


Article

# First-Order Conditions for Set-Constrained Optimization

Steven M. Rovnyak <sup>1,\*</sup>, Edwin K. P. Chong <sup>2</sup>  and James Rovnyak <sup>3</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, Indiana University–Purdue University, Indianapolis, IN 46202, USA

<sup>2</sup> Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA; edwin.chong@colostate.edu

<sup>3</sup> Department of Mathematics, University of Virginia, Charlottesville, VA 22904, USA; rovnyak@virginia.edu

\* Correspondence: srovnyak@iupui.edu

**Abstract:** A well-known first-order necessary condition for a point to be a local minimizer of a given function is the non-negativity of the dot product of the gradient and a vector in a feasible direction. This paper proposes a series of alternative first-order necessary conditions and corresponding first-order sufficient conditions that seem not to appear in standard texts. The conditions assume a nonzero gradient. The methods use extensions of the notions of gradient, differentiability, and twice differentiability. Examples, including one involving the Karush–Kuhn–Tucker (KKT) theorem, illustrate the scope of the conditions.

**Keywords:** constrained optimization; local minimizer; necessary condition; sufficient condition; KKT theorem

**MSC:** 49K99; 90C46



**Citation:** Rovnyak, S.M.; Chong, E.K.P.; Rovnyak, J. First-Order Conditions for Set-Constrained Optimization. *Mathematics* **2023**, *11*, 4274. <https://doi.org/10.3390/math11204274>

Academic Editors: Chao Zhang, Yanfang Zhang, Yang Zhou and Qiang Ye

Received: 15 September 2023

Revised: 3 October 2023

Accepted: 10 October 2023

Published: 13 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Set-constrained and set-unconstrained optimization use several theorems that include a first-order necessary condition, as well as second-order conditions that are necessary and/or sufficient for a point to be a local minimizer [1,2]. These theorems require the objective function to be once or twice continuously differentiable.

The first-order necessary condition for unconstrained optimization requires the gradient to be zero at a minimizer [3–5]. The first-order necessary condition for set-constrained optimization requires the dot product of the gradient and a vector in a feasible direction to be non-negative. When constraints are defined in terms of differentiable functions, the first-order necessary condition takes the form of the first-order Lagrange and Karush–Kuhn–Tucker (KKT) conditions [2,6]. Reference [5] determined the solution to linear programming problems using a first-order necessary condition. Such conditions have been studied for control systems governed by ordinary differential equations [7], stochastic differential equations [8], and stochastic evolution equations [9].

Throughout, we assume a nonzero gradient. Our main results present a series of four sufficient conditions and four corresponding necessary conditions for a point to be a local minimizer of a given function  $f$  (Theorems 1–8). Each is of the first-order type, including those that assume twice differentiability. Theorems 1 and 2 describe the behavior of  $f$  in cones determined by the gradient. Theorems 3 and 4 replace the geometrical viewpoint by analytical conditions involving sequences. The analytical versions use generalizations of the notions of gradient and differentiability that simplify statements and proofs. Theorems 5 and 6 are refinements when  $f$  is twice differentiable. A previous version of this paper included a remark that the analytical conditions are unverifiable. However, the last two results, Theorems 7 and 8, which return to the geometrical view, are proved precisely by verifying the analytical conditions. They replace the original cones

by larger regions that we call  $\alpha$ -cones. An  $\alpha$ -cone is an ordinary cone when  $\alpha = 1$  and a paraboloid when  $\alpha = 2$ . The results fail for half-planes, and a paraboloid is a limiting case for what is possible. Example 5 illustrates a class of problems that do not meet criteria for a strict local minimizer in the KKT theory but are covered by Theorem 7.

We remark that the cones used here are different from the cones of descent directions in [2], which are actually half-planes. Our sufficient conditions do not guarantee that a point with a nonzero gradient is a strict local minimizer on a half-plane. Example 1 gives a function with a nonzero gradient that is not a strict local minimizer on a half-plane.

Convex optimization problems require the objective function to be convex, which essentially makes them second-order conditions. The first-order conditions that we propose do not require the objective function to be convex. The requirement for a function to be twice continuously differentiable is different from the condition of convexity, which constrains the values of the second derivatives.

**Notation and terminology.** Points of  $\mathbb{R}^n$  are written as column vectors  $\mathbf{x} = [x_1, \dots, x_n]^T$  with norm  $\|\mathbf{x}\| = \sqrt{|x_1|^2 + \dots + |x_n|^2}$ . A subset of  $\mathbb{R}^n$  is called a **neighborhood** of a point  $\mathbf{x}^*$  if it contains a disk  $\|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$  for some  $\varepsilon > 0$ , and a **neighborhood** of a set  $\Omega \subseteq \mathbb{R}^n$  if it is a neighborhood of every point in  $\Omega$ . A point  $\mathbf{x}^* \in \Omega$  is a **strict local minimizer** of a function  $f: \Omega \rightarrow \mathbb{R}$  if there is an  $\varepsilon > 0$  such that  $f(\mathbf{x}) > f(\mathbf{x}^*)$  whenever  $\mathbf{x} \in \Omega$  and  $0 < \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$ , and a **local minimizer** if the condition holds with  $>$  replaced by  $\geq$ . **Local maximizers** and **strict local maximizers** are defined similarly by reversing the inequalities.

## 2. First-Order Conditions for Local Minimizers

The gradient of a real-valued function  $f$  defined on a subset  $\Omega$  of  $\mathbb{R}^n$  is defined by

$$\nabla f(\mathbf{x}) = \left[ \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right]^T,$$

whenever the partial derivatives exist. Gradients appear in a standard first-order necessary condition for local minimizers. Consider a function  $f: \Omega \rightarrow \mathbb{R}$  that is  $C^1$  on a neighborhood of a set  $\Omega \subseteq \mathbb{R}^n$  and a point  $\mathbf{x}^* \in \Omega$ . If  $\mathbf{x}^*$  is a local minimizer of  $f$ , then  $\mathbf{d}^T \nabla f(\mathbf{x}^*) \geq 0$  for every vector  $\mathbf{d}$  in a **feasible direction**, that is, a direction such that some straight line segment with endpoint  $\mathbf{x}^*$  lies entirely within  $\Omega$  [1] (Theorem 6.1). Hence, if  $\mathbf{d}^T \nabla f(\mathbf{x}^*) < 0$  for some feasible direction  $\mathbf{d}$  at  $\mathbf{x}^*$ , the standard necessary condition implies that  $\mathbf{x}^*$  is not a local minimizer. However, it may occur that  $\mathbf{x}^*$  has no feasible direction within  $\Omega$ , and then it is impossible for the standard necessary condition to give such information. For example, feasible directions are impossible for any set  $\Omega$  whose points have only rational coordinates. For an elementary example, consider the objective function  $f(\mathbf{x}) = -x_1$  on the set

$$\Omega = \{ \mathbf{x} \in \mathbb{R}^2 : 0 \leq x_1 \leq 1, x_1^3 \leq x_2 \leq x_1^2 \}. \tag{1}$$

Then,  $f$  attains a maximum value on  $\Omega$  at  $\mathbf{x}^* = \mathbf{0}$ . The point  $\mathbf{x}^*$  admits no feasible direction within  $\Omega$  (proof: any feasible direction must be along a line segment  $x_2 = cx_1$ ,  $c > 0$ ; then, for all sufficiently small positive  $x_1$ ,  $x_1^3 \leq cx_1 \leq x_1^2$ , and hence  $x_1^2 \leq c \leq x_1$ , which is impossible), and thus the standard necessary condition yields no information.

Feasible directions play no role in our results, and instead all that is needed is that  $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$ . Theorem 2 (FONC 1) is a new first-order necessary condition for local minimizers. It is proved using a corresponding new first-order sufficient condition, Theorem 1 (FOSC 1). Corollary 1 is a companion sufficiency condition for strict local maximizers. In the example  $f(\mathbf{x}) = -x_1$  on the set  $\Omega$  defined by (1) and  $\mathbf{x}^* = \mathbf{0}$ , the gradient is  $\nabla f(\mathbf{0}) = [-1, 0]^T$ , and therefore:

- (1) Theorem 2 is applicable and implies that  $\mathbf{0}$  is not a strict local minimizer, because every opposite cone  $|x_2| \leq cx_1$  contains points arbitrarily near  $\mathbf{0}$ .
- (2) Corollary 1 is applicable and implies that  $\mathbf{0}$  is a strict local maximizer, because  $\Omega$  is entirely contained in the opposite cone  $|x_2| \leq x_1$ .

