Appl. Math. Optim. 11:43–56 (1984)

Applied Mathematics and Optimization

Generalized Hessian Matrix and Second-Order Optimality Conditions for Problems with $C^{1,1}$ Data

Jean-Baptiste Hiriart-Urruty,[1] Jean-Jacques Strodiot,[2] and V. Hien Nguyen[2]

[1]UER Mathématiques, Informatique, Gestion, Université Paul Sabatier (Toulouse III), Toulouse, France and [2]Département de Mathématique, Facultés Universitaires de Namur, Namur, Belgium

Communicated by J. Stoer

Abstract. In this paper, we present a generalization of the Hessian matrix to $C^{1,1}$ functions, i.e., to functions whose gradient mapping is locally Lipschitz. This type of function arises quite naturally in nonlinear analysis and optimization. First the properties of the generalized Hessian matrix are investigated and then some calculus rules are given. In particular, a second-order Taylor expansion of a $C^{1,1}$ function is derived. This allows us to get second-order optimality conditions for nonlinearly constrained mathematical programming problems with $C^{1,1}$ data.

1. Introduction

Characterizing optimal solutions by means of second-order conditions has been a problem of continuing interest in the theory of mathematical programming. Most studies in this area are related to nonlinearly constrained problems with twice continuously differentiable data. However, in the recent past, some attempts have been made to enlarge this framework of study. Indeed, in order to get deeper results for problems with nondifferentiable data it is necessary to refine the first-order conditions, and a way to do this is to use some second-order information. The need of second-order information also appears in the point of view of numerical algorithms. Until now most works done in this direction have consisted in generalizing the classical second-order directional derivative for various classes of nondifferentiable functions; see [3, 4] and the references quoted therein.

In this paper the approach is quite different. The idea is to replace the Hessian matrix by a *set of matrices*. This has been done for the class of all $C^{1,1}$ functions, that is, for the class of all functions which are continuously differentiable and whose gradient mapping is locally Lipschitz. The set of matrices is called the *generalized Hessian matrix* and is defined as the generalized derivative of the vector-valued gradient mapping. At first sight it might appear that the difference between $C^{1,1}$ functions and C^2 functions is not significant. In fact, the class of all $C^{1,1}$ functions plays an important role and must receive the attention. Many problems from nonlinear analysis and optimization give rise to $C^{1,1}$ functions without hoping to get C^2 functions. For example this situation can occur when a penalty strategy is used for solving a constrained nonlinear programming problem. The penalty function $[\max\{f,0\}]^2$ is necessarily a $C^{1,1}$ function even if f is C^2. Another example concerns variational inequalities. It is known that some problems of this type give rise to solutions which are $C^{1,1}$ and for which it is hopeless to get more regularity; see, for example, [17, Ch. 4].

Defining the generalized Hessian matrix as the Clarke generalized derivative of the gradient mapping seems to be quite natural. This idea was suggested by Hiriart-Urruty [11, Ch. 8] as a first attempt for solving second-order problems with non-C^2 data. Araya and Gormaz [1, Ch. 3] followed this suggestion and derived *second-order sufficient* conditions for problems with $C^{1,1}$ data.

In this paper our aim is to study more extensively the properties of the generalized Hessian matrix and to derive second-order necessary conditions for $C^{1,1}$ problems. In this context, some recent results concerning the generalized derivative of a vector-valued locally Lipschitz function [14] have been revealed to be very useful. The paper is divided into three sections. In Sect. 2, the generalized Hessian matrix is defined and its properties are investigated. Then a few examples where $C^{1,1}$ functions arise in optimization or nonlinear analysis, are developed and finally some calculus rules and a second-order Taylor expansion of a $C^{1,1}$ function are given. Sect. 3 is devoted to second-order optimality conditions for problems with $C^{1,1}$ data.

To conclude this introduction we would like to mention the recent work of Ioffe [16] on $C^{1,1}$ problems, although his main concern has been the semi-infinite programming.

2. The Generalized Hessian Matrix for $C^{1,1}$ Functions

In this paper, $\mathbb{R}^p$ will be the vector space of real p-tuples with the usual inner product denoted by $\langle .,. \rangle$. The component functions of $F: \mathbb{R}^p \to \mathbb{R}^q$ are indicated by $f_1, \dots, f_q$ and $F(x)$ is assumed to be represented by the column vector $(f_1(x), \dots, f_q(x))^T$. When F is differentiable at x, the matrix representation of $F'(x)$ with respect to the canonical bases of $\mathbb{R}^p$ and $\mathbb{R}^q$ is given by the so-called Jacobian matrix denoted throughout by $JF(x)$. In the particular case where $F = f$ is real-valued, instead of a representation of $f'(x)$ by a row vector, we shall use the column vector $f'(x)^T$ which is the gradient of f at x and is denoted by $\nabla f(x)$. If the mapping ∇f itself is differentiable at x, the matrix representation of $f''(x)$ is given by the so-called Hessian matrix denoted here by $\nabla^2 f(x)$.

Throughout, the differentiability is always understood in the sense of Fréchet. Hence, for example, $\nabla^2 f(x)$ is a symmetric matrix whenever it is defined.

