Journal of Convex Analysis
Volume 22 (2015), No. 4, 1061–1090

Bundle Method for Nonconvex Nonsmooth Constrained Optimization

Minh Ngoc Dao
Department of Mathematics and Informatics, Hanoi National University of Education, Hanoi, Vietnam; and: Institut de Mathématiques, Université de Toulouse, Toulouse, France
minhdn@hnue.edu.vn

Received: January 20, 2014
Revised manuscript received: January 25, 2015
Accepted: January 27, 2015

The paper develops a nonconvex bundle method based on the downshift mechanism and a proximity control management technique to solve nonconvex nonsmooth constrained optimization problems. We prove its global convergence in the sense of subsequences for both classes of lower-C^1 and upper-C^1 functions.

Keywords: Nonsmooth optimization, constrained optimization, bundle method, lower-C^1 function, upper-C^1 function

1. Introduction

Nonsmooth optimization problems appear frequently in practical applications such as economics, mechanics, and control theory. There are several methods for solving nonsmooth optimization problems, and they can be divided into two main groups: subgradient methods and bundle methods. We want to mention the latter ones because of their proven efficiency in solving relevant problems. Bundle methods were first introduced by Lemaréchal [12] and have been developed over the years based on subsequent works of Kiwiel [9], Lemaréchal, Nemirovskii, and Nesterov [13]. The main idea of bundle methods is to estimate the Clarke subdifferential [3] of the objective function by accumulating subgradients from past iterations into a bundle, and then to generate a trial step by a quadratic tangent program using information stored in the bundle.

Extending Lemaréchal's algorithm to the nonconvex case, Mifflin [16] gives a bundle algorithm using the so-called downshift mechanism for the nonsmooth minimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) \leqslant 0 \end{array}$$

where f and h are real-valued locally Lipschitz but potentially nonconvex functions on $\mathbb{R}^n$. Subsequently, Mäkelä and Neittaanmäki [14] present a proximal

ISSN 0944-6532 / \$ 2.50

bundle method for the above problem adding linear constraints. This method uses the improvement function

$$F(y, x) = \max\{f(y) - f(x), h(y)\}$$

for the handling of the constraints. While these works use a line search procedure which admits only weak convergence certificates in the sense that at least one of the accumulation points of the sequence of serious iterates is critical, we are interested in using a nonconvex bundle technique along with a suitable backtracking strategy. This brings to stronger convergence certificates, where *every* accumulation point of the sequence of serious iterates is critical. Recently, Gabarrou, Alazard and Noll [7] showed a strong convergence for the case where f and h are lower-C^1 functions in the sense of [23, 22]. However, a convergence proof for upper-C^1 functions still remains open.

In present framework we consider a more general constrained optimization problem of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) \leqslant 0 \\ & x \in \mathbf{C} \end{array} \tag{1}$$

where f and h are real-valued locally Lipschitz but potentially nonsmooth and nonconvex functions, and where $\mathbf{C}$ is a closed convex set of $\mathbb{R}^n$. For solving this problem, we propose a nonconvex bundle method based on downshifted tangents and a proximity control management mechanism, in which a strong convergence certificate is valid for both classes of lower-C^1 and upper-C^1 functions. Our work follows a classical line of bundling techniques [12, 9, 13, 19, 17], and develops the optimization method for constrained problems presented in [16, 14]. Instead of using the improvement function to deal with the presence of constraints, we address problem (1) through a progress function, which is motivated by an idea for smooth problems in [21] and was successfully extended to nonsmooth cases in [1, 7]. We slightly refine some elements of the bundle method, such as the exactness plane, the cutting plane and the management of the proximity control parameter in order to offer several advantages over previous methods.

The motivation of this paper relies on the fact that many application problems are addressed by minimizing lower-C^1 functions. For instance, some problems in the context of automatic control are quite successfully solved in [19, 17, 18, 7, 5] by applying optimization techniques of bundle type to lower-C^1 functions. In particular, the problem of maximizing the memory of a system [5] can be reformulated as minimizing upper-C^1 functions.

The rest of the paper is organized as follows. Sections 2–5 present elements of the proximity control algorithm. In Section 6 we introduce a theoretical tool in the convergence proof which is referred to as the upper envelope model. Some preparatory information on semismooth, lower-C^1 and upper-C^1 functions is given in Section 7. The central Sections 8, 9 prove global convergence of the algorithm.

2. Progress function

Following an idea of Polak in [21, Section 2.2.2], to solve problem (1) we use the *progress function*

$$F(y,x) = \max\{f(y) - f(x) - \mu h(x)_+, h(y) - h(x)_+\},$$

with $\mu > 0$ a fixed parameter and $h(x)_+ = \max\{h(x), 0\}$. Here x represents the current iterate, and y is the next iterate or a candidate for the next iterate.

Let $\partial f(x)$ denote the Clarke subdifferential of f at x. For functions of two variables, the notation ∂_1 stands for the Clarke subdifferential with respect to the first variable. We first remark that $F(x,x) = 0$. Moreover, $F(\cdot, x)$ is also locally Lipschitz, and by [4, Proposition 2.3.12] (see also [2, Proposition 9]),

$$\begin{cases} \partial_1 F(x,x) = \partial f(x) & \text{if } h(x) < 0, \\ \partial_1 F(x,x) \subset \operatorname{conv}\{\partial f(x) \cup \partial h(x)\} & \text{if } h(x) = 0, \\ \partial_1 F(x,x) = \partial h(x) & \text{if } h(x) > 0, \end{cases} \tag{2}$$

where `conv` signifies convex hull, and where equality holds if f and h are regular at x in the sense of Clarke [3]. Recall that the indicator function of a convex set $\mathbf{C} \subset \mathbb{R}^n$ defined by

$$i_{\mathbf{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathbf{C}, \\ +\infty & \text{otherwise}, \end{cases}$$

we have $i_{\mathbf{C}}(\cdot)$ is a convex function, and $\partial i_{\mathbf{C}}(x)$ is the normal cone to $\mathbf{C}$ at x,

$$N_{\mathbf{C}}(x) = \{g \in \mathbb{R}^n : g^\top (y - x) \leqslant 0 \text{ for all } y \in \mathbf{C}\},$$

if $x \in \mathbf{C}$, and the empty set otherwise. It is worth to notice that if $\mathbf{C}$ is a polyhedral set having the form

$$\mathbf{C} = \{x \in \mathbb{R}^n : a_i^\top x \leqslant b_i, i = 1, \ldots, m\},$$

where a_i and b_i are respectively given vectors and scalars, then

$$\partial i_{\mathbf{C}}(x) = N_{\mathbf{C}}(x) = \{\lambda_1 a_1 + \cdots + \lambda_m a_m : \lambda_i \geqslant 0, \lambda_i = 0 \text{ if } a_i^\top x < b_i\}$$

for all $x \in \mathbf{C}$ (see [22, Theorem 6.46]). Motivated by [1, Lemma 5.1] and [2, Theorem 1], we now establish the following result.

Lemma 2.1. *Let f and h be locally Lipschitz functions, then the following statements hold.*

(*i*) *If x^* is a local minimum of problem (1), it is also a local minimum of $F(\cdot, x^*)$ on $\mathbf{C}$, and then $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$. Furthermore, if x^* is an F. John critical point of (1) then $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$ in the case where f and h are regular at x^*.*

(*ii*) *Conversely, if* $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$ *for some* $x^* \in \mathbf{C}$ *then only one of the following situations occurs.*
  (*a*) $h(x^*) > 0$*, in which case* x^* *is a critical point of* h*, called a critical point of constraint violation.*
  (*b*) $h(x^*) \leqslant 0$*, in which case* x^* *is an F. John critical point of* (1)*. In addition, we have either* $h(x^*) = 0$ *and* $0 \in \partial h(x^*) + \partial i_{\mathbf{C}}(x^*)$*, or* x^* *is a Karush–Kuhn–Tucker point of* (1)*.*

Proof. (*i*) Let x^* be a local minimum of problem (1), then $h(x^*) \leqslant 0, x \in \mathbf{C}$, which gives $h(x^*)_+ = 0$, and so

$$F(y, x^*) = \max\{f(y) - f(x^*), h(y)\}.$$

Moreover, there exists a neighborhood U of x^* such that $f(y) \geqslant f(x^*)$ for all $y \in U \cap \mathbf{C}$ satisfying $h(y) \leqslant 0$. We will show that $F(y, x^*) \geqslant F(x^*, x^*)$ for all $y \in U \cap \mathbf{C}$. Indeed, if $h(y) > 0$ then

$$F(y, x^*) \geqslant h(y) > 0 = F(x^*, x^*).$$

If $h(y) \leqslant 0$ then $f(y) \geqslant f(x^*)$, and therefore

$$F(y, x^*) \geqslant f(y) - f(x^*) \geqslant 0 = F(x^*, x^*).$$

This means that x^* is a local minimum of $F(\cdot, x^*)$ on $\mathbf{C}$, which implies $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$.

Now assume that x^* is an F. John critical point of (1), i.e., there exist constants λ_0, λ_1 such that

$$\begin{gathered} 0 \in \lambda_0 \partial f(x^*) + \lambda_1 \partial h(x^*) + \partial i_{\mathbf{C}}(x^*), \\ \lambda_0 \geqslant 0,\ \lambda_1 \geqslant 0,\ \lambda_0 + \lambda_1 = 1, \\ \lambda_1 h(x^*) = 0. \end{gathered}$$

Then if $h(x^*) < 0$, we have $\lambda_1 = 0$, $\lambda_0 = 1$, and by using (2), $\partial_1 F(x^*, x^*) = \partial f(x^*)$, which implies $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$. In the case where f and h are regular at x^*, if $h(x^*) = 0$ then $\partial_1 F(x^*, x^*) = \operatorname{conv}\{\partial f(x^*) \cup \partial h(x^*)\}$, and thus

$$0 \in \lambda_0 \partial f(x^*) + \lambda_1 \partial h(x^*) + \partial i_{\mathbf{C}}(x^*) \subset \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*).$$

(*ii*) Suppose that $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$ for some $x^* \in \mathbf{C}$. Then by (2), there exist constants λ_0, λ_1 such that

$$\begin{gathered} 0 \in \lambda_0 \partial f(x^*) + \lambda_1 \partial h(x^*) + \partial i_{\mathbf{C}}(x^*), \\ \lambda_0 \geqslant 0,\ \lambda_1 \geqslant 0,\ \lambda_0 + \lambda_1 = 1. \end{gathered}$$

If $h(x^*) > 0$ then $\partial_1 F(x^*, x^*) = \partial h(x^*)$, and so $0 \in \partial h(x^*) + \partial i_{\mathbf{C}}(x^*)$, that is, x^* is a critical point of h.

If $h(x^*) < 0$ then $\partial_1 F(x^*, x^*) = \partial f(x^*)$, which gives $\lambda_1 = 0$, and therefore x^* is a Karush–Kuhn–Tucker point and also an F. John critical point of (1).

In the case of $h(x^*) = 0$, we see immediately that x^* is an F. John critical point of (1). If x^* fails to be a Karush–Kuhn–Tucker point then $\lambda_0 = 0$ and we get $0 \in \partial h(x^*) + \partial i_{\mathbf{C}}(x^*)$. The lemma is proved completely. $\square$

3. Tangent program and acceptance test

In accordance with Lemma 2.1, it is reasonable to seek for points x^* satisfying $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$. We present our nonconvex bundle method for finding solutions of problem (1), which generates a sequence x^j of estimates converging to a solution x^* in the sense of subsequences.

Denote the current iterate of the outer loop by x, or x^j if the outer loop counter j is used. When a new iterate of the outer loop is found, it will be denoted by x^+, or x^{j+1}. At the current iterate x, we build first-order *working models* $\phi_k(\cdot, x)$, which approximate $F(\cdot, x)$ in a neighborhood of x. These working models are updated iteratively during the inner loop, and have to satisfy the following properties for every k:

- $\phi_k(\cdot, x)$ is convex;
- $\phi_k(x, x) = F(x, x) = 0$ and $\partial_1 \phi_k(x, x) \subset \partial_1 F(x, x)$.

The latter is valid when the so-called *exactness plane* $m_0(\cdot, x) = g(x)^\top(\cdot - x)$ with $g(x) \in \partial_1 F(x, x)$ is an affine minorant of $\phi_k(\cdot, x)$. Note that due to (2) we can choose $g(x) \in \partial f(x)$ if $h(x) \leqslant 0$, and $g(x) \in \partial h(x)$ if $h(x) > 0$.

Once the first-order working model $\phi_k(\cdot, x)$ has been decided on, we define an associated second-order working model

$$\Phi_k(\cdot, x) = \phi_k(\cdot, x) + \frac{1}{2}(\cdot - x)^\top Q(x)(\cdot - x),$$

where $Q(x)$ is a symmetric matrix depending only on the current iterate x. Now we find a new *trial step* y^k via the *tangent program*

$$\begin{array}{ll} \text{minimize} & \Phi_k(y, x) + \frac{\tau_k}{2}\|y - x\|^2 \\ \text{subject to} & y \in \mathbf{C} \end{array} \tag{3}$$

where $\tau_k > 0$ is called the *proximity control parameter*. Note that this program is strictly convex and has a unique solution as soon as we assure $Q(x) + \tau_k I \succ 0$.

