A note on optimality conditions for optimization problems with empty limiting subdifferentials*

R. Andreani† V. R. Chagas† M. da Rosa† L. D. Secchin‡

October 2, 2026 (last updated October 10, 2026)

Abstract

We study first-order optimality for constrained composite optimization problems whose objective is the sum of a locally Lipschitz function and a proper lower semicontinuous function. We focus on the case in which the limiting subdifferential of the latter function is empty at a point of interest, so that the usual KKT conditions are unavailable. We examine what optimality conditions remain in this setting. We show that any local minimizer must have an active constraint and satisfies a Fritz–John condition based on the horizon subdifferential, independently of any constraint qualification. We then establish an approximate KKT condition that retains information from the Lipschitzian part of the objective. Under differentiability away from the point of interest, a safeguarded augmented Lagrangian method generates a sequence of distinct points satisfying this condition and converging to the minimizer. If the function values of the lower semicontinuous term also converge, the sequence yields the horizon-based condition.

1 Introduction

We investigate constrained composite optimization problems of the form

$$\min_x \ f(x) + \Phi(x) \qquad \text{s.t.} \quad h(x) = 0,\ g(x) \le 0, \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$ are locally Lipschitz, and $\Phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lower semicontinuous, but not necessarily locally Lipschitz at a minimizer x^*.

The nonsmooth terms in (P) call for generalized subdifferentials in the formulation of first-order optimality conditions. For general nonconvex problems, the Fréchet and Mordukhovich subdifferentials are commonly used [8], while the Clarke subdifferential is useful for locally Lipschitz functions [9]. In particular, KKT-type conditions involving the Mordukhovich subdifferential of Φ, also called *limiting subdifferential* and denoted by $\partial\Phi$, have been established for (P) under suitable constraint qualifications (CQs) and subdifferential sum-rule conditions [8, 9].

When $\partial\Phi(x^*) \neq \emptyset$, optimality conditions from the smooth setting can be adapted by replacing gradients with $\partial\Phi(x^*)$. However, the case $\partial\Phi(x^*) = \emptyset$ is more delicate. In the literature, Fritz–John-type conditions for this case are derived using the so-called *horizon* subdifferential $\partial^\infty\Phi(x^*)$ [9]; see [7, 10]. Geometrically, $\partial^\infty\Phi(x^*)$ captures arbitrarily steep variations of Φ around x^*, which cannot be represented by finite subgradients (see Definition 1). Although the condition $\partial\Phi(x^*) = \emptyset$ may appear exceptional, it can occur in elementary problems; see Example 1.

However, optimality using $\partial^\infty\Phi(x^*)$ has an important limitation: the Lipschitzian term f is not present, thereby losing the complete first-order structure of the objective. This motivates the

*This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2023/08706-1 and 2026/08792-3), CNPq (grants 302213/2026-0, 407147/2023-3, and 302520/2025-2), and CAPES (grant 88887.001429/2024-00).

†Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil. Email: andreani@ime.unicamp.br, m265027@dac.unicamp.br, v291045@dac.unicamp.br

‡Department of Applied Mathematics, Federal University of Espírito Santo, ES, Brazil. Email: leonardo.secchin@ufes.br

sequential analyses developed in this work. We investigate how a local minimizer with an empty limiting subdifferential can be approached by points at which ordinary first-order information is still available. This allows us to preserve the contribution of f and, at the same time, establish a connection with sequences generated by optimization methods. Specifically, our contributions regarding (P) are:

- we show that, for x^* to be a local minimizer with $\partial\Phi(x^*) = \emptyset$, some constraint must be active at x^*;
- we clarify the Fritz–John alternative when $\partial\Phi(x^*) = \emptyset$, showing that it necessarily holds without any CQ. We also explain why horizon-adapted CQs, as proposed in [7, 10], are unnecessary for obtaining this certificate;
- we prove that every local minimizer with $\partial\Phi(x^*) = \emptyset$ admits an approximate KKT (AKKT) sequence $\{x^k\}$ with $x^k \neq x^*$, in the spirit of the construction used in smooth optimization [2]. Unlike the horizon condition, this sequential condition retains the first-order information of f;
- we connect AKKT with a safeguarded augmented Lagrangian (AL) method discussed, e.g., in [1, 5, 6]. Under the additional hypothesis $\Phi(x^k) \to \Phi(x^*)$, we are able to recover the horizon-based optimality. We also show by an example that this function-value convergence cannot be omitted.