We need a few more preliminaries. A real-valued function  $f$  on a subset  $\Omega$  of  $\mathbb{R}^n$  is said to be **differentiable** at  $\mathbf{x}^*$  if the domain of  $f$  contains a neighborhood of  $\mathbf{x}^*$ , all partial derivatives of  $f$  exist at  $\mathbf{x}^*$ , and the function  $r(\mathbf{x})$  defined by

$$f(\mathbf{x}) - f(\mathbf{x}^*) = (\mathbf{x} - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) + r(\mathbf{x}) \tag{2}$$

satisfies

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}^*} \frac{r(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}^*\|} = 0. \tag{3}$$

When  $n = 1$ , this is equivalent to the existence of a derivative at the point, but in general the simple existence of partial derivatives does not imply differentiability. A convenient sufficient condition for differentiability is that  $f$  is defined and  $C^1$  on some neighborhood of  $\mathbf{x}^*$  [10] (Th. 9 on p. 113). Differentiability is equivalent to the existence of a first-order Taylor approximation, as in [1] (Th. 5.5 on pp. 64–65) or [10] (Th. 2 on p. 160).

Any two nonzero vectors  $\mathbf{d}_1, \mathbf{d}_2$  in  $\mathbb{R}^n$  determine an acute angle  $\theta$  such that

$$\cos(\theta) = \frac{\mathbf{d}_1^\top \mathbf{d}_2}{\|\mathbf{d}_1\| \|\mathbf{d}_2\|}.$$

This notion is implicit in the definition of a cone in  $\mathbb{R}^n$ .

**Definition 1.** If  $\mathbf{x}^*, \mathbf{d} \in \mathbb{R}^n, \mathbf{d} \neq 0$ , and  $0 < \delta < 1$ , the set consisting of  $\mathbf{x}^*$  together with all points  $\mathbf{x} \neq \mathbf{x}^*$  in  $\mathbb{R}^n$  that satisfy

$$\frac{(\mathbf{x} - \mathbf{x}^*)^\top \mathbf{d}}{\|\mathbf{x} - \mathbf{x}^*\| \|\mathbf{d}\|} \geq \delta$$

is denoted by  $K_\delta(\mathbf{x}^*, \mathbf{d})$  and called a **cone** with vertex  $\mathbf{x}^*$  and direction  $\mathbf{d}$ . The **opposite cone** is defined by

$$-K_\delta(\mathbf{x}^*, \mathbf{d}) = K_\delta(\mathbf{x}^*, -\mathbf{d}).$$

Our first result is a first-order sufficient condition (FOSC) for a local minimizer.

**Theorem 1 (FOSC 1).** Let  $f$  be a real-valued function that is defined and  $C^1$  on a neighborhood of a point  $\mathbf{x}^* \in \mathbb{R}^n$  such that  $\mathbf{d}^* = \nabla f(\mathbf{x}^*) \neq 0$ . Assume that  $\Omega$  is a set in the domain of  $f$  that contains  $\mathbf{x}^*$  and is contained in some cone  $K_\delta(\mathbf{x}^*, \mathbf{d}^*)$ ,  $\delta \in (0, 1)$ . Then,  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on  $\Omega$ .

We note a consequence that will be useful in Theorem 2.

**Corollary 1.** For  $f$  as in Theorem 1,  $\mathbf{x}^*$  is a strict local maximizer of  $f$  on any set in the domain of  $f$  that contains  $\mathbf{x}^*$  and is contained in  $-K_\delta(\mathbf{x}^*, \mathbf{d}^*)$  for some  $\delta \in (0, 1)$ .

The corollary follows by applying Theorem 1 with  $f$  replaced by  $-f$ .

**Proof of Theorem 1.** Let  $\Omega$  be a set in the domain of  $f$  that contains  $\mathbf{x}^*$  and is contained in  $K_\delta(\mathbf{x}^*, \mathbf{d}^*)$ ,  $\delta \in (0, 1)$ . Since  $f$  is  $C^1$  on a neighborhood of  $\mathbf{x}^*$ ,  $f$  is differentiable at  $\mathbf{x}^*$ . Thus,

$$f(\mathbf{x}) - f(\mathbf{x}^*) = (\mathbf{x} - \mathbf{x}^*)^\top \mathbf{d}^* + r(\mathbf{x}), \tag{4}$$

where  $\lim_{\mathbf{x} \rightarrow \mathbf{x}^*} r(\mathbf{x}) / \|\mathbf{x} - \mathbf{x}^*\| = 0$ . Therefore, we may choose  $\eta > 0$  so that the punctured disk  $0 < \|\mathbf{x} - \mathbf{x}^*\| < \eta$  is contained in the domain of  $f$  and

$$\frac{|r(\mathbf{x})|}{\|\mathbf{x} - \mathbf{x}^*\|} < \delta \|\mathbf{d}^*\| \tag{5}$$

whenever  $0 < \|\mathbf{x} - \mathbf{x}^*\| < \eta$ . Suppose  $x \in \Omega$  and  $0 < \|\mathbf{x} - \mathbf{x}^*\| < \eta$ . Then,  $\mathbf{x} \in K_\delta(\mathbf{x}^*, \mathbf{d}^*)$ , so  $(\mathbf{x} - \mathbf{x}^*)^\top \mathbf{d}^* \geq \|\mathbf{x} - \mathbf{x}^*\| \|\mathbf{d}^*\|$ . By (4),

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \delta \|\mathbf{x} - \mathbf{x}^*\| \|\mathbf{d}^*\| + r(\mathbf{x}). \tag{6}$$

Since  $0 < \|\mathbf{x} - \mathbf{x}^*\| < \eta$ , by (5),  $|r(\mathbf{x})| < \delta \|\mathbf{d}^*\| \|\mathbf{x} - \mathbf{x}^*\|$ , and hence by (6),

$$f(\mathbf{x}) - f(\mathbf{x}^*) > 0.$$

Therefore,  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on  $\Omega$ .  $\square$

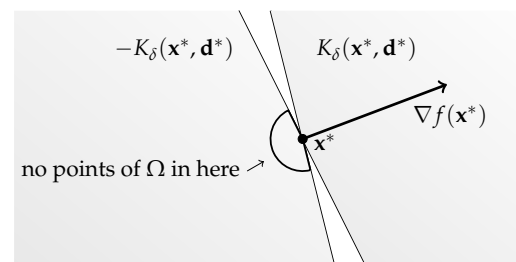
A corresponding first-order necessary condition (FONC) is deduced with the aid of Corollary 1. We first introduce some useful terminology.

**Definition 2.** A point  $\mathbf{x}^*$  of  $\Omega$  is called *isolated* if there exists an  $\varepsilon > 0$  such that the punctured disk  $0 < \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$  contains no points of  $\Omega$ , or, equivalently, there is no sequence  $\{\mathbf{x}_n\}_{n=1}^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  such that  $\mathbf{x}_n \rightarrow \mathbf{x}^*$ .

In the extreme case that a set consists of just one point, that point is isolated in the set because the alternative would imply the existence of other points in the set. Isolated points occur in our setting in the next result, which is illustrated in Figure 1.

**Theorem 2 (FONC 1).** Let  $f$  be a real-valued function that is defined and  $C^1$  on a neighborhood of a point  $\mathbf{x}^* \in \mathbb{R}^n$  such that  $\mathbf{d}^* = \nabla f(\mathbf{x}^*) \neq 0$ . If  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on some subset  $\Omega$  of  $\mathbb{R}^n$ , then  $\mathbf{x}^*$  is an isolated point of  $-K_\delta(\mathbf{x}^*, \mathbf{d}^*) \cap \Omega$  for every  $\delta \in (0, 1)$ .

