Operations Research Letters 50 (2022) 517–523

Contents lists available at ScienceDirect

Operations Research Letters

www.elsevier.com/locate/orl

The regularized feasible directions method for nonconvex optimization

Amir Beck [a,*], Nadav Hallak [b]

[a] *School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv 69978, Israel*
[b] *Faculty of Industrial Engineering and Management, Technion Israel Institute of Technology, Haifa 3200003, Israel*

ARTICLE INFO

Article history:
Received 12 May 2022
Received in revised form 15 June 2022
Accepted 15 July 2022
Available online 29 July 2022

Keywords:
Feasible directions
Nonconvex optimization
Nonsmooth optimization
Constrained optimization
Convergence analysis

ABSTRACT

This paper develops and studies a feasible directions approach for the minimization of a continuous function over linear constraints in which the update directions belong to a predetermined finite set spanning the feasible set. These directions are recurrently investigated in a cyclic semi-random order, where the stepsize of the update is determined via univariate optimization. We establish that any accumulation point of this optimization procedure is a stationary point of the problem, meaning that the directional derivative in any feasible direction is nonnegative. To assess and establish a rate of convergence, we develop a new optimality measure that acts as a proxy for the stationarity condition, and substantiate its role by showing that it is coherent with first-order conditions in specific scenarios. Finally we prove that our method enjoys a sublinear rate of convergence of this optimality measure in expectation.

1. Introduction

1.1. Problem formulation

In this paper, we seek to address the nonconvex, and possibly nonsmooth, problem

$$\min\{f(\mathbf{x}) := h(\mathbf{x}) - g(\mathbf{x}) : \mathbf{A}\mathbf{x} \le \mathbf{b}\}, \tag{P}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$, and $C := \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \le \mathbf{b}\right\}$ is a nonempty feasible set, and we assume throughout the paper that

1. $f : C \to \mathbb{R}$ is lower-bounded over C by $\inf_{\mathbf{x}\in C} f = \bar{f} > -\infty$,
2. $h : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable,
3. $g : \mathbb{R}^n \to \mathbb{R}$ is convex.

The underlying model (P) captures a wide variety of problems, most prominently those in which f is a non-smooth concave function or a continuously differentiable function that does not necessarily have a Lipschitz continuous gradient.

Considering the nonconvexity of (P) and the fact that the objective may not be differentiable, we aim to obtain a *stationary point* of (P), where stationarity at a point $\mathbf{x} \in C$ means that the directional derivative with respect to any feasible direction at $\mathbf{x}$ is nonnegative; this is defined formally in the following Section 2. We seek to achieve stationarity by using an optimization protocol that is oblivious to full first-order information, making it applicable to non-differentiable nonconvex problems, and robust to differentiable problems with no Lipschitz continuity of the gradient.

Literature. The feasible directions approach, see e.g. [18, Chapter 13], [8, Chapter 2], [9, Chapter 7], or the review paper [14], is a classical methodology in which the optimization procedure is updated along a chosen feasible direction. Variations of this approach include zeroth-order methods using directions generator such as those described in [14], first-order procedures such as the ε-perturbation method [18, Chapter 13.4] and the conditional gradient [8, Chapter 2.2], or second-order methods such as the general algorithm [2] and the two directions method [12].

A common difficulty in feasible directions methods for constrained optimization is the jamming phenomenon in which the algorithm gets stuck in sub-optimal points as a result of the fact that the mapping defining the update procedure is not closed [18, Chapter 13]. The remedy for this unwanted scenario comes in the form of taking into account sufficiently close constraints, as done in the ε-perturbation method [18, Chapter 13.4] or in the GFD/RFD methods [6].

The starting point of this work is our recent work [6] in which both deterministic (GFD) and random (RFD) methods were developed to address problem (P). The optimization procedures in [6] first compute an ε-feasible direction (via approximately active constraints due to the jamming phenomenon), and then minimize a consistent majorizer (cf. [7,16]) along that direction. Almost

[*] Corresponding author.
E-mail addresses: becka@tauex.tau.ac.il (A. Beck), ndvhllk@technion.ac.il (N. Hallak).

https://doi.org/10.1016/j.orl.2022.07.005
0167-6377/© 2022 Elsevier B.V. All rights reserved.

sure subsequential convergence to stationary points was proved in [6] under the assumption that the proximity parameter ε with which the ε-feasible directions are generated is smaller than half the distance between the accumulation point and its closest non-active constraint. This assumption on ε incorporates a difficulty as, in some sense, it requires unavailable information on accumulation points of the generated sequence. Additionally, the work [6] does not establish any rates of convergence of the suggested method.

In this paper we introduce a new feasible directions method (see Section 3), called the *regularized feasible directions search (RFDS) method* that optimizes at each update step a regularized version of the objective function along one of the directions in the underlying directions set. The method avoids the challenges of the jamming phenomenon, and any required treatment, by always updating along directions from a predetermined finite set of directions, so that the update direction does not depend on the given point. We establish in Section 5 that the RFDS method achieves subsequence convergence to a stationary point without any conditions on the parameters of the method, and although the method contains a random element, the convergence result is deterministic. Moreover, we develop a new optimality measure in Section 4 that can be thought of as a measure for stationarity, and show that in expectation, it converges to zero in rate of $O(1/k)$.

Notation. Matrices and vectors are denoted by boldface letters. The nonnegative orthant is denoted by $\mathbb{R}^n_+$. For any natural number l, we denote $[l] = \{1, 2, \dots, l\}$. The vector $\mathbf{e}_i$ has 1 in the i-th component and zeros elsewhere. We assume throughout the paper that the underlying norm on $\mathbb{R}^n$ is the l_2-norm. For $\alpha \in \mathbb{R}$, we use the notation $[\alpha]_+ := \max\{\alpha, 0\}$.

2. Mathematical preliminaries: feasible directions, stationarity and spanning sets

Adopting the classical terminology of [18, Section 2.4], a vector $\mathbf{d} \in \mathbb{R}^n$ is a feasible direction at $\mathbf{x} \in C$ if there exists $\sigma > 0$ such that $\mathbf{x} + \tau\mathbf{d} \in C$ for all $\tau \in [0, \sigma]$. The set of all feasible directions at a given point $\mathbf{x} \in C$ is a cone called the *cone of feasible directions*, and by exploiting the specific structure of C, it can be shown (see e.g., [4, Lemma 10.1.2]) to have the following form:

$$D_{\mathbf{x}} = \{\mathbf{d} \in \mathbb{R}^n : \mathbf{a}_i^T\mathbf{d} \le 0, i \in I(\mathbf{x})\}, \tag{2.1}$$

where $I(\mathbf{x}) \equiv \{i \in [m] : \mathbf{a}_i^T\mathbf{x} = b_i\}$ is the set of active constraints at $\mathbf{x}$. Obviously, if $I(\mathbf{x}) = I(\mathbf{y})$, then $D_{\mathbf{x}} = D_{\mathbf{y}}$.

