On the Frank–Wolfe algorithm for non-compact constrained optimization problems

O. P. Ferreira\textsuperscript{a} and W. S. Sosa\textsuperscript{b}

\textsuperscript{a}Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia-GO, Brazil; \textsuperscript{b}Programa de Pós-Graduação em Economia, Universidade Católica de Brasília, Brasília-DF, Brazil

\textbf{ABSTRACT}
This paper deals with the Frank–Wolfe algorithm to solve a special class of non-compact constrained optimization problems. The notion of asymptotic cone is one the main concept used to introduce the class of problems considered as well as to establish the well definition of the algorithm. This class of optimization problems, with closed and convex constraint set, are characterized by two conditions on the gradient of the objective function. The first one establishes that the gradient of the objective function is Lipschitz continuous, which is quite usual in the analysis of this algorithm. The second one, which is new in this subject, establishes that the gradient belongs to the interior of dual asymptotic cone of the constraint set. Classical results on asymptotic behaviour and iteration complexity bounds for the sequence generated by Frank–Wolfe algorithm are extended to this new class of problems. Some examples of problems with non-compact constraints and objective functions satisfying the aforementioned conditions are provided.

\textbf{ARTICLE HISTORY}
Received 29 November 2019
Accepted 27 November 2020

\textbf{KEYWORDS}
Frank–Wolfe method; constrained optimization problem; non-compact constraint

\section{Introduction}
The Frank–Wolfe algorithm or conditional gradient method is one of the oldest methods for finding minimizer of a differentiable functions onto compact convex sets, dating back to the 1950s. This method was initially proposed in [1] for solving quadratic programming problems with linear constraints; see also [2,3]. Since then, this method has attracted the attention of the scientific community working on this subject. One of the factors that explains this interest is its simplicity and ease of implementation, only requiring access to a linear minimization oracle over the constraint set. In particular, allowing a low cost of storage and ready exploitation of separability and sparsity. Indeed, it makes its application in large scale problems very attractive. Recently, there has been an increase in the popularity of this method due to the emergence of machine learning applications; see [4–6]. For these reasons, several variants of this method have emerged and
properties of it have been discovered throughout the years, resulting in a wide
literature on the subject; see, for example [7–14] and references therein.

The aim of this paper is to extend, from theoretical point of view the classical
analysis of Frank–Wolfe algorithm to special class of constrained optimization
problem Minimize \( x \in C f(x) \), where \( f: \mathbb{R}^n \to \mathbb{R} \) is a continuously differentiable
function and \( C \subset \mathbb{R}^n \) is a closed and convex set, but not necessarily compact.
Besides the classical assumptions, namely, \( f \) is gradient Lipschitz continuous,
we also assume that \( \nabla f(x) \in \text{int}(C_\infty)^* \), for all \( x \in C \), where \( \nabla f \) and \( \text{int}(C_\infty)^* \)
denote the gradient of \( f \) and the interior of positive dual asymptotic cone of
\( C \), respectively. For this class of functions, with classical assumptions, we also
extend results on the asymptotic behaviour and iteration-complexity bounds for
the sequence generated by Frank–Wolfe algorithm.

The organization of this paper is as follows. In Section 2, some notations and
auxiliary results, used throughout of the paper, are presented. In Section 3 we
state the Frank–Wolfe algorithm. In Section 3.1 we establish the well-definition
of the sequence generated for this algorithm. Section 3.2 is devoted to the study
of the asymptotic convergence properties of Algorithm 3 and Section 3.3 to the
study of the iteration-complexity bounds. In Section 4 we present some examples.
We conclude the paper with some remarks in Section 5.

2. Preliminaries

In this section, we present some notations, definitions, and results used through-
out the paper. Let \( \mathbb{R}^n \) be the \( n \)-dimensional Euclidean space with the canonical
inner product \( \langle \cdot, \cdot \rangle \) and norm \( \| \cdot \| \). Denote by \( \mathbb{R}^{m \times n} \) the set of all \( m \times n \) matrices
with real entries, \( \mathbb{R}^n \equiv \mathbb{R}^{n \times 1} \), by \( e^i \) the \( i \)-th canonical unit vector in \( \mathbb{R}^n \), and
by \( I_n \) the \( n \times n \) identity matrix. A set \( K \subset \mathbb{R}^n \) is called a cone if for any \( \alpha > 0 \)
and \( x \in K \) we have \( \alpha x \in K \). A cone \( K \subset \mathbb{R}^n \) is called a convex cone if for any
\( x, y \in K, \) we have \( x + y \in K \). The positive dual cone of a cone \( K \subset \mathbb{R}^n \) is the
cone \( K^* := \{ x \in \mathbb{R}^n : x^T y \geq 0, \forall y \in K \} \) and its interior is denoted by \( \text{int}K^* := \{ x \in \mathbb{R}^n : x^T y > 0, \forall y \in K \setminus \{0\} \} \). Let \( C \subset \mathbb{R}^n \) be a closed convex set, we define the
asymptotic cone of \( C \) by

\[
C_\infty := \left\{ d \in \mathbb{R}^n : \exists (t_k)_{k \in \mathbb{N}} \subset (0, \infty), \exists (x^k)_{k \in \mathbb{N}} \subset C; \lim_{k \to \infty} t_k = 0, \lim_{k \to \infty} t_k x^k = d \right\},
\]

or equivalently, \( C_\infty := \{ d \in \mathbb{R}^n : x + td \in C, \forall x \in C, \forall t \geq 0 \} \), see [15, p.39].
Let \( f: \mathbb{R}^n \to \mathbb{R} \) be a continuously differentiable function. Consider the problem
of finding a optimum point of \( f \) in a closed convex set \( C \subset \mathbb{R}^n \), i.e. a point \( x^* \in C \)
such that \( f(x^*) \leq f(x) \) for all \( x \in C \). We denote this constrained problem as

\[
\text{Minimize}_{x \in C} f(x)
\]
The optimum value of \( f \) on \( C \) is denoted by \( f^* \), i.e. \( f^* := \inf_{x \in C} f(x) \). The first-order optimality condition for problem (1) is stated as

\[
\nabla f(x^*)^T (x - x^*) \geq 0, \quad \forall x \in C.
\]  

(2)

In general, the condition (2) is necessary, but not sufficient for optimality. Thus, a point \( x^* \in C \) satisfying condition (2) is called stationary point to problem (1).