In the sequel, we write $\partial_1(\phi(y, x) + i_{\mathbf{C}}(y))$ for the Clarke subdifferential of $\phi(y, x) + i_{\mathbf{C}}(y)$ with respect to the first variable at y. Let us note that $\partial_1(\phi(y, x) + i_{\mathbf{C}}(y)) \subset \partial_1 \phi(y, x) + \partial i_{\mathbf{C}}(y)$, and that equality need not hold. The necessary optimality condition for tangent program (3) gives

$$0 \in \partial_1(\phi_k(y^k, x) + i_{\mathbf{C}}(y^k)) + (Q(x) + \tau_k I)(y^k - x).$$

Therefore, if $y^k = x$ then $0 \in \partial_1 \phi_k(x,x) + \partial i_{\mathbf{C}}(x)$, and so $0 \in \partial_1 F(x,x) + \partial i_{\mathbf{C}}(x)$ due to the fact that $\partial_1 \phi_k(x,x) \subset \partial_1 F(x,x)$. The consequence of this argument is that once $0 \notin \partial_1 F(x,x) + \partial i_{\mathbf{C}}(x)$, the trial step y^k will always bring something new. From this time forth we suppose that $0 \notin \partial_1 F(x,x) + \partial i_{\mathbf{C}}(x)$. Then $y^k \neq x$, and since y^k is the optimal solution of program (3), we have $\Phi_k(y^k, x) + \frac{\tau_k}{2}\|y^k - x\|^2 \leqslant \Phi_k(x,x)$, which gives $\Phi_k(y^k, x) < \Phi_k(x,x) = 0$. In other words, there is always a progress predicted by the working model $\Phi_k(\cdot, x)$, unless x is already a critical point of (1) in the sense that $0 \in \partial_1 F(x,x) + \partial i_{\mathbf{C}}(x)$.

Following standard terminology, y^k is called a *serious step* if it is accepted as the new iterate, and a *null step* otherwise. In order to decide whether y^k is accepted or not, we compute the *test quotient*

$$\rho_k = \frac{F(y^k, x)}{\Phi_k(y^k, x)},$$

which measures the agreement between $F(\cdot, x)$ and $\Phi_k(\cdot, x)$ at y^k. If the current model Φ_k represents F precisely at y^k, it is awaited that $\rho_k \approx 1$. Fixing a constant $0 < \gamma < 1$, we accept the trial step y^k already as the new serious step x^+ if $\rho_k \geqslant \gamma$. Here the inner loop ends. Otherwise y^k is rejected and the inner loop continues.

Remark 3.1. The algorithm assures that $x^+ \in \mathbf{C}$, $F(x^+, x) < 0$ and $h(x^+) < h(x)_+$. Additionally, if the current iterate x is feasible, then the serious step x^+ is strictly feasible and $f(x^+) < f(x)$, otherwise, $h(x^+) < h(x)$. Indeed, assume that the serious step x^+ is accepted at inner loop counter k, which means $x^+ = y^k \in \mathbf{C}$ with $\rho_k \geqslant \gamma > 0$. This combined with $\Phi_k(y^k, x) < 0$ gives $F(x^+, x) < 0$, and so $h(x^+) < h(x)_+$, using the fact that $h(x^+) - h(x)_+ \leqslant F(x^+, x)$. If x is feasible, then $F(x^+, x) = \max\{f(x^+) - f(x), h(x^+)\} < 0$, which implies that $f(x^+) < f(x)$ and $h(x^+) < 0$. Otherwise, $h(x^+) < h(x)_+ = h(x)$, and in this case, a slight increase of f might occur but not exceeding $\mu h(x)$. This helps our approach to avoid being trapped at an infeasible local minima of f alone, which is a possible advantage.

4. Working model update

Suppose that y^k is a null step, we will improve the next model $\phi_{k+1}(\cdot, x)$. Notice that the exactness plane is always kept in first-order working models. To make $\phi_{k+1}(\cdot, x)$ better than $\phi_k(\cdot, x)$, we need two more elements, referred to as cutting and aggregate planes. Let us first look at the cutting plane generation.

The cutting plane $m_k(\cdot, x)$ is a basic element in bundle methods which cuts away the unsuccessful trial step y^k. The idea is to construct $m_k(\cdot, x)$ in the way that y^k is no longer solution of the new tangent program as soon as $m_k(\cdot, x)$ is an affine minorant of $\phi_{k+1}(\cdot, x)$. For each subgradient $g_k \in \partial_1 F(y^k, x)$, the tangent $t_k(\cdot) = F(y^k, x) + g_k^\top(\cdot - y^k)$ to $F(\cdot, x)$ at y^k is used as a cutting plane in the case where $F(\cdot, x)$ is convex. Without convexity, tangent planes may be useless, and a substitute has to be found. We exploit a mechanism first described in [16], which consists in shifting the tangent down until it becomes useful for $\phi_{k+1}(\cdot, x)$. Fixing a parameter $c > 0$ once and for all, we define the *downshift* as $s_k = [t_k(x) + c\|y^k - x\|^2]_+$, and introduce the *cutting plane*

$$m_k(\cdot, x) = t_k(\cdot) - s_k = a_k + g_k^\top(\cdot - x),$$

with $a_k = m_k(x, x) = t_k(x) - s_k = \min\{t_k(x), -c\|y^k - x\|^2\}$. Note that since $y^k \neq x$, $a_k \leqslant -c\|y^k - x\|^2 < 0$.

Remark 4.1. Let $\phi_{k+1}(\cdot, x) = \max\{m_i(\cdot, x) : i = 0, \dots, k\}$, then $\phi_{k+1}(\cdot, x)$ is convex, and $\phi_{k+1}(x, x) = F(x, x) = 0$, $\partial_1\phi_{k+1}(x, x) \subset \partial_1 F(x, x)$. Indeed, since $\phi_{k+1}(\cdot, x)$ is a maximum of affine planes, and $m_i(x, x) = a_i < 0 = m_0(x, x)$ for $i \geqslant 1$, we get convexity of $\phi_{k+1}(\cdot, x)$, and also $\phi_{k+1}(x, x) = 0$, $\partial_1\phi_{k+1}(x, x) = \partial_1 m_0(x, x) = \{g(x)\} \subset \partial_1 F(x, x)$.

Next we see that the optimality condition for (3) can be written as

$$(Q(x) + \tau_k I)(x - y^k) = g_k^* + h_k^*, \quad \text{for } g_k^* \in \partial_1\phi_k(y^k, x), h_k^* \in \partial i_{\mathbf{C}}(y^k). \tag{4}$$

If $\phi_k(\cdot, x) = \max\{m_i(\cdot, x) : i = 0, \dots, r\}$ then there exist $\lambda_0, \dots, \lambda_r$ being non-negative and summing up to 1 such that

$$g_k^* = \sum_{i=0}^{r} \lambda_i g_i, \qquad \phi_k(y^k, x) = \sum_{i=0}^{r} \lambda_i m_i(y^k, x).$$

We call g_k^* the *aggregate subgradient* as traditional, and build the *aggregate plane*

$$m_k^*(\cdot, x) = a_k^* + g_k^{*\top}(\cdot - x)$$

with $a_k^* = \sum_{i=0}^{r} \lambda_i a_i = \phi_k(y^k, x) + g_k^{*\top}(x - y^k)$. Then $\phi_k(y^k, x) = m_k^*(y^k, x) \leqslant \phi_{k+1}(y^k, x)$ if we require that $m_k^*(\cdot, x)$ is an affine minorant of $\phi_{k+1}(\cdot, x)$ at y^k. To avoid overflow, when generating the new working model $\phi_{k+1}(\cdot, x)$, we may replace all older planes corresponding to $\lambda_i > 0$ by the aggregate plane. This construction follows the original lines as proposed in [9]. It does not change neither the conclusion of Remark 4.1, nor the definition of aggregate planes.

Remark 4.2. Typically, the new working model $\phi_{k+1}(\cdot, x)$ can be given by

$$\phi_{k+1}(\cdot, x) = \max\{m_0(\cdot, x), m_k(\cdot, x), m_k^*(\cdot, x)\},$$

which satisfies the required properties of a first-order working model.

As we pass from x to a new serious step x^+, the planes $m(\cdot, x) = a + g^\top(\cdot - x)$ from previous serious steps may become useless at x^+ since we have no guarantee that $m(x^+, x) \leqslant F(x^+, x^+) = 0$. But we can *recycle* the old planes by using again the downshift mechanism as

$$m(\cdot, x^+) = m(\cdot, x) - s^+, \qquad s^+ = [m(x^+, x) + c\|x^+ - x\|^2]_+.$$

For more details, we refer to [17].

5. Proximity control management

The management of the proximity control parameter τ_k is a major difference between the convex and nonconvex bundle methods. In the convex case, the proximity control can remain unchanged during the inner loop. In the absence of convexity, the parameter τ_k has to follow certain basic rules to assure convergence of the algorithm. The central rule which we have to respect is that during the inner loop, the parameter may only increase infinitely often due to the strong discrepancy between the current working model ϕ_k and the best possible working model. Assuming the trial step y^k is a null step, as a means to decide when to increase τ_k or not, we compute the *secondary control parameter*

$$\tilde{\rho}_k = \frac{M_k(y^k, x)}{\Phi_k(y^k, x)},$$

where $M_k(\cdot, x) = \max\{m_0(\cdot, x), m_k(\cdot, x)\} + \frac{1}{2}(\cdot - x)^\top Q(x)(\cdot - x)$ with $m_0(\cdot, x)$ the exactness plane at the current iterate x, and $m_k(\cdot, x)$ the cutting plane at x and y^k. If $\tilde{\rho}_k \approx 1$ which indicates that little to no progress is achieved by adding the cutting plane, the proximity parameter must be increased to force smaller steps. In the case where $\tilde{\rho}_k$ is too far from 1, we hope that the situation will be improved without having to increase the proximity parameter. Fixing parameters $\tilde{\gamma}$ and θ with $0 < \gamma < \tilde{\gamma} < 1 < \theta < +\infty$, we make the following decision

$$\tau_{k+1} = \begin{cases} \tau_k & \text{if } \tilde{\rho}_k < \tilde{\gamma}, \\ \theta\tau_k & \text{if } \tilde{\rho}_k \geqslant \tilde{\gamma}. \end{cases}$$

Let us next consider the management of the proximity parameter between serious steps $x \to x^+$, respectively, $x^j \to x^{j+1}$. To do this we use a *memory element* $\tau_j^\sharp$, which is computed as soon as a serious step is made. Suppose that the serious step x^{j+1} is achieved at inner loop counter k_j, that is $x^{j+1} = y^{k_j}$ with $\rho_{k_j} \geqslant \gamma$. We consider the test

$$\rho_{k_j} = \frac{F(y^{k_j}, x^j)}{\Phi_{k_j}(y^{k_j}, x^j)} \overset{?}{\geqslant} \Gamma,$$

where $0 < \gamma < \Gamma < 1$ is fixed throughout. If $\rho_{k_j} < \Gamma$ then we memorize the last parameter used, that means $\tau_{j+1}^\sharp = \tau_{k_j}$. On the other hand, if $\rho_{k_j} \geqslant \Gamma$ then we may trust the model and store $\tau_{j+1}^\sharp = \theta^{-1}\tau_{k_j} < \tau_{k_j}$. At the first inner loop of the jth outer loop, the memory element $\tau_j^\sharp$ serves to initialize $\tau_1 = \max\{\tau_j^\sharp, -\lambda_{\min}(Q_j) + \kappa\}$ or $\tau_1 = T > q + \kappa$ with $\lambda_{\min}(\cdot)$ the minimum eigenvalue of a symmetric matrix, and $0 < \kappa \ll 1$ fixed, which assures always that $Q_j + \tau_k I \succ 0$ during the jth outer loop.

Figure 5.1 shows a flowchart of the algorithm, while the detailed statement is presented as Algorithm 5.1.