The paper is organized as follows. Section 2 recalls useful concepts from variational analysis and Fritz–John conditions for (P). In Section 3, we present our sequential optimality condition and investigate its connection with AL methods. Finally, Section 4 brings our conclusions.

Notation $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$ and $\|\cdot\|_\infty$ denotes the maximum norm. The open and closed balls centered at x with radius ε are denoted by $B(x,\varepsilon)$ and $B[x,\varepsilon]$, respectively; unless otherwise specified, the norm will be arbitrary. $\mathbb{R}_+$ is the set of nonnegative real numbers and $\overline{\mathbb{R}} = \mathbb{R}\cup\{-\infty,\infty\}$. Given $z \in \mathbb{R}^n$, we write $z_+ = (\max\{0, z_1\}, \ldots, \max\{0, z_n\})$. For sets A and B, $A + B = \{a + b \mid a \in A,\ b \in B\}$. We write $t_k \downarrow 0$ to denote that the sequence $\{t_k\}$ of positive scalars converges to zero. We write $x^k \to_\Psi x^*$ to denote $\lim_k x^k = x^*$ and $\lim_k \Psi(x^k) = \Psi(x^*)$.

2 Mathematical background

2.1 Useful concepts and results from variational analysis

Following [9], a function $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *lower semicontinuous at* x if, for any sequence $\{x^k\}$ converging to x, we have

$$\liminf_{k\to\infty} \Psi(x^k) \geq \Psi(x).$$

If Ψ is lower semicontinuous at all points of its domain, we simply say that Ψ is *lower semicontinuous* (l.s.c. for short). Moreover, we say that Ψ is *proper* if $\Psi(x) < \infty$ for at least one x and $\Psi(x) > -\infty$ for all $x \in \mathbb{R}^n$. $\Psi(x)$ is said to be *finite* whenever $\Psi(x) \in \mathbb{R}$. Finally, a finite sum of proper and l.s.c. functions is l.s.c.; see [9, Proposition 1.39].

Definition 1 ([9, Definition 8.3]). *Let $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$. For any $x \in \mathbb{R}^n$ such that $\Psi(x)$ is finite,*

- *the* regular *(or* Fréchet*)* subdifferential of Ψ at x *is defined by*

$$\widehat{\partial}\Psi(x) := \{v \mid \Psi(y) \geq \Psi(x) + v^T(y - x) + o(\|y - x\|)\ \forall y\};$$

- *the* limiting *(or* Mordukhovich*)* subdifferential of Ψ at x *is defined by*

$$\partial\Psi(x) := \{v \mid \exists x^k \to_\Psi x,\ v^k \in \widehat{\partial}\Psi(x^k)\ s.t.\ v^k \to v\} = \limsup_{y\to_\Psi x} \widehat{\partial}\Psi(y);$$

- *and the* horizon subdifferential of Ψ at x *is defined by*

$$\partial^\infty \Psi(x) := \{v \mid \exists x^k \to_\Psi x,\ v^k \in \widehat{\partial}\Psi(x^k),\ t_k \downarrow 0,\ \text{such that } t_k v^k \to v\}.$$

Proposition 1 (Subdifferential properties [9, Theorem 8.6, Exercise 8.8]). *For every $x \in \mathbb{R}^n$ such that $\Psi(x)$ is finite we have*

(i) $\widehat{\partial}\Psi(x) \subset \partial\Psi(x)$;

(ii) if additionally Ψ is smooth on an open neighborhood of x, then $\widehat{\partial}\Psi(x) = \partial\Psi(x) = \{\nabla\Psi(x)\}$ *and* $\partial^\infty\Psi(x) = \{0\}$;

(iii) if $\Psi = \Psi_0 + \Psi_1$, where Ψ_0 is smooth on an open neighborhood of x and $\Psi_1(x)$ is finite, then $\partial\Psi(x) = \nabla\Psi_0(x) + \partial\Psi_1(x)$ *and* $\partial^\infty\Psi(x) = \partial^\infty\Psi_1(x)$.