Using the set of feasible directions, we can define the *stationarity* condition.

Definition 2.1 *(Stationarity).* A point $\mathbf{x} \in C$ is called a stationary point of (P) if

$$f'(\mathbf{x}; \mathbf{d}) \ge 0 \text{ for any } \mathbf{d} \in D_{\mathbf{x}}.$$

Obviously, stationarity is a necessary optimality condition (see e.g., [18, Lemma 2.11]).

Lemma 2.1. *Let $\mathbf{x}^* \in C$ be a local minimum of (P). Then $\mathbf{x}^*$ is a stationary point of (P).*

We address the task of finding stationary points of (P) by exploring a finite number of directions that *positively span* (cf. [10,17,14]) the set of feasible directions. We first recall the definition of the positive span.

Definition 2.2 *(Positive span [17, Definition 2.3]).* The **positive span** of a finite set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\} \subseteq \mathbb{R}^n$, denoted by $\mathrm{pos}(S)$, is the convex cone given by

$$\mathrm{pos}(S) := \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{v}_i : \lambda_i \ge 0, i = 1, 2, \dots, k \right\}.$$

A linear combination with nonnegative coefficients is called a *positive linear combination*, and thus $\mathrm{pos}(S)$ comprises all positive linear combinations of vectors from S.

Definition 2.3 *(Positive spanning set [17, Definition 2.4]).* A finite set $S \subseteq \mathbb{R}^n$ is **a positive spanning set** of a convex cone $C \subseteq \mathbb{R}^n$ if $\mathrm{pos}(S) = C$. In this case, S is said to **positively span** C.

A finite set of vectors that spans the set of feasible directions at a point $\mathbf{x} \in C$ is called a *positive spanning feasible directions (PSD) set*. The notion of PSD sets is borrowed from [6] with the small adjustment that here the directions in the PSD set are normalized.

Definition 2.4 *(PSD sets).* Let $\mathbf{x} \in C$ and let $D_{\mathbf{x}}$ be the corresponding cone of feasible directions. Then a finite set $V_{\mathbf{x}} \subseteq D_{\mathbf{x}}$ containing normalized vectors that positively span $D_{\mathbf{x}}$ is called a **positive spanning feasible directions (PSD) set of C at $\mathbf{x}$.**

The key for using our approach is the fact that stationarity of a point $\mathbf{x} \in C$ w.r.t. problem (P) can be verified by checking that the directional derivatives of h at $\mathbf{x}$ in all the directions of a given PSD set of C at $\mathbf{x}$ are nonnegative. This is a significant simplification of the stationarity condition (Definition 2.1) that requires the verification of the directional derivative condition (2.1) with respect to *all* the feasible directions.

Theorem 2.1 *(Stationarity via PSD sets [6, Theorem 3.1]). Let $\mathbf{x}^* \in C$ and $V_{\mathbf{x}^*} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a PSD set of C at $\mathbf{x}^*$. Then $\mathbf{x}^*$ is a stationary point of (P) if and only if*

$$f'(\mathbf{x}^*; \mathbf{v}_i) \ge 0, \quad i = 1, 2, \dots, n. \tag{2.2}$$

Example 2.1 *(Box constraints).* Consider the box set $C = \mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}] := \{\mathbf{x} : \ell_i \le x_i \le u_i, i \in [n]\}$, where here $\boldsymbol{\ell}, \mathbf{u} \in \mathbb{R}^n$ are such that $\boldsymbol{\ell} < \mathbf{u}$. Given $\mathbf{x} \in C$, the set of feasible directions is given by

$$D_{\mathbf{x}} = \{\mathbf{d} \in \mathbb{R}^n : d_i \ge 0, d_j \le 0, i \in I_l(\mathbf{x}), j \in I_u(\mathbf{x})\},$$

where $I_l(\mathbf{x}) = \{i : x_i = \ell_i\}$, $I_u(\mathbf{x}) = \{i : x_i = u_i\}$. A PSD set of C at $\mathbf{x}$ is

$$V_{\mathbf{x}} = \{\mathbf{e}_i : i \in I_l(\mathbf{x})\} \cup \{-\mathbf{e}_i : i \in I_u(\mathbf{x})\} \cup \{\pm\mathbf{e}_i : i \in I_b(\mathbf{x})\},$$

where $I_b(\mathbf{x}) = \{i : \ell_i < x_i < u_i\}$. By Theorem 2.1, a point $\mathbf{x}^* \in \mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]$ is a stationary point of (P) if and only if for $i \in I_l(\mathbf{x}^*)$, $j \in I_u(\mathbf{x}^*)$, $k \in I_b(\mathbf{x}^*)$,

$$f'(\mathbf{x}^*, \mathbf{e}_i) \ge 0,\ f'(\mathbf{x}^*, -\mathbf{e}_j) \ge 0,\ f'(\mathbf{x}^*, \pm\mathbf{e}_k) \ge 0.$$

In the case where $g \equiv 0$, meaning that $f = h$ is continuously differentiable, the above translates to the well known condition

$$\nabla_i f(\mathbf{x}^*) \begin{cases} \ge 0, & i \in I_l(\mathbf{x}^*), \\ \le 0, & i \in I_u(\mathbf{x}^*), \\ = 0, & i \in I_b(\mathbf{x}^*). \end{cases}$$

Example 2.2 *(Affine constraints).* Consider the affine subspace $C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{D}\mathbf{x} = \mathbf{b}\}$, where $\mathbf{D} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then at any point

$\mathbf{x} \in C$, it holds that the cone of feasible directions is given by $D_{\mathbf{x}} = \{\mathbf{d} : \mathbf{D}\mathbf{d} = \mathbf{0}\}$. A PSD set of C at *any* $\mathbf{x}$ is $\{\pm\mathbf{v}_1, \pm\mathbf{v}_2, \ldots, \pm\mathbf{v}_k\}$, where $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a basis for the null space of $\mathbf{D}$. By Theorem 2.1, $\mathbf{x}^* \in C$ is a stationary point of (P) if and only if

$$f'(\mathbf{x}^*, \pm\mathbf{v}_i) \geq 0, i = 1, 2, \ldots, k.$$

In the case where $g \equiv 0$, meaning $f = h$ is continuously differentiable, the above translates to

$$\langle \nabla f(\mathbf{x}^*), \mathbf{v}_i \rangle = 0, i = 1, 2, \ldots, k.$$