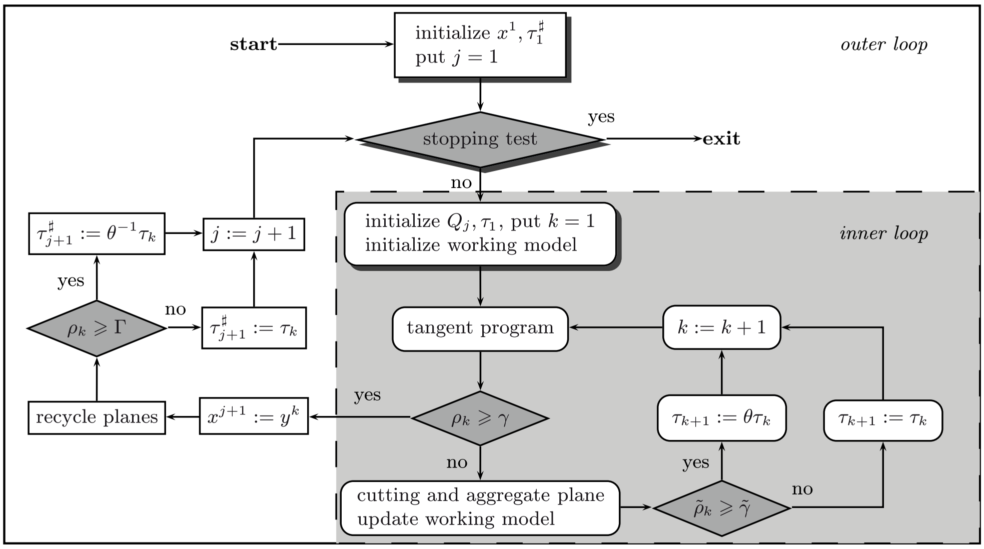

Figure 5.1: Flowchart of proximity control algorithm. Inner loop is shown in the lower right box

Algorithm 5.1. Proximity control algorithm with downshifted tangents.

Parameters: $0 < \gamma < \tilde\gamma < 1$, $0 < \gamma < \Gamma < 1$, $1 < \theta < +\infty$, $0 < \kappa \ll 1$, $0 < q < +\infty$, $q + \kappa < T < +\infty$.

▷ **Step 1** (`Outer loop initialization`)**.** Choose initial guess $x^1 \in \mathbf{C}$, fix memory control parameter $\tau_1^\sharp$, and put outer loop counter $j = 1$.

◇ **Step 2** (`Stopping test`)**.** At outer loop counter j, stop the algorithm if $0 \in \partial_1 F(x^j, x^j) + \partial i_{\mathbf{C}}(x^j)$. Otherwise, take a symmetric matrix Q_j respecting $-qI \preceq Q_j \preceq qI$, and goto inner loop.

▷ **Step 3** (`Inner loop initialization`)**.** Put inner loop counter $k = 1$ and initialize control parameter $\tau_1 = \max\{\tau_j^\sharp, -\lambda_{\min}(Q_j) + \kappa\}$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a symmetric matrix. Reset $\tau_1 = T$ if $\tau_1 > T$, and choose initial working model $\phi_1(\cdot, x^j)$ using the exactness plane $m_0(\cdot, x^j)$ and possibly recycling some planes from previous loop.

▷ **Step 4** (`Tangent program`)**.** At inner loop counter k, let

$$\Phi_k(\cdot, x^j) = \phi_k(\cdot, x^j) + \frac{1}{2}(\cdot - x^j)^\top Q_j(\cdot - x^j)$$

and find solution y^k (`trial step`) of the tangent program

$$\begin{array}{ll} \text{minimize} & \Phi_k(y, x^j) + \frac{\tau_k}{2}\|y - x^j\|^2 \\ \text{subject to} & y \in \mathbf{C}. \end{array}$$

◇ **Step 5** (`Acceptance test`)**.** Compute the quotient

$$\rho_k = \frac{F(y^k, x^j)}{\Phi_k(y^k, x^j)}.$$

If $\rho_k \geqslant \gamma$ (`serious step`), put $x^{j+1} = y^k$, compute new memory element

$$\tau^\sharp_{j+1} = \begin{cases} \tau_k & \text{if } \rho_k < \Gamma, \\ \theta^{-1}\tau_k & \text{if } \rho_k \geqslant \Gamma, \end{cases}$$

and quit inner loop. Increase outer loop counter j and loop back to step 2. If $\rho_k < \gamma$ (`null step`), continue inner loop with step 6.

$\triangleright$ **Step 6** (`Working model update`). Generate a cutting plane $m_k(\cdot, x^j)$ at null step y^k and counter k using downshifted tangents. Compute aggregate plane $m^*_k(\cdot, x^j)$ at y^k, and then build a new working model $\phi_{k+1}(\cdot, x^j)$ by adding the new cutting plane, keeping the exactness plane and using aggregation to avoid overflow.

$\triangleright$ **Step 7** (`Proximity control management`). Compute secondary control parameter

$$\tilde{\rho}_k = \frac{M_k(y^k, x^j)}{\Phi_k(y^k, x^j)},$$

with $M_k(\cdot, x^j) = \max\{m_0(\cdot, x^j), m_k(\cdot, x^j)\} + \frac{1}{2}(\cdot - x^j)^\top Q_j(\cdot - x^j)$, and then put

$$\tau_{k+1} = \begin{cases} \tau_k & \text{if } \tilde{\rho}_k < \tilde{\gamma}, \\ \theta\tau_k & \text{if } \tilde{\rho}_k \geqslant \tilde{\gamma}. \end{cases}$$

Increase inner loop counter k and loop back to step 4.

6. Upper envelope model

To analyse the convergence of the algorithm, we adapt a notion from [17, 18] for the progress function F. At the current iterate x of the outer loop, the *upper envelope model* is defined as

$$\phi^\uparrow(y, x) = \sup\{m_{y^+,g}(y, x) : y^+ \in B(x, M), g \in \partial_1 F(y^+, x)\},$$

where $B(x, M)$ is a fixed ball large enough to contain all possible trial steps during the inner loop, and where $m_{y^+,g}(\cdot, x)$ is the cutting plane at serious iterate x and trial step y^+ with subgradient $g \in \partial_1 F(y^+, x)$. We see immediately that $\phi^\uparrow(\cdot, x)$ is well defined due to boundedness of $B(x, M)$ and boundedness of all possible trial steps during the inner loop which will be proved without using the notion $\phi^\uparrow$ in Lemma 8.1(i) and Lemma 8.2(i). Furthermore, we have the following result.

Lemma 6.1. *Let f and h be locally Lipschitz functions, then the following statements hold.*

(i) $\phi^\uparrow(\cdot, x)$ *is a convex function and* $\phi_k(\cdot, x) \leqslant \phi^\uparrow(\cdot, x)$ *for all counters* k.
(ii) $\phi^\uparrow(x, x) = 0$ *and* $\partial_1\phi^\uparrow(x, x) = \partial_1 F(x, x)$.
(iii) $\phi^\uparrow$ *is jointly upper semi-continuous.*

Proof. (i) The first statement is followed from the definition of $\phi^{\uparrow}(\cdot,x)$ and the construction of $\phi_k(\cdot,x)$.

(ii) By construction, $m_{y^+,g}(x,x) \leqslant 0$ and $m_{x,g}(x,x) = 0$, which implies $\phi^{\uparrow}(x,x) = 0$.

We now take an arbitrary $\bar{g} \in \partial_1\phi^{\uparrow}(x,x)$ and the tangent plane $\bar{m}(\cdot,x) = \bar{g}^\top(\cdot - x)$ to the graph of $\phi^{\uparrow}(\cdot,x)$ at x associated with $\bar{g}$. Since $\phi^{\uparrow}(\cdot,x)$ is a convex function, $\bar{m}(\cdot,x) \leqslant \phi^{\uparrow}(\cdot,x)$. Fixing a vector $v \in \mathbb{R}^n$, for each $t > 0$, by the definition of $\phi^{\uparrow}(\cdot,x)$, there exists a cutting plane at trial step y_t with subgradient $g_t \in \partial_1 F(y_t,x)$ such that $\phi^{\uparrow}(x+tv,x) \leqslant m_{y_t,g_t}(x+tv,x)+t^2$. Note that $m_{y_t,g_t}(\cdot,x)$ can be represented as $m_{y_t,g_t}(\cdot,x) = m_{y_t,g_t}(x,x) + g_t^\top(\cdot - x)$ and $m_{y_t,g_t}(x,x) \leqslant -c\|y_t - x\|^2$. This gives

$$\begin{aligned} t\bar{g}^\top v = \bar{m}(x+tv,x) &\leqslant \phi^{\uparrow}(x+tv,x) \leqslant m_{y_t,g_t}(x+tv,x) + t^2 \\ &\leqslant -c\|y_t - x\|^2 + tg_t^\top v + t^2. \end{aligned}$$

Let $t \to 0^+$, we get $y_t \to x$. By passing to a subsequence and using the upper semi-continuity of the Clarke subdifferential, we may assume that $g_t \to g$ for some $g \in \partial_1 F(x,x)$. In addition, the above estimate also gives $\bar{g}^\top v \leqslant g_t^\top v + t$ for all $t > 0$, which infers $\bar{g}^\top v \leqslant g^\top v$, and so

$$\bar{g}^\top v \leqslant \max\{g^\top v : g \in \partial_1 F(x,x)\}.$$

The expression on the right is the Clarke directional derivative of $F(\cdot,x)$ at x in direction v. Since this relation holds true for every $v \in \mathbb{R}^n$, $\bar{g} \in \partial_1 F(x,x)$. Hence, $\partial_1\phi^{\uparrow}(x,x) \subset \partial_1 F(x,x)$.

It only remain to show $\partial_1 F(x,x) \subset \partial_1\phi^{\uparrow}(x,x)$. In order to do this, we consider the limit set

$$\underset{\rightarrow}{\partial_1} F(x,x) = \left\{ \lim_{k\to+\infty} \nabla_1 F(y^k,x) : y^k \to x,\ F(\cdot,x) \text{ is differentiable at } y^k \right\}.$$

Here $\nabla_1 F(y^k,x)$ denotes the subgradient of $F(\cdot,x)$ at y^k in the case where $F(\cdot,x)$ is differentiable at y^k. We use the symbol $\underset{\rightarrow}{\partial}$ for the limit set, following Hiriart-Urruty [8]. By [2, Proposition 5] (see also [4, Theorem 2.5.1]), $\partial_1 F(x,x) = \mathrm{conv}(\underset{\rightarrow}{\partial_1} F(x,x))$. We will prove that $\underset{\rightarrow}{\partial_1} F(x,x) \subset \partial_1\phi^{\uparrow}(x,x)$. Indeed, take $g \in \underset{\rightarrow}{\partial_1} F(x,x)$, there exist $y^k \to x$ and $g_k = \nabla_1 F(y^k,x) \in \partial_1 F(y^k,x)$ such that $g_k \to g$. Let $m_k(\cdot,x)$ be the cutting plane drawn at y^k with subgradient g_k, then $m_k(y,x) \leqslant \phi^{\uparrow}(y,x)$ for all $y \in \mathbb{R}^n$ and

$$m_k(\cdot,x) = a_k + g_k^\top(\cdot - x), \quad a_k = \min\{t_k(x), -c\|y^k - x\|^2\},$$

where $t_k(x) = F(y^k,x) + g_k^\top(x - y^k)$. From $y^k \to x$, $g_k \to g$ and $F(x,x) = 0$, it follows that $a_k \to 0$, and so $m_k(y,x) \to g^\top(y - x)$, which implies $g^\top(y-x) \leqslant \phi^{\uparrow}(y,x)$ for all y. This together with $\phi^{\uparrow}(x,x) = 0$ gives $g \in \partial_1\phi^{\uparrow}(x,x)$. We obtain

$\underset{\rightarrow}{\partial_1} F(x,x) \subset \partial_1 \phi^\uparrow(x,x)$ and then $\partial_1 F(x,x) = \operatorname{conv}(\underset{\rightarrow}{\partial_1} F(x,x)) \subset \partial_1 \phi^\uparrow(x,x)$ due to the convexity of $\partial_1 \phi^\uparrow(x,x)$.