**Proof.** Suppose  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on some subset  $\Omega$  of  $\mathbb{R}^n$ . We argue by contradiction to show that  $\mathbf{x}^*$  is an isolated point of  $-K_\delta(\mathbf{x}^*, \mathbf{d}^*) \cap \Omega$  for every  $\delta \in (0, 1)$ . Assume that this conclusion is false for some  $\delta \in (0, 1)$ . Then, the cone  $K_\delta(\mathbf{x}^*, -\mathbf{d}^*)$  contains a sequence  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots$  in  $\Omega \setminus \{\mathbf{x}^*\}$  that converges to  $\mathbf{x}^*$ . Since  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on  $\Omega$ , there exists  $\varepsilon > 0$  such that  $f(\mathbf{x}) > f(\mathbf{x}^*)$  whenever  $0 < \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$  and  $\mathbf{x} \in \Omega$ . By the definition of convergence,  $\|\mathbf{x}_n - \mathbf{x}^*\| < \varepsilon$  for all sufficiently large  $n$ , and hence  $f(\mathbf{x}_n) > f(\mathbf{x}^*)$  for all sufficiently large  $n$ . However, by Corollary 1,  $\mathbf{x}^*$  is a strict local maximizer for  $f$  on  $K_\delta(\mathbf{x}^*, -\mathbf{d}^*) \cap \Omega$ , so  $f(\mathbf{x}_n) < f(\mathbf{x}^*)$  for a sufficiently large  $n$ , which is a contradiction. The result follows.  $\square$



**Figure 1.** This figure illustrates Theorem 2 (FONC 1). If  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on some subset  $\Omega$  of  $\mathbb{R}^n$ , then  $\mathbf{x}^*$  is an isolated point of  $-K_\delta(\mathbf{x}^*, \mathbf{d}^*) \cap \Omega$  for every  $\delta \in (0, 1)$ .

### 3. Analytical Versions of the Conditions and Generalized Differentiability

In this section, we present analytical versions of the conditions for local minimizers given in Section 2. They are stated in a more general setting that uses extensions of the notions of differentiability and gradient to arbitrary sets. The analytical versions are in some ways more transparent and lead to generalizations of Theorems 1 and 2 in Section 5.

**Definition 3** (Generalized Differentiability). Let  $\Omega$  be a subset of  $\mathbb{R}^n$ . We say that a function  $f: \Omega \rightarrow \mathbb{R}$  is **differentiable** at a point  $\mathbf{x}^* \in \Omega$  if (1)  $\mathbf{x}^*$  is not an isolated point of  $\Omega$ , and (2) there is a vector  $\mathbf{g}(\mathbf{x}^*)$  in  $\mathbb{R}^n$  such that the function  $r(\mathbf{x})$  defined by

$$f(\mathbf{x}) - f(\mathbf{x}^*) = (\mathbf{x} - \mathbf{x}^*)^\top \mathbf{g}(\mathbf{x}^*) + r(\mathbf{x}) \tag{7}$$

satisfies

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}^*} \frac{r(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}^*\|} = 0, \tag{8}$$

or, equivalently, for every sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  converging to  $\mathbf{x}^*$ , the sequence  $\{r_n\}_1^\infty$  defined by

$$f(\mathbf{x}_n) - f(\mathbf{x}^*) = (\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{g}(\mathbf{x}^*) + r_n \tag{9}$$

satisfies

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}^*} \frac{r_n}{\|\mathbf{x}_n - \mathbf{x}^*\|} = 0. \tag{10}$$

Any such vector  $\mathbf{g}(\mathbf{x}^*)$  is denoted by  $\nabla f(\mathbf{x}^*)$  and called a **gradient** of  $f$  at  $\mathbf{x}^*$ .

By [10] (Th. 9 on p. 113), the condition (8) is automatically met if  $f$  has an extension to a function  $\tilde{f}$  which is  $C^1$  on a neighborhood of  $\mathbf{x}^*$ , and then we can choose  $\mathbf{g}(\mathbf{x}^*) = \nabla \tilde{f}(\mathbf{x}^*)$ . In general, gradients are not unique, but for our purpose any choice will work.

For a fixed  $f$  and  $\mathbf{x}^*$ , the set of all gradients  $\mathbf{g}(\mathbf{x}^*)$  is a closed convex set. We shall not need this fact and omit a proof.

**Theorem 3** (FOSC 2). Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f: \Omega \rightarrow \mathbb{R}$  be differentiable at some point  $\mathbf{x}^* \in \Omega$  with gradient  $\nabla f(\mathbf{x}^*) \neq 0$ . Assume that for every sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  with  $\mathbf{x}_n \rightarrow \mathbf{x}^*$  there is a  $\delta > 0$  such that

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) \geq \delta \|\mathbf{x}_n - \mathbf{x}^*\| \tag{11}$$

for all sufficiently large  $n$ . Then,  $\mathbf{x}^*$  is a strict local minimizer of  $f$  over  $\Omega$ .

**Proof.** Argue by contradiction. If  $\mathbf{x}^*$  is not a strict local minimizer of  $f$ , there is a sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  with  $\mathbf{x}_n \rightarrow \mathbf{x}^*$  such that  $f(\mathbf{x}_n) \leq f(\mathbf{x}^*)$  for all  $n$ . Define  $\{r_n\}_1^\infty$  by (9). Then, by (11),

$$f(\mathbf{x}_n) - f(\mathbf{x}^*) = (\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) + r_n \geq \delta \|\mathbf{x}_n - \mathbf{x}^*\| + r_n \tag{12}$$

for all sufficiently large  $n$ . Since  $f$  is differentiable at  $\mathbf{x}^*$ ,  $r_n / \|\mathbf{x}_n - \mathbf{x}^*\| \rightarrow 0$  as  $n \rightarrow \infty$  by (10). Dividing by  $\|\mathbf{x}_n - \mathbf{x}^*\|$ , we obtain

$$\frac{f(\mathbf{x}_n) - f(\mathbf{x}^*)}{\|\mathbf{x}_n - \mathbf{x}^*\|} \geq \delta + \frac{r_n}{\|\mathbf{x}_n - \mathbf{x}^*\|} > 0 \tag{13}$$

for all sufficiently large  $n$ . The result that  $f(\mathbf{x}_n) > f(\mathbf{x}^*)$  for all sufficiently large  $n$  contradicts our choice of the sequence  $\{\mathbf{x}_n\}_1^\infty$  to satisfy  $f(\mathbf{x}_n) \leq f(\mathbf{x}^*)$  for all  $n$ . The theorem follows.  $\square$

The corresponding necessary condition is conveniently stated in contrapositive form.

**Theorem 4** (FONC 2). Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f: \Omega \rightarrow \mathbb{R}$  be differentiable at some point  $\mathbf{x}^* \in \Omega$  with gradient  $\nabla f(\mathbf{x}^*) \neq 0$ . If there exists a sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  converging to  $\mathbf{x}^*$  and number  $\delta > 0$  such that

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) \leq -\delta \|\mathbf{x}_n - \mathbf{x}^*\| \tag{14}$$

for all sufficiently large  $n$ , then  $\mathbf{x}^*$  is not a local minimizer of  $f$  on  $\Omega$ .

Equivalently, if  $\mathbf{x}^*$  is a local minimizer of  $f$  over  $\Omega$ , then there exists no such sequence  $\{\mathbf{x}_n\}_1^\infty$  and number  $\delta > 0$ .

