee that  $F \subseteq \{\mathbf{x} \in \mathcal{C} \mid \langle \mathbf{e}_j, \mathbf{x} \rangle = 0\}$ . Furthermore, the choice of  $j \in J(F)$  guarantees that  $\{\mathbf{x} \in \mathcal{C} \mid \langle \mathbf{e}_j, \mathbf{x} \rangle = 0\}$  is a proper face of  $\mathcal{C}$ . Since  $F \subseteq \{\mathbf{x} \in \mathcal{C} \mid \langle \mathbf{e}_j, \mathbf{x} \rangle = 0\}$  and  $F$  is a facet of  $\mathcal{C}$ , these two sets must be the same, that is, (6.1) holds.  $\square$

**Example 6.7.** Suppose  $\mathcal{C} \subseteq \mathbb{R}^n$  is a *simplex-like* polytope. Proposition 6.6 implies that each facet of  $\mathcal{C}$  has a face-exposing direction of the form  $\mathbf{e}_j$  for some  $j \in [n]$  and in addition  $\sigma(\mathbf{e}_j) = 1$  because  $\text{vert}(\mathcal{C}) \subseteq \{0, 1\}^n$ . For  $F \in \text{faces}(\mathcal{C})$ , the size of the set  $I(F)$  is the number of zeros of the most dense point in  $F$ . In other words,  $|I(F)|$  is a measure of the sparsity of  $F$ . In particular,  $|I(F)| \leq n$  for all  $F \in \text{faces}(\mathcal{C})$ . Therefore Theorem 4.1 implies that for all  $F \in \text{faces}(\mathcal{C})$  with  $F \subsetneq \mathcal{C}$

$$\text{dist}(\text{aff}(F), \mathcal{C}[F]) \geq \frac{1}{\|\sum_{j \in I(F)} \mathbf{e}_j\|^*} = \frac{1}{\sqrt{|I(F)|}} \geq \frac{1}{\sqrt{n}},$$

and consequently Lemma 3.3 implies that

$$\Phi(\mathcal{C}) \geq \frac{1}{\sqrt{n}}.$$

## 7 Upper bound conjecture on the facial distance

Some straightforward calculations shows that in the examples in Section 6 the ratio  $\Phi(\mathcal{C})/\text{diam}(\mathcal{C})$  is bounded below as follows:  $\sqrt{2}/\sqrt{n}$  for the standard simplex,  $1/(2\sqrt{n})$  for the  $\ell_1$  unit ball in  $\mathbb{R}^n$ ,  $1/n$  for the  $\ell_\infty$  ball in  $\mathbb{R}^n$ ,  $1/(n\sqrt{2n})$  for the  $n \times n$  Birkhoff polytope, and  $1/n$  for a simplex-like polytope in  $\mathbb{R}^n$ . In addition, lower complexity results for the Frank-Wolfe algorithm suggest that  $\Phi(\mathcal{C})/\text{diam}(\mathcal{C})$  should be bounded above. The following conjecture thus arises.

**Conjecture 7.1.** *Suppose  $\mathcal{C}$  is a polytope of dimension  $n$  and the ambient space is endowed with the Euclidean norm. Then*

$$\frac{\Phi(\mathcal{C})}{\text{diam}(\mathcal{C})} \leq \frac{C}{\sqrt{n}}.$$

for some constant  $C > 0$ .

In addition to the examples discussed in Section 6, the next two partial results provide support for this conjecture.

**Proposition 7.2.** *Suppose  $\mathcal{C}$  is an  $n$ -dimensional simplex and  $n$  is an odd number. Then*

$$\frac{\Phi(\mathcal{C})}{\text{diam}(\mathcal{C})} \leq \frac{2\sqrt{2}}{\sqrt{n+1}}. \quad (7.1)$$

*Proof.* Since  $\mathcal{C}$  is an  $n$ -dimensional simplex, it is of the form  $\mathcal{C} = \text{conv}\{\mathbf{x}_1, \dots, \mathbf{x}_{n+1}\}$  where  $\mathbf{x}_1, \dots, \mathbf{x}_{n+1}$  are affine independent. Furthermore, by scaling and shifting if needed, without loss of generality we can assume that  $\text{diam}(\mathcal{C}) = 1$  and  $\mathbf{x}_1 + \dots + \mathbf{x}_{n+1} = \mathbf{0}$ .

Since  $\mathcal{C}$  is a simplex, it follows that every  $F \in \text{faces}(\mathcal{C})$  is of the form  $F = \text{conv}\{\mathbf{x}_i \mid i \in I\}$  for some nonempty  $I \subseteq [n+1]$  and in that case  $\mathcal{C}[F] = \text{conv}\{\mathbf{x}_j \mid j \in [n+1] \setminus I\}$ . Therefore to prove (7.1) it suffices to show the following claim.

