



The Iterates of the Frank-Wolfe Algorithm May Not Converge

Jérôme Bolte, Cyrille W. Combettes, Edouard Pauwels

► To cite this version:

Jérôme Bolte, Cyrille W. Combettes, Edouard Pauwels. The Iterates of the Frank-Wolfe Algorithm May Not Converge. 2022. hal-03579585

HAL Id: hal-03579585

<https://hal.science/hal-03579585>

Preprint submitted on 18 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Iterates of the Frank-Wolfe Algorithm May Not Converge

Jérôme Bolte¹

jerome.bolte@ut-capitole.fr

Cyrille W. Combettes¹

cyrille.combettes@tse-fr.eu

Édouard Pauwels^{2 3}

edouard.pauwels@irit.fr

¹ Toulouse School of Economics, Université Toulouse 1 Capitole, Toulouse, France

² Centre National de la Recherche Scientifique, France

³ Institut de Recherche en Informatique de Toulouse, Université Toulouse 3 Paul Sabatier, Toulouse, France

Abstract

The Frank-Wolfe algorithm is a popular method for minimizing a smooth convex function f over a compact convex set \mathcal{C} . While many convergence results have been derived in terms of function values, hardly nothing is known about the convergence behavior of the sequence of iterates $(x_t)_{t \in \mathbb{N}}$. Under the usual assumptions, we design several counterexamples to the convergence of $(x_t)_{t \in \mathbb{N}}$, where f is d -time continuously differentiable, $d \geq 2$, and $f(x_t) \rightarrow \min_{\mathcal{C}} f$. Our counterexamples cover the cases of open-loop, closed-loop, and line-search step-size strategies. We do not assume *misspecification* of the linear minimization oracle and our results thus hold regardless of the points it returns, demonstrating the fundamental pathologies in the convergence behavior of $(x_t)_{t \in \mathbb{N}}$.

1 Introduction

The Frank-Wolfe algorithm [11], a.k.a. conditional gradient algorithm [23], addresses the optimization problem

$$\min_{x \in \mathcal{C}} f(x),$$

where $\mathcal{C} \subset \mathbb{R}^n$ is a nonempty compact convex set, $f: \mathcal{D} \rightarrow \mathbb{R}$ is a convex function with Lipschitz-continuous gradient, and $\mathcal{D} \subset \mathbb{R}^n$ is an open convex set containing \mathcal{C} . It does not require projections onto \mathcal{C} to ensure the feasibility of its iterates and uses linear minimizations over \mathcal{C} instead, which can be significantly cheaper to compute [6]. Another advantage is that it may generate iterates that are sparse with respect to the vertices of \mathcal{C} [4]. Both properties have led to many applications of the Frank-Wolfe algorithm [22, 15, 20, 17, 24, 7, 9].

With a suitable step-size strategy, the sequence of function values $(f(x_t))_{t \in \mathbb{N}}$ converges to $\min_{\mathcal{C}} f$ at a rate $\mathcal{O}(1/t)$ [23, 10, 16], and this result has been extensively complemented: lower bounds have been established [3, 16, 21], faster rates under additional assumptions have been derived [23, 14, 12, 18], and variants speeding up the algorithm have been developed [19, 13, 2, 5]. However, it is not known whether the sequence of iterates $(x_t)_{t \in \mathbb{N}}$ converges or not, apart from the trivial case where the solution set $\arg \min_{\mathcal{C}} f$ is a singleton [23, 10]. This is in contrast with other popular convex optimization methods such as the gradient descent, coordinate descent, mirror descent, proximal point, forward-backward, or Douglas-Rachford algorithm, for which convergence results for the sequence of iterates have been proved [8] or disproved [1].

Note that counterexamples to the convergence of $(x_t)_{t \in \mathbb{N}}$ could be designed trivially by assuming that the linear minimization oracle is *misspecified* (Definition 2.1), i.e., that it does not necessarily return the same output each time the same input is entered. A simple illustration is given by the minimization of $f = 0$ over $\mathcal{C} = [0, 1]$ using an open-loop strategy (i)–(i'). As misspecification allows to move towards 0 or towards 1 at each iteration, $(x_t)_{t \in \mathbb{N}}$ may follow any arbitrary trajectory in $[0, 1]$. In \mathbb{R}^2 , one can also easily build counterexamples where $(x_t)_{t \in \mathbb{N}}$ does not converge while remaining out of $\arg \min_{\mathcal{C}} f$, for any step-size strategy (i)–(iii): consider, e.g., minimizing $f = \text{dist}(\cdot, [(-1/2, 0), (1/2, 0)])^2$ over $\mathcal{C} = \text{conv}\{(-2, 1/4), (-1, 0), (1, 0), (0, 1)\}$. These counterexamples are artificial and somehow vain, as they rely on the misspecification assumption and a very specific adversarial choice of the points returned by the oracle. Our work does not assume misspecification of the oracle to demonstrate that $(x_t)_{t \in \mathbb{N}}$ may not converge and relies instead on the joint geometry of \mathcal{C} and f .

Contributions. We show that the sequence of iterates $(x_t)_{t \in \mathbb{N}}$ generated by the Frank-Wolfe algorithm does not converge in general. We design several instances of \mathcal{C} and f satisfying the usual assumptions, i.e., \mathcal{C} is compact and convex and f is convex with Lipschitz-continuous gradient, for which the sequence $(x_t)_{t \in \mathbb{N}}$ does not converge, while $f(x_t) \rightarrow \min_{\mathcal{C}} f$. We cover all step-size strategies for which convergence results for the function values have been established, i.e., the so-called open-loop, closed-loop, and line-search strategies. Furthermore, the functions f designed are d -time continuously differentiable, where $d \geq 2$ is arbitrarily large. We use key results from [1] and do not assume *misspecification* of the linear minimization oracle, thus demonstrating the fundamental pathologies in the convergence behavior of $(x_t)_{t \in \mathbb{N}}$.

2 Preliminaries

2.1 Notation and definitions

We consider the Euclidean space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ equipped with the Euclidean norm $\|\cdot\|$. Given a convex set $\mathcal{C} \subset \mathbb{R}^n$ and $x \in \mathcal{C}$, the tangent cone to \mathcal{C} at x is $\mathcal{T}_{\mathcal{C}}x = \text{cl}\{\lambda(y - x) \mid \lambda > 0, y \in \mathcal{C}\}$ and the normal cone to \mathcal{C} at x is $\mathcal{N}_{\mathcal{C}}x = \{u \in \mathbb{R}^n \mid \forall y \in \mathcal{C}, \langle y - x, u \rangle \leq 0\}$. The diameter and the set of vertices of \mathcal{C} are denoted by $\text{diam } \mathcal{C}$ and $\text{vert } \mathcal{C}$ respectively. The Hausdorff distance between two nonempty compact sets $\mathcal{A}, \mathcal{B} \subset \mathbb{R}^n$ is denoted by $\text{dist}(\mathcal{A}, \mathcal{B})$. The canonical vectors in \mathbb{R}^2 are $e_1 = (1, 0)$ and $e_2 = (0, 1)$. The unit sphere in \mathbb{R}^2 is $\mathbb{S} = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$. Given three points $A, B, C \in \mathbb{R}^2$, the angle at B formed by the rays $[BA)$ and $[BC)$ is denoted by \widehat{ABC} .

2.2 The Frank-Wolfe algorithm

The Frank-Wolfe algorithm [11] is presented in Algorithm 2. It assumes access to a linear minimization oracle (LMO, Oracle 1) solving linear minimization problems over the constraint set \mathcal{C} .

Oracle 1 LMO

Input: Vector $u \in \mathbb{R}^n$.

Output: Point $v \in \arg \min_{v \in \mathcal{C}} \langle v, u \rangle$.

The linear minimization oracle does not ensure a priori that the same output is returned each time the same input is entered. When $\arg \min_{v \in \mathcal{C}} \langle v, u \rangle$ is not a singleton, it makes a choice on which element to return, based on, e.g., a proximity score, an ordering of vertices, or a random seed. This leads to the notion of *misspecification* for the oracle.

Definition 2.1. The linear minimization oracle LMO is specified if $u \in \mathbb{R}^n \mapsto \text{LMO}(u)$ is a well-defined function and satisfies $\text{LMO}(u) = \text{LMO}(u')$ for all positively colinear $u, u' \in \mathbb{R}^n$. Otherwise, it is misspecified.