**Proof.** Assume we are given such a sequence  $\{\mathbf{x}_n\}_1^\infty$  and  $\delta > 0$ . Define  $\{r_n\}_1^\infty$  by (9). Then, by (14),

$$f(\mathbf{x}_n) - f(\mathbf{x}^*) = (\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) + r_n \leq -\delta \|\mathbf{x}_n - \mathbf{x}^*\| + r_n \tag{15}$$

for all sufficiently large  $n$ . Since  $f$  is differentiable at  $\mathbf{x}^*$ ,  $r_n / \|\mathbf{x}_n - \mathbf{x}^*\| \rightarrow 0$  as  $n \rightarrow \infty$  by (10). Thus, dividing (15) by  $\|\mathbf{x}_n - \mathbf{x}^*\|$ , we see that

$$\frac{f(\mathbf{x}_n) - f(\mathbf{x}^*)}{\|\mathbf{x}_n - \mathbf{x}^*\|} \leq -\delta + \frac{r_n}{\|\mathbf{x}_n - \mathbf{x}^*\|} < 0 \tag{16}$$

for all sufficiently large  $n$ . Hence,  $f(\mathbf{x}_n) < f(\mathbf{x}^*)$  for all sufficiently large  $n$ . Since  $\mathbf{x}_n^* \rightarrow \mathbf{x}^*$ ,  $\mathbf{x}^*$  is not a strict local minimizer.  $\square$

**Example 1.** The inequality (11) in Theorem 3 (FOSC 2) cannot be weakened to

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) > 0. \tag{17}$$

Choose  $\mathbf{x}^* = \mathbf{0}$ ,  $\Omega = \{\mathbf{0}\} \cup \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 0\}$ , and  $f(\mathbf{x}) = x_1 - x_2^2$ ,  $\mathbf{x} \in \Omega$ . Then, we can take  $\nabla f(\mathbf{x}^*) = [1, 0]^\top$ , and any sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{0}\}$  with  $\mathbf{x}_n \rightarrow \mathbf{0}$  satisfies (17). However,  $\mathbf{0}$  is not a local minimizer because  $f(\mathbf{0}) = 0$ , and the sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{0}\}$  defined by  $\mathbf{x}_n = [1/(2n^2), 1/n]$ ,  $n \geq 1$ , converges to  $\mathbf{0}$  and satisfies  $f(\mathbf{x}_n) < 0$  for all  $n \geq 1$ .

**Example 2.** We cannot relax the inequality (14) in Theorem 4 (FONC 2) to

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) < 0. \tag{18}$$

Choose  $\mathbf{x}^* = \mathbf{0}$ ,  $\Omega = \{\mathbf{x} \in \mathbb{R}^2 : 2x_1 \geq -x_2^2\}$ ,  $f(\mathbf{x}) = x_1 + x_2^2$ ,  $\mathbf{x} \in \Omega$ , and  $\nabla f(\mathbf{x}^*) = [1, 0]^\top$ . See Figure 2. The sequence  $\mathbf{x}_n = [-1/(2n^2), 1/n]$ ,  $n \geq 1$ , belongs to  $\Omega \setminus \{\mathbf{0}\}$  and converges to  $\mathbf{0}$ . It satisfies (18) because

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) = [-1/(2n^2), 1/n] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -\frac{1}{2n^2} < 0, \quad n \geq 1.$$

Nevertheless,  $f(\mathbf{0}) = 0$  is the minimum value of  $f$  attained on  $\Omega$ . For, if  $\mathbf{x} \in \Omega \setminus \{\mathbf{0}\}$  and  $x_2 \neq 0$ , then  $f(\mathbf{x}) = x_1 + x_2^2 \geq -\frac{1}{2}x_2^2 + x_2^2 > 0$ , and  $f(\mathbf{x}) = x_1 > 0$  whenever  $\mathbf{x} \in \Omega \setminus \{\mathbf{0}\}$  and  $x_2 = 0$ .

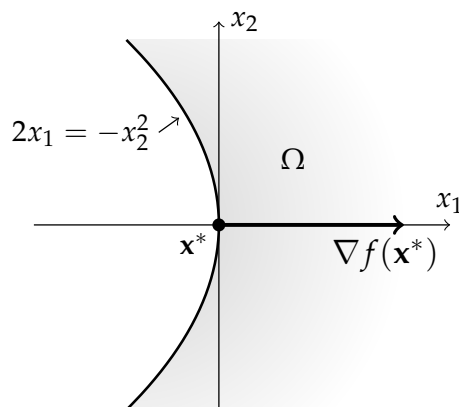


Figure 2. Region  $\Omega$ , point  $\mathbf{x}^*$ , and gradient  $\nabla f(\mathbf{x}^*)$  for Example 2.

#### 4. Refinements Based on Twice Differentiability

The additional smoothness of  $f$  yields stronger results.

**Definition 4** (Generalized Twice Differentiability). Let  $\Omega$  be a subset of  $\mathbb{R}^n$ . We say that a function  $f: \Omega \rightarrow \mathbb{R}$  is **twice differentiable** at a point  $\mathbf{x}^* \in \Omega$  if (1)  $\mathbf{x}^*$  is not an isolated point of  $\Omega$ , and (2) there exist  $\mathbf{g}(\mathbf{x}^*) \in \mathbb{R}^n$  and  $\mathbf{H}(\mathbf{x}^*) \in \mathbb{R}^{n \times n}$  such that the function  $r(\mathbf{x})$  defined by

$$f(\mathbf{x}) - f(\mathbf{x}^*) = (\mathbf{x} - \mathbf{x}^*)^\top \mathbf{g}(\mathbf{x}^*) + (\mathbf{x} - \mathbf{x}^*)^\top \mathbf{H}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + r(\mathbf{x}) \tag{19}$$

satisfies

$$\lim_{n \rightarrow \infty} \frac{r(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}^*\|^2} = 0, \tag{20}$$

or, equivalently, such that for any sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  converging to  $\mathbf{x}^*$ ,

$$f(\mathbf{x}_n) - f(\mathbf{x}^*) = (\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{g}(\mathbf{x}^*) + (\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{H}(\mathbf{x}^*)(\mathbf{x}_n - \mathbf{x}^*) + r_n, \tag{21}$$

where

$$\lim_{n \rightarrow \infty} \frac{r_n}{\|\mathbf{x}_n - \mathbf{x}^*\|^2} = 0. \tag{22}$$

Any such vector  $\mathbf{g}(\mathbf{x}^*)$  is denoted  $\nabla f(\mathbf{x}^*)$  and called a **gradient** of  $f$  at  $\mathbf{x}^*$ , and any such matrix  $\mathbf{H}(\mathbf{x}^*)$  is called a **Hessian** of  $f$  at  $\mathbf{x}^*$ .

Twice differentiability implies differentiability according to Lemma 1 below. By Th. 3 on p. 160 of [10],  $f$  is twice differentiable at  $\mathbf{x}^*$  if it has an extension to a function  $\tilde{f}$  that is  $C^2$  on a neighborhood of  $\mathbf{x}^*$ . In this case,  $\mathbf{g}(\mathbf{x}^*)$  and  $\mathbf{H}(\mathbf{x}^*)$  can be chosen as the usual gradient and Hessian of  $\tilde{f}$ .

**Definition 5.** Given sequences  $\{a_n\}_{n=1}^\infty$  and  $\{b_n\}_{n=1}^\infty$  of real numbers, we write (1)  $a_n = \mathcal{O}(b_n)$  to mean that there exists  $M > 0$  such that  $|a_n| \leq M|b_n|$  for all sufficiently large  $n$ , and (2)  $a_n = \mathcal{o}(b_n)$  if  $b_n \neq 0$  for all sufficiently large  $n$  and  $\lim_{n \rightarrow \infty} a_n/b_n = 0$ .

**Lemma 1.** Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f: \Omega \rightarrow \mathbb{R}$  be twice differentiable at some point  $\mathbf{x}^* \in \Omega$  with gradient  $\nabla f(\mathbf{x}^*)$ . Then, for any sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  converging to  $\mathbf{x}^*$ ,

$$f(\mathbf{x}_n) - f(\mathbf{x}^*) = (\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) + \mathcal{O}(\|\mathbf{x}_n - \mathbf{x}^*\|^2). \tag{23}$$

**Proof.** We start with (21), assuming some choice of Hessian  $\mathbf{H}(\mathbf{x}^*)$ . Let  $M = \|\mathbf{H}(\mathbf{x}^*)\|$  be the matrix bound of  $\mathbf{H}(\mathbf{x}^*)$ . Then, by the triangle inequality and Cauchy–Schwarz inequality,

$$\begin{aligned} |(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{H}(\mathbf{x}^*)(\mathbf{x}_n - \mathbf{x}^*) + r_n| &\leq |(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{H}(\mathbf{x}^*)(\mathbf{x}_n - \mathbf{x}^*)| + |r_n| \\ &\leq M \|\mathbf{x}_n - \mathbf{x}^*\|^2 + \frac{|r_n|}{\|\mathbf{x}_n - \mathbf{x}^*\|^2} \|\mathbf{x}_n - \mathbf{x}^*\|^2. \end{aligned}$$

By (22),  $r_n/\|\mathbf{x}_n - \mathbf{x}^*\|^2 \rightarrow 0$ , and hence  $r_n/\|\mathbf{x}_n - \mathbf{x}^*\|^2 \leq 1$  for all sufficiently large  $n$ . Therefore  $|(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{H}(\mathbf{x}^*)(\mathbf{x}_n - \mathbf{x}^*) + r_n| \leq (M + 1)\|\mathbf{x}_n - \mathbf{x}^*\|^2$  for all sufficiently large  $n$ .  $\square$

**Theorem 5** (FOSC 3). Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f: \Omega \rightarrow \mathbb{R}$  be twice differentiable at some point  $\mathbf{x}^* \in \Omega$ , with gradient  $\nabla f(\mathbf{x}^*) \neq 0$ . Assume that for every sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  with  $\mathbf{x}_n \rightarrow \mathbf{x}^*$ , there is a sequence  $\{\delta_n\}_{n=1}^\infty$  of positive numbers such that  $\|\mathbf{x}_n - \mathbf{x}^*\|^2 = \mathcal{O}(\delta_n)$  and

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) \geq \delta_n \tag{24}$$

for all sufficiently large  $n$ . Then,  $\mathbf{x}^*$  is a strict local minimizer of  $f$  over  $\Omega$ .