A useful characterization of local Lipschitz continuity is given in terms of the horizon subdifferential [10]: if $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. and finite at x^*, then

$$\Psi \text{ is locally Lipschitz around } x^* \quad \Longleftrightarrow \quad \partial^\infty\Psi(x^*) = \{0\}. \tag{1}$$

Below, we list other useful results.

Proposition 2 (Lipschitz sum rule [10, Proposition 2(i)]). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz around x^*, and let $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be l.s.c. and finite at x^*. Then, for any $\alpha, \beta > 0$, we have* $\partial(\alpha f + \beta\Phi)(x^*) \subset \alpha\partial f(x^*) + \beta\partial\Phi(x^*)$.

Proposition 3 ([9, Theorem 1.9.]). *Let $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a l.s.c. and proper function such that the level set $\{x \mid \Psi(x) \le \alpha\}$ is bounded for all $\alpha \in \mathbb{R}$. Then* $\inf \Psi(x) > -\infty$ *and the set of minima of Ψ is nonempty and compact.*

Proposition 4 (Fermat rule [9, Theorem 10.1]). *If x^* is a local minimizer of a proper function $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$, then* $0 \in \widehat{\partial}\Psi(x^*)$.

2.2 Fritz–John optimality conditions for (P)

Fritz–John conditions for (P) were established in [7]. Specifically, if x^* is a local minimizer, then there exist multipliers $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}_+^{|I(x^*)|}$, not all zero, such that at least one of the following conditions holds:

$$0 \in \partial^\infty\Phi(x^*) + \sum_{j=1}^m \partial(\lambda_j h_j)(x^*) + \sum_{i \in I(x^*)} \mu_i \partial g_i(x^*), \tag{2}$$

$$0 \in \partial(f + \Phi)(x^*) + \sum_{j=1}^m \partial(\lambda_j h_j)(x^*) + \sum_{i \in I(x^*)} \mu_i \partial g_i(x^*), \tag{3}$$

where $I(x^*) = \{i \mid g_i(x^*) = 0\}$. Expression (3) is known as the KKT condition for (P). Since this condition cannot hold when $\partial\Phi(x^*) = \emptyset$, the assumptions imposed in [7] to guarantee KKT at a local minimizer necessarily exclude this case. Note that when Φ is Lipschitz around x^*, we have $\partial^\infty\Phi(x^*) = \{0\}$ by (1), and thus (2) becomes a linear independence-type condition between the subdifferentials. In particular, (2)–(3) recover the classical necessary condition "KKT or not LICQ" when all data functions are smooth ("LICQ" stands for the *linear independence CQ*).

As we mentioned in the introduction, we focus on the case $\partial\Phi(x^*) = \emptyset$, for which KKT condition (3) is automatically ruled out. The next example shows that this phenomenon may occur at a local minimizer even when the constraints satisfy LICQ. Nevertheless, conditions (2) may still hold, which motivates our analysis of its role in this setting.

Example 1. *Consider the problem of minimizing $x + x^{1/3}$ subject to $-x \le 0$. In view of (P), we set $f(x) = x$ and $\Phi(x) = x^{1/3}$. At the unique minimizer $x^* = 0$, the gradient of the active constraint is non-zero, and thus LICQ holds.*

However, Proposition 2 implies $\partial(f+\Phi)(0) \subset \partial f(0) + \partial\Phi(0) = \emptyset$ *because* $\partial\Phi(0) = \emptyset$ *To verify that* $\partial\Phi(0) = \emptyset$*, note that for any* $y < 0$ *we have*

$$\frac{\Phi(y) - \Phi(0) - v(y-0)}{|y-0|} = -\frac{1}{y^{2/3}} + v, \quad \text{for all } v \in \mathbb{R},$$

which converges to $-\infty$ *as* $y \to 0$*. This shows that* $\widehat{\partial}\Phi(0) = \emptyset$*. Furthermore, for* $x \neq 0$ *we have* $\partial\Phi(x) = \{\nabla\Phi(x)\} = \left\{\frac{1}{3}x^{-2/3}\right\}$*, which tends to* ∞ *as* $x \to 0$*. Thus,* $\partial\Phi(0) = \emptyset$ *(see Definition 1), and (3) does not hold.*

On the other hand, for any given $v \geq 0$*, we choose* $x^k := 1/k$*,* $v^k := \nabla\Phi(x^k) = \frac{1}{3}k^{2/3} \in \widehat{\partial}\Phi(x^k)$*, and* $t_k := v/v^k$ *and use the last item of Definition 1 to conclude that* $\partial^\infty\Phi(0) = [0, \infty)$*. Thus, (2) holds for any* $\mu > 0$*.*