(iii) Let $(y^j, x^j) \to (y,x)$, we have to prove that $\limsup \phi^\uparrow(y^j, x^j) \leqslant \phi^\uparrow(y,x)$. Pick a sequence $\varepsilon_j \to 0^+$, by the definition of $\phi^\uparrow$, there exist cutting planes $m_{z^j,g_j}(\cdot, x^j) = t_{z^j}(\cdot) - s_j$ at serious iterate x^j, drawn at $z^j \in B(x^j, M)$ with $g_j \in \partial_1 F(z^j, x^j)$ such that

$$\phi^\uparrow(y^j, x^j) \leqslant m_{z^j,g_j}(y^j, x^j) + \varepsilon_j,$$

where $t_{z^j}(\cdot) = F(z^j, x^j) + g_j^\top(\cdot - z^j)$ and $s_j = [t_{z^j}(x^j) + c\|z^j - x^j\|^2]_+$. Since $x^j \to x$ and $z^j \in B(x^j, M)$, the sequence z^j is bounded. Passing to a subsequence, we may assume without loss that $z^j \to z \in B(x,M)$ and $g_j \to g \in \partial_1 F(z,x)$ by the upper semi-continuity of the Clarke subdifferential. This gives $t_{z^j}(\cdot) \to t_z(\cdot) = F(z,x) + g^\top(\cdot - z)$, and so $s_j \to s = [t_z(x) + c\|z - x\|^2]_+$. It follows that

$$m_{z^j,g_j}(\cdot, x^j) = t_{z^j}(\cdot) - s_j \to t_z(\cdot) - s = m_{z,g}(\cdot, x)$$

as $i \to +\infty$, and then also

$$m_{z^j,g_j}(y^j, x^j) = t_{z^j}(y^j) - s_j \to t_z(y) - s = m_{z,g}(y,x),$$

where uniformity comes from boundedness of the g_j. Therefore,

$$\limsup_{j\to+\infty} \phi^\uparrow(y^j, x^j) \leqslant m_{z,g}(y,x) \leqslant \phi^\uparrow(y,x). \qquad \square$$

7. Lower-C^1 and upper-C^1 functions

According to Mifflin [15], a function $f : \mathbb{R}^n \to \mathbb{R}$ is *semismooth* at $x \in \mathbb{R}^n$ if f is Lipschitz on a ball near x, and for $d \in \mathbb{R}^n$, $\{t_k\} \subset \mathbb{R}_+$, $\{\theta_k\} \subset \mathbb{R}^n$, $\{g_k\} \subset \mathbb{R}^n$ satisfying $t_k \downarrow 0$, $\theta_k/t_k \to 0 \in \mathbb{R}^n$, $g_k \in \partial f(x + t_k d + \theta_k)$, the sequence $g_k^\top d$ has exactly one accumulation point. The following lemma can be seen as a generalization of [15, Lemma 2].

Lemma 7.1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ Lipschitz near x is semismooth at x if and only if for any $\{d_k\} \subset \mathbb{R}^n$, $\{t_k\} \subset \mathbb{R}_+$, $\{g_k\} \subset \mathbb{R}^n$ satisfying $d_k \to d \in \mathbb{R}^n$, $t_k \downarrow 0$, $g_k \in \partial f(x + t_k d_k)$, we have*

$$\lim_{k\to+\infty} g_k^\top d_k = f'(x; d).$$

Proof. Assume that f is semismooth at x. Taking $s_k \downarrow 0$, by Lebourg's mean value theorem established in [10, Theorem 2.1] and proved in [11, Theorem 1.7], there exist $t_k^* \in (0, s_k)$ and $g_k^* \in \partial f(x + t_k^* d_k)$ such that

$$f(x + s_k d_k) - f(x) = g_k^{*\top} s_k d_k.$$

Then $t_k^* \downarrow 0$, $x + t_k^* d_k \to x$, and by [22, Theorem 9.13], the sequence g_k^* is bounded, which gives $g_k^{*\top}(d_k - d) \to 0$. Observing that $g_k^* \in \partial f(x + t_k^* d + \theta_k)$ with

$\theta_k = t_k^*(d_k - d)$, $\theta_k/t_k^* = d_k - d \to 0$, due to semismoothness of f, the sequence $g_k^{*\top} d$ has exactly one accumulation point, and so does $g_k^{*\top} d_k = g_k^{*\top} d + g_k^{*\top}(d_k - d)$. On the other hand,

$$\begin{aligned} g_k^{*\top} d_k &= \frac{f(x + s_k d_k) - f(x)}{s_k} \\ &= \frac{f(x + s_k d) - f(x)}{s_k} + \frac{f(x + s_k d_k) - f(x + s_k d)}{s_k}. \end{aligned}$$

The second term tends to 0 as $k \to +\infty$ since f is Lipschitz near x and $d_k \to d$. This implies that $\lim_{k\to+\infty} g_k^{*\top} d_k = f'(x; d)$. Now for any sequence $t_k \downarrow 0$, $g_k \in \partial f(x + t_k d_k)$, then $g_k \in \partial f(x + t_k d + \theta_k)$ with $\theta_k = t_k(d_k - d)$, $\theta_k/t_k = d_k - d \to 0$. We merge two sequences $\{t_k\}$ and $\{t_k^*\}$ into $\{\tau_k\}$, and two corresponding sequences $\{g_k\}$ and $\{g_k^*\}$ into $\{\gamma_k\}$ such that $\tau_k \downarrow 0$. Then again by semismoothness of f and the above argument, all three sequences $g_k^\top d_k$, $g_k^{*\top} d_k$ and $\gamma_k^\top d_k$ have exactly one accumulation point, which implies $\lim_{k\to+\infty} g_k^\top d_k = \lim_{k\to+\infty} g_k^{*\top} d_k = f'(x; d)$. Conversely, writing $t_k d + \theta_k = t_k(d + \theta_k/t_k)$ with $d_k = d + \theta_k/t_k \to d$, we complete the proof of the lemma. $\square$

Corollary 7.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be semismooth at $x \in \mathbb{R}^n$. Then for any $y^k \to x$, $g_k \in \partial f(y^k)$ and for $\varepsilon > 0$,*

$$g_k^\top (x - y^k) \leqslant f(x) - f(y^k) + \varepsilon \|x - y^k\|$$

for infinitely many k.

Proof. Let $y^k \to x$ and $g_k \in \partial f(y^k)$. Passing to a subsequence, we may assume without loss of generality that $d_k = \frac{y^k - x}{\|y^k - x\|} \to d$ as $k \to +\infty$. Set $t_k = \|y^k - x\|$, then $y^k = x + t_k d_k$ and by Lemma 7.1,

$$\lim_{k\to+\infty} \frac{g_k^\top (y^k - x)}{\|y^k - x\|} = f'(x; d).$$

We also have

$$\begin{aligned} \frac{f(y^k) - f(x)}{\|y^k - x\|} &= \frac{f(x + t_k d_k) - f(x)}{t_k} \\ &= \frac{f(x + t_k d) - f(x)}{t_k} + \frac{f(x + t_k d_k) - f(x + t_k d)}{t_k} \end{aligned}$$

converges to $f'(x; d)$ as $k \to +\infty$ due to Lipschitzness of f near x. Hence,

$$\frac{g_k^\top (x - y^k)}{\|x - y^k\|} - \frac{f(x) - f(y^k)}{\|x - y^k\|} \to 0$$

as $k \to +\infty$, which completes the proof. $\square$

We recall here the notion of lower-C^1 and upper-C^1 functions introduced in [23] and [22]. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *lower-*C^1 at (or around) $x_0 \in \mathbb{R}^n$, if there are a compact set S, a neighborhood U of x_0, and a jointly continuous function $g : U \times S \to \mathbb{R}$ whose partial derivative with respect to the first variable is also jointly continuous, such that

$$f(x) = \max_{s \in S} g(x, s)$$

for all $x \in U$. The function f is *upper-*C^1 at x_0 if $-f$ is lower-C^1 at x_0. For the following, we collect some facts on lower-C^1 and upper-C^1 functions.

Remark 7.3. According to Proposition 2.4 and Theorem 3.9 in [23], if f is lower-C^1 at x_0 then f is regular and semismooth at x_0, but the converse need not be true.

Lemma 7.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. For all $x_0 \in \mathbb{R}^n$, the following statements are equivalent.*

(i) *f is lower-C^1 at x_0.*

(ii) *∂f is strictly submonotone at x_0 in the sense that*

$$\liminf_{\substack{x \neq y \\ x, y \to x_0}} \frac{(g_x - g_y)^\top (x - y)}{\|x - y\|} \geqslant 0,$$

whenever $g_x \in \partial f(x)$, $g_y \in \partial f(y)$.

(iii) *For every $\varepsilon > 0$ and x, y close enough to x_0,*

$$g_y^\top (x - y) \leqslant f(x) - f(y) + \varepsilon \|x - y\|,$$

whenever $g_y \in \partial f(y)$.

Proof. The equivalence of (i) and (ii) is already established in [23, Theorem 3.9]. We will show that (ii) and (iii) are equivalent.

$\star$ (ii) $\Rightarrow$ (iii). For any distinct x, y, by Lebourg's mean value theorem, there exist $\lambda \in (0, 1)$ and $g_z \in \partial f(z)$ with $z = \lambda x + (1 - \lambda) y$ such that $f(x) - f(y) = g_z^\top (x - y)$. Take arbitrary $g_y \in \partial f(y)$ and note that $z - y = \lambda(x - y)$, we can write

$$\begin{aligned} f(x) - f(y) &= g_y^\top (x - y) + (g_z - g_y)^\top (x - y) \\ &= g_y^\top (x - y) + \frac{(g_z - g_y)^\top (z - y)}{\|z - y\|} \|x - y\|. \end{aligned}$$

Assume that ∂f is strictly submonotone. Then, for fixed $x_0 \in \mathbb{R}^n$ and $\varepsilon > 0$, there exists $\delta > 0$ such that for any distinct $z, y \in B(x_0, \delta)$,

$$\frac{(g_z - g_y)^\top (z - y)}{\|z - y\|} \geqslant -\varepsilon.$$

Now for every $x, y \in B(x_0, \delta)$, $x \neq y$, we also have $z, y \in B(x_0, \delta)$, $z \neq y$, and thus (iii) holds due to the above expression and estimate.

$\star$ $(iii) \Rightarrow (ii)$. Let $x_0 \in \mathbb{R}^n$ and $\varepsilon > 0$ be fixed. If (iii) holds true, we can pick x, y in a neighborhood of x_0 such that

$$g_y^\top (x - y) \leqslant f(x) - f(y) + \frac{\varepsilon}{2}\|x - y\|,$$

and also

$$g_x^\top (y - x) \leqslant f(y) - f(x) + \frac{\varepsilon}{2}\|y - x\|.$$

After adding these inequalities, reversing the sign and taking the limit inferior, we get (ii). □

By applying Lemma 7.4 to function $-f$, we obtain immediately the following

Corollary 7.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Then f is upper-C^1 at $x_0 \in \mathbb{R}^n$ if and only if for every $\varepsilon > 0$ and x, y close enough to x_0,*

$$g_x^\top (x - y) \leqslant f(x) - f(y) + \varepsilon\|x - y\|,$$

whenever $g_x \in \partial f(x)$.

8. Analysis of the inner loop

In this section we show that the inner loop terminates with a serious iterate after a finite number of steps. The current iterate x is fixed, and so is $Q := Q(x)$. Assume that the inner loop at serious iterate x turns infinitely, then either τ_k is increased infinitely often, or τ_k is frozen from some counter k_0 onwards. These two scenarios will be analyzed in Lemmas 8.1 and 8.2. Denote by $\mathcal{F}$ the feasible set of problem (1), i.e., $\mathcal{F} = \{x \in \mathbf{C} : h(x) \leqslant 0\}$, we have the following results.

Lemma 8.1. *Let f and h be locally Lipschitz on $\mathbb{R}^n$ such that at every point of $\mathcal{F}$, f is semismooth or upper-C^1, and h is semismooth. Suppose that $x^1 \in \mathcal{F}$ and that the inner loop at serious iterate x produces an infinite sequence of null step y^k and the proximity control parameter is increased infinitely often. Then the following statements hold.*

(i) $y^k \to x$ *and* $\Phi_k(y^k, x) \to F(x, x) = 0$ *as* $k \to +\infty$*.*
(ii) $0 \in \partial_1 F(x, x) + \partial i_{\mathbf{C}}(x)$*.*

Proof. (i) We see that the proximity parameter τ_k is never decreased in the inner loop, which combines with the assumption on τ_k implies that $\tau_k \to +\infty$. Since y^k is the optimal solution of the tangent program (3),

$$\tau_k(x - y^k) \in \partial_1(\Phi_k(y^k, x) + i_{\mathbf{C}}(y^k)).$$

Using the subgradient inequality and noting that $\Phi_k(x, x) = 0, x \in \mathbf{C}, y^k \in \mathbf{C}$, we get

$$\tau_k\|x - y^k\|^2 \leqslant \Phi_k(x, x) + i_{\mathbf{C}}(x) - \Phi_k(y^k, x) - i_{\mathbf{C}}(y^k) = -\Phi_k(y^k, x),$$

which implies

$$0 \leqslant \frac{\tau_k}{2}\|x-y^k\|^2 \leqslant -\phi_k(y^k,x) - \frac{1}{2}(x-y^k)^\top(Q+\tau_k I)(x-y^k) \\ \leqslant -\phi_k(y^k,x) \leqslant \|g(x)\|\|x-y^k\|.$$

Here we recall that $Q+\tau_k I \succ 0$ and $m_0(\cdot,x) \leqslant \phi_k(\cdot,x)$ with $m_0(\cdot,x) = g(x)^\top(\cdot - x)$ the exactness plane at x. It thus follows $\tau_k\|x-y^k\| \leqslant 2\|g(x)\|$. This gives $y^k \to x$ since $\tau_k \to +\infty$. Using again the above estimate, we have $\phi_k(y^k,x) \to 0$, and so $\Phi_k(y^k,x) \to 0$.