Algorithm 2 Frank-Wolfe

Input: Start point $x_0 \in \mathcal{C}$, step-size strategy $(\gamma_t)_{t \in \mathbb{N}} \in [0, 1]^{\mathbb{N}}$.

Output: Sequence of iterates $(x_t)_{t \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$.

```

1: for  $t = 0$  to  $T - 1$  do
2:    $v_t \leftarrow \text{LMO}(\nabla f(x_t))$ 
3:    $x_{t+1} \leftarrow (1 - \gamma_t)x_t + \gamma_tv_t$ 
4: end for

```

Remark 2.2. (i) The step-size strategy $(\gamma_t)_{t \in \mathbb{N}}$ need not be numerically defined before running Algorithm 2. More precisely, we may say that a sequence $(x_t)_{t \in \mathbb{N}} \in (\mathbb{R}^n)^{\mathbb{N}}$ is generated by the Frank-Wolfe algorithm if $x_0 \in \mathcal{C}$ and for all $t \in \mathbb{N}$, there exist $v_t \in \arg \min_{v \in \mathcal{C}} \langle v, \nabla f(x_t) \rangle$ and $\gamma_t \in [0, 1]$ such that $x_{t+1} = (1 - \gamma_t)x_t + \gamma_tv_t$. (ii) When $\arg \min_{v \in \mathcal{C}} \langle v, \nabla f(x_t) \rangle$ is not a singleton, the choice for v_t is made by the linear minimization oracle. If the oracle is misspecified, then $(v_t)_{t \in \mathbb{N}}$, and thus $(x_t)_{t \in \mathbb{N}}$, may not be uniquely determined by x_0 and $(\gamma_t)_{t \in \mathbb{N}}$.

Since x_{t+1} is generated by convex combination of x_t and v_t (Line 3), it is feasible and there is no need for a projection step back onto \mathcal{C} . Note that $x_{t+1} = x_t + \gamma_t(v_t - x_t)$ is obtained by moving from x_t towards v_t . Furthermore, assuming the linear minimization oracle returns only vertices of \mathcal{C} , at most one new vertex is added to the convex decomposition of x_t to obtain x_{t+1} , making the sequence $(x_t)_{t \in \mathbb{N}}$ sparse with respect to the vertices of \mathcal{C} .

There are mainly four step-size strategies that have been considered and for which the rate at which $f(x_t) \rightarrow \min_{\mathcal{C}} f$ has been established (Theorem 2.4). These are:

(i) an open-loop strategy

$$\forall t \in \mathbb{N}, \gamma_t = \frac{1}{t+1};$$

(i') another open-loop strategy

$$\forall t \in \mathbb{N}, \gamma_t = \frac{2}{t+2};$$

(ii) the closed-loop strategy

$$\forall t \in \mathbb{N}, \gamma_t = \min \left\{ \frac{\langle x_t - v_t, \nabla f(x_t) \rangle}{L \|x_t - v_t\|^2}, 1 \right\};$$

(iii) the line-search strategy

$$\forall t \in \mathbb{N}, \gamma_t \in \arg \min_{\gamma \in [0, 1]} f(x_t + \gamma(v_t - x_t)).$$

Assumption 2.3 collects the general set of assumptions for the Frank-Wolfe algorithm.

Assumption 2.3. Let $\mathcal{C} \subset \mathbb{R}^n$ be a nonempty compact convex set and $f: \mathcal{D} \rightarrow \mathbb{R}$ be a convex function with L -Lipschitz-continuous gradient, where $L > 0$ and $\mathcal{D} \subset \mathbb{R}^n$ is an open convex set containing \mathcal{C} .

Theorem 2.4 ([23, 10, 16]). Consider Assumption 2.3. Then the Frank-Wolfe algorithm satisfies for all $t \geq 1$,

$$f(x_t) - \min_C f \leq \begin{cases} L(\text{diam } C)^2(1 + \ln t)/(2t) & \text{if } (\gamma_t)_{t \in \mathbb{N}} \text{ follows (i);} \\ 2L(\text{diam } C)^2/(t+2) & \text{if } (\gamma_t)_{t \in \mathbb{N}} \text{ follows (i');} \\ 4L(\text{diam } C)^2/(t+2) & \text{if } (\gamma_t)_{t \in \mathbb{N}} \text{ follows (ii);} \\ 4L(\text{diam } C)^2/(t+2) & \text{if } (\gamma_t)_{t \in \mathbb{N}} \text{ follows (iii).} \end{cases}$$

3 The Frank-Wolfe algorithm may not converge

We turn on to the main results of the paper. We design several instances of C and f satisfying Assumption 2.3 and for which the sequence of iterates $(x_t)_{t \in \mathbb{N}}$ generated by the Frank-Wolfe algorithm using a usual step-size strategy (i)–(iii) does not converge, while Theorem 2.4 still holds. The strength of these counterexamples is to be valid against specified linear minimization oracles, demonstrating that nonconvergence is intrinsic to the Frank-Wolfe algorithm in the smooth convex optimization setting. We will use a key result from [1] which we restate in Theorem 3.1. It shows that we can design a convex function, arbitrarily smooth, from polygonal sketches, and that we can further choose its gradient directions at specific points from a set of directions described by the geometry of the sketches.

Given a polytope $\mathcal{P} \subset \mathbb{R}^2$ and a vertex $V \in \text{vert } \mathcal{P}$, we denote by $\mathcal{K}_{\mathcal{P}}V = \mathcal{N}_{\mathcal{P}}V \cap (-\mathcal{T}_{\mathcal{P}}V)$ the cone of admissible directions at V for \mathcal{P} .

Theorem 3.1 ([1]). In \mathbb{R}^2 , let $(\mathcal{P}_\ell)_{\ell \in \mathbb{N}}$ be a sequence of polytopes, $(u_k)_{k \in \mathbb{N}} \in \mathbb{S}^{\mathbb{N}}$ be a sequence of unit vectors, $(V_k)_{k \in \mathbb{N}} \in (\bigcup_{\ell \in \mathbb{N}} \text{vert } \mathcal{P}_\ell)^{\mathbb{N}}$ be a sequence of vertices of the polytopes, and $(\delta_\ell)_{\ell \in \mathbb{N}} \in]0, 1]^{\mathbb{N}}$ be such that

- (i) for all $k \in \mathbb{N}$, there exists $\ell \in \mathbb{N}$ such that $u_k \in \mathcal{K}_{\mathcal{P}_\ell}V_k$;
- (ii) for all $\ell \in \mathbb{N}$ and $\lambda \in [1 - \delta_\ell, 1 + \delta_\ell]$, $\lambda\mathcal{P}_{\ell+1} \subset \text{int } \mathcal{P}_\ell$;
- (iii) for all $\ell \in \mathbb{N}$, $0 \in \text{int } \mathcal{P}_\ell$.

Then for any open convex set $\mathcal{D} \subset \mathbb{R}^2$ and $d \in \mathbb{N}$ such that $\mathcal{P}_0 \subset \mathcal{D}$ and $d \geq 2$ respectively, there exist a d -time continuously differentiable convex function $f: \mathcal{D} \rightarrow \mathbb{R}$ and a sequence $(\eta_\ell)_{\ell \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ such that for all $k, \ell \in \mathbb{N}$,

- (iv) $\eta_{\ell+1} < \eta_\ell$;
- (v) $\mathcal{P}_\ell \subset \{x \in \mathbb{R}^2 \mid f(x) \leq \eta_\ell\}$;
- (vi) $f(V) = \eta_\ell$ for all $V \in \text{vert } \mathcal{P}_\ell$;
- (vii) $\text{dist}(\mathcal{P}_\ell, \{x \in \mathbb{R}^2 \mid f(x) \leq \eta_\ell\}) \leq \delta_\ell$;
- (viii) $\{x \in \mathbb{R}^2 \mid f(x) \leq \eta_\ell\}$ has positive curvature;
- (ix) $\nabla f(V_k)$ is positively colinear to u_k ;
- (x) $\nabla^2 f$ is positive definite outside of $\arg \min_{\mathcal{D}} f = \bigcap_{\ell \in \mathbb{N}} \mathcal{P}_\ell$.