**Proof.** If the conclusion is not true, there exists a sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  with  $\mathbf{x}_n \rightarrow \mathbf{x}^*$  such that  $f(\mathbf{x}_n) \leq f(\mathbf{x}^*)$  for all  $n$ . Then, by hypothesis, there is a sequence  $\{\delta_n\}_{n=1}^\infty$  of positive numbers satisfying (24) such that  $\|\mathbf{x}_n - \mathbf{x}^*\|^2 = \mathcal{O}(\delta_n)$ . By Lemma 1 and (24),

$$\begin{aligned} f(\mathbf{x}_n) - f(\mathbf{x}^*) &= (\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) + \mathcal{O}(\|\mathbf{x}_n - \mathbf{x}^*\|^2) \\ &\geq \delta_n + \mathcal{O}(\|\mathbf{x}_n - \mathbf{x}^*\|^2) \end{aligned}$$

for all sufficiently large  $n$ . Hence, since  $\|\mathbf{x}_n - \mathbf{x}^*\|^2 = \mathcal{O}(\delta_n)$ ,

$$\frac{f(\mathbf{x}_n) - f(\mathbf{x}^*)}{\delta_n} \geq 1 + \frac{\mathcal{O}(\|\mathbf{x}_n - \mathbf{x}^*\|^2)}{\delta_n} > 0 \tag{25}$$

for all sufficiently large  $n$ . Thus,  $f(\mathbf{x}_n) > f(\mathbf{x}^*)$  for all sufficiently large  $n$ , contradicting our choice of the sequence  $\{\mathbf{x}_n\}_1^\infty$ . The result follows.  $\square$

**Theorem 6** (FONC 3). *Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f : \Omega \rightarrow \mathbb{R}$  be twice differentiable at some point  $\mathbf{x}^* \in \Omega$ , with gradient  $\nabla f(\mathbf{x}^*) \neq 0$ . If there exist a sequence  $\{\mathbf{x}_n\}_1^\infty$  in  $\Omega \setminus \{\mathbf{x}^*\}$  converging to  $\mathbf{x}^*$  and a sequence  $\{\delta_n\}_{n=1}^\infty$  of positive numbers such that  $\|\mathbf{x}_n - \mathbf{x}^*\|^2 = \mathcal{O}(\delta_n)$  and*

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) \leq -\delta_n \tag{26}$$

for all sufficiently large  $n$ , then  $\mathbf{x}^*$  is not a local minimizer of  $f$  on  $\Omega$ .

Equivalently, if  $\mathbf{x}^*$  is a local minimizer of  $f$  on  $\Omega$ , no such sequences  $\{\mathbf{x}_n\}_1^\infty$  and  $\{\delta_n\}_{n=1}^\infty$  can exist.

**Proof.** Let  $\{\mathbf{x}_n\}_1^\infty$  and  $\{\delta_n\}_{n=1}^\infty$  be sequences with the properties stated in the theorem. By Lemma 1 and (26),

$$\begin{aligned} f(\mathbf{x}_n) - f(\mathbf{x}^*) &= (\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) + \mathcal{O}(\|\mathbf{x}_n - \mathbf{x}^*\|^2) \\ &\leq -\delta_n + \mathcal{O}(\|\mathbf{x}_n - \mathbf{x}^*\|^2) \end{aligned} \tag{27}$$

for all sufficiently large  $n$ . Hence, since  $\|\mathbf{x}_n - \mathbf{x}^*\|^2 = \mathcal{O}(\delta_n)$ ,

$$\frac{f(\mathbf{x}_n) - f(\mathbf{x}^*)}{\delta_n} \leq -1 + \frac{\mathcal{O}(\|\mathbf{x}_n - \mathbf{x}^*\|^2)}{\delta_n} < 0$$

for all sufficiently large  $n$ . Thus,  $f(\mathbf{x}_n) < f(\mathbf{x}^*)$  for all sufficiently large  $n$ , and therefore  $\mathbf{x}^*$  is not a local minimizer.  $\square$

**Remark 1.** *If  $f, \Omega, \mathbf{x}^*, \nabla f(\mathbf{x}^*) \neq 0$  satisfy the conditions for a strict local minimizer in Theorem 3 (FOSC 2), they satisfy the conditions in Theorem 5 (FOSC 3) by choosing  $\delta_n = \delta \|\mathbf{x}_n - \mathbf{x}^*\|$ ,  $n \geq 1$ . Similarly, if  $f, \Omega, \mathbf{x}^*, \nabla f(\mathbf{x}^*) \neq 0$  meet the conditions for a local non-minimizer in Theorem 4 (FONC 2), they meet the conditions in Theorem 6 (FONC 3). Examples 3 and 4 show that both converse statements fail.*

**Example 3.** *See Figure 3. Set  $f(\mathbf{x}) = x_1$ ,  $\mathbf{x} = [x_1, x_2]^\top$ , on  $\Omega = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq |x_2|^{3/2}\}$  and let  $\mathbf{x}^* = \mathbf{0}$ . Then,  $f$  is twice differentiable at  $\mathbf{x}^*$  with gradient  $\nabla f(\mathbf{x}^*) = [1, 0]^\top$ . The point  $\mathbf{x}^*$  is a strict local minimizer because  $f(\mathbf{0}) = 0$  and  $f(\mathbf{x}) > 0$  for every other  $\mathbf{x} \in \Omega$ . The condition for a strict local minimizer in Theorem 3 (FOSC 2) is not satisfied. For example, the sequence*

$$\mathbf{x}_n = \left[ \frac{1}{n}, \frac{1}{n^{2/3}} \right], \quad n \geq 1,$$

is in  $\Omega \setminus \{\mathbf{x}^*\}$ ,  $\mathbf{x}_n^* \rightarrow \mathbf{0}$ , and the inequality (11) implies  $1/n \geq \delta [1/n^2 + 1/n^{4/3}]^{1/2}$ , which is impossible. On the other hand, the condition in Theorem 5 (FOSC 3) is satisfied. Consider any



sequence  $\mathbf{x}_n = [x_{n,1}, x_{n,2}]^\top$ ,  $n \geq 1$ , in  $\Omega \setminus \{\mathbf{x}^*\}$  with  $\mathbf{x}_n \rightarrow \mathbf{x}^*$ . Choosing  $\delta_n = x_{n,1}$ , we readily find that  $(\mathbf{x}_n - \mathbf{x}^*)^\top \nabla f(\mathbf{x}^*) \geq \delta_n$ , and  $\|\mathbf{x}_n - \mathbf{x}^*\|^2 = o(\delta_n)$ .

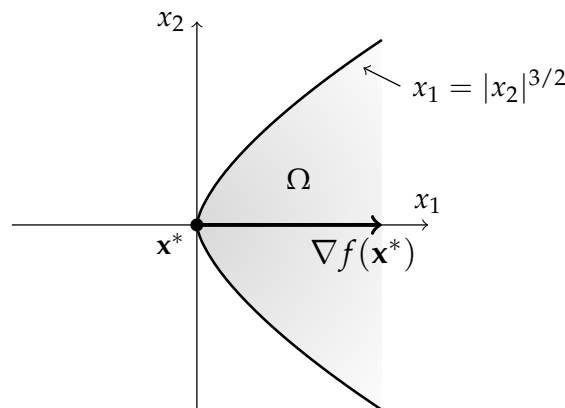


Figure 3. Region  $\Omega$ , point  $\mathbf{x}^*$ , and gradient  $\nabla f(\mathbf{x}^*)$  for Examples 3 and 5.