3 Optimality under empty limiting subdifferential

Since the usual KKT condition (3) is unavailable when $\partial\Phi(x^*) = \emptyset$, we examine what can still be said about local minimizers in this setting. We first show that a local minimizer must have an active constraint and that condition (2) holds independently of any CQ. We then introduce a sequential optimality condition that retains the first-order information of the Lipschitzian term f. Throughout the remainder of the paper, we assume that $\Phi(x^*)$ is finite.

3.1 Basic facts and the role of CQs

In Example 1, the unique constraint is active at the minimizer and, if it is removed, $x^* = 0$ is no longer a minimizer of $f + \Phi$. Geometrically, the bounded slope of the graph of any Lipschitz function f is insufficient to compensate for the vertical tangent to the graph of Φ at x^*. In other words, for x^* to be a local minimizer of $f + \Phi$ with $\partial\Phi(x^*) = \emptyset$, some constraint must be active. The next result shows that this is a general phenomenon.

Theorem 1. *If* $\partial\Phi(x^*) = \emptyset$*, then* x^* *cannot be an unconstrained local minimizer of* $f + \Phi$*.*

Proof. If x^* were an unconstrained local minimizer of $f + \Phi$, then Propositions 1, 2, and 4 would imply $0 \in \widehat{\partial}(f+\Phi)(x^*) \subset \partial(f+\Phi)(x^*) \subset \partial f(x^*) + \partial\Phi(x^*) = \emptyset$, which is a contradiction. □

In view of the above result, we henceforth focus on constrained minimization problems. Together with $\partial\Phi(x^*) = \emptyset$, the very existence of a non-null multiplier vector (λ, μ) satisfying (2), as discussed in [7], shows that the constraints need not satisfy a CQ for (2) to hold. In this case, $\partial^\infty\Phi(x^*)$ provides a "legitimate certificate of optimality". This is summarized in the next result.

Theorem 2. *If* x^* *is a local minimizer of (P) with* $\partial\Phi(x^*) = \emptyset$*, then (2) holds for some* $0 \neq (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^{|I(x^*)|}$*, independently of any CQ.*

Remark 1. *In [7, Definition 1], the so-called* ∂^∞-quasinormality *CQ was defined, which states that there is no nonzero vector* (λ, μ)*,* $\mu \geq 0$*, such that there exists a sequence* $\{x^k\}$ *converging to* x^* *for which (2) holds,* $\lambda_i h_i(x^k) > 0$*, and* $g_j(x^k) > 0$ *for all* k*,* $i \in \{i \mid \lambda_i \neq 0\}$*, and* $j \in \{j \mid \mu_j > 0\}$*. In other words,* ∂^∞*-quasinormality consists of using (2) instead of the classical expression, without* $\partial^\infty\Phi(x^*)$*. However, in view of Theorem 2, such adaptations are not necessary since CQs only need to be imposed to ensure (3). The same comment is valid for other adapted CQs, like* ∂^∞*-RCPLD and* ∂^∞*-NNAMCQ [7].*

3.2 Approximate optimality

Note that expression (2) does not contain the Lipschitzian part of the objective function, f. In this section, we provide an approximate characterization of optimality that takes into account f and encompasses (3). Approximate optimality conditions have been widely used in smooth optimization to establish the global convergence of different methods, especially the (safeguarded) AL method developed in [1]. See [3] and references therein.

Definition 2. *We say that a feasible x^* for* (P) *satisfies the* approximate KKT *(AKKT) condition, or that x^* is an* AKKT point*, if there are sequences $\{x^k\} \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$, $\{u^k\} \subset \mathbb{R}^n$, and $\{\varepsilon_k\} \downarrow 0$ such that, for all k, $x^k \neq x^*$,*

$$u^k \in \partial f(x^k) + \partial \Phi(x^k) + \sum_{i=1}^{m} \partial(\lambda_i^k h_i)(x^k) + \sum_{j=1}^{p} \mu_j^k \partial g_j(x^k), \qquad \|u^k\| \leq \varepsilon_k,$$

$\|(h(x^k), g(x^k)_+)\|_\infty \leq \varepsilon_k$, *and* $\mu_j^k = 0$ *if* $g_j(x^k) \leq -\varepsilon_k$, $j = 1, \ldots, p$.