We note that finding a positive spanning set for convex cones such as the $D_{\mathbf{x}}$ set, is a classical task known in the literature as the representation conversion problem for convex cones, from the so-called $\mathcal{H}$-representation to the so-called $\mathcal{V}$-representation (this conversion problem is also known as the *vertex enumeration problem*), and as such, has well established methods and implementations; for more details see [19, Chapter 1] or [11, Section 9], as well as [3,1], or [14, Section 8]. We further note that the vertex enumeration problem can be tractable or intractable depending on the geometry of the feasible set; tractability is achieved, for example, when the number of constraints is small for example; for a class of sets in which the vertex enumeration problem is intractable see [13].

3. The regularized feasible directions search (RFDS) method

3.1. Complete feasible directions sets

For any $\mathbf{x} \in C$, we denote by $V_{\mathbf{x}}$ a PSD set of C at $\mathbf{x}$. Recall that $V_{\mathbf{x}}$ is a positive spanning set of $D_{\mathbf{x}}$ and that $D_{\mathbf{x}}$ is essentially determined by its active set $I(\mathbf{x}) \subseteq [m]$ in the sense that $D_{\mathbf{x}} = D_{\mathbf{y}}$ whenever $I(\mathbf{x}) = I(\mathbf{y})$. We assume that $V_{\mathbf{x}}$ is arbitrarily chosen, but that the choice is only determined by the set of active constraints $I(\mathbf{x})$, meaning that $V_{\mathbf{x}} = V_{\mathbf{y}}$ if $I(\mathbf{x}) = I(\mathbf{y})$.

Since the number of subsets of $[m]$ is finite (2^m), it follows that the number of possible PSD sets $V_{\mathbf{x}}, \mathbf{x} \in C$, is bounded. We can thus define the *complete feasible directions set* of C as the union of the PSD sets over all feasible points:

$$V_C = \bigcup_{\mathbf{x} \in C} V_{\mathbf{x}}.$$

Due to the fact that any PSD set is finite, and, as was discussed above, the number of PSD sets is also finite, it follows that V_C is finite. This set will be a fundamental ingredient in our algorithm.

Example 3.1 *(Box constraints).* Continuing Example 2.1, the complete feasible directions set for the box set is $V_C = \{\pm\mathbf{e}_1, \pm\mathbf{e}_2, \ldots, \pm\mathbf{e}_n\}$.

Example 3.2 *(Affine constraints).* Continuing Example 2.2, since $V_{\mathbf{x}} = \{\pm\mathbf{v}_1, \pm\mathbf{v}_2, \ldots, \pm\mathbf{v}_k\}$ for any $\mathbf{x}$ in the affine subspace, it follows that $V_C = \{\pm\mathbf{v}_1, \pm\mathbf{v}_2, \ldots, \pm\mathbf{v}_k\}$.

3.2. The RFDS method

To find a stationary point of (P), our proposed method described by Algorithm 1 executes a recurrent univariate optimization in directions belonging to a complete feasible directions set

$$V_C = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$$

that is computed in a preprocess. The optimization is done with respect to a regularized version of the objective function, and thus the method is called the *regularized feasible directions search* (RFDS) method. The process is composed of cycles where in each cycle, every direction is examined exactly once. The order is determined by an arbitrary permutation and the only requirement imposed on the permutation is that the first direction is chosen randomly.

Algorithm 1: Regularized Feasible Directions Search (RFDS).

```
Input: $\mathbf{x}^0 \in C$; $r \in (0, \infty]$, $\gamma > 0$.
Initialize: Generate a complete feasible directions set: $V_C = \{v_1, v_2, \ldots, v_s\}$;
for any $k \geq 0$ do
    choose the first direction $p_1 \in [s]$ randomly via a uniform distribution
      over $[s]$ ;
    determine the order of the remaining directions $\{p_i\}_{i=2}^s$ s.t.
      $(p_1, p_2, \ldots, p_s)$ constitutes a permutation of the set $[s]$ ;
    set $\mathbf{y}_0^k \leftarrow \mathbf{x}^k$;
    for any $i \in [s]$ do
        calculate the stepsize for direction $\mathbf{v}_{p_i}$:
          $q(i,k) \in \operatorname{argmin}\left\{f(\mathbf{y}_{i-1}^k + t\mathbf{v}_{p_i}) + \frac{\gamma}{2}t^2 : t \in C_{t,i}\right\}$, where
          $C_{t,i} := \{t \in [0,r] : \mathbf{A}(\mathbf{y}_{i-1}^k + t\mathbf{v}_{p_i}) \leq \mathbf{b}\}$;
        update $\mathbf{y}_i^k \leftarrow \mathbf{y}_{i-1}^k + q(i,k)\mathbf{v}_{p_i}$;
    end
    update: $\mathbf{x}^{k+1} \leftarrow \mathbf{y}_s^k$;
end
```

Several comments regarding the schematics of Algorithm 1 are in order.

Remark 3.1 *(Case $r = \infty$).* We allow r to be equal to ∞, and in this case the one-dimensional optimization is performed over the entire nonnegative part of the real line. We note that in this case, with admittedly some abuse of notation, the relation $t \in [0, r]$ means that $t \in [0, \infty)$.

Remark 3.2 *(Stepsize for infeasible directions).* The stepsize is determined by solving the following univariate optimization problem:

$$\min\left\{f(\mathbf{x}^k + t\mathbf{d}) + \frac{\gamma}{2}t^2 : t\mathbf{A}\mathbf{d} \leq \mathbf{b} - \mathbf{A}\mathbf{x}^k, t \in [0, r]\right\}. \tag{3.1}$$

Hence, if the direction $\mathbf{d}$ is not feasible, then the solution to (3.1) is simply $t = 0$. Since the order of the directions in the cycle, except for the first direction, can be chosen arbitrarily, in practice directions can be omitted from the cycle if at one point in the cycle they are not feasible directions.

Remark 3.3 *(Grouping opposite directions).* Sometimes, e.g., as occurs in Examples 3.1 and 3.2, we have the situation in which $\mathbf{v}, -\mathbf{v} \in V_C$ for some directions $\mathbf{v}$. In this case, since the order of the directions in each cycle can be chosen arbitrarily, we can group these pairs of opposite directions together, and then perform the one-dimensional optimization with respect to $\mathbf{v}$ over the extended interval $[-r, r]$ instead of $[0, r]$.