(ii) Let

$$\mathbf{g}_k^* := (Q+\tau_k I)(x-y^k) \in \partial_1(\phi_k(y^k,x) + i_{\mathbf{C}}(y^k)),$$

then the sequence $\mathbf{g}_k^*$ is bounded since $\|\mathbf{g}_k^*\|$ is proportional to $\tau_k\|x-y^k\| \leqslant 2\|g(x)\|$ for k large enough. Passing to a subsequence if necessary, we may assume without loss that $\mathbf{g}_k^* \to \mathbf{g}^*$ for some $\mathbf{g}^*$. We claim that $\mathbf{g}^* \in \partial_1 F(x,x) + \partial i_{\mathbf{C}}(x)$. For all $y \in \mathbb{R}^n$, the subgradient inequality gives

$$\mathbf{g}_k^{*\top}(y-y^k) \leqslant \phi_k(y,x) + i_{\mathbf{C}}(y) - \phi_k(y^k,x) - i_{\mathbf{C}}(y^k) \\ \leqslant \phi^\uparrow(y,x) - \phi_k(y^k,x) + i_{\mathbf{C}}(y),$$

due to Lemma 6.1 and the fact that $i_{\mathbf{C}}(y^k) = 0$. Passing to the limit in the above estimate and using the results in part *(i)*, we get

$$\mathbf{g}^{*\top}(y-x) \leqslant \phi^\uparrow(y,x) + i_{\mathbf{C}}(y)$$

for all $y \in \mathbb{R}^n$. This together with $\phi^\uparrow(x,x) = 0$ and $i_{\mathbf{C}}(x) = 0$ gives $\mathbf{g}^* \in \partial_1(\phi^\uparrow(x,x) + i_{\mathbf{C}}(x))$. Using again Lemma 6.1, it implies that $\mathbf{g}^* \in \partial_1 F(x,x) + \partial i_{\mathbf{C}}(x)$.

We now prove $\mathbf{g}^* = 0$. Since the inner loop at serious iterate x turns infinitely, $\rho_k < \gamma$ for all k. Moreover, the proximity parameter τ_k is increased infinitely, so there is an infinity of counters k where $\tilde{\rho}_k \geqslant \tilde{\gamma}$. Therefore,

$$\tilde{\gamma} - \gamma < \tilde{\rho}_k - \rho_k = \frac{F(y^k,x) - M_k(y^k,x)}{-\Phi_k(y^k,x)}. \tag{5}$$

It has already been shown in part *(i)* that $-\Phi_k(y^k,x) \geqslant \tau_k\|x-y^k\|^2$. Fixing $0<\delta<1$ and using $\tau_k \to +\infty$ we have

$$\|\mathbf{g}_k^*\| \leqslant (1+\delta)\tau_k\|x-y^k\|,$$

and then

$$-\Phi_k(y^k,x) \geqslant \frac{1}{1+\delta}\|\mathbf{g}_k^*\|\|x-y^k\| \tag{6}$$

for k large enough. Next we estimate the difference $F(y^k,x) - M_k(y^k,x)$. By construction,

$$M_k(y^k,x) \geqslant m_k(y^k,x) + \frac{1}{2}(y^k-x)^\top Q(y^k-x)$$

with $m_k(\cdot, x) = t_k(\cdot) - [t_k(x) + c\|y^k - x\|^2]_+$, where $t_k(\cdot) = F(y^k, x) + g_k^\top(\cdot - y^k)$ and $g_k \in \partial_1 F(y^k, x)$. This gives

$$F(y^k, x) - M_k(y^k, x) \leqslant [t_k(x) + c\|y^k - x\|^2]_+ - \frac{1}{2}(y^k - x)^\top Q(y^k - x).$$

Let us first consider the case when f and h are semismooth at x. Then $F(\cdot, x)$ is semismooth at x due to [15, Theorem 6]. For each $\varepsilon > 0$, using $y^k \to x$ and Corollary 7.2, and passing to a subsequence, we find $k(\varepsilon)$ such that for $k \geqslant k(\varepsilon)$,

$$g_k^\top(x - y^k) \leqslant F(x, x) - F(y^k, x) + \varepsilon\|x - y^k\|,$$

which implies

$$t_k(x) = F(y^k, x) + g_k^\top(x - y^k) \leqslant \varepsilon\|x - y^k\|,$$

and then for k large enough,

$$F(y^k, x) - M_k(y^k, x) \leqslant (1 + \delta)\varepsilon\|x - y^k\|. \tag{7}$$

In the case where f is upper-C^1 and h is semismooth at x, notice that

$$F(y^k, x) = \max\{f(y^k) - f(x), h(y^k)\}$$

since the algorithm assures the feasibility of x when the starting point is feasible. If $f(y^k) - f(x) < h(y^k)$ then $F(y^k, x) = h(y^k)$, $\partial_1 F(y^k, x) = \partial h(y^k)$, and so $t_k(x) = h(y^k) + g_k^\top(x - y^k)$ with $g_k \in \partial h(y^k)$. Using again Corollary 7.2, for k large enough,

$$t_k(x) \leqslant h(x) + \varepsilon\|x - y^k\| \leqslant \varepsilon\|x - y^k\|,$$

which yields (7). On the other hand, noting that the exactness plane $m_0(\cdot, x) = g(x)^\top(\cdot - x)$ is based on $g(x) \in \partial f(x)$ since $h(x) \leqslant 0$, and then applying Corollary 7.5, we get

$$m_0(y^k, x) = g(x)^\top(y^k - x) \geqslant -f(x) + f(y^k) - \varepsilon\|x - y^k\|$$

for k large enough. Now if $f(y^k) - f(x) \geqslant h(y^k)$ then $F(y^k, x) = f(y^k) - f(x)$, and (7) holds true due to the fact that $M_k(y^k, x) \geqslant m_0(y^k, x) + \frac{1}{2}(y^k - x)^\top Q(y^k - x)$.

From (5), (6) and (7), we obtain

$$\|\mathbf{g}_k^*\| \leqslant \frac{(1 + \delta)^2}{\tilde{\gamma} - \gamma}\varepsilon$$

for k large enough. This holds for all $\varepsilon > 0$, so $\mathbf{g}^* = 0$, and the lemma is proved. □

Lemma 8.2. *Let f and h be locally Lipschitz functions. Suppose that the inner loop at serious iterate x produces an infinite sequence of null step y^k and the proximity control parameter is increased finitely often. Then the following statements hold.*

(*i*) $y^k \to x$ *and* $\Phi_k(y^k, x) \to F(x, x) = 0$ *as* $k \to +\infty$.

(*ii*) $0 \in \partial_1 F(x, x) + \partial i_{\mathbf{C}}(x)$.

Proof. (*i*) Since the control parameter τ_k is increased finitely often, it remains unchanged from counter k_0 onwards, i.e., $\tau_k = \tau_{k_0} := \tau$ for all $k \geqslant k_0$. This means that $\rho_k < \gamma$ and $\tilde{\rho}_k < \tilde{\gamma}_k$ for all $k \geqslant k_0$. We consider the objective function of tangent program (3) for $k \geqslant k_0$,

$$\Psi_k(y, x) = \Phi_k(y, x) + \frac{\tau}{2}\|y - x\|^2 = \phi_k(y, x) + \frac{1}{2}\|y - x\|^2_{Q+\tau I},$$

where $\|\cdot\|_{Q+\tau I}$ denotes the Euclidean norm derived from the positive definite matrix $Q + \tau I$. Then

$$\Psi_{k+1}(y, x) = \phi_{k+1}(y, x) + \frac{1}{2}\|y - x\|^2_{Q+\tau I}.$$

It follows from the construction of $\phi_{k+1}(\cdot, x)$ that $\phi_{k+1}(y, x) \geqslant m_k^*(y, x)$ with $m_k^*(\cdot, x)$ the aggregate plane at null step y^k. For $y \in \mathbf{C}$, we have

$$\begin{aligned} m_k^*(y, x) &= \phi_k(y^k, x) + g_k^{*\top}(y - y^k) \\ &= \phi_k(y^k, x) + [(Q + \tau I)(x - y^k)]^\top (y - y^k) - h_k^{*\top}(y - y^k) \\ &\geqslant \phi_k(y^k, x) + (x - y^k)^\top (Q + \tau I)(y - y^k), \end{aligned}$$

by using (4) and noting that $h_k^{*\top}(y - y^k) \leqslant i_{\mathbf{C}}(y) - i_{\mathbf{C}}(y^k) = 0$ due to the subgradient inequality. In addition,

$$\begin{aligned} \|y - x\|^2_{Q+\tau I} &= \|(y^k - x) + (y - y^k)\|^2_{Q+\tau I} \\ &= \|y^k - x\|^2_{Q+\tau I} + \|y - y^k\|^2_{Q+\tau I} - 2(x - y^k)^\top (Q + \tau I)(y - y^k), \end{aligned}$$

using the fact that $(x - y)^\top (Q + \tau I)(y - y^k) = (y - y^k)^\top (Q + \tau I)(x - y)$. Hence, for $y \in \mathbf{C}$,

$$\begin{aligned} \Psi_{k+1}(y, x) &\geqslant \phi_k(y^k, x) + \frac{1}{2}\|y^k - x\|^2_{Q+\tau I} + \frac{1}{2}\|y - y^k\|^2_{Q+\tau I} \\ &= \Psi_k(y^k, x) + \frac{1}{2}\|y - y^k\|^2_{Q+\tau I}. \end{aligned}$$

Substituting $y = y^{k+1}$ and remarking that y^{k+1} is the minimizer of $\Psi_{k+1}(y, x)$, we have

$$\begin{aligned} \Psi_k(y^k, x) + \frac{1}{2}\|y^{k+1} - y^k\|^2_{Q+\tau I} \leqslant \Psi_{k+1}(y^{k+1}, x) &\leqslant \Psi_{k+1}(x, x) \\ &= \Phi_{k+1}(x, x) = 0. \end{aligned}$$

This shows that the sequence $\Psi_k(y^k, x)$ is increasing and bounded above by 0, so $\Psi_k(y^k, x) \to \Psi^*$ as $k \to +\infty$ for some $\Psi^* \leqslant 0$. Letting $k \to +\infty$ in the above inequality, we obtain $\frac{1}{2}\|y^{k+1} - y^k\|^2_{Q+\tau I} \to 0$, which implies

$$\|y^{k+1} - y^k\| \to 0 \quad \text{as } k \to +\infty. \tag{8}$$

On the other hand, proceeding as in the proof of Lemma 8.1, we have

$$\tau\|x - y^k\| \leqslant 2\|g(x)\|, \quad k \geqslant k_0,$$

which proves that the sequence of trial steps y^k is bounded. By combining with (8),

$$\begin{aligned} &\|y^{k+1} - x\|^2_{Q+\tau I} - \|y^k - x\|^2_{Q+\tau I} \\ &= (y^k - y^{k+1})^\top (Q + \tau I)[(y^{k+1} - x) + (y^k - x)] \to 0 \quad \text{as } k \to +\infty. \end{aligned}$$

Recalling that $\phi_k(y, x) = \Psi_k(y, x) - \frac{1}{2}\|y - x\|^2_{Q+\tau I}$ and using the above convergence results, we get

$$\begin{aligned} &\phi_{k+1}(y^{k+1}, x) - \phi_k(y^k, x) \\ &= \Psi_{k+1}(y^{k+1}, x) - \Psi_k(y^k, x) - \frac{1}{2}\left(\|y^{k+1} - x\|^2_{Q+\tau I} - \|y^k - x\|^2_{Q+\tau I}\right) \end{aligned} \tag{9}$$

converges to 0 as $k \to +\infty$.