The requirement $x^k \neq x^*$ is essential in our setting. Since $\partial\Phi(x^*) = \emptyset$, stationarity cannot be verified at x^* itself, and a constant AKKT sequence is not available. The analysis must therefore be done through nearby points at which $\partial\Phi$ is nonempty. Moreover, we impose the following assumption:

(H1) f, g, and h are continuously differentiable. Moreover, there exists $\delta > 0$ such that Φ is differentiable on $B(x^*, \delta) \setminus \{x^*\}$ and $\partial\Phi(x^*) = \emptyset$.

Under H1, we show that every local minimizer admits an AKKT sequence. By approaching x^* through distinct points, this sequence retains the first-order information of both f and Φ.

Theorem 3. *Let x^* be a local minimizer of* (P)*, and suppose that* (H1) *holds. Then, x^* is an AKKT point.*

Proof. Denote the feasible set of (P) by $\Omega := \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \leq 0\}$. As x^* is a local minimizer, there is $r \in (0, \delta)$ such that

$$(f + \Phi)(x^*) \leq (f + \Phi)(x) \quad \text{for all} \quad x \in \Omega \cap B[x^*, r]. \tag{4}$$

By H1, $\partial\Phi(x) \neq \emptyset$ for all $x \in B := B[x^*, r]$. Consider the penalized problem

$$\min_{x \in B} F_k(x) := (f + \Phi)(x) + \frac{\rho_k}{2}\|h(x)\|^2 + \frac{\rho_k}{2}\|g(x)_+\|^2 + \frac{1}{2}\|x - x^*\|^2. \tag{5}$$

As f, g, h are continuous by H1, and Φ is l.s.c. and proper, it follows that F_k is l.s.c. for all k and thus $\liminf_k F_k(x^k) > -\infty$. Then, as B is compact, (5) admits a global minimizer $x^k \in B$ by Proposition 3. Therefore, we have

$$(f + \Phi)(x^k) + \frac{\rho_k}{2}\|h(x^k)\|^2 + \frac{\rho_k}{2}\|g(x^k)_+\|^2 + \frac{1}{2}\|x^k - x^*\|^2 \leq (f + \Phi)(x^*) \tag{6}$$

for all k. Taking $\rho_k \to \infty$, we arrive at

$$\lim_k \|h(x^k)\| = 0 \quad \text{and} \quad \lim_k \|g(x^k)_+\| = 0. \tag{7}$$

As B is compact, $\{x^k\}$ admits a feasible accumulation point, let us say, $\bar{x} = \lim_{k \in K} x^k \in \Omega \cap B$. Taking the limit in (6) over K and using (4), we have $(f + \Phi)(x^*) \leq (f + \Phi)(\bar{x}) + \frac{1}{2}\|\bar{x} - x^*\|^2 \leq (f + \Phi)(x^*)$. Thus, $\bar{x} = x^*$ and we can suppose without loss of generality that $x^k \in B(x^*, r)$ for all $k \in K$.

We affirm that $x^k \neq x^*$ for all k. In fact, suppose that $x^k = x^*$. As $x^k \in B(x^*, r)$, (5) becomes a locally unconstrained problem, and then Propositions 2 and 4, together with $\partial\Phi(x^*) = \emptyset$, imply $0 \in \partial(f + \Phi)(x^*) \subset \partial\Phi(x^*) + \partial f(x^*) = \emptyset$, which is absurd. Thus, $x^k \neq x^*$ for all k. Consequently, Proposition 4 and the differentiability of F_k on $B(x^*, \delta)$ allow us to write

$$0 \in \nabla F_k(x^k) = \nabla f(x^k) + \nabla\Phi(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) + (x^k - x^*),$$

where $\lambda_i^k := \rho_k h_i(x^k)$, $i = 1, \ldots, m$, and $\mu_j^k := \rho_k g_j(x^k)_+$, $j = 1, \ldots, p$. Taking $\varepsilon^k := \max\{\|x^k - x^*\|_\infty, \|h(x^k)\|_\infty, \|g(x^k)_+\|_\infty\} \downarrow 0$, the statement follows by considering (7) and the subsequence $\{x^k\}_{k \in K}$. □