Remark 3.4 *(Worst-case single loop complexity).* We note that the number of constraints forming the feasible set can lead to a non-polynomial number of vectors in the set V_C, allowing for a scenario in which a single loop might examine a non-polynomial number of directions. In that regard, Algorithm 1 is suited for problems with tractable complete feasible directions sets.

Example 3.3 *(Coordinate-wise optimization).* Suppose that $V = \text{Box}[\boldsymbol{\ell}, \mathbf{u}]$ as defined in Example 2.1. Assuming that we group opposite directions as discussed in Remark 3.3, then Algorithm 1 is a coordinate descent-type method in which the first coordinate at each cycle is picked randomly via a uniform distribution and all other coordinates are picked arbitrary. One difference between the suggested method and "standard" coordinate descent methods is that the regularization term $\frac{\gamma}{2}t^2$ is added in the one-dimensional optimization procedure.

Example 3.4 *(Quadratic minimization over linear constraints).* Suppose that $f(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x}+\mathbf{g}^T\mathbf{x}+c$. Then the stepsize is calculated by solving a univariate quadratic problem of the form $\min_t\{c_1t^2+c_2t+c_3 : t \in [0, l_{\mathbf{x},\mathbf{v}}]\}$, where $l_{\mathbf{x},\mathbf{v}} \in [0,r]$ is determined according to the point $\mathbf{x}$ and the direction $\mathbf{v}$. Hence, the stepsize receives one of the values in the set $\{0, -c_2/(2c_1), l_{\mathbf{x},\mathbf{v}}\}$, giving it a closed-form solution.

4. The optimality measure

4.1. Definition

One aspect of the theoretical performance guarantees of an algorithm is the rate at which it converges to the designated optimality condition. The rate is given with respect to a proxy of the optimality condition, usually referred to as the *optimality measure*, see e.g., the gradient mapping [15, Section 2.2.3], the conditional gradient norm [5, Definition 13.2], or the first/second-order optimality measures in [12, Section 4.1]. Optimality measures should be nonnegative, and be equal zero only at points satisfying the designated optimality condition.

To analyze the rate of the RFDS method, we define an optimality measure that quantifies the best improvement in the regularized function value in all directions belonging to a given complete feasible directions set.

Definition 4.1 *(Optimality measure).* Given $r \in (0,\infty]$, $\mathbf{x} \in C$, and a corresponding complete feasible directions set $V_C = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$, define for any $i \in [s]$,

$$\eta_i^{\gamma}(\mathbf{x}; V, r) = \min\left\{z : z \in \operatorname*{argmin}_{t\in C_i}\left\{f(\mathbf{x}+t\mathbf{v}_i)+\frac{\gamma}{2}t^2\right\}\right\}, \tag{4.1}$$
$$C_i = \{t \in [0,r] : \mathbf{A}(\mathbf{x}+t\mathbf{v}_i) \le \mathbf{b}\}.$$

That is, $\eta_i^{\gamma}(\mathbf{x}; V, r)$ is the smallest minimizer of the function $t \mapsto f(\mathbf{x}+t\mathbf{v}_i)+\frac{\gamma}{2}t^2$ over a closed interval. The optimality measure $\mathcal{O}^{\gamma}_{V_C,r}$ at $\mathbf{x}$ with a radius $r \in (0,\infty]$ is defined by $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) = \sum_{i=1}^{s}\eta_i^{\gamma}(\mathbf{x}; V, r)^2$.

To reduce clutter, we will omit the parameter γ from the notation η_i^{γ} and just write η_i, assuming that the identity of γ is clear from context, mainly from the notation $\mathcal{O}^{\gamma}_{V_C,r}$.

Lemma 4.1 verifies that $\mathcal{O}^{\gamma}_{V_C,r}$ indeed satisfies the basic properties of optimality measures, regardless of the choices of V_C, r and γ.

Lemma 4.1. *Let $r \in (0,\infty]$, $\gamma > 0$ and $V_C = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ be a complete feasible directions set of C. Then $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) \ge 0$ for any $\mathbf{x} \in C$ and if $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) = 0$, then $\mathbf{x}$ is a stationary point of (P).*

Proof. Obviously $\mathcal{O}^{\gamma}_{V_C,r}$ is nonnegative since it is defined as a sum of squares. Suppose that $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) = 0$. Since V_C is a complete feasible directions set of C, it follows that there exists a PSD set $V_{\mathbf{x}} \subseteq V_C$ that spans all feasible directions of C at $\mathbf{x}$. Since $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) = 0$, we have that

$$f(\mathbf{x}) \le f(\mathbf{x}+t\mathbf{v}) + \frac{\gamma}{2}t^2 \text{ for all } t \in [0,r], \mathbf{v} \in V_{\mathbf{x}}.$$

Rearranging, dividing by $t \in (0,r)$, and taking the limit $t \to 0^+$ then yields $f'(\mathbf{x}; \mathbf{v}) \ge 0$ for all $\mathbf{v} \in V_{\mathbf{x}}$. Consequently, by Theorem 2.1, $\mathbf{x}$ is a stationary point of (P). □

4.2. Relation to explicit optimality measures

In this section we assume that $r = \infty$ and that $g = 0$, meaning that the objective function $f = h$ is smooth. To give some more insight into the definition of the optimality measure, we will establish a direct relation between $\mathcal{O}^{\gamma}_{V_C,r}$ and an optimality measure that is expressed in terms of the directional derivatives of the objective function in the directions of V_C.