**Definition 2.1:** Let \( C \subset \mathbb{R}^n \) be a convex set. A function \( f : \mathbb{R}^n \to \mathbb{R} \) is called \( M \)-strongly convex with parameter \( M \geq 0 \) on \( C \) when the following inequality holds

\[
f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{1}{2}Mt(1-t)\|y - x\|^2,
\]

for all \( t \in [0,1] \) and \( x, y \in C \). In particular, for \( M = 0 \), we simply say that \( f \) is convex instead of 0-strongly convex.

In the following results we give a useful characterization for convex/strongly convex differentiable functions, see the proof in [16, Theorem 4.1.1, p.183].

**Proposition 2.2:** Let \( C \subset \mathbb{R}^n \) be a convex set, \( f : \mathbb{R}^n \to \mathbb{R} \) be a continuously differentiable function and \( M \geq 0 \). Then, \( f \) is \( M \)-strongly convex in \( C \) if and only if,

\[
f(y) \geq f(x) + \nabla f(x)^T (y - x) + M\|x - y\|/2, \quad \text{for all } x, y \in C.
\]

**Remark 2.1:** It is well known that if \( f \) is \( M \)-strong convex, then (2) is sufficient for optimality, i.e. any point \( x^* \in C \) satisfying (2) is a minimizer to problem (1).

The proof of the next lemma can be found in [17, Lemma 6].

**Lemma 2.3:** Let \( \{a_k\} \) be a non-negative sequence of real numbers. If \( \Gamma a_k^2 \leq a_k - a_{k+1}, \) for some \( \Gamma > 0 \) and for any \( k = 1, \ldots, \ell \), then \( a_{\ell} \leq a_0/(1 + \Gamma a_0 \ell) < 1/(\Gamma \ell) \).

### 3. Frank–Wolfe algorithm

In this section we state the Frank–Wolfe algorithm to solve problem (1). For that, henceforward we assume that the constraint set \( C \subset \mathbb{R}^n \) is closed and convex (not necessarily compact), the objective function \( f : \mathbb{R}^n \to \mathbb{R} \) of problem (1) is continuously differentiable and its gradient satisfies the following condition:

\[
\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in C \text{ and } L > 0.
\]

**Remark 3.1:** In Section 4 will be presented some examples of problem (1) with objective function satisfying (A).

To state the Frank–Wolfe algorithm we need to assume an existence of a linear optimization oracle (LO oracle) capable of minimizing linear functions over \( C \).
Algorithm 1 CondG$_{C,f}$ method

(0) Take $x^0 \in C$. Set $k = 0$.

(1) Use a ‘LO oracle’ to compute an optimal solution $p_k$ and the optimal value $v^*_k$ as

$$p_k \in \arg\min_{p \in C} \nabla f(x^k)^T (p - x^k), \quad v^*_k := \nabla f(x^k)^T (p_k - x^k).$$

(3)

(2) If $v^*_k = 0$, then stop; otherwise, compute the step-size $\lambda_k \in (0,1]$ as

$$\lambda_k := \min \left\{ 1, \frac{|v^*_k|}{L \|p_k - x^k\|^2} \right\} = \min_{\lambda \in (0,1]} \left\{ v^*_k \lambda + \frac{L}{2} \|p(x^k) - x^k\|^2 \lambda^2 \right\}$$

and set the next iterated $x^{k+1}$ as

$$x^{k+1} := x^k + \lambda_k (p_k - x^k).$$

(5)

(3) Set $k \leftarrow k + 1$, and go to step 1.

We end this section by stating a basic inequality for functions satisfying assumption (A); see [18, Lemma 2.4.2].

Lemma 3.1: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function satisfying condition (A), $x \in C$ and $\lambda \in [0,1]$. Then

$$f(x + \lambda (p - x)) \leq f(x) + \nabla f(x)^T (p - x) \lambda + \frac{L}{2} \|p - x\|^2 \lambda^2, \quad \forall p \in C.$$  

(6)

3.1. Well definition

In this section we establish the well-definition of the sequence $(x^k)_{k \in \mathbb{N}}$ generated by Algorithm 3. For that, we assume that gradient of the objective function $f : \mathbb{R}^n \to \mathbb{R}$ of problem (1) satisfies the following condition:

(B) $\nabla f(x) \in \text{int}(C_\infty)^*$, for all $x \in C$.

Remark 3.2: It follows from [15, Proposition 2.2.3] that a closed convex set $C \subset \mathbb{R}^n$ is compact if and only if $C_\infty = \{0\}$. In this case, $\text{int}(C_\infty)^* = \mathbb{R}^n$. Therefore, (B) holds trivially whenever $C$ is compact. In Section 4 will be presented some examples of problem (1) with $\nabla f$ satisfying (B) with constraint set $C$ unbounded.

Next we use (B) to prove a general result, which imply existence of solution of the problem (3). It is worth mentioning that we will not need (A) in its proof.

Proposition 3.2: The following three conditions hold:
(i) For each \( x \in C \), the set \( D_x := \{ p \in C : \nabla f(x)^T (p - x) \leq 0 \} \) is compact.

(ii) For each \( x \in C \), the following linear problem

\[
\text{Minimize}_{p \in C} \nabla f(x)^T (p - x)
\]

has a solution.

(iii) If \( D \subset C \) is a bounded set, then the set

\[
\bigcup_{x \in D} \{ q_x \in C : q_x \in \arg\min_{p \in C} \nabla f(x)^T (p - x) \},
\]

is also bounded.

