Daidai *Journal of Inequalities and Applications* (2017) 2017:212
DOI 10.1186/s13660-017-1487-8

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

Second-order optimality conditions for nonlinear programs and mathematical programs

Ikram Daidai*

*Correspondence:
ikram_daidai_57@yahoo.fr
Department of Mathematics,
Faculté des Sciences et Techniques,
University Cadi Ayyad, B.P. 549,
Marrakech, Morocco

Abstract

It is well known that second-order information is a basic tool notably in optimality conditions and numerical algorithms. In this work, we present a generalization of optimality conditions to strongly convex functions of order γ with the help of first- and second-order approximations derived from (Optimization 40(3):229-246, 2011) and we study their characterization. Further, we give an example of such a function that arises quite naturally in nonlinear analysis and optimization. An extension of Newton's method is also given and proved to solve Euler equation with second-order approximation data.

Keywords: strong convexity of order γ; second-order approximation; $C^{1,1}$ functions; Newton's method

1 Introduction

The concept of approximations of mappings was introduced by Thibault [2]. Sweetser [3] considered approximations by subsets of the space of continuous linear maps $L(X, Y)$, where X and Y are Banach spaces, and Ioffe [4] by the so-called fans. This approach was revised by Jourani and Thibault [5]. Another approach belongs to Allali and Amahroq [1]. Following the same ideas, Amahroq and Gadhi [6, 7] have established optimality conditions to some optimization problems under set-valued mapping constraints.

In this work, we explore the notion of strongly convex functions of order γ; see, for instance, [8–15] and references therein. Let f be a mapping from a Banach space X into $\mathbb{R}$, and let $C \subset X$ be a closed convex set. It is well known that the notion of strong convexity plays a central role. On the one hand, it ensures the existence and uniqueness of the optimal solution for the problem

$$(\mathcal{P}) \qquad \min_{x \in C} f(x).$$

On the other hand, if f is twice differentiable, then the strong convexity of f implies that its Hessian matrix is nonsingular, which is an important tool in numerical algorithms. Here we adopt the definition of a second-order approximation [1] to detect some equivalent properties of strongly convex functions of order γ and to characterize the latter. Further-more, for a $C^{1,1}$ function f on a finite-dimensional setting, we show some simple facts. We also provide an extension of Newton's method to solve an Euler equation with second-order approximation data.

The rest of the paper is written as follows. Section 2 contains basic definitions and preliminary results. Section 3 is devoted to mains results. In Section 4, we point out an extension of Newton's method and prove its local convergence.

2 Preliminaries

Let X and Y be two Banach spaces. We denote by $\mathcal{L}(X, Y)$ the set of all continuous linear mappings from X into Y, by $\mathcal{B}(X \times X, Y)$ the set of all continuous bilinear mappings from $X \times X$ into Y, and by $\mathbb{B}_Y$ the closed unit ball of Y centered at the origin.

Throughout this paper, X^* and Y^* denote the continuous duals of X and Y, respectively, and we write $\langle \cdot, \cdot \rangle$ for the canonical bilinear forms with respect to the dualities $\langle X^*, X\rangle$ and $\langle Y^*, Y\rangle$.

Definition 1 ([1]) Let f be a mapping from X into Y, $\bar{x} \in X$. A set of mappings $\mathcal{A}_f(\bar{x}) \subset \mathcal{L}(X, Y)$ is said to be a first-order approximation of f at $\bar{x}$ if there exist $\delta > 0$ and a function $r : X \to \mathbb{R}$ satisfying $\lim_{x\to\bar{x}} r(x) = 0$ such that

$$f(x) - f(\bar{x}) \in \mathcal{A}_f(\bar{x})(x - \bar{x}) + \|x - \bar{x}\| r(x)\mathbb{B}_Y \tag{1}$$

for all $x \in \bar{x} + \delta\mathbb{B}_X$.

It is easy to check that Definition 1 is equivalent to the following: for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$f(x) - f(\bar{x}) \in \mathcal{A}_f(\bar{x})(x - \bar{x}) + \varepsilon\|x - \bar{x}\|\mathbb{B}_Y \tag{2}$$

for all $x \in \bar{x} + \delta\mathbb{B}_X$.

