JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 47, No. 4, DECEMBER 1985

TECHNICAL NOTE

Directional Derivatives in Nonsmooth Optimization[1]

A. BEN-TAL[2] AND J. ZOWE[3]

Communicated by O. L. Mangasarian

Abstract. In this note, we consider two notions of second-order directional derivatives and discuss their use in the characterization of minimal points for nonsmooth functions.

Key Words. First-order and second-order directional derivatives, optimality conditions, nonsmooth optimization.

1. First-Order and Second-Order Directional Derivatives

Let f be a real functional on a real vector space X. We assume throughout that the following directional derivatives exist at $\bar{x}$ in all directions h and z

$$f'(\bar{x}; h) := \lim_{\lambda \to 0^+} \lambda^{-1}[f(\bar{x}+\lambda h) - f(\bar{x})], \tag{1}$$

$$f''(\bar{x}; h) := \lim_{\lambda \to 0^+} \lambda^{-2}[f(\bar{x}+\lambda h) - f(\bar{x}) - \lambda f'(\bar{x}; h)], \tag{2}$$

$$f''(\bar{x}; h, z) := \lim_{\lambda \to 0^+} \lambda^{-2}[f(\bar{x}+\lambda h + \lambda^2 z) - f(\bar{x}) - \lambda f'(\bar{x}; h)]. \tag{3}$$

(1) is the classical directional derivative. The proposal (2) for a second directional derivative is studied, e.g., in Demyanov and Pevnyi (Ref. 1) and Hiriart-Urruty (Ref. 2). The stronger notion (3) of a curved second directional derivative was introduced by the authors (see Ref. 3). We emphasize that the above limits exist for a large class of nonsmooth functions arising

[1] This research was supported by NSF Grant No. ECS-8214081, by the Fund for Promotion of Research at the Technion, and by Deutsche Forschungsgemeinschaft.

[2] Associate Professor, Technion-Israel Institute of Technology, Haifa, Israel.
[3] Professor, University of Bayreuth, Bayreuth, West Germany.

0022-3239/85/1200-0483$04.50/0

in applications. This class includes, e.g., the l_1-function, the max-function, and the exact penalty function; see Ben-Tal and Zowe (Ref. 4). If f is C^1 with Fréchet-derivative $f'(x)$, then the mean-value theorem gives, for small $\lambda > 0$ and suitable $0 \leq \theta_\lambda \leq 1$,

$$\begin{aligned} &\lambda^{-2}[f(\bar{x}+\lambda h+\lambda^2 z)-f(\bar{x})-\lambda f'(\bar{x}; h)] \\ &\quad =\lambda^{-2}[f(\bar{x}+\lambda h+\lambda^2 z)-f(\bar{x}+\lambda h)] \\ &\quad +\lambda^{-2}[f(\bar{x}+\lambda h)-f(\bar{x})-\lambda f'(\bar{x}; h)] \\ &\quad =f'(\bar{x}+\lambda h+\theta_\lambda \lambda^2 z)z+\lambda^{-2}[f(\bar{x}+\lambda h)-f(\bar{x})-\lambda f'(\bar{x}; h)]. \end{aligned}$$

Hence, for C^1-functions,

$$f''(\bar{x}; h, z)=f'(\bar{x})z+f''(\bar{x}; h). \tag{4}$$

If f is even C^2 with second Fréchet-derivative $f''(x)$, then a Taylor expansion of the bracketed terms in (1)-(3) shows that

$$f'(\bar{x}; h)=f'(\bar{x})h, \tag{5a}$$

$$f''(\bar{x}; h)=\tfrac{1}{2}f''(\bar{x})(h, h), \tag{5b}$$

$$f''(\bar{x}; h, z)=f'(\bar{x})z+\tfrac{1}{2}f''(\bar{x})(h, h). \tag{5c}$$

In the next two sections, we will characterize the minimality of $\bar{x}$ via the above directional derivatives. The necessary conditions (6), (7), and (8) are certainly not new. We cite them here only to show what has to be added to reach sufficient conditions. The sufficient conditions of Theorem 3.2 are new. Furthermore, it does not seem to be well known in the literature that the Lipschitz condition is essential for the sufficient conditions stated in Theorem 3.1. Examples and counterexamples are given.