2.1. *Definition of $\partial^2 f$ for $C^{1,1}$ Functions f*

Let $\mathcal{O}$ be a nonempty open subset of $\mathbb{R}^p$; we denote by $C^{1,1}(\mathcal{O})$ the class of all real-valued functions f which are differentiable on $\mathcal{O}$ and whose gradient (mapping) ∇f is locally Lipschitz on $\mathcal{O}$ (i.e., satisfies a Lipschitz property in a neighborhood of each point $x_0 \in \mathcal{O}$). ∇f is therefore differentiable almost everywhere on $\mathcal{O}$ so that its generalized derivative (or generalized Jacobian matrix) in Clarke's sense [8] can be defined everywhere on $\mathcal{O}$. This is precisely the *generalized Hessian matrix* of f whose exact definition comes as follows:

Definition 2.1. Let $f \in C^{1,1}(\mathcal{O})$ and let $x_0 \in \mathcal{O}$. The generalized Hessian matrix of f at x_0, denoted by $\partial^2 f(x_0)$, is the set of matrices defined as the convex hull of the set

$\{M \mid \exists\, x_i \to x_0$ with f twice differentiable at x_i and $\nabla^2 f(x_i) \to M\}$.

The space of $p \times p$ matrices is topologized by taking a matricial norm $|||.|||$ on it. By construction itself, $\partial^2 f(x_0)$ is a *nonempty compact convex* set of *symmetric* matrices which reduces to $\{\nabla^2 f(x_0)\}$ whenever ∇f is strictly (or strongly) differentiable at x_0 (see [19, p. 71] for the definition).

Example 2.1. Let $g: \mathcal{O} \subset \mathbb{R}^p \to \mathbb{R}$ be twice continuously differentiable on $\mathcal{O}$ ($g \in C^2(\mathcal{O})$) and consider f defined on $\mathcal{O}$ as

$$f(x) = [g^+(x)]^2 \quad \text{where } g^+(x) = \max\{g(x), 0\}.$$

Clearly $f \in C^{1,1}(\mathcal{O})$ and it is easy to check that, for all $x_0 \in \mathcal{O}$,

$$\partial^2 f(x_0) = \begin{cases} \{2g(x_0)\nabla^2 g(x_0) + 2\nabla g(x_0)\nabla g(x_0)^T\} & \text{if } g(x_0) > 0, \\ \{0\} & \text{if } g(x_0) < 0, \\ \{2\alpha \nabla g(x_0)\nabla g(x_0)^T \mid \alpha \in [0,1]\} & \text{if } g(x_0) = 0. \end{cases}$$

The properties of $\partial^2 f$ are derived from those of the generalized derivative for vector-valued mappings (see [8, 14] and references therein); let us recall the two basic properties which will be of constant use in the sequel:

(P_1). The set-valued mapping $x \rightrightarrows \partial^2 f(x)$ is *locally bounded*, that is to say: there exists a neighborhood V of x_0 and a constant K such that

$$\sup\{|||M||| \mid M \in \partial^2 f(x), x \in V\} \leqslant K;$$

(P_2). $\partial^2 f$ is an *upper-semicontinuous* (or *closed*) set-valued mapping in the following sense: if $x_n \to x_0$ and if $M_n \to M_0$ with $M_n \in \partial^2 f(x_n)$ for all n, then $M_0 \in \partial^2 f(x_0)$.

The use of $\partial^2 f(x_0)$ will often arise by means of the collection of images $\partial^2 f(x_0)u = \{Mu \mid M \in \partial^2 f(x_0)\}$ for all $u \in \mathbb{R}^p$. The bivariate function

$$(u, v) \in \mathbb{R}^p \times \mathbb{R}^p \to \max\{\langle Mu, v\rangle \mid M \in \partial^2 f(x_0)\} \tag{2.1}$$

is what Hiriart-Urruty called elsewhere [14] the support bifunction of $\partial^2 f(x_0)$; we shall denote it here by $f^{00}(x_0; u, v)$. Clearly, $f^{00}(x_0; u, v)$ is symmetric in u and v and plays the role of a *generalized second-order directional derivative* for f at x_0. The next result supplies an analytic formulation of $f^{00}(x_0; u, v)$ in terms of the gradient mapping ∇f.

Theorem 2.1 [14, Thm. 2.1]. *Let $f \in C^{1,1}(\mathcal{O})$ and $x_0 \in \mathcal{O}$. Then*

$$f^{00}(x_0; u, v) = \limsup_{\substack{x \to x_0 \\ \lambda \to 0^+}} \frac{\langle \nabla f(x + \lambda u), v \rangle - \langle \nabla f(x), v \rangle}{\lambda}$$

for all $(u, v) \in \mathbb{R}^p \times \mathbb{R}^p$.

As for example, if f is like in Ex. 2.1,

$$f^{00}(x_0; u, v) = 2[\langle \nabla g(x_0), u \rangle \langle \nabla g(x_0), v \rangle]^+$$

when $g(x_0) = 0$.

2.2. *Examples Giving Rise to $C^{1,1}$ Functions*

In several areas of differential analysis or optimization, $C^{1,1}$ functions arise quite naturally. Some examples will be detailed below. Clearly, the properties of $\partial^2 f$ are inherited from those of $\nabla^2 f(x_i)$ at the points where the latter exists.

A set $\mathcal{S}$ of matrices will be said to satisfy a matricial property (π) if all the matrices $M \in \mathcal{S}$ satisfy (π).

Example 2.2. $C^{1,1}$ *convex functions.* Let $\mathcal{O}$ be a convex open set and consider $f \in C^{1,1}(\mathcal{O})$. f is known to be convex on $\mathcal{O}$ if and only if ∇f is a *monotone* mapping on $\mathcal{O}$. Translated in terms of $\partial^2 f(x)$, we get the following:

- f is convex on $\mathcal{O}$ if and only if $\partial^2 f(x)$ is positive semidefinite for all $x \in \mathcal{O}$;
- a sufficient condition for f to be strictly convex on $\mathcal{O}$ is that $\partial^2 f(x)$ be positive definite for all $x \in \mathcal{O}$.