Remark 1 If $\mathcal{A}_f(\bar{x})$ is a first-order approximation of f at $\bar{x}$, then (2) means that for any $x \in \bar{x} + \delta\mathbb{B}_X$, there exist $A(x) \in \mathcal{A}_f(\bar{x})$ and $b \in \mathbb{B}_Y$ such that

$$f(x) - f(\bar{x}) = A(x)(x - \bar{x}) + \varepsilon\|x - \bar{x}\|b.$$

Hence, for any $x \in \mathbb{B}(\bar{x}, \delta)$ and $A(x) \in \mathcal{A}_f(\bar{x})$,

$$\big\|f(x) - f(\bar{x}) - A(x)(x - \bar{x})\big\| \le \varepsilon\|x - \bar{x}\|. \tag{3}$$

If $\mathcal{A}_f(\bar{x})$ is norm-bounded (resp. compact), then it is called a bounded (resp. compact) first-order approximation. Recall that $\mathcal{A}_f(\bar{x})$ is a singleton if and only if f is Fréchet differentiable at $\bar{x}$.

The following proposition proved by Allali and Amahroq [1] plays an important role in the sequel in a finite-dimensional setting.

Proposition 1 ([1]) *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a locally Lipschitz function at $\bar{x}$. Then the Clarke subdifferential of f at $\bar{x}$,*

$$\partial_c f(\bar{x}) := \operatorname{co}\bigl\{\lim \nabla f(x_n) : x_n \in \operatorname{dom} \nabla f \text{ and } x_n \to \bar{x}\bigr\}, \tag{4}$$

is a first-order approximation of f at $\bar{x}$.

In [6], it is also shown that when f is a continuous function, it admits as an approximation the symmetric subdifferential defined and studied in [16].

The next proposition shows that Proposition 1 holds also when f is a vector-valued function. Let us first recall the definition of the generalized Jacobian for a vector-valued function (see [17, 18] for more details) and the definition of upper semicontinuity.

Definition 2 The generalized Jacobian of a function $g : \mathbb{R}^p \to \mathbb{R}^q$ at $\bar{x}$, denoted $\partial_c g(\bar{x})$, is the convex hull of all matrices M of the form

$$M = \lim_{n \to +\infty} Jg(x_n),$$

where $x_n \to \bar{x}$, g is differentiable at x_n for all n, and Jg denotes the $q \times p$ usual Jacobian matrix of partial derivatives.

Definition 3 A set-valued mapping $F : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is said to be upper semicontinuous at a point $\bar{x} \in \mathbb{R}^p$ if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$F(x) \subset F(\bar{x}) + \varepsilon \mathbb{B}$$

for every $x \in \mathbb{R}^p$ such that $\|x - \bar{x}\| < \delta$.

Proposition 2 *Let $g : \mathbb{R}^p \to \mathbb{R}^q$ be a locally Lipschitz function at $\bar{x}$. Then the generalized Jacobian $\partial_c g(\bar{x})$ of g at $\bar{x}$ is a first-order approximation of g at $\bar{x}$.*

Proof Since the set-valued mapping $\partial_c g(\cdot)$ is upper semicontinuous, for all $\varepsilon > 0$, there exists $r_0 > 0$ such that

$$\partial_c g(x) \subset \partial_c g(\bar{x}) + \varepsilon \mathbb{B}_{\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)} \quad \text{for all } x \in \bar{x} + r_0 \mathbb{B}_{\mathbb{R}^p}.$$

We may assume that g is Lipschitzian in $\bar{x} + r_0 \mathbb{B}_{\mathbb{R}^p}$. Let $x \in \bar{x} + r_0 \mathbb{B}_{\mathbb{R}^p}$. We apply [17], Prop. 2.6.5, to derive that there exits $c \in \,]x, \bar{x}[$ such that

$$g(x) - g(\bar{x}) \in \partial_c g(c)(x - \bar{x}) \subset \partial_c g(\bar{x})(x - \bar{x}) + \varepsilon \mathbb{B}_{\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)}(x - \bar{x}).$$

Since

$$\mathbb{B}_{\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)}(x - \bar{x}) \subset \|x - \bar{x}\| \mathbb{B}_{\mathbb{R}^q},$$

we have

$$g(x) - g(\bar{x}) \in \partial_c g(\bar{x})(x - \bar{x}) + \varepsilon \|x - \bar{x}\| \mathbb{B}_{\mathbb{R}^q},$$

which means that $\partial_c g(\bar{x})$ is a first-order approximation of g at $\bar{x}$. □

Recall that a mapping $f : X \to Y$ is said to be $C^{1,1}$ at $\bar{x}$ if it is Fréchet differentiable in neighborhood of $\bar{x}$ and if its Fréchet derivative $\nabla f(\cdot)$ is Lipschitz at $\bar{x}$.