**Proof:** To prove item (i), let \( x \in C \). Assume by contradiction that \( D_x \) is unbounded. Thus, let \((q^k)_{k \in \mathbb{N}} \subset D_x\) such that \( \lim_{k \to \infty} \|q^k\| = \infty \). Let \((t_k)_{k \in \mathbb{N}} \subset (0, \infty)\) be the sequence defined by \( t_k := 1/\|q^k\| \), for all \( k = 0, 1, \ldots \). Thus, we have \( \lim_{k \to \infty} t_k = 0 \). Since \( t_k q^k = q^k/\|q^k\| \), we conclude \( \|t_k p_k\| = 1 \), for all \( k = 0, 1, \ldots \). Hence, there exist subsequences \((q^{kj})_{j \in \mathbb{N}} \subset D_x\) and \((t_kj)_{j \in \mathbb{N}} \subset (0, \infty)\) such that \( \lim_{j \to \infty} t_kj q^{kj} = d \in C_\infty \). Thus, the definition of \( D_x \) implies

\[
\nabla f(x)^T \left( t_kj q^{kj} - t_kj x \right) = \nabla f(x)^T \left( \frac{q^{kj}}{\|q^{kj}\|} - \frac{x}{\|q^{kj}\|} \right) \leq 0.
\]

Taking limit in the last inequality as \( j \) goes to \( \infty \) we conclude that \( \nabla f(x)^T d \leq 0 \), which is an absurd. Indeed, assumption (B) implies that \( \nabla f(x) \in \text{int}(C_\infty)^* \) and considering that \( d \in C_\infty \) we have \( \nabla f(x)^T d > 0 \). Therefore, item (i) is proved. For proving item (ii) it is sufficient to note that the problem (7) has as a sublevel \( D_x \), which by item (i) is compact. We proceed with the prove of item (iii). For that, we assume by contradiction that the set in (8) is unbounded. Thus, there exist sequences \((x^k)_{k \in \mathbb{N}} \subset D\) and \((q_{x^k})_{k \in \mathbb{N}} \subset C\) such that \( \lim_{k \to \infty} \|q_{x^k}\| = \infty \). Thus, considering that \( D \) is bounded we have

\[
\lim_{k \to \infty} \tau_k = 0, \quad \text{where } \tau_k := \frac{1}{\|q_{x^k} - x^k\|}, \quad k = 0, 1, \ldots
\]

On the other hand, since \( C \) is convex and \((x^k)_{k \in \mathbb{N}} \) and \((q_{x^k})\) belong to \( C \) we have

\[
x^k + t \left( q_{x^k} - x^k \right) \in C, \quad k = 0, 1, \ldots
\]

for any \( t \in (0, 1) \). Since \( \tau_k (q_{x^k} - x^k) = (q_{x^k} - x^k)/\|q_{x^k} - x^k\| \), we have \( \|\tau_k (q_{x^k} - x^k)\| = 1 \), for all \( k = 0, 1, \ldots \). Thus, there exist, subsequences \((x^{kj})_{j \in \mathbb{N}} \subset D\),
\((q_{x^{kj}})_{j \in \mathbb{N}} \subset C\) and \((t_{kj})_{j \in \mathbb{N}} \subset (0, \infty)\) such that
\[
\lim_{k_j \to \infty} \tau_{kj}(q_{x^{kj}} - x^{kj}) = \nu. \tag{11}
\]

Using that \((x^{kj})_{j \in \mathbb{N}} \subset D\) and \(D\) is bounded, the combination of (9), (10) and (11) yield
\[
\lim_{k_j \to \infty} \tau_{kj} \left[ x^{kj} + t \left( q_{x^{kj}} - x^{kj} \right) \right] = \lim_{k_j \to \infty} \left[ \tau_{kj}x^{kj} + t\tau_{kj} \left( q_{x^{kj}} - x^{kj} \right) \right] = tv \in C_\infty. \tag{12}
\]

Since \(q_{x^{kj}} \in \arg\min_{p \in C} \nabla f(x^{kj})^T(p - x^{kj})\) and \(x^{kj} \in C\), we have \(\nabla f(x^{kj})^T(q_{x^{kj}} - x^{kj}) \leq 0\), for all \(j = 0, 1, \ldots\). Then, due to \((t_{kj})_{j \in \mathbb{N}} \subset (0, \infty)\) we have
\[
\nabla f(x^{kj})^T \left( \tau_{kj}(q_{x^{kj}} - x^{kj}) \right) \leq 0, \quad j = 0, 1, \ldots.
\]

Considering that \(D\) is bounded and \((x^{kj})_{j \in \mathbb{N}} \subset D\), we can assume without lose of generality that \(\lim_{k_j \to \infty} x^{kj} = \bar{x}\). Thus, taking limit in the last inequality as \(j\) goes to \(\infty\) and using (11) we obtain \(\nabla f(\bar{x})^T v \leq 0\), which is an absurd because by (12) we have \(v \in C_\infty\) and by assumption (B) we have \(\nabla f(\bar{x}) \in \text{int}(C_\infty)^*\). Therefore, the proof of item (iii) is concluded.

In the next lemma we establish the well definition of the sequence \((x^k)_{k \in \mathbb{N}}\) generated by Algorithm 3 and some results related to the optimal value \(v^*_k\) defined in (3).

**Lemma 3.3:** The sequences \(\{p^k\}_{k \in \mathbb{N}} \subset C\) and \((x^k)_{k \in \mathbb{N}} \subset C\) are well defined. Moreover, the following statements hold:

(i) \(v^*_k \leq 0\), for all \(k = 0, 1, \ldots\);
(ii) \(v^*_k = 0\) if and only if \(x^k\) is a stationary point of problem (1);
(iii) \(v^*_k < 0\) if and only if \(\lambda_k > 0\) and \(p^k \neq x^k\).