In the first result, an equivalent condition evidently is that $\nabla^2 f(x)$ be positive semidefinite whenever f is twice-differentiable at $x \in \mathcal{O}$. As for the second result, the sufficient condition can be refined in the same way as it is done for C^2 functions ([6], [21, pp. 152–153]). The proofs are, like in the twice-differentiable case, based on second-order Taylor expansions (cf. 2.3 for details).

Since a convex function is C^1 on $\mathcal{O}$ whenever it is differentiable on $\mathcal{O}$, the local Lipschitz property on ∇f is a mild requirement on such a function. A classical example of a differentiable convex function f which is not $C^{1,1}$ on $\mathbb{R}$ is $f(x) = |x|^{3/2}$.

Example 2.3. Primitive functions of the proximal mapping. Let $f: \mathbb{R}^p \to (-\infty, +\infty]$ be a lower-semicontinuous convex function which is not identically equal to $+\infty$. The unique point where the function $u \to f(u) + \frac{1}{2}\|x - u\|^2$ achieves its minimum value is called the "proximal point of x relative to f" and denoted by $\mathrm{prox}_f x$. Among the main properties of the proximal mapping prox_f (for a thorough account of these facts, see Moreau's pithy paper [18]), there is one which is of particular importance and which says that prox_f *is a gradient mapping* satisfying

$$\|\mathrm{prox}_f x - \mathrm{prox}_f x'\| \leqslant \|x - x'\|, \qquad x, x' \in \mathbb{R}^p.$$

Moreau gave a detailed account of the relations between $\mathrm{prox}_f x$ and $\mathrm{prox}_{f^*} x$ (i.e., relative to the conjugate function f^*). In particular, he proved that the *convex* function

$$\begin{aligned}\phi : x \mapsto \phi(x) &= \tfrac{1}{2}\|\mathrm{prox}_f x\|^2 + f^*(\mathrm{prox}_{f^*} x)\\ &= \tfrac{1}{2}\left[\|x\|^2 - \|x - \mathrm{prox}_f x\|^2\right] - f(\mathrm{prox}_f x)\end{aligned} \tag{2.2}$$

is differentiable on all of $\mathbb{R}^p$ with

$$\nabla\phi(x) = \mathrm{prox}_f x, \qquad x \in \mathbb{R}^p. \tag{2.3}$$

Thus ϕ is a $C^{1,1}$ convex function to which we refer as the primitive function of prox_f. As a consequence, we observe that the ϕ_n defined by

$$\phi_n(x) = \inf_{u \in \mathbb{R}^p} \left\{ f(u) + \frac{1}{2n}\|x - u\|^2 \right\}, \qquad x \in \mathbb{R}^p, \tag{2.4}$$

yield a sequence of $C^{1,1}$ convex functions since, following (2.2) and (2.3),

$$\nabla\phi_n(x) = \frac{1}{n}(x - \bar{x}) \in \partial f(\bar{x}), \qquad x \in \mathbb{R}^p,$$

where $\bar{x}$ is the unique solution of the minimization problem (2.4). It is a classical result that the ϕ_n converge pointwise to f [5].

If f is the indicator function ψ_Q of a (nonempty) closed convex set Q (i.e., $\psi_Q(x) = 0$ if $x \in Q$, $+\infty$ if not), prox_f is nothing else than the projection mapping P_Q. According to (2.2), the primitive function of P_Q is defined as

$$x \mapsto \phi(x) = \tfrac{1}{2}\left[\|x\|^2 - d_Q^2(x)\right].$$

The elements M of $\partial^2\phi(x)$ (or equivalently those of $\partial^2 d_Q^2(x)$) enjoy some elementary properties which do not depend on Q. Clearly

$$\partial^2\phi(x) = I_p - \tfrac{1}{2}\partial^2 d_Q^2(x), \qquad x \in \mathbb{R}^p,$$

and since both ϕ and d_Q^2 are convex, we have that

$$\begin{aligned}&0 \leqslant \langle Mh, h\rangle \leqslant \|h\|^2, \qquad h \in \mathbb{R}^p,\\ &M\left[x - P_Q(x)\right] = 0,\end{aligned}$$

whatever $M \in \partial^2\phi(x)$. Although it is, as a general rule, difficult to calculate $\partial^2\phi(x)$ explicitly, the main properties of the $M \in \partial^2\phi(x)$ can be derived from the properties of the projection mapping P_Q. The question whether ϕ might be C^2 on $\mathcal{O} = \mathbb{R}^p \setminus Q$ (hence $\partial^2\phi(x) = \{\nabla^2\phi(x)\}$ for all $x \in \mathcal{O}$) has been settled by Holmes [15] and, more recently, by Fitzpatrick and Phelps [9]. They proved that, to secure that ϕ is C^2 on $\mathcal{O}$, it is (almost) necessary and certainly sufficient to assume that the boundary of Q is C^2.