Let $\bar{x} \in \mathbb{R}^p$, and let $f : \mathbb{R}^p \to \mathbb{R}$ be a $C^{1,1}$ function at $\bar{x}$. The generalized Hessian matrix of f at $\bar{x}$ was introduced and studied by Hiriart-Urruty et al. [19] is the compact nonempty convex set

$$\partial_H^2 f(\bar{x}) := \mathrm{co}\left\{\lim \nabla^2 f(x_n) : (x_n) \in \mathrm{dom}\, \nabla^2 f \textit{ and } x_n \to \bar{x}\right\}, \tag{5}$$

where $\mathrm{dom}\, \nabla^2 f$ is the effective domain of $\nabla^2 f(\cdot)$.

Corollary 1 *Let $\bar{x} \in \mathbb{R}^p$, and $f : \mathbb{R}^p \to \mathbb{R}$ be a $C^{1,1}$ function at $\bar{x}$. Then, ∇f admits $\partial_H^2 f(\bar{x})$ as a first-order approximation at $\bar{x}$.*

Definition 4 ([1]) We say that $f : X \to Y$ admits a second-order approximation at $\bar{x}$ if there exit two sets $\mathcal{A}_f(\bar{x}) \subset \mathcal{L}(X, Y)$ and $\mathcal{B}_f(\bar{x}) \subset \mathcal{B}(X \times X, Y)$ such that

(i) $\mathcal{A}_f(\bar{x})$ is a first-order approximation of f at $\bar{x}$;

(ii) For all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$f(x) - f(\bar{x}) \in \mathcal{A}_f(\bar{x})(x - \bar{x}) + \mathcal{B}_f(\bar{x})(x - \bar{x})(x - \bar{x}) + \varepsilon \|x - \bar{x}\|^2 \mathbb{B}_Y$$

for all $x \in \bar{x} + \delta \mathbb{B}_X$.

In this case the pair $(\mathcal{A}_f(\bar{x}), \mathcal{B}_f(\bar{x}))$ is called a second-order approximation of f at $\bar{x}$. It is called a compact second-order approximation if $\mathcal{A}_f(\bar{x})$ and $\mathcal{B}_f(\bar{x})$ are compacts.

Every C^2 mapping $f : X \to Y$ at $\bar{x}$ admits $(\nabla f(\bar{x}), \nabla^2 f(\bar{x}))$ as a second-order approximation, where $\nabla f(\bar{x})$ and $\nabla^2 f(\bar{x})$ are, respectively, the first- and second-order Fréchet derivatives of f at $\bar{x}$.

Proposition 3 ([1]) *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a $C^{1,1}$ function at $\bar{x}$. Then f admits $(\nabla f(\bar{x}), \frac{1}{2}\partial_H^2 f(\bar{x}))$ as a second-order approximation at $\bar{x}$.*

Proposition 4 *Let $f : X \to Y$ be a Fréchet-differentiable mapping. If $(\nabla f(\bar{x}), \mathcal{B}_f(\bar{x}))$ is a bounded second-order approximation of f at $\bar{x}$. Then $\nabla f(\cdot)$ is stable at $\bar{x}$, that is, there exist $c, r > 0$ such that*

$$\left\| \nabla f(x) - \nabla f(\bar{x}) \right\| \leq c\|x - \bar{x}\| \tag{6}$$

for all $x \in \bar{x} + r\mathbb{B}_X$.

To derive some results for γ-strong convex functions, the following notions are needed.

Definition 5 ([8]) Let $\gamma > 0$. We say that a map $f : X \to \mathbb{R} \cup \{+\infty\}$ is γ-strongly convex if there exist $c \geq 0$ and $g : [0,1] \to \mathbb{R}^+$ satisfying

$$g(0) = g(1) = 0 \quad \text{and} \quad \lim_{\theta \to 0} \frac{g(\theta)}{\theta} = 1 \tag{7}$$

and such that

$$f\big(\theta x + (1-\theta)y\big) \leq \theta f(x) + (1-\theta)f(y) - cg(\theta)\|x-y\|^{\gamma} \tag{8}$$

for all $\theta \in [0,1]$ and $x, y \in X$.