2. Necessary Conditions

As an immediate consequence of the definitions (1), (2), and (3), we obtain the following theorems.

Theorem 2.1. If $\bar{x}$ is a local minimizer of f, then

$$f'(\bar{x}; h) \geq 0, \qquad \text{for all } h \in X. \tag{6}$$

Theorem 2.2. If $\bar{x}$ is a local minimizer of f, then

$$f'(\bar{x}; h) \geq 0, \qquad \text{for all } h \in X, \tag{7a}$$

and

$$f'(\bar{x}; h) = 0 \text{ implies } f''(\bar{x}; h) \geq 0. \tag{7b}$$

Theorem 2.3. If $\bar{x}$ is a local minimizer of f, then

$$f'(\bar{x}; h) \geq 0, \qquad \text{for all } h \in X, \tag{8a}$$

and

$$f'(\bar{x}; h) = 0 \text{ implies } f''(\bar{x}; h, z) \geq 0, \text{ for all } z \in X. \tag{8b}$$

For a C^2-function, we get from (5) that (7) and (8) coincide and reduce to the standard necessary conditions in smooth optimization

$$f'(\bar{x}) = 0 \text{ and } f''(\bar{x})(h, h) \geq 0, \text{ for all } h \in X.$$

For a C^1-function, the conditions (7) and (8) become identical, because of (4). However, for general nonsmooth f, condition (8) is superior to (7); it can exclude nonoptimal points not excluded by (7). We give two examples.

Example 2.1. Let $X = \mathbb{R}^2$ and

$$f(x) = x_1 + |2x_1 + x_1^2 + x_2^2|.$$

The origin $\bar{x} = (0, 0)^T$ is not locally optimal, since, for $-2 \leq x_1 < 0$,

$$f(x_1, \sqrt{-2x_1 - x_1^2}) = x_1 < 0 = f(\bar{x}).$$

The directional derivatives for some special h and z are (we omit some straightforward arithmetic)

$$f'(\bar{x}; (h_1, h_2)^T) = h_1 + 2|h_1|, \qquad \text{for all } (h_1, h_2),$$

$$f''(\bar{x}; (0, h_2)^T) = h_2^2,$$

$$f''(\bar{x}; (0, h_2)^T, (\alpha, 0)^T) = -\alpha - h_2^2, \qquad \alpha < -h_2^2/2.$$

Obviously, condition (8) is violated for the choice $\alpha \in (-h_2^2, -h_2^2/2)$, whereas (7) does not exclude the nonoptimal $\bar{x}$.

Suppose that the domain of definition of f is a proper subset S of X and $\bar{x}$ is a local minimizer of f on S. Then, the above inequalities (6), (7), and (8), respectively, hold for all directions h and z for which the corresponding directional derivatives exist. Again, the curved directional derivative $f''(\bar{x}; \cdot, \cdot)$ provides a more refined tool than $f''(\bar{x}; \cdot)$.

Example 2.2. Let $X = \mathbb{R}^2$, $a = (-1, 0)^T$,

$$S_\alpha = \{(x_1, x_2)^T | (x_1 - \alpha)^2 + x_2^2 \geq \alpha^2\},$$

with fixed parameter $\alpha < 0$, and define a function f_α on S_α by

$$f_\alpha(x) = (x-a)^T(x-a), \qquad \text{for } x \in S_\alpha.$$

Obviously, the origin $\bar{x} = (0,0)^T$ is the point in S_α closest to $a = (-1,0)^T$ in the l_2-norm whenever $\alpha \leqslant -1$; i.e., $\bar{x}$ minimizes f_α on S_α if and only if $\alpha \leqslant -1$.