Remark 3.2. In Theorem 3.1, ∇f is Lipschitz-continuous on \mathcal{D} by the mean value theorem, since $d \geq 2$. Thus, f satisfies Assumption 2.3.

Lemma 3.3. Consider the function f and all other variables defined in Theorem 3.1. Let $\mathcal{C} \subset \mathcal{D}$ be a compact convex set and consider minimizing f over \mathcal{C} using the Frank-Wolfe algorithm with the line-search strategy (iii). Let $t \in \mathbb{N}$ and suppose that there exists $k \in \mathbb{N}$ such that $V_k \in [x_t, v_t]$ and $\langle v_t - x_t, u_k \rangle = 0$. Then $x_{t+1} = V_k$.

Proof. We have $x_{t+1} = (1 - \gamma_t)x_t + \gamma_tv_t$ where $\gamma_t \in \arg \min_{\gamma \in [0,1]} f(x_t + \gamma(v_t - x_t))$, i.e., $x_{t+1} \in [x_t, v_t]$ and $\langle v_t - x_t, \nabla f(x_{t+1}) \rangle = 0$. By Theorem 3.1(ix), u_k and $\nabla f(V_k)$ are positively colinear so $\langle v_t - x_t, \nabla f(V_k) \rangle = 0$. By Theorem 3.1(viii), $\{x \in \mathbb{R}^2 \mid f(x) \leq f(V_k)\}$ has positive curvature, which yields $x_{t+1} = V_k$. \square

Lemma 3.4. Consider the function f and all other variables defined in Theorem 3.1. Let $k_1, k_2, \ell_1, \ell_2 \in \mathbb{N}$, $\ell_1 > \ell_2$, and $u \in \mathbb{S}$, and suppose that $V_{k_1} \in \mathcal{P}_{\ell_1}$ and $V_{k_2} \in \mathcal{P}_{\ell_2}$ are such that 0, V_{k_1} , V_{k_2} are aligned and $u_{k_1} = u_{k_2} = u$. Then $\nabla f(x)$ is positively colinear to u for all $x \in [V_{k_1}, V_{k_2}]$.

Proof. Let $\mathcal{L}_1 = \{x \in \mathbb{R}^2 \mid f(x) \leq f(V_{k_1})\}$ and $\mathcal{L}_2 = \{x \in \mathbb{R}^2 \mid f(x) \leq f(V_{k_2})\}$. By Theorem 3.1(viii), they have positive curvature. Let $n_1: \text{bd } \mathcal{L}_1 \rightarrow \mathbb{S}$ and $n_2: \text{bd } \mathcal{L}_2 \rightarrow \mathbb{S}$ be their Gauss maps respectively, which are diffeomorphisms that map a point on the boundary of the set to the normal to the set at that point [25, Sec. 2.5]. By Theorem 3.1(ix), $n_1(V_{k_1}) = u = n_2(V_{k_2})$. Let $x' \in [V_{k_1}, V_{k_2}]$ and $\mathcal{L} = \{x \in \mathbb{R}^2 \mid f(x) \leq f(x')\}$. The construction from [1] leading to Theorem 3.1 gives $\mathcal{L} = \alpha\mathcal{L}_1 + \beta\mathcal{L}_2$, where $\alpha, \beta \geq 0$ and $\alpha + \beta > 0$. Thus, by alignment with 0, $x' = \alpha V_{k_1} + \beta V_{k_2}$, and, by [1, Lem. 2], \mathcal{L} has positive curvature. Let $n: \text{bd } \mathcal{L} \rightarrow \mathbb{S}$ be its Gauss map. By [1, Lem. 2] again, $n^{-1}(u) = \alpha n_1^{-1}(u) + \beta n_2^{-1}(u) = \alpha V_{k_1} + \beta V_{k_2} = x'$, i.e., $n(x') = u$. \square

3.1 Counterexample 1: Line-search strategy

We start with a counterexample for the line-search strategy (iii).

Counterexample 3.5. There exist \mathcal{C} and f , satisfying Assumption 2.3, and $x_0 \in \mathcal{C}$ such that any sequence $(x_t)_{t \in \mathbb{N}}$ generated by Frank-Wolfe algorithm using the line-search strategy (iii) does not converge.

Let $\mathcal{C} = \text{conv}\{(-1, 0), (0, 1), (1, 0)\}$ and f be defined via Theorem 3.1 with $A_0 = (-1/2, 0)$, $B_0 = (-1/4, 3/4)$, $C_0 = (1/4, 3/4)$, $D_0 = (1/2, 0)$, and, for all $k \in \mathbb{N}$,

$$\begin{aligned} A_{k+1} &= \left(-\frac{1}{4} + \frac{\langle B_{k+1}, e_2 \rangle}{\langle B_k, e_2 \rangle} \left(\langle A_k, e_1 \rangle + \frac{1}{4} \right), 0 \right), \\ B_{k+1} &= \left(-\frac{1}{4}, \frac{3}{5} \langle C_k, e_2 \rangle \right), \\ C_{k+1} &= \left(\frac{1}{4}, \frac{3}{5} \langle B_k, e_2 \rangle \right), \\ D_{k+1} &= \left(\frac{1}{4} + \frac{\langle C_{k+1}, e_2 \rangle}{\langle C_k, e_2 \rangle} \left(\langle D_k, e_1 \rangle - \frac{1}{4} \right), 0 \right), \end{aligned}$$

and $\mathcal{P}_k = \text{conv}\{A_k, B_k, C_k, D_k, -D_k, -C_k, -B_k, -A_k\}$. Then \mathcal{C} and f satisfy Assumption 2.3 so Theorem 2.4 holds. The solution set is $\arg \min_{\mathcal{C}} f = \bigcap_{k \in \mathbb{N}} \mathcal{P}_k = [(-1/4, 0), (1/4, 0)]$. The construction is such that for all $k \in \mathbb{N}$, the points B_k , C_{k+1} , $(1, 0)$, and the points C_k , B_{k+1} , $(-1, 0)$, are aligned, and $\widehat{A_k B_k C_k} = \widehat{A_0 B_0 C_0}$ and $\widehat{B_k C_k D_k} = \widehat{B_0 C_0 D_0}$. An illustration is presented in Figure 1.

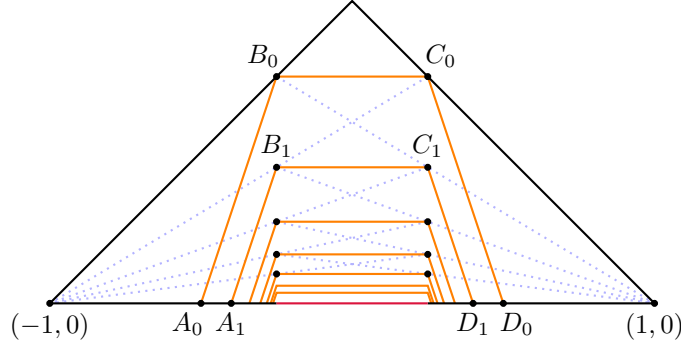


Figure 1: The constraint set (in black), polygonal sketches of the objective function (in orange), and the solution set (in red). For all $k \in \mathbb{N}$, the points $B_k, C_{k+1}, (1, 0)$, and the points $C_k, B_{k+1}, (-1, 0)$, are aligned, and $\widehat{A_k B_k C_k} = \widehat{A_0 B_0 C_0}$ and $\widehat{B_k C_k D_k} = \widehat{B_0 C_0 D_0}$.

The construction allows to choose $(u_k)_{k \in \mathbb{N}}$ such that $u_0 \in \mathcal{K}_{\mathcal{P}_0} C_0$ and, for all $k \geq 1$,

$$\begin{cases} u_k \in \mathcal{K}_{\mathcal{P}_k} C_k \text{ and } \langle C_k - B_{k-1}, u_k \rangle = 0 & \text{if } k \text{ is even;} \\ u_k \in \mathcal{K}_{\mathcal{P}_k} B_k \text{ and } \langle B_k - C_{k-1}, u_k \rangle = 0 & \text{if } k \text{ is odd.} \end{cases}$$

Let $x_0 = C_0$. By induction and using Lemma 3.3, we obtain for all $t \in \mathbb{N}$,

$$x_t = \begin{cases} C_t & \text{if } t \text{ is even;} \\ B_t & \text{if } t \text{ is odd.} \end{cases}$$

Therefore, $\langle x_t, e_1 \rangle = (-1)^t/4$ for all $t \in \mathbb{N}$, so $(x_t)_{t \in \mathbb{N}}$ does not converge. An illustration is presented in Figure 2.