Remark 2. *For clarity, we keep ∂f and $\partial\Phi$ in Definition 2 separate, as in [7]. Although they could be combined as in (2), this makes no difference since, under H1, $\nabla(f+\Phi)(x) = \nabla f(x) + \nabla\Phi(x)$ at $x \neq x^*$. Also, we note that imposing $x^k \neq x^*$ excludes some simple problems beyond the one considered here. For example, our AKKT condition is not suitable for characterizing the minimizer of $f(x) = |x|$, since $\partial f(x) = \{\pm 1\}$ for $x \neq 0$. Establishing approximate stationarity conditions for such cases is a subject for future work.*

3.3 Application to an augmented Lagrangian method

The proof of Theorem 3 uses the external penalty method. Since AL methods are similar [3], we investigate whether they also generate AKKT points.

We consider the AL scheme proposed in [1], whose robustness and theoretical properties are supported by an extensive literature; see, e.g., [3, 4]. It is essentially the same method described in [5], where subproblems are adapted to encompass the composite structure of (P). More precisely, we employ the stationarity criterion considered in [5, Algorithm 3.1] (expression (8)).

Algorithm 1 Safeguarded augmented Lagrangian method

The parameters are $\tau \in [0,1)$, $\gamma > 1$, $-\infty < \lambda_{\min} \leq \lambda_{\max} < \infty$, $0 \leq \mu_{\max} < \infty$, and $\rho_1 > 0$. Let $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^1 \in [0, \mu_{\max}]^p$, and a sequence $\varepsilon_k \downarrow 0$. Initialize $k \leftarrow 1$.

Step 1 (Solving the subproblem): Compute an *ε_k-approximate stationary point* x^k of $\min_x L(x, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)$, that is, x^k satisfying

$$\operatorname{dist}\left(0, \nabla_x \varphi(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) + \partial\Phi(x^k)\right) \leq \varepsilon_k. \tag{8}$$

Step 2 (Updating the penalty parameter): Define $V^k := \min\{-g(x^k), \bar{\mu}^k/\rho_k\}$. If $k = 1$ or

$$\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\},$$

set $\rho_{k+1} := \rho_k$. Otherwise, take $\rho_{k+1} \geq \gamma\rho_k$.

Step 3 (New estimated multipliers): Choose $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^p$.

Step 4: Set $k \leftarrow k+1$ and go to **Step 1**.

A point x^k satisfying (8) can be computed by the nonmonotone proximal-gradient method described in [5, Algorithm 3.2]. In fact, the authors proved its finite termination depending solely on the Lipschitz constant of φ [5, Theorem 3.4]. Note that if Φ is differentiable, (8) reduces to $\|\nabla_x \varphi(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) + \nabla\Phi(x^k)\| \leq \varepsilon_k$, recovering the usual criterion employed in [1, Algorithm 3.1].

We now examine the behavior of Algorithm 1 under the condition $\partial\Phi(x^*) = \emptyset$. In what follows, we show that such a point cannot itself be an iterate of the method. Consequently, a sequence generated by the method approaches x^* through distinct points, as required by Definition 2.

Theorem 4. *Suppose that $\partial\Phi(x^*) = \emptyset$. Then, for every $\varepsilon > 0$, $\rho > 0$, $\bar{\mu} \geq 0$, and $\bar{\lambda}$, the point x^* is not ε-approximate stationary in the sense of (8).*

Proof. Since $\partial\Phi(x^*) = \emptyset$, we have $\nabla_x \varphi(x^*, \bar{\lambda}, \bar{\mu}, \rho) + \partial\Phi(x^*) = \emptyset$, and thus condition (8) cannot hold at x^* for any $\varepsilon_k > 0$. □

The next result shows that, under convergence of the function values, the AL sequence recovers the horizon-based optimality condition (2) at every feasible accumulation point x^* satisfying $\partial\Phi(x^*) = \emptyset$. Its proof is similar to those of [6, Proposition 2.5, Theorem 3.3]. However, the authors impose additional assumptions on the problem data and, in the case $\partial\Phi(x^*) = \emptyset$, do not address the essential condition $x^k \neq x^*$ established in Theorem 4. Furthermore, they prove convergence to a point satisfying only a sequential condition, whereas we prove convergence to a point satisfying a pointwise stationarity condition without any CQ. For these reasons, we provide a complete proof.