The directional derivatives $f'_\alpha(\bar{x}; h)$ and $f''_\alpha(\bar{x}; h)$ exist for all h with $h_1 \geqslant 0$, and one has

$$f'_\alpha(\bar{x}; h) = 2(\bar{x}-a)^T h = 2h_1, \qquad h_1 \geqslant 0,$$

$$f''_\alpha(\bar{x}; h) = h^T h, \qquad h_1 \geqslant 0.$$

Hence, the necessary conditions (6) and (7) hold, no matter if $\alpha \leqslant -1$ or not. Now, consider condition (8). One easily checks that $\bar{x} + \lambda h + \lambda^2 z \in S_\alpha$, for all λ, $h = (0, h_2)^T$, and $z = (h_2^2/2\alpha, 0)^T$. For this special choice of h and z, one gets

$$f''(\bar{x}; h, z) = h^T h + 2(\bar{x}-a)^T z = h_2^2(1 + 1/\alpha).$$

Hence, (8) tells us that $\bar{x}$ cannot be optimal for $\alpha > -1$.

3. Sufficient Conditions

The directional derivatives characterize f along halflines and certain curves only. Hence, one cannot expect that replacement of $\geqslant$ by $>$ in (6), (7), and (8), respectively, provides a sufficient condition. Suitable topological assumptions have to be added. Some progress in this direction is reported in the following theorems. We restrict ourselves to finite-dimensional X; otherwise, the condition (6), for example, has to be strengthened to: $f'(\bar{x}; h) \geqslant \alpha \|h\|$, for all h and some fixed positive α; see Ref. 5. Following the sufficient conditions, we will discuss their interrelations and give some counterexamples.

Theorem 3.1. Suppose that $\dim X < \infty$ and f is Lipschitz continuous in a neighborhood of $\bar{x}$; i.e., there are $L > 0$ and $\varepsilon > 0$ such that

$$|f(x_1) - f(x_2)| \leqslant L\|x_1 - x_2\|, \qquad \text{whenever } \|x_i - \bar{x}\| \leqslant \varepsilon,\ i = 1, 2.$$

If

$$f'(\bar{x}; h) > 0, \qquad \text{for all } h \in X,\ h \neq 0, \tag{9}$$

then $\bar{x}$ is a strict local minimizer of f.

The proof of Theorem 3.1 uses a similar indirect argument as the more difficult proof of the following partial counterpart to Theorem 2.2.

Theorem 3.2. Suppose that $\dim X < \infty$, and let f be Fréchet-differentiable in a neighborhood of $\bar{x}$ with Lipschitz continuous $f'(\cdot)$ at $\bar{x}$; i.e., there are $L > 0$ and $\varepsilon > 0$ such that

$$\|f'(x) - f'(\bar{x})\| \leq L\|x - \bar{x}\|, \qquad \text{whenever } \|x - \bar{x}\| \leq \varepsilon. \tag{10}$$

If

$$f'(\bar{x}) = 0, \tag{11a}$$

and

$$f''(\bar{x}; h) > 0, \qquad \text{for all } h \in X,\ h \neq 0, \tag{11b}$$

then $\bar{x}$ is a strict local minimizer of f.

Proof. We proceed by contradiction. Suppose that there is a sequence $x_1, x_2, \ldots$, with

$$x_n \neq \bar{x} \text{ and } x_n \xrightarrow[n\to\infty]{} \bar{x}, \text{ but } f(x_n) \leq f(\bar{x}), \text{ for all } n.$$

A compactness argument in the finite-dimensional X says that, for a suitable subsequence,

$$(x_n - \bar{x})/\|x_n - \bar{x}\| \xrightarrow[n\to\infty]{} \bar{h} \neq 0.$$

Hence,

$$x_n = \bar{x} + \lambda_n \bar{h} + r_n, \qquad \text{with } \lambda_n := \|x_n - \bar{x}\|,$$

and

$$r_n/\lambda_n \xrightarrow[n\to\infty]{} 0.$$

By assumption,

$$\begin{aligned} 0 &\geq f(x_n) - f(\bar{x}) \\ &= [f(\bar{x} + \lambda_n \bar{h}) - f(\bar{x}) - \lambda_n f'(\bar{x})\bar{h}] + [f(x_n) - f(\bar{x} + \lambda_n \bar{h})]. \end{aligned} \tag{12}$$

The mean-value theorem implies that the second bracket is equal to

$$f'(\bar{x} + \lambda_n \bar{h} + \theta_n r_n) r_n, \qquad \text{for suitable } 0 \leq \theta_n \leq 1.$$