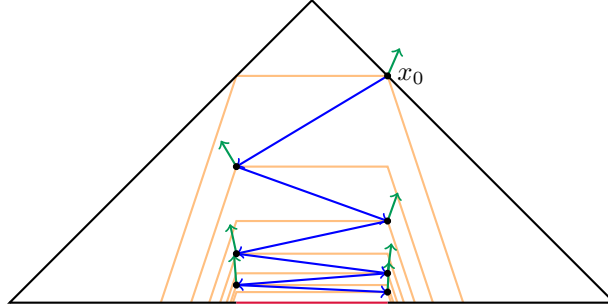


Figure 2: The constraint set (in black), polygonal sketches of the objective function (in orange), the solution set (in red), gradient directions (in green), and the trajectory (in blue) of the sequence $(x_t)_{t \in \mathbb{N}}$ generated by the Frank-Wolfe algorithm with the line-search strategy (iii) starting from x_0 . The abscisse of x_t is $(-1)^t/4$ for all $t \in \mathbb{N}$, so $(x_t)_{t \in \mathbb{N}}$ does not converge.

3.2 Counterexample 2: Line-search strategy and solution set in the interior

In our next counterexample, the solution set $\arg \min_{\mathcal{C}} f$ lies in the interior of the constraint set.

Counterexample 3.6. *There exist \mathcal{C} and f , satisfying Assumption 2.3 and $\arg \min_{\mathcal{C}} f \subset \text{int } \mathcal{C}$, and $x_0 \in \mathcal{C}$ such that any sequence $(x_t)_{t \in \mathbb{N}}$ generated by Frank-Wolfe algorithm using the line-search strategy (iii) does not converge.*

Let $\mathcal{C} = \text{conv}\{(-1, -1), (-1, 1), (1, 1), (1, -1)\}$ and f be defined via Theorem 3.1 with $A_0 = (-1/10, -1)$, $B_0 = (-1, 1/10)$, $C_0 = (1/10, 1)$, $D_0 = (1, -1/10)$, $\lambda_0 = 1$, and, for all $k \in \mathbb{N}$, $\lambda_{k+1} = 110\lambda_k/(90 + 101\lambda_k)$,

$$A_k = \lambda_k A_0, \quad B_k = \lambda_k B_0, \quad C_k = \lambda_k C_0, \quad D_k = \lambda_k D_0,$$

and $\mathcal{P}_k = \text{conv}\{\lambda_k A_0, \lambda_k B_0, \lambda_k C_0, \lambda_k D_0\} = \lambda_k \mathcal{P}_0$. Thus, $(\lambda_k)_{k \in \mathbb{N}}$ is a decreasing sequence that converges to $1/5$. Then \mathcal{C} and f satisfy Assumption 2.3 so Theorem 2.4 holds. The solution set is $\arg \min_{\mathcal{C}} f = \bigcap_{k \in \mathbb{N}} \mathcal{P}_k = (1/5)\mathcal{P}_0$. The construction is such that for all $k \in \mathbb{N}$, the points $A_k, B_{k+1}, (-1, 1)$, the points $B_k, C_{k+1}, (1, 1)$, the points $C_k, D_{k+1}, (1, -1)$, and the points $D_k, A_{k+1}, (-1, -1)$, are aligned. An illustration is presented in Figure 3.

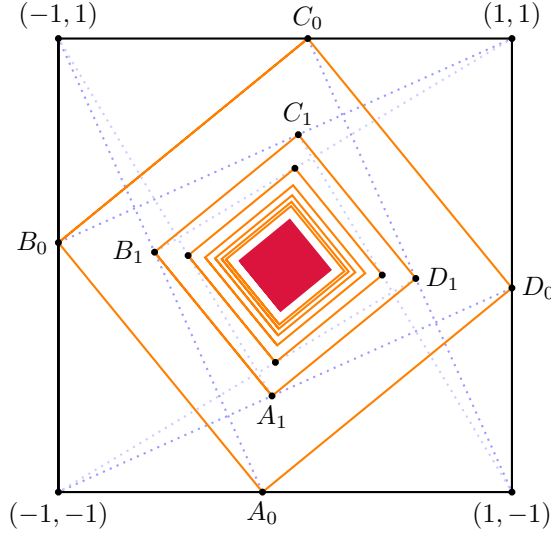


Figure 3: The constraint set (in black), polygonal sketches of the objective function (in orange), and the solution set (in red). For all $k \in \mathbb{N}$, the points $A_k, B_{k+1}, (-1, 1)$, the points $B_k, C_{k+1}, (1, 1)$, the points $C_k, D_{k+1}, (1, -1)$, and the points $D_k, A_{k+1}, (-1, -1)$, are aligned.

The construction allows to choose $(u_k)_{k \in \mathbb{N}}$ such that $u_0 \in \mathcal{K}_{\mathcal{P}_0} A_0$ and, for all $k \geq 1$,

$$\begin{cases} u_k \in \mathcal{K}_{\mathcal{P}_k} A_k \text{ and } \langle A_k - D_{k-1}, u_k \rangle = 0 & \text{if } k \equiv 0 \pmod{4}; \\ u_k \in \mathcal{K}_{\mathcal{P}_k} B_k \text{ and } \langle B_k - A_{k-1}, u_k \rangle = 0 & \text{if } k \equiv 1 \pmod{4}; \\ u_k \in \mathcal{K}_{\mathcal{P}_k} C_k \text{ and } \langle C_k - B_{k-1}, u_k \rangle = 0 & \text{if } k \equiv 2 \pmod{4}; \\ u_k \in \mathcal{K}_{\mathcal{P}_k} D_k \text{ and } \langle D_k - C_{k-1}, u_k \rangle = 0 & \text{if } k \equiv 3 \pmod{4}. \end{cases}$$

Let $x_0 = A_0$. By induction and using Lemma 3.3, we obtain for all $t \in \mathbb{N}$,

$$x_t = \begin{cases} A_t & \text{if } t \equiv 0 \pmod{4}; \\ B_t & \text{if } t \equiv 1 \pmod{4}; \\ C_t & \text{if } t \equiv 2 \pmod{4}; \\ D_t & \text{if } t \equiv 3 \pmod{4}. \end{cases}$$

Therefore, $\|x_{t+1} - x_t\| \geq \sqrt{2}/5$ for all $t \in \mathbb{N}$, where $\sqrt{2}/5$ is the side length of the square $\arg \min_{\mathcal{C}} f = (1/5)\mathcal{P}_0$, so $(x_t)_{t \in \mathbb{N}}$ does not converge. An illustration is presented in Figure 4.

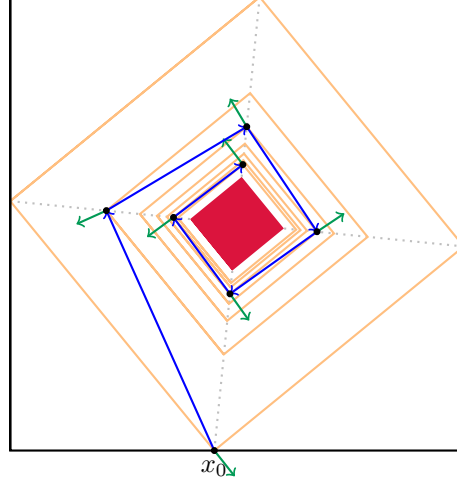


Figure 4: The constraint set (in black), polygonal sketches of the objective function (in orange), the solution set (in red), gradient directions (in green), and the trajectory (in blue) of the sequence $(x_t)_{t \in \mathbb{N}}$ generated by the Frank-Wolfe algorithm with the line-search strategy (iii) starting from x_0 . The sequence $(x_t)_{t \in \mathbb{N}}$ circles around the solution set and does not converge.

3.3 Counterexample 3: Closed-loop strategy

Our next counterexample involves the closed-loop strategy (ii).