Example 2.4. Augmented Lagrangians. Consider the following minimization problem:

$$\text{Minimize } f_0(x) \text{ over all } x \in \mathbb{R}^p \text{ such that } f_1(x) \leqslant 0,\dots,f_m(x) \leqslant 0. \qquad \text{(P)}$$

Letting r denote a positive parameter, the augmented Lagrangian L_r (see [20] and references therein) is defined on $\mathbb{R}^p \times \mathbb{R}^m$ as

$$L_r(x, y) = f_0(x) + \frac{1}{4r}\sum_{i=1}^{m} \{[y_i + 2rf_i(x)]^+\}^2 - y_i^2, \qquad x \in \mathbb{R}^p, \quad y \in \mathbb{R}^m. \tag{2.5}$$

From the general theory of duality which yields L_r as a particular Lagrangian, we know that $L_r(x,.)$ is concave and also that $L_r(., y)$ is convex whenever (P) is a convex minimization problem. By stating $y = 0$ in (2.5), we observe that

$$L_r(x,0) = f_0(x) + r\sum_{i=1}^{m} [f_i^+(x)]^2$$

is the ordinary penalized version of problem (P). L_r is differentiable everywhere on $\mathbb{R}^p \times \mathbb{R}^m$ with

$$\nabla_x L_r(x, y) = \nabla f_0(x) + \sum_{j=1}^{m} [y_j + 2rf_j(x)]^+ \nabla f_j(x),$$

$$\frac{\partial L_r}{\partial y_i}(x, y) = \max\left\{f_i(x), \frac{-y_i}{2r}\right\}, \qquad i = 1,\dots,m.$$

When the f_i are C^2 on $\mathbb{R}^p$, L_r is $C^{1,1}$ on $\mathbb{R}^{p+m}$ and estimates of the generalized Hessian matrices $\partial^2_{xx}L_r(x, y)$, $\partial^2_{yy}L_r(x, y)$, $\partial^2 L_r(x, y)$ of $L_r(., y)$, $L_r(x,.)$ and $L_r(.,.)$ evaluated at x, y, and (x, y), respectively, along the surfaces defined by the equations $y_i + 2rf_i(x) = 0$, are provided by the rules we shall display in the next paragraph.

The dual problem to (P) corresponding to L_r is by definition

$$\text{Maximize } g_r(y) \text{ over } \mathbb{R}^m, \text{ where } g_r(y) = \inf_{x \in \mathbb{R}^p} L_r(x, y). \qquad (\text{D}_r)$$

In the convex case with $r > 0$, g_r is again a $C^{1,1}$ concave function [20, Thm. 14] with the following uniform Lipschitz property on ∇g_r

$$\| \nabla g_r(y) - \nabla g_r(y')\| \leqslant \frac{1}{2r}\|y - y'\|, y, y' \in \mathbb{R}^m.$$

2.3. *Calculus Rules on $\partial^2 f$*

Chain rules. Let $F = (f_1, \ldots, f_q)^T : \mathcal{O} \subset \mathbb{R}^p \to \mathbb{R}^q$ be a (vector-valued) function with each component $f_i \in C^{1,1}(\mathcal{O})$. Assuming that $\phi : \mathcal{O}' \subset \mathbb{R}^q \to \mathbb{R}$ is $C^{1,1}$ on an open set $\mathcal{O}'$ containing $F(\mathcal{O})$, it is clear that the composed function $\phi \circ F$ is $C^{1,1}$ on $\mathcal{O}$. A general estimate of $\partial^2(\phi \circ F)(x_0)u, u \in \mathbb{R}^p$, in terms of the $\partial^2 f_i(x_0)u$, $\partial^2\phi(F(x_0))$ and $JF(x_0)$ can be derived via inequalities between generalized second-order directional derivatives (cf. Thm. 2.1). The next result gives such an inequality.

Theorem 2.2. *Let $x_0 \in \mathcal{O}$. Then for all u, v in $\mathbb{R}^p$, we have that*

$$(\phi \circ F)^{00}(x_0; u, v) \leqslant \sum_{i=1}^{q} \frac{\partial \phi}{\partial y_i}(F(x_0)) f_i^{00}(x_0; u, v) + \phi^{00}(F(x_0); JF(x_0)u, JF(x_0)v). \tag{2.6}$$

Equality holds whenever either $f_i \in C^2(\mathcal{O})$ for all i, or $\phi \in C^2(\mathcal{O}')$ and $q = 1$.

Since $f^{00}(x_0; u, v)$ is by definition the support function of $\partial^2 f(x_0)\, u$ in the v direction, we deduce the following from the theorem above:

$$\partial^2(\phi \circ F)(x_0)u \subset \sum_{i=1}^{q} \frac{\partial \phi}{\partial y_i}(F(x_0))\, \partial^2 f_i(x_0)u + JF(x_0)^T \partial^2\phi(F(x_0)) JF(x_0)u \tag{2.7}$$

for all $u \in \mathbb{R}^p$. Equality holds if all the f_i are C^2 on $\mathcal{O}$. However, the above inclusion, although it holds true for all $u \in \mathbb{R}^p$, does not allow us to derive an inclusion between the sets of matrices! See the discussion in that respect in [14, Sect. II].

Before proving Thm. 2.2, let us illustrate the foregoing results with the aid of some examples.

Example 2.5. Let $F: \mathbb{R}^p \to \mathbb{R}^q$ be a linear mapping represented as $F(x) = Ax$ for a certain $q \times p$ matrix A. Then

$$(\phi \circ A)^{00}(x_0; u, v) = \phi^{00}(Ax_0; Au, Av)$$

for all $u, v \in \mathbb{R}^p$, and

$$\partial^2(\phi \circ A)(x_0)u = A^T \partial^2\phi(Ax_0)Au$$

for all $u \in \mathbb{R}^p$.

Example 2.6. Let $\phi:\mathbb{R}^q\to\mathbb{R}$ be a linear form defined by $\phi(y)=\langle a, y\rangle$ for a certain $a\in\mathbb{R}^q$. Then

$$\partial^2(\langle a, F\rangle)(x_0)u \subset \sum_{i=1}^{q} a_i\partial^2 f_i(x_0)u$$

for all $u\in\mathbb{R}^p$.