**Proof:** For each \(x^k \in C\) it follows from Proposition 3.2 that \(p^k\) and \(v^*_k\) in (3) can be computed and \(p^k \in C\). Since \(x^0 \in C\) and \(0 \leq \lambda \leq 1\), by using (5) and induction argument we conclude that \((\tilde{p}^k)_{k \in \mathbb{N}}, (\nu^*_k)_{k \in \mathbb{N}}\) and \((x^k)_{k \in \mathbb{N}}\) are well defined and \((p^k)_{k \in \mathbb{N}}\) and \((x^k)_{k \in \mathbb{N}}\) belong to \(C\). To prove (i) it sufficient to note that the optimality in (3) implies \(v^*_k \leq \nabla f(x^k)^T(x^k - x^k) = 0\), for all \(k = 0, 1, \ldots\). For proving (ii), note that (3) implies that \(v^*_k \leq \nabla f(x^k)^T(p - x^k)\), for all \(p \in C\). Thus, if \(v^*_k = 0\) we conclude \(0 \leq \nabla f(x^k)^T(p - x^k)\), for all \(p \in C\). Hence, \(x^k\) satisfies (2), i.e. \(x^k\) is a stationary point of problem (1). Reciprocally, if \(x^k\) is a stationary point of problem (1), then (2) implies \(0 \leq \nabla f(x^k)^T(p - x^k)\), for all \(p \in C\). Since \(p^k \in C\), we conclude \(0 \leq \nabla f(x^k)^T(p^k - x^k) = v^*_k\). Therefore, from item (i) we have \(v^*_k = 0\). We proceed to prove (iii). It is immediate from (i) and (4) that \(v^*_k < 0\) if and only if \(\lambda_k > 0\) and \(p^k \neq x^k\), which concludes the proof. ■
It follows from Lemma 3.3 that, either Algorithm 3 generates finite sequence \( \{x^k\}_{k \in \mathbb{N}} \subset C \), with the last iterate being a stationary point of problem (1), or it generates an infinite sequence. Henceforward, let \( \{p^k\}_{k \in \mathbb{N}} \subset C \) and \( \{x^k\}_{k \in \mathbb{N}} \subset C \) be sequences generated by Algorithm 3, which we assume to be infinite.

### 3.2. Asymptotic convergence analysis

In this section we study the asymptotic convergence properties of Algorithm 3. We begin by showing an important inequality for our analysis.

**Lemma 3.4:** The following inequality holds:

\[
f(x^{k+1}) \leq f(x^k) - \frac{1}{2} |v_k^*| \lambda_k, \quad k = 0, 1, \ldots
\]  

As a consequence, \( f(x^k) > f(x^{k+1}) \), for all \( k = 0, 1, \ldots \).

**Proof:** Let \( x^k \in C \) be defined in Algorithm 3 and \( v_k^* \) as in (3). First, remind the we are assuming that \( (x^k)_{k \in \mathbb{N}} \) is infinite. Thus, Lemma 3.3 implies that \( v_k^* < 0 \) and \( p^k \neq x^k \). Applying Lemma 3.1 with \( x = x^k \), \( p = p^k \) and \( \lambda = \lambda_k \), we have

\[
f(x^{k+1}) \leq f(x^k) + v_k^* \lambda_k + \frac{L}{2} \|p^k - x^k\|^2 \lambda_k^2.
\]  

(14)

We will consider two separate cases, namely, \( \lambda_k = |v_k^*|/(L\|p^k - x^k\|^2) \) and \( \lambda_k = 1 \). First, take \( \lambda_k = |v_k^*|/(L\|p^k - x^k\|^2) \). In this case, it follows from (14) that

\[
f(x^{k+1}) \leq f(x^k) - \frac{1}{2} |v_k^*| \lambda_k.
\]  

(15)

On the other hand, taking \( \lambda_k = 1 \), (14) becomes \( f(x^{k+1}) \leq f(x^k) - |v_k^*| + L\|p^k - x^k\|^2/2 \). In this case, (4) gives \( \lambda_k = 1 \leq |v_k^*|/(L\|p^k - x^k\|^2) \). Thus, in this case, we obtain \( f(x^{k+1}) \leq f(x^k) - (|v_k^*|/2)\lambda_k \). Therefore, the combination of this inequality with (15) yields (13). Since we are under the assumption \( v_k^* < 0 \), for all \( k = 0, 1, \ldots \), the second part follows and the proof is concluded. \( \blacksquare \)

Next result is a partial asymptotic convergence property of the Frank–Wolfe algorithm that requires neither convexity nor strong convexity on \( f \).

**Theorem 3.5:** Each limit point \( \bar{x} \in C \) of \( (x^k)_{k \in \mathbb{N}} \) is stationary for the problem (1).

**Proof:** Let \( \bar{x} \in C \) be a limit point of the sequence \( (x^k)_{k \in \mathbb{N}} \) and \( (x^{kj})_{j \in \mathbb{N}} \) be a subsequence of \( (x^k)_{k \in \mathbb{N}} \) such that \( \lim_{j \to \infty} x^{kj} = \bar{x} \). Hence, we have \( \lim_{j \to \infty} f(x^{kj}) = \)
Taking into account that Lemma 3.4 implies that \((f(x^k))_{k \in \mathbb{N}}\) is a decreasing sequence, we conclude that \((f(x^k))_{k \in \mathbb{N}}\) converges to \(f(\bar{x})\). Thus, in particular,

\[
\lim_{k \to \infty} [f(x^k) - f(x^{k+1})] = 0. \tag{16}
\]

Moreover, due to \((x^k)_{j \in \mathbb{N}}\) being bounded, it follows from the combination of the inclusion in (3) with item (iii) of Proposition 3.2 that \((p^{k_j})_{j \in \mathbb{N}}\) is also bounded. Let \((p^{k_j})_{j \in \mathbb{N}}\) be a subsequence of \((p^{k_i})_{j \in \mathbb{N}}\) such that \(\lim_{\ell \to \infty} p^{k_\ell} = \bar{p}\). If \(\bar{x} = \bar{p}\), then from (3) and the continuity of \(\nabla f\) we have \(\lim_{\ell \to \infty} v^*_k = \nabla f(x^{k_\ell})^T (p^{k_\ell} - x^{k_\ell}) = 0\). Now, assume that \(\bar{x} \neq \bar{p}\). Hence, combining Lemma 3.4 with (16), we conclude that \(\lim_{\ell \to \infty} |v^*_k|^2 / \lambda_{k_\ell} = 0\). On the other hand, using (4) we have

\[
|v^*_k|^2 / \lambda_{k_\ell} = \min \left\{ |v^*_k|^2, \frac{|v^*_k|^2}{L_\| p^{k_\ell} - x^{k_\ell} \|^2} \right\}. \tag{17}
\]