Counterexample 3.7. *There exist \mathcal{C} and f , satisfying Assumption 2.3, and $x_0 \in \mathcal{C}$ such that any sequence $(x_t)_{t \in \mathbb{N}}$ generated by Frank-Wolfe algorithm using the closed-loop strategy (ii) does not converge.*

Let $\mathcal{C} = [-1, 1] \times [0, 2^K]$, where $K \in \mathbb{N}$, and f be defined via Theorem 3.1 with, for all $k \in \mathbb{N}$,

$$\begin{aligned}
 A_k &= \left((-1)^{k+1} \frac{61}{35}, \frac{9}{8} \frac{61}{35} \frac{1}{2^{k+1-K}} \right), & B_k &= \left((-1)^k \frac{61}{35}, \frac{17}{16} \frac{61}{35} \frac{1}{2^{k+1-K}} \right), \\
 D_k &= \left((-1)^{k+1}, \frac{9}{8} \frac{1}{2^{k+1-K}} \right), & C_k &= \left((-1)^k, \frac{17}{16} \frac{1}{2^{k+1-K}} \right), \\
 D'_k &= \left((-1)^{k+1}, \frac{1}{2^{k+1-K}} \right), & C'_k &= \left((-1)^k \frac{61}{35}, \frac{1}{2^{k+1-K}} \right), \\
 Y_k &= \begin{cases} \left(-\frac{61}{35} \left(1 + \frac{1}{2^k} \right), 0 \right) & \text{if } k \text{ is even;} \\ \left(-\frac{61}{35} \left(1 + \frac{17}{16} \frac{8}{9} \frac{1}{2^k} \right), 0 \right) & \text{if } k \text{ is odd,} \end{cases} & Y'_k &= \left(-\frac{61}{35} - \frac{8}{9} \frac{1}{2^k}, 0 \right), \\
 Z_k &= \begin{cases} \left(\frac{61}{35} \left(1 + \frac{1}{2^k} \right), 0 \right) & \text{if } k \text{ is even;} \\ \left(\frac{61}{35} \left(1 + \frac{9}{8} \frac{16}{17} \frac{1}{2^k} \right), 0 \right) & \text{if } k \text{ is odd,} \end{cases} & Z'_k &= \left(\frac{61}{35} + \frac{16}{17} \frac{1}{2^k}, 0 \right),
 \end{aligned}$$

and $\mathcal{P}_{2k} = \text{conv}\{Y_k, A_k, B_k, Z_k, X_{2k}\}$ and $\mathcal{P}_{2k+1} = \text{conv}\{Y'_k, D'_k, D_k, C_k, C'_k, Z'_k, X_{2k+1}\}$, where $X_k = (0, -1 - 1/(k+1))$. Then \mathcal{C} and f satisfy Assumption 2.3 so Theorem 2.4 holds. The solution set is $\arg \min_{\mathcal{C}} f = \bigcap_{k \in \mathbb{N}} \mathcal{P}_k \cap \mathcal{C} = [(-1, 0), (1, 0)]$. The construction is such that for all $k \in \mathbb{N}$, the lines $(A_k B_k)$ and $(C_k D_k)$ are parallel, $\langle B_k, e_2 \rangle < \langle A_k, e_2 \rangle$ and $\langle C_k, e_2 \rangle < \langle D_k, e_2 \rangle$, and the points $0, D_k, A_k$, and the points $0, C_k, B_k$, are aligned. An illustration is presented in Figure 5.

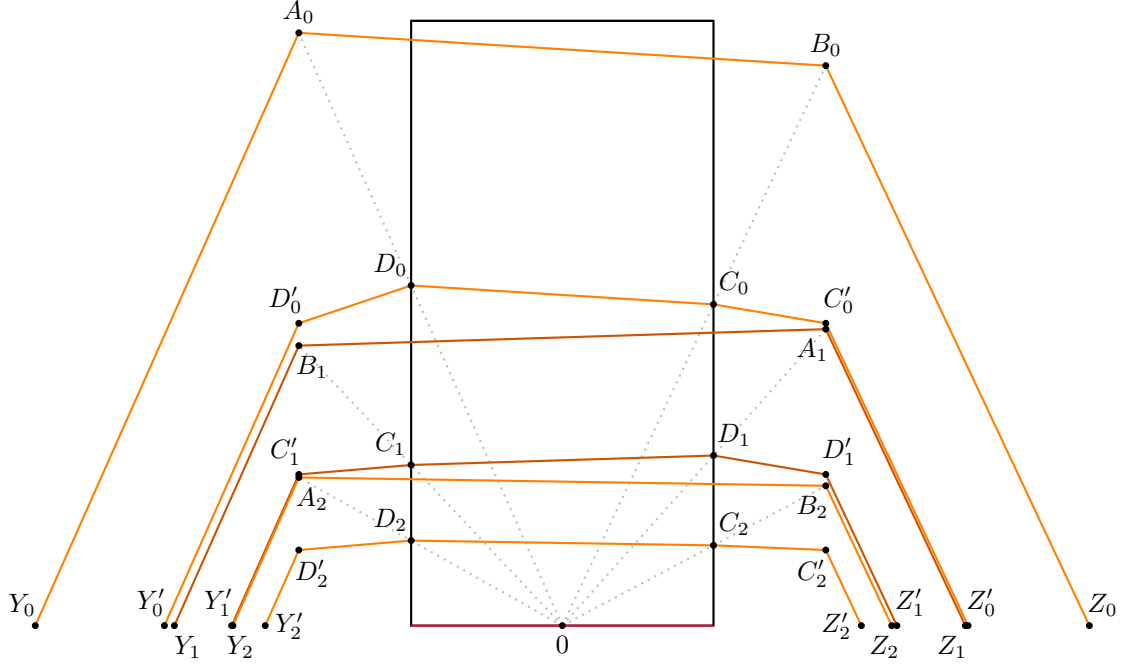


Figure 5: The constraint set (in black), polygonal sketches of the objective function (in orange and dark orange), and the solution set (in red). The polygonal sketches are studied by pairs, corresponding to the color tone, i.e., to the direction in which they lean (see Figure 6). For all $k \in \mathbb{N}$, the lines $(A_k B_k)$ and $(C_k D_k)$ are parallel, $\langle B_k, e_2 \rangle \leq \langle A_k, e_2 \rangle$ and $\langle C_k, e_2 \rangle \leq \langle D_k, e_2 \rangle$, and the points $0, D_k, A_k$, and the points $0, C_k, B_k$, are aligned.

By properties of the construction, for all $k \in \mathbb{N}$,

$$\begin{aligned} \mathcal{K}_{\mathcal{P}_{2k+1}} D_k &\subset \mathcal{K}_{\mathcal{P}_{2k}} A_k, & \mathcal{K}_{\mathcal{P}_{2k+1}} C_k &\subset \mathcal{K}_{\mathcal{P}_{2k}} B_k, \\ \mathcal{K}_{\mathcal{P}_{4k+1}} D_{2k} \cap (\mathbb{R}_+^*)^2 &\neq \emptyset, & \mathcal{K}_{\mathcal{P}_{4k+1}} C_{2k} \cap (\mathbb{R}_+^*)^2 &\neq \emptyset, \\ \mathcal{K}_{\mathcal{P}_{4k+3}} D_{2k+1} \cap (\mathbb{R}_-^* \times \mathbb{R}_+^*) &\neq \emptyset, & \mathcal{K}_{\mathcal{P}_{4k+3}} C_{2k+1} \cap (\mathbb{R}_-^* \times \mathbb{R}_+^*) &\neq \emptyset. \end{aligned}$$

Thus, we can choose $(u_k)_{k \in \mathbb{N}}$ such that for all $k \in \mathbb{N}$,

- (i) $\nabla f(A_k)$ and $\nabla f(D_k)$, and $\nabla f(B_k)$ and $\nabla f(C_k)$, are positively colinear;
- (ii) $\langle e_2, \nabla f(x) \rangle > 0$ and $(-1)^k \langle e_1, \nabla f(x) \rangle > 0$ for all $x \in \{A_k, B_k, C_k, D_k\}$.