We now claim that $\phi_{k+1}(y^k, x) - \phi_k(y^k, x) \to 0$, and then also $\Phi_{k+1}(y^k, x) - \Phi_k(y^k, x) \to 0$ as $k \to +\infty$. By the construction of the model $\phi_{k+1}(\cdot, x)$, there exists a cutting plane $m_{i_k}(\cdot, x) = a_{i_k} + g_{i_k}^\top(\cdot - x)$ at null step y^{i_k}, $i_k \in \{1, \dots, k\}$, with $g_{i_k} \in \partial_1 F(y^{i_k}, x)$ such that $\phi_{k+1}(y^k, x) = m_{i_k}(y^k, x)$. Then

$$\phi_{k+1}(y^k, x) = m_{i_k}(y, x) - g_{i_k}^\top(y - y^k) \leqslant \phi_{k+1}(y, x) - g_{i_k}^\top(y - y^k)$$

for all y. Therefore,

$$\begin{aligned} 0 &\leqslant \phi_{k+1}(y^k, x) - \phi_k(y^k, x) \\ &\leqslant \phi_{k+1}(y^{k+1}, x) - \phi_k(y^k, x) + \|g_{i_k}\|\|y^{k+1} - y^k\| \end{aligned}$$

and this term converges to 0 due to (8), (9) and boundedness of g_{i_k}. Here boundedness of the $g_{i_k} \in \partial_1 F(y^{i_k}, x)$ follows from boundedness of the subdifferential of $F(\cdot, x)$ on the bounded set of trial steps y^k (cf. [22, Theorem 9.13]). We obtain $\phi_{k+1}(y^k, x) - \phi_k(y^k, x) \to 0$, and so

$$\Phi_{k+1}(y^k, x) - \Phi_k(y^k, x) \to 0 \quad \text{as } k \to +\infty. \tag{10}$$

We next show that $\Phi_k(y^k, x) \to F(x, x) = 0$, of course also $\phi_k(y^k, x) \to 0$, and then $y^k \to x$ as $k \to +\infty$. Assume this is not the case, then $\eta :=$

$\limsup_{k\to+\infty} \Phi_k(y^k, x) < 0$. Choose $\varepsilon > 0$ such that $0 < \varepsilon < -(1-\tilde{\gamma})\eta$. Thanks to (10), there exists $k_1 \geqslant k_0$ such that

$$\Phi_{k+1}(y^k, x) \leqslant \Phi_k(y^k, x) + \varepsilon$$

for all $k \geqslant k_1$. Since $\tilde{\rho}_k < \tilde{\gamma}$ for all $k \geqslant k_1 \geqslant k_0$ and $\Phi_k(y^k, x) \leqslant \Phi_k(x, x) = 0$,

$$\tilde{\gamma}\Phi_k(y^k, x) \leqslant M_k(y^k, x) \leqslant \Phi_{k+1}(y^k, x) \leqslant \Phi_k(y^k, x) + \varepsilon,$$

using $M_k(\cdot, x) \leqslant \Phi_{k+1}(\cdot, x)$ by construction. Passing to the limit, we get $\tilde{\gamma}\eta \leqslant \eta + \varepsilon$, which contradicts the choice of ε. That gives $\eta = 0$, as claimed.

By the definitions of Φ_k and y^k we have

$$\Phi_k(y^k, x) + \frac{\tau}{2}\|y^k - x\|^2 = \Psi_k(y^k, x) \leqslant \Psi_k(x, x) = \Phi_k(x, x) = 0.$$

This together with $\Phi_k(y^k, x) \to F(x, x) = 0$ gives $y^k \to x$ as $k \to +\infty$.

(*ii*) We observe that by the necessary optimality condition for (3) and the subgradient inequality,

$$\begin{aligned}(x - y^k)^\top(Q + \tau I)(y - y^k) &\leqslant \phi_k(y, x) + i_{\mathbf{C}}(y) - \phi_k(y^k, x) - i_{\mathbf{C}}(y^k)\\ &\leqslant \phi^\uparrow(y, x) + i_{\mathbf{C}}(y) - \phi_k(y^k, x) - i_{\mathbf{C}}(y^k)\end{aligned}$$

for all y. Passing to the limit and noting that $\phi^\uparrow(x, x) = \phi(x, x) = 0$, $i_{\mathbf{C}}(y^k) = i_{\mathbf{C}}(x) = 0$, we obtain

$$0 \leqslant \phi^\uparrow(y, x) + i_{\mathbf{C}}(y) - \phi^\uparrow(x, x) - i_{\mathbf{C}}(x),$$

which implies $0 \in \partial_1(\phi^\uparrow(x, x) + i_{\mathbf{C}}(x))$, and since $\partial_1\phi^\uparrow(x, x) = \partial_1 F(x, x)$, we are done. □

We end this section with the following conclusion.

Proposition 8.3. *Let f and h be locally Lipschitz on $\mathbb{R}^n$ such that at every point of $\mathcal{F}$, f is semismooth or upper-C^1, and h is semismooth. Suppose that $x^1 \in \mathcal{F}$. Then the inner loop finds a serious iterate after a finite number of trial steps.*

Proof. Suppose that the inner loop at serious iterate x turns infinitely. Then, as proved in Lemmas 8.1 and 8.2, we must have $0 \in \partial_1 F(x, x) + \partial i_{\mathbf{C}}(x)$. This contradicts the fact that the inner loop is only entered when $0 \notin \partial_1 F(x, x) + \partial i_{\mathbf{C}}(x)$. □

9. Convergence of the outer loop

We show in this section a strong convergence of our algorithm under the assumption that at every point of the feasible set $\mathcal{F}$, f is lower-C^1 or upper-C^1, and h is lower-C^1. By Proposition 8.3 and Remark 7.3, this assumption on f and h assures that the inner loop always terminates finitely.

Theorem 9.1. *Assume f and h in problem* (1) *are locally Lipschitz on $\mathbb{R}^n$ such that at every point of the feasible set $\mathcal{F}$, f is lower-C^1 or upper-C^1, and h is lower-C^1. Let $x^1 \in \mathcal{F}$ be such that $\{x \in \mathcal{F} : f(x) < f(x^1)\}$ is bounded, and let x^j be the sequence of serious iterates generated by Algorithm* 5.1. *Then x^j is a sequence of feasible points for* (1), *and one of the following two statements holds.*

(*i*) *The sequence x^j ends finitely at an F. John critical point x^{j^*} of* (1). *In the case $j^* > 1$, x^{j^*} is even a Karush–Kuhn–Tucker point.*

(*ii*) *The sequence x^j is bounded infinite, and every accumulation point x^* is an F. John critical point of* (1). *In particular, x^* is either a critical point of constraint violation, or a Karush–Kuhn–Tucker point.*

We see immediately that the feasibility of sequence x^j follows from the feasibility of x^1 and Remark 3.1. If the sequence x^j is finite, then the first statement of the theorem holds due to the stopping test of Algorithm 5.1 and Lemma 2.1. In the sequel, we focus on the case where the sequence x^j is infinite, and suppose that in the jth outer loop, the serious step is accepted at inner loop counter k_j, that is, $x^{j+1} = y^{k_j}$. At the jth outer loop and the kth inner loop, we denote more precisely the proximity control parameter as τ_k^j, and write τ_{k_j} for $\tau_{k_j}^j$. We also write $Q_j := Q(x^j)$ for the matrix of the second-order model, which depends on the serious iterates x^j.

Lemma 9.2. *Let f and h be locally Lipschitz functions, and let $x^1 \in \mathcal{F}$ be such that $\{x \in \mathcal{F} : f(x) < f(x^1)\}$ is bounded. Then the sequence of serious iterates x^j is bounded. In addition, $F(x^{j+1}, x^j) \to 0$, $\tau_{k_j}\|x^j - x^{j+1}\|^2 \to 0$ and $\|x^j - x^{j+1}\|^2_{Q_j+\tau_{k_j}I} \to 0$ as $j \to +\infty$.*

Proof. Following Remark 3.1, the feasibility of x^1 gives $f(x^{j+1}) < f(x^j)$ and $h(x^{j+1}) < 0$ for all j. Thus, x^j is feasible for all j, and sequence $f(x^j)$ is decreasing. This yields $\{x^j : j = 1, 2, \dots\} \subset \{x \in \mathcal{F} : f(x) < f(x^1)\}$, and so the sequence x^j is bounded.

Now for every accumulation point x^* of the sequence x^j, the local Lipschitz continuity of f implies that $f(x^*)$ is an accumulation point of the sequence $f(x^j)$, and then $f(x^j) \to f(x^*)$ due to the monotone sequence theorem. Therefore,

$$\liminf_{j\to+\infty} F(x^{j+1}, x^j) \geqslant \lim_{j\to+\infty} (f(x^{j+1}) - f(x^j)) = 0.$$

This together with $F(x^{j+1}, x^j) < 0$ gives $F(x^{j+1}, x^j) \to 0$ as $j \to +\infty$.

Since $x^{j+1} = y^{k_j}$ is the optimal solution of tangent program (3),

$$(Q_j + \tau_{k_j} I)(x^j - x^{j+1}) \in \partial_1(\phi_{k_j}(x^{j+1}, x^j) + i_{\mathbf{C}}(x^{j+1})).$$

Using the subgradient inequality, we obtain

$$\begin{aligned}
&(x^j - x^{j+1})^\top (Q_j + \tau_{k_j} I)(x^j - x^{j+1}) \\
\leqslant\ & \phi_{k_j}(x^j, x^j) + i_{\mathbf{C}}(x^j) - \phi_{k_j}(x^{j+1}, x^j) - i_{\mathbf{C}}(x^{j+1}) \\
=\ & -\phi_{k_j}(x^{j+1}, x^j) \\
=\ & -\Phi_{k_j}(x^{j+1}, x^j) + \frac{1}{2}(x^j - x^{j+1})^\top Q_j (x^j - x^{j+1}).
\end{aligned}$$

By noting that $Q_j + \tau_{k_j} I \succ 0$, this implies

$$\frac{1}{2}\|x^j - x^{j+1}\|^2_{Q_j + \tau_{k_j} I} + \frac{1}{2}\tau_{k_j}\|x^j - x^{j+1}\|^2 \leqslant -\Phi_{k_j}(x^{j+1}, x^j).$$

Moreover, $-\gamma\Phi_{k_j}(x^{j+1}, x^j) \leqslant -F(x^{j+1}, x^j)$ due to the acceptance test and the fact that $\Phi_{k_j}(x^{j+1}, x^j) \leqslant 0$. Hence,

$$\frac{1}{2}\|x^j - x^{j+1}\|^2_{Q_j + \tau_{k_j} I} + \frac{1}{2}\tau_{k_j}\|x^j - x^{j+1}\|^2 \leqslant -\frac{1}{\gamma}F(x^{j+1}, x^j).$$

Combining with $F(x^{j+1}, x^j) \to 0$, we complete the proof. □

Lemma 9.3. *Let f and h be locally Lipschitz functions, and let $x^1 \in \mathcal{F}$ be such that $\{x \in \mathcal{F} : f(x) < f(x^1)\}$ is bounded. Suppose there exists an infinite subset $J \subset \mathbb{N}$ such that $x^j \to x^*$, $j \in J$. Let $\mathbf{g}_j^* = (Q_j + \tau_{k_j} I)(x^j - x^{j+1})$ be the aggregate subgradient belonging to x^{j+1} in the jth outer loop. Then if the sequence $(\mathbf{g}_j^*)_{j \in J}$ has a subsequence which converges to 0 we have that $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$.*

Proof. Assume that there exists an infinite subset J' of J such that $\mathbf{g}_j^* \to 0$, $j \in J'$. Since $\mathbf{g}_j^* \in \partial_1(\phi_{k_j}(x^{j+1}, x^j) + i_{\mathbf{C}}(x^{j+1}))$, for any $y \in \mathbb{R}^n$, the subgradient inequality gives

$$\begin{aligned}
&\mathbf{g}_j^{*\top}(y - x^{j+1}) \\
\leqslant\ & \phi_{k_j}(y, x^j) + i_{\mathbf{C}}(y) - \phi_{k_j}(x^{j+1}, x^j) - i_{\mathbf{C}}(x^{j+1}) \\
=\ & \phi_{k_j}(y, x^j) - \Phi_{k_j}(x^{j+1}, x^j) + \frac{1}{2}(x^{j+1} - x^j)^\top Q_j (x^{j+1} - x^j) + i_{\mathbf{C}}(y) \\
\leqslant\ & \phi_{k_j}(y, x^j) - \Phi_{k_j}(x^{j+1}, x^j) + \frac{1}{2}\|x^j - x^{j+1}\|^2_{Q_j + \tau_{k_j} I} + i_{\mathbf{C}}(y) \\
\leqslant\ & \phi^\uparrow(y, x^j) - \frac{1}{\gamma}F(x^{j+1}, x^j) + \frac{1}{2}\|x^j - x^{j+1}\|^2_{Q_j + \tau_{k_j} I} + i_{\mathbf{C}}(y).
\end{aligned}$$

Here the last estimate is obtained by Lemma 6.1 and the acceptance test of the algorithm. By passing to the limit and using the hypothesis $\mathbf{g}_j^* \to 0$ and the results from Lemmas 6.1(*iii*) and 9.2, we get

$$0 \leqslant \phi^\uparrow(y, x^*) + i_{\mathbf{C}}(y).$$

It follows that $0 \in \partial_1(\phi^\uparrow(x^*, x^*) + i_{\mathbf{C}}(x^*))$ since $\phi^\uparrow(x^*, x^*) = 0$ and $i_{\mathbf{C}}(x^*) = 0$. Together with $\partial_1\phi^\uparrow(x^*, x^*) = \partial_1 F(x^*, x^*)$, this ends the proof of the lemma. □

Lemma 9.4. *Under the hypotheses of Lemma 9.3, if $\|\mathbf{g}_j^*\| \geqslant \zeta$ for some $\zeta > 0$ and every $j \in J$ then the following statements hold.*

(*i*) $\tau_{k_j} \to +\infty$ *as* $j \in J$, $j \to +\infty$.