We add $f'(\bar{x}) r_n = 0$, divide by λ_n^2, and use the Lipschitz continuity of f' to see that

$$\begin{aligned} &\lambda_n^{-2}|f(x_n) - f(\bar{x} + \lambda_n \bar{h})| = \lambda_n^{-2}|[f'(\bar{x} + \lambda_n \bar{h} + \theta_n r_n) - f'(\bar{x})] r_n| \\ &\leq \lambda_n^{-2} L \|\lambda_n \bar{h} + \theta_n r_n\| \, \|r_n\| = L\|\bar{h} + \theta_n (r_n/\lambda_n)\| \cdot \|r_n/\lambda_n\| \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

The first bracket in (12) divided by λ_n^2 tends to $f''(\bar{x}; \bar{h})$ as $n \to \infty$. Hence, $f''(\bar{x}; \bar{h}) \leq 0$, which contradicts (11). □

When restricting the objective function f to a special class (including the l_1-function, the max-function, and the exact penalty function), we can establish a direct counterpart to Theorem 2.3; see Ref. 4 for a proof and for explicit formulas of the various directional derivatives.

Theorem 3.3. Suppose that dim $X < \infty$, and let f be of the type

$$f(x) = \sum_{i \in I} q_i(\max_{j \in J_i} f_{ij}(x)), \tag{13}$$

where $q_i \in C^2$, q_i nondecreasing, $f_{ij} \in C^2$, I and J_i for $i \in I$ are finite index sets. If

$$f'(\bar{x}; h) \geq 0, \qquad \text{for all } h, \tag{14a}$$

and

$$f'(\bar{x}; h) = 0 \text{ and } h \neq 0 \text{ implies } f''(\bar{x}; h, z) > 0, \qquad \text{for all } z \in X, \tag{14b}$$

then $\bar{x}$ is a strict local minimizer of f.

We start with an example demonstrating that Lipschitz continuity in Theorem 3.1 cannot be weakened to just continuity.

Example 3.1. Let $X = \mathbb{R}^2$ and

$$f(x) := \begin{cases} |x_2| - 2x_1^2, & \text{if } |x_2| \geq 2x_1^2, \\ (|x_2| - 2x_1^2)(|x_2| - x_1^2), & \text{if } 2x_1^2 \geq |x_2| \geq x_1^2, \\ |x_1| - \sqrt{|x_2|}, & \text{if } x_1^2 \geq |x_2|. \end{cases}$$

It is easily checked that f is continuous, but not Lipschitz continuous, at $\bar{x} = (0, 0)^T$. A simple geometric argument tells us that, for h, with $h_2 \neq 0$,

$$f'(\bar{x}; h) = \lim_{\lambda \to 0^+} \lambda^{-1}[|\lambda h_2| - 2\lambda^2 h_1^2] = |h_2|.$$

Furthermore,

$$f'(\bar{x}; (h_1, 0)^T) = |h_1|.$$

Hence, (9) holds. Nevertheless, $\bar{x}$ is not a local minimizer, since, for example,

$$f(\alpha, \tfrac{3}{2}\alpha^2) = -(\tfrac{1}{2}\alpha^2)(\tfrac{1}{2}\alpha^2) < 0 = f(\bar{x}), \qquad \text{for } \alpha > 0.$$

For C^2-functions, the conditions (11) and (14) coincide and reduce to the standard sufficient conditions in smooth optimization [use (4)],

$$f'(\bar{x}) = 0, \qquad f''(\bar{x})(h, h) > 0, \qquad \text{for all } h \neq 0. \tag{15}$$

But note that there are important functionals which are of type (13) or which are C^1, but not C^2; for these functionals, (14) or (11), respectively, applies but not (15). As an example for the latter, consider

$$f(x) := [\max\{0, g(x)\}]^p, \qquad \text{with } g \in C^2,$$

which plays a role in the context of exterior penalty functions [see, e.g., Auslender (Ref. 6) and Ben-Tal and Zowe (Ref. 4)]. This function is C^1 for $p > 1$, but in general not C^2 for $p \leq 2$.