Then, for all $k \in \mathbb{N}$, let

$$\begin{aligned} E_k &= \left((-1)^{k+1}, \frac{7}{4} \frac{1}{2^{k+1-K}} \right), & F_k &= \left((-1)^k, \frac{7}{4} \frac{1}{2^{k+1-K}} \right), \\ I_k &= \left((-1)^{k+1}, \frac{13}{8} \frac{1}{2^{k+1-K}} \right), & J_k &= \left((-1)^k, \frac{13}{8} \frac{1}{2^{k+1-K}} \right), \\ H_k &= \left((-1)^{k+1}, \frac{5}{4} \frac{1}{2^{k+1-K}} \right), & G_k &= \left((-1)^k, \frac{5}{4} \frac{1}{2^{k+1-K}} \right). \end{aligned}$$

An illustration is presented in Figure 6.

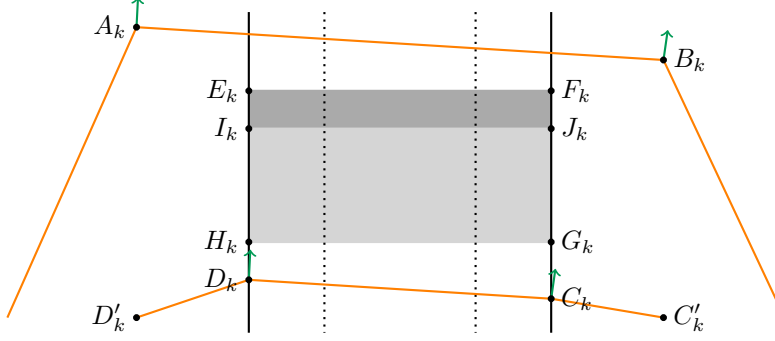


Figure 6: A pair of polygonal sketches (in orange) and gradient directions (in green). Here, $k \in \mathbb{N}$ is even. Using the gradient directions, we show that the linear minimization oracle always returns $(-1, 0)$ in $\text{conv}\{E_k, F_k, G_k, H_k\}$ here (Lemma 3.9(iv)). Then, we show that, if k is large enough, there exists an iterate $x_t \in \text{conv}\{E_k, F_k, J_k, I_k\}$ (Lemma 3.9(v)). Put together and using Thales' theorem, we can measure the minimum horizontal displacement of the iterates when they cross (vertically) the rectangle $\text{conv}\{I_k, J_k, G_k, H_k\}$. This turns out to be a constant number, showing that $(x_t)_{t \in \mathbb{N}}$ does not converge (Lemma 3.9(vi)).

Lemma 3.8. *Let $k \in \mathbb{N}$ and $x \in \mathcal{C}$.*

- (i) *If $\langle x, e_2 \rangle > 0$, then $\langle e_2, \nabla f(x) \rangle > 0$.*
- (ii) *If $x \in \text{conv}\{E_k, F_k, G_k, H_k\}$, then $(-1)^k \langle e_1, \nabla f(x) \rangle > 0$.*

Proof. (i) Let $x^* = x - \langle x, e_2 \rangle e_2 \in [(-1, 0), (1, 0)] = \arg \min_{\mathcal{C}} f$. By convexity of f , $\langle x^* - x, \nabla f(x) \rangle < 0$, i.e., $\langle x, e_2 \rangle \langle e_2, \nabla f(x) \rangle > 0$. Since $\langle x, e_2 \rangle > 0$, we obtain $\langle e_2, \nabla f(x) \rangle > 0$.

- (ii) By Lemma 3.4 and items (i)–(ii) above, $(-1)^k \langle e_1, \nabla f(y) \rangle > 0$ for all $y \in [A_k, D_k] \cup [B_k, C_k]$. There exist $y_1 \in [A_k, D_k]$, $y_2 \in [B_k, C_k]$, and $\gamma' \in [0, 1]$ such that $y_2 - y_1 = \|y_2 - y_1\| e_1$ and $x = (1 - \gamma')y_1 + \gamma'y_2$. Let $\varphi: \gamma \in [0, 1] \mapsto f((1 - \gamma)y_1 + \gamma y_2)$, which is convex. Then $\varphi'(\gamma) = \langle y_2 - y_1, \nabla f((1 - \gamma)y_1 + \gamma y_2) \rangle = \|y_2 - y_1\| \langle e_1, \nabla f((1 - \gamma)y_1 + \gamma y_2) \rangle$ for all $\gamma \in [0, 1]$. Thus, $(-1)^k \varphi'(0) = (-1)^k \|y_2 - y_1\| \langle e_1, \nabla f(y_1) \rangle > 0$ and $(-1)^k \varphi'(1) = (-1)^k \|y_2 - y_1\| \langle e_1, \nabla f(y_2) \rangle > 0$, so, by monotonicity of φ' , $(-1)^k \varphi'(\gamma) > 0$ for all $\gamma \in [0, 1]$. Therefore, $(-1)^k \langle e_1, \nabla f(x) \rangle > 0$. □

Lemma 3.9 completes our counterexample. Note that $L > L_{\mathcal{C}}^* \stackrel{\text{def}}{=} \sup_{x, y \in \mathcal{C}, x \neq y} \|\nabla f(y) - \nabla f(x)\| / \|y - x\|$ is most likely in practice, because the optimal value $L_{\mathcal{C}}^*$ is very hard to estimate precisely in general.

Lemma 3.9. *Consider the Frank-Wolfe algorithm starting from $x_0 \in \mathcal{C} \setminus \arg \min_{\mathcal{C}} f$ with the closed-loop strategy (ii). If $L > L_{\mathcal{C}}^*$, then:*

- (i) *For all $t \in \mathbb{N}$, $x_t \notin \arg \min_{\mathcal{C}} f$;*
- (ii) *For all $t \in \mathbb{N}$, $v_t \in [(-1, 0), (1, 0)]$ and $\gamma_t < 1$;*
- (iii) *$\gamma_t \rightarrow 0$;*
- (iv) *For all $t, k \in \mathbb{N}$, if $x_t \in \text{conv}\{E_k, F_k, G_k, H_k\}$, then $v_t = ((-1)^{k+1}, 0)$;*
- (v) *There exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$, $\{x_t \mid t \in \mathbb{N}\} \cap \text{conv}\{E_k, F_k, J_k, I_k\} \neq \emptyset$;*
- (vi) *$(x_t)_{t \in \mathbb{N}}$ does not converge.*

Proof. Let $x_t = (\alpha_t, \beta_t)$ for all $t \in \mathbb{N}$ and recall that $\arg \min_{\mathcal{C}} f = [(-1, 0), (1, 0)]$.

- (i) We proceed by induction. The base case is satisfied. Let $t \in \mathbb{N}$ be such that $x_t \notin \arg \min_{\mathcal{C}} f$. Then $\beta_t > 0$ and, by Lemma 3.8(i), $v_t \in [(-1, 0), (1, 0)]$, so $\nabla f(v_t) = 0$. By the Cauchy-Schwarz inequality,

$$\gamma_t \leq \frac{\langle x_t - v_t, \nabla f(x_t) \rangle}{L \|x_t - v_t\|^2} \leq \frac{\|\nabla f(x_t)\|}{L \|x_t - v_t\|} = \frac{\|\nabla f(x_t) - \nabla f(v_t)\|}{L \|x_t - v_t\|} \leq \frac{L_{\mathcal{C}}^*}{L} < 1. \quad (1)$$

Therefore, $x_{t+1} \in [x_t, v_t] \subset \mathcal{C} \setminus \arg \min_{\mathcal{C}} f$.

- (ii) See the proof of Lemma 3.9(i).
 (iii) By Theorem 3.1, f is d -time continuously differentiable, where $d \geq 2$. By (1) and the mean value theorem,

$$\gamma_t \leq \frac{\|\nabla f(x_t) - \nabla f(v_t)\|}{L \|x_t - v_t\|} \leq \frac{1}{L} \sup_{x \in [x_t, v_t]} \|\nabla^2 f(x)\|_{\text{op}}, \quad (2)$$

where $\|\cdot\|_{\text{op}}$ denotes the operator norm. Since $\arg \min_{\mathcal{D}} f = \bigcap_{k \in \mathbb{N}} \mathcal{P}_k$ has nonempty interior, $\nabla^2 f(x) = 0$ for all $x \in \arg \min_{\mathcal{D}} f \supset \arg \min_{\mathcal{C}} f$. By Theorem 2.4, $f(x_t) \rightarrow \min_{\mathcal{C}} f$, so $\text{dist}(x_t, \arg \min_{\mathcal{C}} f) \rightarrow 0$, and, by Lemma 3.9(ii), $v_t \in \arg \min_{\mathcal{C}} f$ for all $t \in \mathbb{N}$. By continuity of $\nabla^2 f$, it follows that $\sup_{x \in [x_t, v_t]} \|\nabla^2 f(x)\|_{\text{op}} \rightarrow 0$. By (2), $\gamma_t \rightarrow 0$.