Considering that \(\lim_{\ell \to \infty} x^{k_\ell} = \bar{x}\), \(\lim_{\ell \to \infty} p^{k_\ell} = \bar{p}\) and \(\bar{x} \neq \bar{p}\) we obtain that \(\lim_{\ell \to \infty} \| p^{k_\ell} - x^{k_\ell} \| = \| \bar{x} - \bar{p} \| \neq 0\). Thus, due to \(\lim_{\ell \to \infty} |v^*_k|^2 / \lambda_{k_\ell} = 0\), it follows from (17) that \(\lim_{\ell \to \infty} |v^*_k| = 0\). We also know that the optimality of \(v^*_k\) in (3) yields

\[
v^*_k \leq \nabla f(x^{k_\ell})^T (p - x^{k_\ell}), \quad \forall p \in C. \tag{18}
\]

Since \(\lim_{\ell \to \infty} v^*_k = 0\), taking the limit in (18) and using the continuity of \(\nabla f\) we have \(\nabla f(\bar{x})^T (p - \bar{x}) \geq 0\), for all \(p \in C\). Therefore, \(\bar{x}\) is stationary for the problem (1).

Next we show that under convexity of \(f\) we can improve the previous result.

**Theorem 3.6:** The following statement holds:

(i) If \(f\) is a convex function and \(x^*\) is a cluster point of the sequence \((x^k)_{k \in \mathbb{N}}\), then \(x^*\) is a solution for the problem (1).

(ii) If \(f\) is a \(M\)-strongly convex function, then \((x^k)_{k \in \mathbb{N}}\) converges to a point \(x^* \in C\) which is a solution of problem (1). Moreover, \(\|x^k - x^*\| \leq \sqrt{2(f(x^k) - f(x^*)) / M}\), for all \(k = 0, 1, \ldots\).

**Proof:** For proving (i), assume that \(f\) is convex and \(x^*\) is a cluster point of \((x^k)_{k \in \mathbb{N}}\). Since \(f\) is convex, applying Proposition 2.2 with \(M = 0\) we obtain that \(f(p) \geq f(x^*) + \nabla f(x^*)^T (p - x^*)\), for all \(p \in C\). Therefore, considering that Theorem 3.5 implies that \(\nabla f(x^*)^T (p - x^*) \geq 0\), for all \(p \in C\), we conclude \(f(p) \geq f(x^*)\), for all \(p \in C\). Then, \(x^*\) is a solution for the problem (1). To prove item (ii), we first note that due to \(f\) be \(M\)-strongly convex the level set \(L_{f(x^0)} := \{x \in C : f(x) \leq f(x^0)\}\) is bounded. On the other hand, Lemma 3.4 implies that \((x^k)_{k \in \mathbb{N}} \subset L_{f(x^0)}\). Hence, \((x^k)_{k \in \mathbb{N}}\) is also bounded. Let \(x^* \in C\) a cluster point of \((x^k)_{k \in \mathbb{N}}\). It follows from
item (i) that $x^*$ is a solution of problem (1). Furthermore, combining (2) with Proposition 2.2 we obtain
\[ f(x^k) - f(x^*) \geq \frac{M}{2} \|x^* - x^k\|, \quad k = 0, 1, \ldots \] (19)

Taking into account that Lemma 3.4 implies that $(f(x^k))_{k \in \mathbb{N}}$ is a decreasing sequence, we have $\lim_{k \to +\infty} f(x^k) = f(x^*)$. Therefore, from (19) we conclude that $(x^k)_{k \in \mathbb{N}}$ converges to $x^*$. Finally, note that (19) is equivalent to the inequality in item (ii) and the proof is concluded. 

\[ \text{Theorem 3.7: The following statement holds:} \]

\begin{enumerate}
\item If $f$ is a convex function, then $f(x^k) - f^* \leq \Gamma^{-1}/k$, for all $k = 1, 2, \ldots$.
\item If $f$ is $M$-strongly convex, then $\|x^k - x^*\| \leq \sqrt{2f(\Gamma M)/\sqrt{k}}$, for all $k = 1, 2, \ldots$.
\end{enumerate}

\textbf{Proof:} To prove the item (i), we will prove first the following inequality
\[ \Gamma v_k^* \leq f(x^k) - f(x^{k+1}), \quad \forall \, k = 0, 1, \ldots \] (22)

where $\Gamma$ is defined in (21). By using Lemma 3.4, $\sigma$ defined in (20), and considering that $v_k^* < 0$, we conclude after some algebraic manipulations that
\[ \min \left\{ \frac{1}{2|v_k^*|}, \frac{1}{2L\sigma^2} \right\} v_k^* \leq f(x^k) - f(x^{k+1}), \quad \forall \, k = 0, 1, \ldots \] (23)

On the other hand, by combining (3) with (20) and the second equality in (21) we obtain $0 < |v_k^*| \leq \|\nabla f(x^k)\| \|x^k - p^k\| \leq \gamma \sigma$, for all $k = 0, 1, \ldots$, which implies
\[ \frac{1}{\gamma \sigma} \leq \frac{1}{|v_k^*|}, \quad k = 0, 1, \ldots \]

Thus, (22) follows from (23), previous inequality and (21). Now, it follows from item (ii) of Theorem 3.6 that $x^*$ is a solution of problem (1). Considering that $f$ is
convex, we have \( f^* = f(x^*) \geq f(x_k) + \nabla f(x_k)^T(x^* - x_k) \), for all \( k \). Thus, taking into account (3), \( f^* - f(x_k) \geq \nabla f(x_k)^T(x^* - x_k) \geq f(x_k)^T(p_k - x_k) = \nu_k^* \), for all \( k \). Since \( f^* \leq f(x_k) \) for all \( k \), we conclude that \( \nu_k^* \leq f^* - f(x_k) \leq 0 \) for all \( k \). Therefore, we obtain \( (f(x_k) - f^*)^2 \leq \nu_k^2 \), for all \( k = 0, 1, \ldots \), which together (22) yields

\[
(f(x^k) - f^*) - (f(x^{k+1}) - f^*) \geq \Gamma (f(x^k) - f^*)^2, \quad k = 0, 1, \ldots
\]  