**Example 4.** See Figure 4. Set  $f(\mathbf{x}) = x_1$ ,  $\mathbf{x} = [x_1, x_2]^\top$ , on  $\Omega = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq -|x_2|^{3/2}\}$  and let  $\mathbf{x}^* = \mathbf{0}$ . Then,  $f$  is twice differentiable at  $\mathbf{x}^*$  with gradient  $\nabla f(\mathbf{x}^*) = [1, 0]^\top$ . Since  $f(\mathbf{x}) < 0$  on the curve  $x_1 = -|x_2|^{3/2}$ ,  $\mathbf{x}^*$  is not a local minimizer of  $f$  on  $\Omega$ . It is not possible to show this using Theorem 4 (FONC 2), because it is impossible to find a sequence satisfying the conditions there. However, the conditions of Theorem 6 (FONC 3) are met by choosing

$$\mathbf{x}_n = \left[ -\frac{1}{n^{3/2}}, \frac{1}{n} \right]^\top \quad \text{and} \quad \delta_n = \frac{1}{n^{3/2}}$$

for all  $n \geq 1$ .

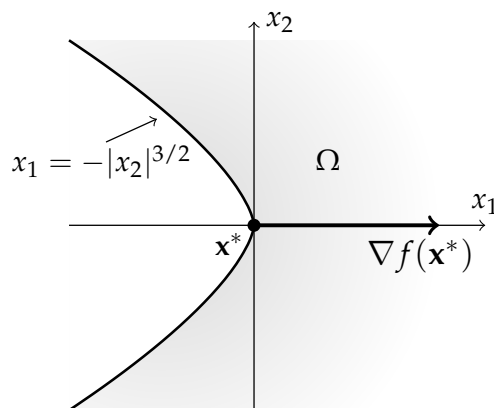


Figure 4. Region  $\Omega$ , point  $\mathbf{x}^*$ , and gradient  $\nabla f(\mathbf{x}^*)$  for Example 4.

### 5. Applications of the Analytical Conditions

Example 3 suggests looking for generalizations of Theorems 1 and 2 to larger regions. In this section, we show that such generalizations exist, again assuming twice differentiability.

**Definition 6.** Let  $\mathbf{x}^*, \mathbf{d} \in \mathbb{R}^n$ ,  $\mathbf{d} \neq \mathbf{0}$ , and  $\alpha > 0$  be given. For each  $\beta > 0$ , define  $K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}) \subseteq \mathbb{R}^n$  as the set consisting of the point  $\mathbf{x}^*$ , together with all points  $\mathbf{x} \neq \mathbf{x}^*$  such that (1)  $\|\mathbf{u}\| \geq \beta \|\mathbf{v}\|^\alpha$  and (2)  $\mathbf{u}^\top \mathbf{d} > 0$ , where

$$\mathbf{u} = \frac{(\mathbf{x} - \mathbf{x}^*)^\top \mathbf{d}}{\|\mathbf{d}\|} \frac{\mathbf{d}}{\|\mathbf{d}\|}$$

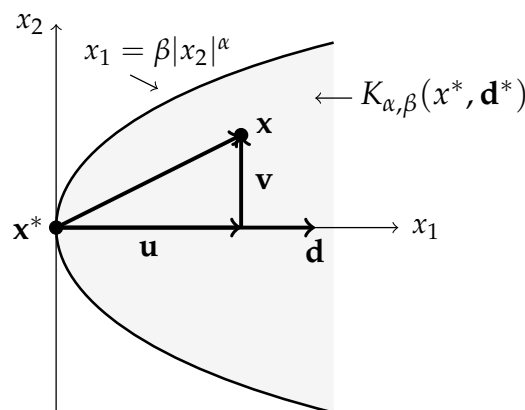
is the projection of  $\mathbf{x} - \mathbf{x}^*$  in the direction  $\mathbf{d}$ , and

$$\mathbf{v} = (\mathbf{x} - \mathbf{x}^*) - \mathbf{u}$$

is a component of  $\mathbf{x} - \mathbf{x}^*$  in the orthogonal direction. We call  $K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d})$  an  $\alpha$ -cone. The opposite  $\alpha$ -cone is the set

$$-K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}) = K_{\alpha,\beta}(\mathbf{x}^*, -\mathbf{d}).$$

See Figure 5. For  $\alpha = 1$ ,  $K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}) \subseteq \mathbb{R}^n$  is a cone as in Definition 1. For  $\alpha = 2$ , it is a paraboloid.



**Figure 5.** This figure illustrates Definition 6 when  $\mathbf{x}^* = \mathbf{0}$  and  $\mathbf{d}$  is in the positive  $x_1$ -direction. The  $\alpha$ -cone  $K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d})$  is the region to the right of the curve together with the curve itself.

**Theorem 7 (FOSC 4).** Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f : \Omega \rightarrow \mathbb{R}$  be twice differentiable at some point  $\mathbf{x}^* \in \Omega$ , with nonzero gradient  $\mathbf{d}^* = \nabla f(\mathbf{x}^*)$ . If  $\Omega \subseteq K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}^*)$  for some  $\alpha \in [1, 2)$  and  $\beta > 0$ , then  $\mathbf{x}^*$  is a strict local minimizer of  $f$  over  $\Omega$ .

As in Theorem 1, we can apply Theorem 7 with  $f$  replaced by  $-f$ .

**Corollary 2.** Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f : \Omega \rightarrow \mathbb{R}$  be twice differentiable at some point  $\mathbf{x}^* \in \Omega$ , with nonzero gradient  $\mathbf{d}^* = \nabla f(\mathbf{x}^*)$ . If  $\Omega \subseteq -K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}^*)$  for some  $\alpha \in [1, 2)$  and  $\beta > 0$ , then  $\mathbf{x}^*$  is a strict local maximizer of  $f$  over  $\Omega$ .

**Remark 2.** Theorem 7 fails for  $\alpha = 2$ . For an example, let  $\mathbf{x}^* = \mathbf{0}$ ,  $\mathbf{x} = [x_1, x_2]^\top$ , and

$$\Omega = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq x_2^2\}.$$

Define

$$f(\mathbf{x}) = x_1 - 2x_2^2, \quad \mathbf{x} \in \Omega.$$

Then,  $f$  is twice differentiable at  $\mathbf{x}^*$ ,  $\mathbf{d}^* = \nabla f(\mathbf{x}^*) = [1, 0]^\top$ , and  $\Omega = K_{2,1}(\mathbf{x}^*, \mathbf{d}^*)$  (see Figure 5). If Theorem 7 were true for  $\alpha = 2$ , then  $\mathbf{x}^* = \mathbf{0}$  would be a strict local minimizer for  $f$  on  $\Omega$ . However,  $f$  assumes positive, negative, and zero values at points of  $\Omega$  arbitrarily close to  $\mathbf{x}^*$ . This contradicts Theorem 7, and therefore  $\alpha = 2$  cannot be allowed in the theorem.

**Proof of Theorem 7.** We prove the theorem by verifying the condition for a strict local minimizer in Theorem 5 (FOSC 3). Let  $\{\mathbf{x}_n\}_1^\infty$  be a sequence in  $\Omega \setminus \{\mathbf{x}^*\}$  such that  $\mathbf{x}_n \rightarrow \mathbf{x}^*$ .