Example 2.7. Let $f\in C^{1,1}(\mathcal{O})$ and let $\phi\in C^2(I')$ with $f(\mathcal{O})$ included in the open interval I' of $\mathbb{R}$. For such a case, we have claimed in Thm. 2.2 that

$$\begin{aligned}(\phi\circ f)^{00}(x_0; u, v) &= \phi'(f(x_0))f^{00}(x_0; u, v)\\ &\quad + \phi''(f(x_0))\langle \nabla f(x_0), u\rangle\langle \nabla f(x_0), v\rangle\end{aligned}$$

for all u, v in $\mathbb{R}^p$. Therefore

$$\partial^2(\phi\circ f)(x_0)u = \phi'(f(x_0))\,\partial^2 f(x_0)u + \phi''(f(x_0))\nabla f(x_0)^T\nabla f(x_0)u$$

for all $u\in\mathbb{R}^p$.

Proof of Thm. 2.2. Let $\Delta_{x,\lambda}(u,v)$ be defined in a neighborhood of $x_0\in\mathcal{O}$ as

$$\Delta_{x,\lambda}(u,v) = \frac{\langle \nabla(\phi\circ F)(x+\lambda u), v\rangle - \langle \nabla(\phi\circ F)(x), v\rangle}{\lambda}.$$

Since $\nabla(\phi\circ F)(x)=JF(x)^T\nabla\phi(F(x))$ for all $x\in\mathcal{O}$, we have that

$$\begin{aligned}\Delta_{x,\lambda}(u,v) &= \frac{\langle [JF(x+\lambda u)-JF(x)]^T\nabla\phi(F(x+\lambda u)), v\rangle}{\lambda}\\ &\quad + \frac{\langle JF(x)^T[\nabla\phi(F(x+\lambda u))-\nabla\phi(F(x))], v\rangle}{\lambda}.\end{aligned} \tag{2.8}$$

F has been assumed to be C^1 on $\mathcal{O}$. Therefore,

$$F(x+\lambda u) = F(x)+\lambda JF(x_0)u + \lambda\varepsilon(x,\lambda).$$

In this development as in the sequel, $\varepsilon(x,\lambda)$ will denote generically an expression such that

$$\lim_{\substack{x\to x_0\\ \lambda\to 0^+}} \varepsilon(x,\lambda) = 0.$$

Now, since $\nabla\phi$ is Lipschitz in a neighborhood of $F(x_0)$, we have that

$$\nabla\phi(F(x+\lambda u)) = \nabla\phi(F(x)+\lambda JF(x_0)u) + \lambda\varepsilon(x,\lambda).$$

Due to this expansion and to the Lipschitz property of JF around x_0, the second expression occurring in the development (2.8) can be rewritten as

$$\frac{\langle \nabla\phi(F(x)+\lambda JF(x_0)u)-\nabla\phi(F(x)), JF(x_0)v\rangle}{\lambda}+\varepsilon(x,\lambda).$$

As for the first expression in (2.8), for the same reasons as above, we note that

$$\begin{aligned}&[JF(x+\lambda u)-JF(x)]^T\nabla\phi(F(x+\lambda u))\\ &\quad=\sum_{i=1}^{q}\frac{\partial\phi}{\partial y_i}(F(x_0))[\nabla f_i(x+\lambda u)-\nabla f_i(x)]+\lambda\varepsilon(x,\lambda).\end{aligned}$$

As a result,

$$\begin{aligned}\Delta_{x,\lambda}(u,v) &= \sum_{i=1}^{q}\frac{\partial\phi}{\partial y_i}(F(x_0))\frac{\nabla f_i(x+\lambda u)-\nabla f_i(x)}{\lambda}\\ &\quad+\frac{\langle \nabla\phi(F(x)+\lambda JF(x_0)u)-\nabla\phi(F(x)), JF(x_0)v\rangle}{\lambda}\\ &\quad+\varepsilon(x,\lambda). \qquad (2.9)\end{aligned}$$

Now, since $(\phi\circ F)^{00}(x_0;\, u,v)$ is equal to

$$\limsup_{\substack{x\to x_0\\ \lambda\to 0^+}}\Delta_{x,\lambda}(u,v)$$

(Thm. 2.2), one readily gets the inequality (2.6) from (2.9). If all the f_i are C^2, or if ϕ is C^2 and $q=1$, one of the two main expressions in (2.9) actually has a limit when $x\to x_0$ and $\lambda\to 0^+$. Whence the announced equalities. □

Second-order Taylor Expansion. Let I be an open interval containing $[0,1]$ and let $\phi\in C^{1,1}(\mathrm{I})$. Then, a classical argument yields that

$$\phi(1)-\phi(0)-\phi'(0)\in\tfrac{1}{2}\partial^2\phi(t) \qquad (2.10)$$

for some $t\in\,]0,1[$. By using the chain rule displayed in Ex. 2.5, one immediately gets the next second-order expansion of f:

Theorem 2.3. *Let $f\in C^{1,1}(\mathcal{O})$ and let $[a,b]\subset\mathcal{O}$. Then there exist $c\in\,]a,b[$ and $M_c\in\partial^2 f(c)$ such that*

$$f(b)=f(a)+\langle\nabla f(a),b-a\rangle+\tfrac{1}{2}\langle M_c(b-a),b-a\rangle. \qquad (2.11)$$

3. Second-Order Optimality Conditions for $C^{1,1}$ Problems

In this section, our aim is to present second-order necessary conditions for a point x_0 to be an optimal solution of a mathematical programming problem with $C^{1,1}$ data. In particular we show how to replace the Hessian matrix of the Lagrangian function when it is not defined.