As usual, we assume that the given complete feasible directions set is $V_C = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$. We define $c_{\mathbf{x},i}$ as the maximal stepsize (possibly ∞) that can be taken when moving from $\mathbf{x} \in C$ towards the direction $\mathbf{v}_i \in V_C$:

$$c_{\mathbf{x},i} \equiv \max\{t : \mathbf{A}(\mathbf{x}+t\mathbf{v}_i) \le \mathbf{b}, t \ge 0\}.$$

Noting that $\mathbf{v}_i \in V_C$ is a feasible direction at $\mathbf{x}$ if and only if $c_{\mathbf{x},i} > 0$, we may interpretate $c_{\mathbf{x},i}$ as the "level" of feasibility in the direction of $\mathbf{v}_i$ at $\mathbf{x}$. The result assumes in addition that $f \in C_L^{1,1}$ [5, Chapter 5], meaning that $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \le L\|\mathbf{x}-\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Theorem 4.1. *Suppose that $g \equiv 0$ and $f \in C_L^{1,1}$ for some $L > 0$, and let $\mathbf{x} \in C$. Then*

$$\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) \ge \frac{1}{(L+\gamma)^2}\sum_{i\in[s]}\min\{[-\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle]_+, (L+\gamma)c_{\mathbf{x},i}\}^2. \tag{4.2}$$

Proof. Let $i \in [s]$ and $\mathbf{v}_i \in V_C$. Assume that $\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle < 0$. Define $\phi(t) \equiv f(\mathbf{x}+t\mathbf{v}_i) + \frac{\gamma}{2}t^2$. Then $\phi'(0) = \langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle < 0$. Also, for any $0 \le t < -\frac{\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle}{L+\gamma}$ it holds that

$$\begin{aligned}\phi'(t) &\le |\phi'(t) - \phi'(0)| + \phi'(0)\\ &\le |\langle\nabla f(\mathbf{x}+t\mathbf{v}_i), \mathbf{v}_i\rangle - \langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle| + \gamma t + \phi'(0)\\ &\le (L+\gamma)t + \phi'(0) < 0,\end{aligned} \tag{4.3}$$

where the third inequality follows from the facts that $f \in C_L^{1,1}$ and $\|\mathbf{v}_i\| = 1$ as well as the Cauchy-Schwarz inequality (recall that all vectors in PSD sets are unitary). The last inequality in (4.3) is a result of the assertion that $t < -\frac{\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle}{L+\gamma}$. Relation (4.3) implies that ϕ is a monotonic decreasing function over $\left[0, -\frac{\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle}{L+\gamma}\right)$, implying that any minimizer of ϕ over $[0, c_{\mathbf{x},i}]$ is bounded below by $\min\left\{-\frac{\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle}{L+\gamma}, c_{\mathbf{x},i}\right\}$. Recalling the definition of $\eta_i(\mathbf{x}, V_c, r)$ as the smallest minimizer of ϕ over the interval $[0, c_{\mathbf{x},i}]$, we conclude that

$$\begin{aligned}\eta_i(\mathbf{x}, V_c, r) &\ge \min\left\{-\frac{\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle}{L+\gamma}, c_{\mathbf{x},i}\right\}\\ &= \frac{1}{L+\gamma}\min\{-\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle, (L+\gamma)c_{\mathbf{x},i}\}.\end{aligned} \tag{4.4}$$

The above holds for indices $i \in [s]$ for which $\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle < 0$. Thus, for any $i \in [s]$,

$$\eta_i(\mathbf{x}, V_c, r) \ge \frac{1}{L+\gamma}\min\{[-\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle]_+, (L+\gamma)c_{\mathbf{x},i}\}. \tag{4.5}$$

Indeed, (4.5) reduces to (4.4) when $\langle\nabla f(\mathbf{x}), \mathbf{v}_i\rangle < 0$ and to the trivial inequality $\eta_i(\mathbf{x}, V_c, r) \ge 0$ otherwise. Summing over i the squares of both left-hand and right-hand sides of (4.5) yield the desired result (4.2). □

The quantity

$$\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x}) := \frac{1}{(L+\gamma)^2}\sum_{i\in[s]} \min\{[-\langle \nabla f(\mathbf{x}), \mathbf{v}_i\rangle]_+, (L+\gamma)c_{\mathbf{x},i}\}^2 \tag{4.6}$$

is always nonnegative, and it is equal to zero if and only if $\mathbf{x}\in C$ is a stationary point (indeed, $\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x}) = 0$ if and only if $\langle \nabla f(\mathbf{x}), \mathbf{v}_i\rangle \geq 0$ for all the feasible directions $\mathbf{v}_i$ in V_C at $\mathbf{x}$). Thus, $\mathcal{G}^{\gamma}_{V_C,r}$ is an optimality measure explicitly expressed in terms of the directional derivatives in the directions of the given complete feasible directions set V_C. Theorem 4.1 states that $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) \geq \mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x})$ for $\mathbf{x}\in C$.

Let us consider the two simple examples of $C=\mathbb{R}^n$ and $C=\mathbb{R}^n_+$.

Example 4.1. In case where $C=\mathbb{R}^n$, the stationarity condition is obviously $\nabla f(\mathbf{x}) = \mathbf{0}$, and thus a possible (and actually quite popular) optimality measure is $\|\nabla f(\mathbf{x})\|^2$. For this case we can take $V_C = \{\pm\mathbf{e}_1, \pm\mathbf{e}_2, \ldots, \pm\mathbf{e}_n\}$ and $\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x})$ is given by (note that $c_{\mathbf{x},i} = \infty$ for all $\mathbf{x}, i$)

$$\begin{aligned}\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x}) &= \frac{1}{(L+\gamma)^2}\sum_{i\in[s]}[-\nabla_i f(\mathbf{x})]_+^2 + \frac{1}{(L+\gamma)^2}\sum_{i\in[s]}[\nabla_i f(\mathbf{x})]_+^2\\ &= \frac{1}{(L+\gamma)^2}\|\nabla f(\mathbf{x})\|^2,\end{aligned}$$

meaning that it is a constant times the "standard" optimality condition $\|\nabla f(\mathbf{x})\|^2$. By Theorem 4.1, we conclude that $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) \geq \frac{1}{(L+\gamma)^2}\|\nabla f(\mathbf{x})\|^2$.

Example 4.2. In the case where $C=\mathbb{R}^n_+$, we can take the complete feasible directions set as $V_C = \{\pm\mathbf{e}_1, \pm\mathbf{e}_2, \ldots, \pm\mathbf{e}_n\}$, and we have that

$$\begin{aligned}\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x}) =& \frac{1}{(L+\gamma)^2}\sum_{i=1}^{n}([\min\{(L+\gamma)x_i, [\nabla_i f(\mathbf{x})]_+\}^2\\ &+ [-\nabla_i f(\mathbf{x})]_+^2).\end{aligned}$$

It is easy to see that $\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x}) = 0$ if and only if it holds that $\nabla_i f(\mathbf{x}) \begin{cases} = 0, & x_i > 0,\\ \geq 0, & x_i = 0,\end{cases}$ which is the well known stationarity conditions for smooth problems over the nonnegative orthant. By Theorem 4.1, $\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}) \geq \mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x})$.