We shall construct a sequence  $\{\delta_n\}_{n=1}^\infty$  of positive numbers satisfying (24) for all sufficiently large  $n$  such that  $\|\mathbf{x}_n - \mathbf{x}^*\|^2 = \mathcal{O}(\delta_n)$ . For all  $n \geq 1$ , set

$$\mathbf{u}_n = \frac{(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^*}{\|\mathbf{d}^*\|} \frac{\mathbf{d}^*}{\|\mathbf{d}^*\|} \quad \text{and} \quad \mathbf{v} = (\mathbf{x}_n - \mathbf{x}^*) - \mathbf{u}_n,$$

Then,  $\|\mathbf{u}_n\| \geq \beta \|\mathbf{v}_n\|^\alpha$  and  $\mathbf{u}_n^\top \mathbf{d}^* > 0$  for all  $n$  by the definition of  $K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}^*)$ . Since  $\mathbf{u}_n$  and  $\mathbf{v}_n$  are orthogonal,

$$\|\mathbf{x}_n - \mathbf{x}^*\|^2 = \|\mathbf{u}_n\|^2 + \|\mathbf{v}_n\|^2 \leq \|\mathbf{u}_n\|^2 + \left(\frac{1}{\beta}\right)^{2/\alpha} \|\mathbf{u}_n\|^{2/\alpha}.$$

Since  $\mathbf{x}_n \rightarrow \mathbf{x}^*$ ,  $\mathbf{u}_n \rightarrow \mathbf{0}$ , and hence  $\|\mathbf{u}_n\| < 1$  for all sufficiently large  $n$ , say  $n \geq n_0$ . Since  $\alpha \geq 1$ ,  $2/\alpha \leq 2$ , and hence  $\|\mathbf{u}_n\|^{2/\alpha} \geq \|\mathbf{u}_n\|^2$  for all  $n \geq n_0$  (because  $\|\mathbf{u}_n\| < 1$ ). Thus,

$$\|\mathbf{x}_n - \mathbf{x}^*\|^2 \leq \|\mathbf{u}_n\|^{2/\alpha} + \left(\frac{1}{\beta}\right)^{2/\alpha} \|\mathbf{u}_n\|^{2/\alpha} = \gamma \|\mathbf{u}_n\|^{2/\alpha}, \quad n \geq n_0.$$

We assume also that  $\alpha < 2$  and hence  $2/\alpha > 1$ , say  $2/\alpha = 1 + \varepsilon$ . Then,

$$\begin{aligned} \|\mathbf{x}_n - \mathbf{x}^*\|^2 &\leq \gamma \|\mathbf{u}_n\|^{2/\alpha} = \gamma \left[ \frac{(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^*}{\|\mathbf{d}^*\|} \right]^{1+\varepsilon} \\ &= \gamma \frac{(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^*}{\|\mathbf{d}^*\|} \left[ \frac{(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^*}{\|\mathbf{d}^*\|} \right]^\varepsilon \\ &\leq \gamma \frac{(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^*}{\|\mathbf{d}^*\|} \|\mathbf{x}_n - \mathbf{x}^*\|^\varepsilon, \end{aligned}$$

by the Cauchy–Schwarz inequality and the fact that  $(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^* = \mathbf{u}_n^\top \mathbf{d}^* > 0$ . Therefore,

$$(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^* \geq \frac{\|\mathbf{x}_n - \mathbf{x}^*\|^{2-\varepsilon} \|\mathbf{d}^*\|}{\gamma}, \quad n \geq n_0.$$

Setting  $\delta_n = \|\mathbf{x}_n - \mathbf{x}^*\|^{2-\varepsilon} \|\mathbf{d}^*\| / \gamma$ , we obtain  $(\mathbf{x}_n - \mathbf{x}^*)^\top \mathbf{d}^* \geq \delta_n$  for all sufficiently large  $n$  and

$$\lim_{n \rightarrow \infty} \frac{\|\mathbf{x}_n - \mathbf{x}^*\|^2}{\delta_n} = \lim_{n \rightarrow \infty} \frac{\gamma \|\mathbf{x}_n - \mathbf{x}^*\|^\varepsilon}{\|\mathbf{d}^*\|} = 0.$$

We have verified the requirements in Theorem 5, and therefore  $\mathbf{x}^*$  is a strict local minimizer of  $f$  over  $\Omega$  by that result.  $\square$

**Theorem 8** (FONC 4). *Let  $\Omega$  be a subset of  $\mathbb{R}^n$ , and let  $f : \Omega \rightarrow \mathbb{R}$  be twice differentiable at some point  $\mathbf{x}^* \in \Omega$ , with nonzero gradient  $\mathbf{d}^* = \nabla f(\mathbf{x}^*)$ . If  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on  $\Omega$ , then  $\mathbf{x}^*$  is an isolated point of  $-K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}^*) \cap \Omega$  for every  $\alpha \in [1, 2)$  and  $\beta > 0$ .*

**Proof.** Assume that  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on  $\Omega$ , and, if possible, that  $\mathbf{x}^*$  is not an isolated point of  $-K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}^*) \cap \Omega$  for some  $\alpha \in [1, 2)$  and  $\beta > 0$ . Then, there is a sequence  $\{\mathbf{x}_n\}_{n=1}^\infty$  that belongs to both  $-K_{\alpha,\beta}(\mathbf{x}^*, \mathbf{d}^*)$  and  $\Omega \setminus \{\mathbf{x}^*\}$  such that  $\mathbf{x}_n \rightarrow \mathbf{x}^*$ . By Corollary 2,  $f(\mathbf{x}_n) < f(\mathbf{x}^*)$  for all sufficiently large  $n$ , contradicting our assumption that  $\mathbf{x}^*$  is a strict local minimizer of  $f$  on  $\Omega$ . The theorem follows.  $\square$

Examples 5 and 6 are set in the context of the Karush–Kuhn–Tucker (KKT) theorem [2,6], which allows constraint conditions to be expressed in terms of inequalities. We follow the account in [1], in which the KKT theorem appears as Theorem 21.1, and Theorem 21.3 is the corresponding second-order sufficient condition (SOSC). Theorems 9 and 10 below are specializations of these results to the cases that concern us here. Theorems 21.1 and 21.3 in [1] allow additional Lagrange-type conditions that play no role in our examples.

**Theorem 9 (KKT Theorem).** Let  $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$  be given  $C^1$  functions. Assume that  $\mathbf{x}^* \in \mathbb{R}^n$  is a local minimizer for  $f$  subject to the condition  $g(\mathbf{x}) \leq 0$ , and that  $\mathbf{x}^*$  is a regular point for  $g$  in the sense that  $\nabla g(\mathbf{x}^*) \neq \mathbf{0}$ . Then, there is a real number  $\mu^* \geq 0$  such that

- (1)  $\mu^* g(\mathbf{x}^*) = 0$ ;
- (2)  $\nabla f(\mathbf{x}^*) + \mu^* \nabla g(\mathbf{x}^*) = \mathbf{0}$ .

The corresponding sufficient condition requires the stronger assumption that the given functions  $f, g$  are  $C^2$ . The Hessians  $F, G$  for  $f, g$  are the  $n \times n$  matrices of second-order partials of  $f, g$ .

**Theorem 10 (SOSC).** Let  $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$  be given  $C^2$  functions. Assume that  $\mathbf{x}^* \in \mathbb{R}^n$  satisfies  $g(\mathbf{x}^*) \leq 0$  and we can find a real number  $\mu^* \geq 0$  satisfying the following conditions:

- (1)  $\mu^* g(\mathbf{x}^*) = 0$ .
- (2)  $\nabla f(\mathbf{x}^*) + \mu^* \nabla g(\mathbf{x}^*) = \mathbf{0}$ .
- (3) If  $F, G$  are the Hessians of  $f, g$  and  $L(\mathbf{x}^*, \mu^*) = F(\mathbf{x}^*) + \mu^* G(\mathbf{x}^*)$ , then  $\mathbf{y}^\top L(\mathbf{x}^*, \mu^*) \mathbf{y} > 0$  for all  $\mathbf{y} \in \mathbb{R}^n$  such that  $\mathbf{y} \neq \mathbf{0}$  and  $\nabla g(\mathbf{x}^*)^\top \mathbf{y} = 0$ .

Then,  $\mathbf{x}^*$  is a strict local minimizer for  $f$  subject to the condition  $g(\mathbf{x}) \leq 0$ .