3.1. Necessary Conditions for Unconstrained Problems

We consider the following problem:

$$\text{Minimize } f(x) \quad \text{over all } x \in \mathcal{O} \tag{U}$$

where $\mathcal{O}$ is an open subset of $\mathbb{R}^p$ and f a $C^{1,1}$ function on $\mathcal{O}$. For this problem the second-order necessary conditions are given in the theorem below.

Theorem 3.1. *Let $x_0 \in \mathcal{O}$ be a local minimum for problem* (U). *Then for each $d \in \mathbb{R}^p$, there exists a matrix $A \in \partial^2 f(x_0)$ such that $\langle Ad, d\rangle \geqslant 0$.*

Proof. Let $d \in \mathbb{R}^p$ be fixed and consider the sequence $\{x_k\}_{k \geqslant 1}$ where $x_k = x_0 + \frac{1}{k}d$. Without loss of generality we can suppose that $x_k \in \mathcal{O}$ for each k. Then the second-order Taylor expansion (Thm. 2.3) of f in a neighborhood of x_0 becomes:

$$f(x_k) = f(x_0) + \frac{1}{k}\langle \nabla f(x_0), d\rangle + \frac{1}{2k^2}\langle A_k d, d\rangle \tag{3.1}$$

where $k \geqslant 1$, $A_k \in \partial^2 f(\bar{x}_k)$ and $\bar{x}_k \in]x_0, x_k[$.

Now, x_0 being a local minimum, the gradient $\nabla f(x_0)$ is identically zero and $f(x_0) \leqslant f(x_k)$ for each k. Hence it follows from (3.1) that

$$\langle A_k d, d\rangle \geqslant 0 \quad \text{for each } k. \tag{3.2}$$

On the other hand, as $\partial^2 f$ is locally bounded (Sect. 2, (P_1)), the sequence $\{A_k\}_k$ is bounded and thus has a subsequence which converges. Let A be the limit of this subsequence. By the upper-semicontinuity of $\partial^2 f$ (Sect. 2, (P_2)), this matrix $A \in \partial^2 f(x_0)$. Moreover, taking the limit in (3.2) gives $\langle Ad, d\rangle \geqslant 0$. □

Remark 3.1. In Thm. 3.1 it is only assured that there exists a matrix $A \in \partial^2 f(x_0)$ such that $\langle Ad, d\rangle \geqslant 0$. In general this inequality cannot be extended to each matrix $A \in \partial^2 f(x_0)$ as illustrated by the following example. Let $f: \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \int_0^{|x|} \phi(t)\, dt \tag{3.3}$$

where

$$\phi(t) = \begin{cases} 2t^2 + t^2 \sin 1/t & \text{if } t > 0, \\ 0 & \text{if } t = 0. \end{cases}$$

The function ϕ has been chosen to satisfy the inequality $\phi(t) \geqslant t^2$. So $f(x) \geqslant |x|^3/3$ for each $x \in \mathbb{R}$ and as $f(0)=0$ it is clear that 0 is a local minimum of f. Moreover $\partial^2 f(0)=[-1,+1]$ and consequently $\langle Ad, d\rangle \ngeqslant 0$ for each $A \in \partial^2 f(0)$.

It would be interesting to characterize the largest subclass of $C^{1,1}$ functions for which the necessary optimality conditions can be written:

$$\min_{A \in \partial^2 f(x_0)} \langle Ad, d\rangle \geqslant 0 \quad \text{for each } d. \tag{3.4}$$

This is the case for the functions described in Ex. 2.1. Indeed for each $x_0 \in \mathbb{R}^p$ such that $g(x_0)=0$ (and x_0 is then a minimum of f) the generalized Hessian $\partial^2 f(x_0)$ is a subset of the set of symmetric positive semidefinite matrices. At this time it appears that the characterization of the class of functions satisfying (3.4) remains an open question.

Remark 3.2. As (3.4) is not true in general, the set $\mathscr{A}(d)=\{A \in \partial^2 f(x_0) | \langle Ad, d\rangle \geqslant 0\}$, i.e., the maximal subset of $\partial^2 f(x_0)$ satisfying $\langle Ad, d\rangle \geqslant 0$, depends on d. This can be illustrated by the following example. Let $g: \mathbb{R}^2 \to \mathbb{R}$ be defined by $g(x, y)=f(x)+f(y)$ where f is the function considered in (3.3). It is easy to see that $x_0=(0,0)$ is a minimum of g and that, for $d=(d_1, d_2)$,

$$\mathscr{A}(d) = \left\{\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \middle| -1 \leqslant \alpha \leqslant 1,\ -1 \leqslant \beta \leqslant 1,\ \alpha d_1^2 + \beta d_2^2 \geqslant 0\right\}.$$

Then $\mathscr{A}((1,0)) \neq \mathscr{A}((0,1))$. Moreover, for this example the set $\bigcap_d \mathscr{A}(d)$ is nonempty.

In general, however, it is not sure that the set $\bigcap_d \mathscr{A}(d)$ is nonempty. For one-dimensional problems it is obvious. For the other cases the question remains open.

3.2. *Necessary Conditions for Constrained Problems*

We now consider the following constrained problem:

$$\left.\begin{array}{lr} \text{minimize}\, f(x) & \\ s.t.\ g_i(x) \leqslant 0, & i=1,\ldots,m, \\ \quad h_j(x)=0, & j=1,\ldots,n, \end{array}\right\} \tag{C}$$

where the functions f, g_i, $i=1,\ldots,m$, and h_j, $j=1,\ldots,n$, are $C^{1,1}$ functions on $\mathbb{R}^p$.