(24)

Considering that \( \Gamma > 0 \) and defining \( a_k := f(x^k) - f^* \) we conclude from (24) that \( a_k - a_{k+1} \geq \Gamma a_k^2 \). Thus, the item (i) follows by applying Lemma 2.3. For proving the item (ii), we first note that the \( M \)-strong convexity of \( f \) and (2) imply that \( \|x^k - x^*\| \leq \sqrt{2(f(x^k) - f(x^*))/M} \), for all \( k = 0, 1, \ldots \). Therefore, item (ii) follows by using the inequality in item (i), and the proof is concluded. \( \blacksquare \)

4. Examples

In this section we present some examples of problem (1) with \( f \) satisfying (A) and (B) and \( C \) unbounded. We first present a general class of function \( f \) satisfying (A) and (B). For that, let \( K \subset \mathbb{R}^n \) be a closed convex cone such that \( \text{int} K \cap \text{int} K^* \neq \emptyset \), \( G : \mathbb{R}^n \to \mathbb{R}^n \) be a differentiable function and \( G' \) its Jacobian. Assume that for constants \( L_1 > 0 \) and \( L_2 > 0 \), the function \( G \) satisfies the following conditions:

\[
\begin{align*}
\text{(C1)} & \quad \|G(x) - G(y)\| \leq L_1\|x - y\|, \text{ for all } x, y \in K; \\
\text{(C2)} & \quad \|G'(x)x - G'(y)y\| \leq L_2\|x - y\|, \text{ for all } x, y \in K; \\
\text{(C3)} & \quad G(x) + G'(x)x \in K^*, \text{ for all } x \in K.
\end{align*}
\]

Let \( a \in \text{int} K^* \) and a function \( f : \mathbb{R}^n \to \mathbb{R} \) be defined by

\[
f(x) := a^Tx + G(x)^T x. \quad (25)
\]

**Lemma 4.1:** Let \( C \subset K \) be a closed and convex. Thus, the gradient \( \nabla f \) of \( f \) defined in (25) satisfies conditions (A) and (B) on \( C \).

**Proof:** First note that the gradient \( \nabla f \) of the function \( f \) defined in (25) is given by

\[
\nabla f(x) = a + G(x) + G'(x)x. \quad (26)
\]

Thus, using (C1) and (C2), we conclude that \( \|\nabla f(x) - \nabla f(y)\| \leq (L_1 + L_2)\|x - y\|, \text{ for all } x, y \in C \). Hence, \( \nabla f \) satisfies (A). Finally, due to \( a \in \text{int} K^* \), (C3) implies that \( \nabla f(x) \in \text{int} K^* \), for all \( x \in C \). On the other hand, since \( C \subset K \), we obtain that \( C_{\infty} \subset K \). Hence, \( K^* \subset C_{\infty}^* \), which implies that \( \text{int} K^* \subset \text{int}(C_{\infty}^*) \). Thus, we conclude that \( \nabla f(x) \in \text{int}(C_{\infty})^* \), for all \( x \in C \). Therefore, \( \nabla f \) satisfies (B) on \( C \). \( \blacksquare \)
Next we recall a well known result about Lipschitz functions, which is an immediate consequence of the mean value inequality.

**Theorem 4.2:** Let $C \subset \mathbb{R}^n$ be a convex set, $F : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function, and $F'$ its Jacobian. Assume that there exists a constant $L \geq 0$ such that $\|F'(x)\| \leq L$, for all $x \in C$. Then, $F$ is Lipschitz continuous with constant $L$ on $C$, i.e. $\|F(x) - F(y)\| \leq L\|x - y\|$, for all $x, y \in C$.

For proceeding we also need a result about characterization of convex functions, which the proof can be found in [16, Theorem 4.1, p.190].

**Theorem 4.3:** Let $C \subset \mathbb{R}^n$ be a convex set, $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function, $\nabla^2 f$ its Hessian and $M \geq 0$ be a constant. Then, $f$ is $M$-strongly convex on $C$ if and only if $\nabla^2 f(x)v \geq M\|v\|^2$, for all $x \in C$ and $v \in \mathbb{R}^n$.

In the following we present specific examples of functions $G$ satisfying (C1)-(C3) on the cone $K = \mathbb{R}^n_+$ such that $f$ in (25) is convex and $\nabla f$ satisfies (A) and (B), for any closed convex set $C \subset \mathbb{R}^n_+$.

**Example 4.4:** Let $Q = (q_{ij}) \in \mathbb{R}^{n \times n}$ with the entries $q_{ij} \geq 0$, for all $i, j$ and $a \in \mathbb{R}^{n_+}$. Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be a linear function defined by $G(x) = Qx$. Direct calculations show that $\|G(x) - G(y)\| = \|G'(x)x + G'(y)y\| \leq \|Q\|\|x - y\|$, for all $x, y \in \mathbb{R}^n$. Thus, $G$ satisfies (C1) and (C2) with $L_1 = L_2 = \|Q\|$, for any cone $K$. Since $q_{ij} \geq 0$ and $a \in \mathbb{R}^{n_+}$, then $G$ also satisfies (C3) in $K = \mathbb{R}^n_+$. Moreover, if $Q$ satisfies the condition $\nabla^2 Qv \geq M\|v\|^2$, for all $v \in \mathbb{R}^n$ and some $M \geq 0$, then Theorem 4.3 implies that the problem (1) with the associated quadratic function $f(x) := a^T x + x^T Q x$, is $M$-strong convex. Moreover, Lemma 4.1 implies that $\nabla f(x) = a + Qx$ satisfies (A) and (B), for any closed and convex set $C \subset \mathbb{R}^n_+$.