5. Convergence analysis

In this section we establish the theoretical performance guarantees of the RFDS method (Algorithm 1) – the expected rate of convergence and the subsequence convergence to a stationary point. Before doing so, we derive some basic properties of the output sequence of the algorithm.

Lemma 5.1. *Let $\{\mathbf{x}^k\}_{k\geq 0}$ be the sequence generated by the RFDS method. Then*

(a) $f(\mathbf{y}^k_{i-1}) - f(\mathbf{y}^k_i) \geq \frac{\gamma}{2}q(i,k)^2 = \frac{\gamma}{2}\|\mathbf{y}^k_{i-1} - \mathbf{y}^k_i\|^2$ for all $i\in[s]$.

(b) $f(\mathbf{x}^k) - f(\mathbf{x}^{k+1}) \geq \frac{\gamma}{2}t_k^2 := \sum_{i\in[s]}\frac{\gamma}{2}q(i,k)^2 \geq \frac{\gamma}{2s}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2$ for all $k\geq 0$.

(c) The sequence $\{f(\mathbf{x}^k)\}_{k\geq 0}$ is non-ascending and convergent.

(d) $\sum_{k=0}^{K} t_k^2 \leq \frac{2(f(\mathbf{x}^0)-\bar{f})}{\gamma}$ for any integer $K>0$.

(e) $\lim\limits_{k\to\infty} t_k = \lim\limits_{k\to\infty}\|\mathbf{x}^{k+1} - \mathbf{x}^k\| = \lim\limits_{k\to\infty}\|\mathbf{y}^k_{i-1} - \mathbf{y}^k_i\| = \lim\limits_{k\to\infty} q(i,k) = 0$ for any $i\in[s]$.

Proof. Part (a) follows immediately from the update rule of $\mathbf{y}^k_i$, and (b) follows from summing the inequality in part (a) over $i\in[s]$ and applying Young's inequality. Part (c) is a direct consequence of part (b) and the underlying assumption that f is lower-bounded. Part (d) is derived by summing the first inequality in (b) and using the fact that the sequence $\{f(\mathbf{x}^k)\}_{k\geq 0}$ is bounded-below by $\bar{f}$. Part (e) readily follows by combining (a), (b) and (c). □

Theorem 5.1 establishes an $O(1/k)$ rate of convergence of the expected value of the optimality measure $\mathcal{O}^{\gamma}_{V_C,r}$ of the sequence generated by the RFDS method. We use the following notation: p_1^k is the random variable that corresponds to the direction $\mathbf{v}_{p_1^k}$ in V_C chosen to be first in the cycle of the k-th iteration and $\xi_k = (p_1^0, p_1^1, \ldots, p_1^k)$ is the multivariate random variable containing the history of these random variables up to cycle k.

Theorem 5.1 *(Expected rate). Let $\{\mathbf{x}^k\}_{k\geq 0}$ be a sequence generated by the RFDS method. Then for any integer $K>0$ it holds that*

$$\min_{k\in[K]}\mathbb{E}_{\xi_{k-1}}\left(\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}^k)\right) \leq \frac{2s(f(\mathbf{x}^0)-\bar{f})}{\gamma K}. \tag{5.1}$$

Proof. By Lemma 5.1(d), we have that

$$2\gamma^{-1}(f(\mathbf{x}^0)-\bar{f}) \geq \sum_{k=0}^{K} t_k^2 \geq \sum_{k=1}^{K} t_k^2 \geq \sum_{k=1}^{K} q(1,k)^2, \tag{5.2}$$

where the last inequality follows from the definition of t_k (Lemma 5.1(b)) as $t_k^2 := \sum_{i\in[s]} q(i,k)^2$. Taking expectation on relation (5.2) w.r.t. ξ_K, using the law of total expectation, and noting that $\mathbb{E}_{\xi_K}(q(1,k)^2) = \mathbb{E}_{\xi_k}(q(1,k)^2)$ yields

$$\begin{aligned}2\gamma^{-1}(f(\mathbf{x}^0)-\bar{f}) &\geq \sum_{k=1}^{K}\mathbb{E}_{\xi_k}(q(1,k)^2)\\ &= \sum_{k=1}^{K}\mathbb{E}_{\xi_{k-1}}\left(\mathbb{E}_{p_1^k}(q(1,k)^2)\right).\end{aligned} \tag{5.3}$$

Since by definition (cf. (4.1)) $\eta_{p_1^k}(\mathbf{x}^k; V, r) \leq q(1,k)$, we have that

$$\begin{aligned}\sum_{k=1}^{K}\mathbb{E}_{\xi_{k-1}}\left(\mathbb{E}_{p_1^k}(q(1,k)^2)\right) &\geq \sum_{k=1}^{K}\mathbb{E}_{\xi_{k-1}}\left(\mathbb{E}_{p_1^k}(\eta_{p_1^k}(\mathbf{x}^k; V, r)^2)\right)\\ &= \frac{1}{s}\sum_{k=1}^{K}\mathbb{E}_{\xi_{k-1}}\left(\sum_{i=1}^{s}\eta_i(\mathbf{x}^k; V, r)^2\right)\\ &\geq \frac{K}{s}\min_{k\in[K]}\mathbb{E}_{\xi_{k-1}}\left(\sum_{i=1}^{s}\eta_i(\mathbf{x}^k; V_C, r)^2\right)\\ &= \frac{K}{s}\min_{k\in[K]}\mathbb{E}_{\xi_{k-1}}\left(\mathcal{O}^{\gamma}_{V_C,r}(\mathbf{x}^k)\right).\end{aligned} \tag{5.4}$$

Combining (5.3) and (5.4), the desired result (5.1) follows. □

Theorem 5.1 together with Theorem 4.1 implies the following rate on the directional derivatives-based optimality measure gradient-feasibility level $\mathcal{G}^{\gamma}_{V_C,r}$ when $f\in C^{1,1}_L$.