- (iv) This follows from Lemma 3.8(ii).

- (v) By Lemma 3.9(ii), $v_t \in [(-1, 0), (1, 0)]$ for all $t \in \mathbb{N}$. Thus,

$$\beta_{t'} = \left(\prod_{\ell=t}^{t'-1} (1 - \gamma_{\ell}) \right) \beta_t, \quad (3)$$

for all $t, t' \in \mathbb{N}$ such that $t' \geq t+1$. Let $c = 1 - \sqrt{13/14} > 0$. By Theorem 2.4, $f(x_t) \rightarrow \min_{\mathcal{C}} f$, so $\beta_t \rightarrow 0$. By Lemma 3.9(iii), $\gamma_t \rightarrow 0$, so there exists $t_0 \in \mathbb{N}$ such that $\beta_t \leq \beta_{t_0}$ and $\gamma_t \leq c$ for all $t \geq t_0$. Let $k_0 \in \mathbb{N}$ be such that $\beta_{t_0} \geq (7/4)(1/2)^{k_0+1-K} = \langle E_{k_0}, e_2 \rangle$. Let $k \geq k_0$, $t \geq t_0$, and $t' \geq t+1$ be such that $\beta_t \geq (7/4)(1/2)^{k+1-K} = \langle E_k, e_2 \rangle$ and $\beta_{t'} \leq (13/8)(1/2)^{k+1-K} = \langle I_k, e_2 \rangle$. By (3),

$$\frac{13/8}{7/4} \geq \prod_{\ell=t}^{t'-1} (1 - \gamma_{\ell}) \geq (1 - c)^{t'-t},$$

so

$$t' - t \geq \frac{\ln(14/13)}{\ln(1/(1 - c))} = 2.$$

Therefore, there exists $t'' \in [t+1, t'-1]$ such that $x_{t''} \in \text{conv}\{E_k, F_k, J_k, I_k\}$.

- (vi) Let $t, k \in \mathbb{N}$ be such that $x_t \in \text{conv}\{E_k, F_k, J_k, I_k\}$. With no loss of generality, we can assume that k is even.

- If $\alpha_t \geq -1/2$, let $t' = \min\{\ell \in \mathbb{N} \mid \beta_{\ell} < (5/4)(1/2)^{k+1-K} = \langle H_k, e_2 \rangle, \ell \geq t\}$. By Lemma 3.9(iv), $v_{\ell} = (-1, 0)$ for every $\ell \in [t, t'-1]$. By Thales' theorem,

$$\frac{\alpha_t - \alpha_{t'}}{\alpha_t - (-1)} = \frac{\beta_t - \beta_{t'}}{\beta_t - 0},$$

where $\alpha_t \geq -1/2$, $(13/8)(1/2)^{k+1-K} = \langle I_k, e_2 \rangle \leq \beta_t \leq (7/4)(1/2)^{k+1-K} = \langle E_k, e_2 \rangle$, and $\beta_{t'} \leq (5/4)(1/2)^{k+1-K} = \langle H_k, e_2 \rangle$. Thus,

$$\alpha_t - \alpha_{t'} \geq \frac{13/8 - 5/4}{7/4} \frac{1}{2} = \frac{3}{28}.$$

- Else, $\alpha_t < -1/2$. There exists $t' \in \mathbb{N}$ such that $x_{t'} \in \text{conv}\{E_{k+1}, F_{k+1}, J_{k+1}, I_{k+1}\}$. Let $t'' = \min\{\ell \in \mathbb{N} \mid \beta_\ell < (5/4)(1/2)^{k+2-K} = \langle H_{k+1}, e_2 \rangle, \ell \geq t'\}$. If $\alpha_{t'} \leq 1/2$, we can reproduce the same argument as above and conclude that

$$\alpha_{t''} - \alpha_{t'} \geq \frac{3}{28}.$$

Else, $\alpha_{t'} \geq 1/2$, so

$$\alpha_{t'} - \alpha_t \geq 1 \geq \frac{3}{28}.$$

Thus, together with Lemma 3.9(v), we conclude that $\text{card}\{t \in \mathbb{N} \mid \exists t' \geq t: |\alpha_{t'} - \alpha_t| \geq 3/28\} = +\infty$. Therefore, $(\alpha_t)_{t \in \mathbb{N}}$ is not a Cauchy sequence, so $(x_t)_{t \in \mathbb{N}}$ does not converge. \square

3.4 Counterexample 4: Open-loop strategies

Our last counterexample involves the open-loop strategies (i)–(i'). It is similar in design to the one in Section 3.1. Note that the counterexample from Section 3.3 would be valid here as well if we had $\gamma_0 < 1$.

Counterexample 3.10. *There exist \mathcal{C} and f , satisfying Assumption 2.3, and $x_0 \in \mathcal{C}$ such that any sequence $(x_t)_{t \in \mathbb{N}}$ generated by Frank-Wolfe algorithm using an open-loop strategy (i)–(i') does not converge.*

Let $(\gamma_t)_{t \in \mathbb{N}}$ be an open-loop strategy, $\mathcal{C} = \text{conv}\{(-2, 1/4), (-1, 0), (0, 1), (1, 0)\}$, and f be defined via Theorem 3.1 with, for all $k \in \mathbb{N}$, $\mathcal{P}_k = \text{conv}\{A_k, B_k, C_k, -B_k\}$ where

$$A_k = \left(-2 - \frac{1}{k+1}, 0\right), \quad C_k = \left(1 + \frac{1}{k+1}, 0\right),$$

$B_0 = (0, 1)$, $B_1 = V_0 = (-2, 1/4)$, $V_1 = (1, 0)$, and, for all $k \geq 1$,

$$B_{k+1} = (1 - \gamma_k)B_k + \gamma_k V_k, \\ V_{k+1} = \begin{cases} -V_k & \text{if } |\langle B_{k+1}, e_1 \rangle| > 1/4 \text{ and } |\langle B_k, e_1 \rangle| \leq 1/4; \\ V_k & \text{else.} \end{cases}$$

Then \mathcal{C} and f satisfy Assumption 2.3 so Theorem 2.4 holds. The solution set is $\arg \min_{\mathcal{C}} f = \bigcap_{k \in \mathbb{N}} \mathcal{P}_k \cap \mathcal{C} = [(-1, 0), (1, 0)]$. The construction is such that the trajectory of $(B_k)_{k \in \mathbb{N}}$ is headed by $(V_k)_{k \in \mathbb{N}}$, whose terms change value each time $(B_k)_{k \in \mathbb{N}}$ crosses the vertical band $\{x \in \mathcal{C} \mid |\langle x, e_1 \rangle| \leq 1/4\}$. An illustration is presented in Figure 7.