Theorem 5. *Let $\{x^k\}$ be a sequence generated by Algorithm 1, and let $x^* = \lim_{k\in K} x^k$ be a feasible accumulation point of it with $\partial\Phi(x^*) = \emptyset$. Suppose that $\lim_{k\in K} \Phi(x^k) = \Phi(x^*)$ and that H1 holds. Then (2) holds for some $(\lambda, \mu) \neq 0$.*

Proof. By Step 1 of Algorithm 1 there exist u^k and $v^k \in \partial\Phi(x^k)$ such that

$$u^k = \nabla f(x^k) + v^k + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k), \tag{9}$$

where $\lim_k \|u^k\| = 0$, $\lambda^k := \bar{\lambda}^k + \rho_k h(x^k)$, and $\mu^k := [\bar{\mu}^k + \rho_k g(x^k)]_+$ for all k.

For each $k \in K$, since $v^k \in \partial\Phi(x^k)$, the definition of $\partial\Phi(x^k)$ allows us to choose y^k and $w^k \in \widehat{\partial}\Phi(y^k)$ such that $\|y^k - x^k\| < 1/k$, $|\Phi(y^k) - \Phi(x^k)| < 1/k$, and $\|w^k - v^k\| < 1/k$ (see Definition 1). Thus, $y^k \to_\Phi x^*$ along K. If $\{v^k\}_{k\in K}$ had a bounded subsequence, passing to a subsequence would yield $v^k \to v$ and $w^k \to v$. Again, the definition of $\partial\Phi(x^k)$ would then give $v \in \partial\Phi(x^*) = \emptyset$, a contradiction. Therefore, $\lim_{k\in K} \|v^k\| = \infty$. On the other hand, (9) gives

$$\|v^k\| \le \|u^k\| + \|\nabla f(x^k)\| + \sum_{i=1}^{m} |\lambda_i^k| \, \|\nabla h_i(x^k)\| + \sum_{j=1}^{p} \mu_j^k \, \|\nabla g_j(x^k)\|.$$

Since the gradients of f, h, and g are bounded near x^* by H1, it follows that $\lim_{k\in K} \|(\lambda^k, \mu^k)\|_\infty = \infty$.

For each $k \in K$, define $M_k := \|(1, \lambda^k, \mu^k)\|_\infty$. Then $\lim_{k\in K} M_k = \infty$ and, after passing to a subsequence, $\lim_{k\in K} (\lambda^k, \mu^k)/M_k = (\lambda, \mu)$ with $\|(\lambda, \mu)\|_\infty = 1$ and $\mu \ge 0$. Dividing (9) by M_k and using the continuity of the gradients, we obtain

$$\lim_{k\in K} \frac{v^k}{M_k} = v^* := -\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) - \sum_{j=1}^{p} \mu_j \nabla g_j(x^*).$$

Since $\|w^k - v^k\| < 1/k$, we also have $\lim_{k\in K} w^k/M_k = v^*$. Together with $y^k \to_\Phi x^*$ and $w^k \in \widehat{\partial}\Phi(y^k)$, this implies $v^* \in \partial^\infty\Phi(x^*)$ by taking $t_k := 1/M_k \downarrow 0$ in Definition 1. If $g_j(x^*) < 0$, then $g_j(x^k) < 0$ for all sufficiently large $k \in K$, and hence $0 \le \mu_j^k = [\bar{\mu}_j^k + \rho_k g_j(x^k)]_+ \le \bar{\mu}_j^k \le \mu_{\max}$. Dividing by $M_k \to \infty$ yields $\mu_j = 0$. Consequently, x^* satisfies (2) with $(\lambda, \mu) \ne 0$. □

The assumption $\Phi(x^k) \to \Phi(x^*)$ cannot be omitted from the above theorem. This has been shown for the case where $\partial\Phi(x^*) \ne \emptyset$ in [6, Example 3.4]. This example also shows that [5, Theorem 3.1] lacks this hypothesis. Next, we adapt the cited example to show that the same holds when $\partial\Phi(x^*) = \emptyset$.