Corollary 5.1. *Suppose that $g\equiv 0$, $f\in C^{1,1}$ and $r=\infty$. Let $\{\mathbf{x}^k\}_{k\geq 0}$ be the sequence generated by the RFDS method. Then for any integer $K>0$ it holds that ($\mathcal{G}^{\gamma}_{V_C,r}$ is given in (4.6))*

$$\min_{k\in[K]}\mathbb{E}_{\xi_{k-1}}\left(\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x}^k)\right) \leq \frac{2s(f(\mathbf{x}^0)-\bar{f})}{\gamma K}.$$

For example, when $C = \mathbb{R}^n$, in the setting of Example 4.1, we have that $\mathcal{G}^{\gamma}_{V_C,r}(\mathbf{x}) = \frac{1}{(L+\gamma)^2}\|\nabla f(\mathbf{x})\|^2$, and we thus obtain by Corollary 5.1 that (note that here $s = 2n$)

$$\min_{k\in[K]} \mathbb{E}_{\xi_{k-1}}\|\nabla f(\mathbf{x}^k)\|^2 \le \frac{4(L+\gamma)^2 n(f(\mathbf{x}^0)-\bar{f})}{\gamma K}.$$

The right-hand side in the above inequality depends on γ through the expression $\frac{(L+\gamma)^2}{\gamma}$ whose minimal value is L, and is attained at $\gamma = L$. Thus, if we employ the RFDS method with $\gamma = L$ we obtain the bound

$$\min_{k\in[K]} \mathbb{E}_{\xi_{k-1}}\|\nabla f(\mathbf{x}^k)\|^2 \le \frac{4Ln(f(\mathbf{x}^0)-\bar{f})}{K}.$$

We conclude the analysis with Theorem 5.2 below that shows a deterministic subsequence convergence guarantee for the sequence generated by the RFDS method. Note that this deterministic result is established even though the RFDS method incorporates randomness. The following technical lemma will be used in the proof, where we denote $B[\mathbf{c}, r] := \{\mathbf{x} : \|\mathbf{x}-\mathbf{c}\| \le r\}$.

Lemma 5.2. *Let $\mathbf{x}^* \in C$, $r > 0$, and $\mathbf{d} \neq \mathbf{0}$ be a normalized feasible direction of C at $\mathbf{x}^*$. Then there exists $\tilde{\varepsilon}, \tilde{r} > 0$ (possibly depending on $\mathbf{x}^*$) for which*

$$\mathbf{x} + t\mathbf{d} \in C \text{ for any } \mathbf{x} \in B[\mathbf{x}^*, \tilde{\varepsilon}] \cap C,\ t \in [0, \tilde{r}]. \tag{5.5}$$

Proof. Denote $I_=(\mathbf{x}^*) \equiv \{i : \mathbf{a}_i^T\mathbf{x}^* = b_i\}$ and $I_<(\mathbf{x}^*) \equiv \{i : \mathbf{a}_i^T\mathbf{x}^* < b_i\}$. If $I_<(\mathbf{x}^*) \neq \emptyset$, we define $\tilde{\varepsilon} = \tilde{r} \equiv \min\left\{\frac{|\mathbf{a}_i^T\mathbf{x}^* - b_i|}{2\|\mathbf{a}_i\|} : i \in I_<(\mathbf{x}^*)\right\}$; otherwise, we choose arbitrarily $\tilde{\varepsilon} = \tilde{r} = 1$. Take $\mathbf{x} \in B[\mathbf{x}^*, \tilde{\varepsilon}] \cap C$ and $t \in [0, \tilde{r}]$. For any $i \in I_<(\mathbf{x}^*)$ (relevant only when $I_<(\mathbf{x}^*) \neq \emptyset$),

$$\begin{aligned}\mathbf{a}_i^T(\mathbf{x}+t\mathbf{d}) - b_i &= \mathbf{a}_i^T\mathbf{x}^* - b_i + \mathbf{a}_i^T(\mathbf{x}-\mathbf{x}^*) + t\mathbf{a}_i^T\mathbf{d}\\ &\le \mathbf{a}_i^T\mathbf{x}^* - b_i + \tilde{\varepsilon}\|\mathbf{a}_i\| + \tilde{r}\|\mathbf{a}_i\|\\ &\le \mathbf{a}_i^T\mathbf{x}^* - b_i + \frac{|\mathbf{a}_i^T\mathbf{x}^* - b_i|}{2} + \frac{|\mathbf{a}_i^T\mathbf{x}^* - b_i|}{2}\\ &= 0.\end{aligned}$$

If $i \in I_=(\mathbf{x}^*)$, then since d is a feasible descent direction of C at $\mathbf{x}^*$, it follows that $\mathbf{a}_i^T\mathbf{d} \le 0$. Since $\mathbf{x} \in C$, it holds that $\mathbf{a}_i^T\mathbf{x} - b_i \le 0$, and consequently, $\mathbf{a}_i^T(\mathbf{x}+t\mathbf{d}) - b_i = \mathbf{a}_i^T\mathbf{x} - b_i + t\mathbf{a}_i^T\mathbf{d} \le 0$. We have thus established that $\mathbf{a}_i^T(\mathbf{x}+t\mathbf{d}) - b_i \le 0$ for all $i \in [m]$, and hence that $\mathbf{x} + t\mathbf{d} \in C$ as required. □

Theorem 5.2 *(Subsequence convergence). Any accumulation point of the sequence $\{\mathbf{x}^k\}_{k\ge 0}$ generated by the RFDS method is a stationary point of (P).*

Proof. Let $\mathbf{x}^*$ be an accumulation point of $\{\mathbf{x}^k\}_{k\ge 0}$. Then there exists a subsequence $\{\mathbf{x}^{k_j}\}_{j\ge 0}$ that converges to $\mathbf{x}^*$. We first show that $\mathbf{y}_i^{k_j} \to \mathbf{x}^*$ for any $i \in [s]$. By telescoping and invoking the triangle inequality we have

$$\begin{aligned}\|\mathbf{x}^* - \mathbf{y}_i^{k_j}\| &= \left\|\mathbf{x}^* - \mathbf{x}^{k_j} + \sum_{l=1}^{i}(\mathbf{y}_{l-1}^{k_j} - \mathbf{y}_l^{k_j})\right\|\\ &\le \left\|\mathbf{x}^* - \mathbf{x}^{k_j}\right\| + \sum_{l=1}^{i}\left\|\mathbf{y}_{l-1}^{k_j} - \mathbf{y}_l^{k_j}\right\|.\end{aligned}$$

Taking a limit and utilizing Lemma 5.1(e), we deduce that (note that $i \le s$ is bounded)

$$\begin{aligned}\lim_{j\to\infty}\|\mathbf{x}^* - \mathbf{y}_i^{k_j}\| &\le \lim_{j\to\infty}\left\|\mathbf{x}^* - \mathbf{x}^{k_j}\right\| + \sum_{l=1}^{i}\lim_{j\to\infty}\|\mathbf{y}_{l-1}^{k_j} - \mathbf{y}_l^{k_j}\|\\ &= 0.\end{aligned}$$