(*ii*) *There exists an infinite subset J^+ of J such that the τ-parameter was increased at least once during the jth outer loop for all $j \in J^+$. Suppose this happened for the last time at stage r_j for some r_j. Then $x^j - y^{r_j} \to 0$ and $\phi_{r_j}(y^{r_j}, x^j) \to 0$ as $j \in J^+$, $j \to +\infty$.*

(*iii*) *If at every point of $\mathcal{F}$, f is lower-C^1 or upper-C^1, and h is lower-C^1, then $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$.*

Proof. (*i*) Suppose on the contrary that the sequence $(\tau_{k_j})_{j\in J}$ has a bounded subsequence, then by passing to a subsequence, we may assume without loss of generality that $(\tau_{k_j})_{j\in J}$ is bounded. By combining with boundedness of the Q_j and boundedness of the serious steps x^j shown in Lemma 9.2, there exists an infinite subset J' of J such that $\tau_{k_j} \to \bar{\tau}$, $Q_j \to \bar{Q}$ and $x^j - x^{j+1} \to \Delta x$ as $j \in J', j \to +\infty$. It follows that $\mathbf{g}_j^* \to (\bar{Q} + \bar{\tau} I)\Delta x$ with $\|(\bar{Q} + \bar{\tau} I)\Delta x\| \geqslant \zeta > 0$ and $\mathbf{g}_j^{*\top}(x^j - x^{j+1}) \to \Delta x^\top (\bar{Q} + \bar{\tau} I)\Delta x$. According to Lemma 9.2, $\mathbf{g}_j^{*\top}(x^j - x^{j+1}) = \|x^j - x^{j+1}\|^2_{Q_j + \tau_{k_j} I} \to 0$, which implies $\Delta x^\top (\bar{Q} + \bar{\tau} I)\Delta x = 0$. Since $\bar{Q} + \bar{\tau} I$ is positive semidefinite symmetric, we deduce $(\bar{Q} + \bar{\tau} I)\Delta x = 0$, that contradicts $\|(\bar{Q} + \bar{\tau} I)\Delta x\| \geqslant \zeta > 0$. Hence, $\tau_{k_j} \to +\infty$ as $j \to +\infty$.

(*ii*) For each outer loop counter $j \in J$, either $\tau_{k_j} > \tau_1^j$ or $\tau_{k_j} = \tau_1^j$ with $\tau_1^j \leqslant T < +\infty$ by the algorithm. But $\tau_{k_j} \to +\infty$ as $j \to +\infty$, $j \in J$, set $J^- = \{j \in J : \tau_{k_j} = \tau_1^j\}$ therefore must be finite, which implies the infinity of set $J^+ = \{j \in J : \tau_{k_j} > \tau_1^j\}$. In the remainder of the proof, we will only work with $j \in J^+$. Suppose that for each j, the τ-parameter was increased for the last time at counter r_j, then $r_j \in \{1, \ldots, k_j - 1\}$ since at inner loop counter k_j the serious step is accepted. That means

$$\tau_{k_j} = \tau_{k_j - 1} = \cdots = \tau_{r_j + 1} = \theta \tau_{r_j}.$$

Conforming to the update of the proximity control parameter the increase at stage r_j is due to the fact that

$$\rho_{r_j} < \gamma \quad \text{and} \quad \tilde{\rho}_{r_j} \geqslant \tilde{\gamma}. \tag{11}$$

Noting that $\tau_{r_j} = \theta^{-1}\tau_{k_j} \to +\infty$ and y^{r_j} is the optimal solution of tangent program (3), we have

$$\tau_{r_j}(x^j - y^{r_j}) \in \partial_1(\Phi_{r_j}(y^{r_j}, x^j) + i_{\mathbf{C}}(y^{r_j})).$$

By the subgradient inequality and the fact that $\Phi_{r_j}(x^j, x^j) = 0$, $i_{\mathbf{C}}(x^j) = i_{\mathbf{C}}(y^{r_j}) = 0$,

$$\tau_{r_j}\|x^j - y^{r_j}\|^2 \leqslant -\Phi_{r_j}(y^{r_j}, x^j). \tag{12}$$

It follows that

$$
\begin{aligned}
0 \leqslant \frac{\tau_{r_j}}{2}\|x^j - y^{r_j}\|^2 &\leqslant -\phi_{r_j}(y^{r_j}, x^j) - \frac{1}{2}(x^j - y^{r_j})^\top (Q_j + \tau_{r_j} I)(x^j - y^{r_j}) \\
&\leqslant -\phi_{r_j}(y^{r_j}, x^j) \leqslant \|g(x^j)\|\|x^j - y^{r_j}\|,
\end{aligned}
$$

where $m_0(\cdot, x^j) = g(x^j)^\top(\cdot - x^j)$ is the exactness plane at x^j. This implies $\tau_{r_j}\|x^j - y^{r_j}\| \leqslant 2\|g(x^j)\|$. Remark that the sequence $g(x^j)$ is bounded due to [22, Theorem 9.13], and then $x^j - y^{r_j} \to 0$ since $\tau_{r_j} \to +\infty$. The term $-\phi_{r_j}(y^{r_j}, x^j)$ therefore is squeezed in between two convergent terms with the same limit 0, which gives $\phi_{r_j}(y^{r_j}, x^j) \to 0$.

(iii) We now consider

$$
\tilde{\mathbf{g}}_j := \left(Q_j + \tau_{r_j} I\right)(x^j - y^{r_j}) \in \partial_1(\phi_{r_j}(y^{r_j}, x^j) + i_{\mathbf{C}}(y^{r_j})),
$$

then as $\tau_{r_j} \to +\infty$ and the Q_j are bounded, $\|\tilde{\mathbf{g}}_j\|$ behaves asymptotically like constant times $\tau_{r_j}\|x^j - y^{r_j}\| \leqslant 2\|g(x^j)\|$, which implies boundedness of the sequence $\tilde{\mathbf{g}}_j$. Therefore, possibly passing to a subsequence, we have $\tilde{\mathbf{g}}_j \to \tilde{\mathbf{g}}$ for some $\tilde{\mathbf{g}}$. By using the subgradient inequality and Lemma 6.1, and noting that $i_{\mathbf{C}}(y^{r_j}) = 0$,

$$
\begin{aligned}
\tilde{\mathbf{g}}_j^\top(y - y^{r_j}) &\leqslant \phi_{r_j}(y, x^j) + i_{\mathbf{C}}(y) - \phi_{r_j}(y^{r_j}, x^j) - i_{\mathbf{C}}(y^{r_j}) \\
&\leqslant \phi^\uparrow(y, x^j) + i_{\mathbf{C}}(y) - \phi_{r_j}(y^{r_j}, x^j)
\end{aligned}
$$

for all $y \in \mathbb{R}^n$. Passing to the limit and using the results in part *(ii)*, we obtain

$$
\tilde{\mathbf{g}}^\top(y - x^*) \leqslant \phi^\uparrow(y, x^*) + i_{\mathbf{C}}(y),
$$

which implies $\tilde{\mathbf{g}} \in \partial_1(\phi^\uparrow(x^*, x^*) + i_{\mathbf{C}}(x^*))$ since $\phi^\uparrow(x^*, x^*) = 0$ and $i_{\mathbf{C}}(x^*) = 0$. By Lemma 6.1, we deduce that $\tilde{\mathbf{g}} \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$.

Let us next show $\tilde{\mathbf{g}} = 0$. Fix $0 < \delta < 1$, it follows from $\tau_{r_j} \to +\infty$ that for j large enough,

$$
\|\tilde{\mathbf{g}}_j\| \leqslant (1 + \delta)\tau_{r_j}\|x^j - y^{r_j}\|,
$$

which combined with (12) gives

$$
\|\tilde{\mathbf{g}}_j\| \leqslant (1 + \delta)\frac{-\Phi_{r_j}(y^{r_j}, x^j)}{\|x^j - y^{r_j}\|}. \tag{13}
$$

On the other hand, from (11) we have

$$
\tilde{\gamma} - \gamma < \tilde{\rho}_{r_j} - \rho_{r_j} = \frac{F(y^{r_j}, x^j) - M_{r_j}(y^{r_j}, x^j)}{-\Phi_{r_j}(y^{r_j}, x^j)}. \tag{14}
$$

Remarking that

$$
M_{r_j}(\cdot, x^j) \geqslant m_{r_j}(\cdot, x^j) + \frac{1}{2}(\cdot - x^j)^\top Q_j(\cdot - x^j),
$$

where $m_{r_j}(\cdot, x^j) = t_{r_j}(\cdot) - [t_{r_j}(x^j) + c\|y^{r_j} - x^j\|^2]_+$, and $t_{r_j}(\cdot) = F(y^{r_j}, x^j) + g_{r_j}^\top(\cdot - y^{r_j})$ with $g_{r_j} \in \partial_1 F(y^{r_j}, x^j)$, we get

$$\begin{aligned} &F(y^{r_j}, x^j) - M_{r_j}(y^{r_j}, x^j) \\ &\leqslant [t_{r_j}(x^j) + c\|y^{r_j} - x^j\|^2]_+ - \frac{1}{2}(y^{r_j} - x^j)^\top Q_j (y^{r_j} - x^j). \end{aligned}$$

For $\varepsilon > 0$ fixed, we distinguish the following two cases.

Case I. The both functions f and h are lower-C^1 at x^*, so is $F(\cdot, x^j)$. By the assumption that $x^j \to x^*$ and the fact that $x^j - y^{r_j} \to 0$ proved in part (ii), thanks to Lemma 7.4, there exists $j(\varepsilon)$ such that

$$g_{r_j}^\top(x^j - y^{r_j}) \leqslant F(x^j, x^j) - F(y^{r_j}, x^j) + \varepsilon\|x^j - y^{r_j}\|$$

for every $j \geqslant j(\varepsilon)$. This implies

$$t_{r_j}(x^j) = F(y^{r_j}, x^j) + g_{r_j}^\top(x^j - y^{r_j}) \leqslant \varepsilon\|x^j - y^{r_j}\|,$$

and thus for j large enough,

$$F(y^{r_j}, x^j) - M_{r_j}(y^{r_j}, x^j) \leqslant (1+\delta)\varepsilon\|x^j - y^{r_j}\|. \tag{15}$$

Case II. The function f is upper-C^1 and the function h is lower-C^1 at x^*. By the feasibility of x^j, if $f(y^{r_j}) - f(x^j) < h(y^{r_j})$ then

$$F(y^{r_j}, x^j) = \max\{f(y^{r_j}) - f(x^j), h(y^{r_j})\} = h(y^{r_j}), \quad \partial_1 F(y^{r_j}, x^j) = \partial h(y^{r_j}),$$

and therefore the tangent $t_{r_j}(\cdot) = h(y^{r_j}) + g_{r_j}^\top(\cdot - y^{r_j})$ with $g_{r_j} \in \partial h(y^{r_j})$. The estimate (15) holds based on the inequality

$$t_{r_j}(x^j) \leqslant h(x^j) + \varepsilon\|x^j - y^{r_j}\| \leqslant \varepsilon\|x^j - y^{r_j}\|, \quad \text{for } j \text{ large enough},$$

using Lemma 7.4. Conversely, if $f(y^{r_j}) - f(x^j) \geqslant h(y^{r_j})$ then $F(y^{r_j}, x^j) = f(y^{r_j}) - f(x^j)$, and by recalling the exactness plane $m_0(\cdot, x^j) = g(x^j)^\top(\cdot - x^j)$ with $g(x^j) \in \partial f(x^j)$, we have

$$\begin{aligned} &M_{r_j}(y^{r_j}, x^j) - \frac{1}{2}(y^{r_j} - x^j)^\top Q_j (y^{r_j} - x^j) \\ &\geqslant m_0(y^{r_j}, x^j) \geqslant -f(x^j) + f(y^{r_j}) - \varepsilon\|x^j - y^{r_j}\| \end{aligned}$$

due to Corollary 7.5. This gives (15).

Now it follows from (13), (14) and (15) that

$$\|\tilde{\mathbf{g}}_j\| \leqslant \frac{(1+\delta)^2}{\tilde{\gamma} - \gamma}\varepsilon$$

for j large enough. Since $\varepsilon > 0$ is arbitrary, we conclude that $\tilde{\mathbf{g}} = 0$, meaning $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$. □

Proof of Theorem 9.1. As discussed just after the statement of the theorem, the sequence x^j consists of feasible points for (1) and verifies statement (i) when it is finite. Suppose that the sequence x^j is infinite, then it is bounded by Lemma 9.2. Let x^* be an accumulation point of the sequence x^j, we have $h(x^*) \leqslant 0, x^* \in \mathbf{C}$ due to feasibility of x^j for all j, continuity of $h(\cdot)$ and closed convexity of $\mathbf{C}$. It follows from Lemmas 9.3 and 9.4 that $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$. This together with Lemma 2.1 gives the last statement of the theorem. □

Notice that Algorithm 5.1 also works for solving the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{C} \end{array} \tag{16}$$

where f is a real-valued locally Lipschitz but potentially nonsmooth and nonconvex function, and $\mathbf{C}$ a closed convex set of $\mathbb{R}^n$. In this case, we use the progress function $F(y, x) = f(y) - f(x)$ which has $\partial_1 F(x, x) = \partial f(x)$, and remark that the feasible set $\mathcal{F} = \mathbf{C}$. An immediate consequence of Theorem 9.1 is as follows.