**Claim.** There exists  $I \subseteq [n+1]$  with  $|I| = (n+1)/2$  such that

$$\left\| \sum_{i \in I} \mathbf{x}_i - \sum_{j \in [n+1] \setminus I} \mathbf{x}_j \right\|^2 \leq 2(n+1). \quad (7.2)$$

Indeed, the above claim implies that for  $F = \text{conv}\{\mathbf{x}_i \mid i \in I\}$  it holds that

$$\begin{aligned} \text{dist}(F, \mathcal{C}[F]) &\leq \left\| \frac{2}{n+1} \sum_{i \in I} \mathbf{x}_i - \frac{2}{n+1} \sum_{j \in [n+1] \setminus I} \mathbf{x}_j \right\| \\ &\leq \frac{2}{n+1} \cdot \sqrt{2(n+1)} \\ &\leq \frac{2\sqrt{2}}{\sqrt{n+1}}, \end{aligned}$$

and hence

$$\frac{\Phi(\mathcal{C})}{\text{diam}(\mathcal{C})} \leq \text{dist}(F, \mathcal{C}[F]) \leq \frac{2\sqrt{2}}{\sqrt{n+1}}.$$

We next prove the claim. To that end, we proceed in two steps.

*Step 1:* Because  $\text{diam}(\mathcal{C}) = 1$  and  $\mathbf{x}_1 + \cdots + \mathbf{x}_{n+1} = \mathbf{0}$ , it holds that

$$\|\mathbf{x}_i\| \leq 1 \text{ for all } i \in [n+1] \text{ and } \sum_{i \neq j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \leq 0. \quad (7.3)$$

For the first inequality in (7.3), proceed by contradiction. Suppose  $\|\mathbf{x}_i\| > 1$  for some  $i \in [n+1]$ . Then

$$\|\mathbf{x}_i - \frac{1}{n} \sum_{j \neq i} \mathbf{x}_j\|^2 \geq \|\mathbf{x}_i\|^2 - \frac{2}{n} \sum_{j \neq i} \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \quad (7.4)$$

But  $0 = \|\mathbf{x}_i + \sum_{j \neq i} \mathbf{x}_j\|^2 = \|\mathbf{x}_i\|^2 + \|\sum_{j \neq i} \mathbf{x}_j\|^2 + 2 \sum_{j \neq i} \langle \mathbf{x}_i, \mathbf{x}_j \rangle$  and so  $\sum_{j \neq i} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \leq -\|\mathbf{x}_i\|^2 < 0$ . Thus (7.4) implies that

$$\|\mathbf{x}_i - \frac{1}{n} \sum_{j \neq i} \mathbf{x}_j\|^2 > \|\mathbf{x}_i\|^2 > 1$$

which contradicts the assumption  $\text{diam}(\mathcal{C}) = 1$ .

For the second inequality in (7.3), observe that  $0 = \|\sum_i \mathbf{x}_i\|^2 = \sum_i \|\mathbf{x}_i\|^2 + 2 \sum_{i \neq j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle$  and so  $\sum_{i \neq j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle = -\sum_i \|\mathbf{x}_i\|^2 / 2 \leq 0$ .

*Step 2:* Let

$$\delta := \min_{\substack{I \subseteq [n+1] \\ |I| = (n+1)/2}} \left\| \sum_{i \in I} \mathbf{x}_i - \sum_{j \in [n+1] \setminus I} \mathbf{x}_j \right\|^2.$$

we next show that  $\delta \leq 2(n+1)$  and thus (7.2) must hold for some  $I \subseteq [n+1]$  with  $|I| = (n+1)/2$ .

To that end, observe that by adding over all  $I \subseteq [n+1]$  with  $|I| = (n+1)/2$  we get

$$\begin{aligned} \binom{n+1}{(n+1)/2} \delta &\leq \sum_{\substack{I \subseteq [n+1] \\ |I| = (n+1)/2}} \left\| \sum_{i \in I} \mathbf{x}_i - \sum_{j \in [n+1] \setminus I} \mathbf{x}_j \right\|^2 \\ &\leq \binom{n+1}{(n+1)/2} \left( \sum_i \|\mathbf{x}_i\|^2 - 2 \sum_{i \neq j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right) \\ &= \binom{n+1}{(n+1)/2} \left( 2 \sum_i \|\mathbf{x}_i\|^2 - \left\| \sum_i \mathbf{x}_i \right\|^2 \right) \\ &= 2 \binom{n+1}{(n+1)/2} \sum_i \|\mathbf{x}_i\|^2. \end{aligned} \quad (7.5)$$