Let x_0 be a local minimum for problem (C). The functions f, g_i, and h_j being C^1 functions, there exists a vector $(\lambda_0, \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n)$ in $\mathbb{R}^{n+m+1}$ *not identically zero* such that the following condition, known as the Fritz-John

Necessary Condition,

$$\left.\begin{aligned}&\lambda_0\nabla f(x_0)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x_0)+\sum_{j=1}^{n}\mu_j\nabla h_j(x_0)=0,\\&\lambda_0\geqslant 0,\quad \lambda_i\geqslant 0\ \text{ and }\ \lambda_i g_i(x_0)=0\ \text{ for } i=1,\dots,m,\end{aligned}\right\}\tag{FJ}$$

is satisfied.

Now if we want the multiplier λ_0 to be positive then an additional regularity hypothesis, called a *constraint qualification*, must be assumed. We denote by (H_1) any constraint qualification. So, under (H_1), the Fritz-John's Condition becomes: There exists a vector $(\lambda_1,\dots,\lambda_m,\mu_1,\dots,\mu_n)$ in R^{m+n} such that the following condition:

$$\left.\begin{aligned}&\nabla f(x_0)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x_0)+\sum_{j=1}^{n}\mu_j\nabla h_j(x_0)=0,\\&\lambda_i\geqslant 0\ \text{ and }\ \lambda_i g_i(x_0)=0\ \text{ for } i=1,\dots,m,\end{aligned}\right\}\tag{KT}$$

is satisfied. It is the Kuhn-Tucker Optimality Condition. The linear independence of the vectors $\nabla g_i(x_0)$, $i\in I(x_0)$, $\nabla h_j(x_0)$, $j=1,\dots,n$, where $I(x_0)=\{i|g_i(x_0)=0\}$, is an example of Condition (H_1).

From now on, we assume that (H_1) is satisfied. So there exists at least one Kuhn-Tucker multiplier (λ,μ). Then, to get second-order necessary conditions, we associate to each multiplier $\lambda=(\lambda_1,\dots,\lambda_m)$ the set $G(\lambda)$ defined as follows:

$$G(\lambda)=\left\{x\ \middle|\ \begin{aligned}&g_i(x)=0 &&\text{for } i \text{ such that } \lambda_i>0\\&g_i(x)\leqslant 0 &&\text{for } i \text{ such that } \lambda_i=0\\&h_j(x)=0 &&\text{for all } j\end{aligned}\right\}\tag{3.5}$$

and we denote by T_λ the Bouligand tangent cone to $G(\lambda)$ at x_0. If we denote by $L(x,\lambda,\mu)$ the usual Lagrangian function

$$L(x,\lambda,\mu)=f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)+\sum_{j=1}^{n}\mu_j h_j(x)\tag{3.6}$$

and by $\partial^2_{xx}L(x_0,\lambda,\mu)$ the generalized Hessian matrix of $L(.,\lambda,\mu)$ at x_0, then the second-order necessary conditions can be expressed as follows:

Theorem 3.2. *Let x_0 be a local minimum of problem* (C) *and let (H_1) be assumed. Then for each Kuhn-Tucker multiplier vector (λ,μ) and for each $d\in T_\lambda$, there exists a matrix $A\in\partial^2_{xx}L(x_0,\lambda,\mu)$ such that $\langle Ad,d\rangle\geqslant 0$.*

Proof. Let λ, μ and d be fixed. By definition of the tangent cone T_λ, there exist sequences $\{\alpha_k\}\to 0^+$ and $\{d_k\}\to d$ such that, for every k, $x_0+\alpha_k d_k\in G(\lambda)$.

Then, by (3.5) and (3.6), we have for every k, that

$$L(x_0+\alpha_k d_k, \lambda, \mu) = f(x_0+\alpha_k d_k). \tag{3.7}$$

On the other hand, by Thm. 2.3, the Lagrangian function admits a second-order Taylor expansion in a neighborhood of x_0, namely,

$$L(x_0+\alpha_k d_k, \lambda, \mu) = L(x_0, \lambda, \mu) + \alpha_k\langle \nabla_x L(x_0, \lambda, \mu), d_k\rangle + \tfrac{1}{2}\alpha_k^2\langle A_k d_k, d_k\rangle \tag{3.8}$$

where $k \geqslant 1$, $A_k \in \partial^2_{xx} L(x_0 + \overline{\alpha_k} d_k, \lambda, \mu)$, and $\overline{\alpha_k}$ verifies $0 < \overline{\alpha_k} < \alpha_k$. But x_0 is a local minimum for problem (C) and thus the following relations

$$\left.\begin{array}{ll} L(x_0, \lambda, \mu) = f(x_0), & \nabla_x L(x_0, \lambda, \mu) = 0, \\ f(x_0+\alpha_k d_k) \geqslant f(x_0) & \text{for } k \text{ sufficiently large} \end{array}\right\} \tag{3.9}$$

are satisfied. Then, gathering (3.8) and (3.9) gives:

$$\langle A_k d_k, d_k\rangle \geqslant 0 \quad \text{for } k \text{ sufficiently large.} \tag{3.10}$$