Further, (4) tells us that (11) and (14) become identical for C^1-functions. But note that we encounter different situations in Theorems 3.2 and 3.3. In Theorem 3.2, we study the C^1-case. This differentiability assumption is essential. For the function f from Example 2.1, which is not C^1, one has

$$f'(\bar{x}; h) \geq 0, \qquad \text{for all } h,$$

and

$$f'(\bar{x}; h) = 0 \text{ and } h \neq 0 \text{ implies } f''(x; h) > 0;$$

nevertheless, $\bar{x}$ is not optimal.

To show that Theorem 3.3 may fail for functions not of type (13), one needs a significantly more tricky function. Here is one.

Example 3.2. Let $X = \mathbb{R}^2$, and consider, with fixed $\varepsilon > 0$,

$$f(x) := \begin{cases} |x_2| - |x_1|^{2+\varepsilon}, & \text{if } |x_2| \geq |x_1|^{2+\varepsilon}, \\ (|x_1|^{2+\varepsilon} - |x_2|)(|x_2| - \frac{1}{2}|x_1|^{2+\varepsilon}), & \text{if } |x_1|^{2+\varepsilon} \geq |x_2| \geq \frac{1}{2}|x_1|^{2+\varepsilon}, \\ (\frac{1}{2}|x_1|^{2+\varepsilon} - |x_2|)^{2/(2+\varepsilon)}, & \text{if } \frac{1}{2}|x_1|^{2+\varepsilon} \geq |x_2|. \end{cases}$$

The directional derivatives at $\bar{x} := (0,0)^T$ exist for all h and z, and a rather straightforward calculation shows that

$$f'(\bar{x}; h) = |h_2|,$$

and

$$f''(\bar{x}; (h_1, 0)^T, z) = \begin{cases} \lim_{\lambda \to 0^+} \lambda^{-2}[|\lambda^2 z_2| - |\lambda h_1 + \lambda^2 z_1|^{2+\varepsilon}] = |z_2|, & \text{if } z_2 \neq 0, \\ \lim_{\lambda \to 0^+} \lambda^{-2}[\frac{1}{2}|\lambda h_1 + \lambda^2 z_1|^{2+\varepsilon}]^{2/(2+\varepsilon)} = \frac{1}{2}|h_1|, & \text{if } z_2 = 0. \end{cases}$$

Hence, (14) is satisfied, but $\bar{x}$ is not optimal since, for example,

$$f(\alpha, \tfrac{2}{3}\alpha^{2+\varepsilon}) = -(\tfrac{1}{3}\alpha^{2+\varepsilon})(\tfrac{1}{6}\alpha^{2+\varepsilon}) < 0 = f(\bar{x}), \qquad \text{for } \alpha > 0.$$

References

1. DEMYANOV, V. F., and PEVNYI, A. B., *Expansion with Respect to a Parameter of the Extremal Values of Game Problems*, USSR Computational Mathematics and Mathematical Physics, Vol. 14, pp. 33–45, 1974.
2. HIRIART-URRUTY, J. B., *Limiting Behavior of the Approximate First-Order and Second-Order Directional Derivatives for a Convex Function*, Nonlinear Analysis: Theory, Methods, and Applications, Vol. 6, pp. 1309–1326, 1982.
3. BEN-TAL, A., and ZOWE, J., *A Unified Theory of First and Second-Order Conditions for Extremum Problems in Topological Vector Spaces*, Mathematical Programming Study, Vol. 19, pp. 39–77, 1982.
4. BEN-TAL, A., and ZOWE, J., *Necessary and Sufficient Optimality Conditions for a Class of Nonsmooth Minimization Problems*, Mathematical Programming, Vol. 24, pp. 70–92, 1982.
5. MAURER, H., and ZOWE, J., *First and Second-Order Necessary and Sufficient Optimality Conditions for Infinite-Dimensional Programming Problems*, Mathematical Programming, Vol. 16, pp. 98–110, 1979.
6. AUSLENDER, A., *Penalty Methods for Computing Points That Satisfy Second-Order Necessary Conditions*, Mathematical Programming, Vol. 17, pp. 229–238, 1979.