**Example 5.** Set  $\mathbf{x}^* = \mathbf{0}$ ,  $f(\mathbf{x}) = x_1 - x_1^2 - x_2^2$ , and  $g(\mathbf{x}) = |x_2|^{3/2} - x_1$  for all  $\mathbf{x} = [x_1, x_2]^\top$  in  $\mathbb{R}^2$ . Then,  $f \in C^2$  and  $g \in C^1$  on  $\mathbb{R}^2$ . The set  $\Omega = \{\mathbf{x} \in \mathbb{R}^2: g(\mathbf{x}) \leq 0\}$  is an alpha-cone with  $\alpha = 3/2$  in the direction  $\nabla f(\mathbf{x}^*) = [1, 0]^\top$  (see Figure 3). Therefore, by Theorem 7 (FOSC 4),  $\mathbf{x}^*$  is a strict local minimizer for  $f$  subject to the constraint  $g(\mathbf{x}) \leq 0$ . However, this cannot be shown with Theorem 10 because  $g \notin C^2$ , which is a hypothesis in Theorem 10. To see why this is a problem, consider the form  $L(\mathbf{x}, \mu^*) = F(\mathbf{x}) + \mu^* G(\mathbf{x})$  that appears in condition (3). At any point  $\mathbf{x} = [x_1, x_2]^\top$  with  $x_2 \neq 0$ , the Hessian of  $g$  is given by

$$G(\mathbf{x}) = \left[ \frac{\partial^2 g}{\partial x_i \partial x_j} \right]_{i,j=1}^2 = \begin{bmatrix} 0 & 0 \\ 0 & \frac{3}{4} |x_2|^{-1/2} \end{bmatrix}.$$

The second partial  $\partial^2 g / \partial x_2^2$  does not exist at any point on the line  $x_2 = 0$ . Thus,  $G(\mathbf{x}^*)$  is undefined. Hence, the Lagrangian  $L(\mathbf{x}^*, \mu^*) = F(\mathbf{x}^*) + \mu^* G(\mathbf{x}^*)$  in condition (3) of Theorem 10 is undefined, and therefore Theorem 10 cannot be applied. We remark that this example is within the scope of Theorem 9 (KKT Theorem), and the conditions (1) and (2) there are satisfied with  $\mu^* = 1$ .

**Example 6.** Set  $\mathbf{x}^* = \mathbf{0}$ ,  $f(\mathbf{x}) = x_1 - x_1^2 - x_2^2$ , and  $g(\mathbf{x}) = x_2^2 - x_1$  for all  $\mathbf{x} = [x_1, x_2]^\top$  in  $\mathbb{R}^2$ . Then,  $f, g \in C^2$  on  $\mathbb{R}^2$ . In this example,  $\mathbf{x}^*$  is not a local minimizer of  $f$  subject to the constraint  $g(\mathbf{x}) \leq 0$ . For example, for  $\mathbf{x} \neq \mathbf{x}^*$  on the boundary  $x_1 = x_2^2$  of the constraint set,

$$f(\mathbf{x}) = x_2^2 - x_1^2 + x_2^2 = -x_1^2 < 0.$$

Might this example contradict Theorem 7 or Theorem 10? Fortunately, no, and it is instructive to see why. Theorem 7 is not applicable because the constraint set  $\Omega = \{\mathbf{x} \in \mathbb{R}^2: g(\mathbf{x}) \leq 0\}$  is an alpha-cone with  $\alpha = 2$ , and it is shown in Remark 2 that Theorem 7 fails for  $\alpha = 2$ . To see that Theorem 10 is also not applicable, let us check the required conditions (1)–(3):

- (1) Since  $g(\mathbf{x}^*) = 0$ ,  $\mu^* g(\mathbf{x}^*) = 0$  for all  $\mu^* \geq 0$ .
- (2) For  $\mu^* = 1$ ,  $\nabla g(\mathbf{x}^*) + \mu^* \nabla f(\mathbf{x}^*) = [-1, 0]^\top + \mu^* [1, 0]^\top = \mathbf{0}$ .
- (3) In our example,

$$L(\mathbf{x}^*, \mu^*) = F(\mathbf{x}^*) + \mu^* G(\mathbf{x}^*) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore,

$$\mathbf{y}^\top L(\mathbf{x}^*, \mu^*) \mathbf{y} = \begin{bmatrix} 0 & y_2 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ y_2 \end{bmatrix} = 0 \quad (28)$$

for every  $\mathbf{y} = [0, y_2]^\top$  such that  $y_2 \neq 0$ , that is, for all  $\mathbf{y} \neq \mathbf{0}$  such that  $\nabla g(x^*)^\top \mathbf{y} = 0$ .

In view of (28), the positive definiteness condition in (3) fails, and hence Theorem 7 cannot be applied to this example.

## 6. Conclusions

The first-order necessary conditions in this paper contribute to the literature on first-order optimality conditions by introducing stronger results than those in the current literature. The new first-order necessary conditions imply the standard first-order necessary conditions, including those in [1,5]. We introduced first-order sufficient conditions that we did not find elsewhere in the literature. Our explanation of why the new conditions are stronger used examples that were two-dimensional. However, the method is applicable to general  $n$ -dimensional problems including linear programming.

We proposed first-order sufficient conditions for set-constrained optimization that do not require the objective function to be convex or the constraint equations to be differentiable. Conditions that require the function to be convex are essentially second-order conditions. Our conditions only require the gradient of the objective function to be nonzero at a candidate minimizer, and they are essentially first-order conditions even when we apply them to problems where the objective function is twice differentiable.

When the given function is continuously differentiable at  $x^*$  and the gradient is nonzero, the simplest form of the sufficient condition says that there is a cone with a vertex at  $x^*$ , and  $x^*$  is a strict local minimizer on the cone. This sufficient condition was employed to prove a corresponding necessary condition that does not use feasible directions and instead uses the topological notion of an isolated point in a set.

We introduced generalized differentiability and reformulated the first-order conditions in terms of convergent sequences. The new differentiability does not require the objective function to be defined on an open neighborhood of  $x^*$ . It only requires the function to be defined on the constraint set.

We refined the first-order conditions for a minimizer to twice differentiable functions in terms of  $\alpha$ -cones. The sufficiency version says that a twice differentiable function with a nonzero gradient has a strict local minimizer at the vertex of an  $\alpha$ -cone whose axis is the gradient direction. We presented a problem with an  $\alpha$ -cone constraint set where the new sufficiency condition shows that the candidate point is a strict local minimizer. This problem satisfies the necessary condition of the KKT method but not the sufficient condition, because the Hessian is undefined at the candidate minimizer.

**Author Contributions:** Conceptualization, S.M.R. and E.K.P.C.; methodology, S.M.R., E.K.P.C. and J.R.; writing—original draft, S.M.R.; writing—review and editing, J.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** S.M. Rovnyak was supported in part by the National Science Foundation under grant ECCS-1711521. E.K.P. Chong was supported in part by the National Science Foundation under grant CCF-2006788.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** The authors thank Henry Rovnyak for help with the LaTeX document and TikZ figures.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Chong, E.K.P.; Zak, S.H. *An Introduction to Optimization*, 4th ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2013.
2. Forst, W.; Hoffmann, D. *Optimization—Theory and Practice*; Springer Undergraduate Texts in Mathematics and Technology; Springer: New York, NY, USA, 2010.
3. Sioshansi, R.; Conejo, A.J. *Optimization in Engineering: Models and Algorithms*; Springer Optimization and Its Applications; Springer: Cham, Switzerland, 2017; Volume 120.
4. Butenko, S.; Pardalos, P.M. *Numerical Methods and Optimization*; Chapman & Hall/CRC Numerical Analysis and Scientific Computing; CRC Press: Boca Raton, FL, USA, 2014.
5. Kochenderfer, M.J.; Wheeler, T.A. *Algorithms for Optimization*; MIT Press: Cambridge, MA, USA, 2019.
6. Luenberger, D.G. *Optimization by Vector Space Methods*; John Wiley & Sons, Inc.: New York, NY, USA, 1969.
7. Lewis, A.D. Maximum Principle. Online Lecture Notes. 2006. Available online: <https://mast.queensu.ca/~andrew/teaching/pdf/maximum-principle.pdf> (accessed on 9 October 2023).
8. Peng, S.G. A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.* **1990**, *28*, 966–979. [[CrossRef](#)]
9. Lu, Q. Second order necessary conditions for optimal control problems of stochastic evolution equations. In Proceedings of the 35th Chinese Control Conference (CCC), Chengdu, China, 27–29 July 2016.
10. Marsden, J.E.; Tromba, A.J. *Vector Calculus*, 6th ed.; W.H. Freeman & Company: New York, NY, USA, 2012.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.