Finally, as $\partial^2 L$ is locally bounded (Sect. 2, (P_1)), the sequence $\{A_k\}$ is bounded and thus has a subsequence, again denoted by $\{A_k\}$, which converges to a matrix A. Now as $\{\alpha_k\} \to 0^+$ and $\{d_k\} \to d$, the sequence $\{x_0 + \overline{\alpha_k} d_k\}$ converges to x_0 and, by the upper-semicontinuity of $\partial^2 L$ (Sect. 2, (P_2)), the matrix $A \in \partial^2_{xx} L(x_0, \lambda, \mu)$. Passing to the limit in (3.10) gives the inequality $\langle Ad, d\rangle \geqslant 0$. □

Remark 3.3. If (H_1) is not assumed, then there exist $\lambda_0, \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n$ not all zero satisfying the Fritz-John Condition and if the Lagrangian function is defined by

$$L(x, \lambda_0, \lambda, \mu) = \lambda_0 f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{n} \mu_j h_j(x)$$

then a similar result can be obtained, namely, for each $d \in T_\lambda$, there exists a matrix $A \in \partial^2_{xx} L(x_0, \lambda_0, \lambda, \mu)$ such that $\langle Ad, d\rangle \geqslant 0$.

One way to obtain a more tractable expression of the tangent cone T_λ is to express it in terms of the gradients of the functions g_i and h_j. It is easy to see that

$$T_\lambda \subseteq \left\{ d \,\middle|\, \begin{array}{ll} \langle \nabla g_i(x_0), d\rangle = 0 & \text{for } i \text{ such that } \lambda_i > 0 \\ \langle \nabla g_i(x_0), d\rangle \leqslant 0 & \text{for } i \text{ such that } \lambda_i = 0 \text{ and } g_i(x_0) = 0 \\ \langle \nabla h_j(x_0), d\rangle = 0 & \text{for } j = 1, \ldots, n \end{array} \right\}.$$

But, in order to get the equality between these two sets, a second-order regularity condition must be imposed. Let (H_2) be any condition of this type. It is well known that both (H_1) and (H_2) are satisfied if the following vectors

$$\nabla g_i(x_0), \quad i \in I(x_0), \quad \nabla h_j(x_0),\ j = 1, \ldots, n,$$

are linearly independent (see, for example, [10, Sect. 7 and 10]).

Corollary 3.1. *Let x_0 be a local minimum for problem (C) and let (H_1) and (H_2) be assumed. Then for each multiplier (λ, μ) and for each d such that*

$$\begin{cases} \langle \nabla g_i(x_0), d\rangle = 0 & \text{for } i \text{ such that } \lambda_i > 0, \\ \langle \nabla g_i(x_0), d\rangle \leqslant 0 & \text{for } i \text{ such that } \lambda_i = 0 \text{ and } g_i(x_0) = 0, \\ \langle \nabla h_j(x_0), d\rangle = 0 & \text{for } j = 1, \ldots, n, \end{cases}$$

there exists a matrix $A \in \partial^2_{xx} L(x_0, \lambda, \mu)$ such that $\langle Ad, d\rangle \geqslant 0$.

References

1. Araya Schulz R, Gormaz Arancibia R (1979) Problemas localmente Lipschitzianos en optimizacion. (Master's thesis) University of Chile at Santiago
2. Auslender A (1979) Penalty methods for computing points that satisfy second order necessary conditions. Math Programming 17:229–238
3. Auslender A (1981) Stability in mathematical programming with nondifferentiable data; second order directional derivative for lower-C^2 functions.
4. Ben-Tal A, Zowe J (1982) Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems. Math Programming 24:70–91
5. Brezis H (1973) Opérateurs maximaux monotones. North-Holland, Amsterdam
6. Bernstein B, Toupin R (1962) Some properties of the Hessian matrix of a strictly convex function. Journal für die Reine und Angewandte Mathematik 210:65–72
7. Clarke FH (1975) Generalized gradients and applications. Trans Amer Math Soc 205:247–262
8. Clarke FH (1976) On the inverse function theorem. Pacif J Math 64:97–102
9. Fitzpatrick S, Phelps R (to appear) Differentiability of the metric projection in Hilbert spaces.
10. Hestenes M (1975) Optimization theory: The finite dimensional case. Wiley, New York
11. Hiriart-Urruty J-B (1977) Contributions à la programmation mathématique: Cas déterministe et stochastique. (Doctoral thesis) Sciences Mathématiques, Université de Clermont-Ferrand II
12. Hiriart-Urruty J-B (1979) Refinements of necessary optimality conditions in nondifferentiable programming I. Appl Math Optim 5:63–82
13. Hiriart-Urruty J-B (1982a) Refinements of necessary optimality conditions in nondifferentiable programming II. Math Programming Study 19:120–139
14. Hiriart-Urruty J-B (1982b) Characterizations of the plenary hull of the generalized Jacobian matrix. Math Programming Study 17:1–12
15. Holmes R (1973) Smoothness of certain metric projections on Hilbert spaces. Trans Amer Math Soc 184:87–100
16. Ioffe A (1981) Second order conditions in nonlinear nonsmooth problems of semi-infinite programming.
17. Kinderlehrer D, Stampacchia G (1980) An introduction to variational inequalities and their applications. Academic Press, New York
18. Moreau J-J (1965) Proximité et dualité dans un espace hilbertien. Bulletin de la Société Mathématique de France 93:273–299
19. Ortega JM, Rheinboldt WC (1970) Iterative solution of nonlinear equations in several variables. Academic Press, New York
20. Rockafellar RT (1976) Solving a nonlinear programming problem by way of a dual problem. Symposia Mathematica 19:135–160
21. Stoer J, Witzgall C (1970) Convexity and optimization in finite dimensions I. Springer-Verlag, Berlin

Accepted 10 August 1983