Example 2. *Consider the problem* (P) *in one variable, where*

$$f(x) := x, \quad g(x) := x, \quad \text{and} \quad \Phi(x) := \begin{cases} x^{1/3}, & x \le 0 \\ 1 - x, & x > 0 \end{cases}.$$

The function Φ is proper, l.s.c., and differentiable on $\mathbb{R} \setminus \{0\}$. Moreover, similar to Example 1, we can conclude that $\partial\Phi(0) = \emptyset$ and $\partial^\infty\Phi(0) = [0, \infty)$.

The point $x^* = 0$ is feasible and can be achieved by Algorithm 1. In fact, choosing $\gamma > 1$ and $\tau \in (0, 1/\gamma] \subset [0, 1)$, we set

$$\{x^k := \tau^k\} \qquad \text{and} \qquad \{\bar{\mu}^k := 0\}.$$

Supposing that in Step 2 of Algorithm 1 we choose $\rho_{k+1} = \gamma\rho_k$ when ρ must grow, we have $\rho_k \le \gamma\rho_{k-1} \le \cdots \le \gamma^{k-1}\rho_1 \le \gamma^k\rho_1$, and thus $\rho_k\tau^k \le (\gamma\tau)^k\rho_1 \downarrow 0$. Therefore, we can define $\varepsilon_k := (\gamma\tau)^k\rho_1$ for all k. Since $x^k > 0$, we have $f'(x^k) = 1$, $\partial\Phi(x^k) = \{-1\}$, and $g'(x^k) = 1$. Consequently, $\operatorname{dist}(0, \rho_k\tau^k) = \rho_k\tau^k \le \varepsilon_k$, *and therefore* (8) *holds. Moreover, $V^k = \min\{-\tau^k, 0\} = -\tau^k$ for all k, and hence $|V^{k+1}| = \tau|V^k|$ for all k. This shows that, actually, $\rho_k = \rho_1$ for all k. Thus, x^* may be reached by Algorithm 1.*

However, (2) *does not hold at x^*: as $g'(0) = 1$, it requires $0 \in \partial^\infty\Phi(0) + \mu g'(0) = [0, \infty) + \mu$ for some $\mu > 0$, which is impossible. Note that $\{\Phi(x^k)\}$ does not accumulate at $\Phi(x^*)$ since $\Phi(x^k) = 1 - \tau^k \to 1 \ne 0 = \Phi(0)$.*

4 Conclusions

We studied the constrained minimization of a composite function $f + \Phi$ at local minimizers x^* for which the limiting subdifferential $\partial\Phi(x^*) = \emptyset$. We showed that some constraint must be active at x^*, and that the horizon-based Fritz–John condition [7] holds without any CQ. Under differentiability of Φ near x^*, we established an approximate KKT condition that retains the first-order contribution of the Lipschitzian term f. Finally, for an augmented Lagrangian method, we showed that every feasible accumulation point x^* satisfying $\partial\Phi(x^*) = \emptyset$ satisfies the horizon-based condition, provided that the corresponding values of Φ converge to $\Phi(x^*)$. We exhibit an example showing that this function-value convergence assumption is essential for this conclusion.

AI assistance AI was used to assist with language editing and the preparation of examples. The authors verified all content and remain fully responsible for it.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2007.

[2] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011.

[3] R. Andreani, G. Haeser, L. M. Mito, A. Ramos, and L. D. Secchin. On the best achievable quality of limit points of augmented lagrangian schemes. *Numerical Algorithms*, 90(2):851–877, 2022.

[4] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[5] X. Chen, L. Guo, Z. Lu, and J. J. Ye. An augmented Lagrangian method for non-Lipschitz nonconvex programming. *SIAM Journal on Numerical Analysis*, 55(1):168–193, 2017.

[6] A. De Marchi, X. Jia, C. Kanzow, and P. Mehlitz. Constrained composite optimization and augmented Lagrangian methods. *Mathematical Programming*, 201(1):863–896, 2023.

[7] L. Guo and J. J. Ye. Necessary optimality conditions and exact penalization for non-Lipschitz nonlinear programs. *Mathematical Programming*, 168(1):571–598, 2018.

[8] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation, vol 1. Basic Theory*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag Berlin Heidelberg, New York, 2006.

[9] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1 edition, 1998.

[10] J. J. Ye and J. Zhang. Enhanced Karush-Kuhn-Tucker condition and weaker constraint qualifications. *Mathematical Programming*, 139(1):353–381, 2013.