Of course, when $c = 0$, f is called a convex function. Otherwise, f is said γ-strongly convex. This class has been introduced by Polyak [11] when $\gamma = 2$ and $g(\theta) = \theta(1-\theta)$ and studied by many authors. Recently, a characterization of γ-strongly convex functions has been shown in [8]. For example, if f is C^1 and $\gamma \geq 1$, then (8) is equivalent to

$$\big\langle \nabla f(x), y-x \big\rangle \leq f(y) - f(x) - \frac{c}{\gamma}\|y-x\|^{\gamma}, \quad \forall x, y \in X. \tag{9}$$

Let $f : X \to \mathbb{R} \cup \{+\infty\}$ and $\bar{x} \in \operatorname{dom} f := \{x \in X, f(x) < +\infty\}$ (the effective domain of f). The Fenchel-subdifferential of f at $\bar{x}$ is the set

$$\partial_{\mathrm{Fen}} f(\bar{x}) = \big\{x^* \in X^* : \big\langle x^*, y-\bar{x}\big\rangle \leq f(y) - f(\bar{x}), \forall y \in X\big\}. \tag{10}$$

Let $\gamma > 0$ and $c > 0$. The (γ, c)-subdifferential of f at $\bar{x}$ is the set

$$\partial_{(\gamma,c)} f(\bar{x}) = \big\{x^* \in X^* : \big\langle x^*, y-\bar{x}\big\rangle \leq f(y) - f(\bar{x}) - c\|\bar{x}-y\|^{\gamma}, \forall y \in X\big\}. \tag{11}$$

For more details on (γ, c)-subdifferential, see [8]. Note that if $x \notin \operatorname{dom} f$, then $\partial_{(\gamma,c)} f(\bar{x}) = \partial_{\mathrm{Fen}} f(\bar{x}) = \emptyset$. Clearly, we have $\partial_{(\gamma,c)} f(\bar{x}) \subset \partial_{\mathrm{Fen}} f(\bar{x})$. Note that the Fenchel-subdifferential defined by (10) coincides with the Clarke subdifferential of f at $\bar{x}$ if the function f is convex. We also need to recall the following definitions.

Definition 6 ([20]) We say that a map $f : X \to \mathbb{R} \cup \{+\infty\}$ is 2-paraconvex if there exists $c > 0$ such that

$$f\big(\theta x + (1-\theta)y\big) \leq \theta f(x) + (1-\theta)f(y) + c\min(\theta, 1-\theta)\|x-y\|^{2} \tag{12}$$

for all $\theta \in [0,1]$ and $x, y \in X$.

It has been proved in [20] that if f is a C^1 mapping, then (12) is equivalent to

$$\big\langle \nabla f(x), y-x \big\rangle \leq f(y) - f(x) + c\|y-x\|^{2}, \quad \forall x, y \in X. \tag{13}$$

3 Main results

In this section, we obtain the main results of the paper related to strongly convex functions of order γ defined by (7)-(8). We begin by showing some interesting facts of functions that admit a first-order approximation.

For any subset A of X^*, we define the support function of A as

$$s(A, x) = \sup\big\{\big\langle x^*, x\big\rangle, x^* \in A\big\}. \tag{14}$$

It is well known that, for any convex function $f : X \to \mathbb{R} \cup \{+\infty\}$, the 'right-hand' directional derivative at x in dom$f$ (the domain of f) exists and, for each $h \in X$, is

$$d^+ f(x)(h) = \lim_{t\to 0^+} \frac{f(x+th) - f(x)}{t}.$$

Theorem 1 *Let $\bar{x} \in X$. If $f : X \to \mathbb{R} \cup \{+\infty\}$ is convex and continuous at $\bar{x}$ and if $\mathcal{A}_f(\bar{x}) \subset X^*$ is a convex $w(X^*, X)$-closed approximation of f at $\bar{x}$, then*