By construction, for all $k \in \mathbb{N}$,

$$\mathcal{K}_{\mathcal{P}_k} B_k \cap (\mathbb{R}_-^* \times \mathbb{R}_+^*) \neq \emptyset \quad \text{and} \quad \mathcal{K}_{\mathcal{P}_k} B_k \cap (\mathbb{R}_+^*)^2 \neq \emptyset.$$

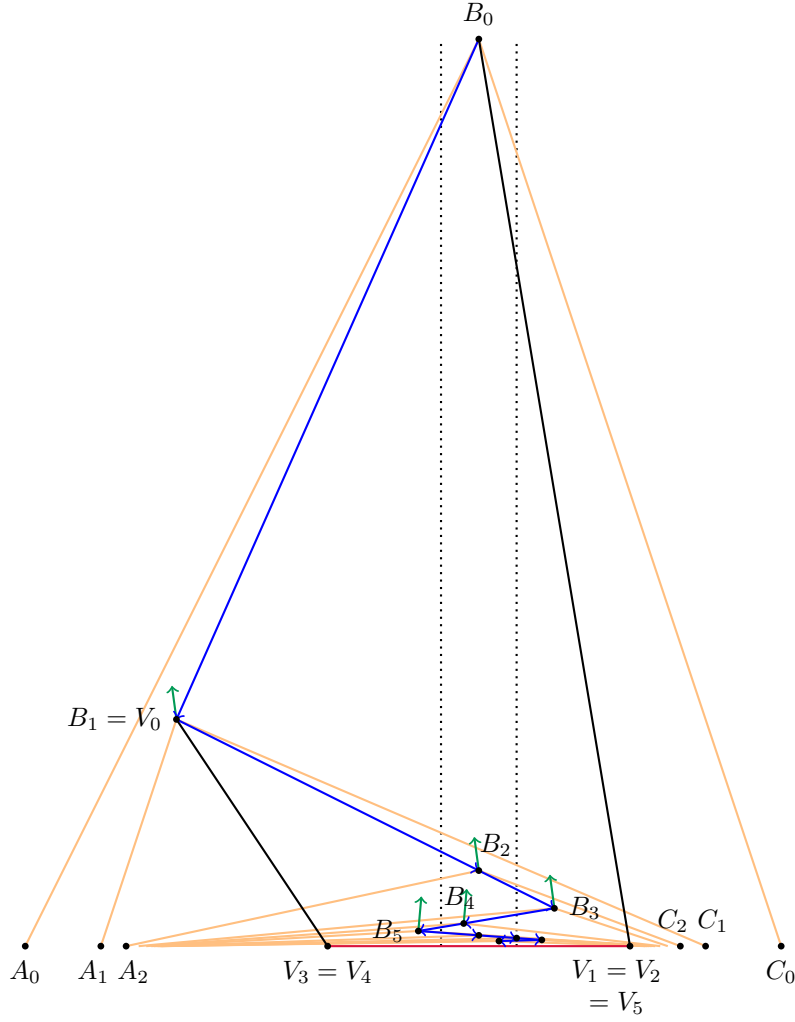


Figure 7: *Stretched for visualization purposes.* The constraint set (in black), polygonal sketches of the objective function (in orange) built using the open-loop strategy (i'), the solution set (in red), gradient directions (in green), and the trajectory (in blue) of $(B_k)_{k \in \mathbb{N}}$. For all $k \geq 1$, $B_{k+1} \in [B_k, V_k]$ where $V_k = V_{k-1}$ or $V_k = -V_{k-1}$, depending on whether $(B_k)_{k \in \mathbb{N}}$ has just crossed the vertical band delimited by the dashed lines or not.

Thus, we can choose $(u_k)_{k \in \mathbb{N}}$ such that for all $k \in \mathbb{N}$,

$$u_k \in \mathcal{K}_{\mathcal{P}_k} B_k \quad \text{and} \quad V_k \in \arg \min_{v \in \mathcal{C}} \langle v, u_k \rangle.$$

Let $x_0 = B_0$. By induction, $x_t = B_t$ for all $t \in \mathbb{N}$. Therefore, $(x_t)_{t \in \mathbb{N}}$ crosses the vertical band $\{x \in \mathcal{C} \mid |\langle x, e_1 \rangle| \leq 1/4\}$ indefinitely, i.e., $\text{card}\{t \in \mathbb{N} \mid \langle x_t, e_1 \rangle \leq -1/4\} = +\infty$ and $\text{card}\{t \in \mathbb{N} \mid \langle x_t, e_1 \rangle \geq 1/4\} = +\infty$. Since $\gamma_t < 1$ for all $t \geq 1$, $\langle x_t, e_2 \rangle > 0$ for all $t \in \mathbb{N}$, so $(x_t)_{t \in \mathbb{N}}$ does not converge.

Acknowledgment

The authors acknowledge the support of the AI Interdisciplinary Institute ANITI funding, through the French “Investments for the Future – PIA3” program under the grant agreement ANR-19-PI3A0004, Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant numbers FA9550-19-1-7026, and ANR MaSDOL 19-CE23-0017-0. Jérôme Bolte also acknowledges the support of ANR Chess, grant ANR-17-EURE-0010, and TSE-P.

References

- [1] J. Bolte and E. Pauwels. Curiosities and counterexamples in smooth convex optimization. *Mathematical Programming*, to appear.
- [2] G. Braun, S. Pokutta, D. Tu, and S. Wright. Blended conditional gradients: the unconditioning of conditional gradients. In *Proceedings of the 36th International Conference on Machine Learning*, pages 735–743, 2019.
- [3] M. D. Canon and C. D. Cullum. A tight upper bound on the rate of convergence of Frank-Wolfe algorithm. *SIAM Journal on Control*, 6(4):509–516, 1968.
- [4] K. L. Clarkson. Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. *ACM Transactions on Algorithms*, 6(4):1–30, 2010.
- [5] C. W. Combettes and S. Pokutta. Boosting Frank-Wolfe by chasing gradients. In *Proceedings of the 37th International Conference on Machine Learning*, pages 2111–2121, 2020.
- [6] C. W. Combettes and S. Pokutta. Complexity of linear minimization and projection on some sets. *Operations Research Letters*, 49(4):565–571, 2021.
- [7] C. W. Combettes and S. Pokutta. Revisiting the approximate Carathéodory problem via the Frank-Wolfe algorithm. *Mathematical Programming*, to appear.
- [8] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.
- [9] V. Do, S. Corbett-Davies, J. Atif, and N. Usunier. Two-sided fairness in rankings via Lorenz dominance. In *Advances in Neural Information Processing Systems*, volume 34, 2021. To appear.
- [10] J. C. Dunn and S. Harshbarger. Conditional gradient algorithms with open loop step size rules. *Journal of Mathematical Analysis and Applications*, 62(2):432–444, 1978.
- [11] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1–2):95–110, 1956.
- [12] D. Garber and E. Hazan. Faster rates for the Frank-Wolfe method over strongly-convex sets. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 541–549, 2015.
- [13] D. Garber and O. Meshi. Linear-memory and decomposition-invariant linearly convergent conditional gradient algorithm for structured polytopes. In *Advances in Neural Information Processing Systems*, volume 29, pages 1001–1009, 2016.
- [14] J. Guélat and P. Marcotte. Some comments on Wolfe’s ‘away step’. *Mathematical Programming*, 35(1):110–119, 1986.
- [15] E. Hazan. Sparse approximate solutions to semidefinite programs. In *Proceedings of the 8th Latin American Symposium on Theoretical Informatics*, pages 306–316, 2008.
- [16] M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning*, pages 427–435, 2013.

- [17] A. Joulin, K. Tang, and L. Fei-Fei. Efficient image and video co-localization with Frank-Wolfe algorithm. In *European Conference on Computer Vision*, pages 253–268, 2014.
- [18] T. Kerdreux, A. d’Aspremont, and S. Pokutta. Projection-free optimization on uniformly convex sets. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, pages 19–27, 2021.
- [19] S. Lacoste-Julien and M. Jaggi. On the global linear convergence of Frank-Wolfe optimization variants. In *Advances in Neural Information Processing Systems*, volume 28, pages 496–504, 2015.
- [20] S. Lacoste-Julien, M. Jaggi, M. Schmidt, and P. Pletscher. Block-coordinate Frank-Wolfe optimization for structural SVMs. In *Proceedings of the 30th International Conference on Machine Learning*, pages 53–61, 2013.
- [21] G. Lan. The complexity of large-scale convex programming under a linear optimization oracle. Technical report, Department of Industrial and Systems Engineering, University of Florida, 2013.
- [22] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla. An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research*, 9(5):309–318, 1975.
- [23] E. S. Levitin and B. T. Polyak. Constrained minimization methods. *USSR Computational Mathematics and Mathematical Physics*, 6(5):1–50, 1966.
- [24] G. Luise, S. Salzo, M. Pontil, and C. Ciliberto. Sinkhorn barycenters with free support via Frank-Wolfe algorithm. In *Advances in Neural Information Processing Systems*, volume 32, pages 9322–9333, 2019.
- [25] R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge University Press, 2nd edition, 2013.