**Example 4.5:** Let $e := (1, \ldots, n) \in \mathbb{R}^n$ be a vector, $\alpha > 0$ and $\beta > 0$ be constants satisfying $2\alpha > 3\beta^{3/2}/\sqrt{n}$. Then, the function $G : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$G(x) := \alpha x + \frac{\beta}{\sqrt{1 + \beta x^T x}} e,$$  

satisfies (C1)-(C3) in $K = \mathbb{R}^n_+$. Moreover, the problem (1) with the associated function

$$f(x) := a^T x + \alpha x^T x + \frac{\beta}{\sqrt{1 + \beta x^T x}} e^T x,$$  

with $a \in \mathbb{R}^{n_+}$ is $M$-strong convex. Moreover, $\nabla f$ satisfies (A) and (B). Indeed, we first note that $f(x) := a^T x + G(x)^T x$. Some calculations show that

$$G'(x) = \alpha I_n + \frac{-\beta^2}{(1 + \beta x^T x)^{3/2}} e x^T, \quad G(x) + G'(x)x = 2\alpha x + \frac{\beta}{(1 + \beta x^T x)^{3/2}} e.$$  

(29)
where $I_n \in \mathbb{R}^{n \times n}$ is the identity matrix. Thus, $\nabla f(x) = a + G(x) + G'(x)x$. Hence, after some calculations, we have

$$\nabla^2 f(x) = 2\alpha I_n + \frac{-3\beta^2}{(1 + \beta x^T x)^{3/2}} ex^T. \tag{30}$$

The first equality in (29) yields $\|G'(x)\| \leq \beta^{3/2}\sqrt{n} + \alpha$, for all $x \in C$ and (30) implies

$$0 < (2\alpha - 3\beta^{3/2}\sqrt{n})|v|^2 \leq v^T \nabla^2 f(x)v \leq (2\alpha + 3\beta^{3/2}\sqrt{n})|v|^2, \tag{31}$$

for all $x \in C$ and all $v \in \mathbb{R}^n$. Considering that $\|G'(x)\| \leq \beta^{3/2}\sqrt{n} + \alpha$, it follows from Theorem 4.2 that $G$ also satisfies (C1) with $L_1 = \beta^{3/2}\sqrt{n} + \alpha$. In particular, (31) implies that $\|\nabla^2 f(x)\| \leq 2\alpha + 3\beta^{3/2}\sqrt{n}$, for all $x \in C$. Hence, taking into account that $\|G'(x)\| \leq \beta^{3/2}\sqrt{n} + \alpha$, for all $x \in C$, Theorem 4.2 also implies that

$$\|G'(x)x - G'(y)y\| \leq \|G(x) - G(y)\| + \|\nabla f(x) - \nabla f(y)\|$$

$$\leq (4\beta^{3/2}\sqrt{n} + 3\alpha) \|x - y\|,$$

for all $x, y \in C$. Thus, $G$ also satisfies (C2) with $L_2 = 4\beta^{3/2}\sqrt{n} + 3\alpha$. Second inequality in (29) implies that $G(x) + G'(x)x \in \mathbb{R}^n_+$, for all $x \in \mathbb{R}^n_+$. Since $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, $G$ satisfies (C3). Therefore, Lemma 4.1 implies that $\nabla f$ of $f$ in (28) satisfies conditions (A) and (B), for any closed and convex set $C \subset \mathbb{R}^n_+$. Finally, using (31), it follows from Theorem 4.3 that $f$ in (28) is $M$-strong convex with $M = 2\alpha - 3\beta^{3/2}\sqrt{n}$.

In the next example we present directly a convex function $f$ satisfying (A) and (B) in closed convex sets $C \subset \mathbb{R}^n_+$.

**Example 4.6:** Let $\beta > 0, a \in \mathbb{R}^n_+$ and $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) := a^T x + \sqrt{1 + \beta x^T x}. \tag{32}$$

Note that, in this case, the gradient and the Hessian of $f$ are given, respectively, by

$$\nabla f(x) = a + \frac{\beta}{\sqrt{1 + \beta x^T x}}x, \quad \nabla^2 f(x) = \frac{\beta}{\sqrt{1 + \beta x^T x}}I_n - \frac{\beta^2}{(1 + \beta x^T x)^{3/2}}xx^T,$$

where $I_n \in \mathbb{R}^{n \times n}$ is the identity matrix. Some calculations show that

$$\frac{\beta}{(1 + \beta x^T x)^{3/2}}v^T v \leq v^T \nabla^2 f(x)v \leq \beta v^T v, \quad \forall v \in \mathbb{R}^n,$$

which implies that $\nabla^2 f(x)$ is positive definite and $\|\nabla^2 f(x)\| \leq \beta$. Thus, using Theorems 4.3 and 4.2, we conclude that $f$ is convex and $\nabla f$ is Lipschitz continuous with constant $\beta$. On the other hand, for any closed and convex set
C ⊂ \mathbb{R}^n_+, we have \( C_\infty \subset \mathbb{R}^n_+ \). Hence, \( \mathbb{R}^n_+ = (\mathbb{R}^n_+)^* \subset C_\infty^* \), which implies that \( \mathbb{R}^n_+ \subset int(C_\infty^*) \). Considering that \( \nabla f(x) \in \mathbb{R}^n_+ \), for all \( x \in C \), we conclude that \( \nabla f(x) \in int(C_\infty^*) \), for all \( x \in C \). Finally, due \( g \) be convex, we have that \( C \) is convex. Therefore, the problem (1) with the objective function (32) is convex and \( f \) satisfies conditions (B) and (A).

Let us preset two more examples of convex functions and the respective convex sets satisfying conditions (B) and (A).

1. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be defined by \( f(x) = \ln(e^{x_1} + e^{x_2} + \ldots + e^{x_n}) \). This function satisfies (A); see [19, Example 5.15, p.115]. Some calculations show that this function also satisfies (B), for any convex set \( C \subset \mathbb{R}^n_+ \).

2. Let \( C \subset \mathbb{R}^n_+ \) be closed convex and \( d_C(x) := \min_{y \in C} \|x - y\| \), for \( x \in \mathbb{R}^n \). Define the convex function \( \psi_C : \mathbb{R}^n \to \mathbb{R} \) by \( \psi_C(x) := \frac{1}{2}\|x\|^2 - \frac{1}{2}d_C^2(x) \), see [19, Example 2.17.4, p.22]. We can prove that \( \nabla \psi_C(x) = P_C(x) \), where \( P_C \) denotes the orthogonal projection onto \( C \); see [19, Example 3.49, p.61]. Since \( \mathbb{R}^n_+ \subset int(C_\infty^*) \) and \( C \subset \mathbb{R}^n_+ \) we conclude that \( \nabla \psi_C(x) \in C_\infty^* \), which implies that \( \psi_C \) satisfies (B). Moreover, the non-expansivity of the projection implies that \( \psi_C \) also satisfies (A).