Proposition 9.5. *Assume f in problem* (16) *is locally Lipschitz on $\mathbb{R}^n$ such that at every $x \in \mathbf{C}$, f is either lower-C^1 or upper-C^1. Let x^1 be such that $\{x \in \mathbf{C} : f(x) < f(x^1)\}$ is bounded. Then every accumulation point x^* of the sequence of serious iterates x^j generated by Algorithm* 5.1 *is a critical point of* (16) *in the sense that* $0 \in \partial f(x^*) + \partial i_{\mathbf{C}}(x^*)$.

Remark 9.6. Theorem 9.1 requires a feasible starting point x^1 that can be computed by minimizing $h(x)$ on $\mathbf{C}$, which is in the form of problem (16). This, however, does not mean that we are interested in the global minimum of $h(x)$ on $\mathbf{C}$. A nonpositive value of h is all what we need, so $h(x)$ is minimized until an iterate $x^1 \in \mathbf{C}$ with $h(x^1) \leqslant 0$ is found.

In the case where the starting point is not necessarily feasible for (1), we will work with the following assumptions from [1] and [7]:

(A1) f is weakly coercive[1] on the feasible set $\mathcal{F}$ in the sense that $f(x^j)$ is not strictly decreasing as $x^j \in \mathcal{F}$, $\|x^j\| \to +\infty$.

(A2) h is weakly coercive on $\mathbf{C}$ in the sense that $h(x^j)$ is not strictly decreasing as $x^j \in \mathbf{C}$, $\|x^j\| \to +\infty$.

Note that assumption (A1) holds when $\{x \in \mathcal{F} : f(x) < f(x^1)\}$ is bounded, and likewise, assumption (A2) is satisfied if $\{x \in \mathbf{C} : h(x) < h(x^1)\}$ is bounded. The following result can be seen as an extension of convergence theorems in [1, 7] to our more general setting.

Theorem 9.7. *Assume f and h in problem* (1) *are locally Lipschitz on $\mathbb{R}^n$ and lower-C^1 on $\mathbf{C}$ such that* (A1) *and* (A2) *hold. Then the sequence of serious iterates x^j generated by Algorithm* 5.1 *either ends finitely at a point $x^* = x^{j^*} \in \mathbf{C}$*

[1]A function $f : \mathbb{R}^n \to \mathbb{R}$ is coercive on $D \subset \mathbb{R}^n$ if $f(x) \to +\infty$ as $x \in D$, $\|x\| \to +\infty$.

with $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$, *or is bounded infinite in which every accumulation point* x^* *satisfies* $x^* \in \mathbf{C}$ *and* $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$. *Furthermore,* x^* *is either a critical point of constraint violation, or a Karush–Kuhn–Tucker point of* (1).

Proof. We see in the proof of Theorem 9.1 that the feasibility of x^1 is only used when f is upper-C^1 in Lemmas 8.1 and 9.4, and when showing the boundedness of the sequence x^j in Lemma 9.2. The proof thus follows from Theorem 9.1 as soon as the boundedness of the sequence x^j is shown. Let us first note that $x^j \in \mathbf{C}$ for all j due to Remark 3.1. If $h(x^j) > 0$ for all j, then also by Remark 3.1, the sequence $h(x^j)$ is strictly decreasing, and assumption (A2) implies that the sequence x^j is bounded. If $h(x^{j_0}) \leqslant 0$ for some $j_0 \in \mathbb{N}$, using again Remark 3.1 we have $f(x^{j+1}) < f(x^j)$ and $h(x^{j+1}) < 0$ for all $j \geqslant j_0$. Therefore, $x^j \in \mathcal{F}$ and the sequence $f(x^j)$ is strictly decreasing from j_0 onwards. This together with assumption (A1) gives the boundedness of the sequence x^j. □

In practice, a challenge is the lack of convexity, by which it is difficult to guarantee convergence to a single critical point. Some satisfactory results can nevertheless be obtained from the following corollaries.

Corollary 9.8. *Under the hypotheses of Theorem* 9.1, *for every* $\varepsilon > 0$ *there exists an index* $j_0(\varepsilon) \in \mathbb{N}$ *such that every* $j \geqslant j_0(\varepsilon)$, x^j *is within* ε*-distance of the set*

$$L = \{x^* \in \mathbf{C} : 0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)\}.$$

Proof. Suppose there exists $\varepsilon > 0$ and an infinite subsequence x^j, $j \in J$, such that $\|x^j - x^*\| > \varepsilon$ for all $j \in J$ and all $x^* \in L$. Since the sequence x^j, $j \in J$, is bounded, it has an accumulation point x^*, and by Theorem 9.1, $x^* \in L$. That is a contradiction. □

Corollary 9.9. *Under the hypotheses of Theorem* 9.1, *if the set* L *in Corollary* 9.8 *is totally disconnected in the sense of* [6, *Definition* 9.4.1], *then the sequence* x^j *converges to a single point* $x^* \in \mathbf{C}$ *with* $0 \in \partial_1 F(x^*, x^*) + \partial i_{\mathbf{C}}(x^*)$.

Proof. Recall that $\|x^j - x^{j+1}\|^2_{Q_j + \tau_{k_j} I} \to 0$ as $j \to +\infty$ due to Lemma 9.2. In each outer loop counter j, since $T > q + \kappa \geqslant -\lambda_{\min}(Q_j) + \kappa$, so

$$\tau_{k_j} \geqslant \tau_1 \geqslant -\lambda_{\min}(Q_j) + \kappa,$$

and therefore $\lambda_{\min}(Q_j + \tau_{k_j} I) \geqslant \kappa$, which implies

$$\|x^j - x^{j+1}\|^2_{Q_j + \tau_{k_j} I} \geqslant \kappa \|x^j - x^{j+1}\|^2.$$

It follows that $\|x^j - x^{j+1}\|^2 \to 0$, and also $x^j - x^{j+1} \to 0$ as $j \to +\infty$. Now let K be the set of accumulation points of the sequence x^j. Then K is either a singleton or a continuum (see [20, Theorem 26.1] and its following remark). On the other hand, $K \subset L$ thanks to Theorem 9.1. Since L is totally disconnected, i.e., for any two distinct points u and v of L, there exist disjoint open sets U and V such that $u \in U$, $v \in V$ and $L \subset U \cup V$, we can decompose K into the union of two disjoint closed sets, and so K cannot be a continuum. Hence, K must be a singleton. □

In the case where subgradients are inexact, working with the approximate subdifferential

$$\partial^\varepsilon f(x) = \partial f(x) + \varepsilon B,$$

where ∂ is the exact Clarke subdifferential, and B the unit ball in some fixed Euclidean norm, we have the following

Corollary 9.10. *Assume f and h in problem* (1) *are locally Lipschitz on $\mathbb{R}^n$ such that at every point of $\mathcal{F}$, f is lower-C^1 or upper-C^1, and h is lower-C^1. Let $x^1 \in \mathcal{F}$ be such that $\{x \in \mathcal{F} : f(x) < f(x^1)\}$ is bounded, and assume that subgradients are drawn from $\partial_1^\varepsilon F(y,x)$, whereas function values are exact. Then the sequence of serious iterates x^j is a bounded sequence of feasible points for* (1)*, and every accumulation point x^* of the x^j satisfies $h(x^*) \leqslant 0, x^* \in \mathbf{C}$ and $0 \in \partial_1^{\tilde{\varepsilon}} F(x^*,x^*) + \partial^{\tilde{\varepsilon}} i_{\mathbf{C}}(x^*)$, where $\tilde{\varepsilon} = (1 + (\tilde{\gamma} - \gamma)^{-1})\varepsilon$.*

Proof. Noting that in this case $\partial_1 \phi^\uparrow(x,x) = \partial_1^\varepsilon F(x,x)$, we proceed as in the proof of Theorem 9.1, and have just to replace (7) and (15) by the following estimates for every $\varepsilon' > 0$,

$$F(y^k,x) - M_k(y^k,x) \leqslant (1+\delta)(\varepsilon' + \varepsilon)\|x - y^k\| \quad \text{for } k \text{ large enough,}$$
$$F(y^{r_j},x^j) - M_{r_j}(y^{r_j},x^j) \leqslant (1+\delta)(\varepsilon' + \varepsilon)\|x^j - y^{r_j}\| \quad \text{for } j \text{ large enough.}$$

For a detailed proof in the case of unconstrained optimization, we refer to [18]. □

10. Conclusion

We have presented a nonconvex bundle method using downshifted tangents and the management of proximity control, which is adapted for nonconvex nonsmooth constrained optimization problems with lower-C^1 and upper-C^1 functions. A global convergence of the algorithm was proved in the sense that every accumulation point of the sequence of serious iterates is critical. Some satisfactory convergence results for practical purpose have been given as corollaries.

Acknowledgements. The author thanks Professor Dominikus Noll for many useful discussions, and acknowledges the anonymous reviewers for valuable comments.

References

[1] P. Apkarian, D. Noll, A. Rondepierre: Mixed H_2/H_∞ control via nonsmooth optimization, SIAM J. Control Optim. 47(3) (2008) 1516–1546.

[2] F. H. Clarke: A new approach to Lagrange multipliers, Math. Oper. Res. 1(2) (1976) 165–174.

[3] F. H. Clarke: Generalized gradients of Lipschitz functionals, Adv. Math. 40(1) (1981) 52–67.

[4] F. H. Clarke: Optimization and Nonsmooth Analysis, Canad. Math. Soc. Ser. Monogr. Adv. Texts, John Wiley & Sons, New York (1983).

[5] M. N. Dao, D. Noll: Minimizing memory effects of a system, Math. Control Signals Syst. 27(1) (2015) 77–110.

[6] K. R. Davidson, A. P. Donsig: Real Analysis with Real Applications, Prentice Hall, Upper Saddle River (2002).

[7] M. Gabarrou, D. Alazard, D. Noll: Design of a flight control architecture using a non-convex bundle method, Math. Control Signals Syst. 25(2) (2013) 257–290.

[8] J.-B. Hiriart-Urruty: Mean value theorems in nonsmooth analysis, Numer. Funct. Anal. Optim. 2(1) (1980) 1–30.

[9] K. C. Kiwiel: An aggregate subgradient method for nonsmooth convex minimization, Math. Program. 27(3) (1983) 320–341.

[10] G. Lebourg: Valeur moyenne pour gradient généralisé, C. R. Acad. Sci. Paris Sér. A 281(19) (1975) 795–797.

[11] G. Lebourg: Generic differentiability of Lipschitzian functions, Trans. Amer. Math. Soc. 256 (1979) 125–144.

[12] C. Lemaréchal: Bundle methods in nonsmooth optimization, in: Nonsmooth Optimization (Laxenburg, 1977), C. Lemaréchal, R. Mifflin (eds.), IIASA Proc. Ser. 3, Pergamon Press, Oxford (1978) 79–102.

[13] C. Lemaréchal, A. Nemirovskii, Y. Nesterov: New variants of bundle methods, Math. Program., Ser. B 69(1) (1995) 111–147.

[14] M. M. Mäkelä, P. Neittaanmäki: Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control, World Scientific, Singapore (1992).

[15] R. Mifflin: Semismooth and semiconvex functions in constrained optimization, SIAM J. Control Optim. 15(6) (1977) 959–972.

[16] R. Mifflin: A modification and extension of Lemarechal's algorithm for nonsmooth minimization, in: Nondifferentiable and Variational Techniques in Optimization (Lexington, 1980), D. C. Sorensen, R. J.-B. Wets (eds.), Math. Programming Stud. 17, North-Holland, Amsterdam (1982) 77–90.

[17] D. Noll: Cutting plane oracles to minimize non-smooth non-convex functions, Set-Valued Var. Anal. 18(3–4) (2010) 531–568.

[18] D. Noll: Bundle method for non-convex minimization with inexact subgradients and function values, in: Computational and Analytical Mathematics, D. H. Bailey et al. (eds.), Springer Proc. Math. Stat. 50, Springer, New York (2013) 555–592.

[19] D. Noll, O. Prot, A. Rondepierre: A proximity control algorithm to minimize nonsmooth and nonconvex functions, Pacific J. Optim. 4(3) (2008) 571–604.

[20] A. M. Ostrowski: Solutions of Equations in Euclidean and Banach Spaces, Pure and Applied Mathematics 9, Academic Press, New York (1973).

[21] E. Polak: Optimization: Algorithms and Consistent Approximations, Appl. Math. Sci. 124, Springer, New York (1997).

[22] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Springer, Berlin (1998).

[23] J. E. Spingarn: Submonotone subdifferentials of Lipschitz functions, Trans. Amer. Math. Soc. 264(1) (1981) 77–89.