$$\partial_{(\gamma,c)} f(\bar{x}) \subset \mathcal{A}_f(\bar{x}).$$

Proof By the definition of $\mathcal{A}_f(\bar{x})$, there exist $\delta > 0$ and $r : X \to \mathbb{R}$ with $\lim_{x\to\bar{x}} r(x) = 0$ such that, for all $x \in \bar{x} + \delta\mathbb{B}_X$, $t \in]0, \delta[$, and $h \in X$, there exist $A \in \mathcal{A}_f(\bar{x})$ and $b \in [-1, 1]$ satisfying

$$\frac{f(\bar{x}+th) - f(\bar{x})}{t} - \|h\| r(\bar{x}+th) b = \langle A, h\rangle \le s\big(\mathcal{A}_f(\bar{x}); h\big).$$

By letting $t \to 0^+$ the directional derivative of f at $\bar{x}$ satisfies

$$d^+ f(\bar{x})(h) \le s\big(\mathcal{A}_f(\bar{x}); h\big), \quad \forall h \in X. \tag{15}$$

Using [21], Prop. 2.24, we get

$$s\big(\partial_{\mathrm{Fen}} f(\bar{x}); h\big) \le s\big(\mathcal{A}_f(\bar{x}); h\big).$$

Since $\partial_{(\gamma,c)} f(\bar{x}) \subset \partial_{\mathrm{Fen}} f(\bar{x})$, we deduce that

$$s\big(\partial_{(\gamma,c)} f(\bar{x}); h\big) \le s\big(\mathcal{A}_f(\bar{x}); h\big).$$

Hence we conclude that $\partial_{(\gamma,c)} f(\bar{x}) \subset \mathcal{A}_f(\bar{x})$. □

Proposition 5 *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a γ-strongly convex function. Assume that $\mathcal{A}_f(\bar{x})$ is a compact approximation at $\bar{x}$. Then $\mathcal{A}_f(\bar{x}) \cap \partial_{(\gamma,c)} f(\bar{x}) \neq \emptyset$.*

Proof Let $d \in X$ be fixed and define $x_n := \bar{x} + \frac{1}{n} d$. Using Definition 1, we get, for n large enough, $A_n \in \mathcal{A}_f(\bar{x})$ and $b_n \in [-1, 1]$ such that

$$\frac{1}{n}\langle A_n, d\rangle = f\left(\bar{x} + \frac{1}{n} d\right) - f(\bar{x}) - \frac{1}{n}\|d\| r(x_n) b_n.$$

By γ-strong convexity we obtain

$$\frac{1}{n}\langle A_n, d\rangle \le \frac{1}{n}\big(f(\bar{x}+d) - f(\bar{x})\big) - cg\left(\frac{1}{n}\right)\|d\|^{\gamma} - \frac{1}{n}\|d\| r(x_n) b_n.$$

By the compactness of $\mathcal{A}_f(\bar{x})$, extracting a subsequence if necessary, we may assume that there exists $A \in \mathcal{A}_f(\bar{x})$ such that $\langle A_n, d\rangle \to \langle A, d\rangle$; and hence we obtain

$$\langle A, d\rangle \le f(\bar{x}+d) - f(\bar{x}) - c\|d\|^{\gamma}. \tag{16}$$

Assume that $A \in \mathcal{A}_f(\bar{x}) \cap \partial_{(\gamma,c)} f(\bar{x})$. By the separation theorem there exists $h \in X$ with $\|h\| = 1$ such that

$$\min_{A \in \mathcal{A}_f(\bar{x})} \langle A, h \rangle > \sup_{x^* \in \partial_{(\gamma,c)} f(\bar{x})} \langle x^*, h \rangle.$$

Let $t > 0$ sufficiently small, so that

$$\min_{A \in \mathcal{A}_f(\bar{x})} \langle A, h \rangle > \frac{f(\bar{x} + th) - f(\bar{x})}{t},$$

in contradiction with relation (16) by taking $d = th$. □

Following a result by Rademacher, which states that a locally Lipschitzian function between finite-dimensional spaces is differentiable (Lebesgue) almost everywhere, we can prove the following result.