We end this section presenting examples of unbounded convex set that appear as constraints of optimization problems.

1. \( C := \{x \in \mathbb{R}^n_+ : 1 \leq x_1 \ldots x_n\} \);
2. \( C := \{x \in \mathbb{R}^n_+ : 1 \leq x_1 + \ldots + x_n\} \);
3. \( C := \{x \in \mathbb{R}^n_+ : b \leq Ax\} \), where \( A = (a_{ij}) \in \mathbb{R}^{n \times n} \) with \( a_{ij} > 0 \) and \( b \in \mathbb{R}^n_+ \).

Let \( \Omega \subset \mathbb{R}^n \) be a closed and convex set and \( g : \Omega \to \mathbb{R} \) be a convex function. The epigraph of \( g \) is defined by

\[
\text{epi}(g) := \{(x, t) \in \Omega \times \mathbb{R} : g(x) \leq t\}.
\]

The set \( C = \text{epi}(g) \subset \mathbb{R}^n \times \mathbb{R} \) is convex and unbounded. Next some specific examples.

1. (Lorentz cone) \( \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_2 \leq t\} \), where \( \| \cdot \|_2 \) denotes the 2-norm;
2. \( \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_1 \leq t\} \), where \( \| \cdot \|_1 \) denotes the 1-norm;
3. \( \{(x_1, \ldots, x_n, t) \in \mathbb{R}^n_+ \times \mathbb{R} : 1/(x_1 \ldots x_n) \leq t\} \).

We point out that projecting on above sets is not a very expensive task; see [19, Chapter 6].
5. Conclusions

In this paper, we consider the classical Frank–Wolf algorithm for non-empty closed convex restriction, no necessarily compact. In order to provide its convergence properties, we develop results using recession techniques. The examples, given in Section 4, show that the Frank–Wolf algorithm, in fact, can be applied to solve many optimization problems, non-necessarily convex, with non-compact constraint sets.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The first author was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grants 305158/2014-7, 302473/2017-3, FAPEG/PRONEM-201710267000532 and CAPES. The second author was supported in part by Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF) by the grant 0193.001695/2017, PDE 05/2018. This research was carried out, in part, during (the state of alert in the Catalonia) a visit, of the second author to the Centro de Recerca Matemática (CRM), in the framework of the Research in pairs call in 2020. The CRM is a paradise for research, the author appreciates the hospitality and all the support received from CRM.

ORCID

W. S. Sosa http://orcid.org/0000-0001-5697-5710

References

[1] Frank M, Wolfe P. An algorithm for quadratic programming. Nav Res Log. 1956;3:95–110.
[2] Demyanov VF, Rubinov AM. Approximate methods in optimization problems. New York: American Elsevier Publishing Co. Inc.; 1970.
[3] Levitin E, Polyak B. Constrained minimization methods. USSR Comput Math Math Phys. 1966;6(5):1–50.
[4] Jaggi M. Revisiting Frank–Wolfe: projection-free sparse convex optimization. Proceedings of the 30th International Conference on International Conference on Machine Learning; Vol. 28, ICML’13; 2013. p. I–427–I–435.
[5] Lacoste-Julien S, Jaggi M. On the global linear convergence of Frank–Wolfe optimization variants; 2015. arXiv:1511.05932.
[6] Lan G. The complexity of large-scale convex programming under a linear optimization oracle; 2013. arXiv:1309.5550.
[7] Beck A, Teboulle M. A conditional gradient method with linear rate of convergence for solving convex linear systems. Math Methods Oper Res. 2004;59(2):235–247.
[8] Harchaoui Z, Juditsky A, Nemirovski A. Conditional gradient algorithms for norm-regularized smooth. Convex Optim Math Program. 2015;152(1–2, Ser. A):75–112.
[9] Boyd N, Schiebinger G, Recht B. The alternating descent conditional gradient method for sparse inverse problems. SIAM J Optim. 2017;27(2):616–639.
[10] Luss R, Teboulle M. Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint. SIAM Rev. 2013;55(1):65–98.

[11] Freund RM, Grigas P, Mazumder R. An extended Frank-Wolfe method with ‘in-face’ directions, and its application to low-rank matrix completion. SIAM J Optim. 2017;27(1):319–346.

[12] Konnov IV. Simplified versions of the conditional gradient method. Optimization. 2018;67(12):2275–2290.

[13] Ghadimi S. Conditional gradient type methods for composite nonlinear and stochastic optimization. Math Program. 2019;173(1–2, Ser. A):431–464.

[14] Lan G, Zhou Y. Conditional gradient sliding for convex optimization. SIAM J Optim. 2016;26(2):1379–1409.

[15] Hiriart-Urruty JB, Lemaréchal C. Fundamentals of convex analysis. Berlin: Springer-Verlag; 2001. Grundlehren Text Editions; abridged version of it Convex analysis and minimization algorithms. I, Springer, Berlin, 1993.

[16] Hiriart-Urruty JB, Lemaréchal C. Convex analysis and minimization algorithms I. Berlin: Springer-Verlag; 1993. (Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]; 305).

[17] Polyak BT. Introduction to optimization. New York: Optimization Software; 1987. (Translations Series in Mathematics and Engineering).

[18] Dennis Jr JE, Schnabel RB. Numerical methods for unconstrained optimization and nonlinear equations. Philadelphia (PA): Society for Industrial and Applied Mathematics (SIAM); 1996. (Classics in Applied Mathematics; 16)

[19] Beck A. First-order methods in optimization. Philadelphia (PA): Society for Industrial and Applied Mathematics (SIAM); 2017. (Mathematical Optimization Society(MOS)-SIAM Series on Optimization; 25).