The second step above holds because  $\sum_{i \neq j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \leq 0$ . Inequality (7.5) in turn implies that

$$\delta \leq 2 \sum_i \|\mathbf{x}_i\|^2 \leq 2(n+1).$$

□

**Proposition 7.3.** Suppose that for  $i = 1, \dots, k$  the sets  $\mathcal{C}_i \subseteq \mathbb{R}^{n_i}$  are polytopes and let  $\mathcal{C} := \mathcal{C}_1 \times \dots \times \mathcal{C}_k$ . If each  $\mathbb{R}^{n_i}$  and  $\mathbb{R}^{n_1 + \dots + n_k}$  are endowed with the Euclidean norms then

$$\frac{\text{diam}(\mathcal{C})}{\Phi(\mathcal{C})} = \|(\text{diam}(\mathcal{C}_1), \dots, \text{diam}(\mathcal{C}_k))\| \cdot \left\| \left( \frac{1}{\Phi(\mathcal{C}_1)}, \dots, \frac{1}{\Phi(\mathcal{C}_k)} \right) \right\|. \quad (7.6)$$

In particular if each  $\frac{\Phi(\mathcal{C}_i)}{\text{diam}(\mathcal{C}_i)} \leq \frac{C}{\sqrt{n_i}}$  then  $\frac{\Phi(\mathcal{C})}{\text{diam}(\mathcal{C})} \leq \frac{C}{\sqrt{n_1 + \dots + n_k}}$ .

*Proof.* Identity (7.6) is an immediate consequence of Proposition 3.2 and Theorem 3.1. The second statement follows from Cauchy-Schwarz inequality and the concavity of the square root function:

$$\begin{aligned} \frac{\text{diam}(\mathcal{C})}{\Phi(\mathcal{C})} &= \|(\text{diam}(\mathcal{C}_1), \dots, \text{diam}(\mathcal{C}_k))\| \cdot \left\| \left( \frac{1}{\Phi(\mathcal{C}_1)}, \dots, \frac{1}{\Phi(\mathcal{C}_k)} \right) \right\| \\ &\geq \sum_{i=1}^k \frac{\text{diam}(\mathcal{C}_i)}{\Phi(\mathcal{C}_i)} \\ &\geq \sum_{i=1}^k \frac{\sqrt{n_i}}{C} \\ &\geq \frac{\sqrt{n_1 + \dots + n_k}}{C}. \end{aligned}$$

□

In spite of the above two promising results, we have not been able to prove Conjecture 7.1. It is natural to try to leverage Proposition 7.2 via an induction argument: The initial inductive step would be Proposition 7.2. This proposition shows the conjecture for any polytope  $\mathcal{C}$  of dimension  $\dim(\mathcal{C}) = n$  with the minimal number of vertices, namely  $n + 1$ . The induction would then carry on if we could show that the ratio  $\Phi(\mathcal{C})/\text{diam}(\mathcal{C})$  does not increase as the number of vertices of  $\mathcal{C}$  increases while  $\dim(\mathcal{C})$  remains constant. Unfortunately, the second step of this argument fails as illustrated by the following example.

**Example 7.4.** Suppose  $0 \leq \epsilon \leq 1/4$  and consider the following distorted  $\ell_1$  unit ball in  $\mathbb{R}^3$ :

$$\mathcal{C} = \text{conv}\{\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_2 - \epsilon\mathbf{e}_3, -\mathbf{e}_2 - \epsilon\mathbf{e}_3, \mathbf{e}_3, -\mathbf{e}_3\}.$$

Observe that this is exactly the  $\ell_1$  unit ball when  $\epsilon = 0$ . In that case Example 6.2 shows that  $\Phi(\mathcal{C}) = 1/\sqrt{2}$ . It thus follows that  $\Phi(\mathcal{C}) \approx 1/\sqrt{2}$  when  $0 < \epsilon \ll 1/4$ . A more careful calculation shows that when  $0 < \epsilon \ll 1/4$  it holds that

$$\Phi(\mathcal{C}) = \text{dist}(F, \mathcal{C}[F]) = \frac{1 - \epsilon}{\sqrt{2}}$$

for  $F = [\mathbf{e}_1, -\mathbf{e}_3] \in \text{faces}(\mathcal{C})$ . Now consider the polytope obtained after we remove one of the vertices of  $\mathcal{C}$ :

$$\mathcal{C}' = \text{conv}\{\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_2 - \epsilon\mathbf{e}_3, -\mathbf{e}_2 - \epsilon\mathbf{e}_3, \mathbf{e}_3\}.$$