Now, let us consider the subset $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{s_*}\} = V_{\mathbf{x}^*} \subseteq V_C$, which is a PSD set of C at $\mathbf{x}^*$. By Lemma 5.2, there exist $\tilde{\varepsilon} > 0$ and $\tilde{r} \in (0, r]$ such that (5.5) holds. Since $\mathbf{y}_i^{k_j} \to \mathbf{x}^*$, there exists $J > 0$ such that for any $j \ge J$ it holds that

$$\|\mathbf{x}^* - \mathbf{y}_i^{k_j}\| < \tilde{\varepsilon}.$$

Denote by $n(k,i)$ the step in which the i-th direction in $V_{\mathbf{x}^*}$ is chosen during the computation of the k-th iteration. Then

$$\mathbf{y}_{n(k_j,i)}^{k_j} + t\mathbf{w}_i \in C \text{ for any } j \ge J, t \in [0, \tilde{r}], i \in [s_*].$$

The above together with the update procedure of the RFDS method implies that for any $t \in [0, \tilde{r}]$

$$\begin{aligned}f(\mathbf{y}_{n(k_j,i)}^{k_j} + q(n(k_j,i),k_j)\mathbf{w}_i) + \frac{\gamma}{2}q(n(k_j,i),k_j)^2&\\ \le f(\mathbf{y}_{n(k_j,i)}^{k_j} + t\mathbf{w}_i) + \frac{\gamma}{2}t^2.&\end{aligned} \tag{5.6}$$

We have that $\mathbf{y}_{n(k_j,i)}^{k_j} \to \mathbf{x}^*$ as $j \to \infty$, and by Lemma 5.1(e), $q(n(k_j,i),k_j) \to 0$. Thus, due the continuity of f (recall that it is the sum of a continuous function and a concave function), $f(\mathbf{y}_{n(k_j,i)}^{k_j} + q(n(k_j,i),k_j)\mathbf{w}_i) \to f(\mathbf{x}^*)$. Consequently, taking the limit $j \to \infty$ in (5.6), yields that

$$f(\mathbf{x}^*) \le f(\mathbf{x}^* + t\mathbf{w}_i) + \frac{\gamma}{2}t^2 \text{ for all } t \in [0, \tilde{r}],$$

which is the same as

$$\frac{f(\mathbf{x}^* + t\mathbf{w}_i) - f(\mathbf{x}^*)}{t} \ge -\frac{\gamma}{2}t \text{ for all } t \in (0, \tilde{r}].$$

Taking the limit $t \to 0^+$ in the above, we obtain that

$$f'(\mathbf{x}^*; \mathbf{w}_i) = \lim_{t\to 0^+}\frac{f(\mathbf{x}^* + t\mathbf{w}_i) - f(\mathbf{x}^*)}{t} \ge 0.$$

Since the above holds for any direction in the PSD set $V_{\mathbf{x}^*}$ of C at $\mathbf{x}^*$, Theorem 2.1 implies that $\mathbf{x}^*$ is a stationary point of (P). □

Data availability

No data was used for the research described in the article.

Acknowledgements

Nadav Hallak's research was supported by the Israel Science Foundation (grant No. 637/21). Amir Beck's research was supported by the Israel Science Foundation (grant No. 926/21).

References

[1] B. Assarf, E. Gawrilow, K. Herr, M. Joswig, B. Lorenz, A. Paffenholz, T. Rehn, Computing convex hulls and counting integer points with polymake, Math. Program. Comput. 9 (1) (may 2016) 1–38.

[2] A. Auslender, Computing points that satisfy second order necessary optimality conditions for unconstrained minimization, SIAM J. Optim. 20 (4) (2010) 1868–1884.

[3] D. Avis, A revised implementation of the reverse search vertex enumeration algorithm, in: Polytopes – Combinatorics and Computation, Birkhäuser, Basel, 2000, pp. 177–198.

[4] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear Programming: Theory and Algorithms, John Wiley & Sons, 2006.

[5] A. Beck, First-Order Methods in Optimization, vol. 25, SIAM, 2017.

[6] A. Beck, N. Hallak, On the convergence to stationary points of deterministic and randomized feasible descent directions methods, SIAM J. Optim. 30 (1) (2020) 56–79.

[7] A. Beck, D. Pan, Convergence of an inexact majorization-minimization method for solving a class of composite optimization problems, in: P. Gisellson, A. Rantzer (Eds.), Large Scale and Distributed Optimization, Springer, 2018.

[8] D.P. Bertsekas, Nonlinear Programming, second ed., Athena Scientific Optimization and Computation Series, Athena Scientific, Belmont, MA, 1999.

[9] A.R. Conn, K. Scheinberg, L.N. Vicente, Introduction to Derivative-Free Optimization, MPS/SIAM Series on Optimization, vol. 8, Society for Industrial and Applied Mathematics (SIAM)/Mathematical Programming Society (MPS), Philadelphia, PA/Philadelphia, PA, 2009.

[10] C. Davis, Theory of positive linear dependence, Am. J. Math. 76 (1954) 733–746.

[11] K. Fukuda, Lecture: Polyhedral Computation, ETH Zurich, 2014.

[12] N. Hallak, M. Teboulle, Finding second-order stationary points in constrained minimization: a feasible direction approach, J. Optim. Theory Appl. 186 (2) (2020) 480–503.

[13] L. Khachiyan, E. Boros, K. Borys, V. Gurvich, K. Elbassioni, Generating all vertices of a polyhedron is hard, Discrete Comput. Geom. 39 (2008) 174–190.

[14] T.G. Kolda, R.M. Lewis, V. Torczon, Optimization by direct search: new perspectives on some classical and modern methods, SIAM Rev. 45 (3) (2003) 385–482.

[15] Y. Nesterov, Introductory Lectures on Convex Optimization, Applied Optimization, vol. 87, Kluwer Academic Publishers, Boston, MA, 2004.

[16] M. Razaviyayn, M. Hong, Z.Q. Luo, A unified convergence analysis of block successive minimization methods for nonsmooth optimization, SIAM J. Optim. 23 (2) (2013) 1126–1153.

[17] R.G. Regis, On the properties of positive spanning sets and positive bases, Optim. Eng. 17 (1) (2016) 229–262.

[18] W.I. Zangwill, Nonlinear Programming: a Unified Approach, vol. 196, Prentice-Hall, Englewood Cliffs, NJ, 1969.

[19] G.M. Ziegler, Faces of polytopes, in: Graduate Texts in Mathematics, Springer, New York, 1995, pp. 51–76.