A simple calculation shows that when  $0 < \epsilon \ll 1/4$  it holds that

$$\Phi(\mathcal{C}') = \text{dist}(G, \mathcal{C}'[G]) = \epsilon \ll \Phi(\mathcal{C})$$

for  $G = [\mathbf{e}_2 - \epsilon\mathbf{e}_3, -\mathbf{e}_2 - \epsilon\mathbf{e}_3] \in \text{faces}(\mathcal{C}')$ . As a consequence  $\Phi(\mathcal{C}')/\text{diam}(\mathcal{C}') \ll \Phi(\mathcal{C})/\text{diam}(\mathcal{C})$ . This example shows that  $\Phi(\mathcal{C})/\text{diam}(\mathcal{C})$  may increase drastically when the number of vertices increases while  $\dim(\mathcal{C})$  remains constant.

We also have not been able to prove Conjecture 7.1 when  $\mathcal{C}$  is the Birkhoff polytope. We hope Conjecture 7.1 and the other developments in this paper motivate future work on a better understanding of the facial distance.

**Remark 7.5.** Example 7.4 disproves a conjecture formulated by Lacoste-Julien and Jaggi in [8, Sect 3.1]. They conjectured that the pyramidal width (facial distance) is non-increasing when a vertex is added to a polytope, as long as the vertices of the old polytope remain vertices of the new one. The very recent manuscript [13] also provides a counterexample to the same conjecture.

## References

- [1] Immanuel M Bomze, Francesco Rinaldi, and Damiano Zeffiro. Frank–Wolfe and friends: a journey into projection-free first-order optimization methods. *Annals of Operations Research*, 343(2):607–638, 2024.
- [2] Gábor Braun, Alejandro Carderera, Cyrille W Combettes, Hamed Hassani, Amin Karbasi, Aryan Mokhtari, and Sebastian Pokutta. *Conditional gradient methods: From core principles to AI applications*. SIAM, 2025.
- [3] Antoine Deza, Zhongyuan Liu, and Lionel Pournin. Small kissing polytopes. *Vietnam Journal of Mathematics*, 53(4):901–913, 2025.
- [4] Antoine Deza, Zhongyuan Liu, and Lionel Pournin. Kissing polytopes in dimension 3. *Experimental Mathematics*, pages 1–9, 2026.
- [5] Antoine Deza, Shmuel Onn, Sebastian Pokutta, and Lionel Pournin. Kissing polytopes. *SIAM Journal on Discrete Mathematics*, 38(4):2643–2664, 2024.
- [6] Dan Garber and Ofer Meshi. Linear-memory and decomposition-invariant linearly convergent conditional gradient algorithm for structured polytopes. In *Proceedings of the International Conference on Neural Information Processing Systems*, pages 1009–1017. PMLR, 2016.
- [7] Gabriele Iommazzo, David Martínez-Rubio, Francisco Criado, Elias Wirth, and Sebastian Pokutta. Linear convergence of the Frank-Wolfe algorithm over product polytopes. *arXiv preprint arXiv:2505.11259*, 2025.
- [8] Simon Lacoste-Julien and Martin Jaggi. On the global linear convergence of Frank-Wolfe optimization variants. In *Proceedings of Advances in Neural Information Processing Systems*, pages 496–504, 2015.
- [9] Javier Pena and Daniel Rodriguez. Polytope conditioning and linear convergence of the Frank–Wolfe algorithm. *Mathematics of Operations Research*, 44(1):1–18, 2019.
- [10] Francesco Rinaldi and Damiano Zeffiro. Avoiding bad steps in Frank-Wolfe variants. *Computational Optimization and Applications*, 84(1):225–264, 2023.
- [11] Kazuma K Tsuji, Ken’ichiro Tanaka, and Sebastian Pokutta. Pairwise conditional gradients without swap steps and sparser kernel herding. In *Proceedings of the International Conference on Machine Learning*, pages 21864–21883. PMLR, 2022.

- [12] Elias Wirth, Javier Pena, and Sebastian Pokutta. Fast convergence of Frank-Wolfe algorithms on polytopes. *Mathematics of Operations Research*, 51(2):1463–1485, 2026.
- [13] Jinze Zhao. Pyramidal width can increase under vertex insertion. *arXiv preprint arXiv:2607.29555*, 2026.
- [14] Renbo Zhao. New analysis of an away-step Frank–Wolfe method for minimizing log-homogeneous barriers. *Mathematics of Operations Research*, 51(1):35–59, 2026.