Proposition 6 *Let $\gamma \geq 1$, $\bar{x} \in \mathbb{R}^p$, and let $f : \mathbb{R}^p \to \mathbb{R}$ be continuous at $\bar{x}$. Assume that f is a γ-strongly convex function. Then $\partial_c f(\bar{x}) = \partial_{(\gamma,c)} f(\bar{x})$.*

Proof Obviously, we have $\partial_{(\gamma,c)} f(\bar{x}) \subset \partial_c f(\bar{x})$. Now let $A \in \partial_c f(\bar{x})$. For all n, there exists $x_n \in \operatorname{dom} \nabla f$ such that $x_n \to \bar{x}$ and $\nabla f(x_n) \to A$. Since f is γ-strongly convex and Fréchet differentiable at x_n for all $n \in \mathbb{N}$, it follows by (9) that

$$\langle \nabla f(x_n), y - x_n \rangle \leq f(y) - f(x_n) - c\|y - x_n\|^\gamma, \quad \forall y \in \mathbb{R}^p, \forall n \in \mathbb{N}.$$

Letting $n \to +\infty$, we get

$$\langle A, y - \bar{x} \rangle \leq f(y) - f(\bar{x}) - c\|y - \bar{x}\|^\gamma, \quad \forall y \in \mathbb{R}^p,$$

which means that $\partial_c f(\bar{x}) \subset \partial_{(\gamma,c)} f(\bar{x})$. □

Corollary 2 *Let $\gamma \geq 1$, $\bar{x} \in \mathbb{R}^p$, and let $f : \mathbb{R}^p \to \mathbb{R}$ be continuous at $\bar{x}$. Assume that f is a γ-strongly convex function. Then, for all $\varepsilon > 0$, there exists $r > 0$ such that*

$$f(x) - f(\bar{x}) \in \partial_{(\gamma,c)} f(\bar{x})(x - \bar{x}) + \varepsilon \|x - \bar{x}\| \mathbb{B}_{\mathbb{R}} \tag{17}$$

for all $x \in \bar{x} + r\mathbb{B}_{\mathbb{R}^p}$, which means that $\partial_{(\gamma,c)} f(\bar{x})$ is a first-order approximation of f at $\bar{x}$.

Proof It is clear that $\partial_c f(\bar{x})$ is a first-order approximation of at $\bar{x}$. We end the proof by Propositions 1 and 6. □

The converse of Proposition 5 holds if (16) is valid for any $A \in \mathcal{A}_f(x)$ and $x \in X$.

Proposition 7 *Let $\gamma \geq 1$ and $f : X \to \mathbb{R} \cup \{+\infty\}$. Assume that, for each $x \in X$, f admits a first-order approximation $\mathcal{A}_f(x)$ such that $\mathcal{A}_f(x) \subset \partial_{(\gamma,c)} f(x)$. Then f is γ-strongly convex.*

Proof Define $x_\theta := \theta u + (1-\theta)v$ for $\theta \in [0,1]$ and $u, v \in X$. Let us take $A \in \mathcal{A}_f(x_\theta)$. Then

$$\langle A, u - x_\theta \rangle \leq f(u) - f(x_\theta) - c\|u - x_\theta\|^\gamma.$$

Multiplying this inequality by θ, we obtain

$$(\mathrm{a}') \quad \theta(1-\theta)\langle A, u - v\rangle \leq \theta f(u) - \theta f(x_\theta) - c(1-\theta)^\gamma \theta \|u - v\|^\gamma.$$

In a similar way, since

$$\langle A, v - x_\theta \rangle \leq f(v) - f(x_\theta) - c\|v - x_\theta\|^\gamma,$$

we get

$$(\mathrm{a}'') \quad -\theta(1-\theta)\langle A, u - v\rangle \leq (1-\theta)f(v) - (1-\theta)f(x_\theta) - c(1-\theta)\theta^\gamma \|u - v\|^\gamma.$$

We deduce by addition of (a′) and (a″) that

$$f(x_\theta) \leq \theta f(u) + (1-\theta)f(v) - cg(\theta)\|u - v\|^\gamma \quad \text{for all } u, v \in X,$$

where $g(\theta) = (1-\theta)\theta^\gamma + (1-\theta)^\gamma\theta$, so that f is γ-strongly convex. □

The next results are devoted to presenting some useful properties of the generalized Hessian matrix for a $C^{1,1}$ function in the finite-dimensional setting and a characterization of γ-strongly convex functions with the help of a second-order approximation.

Proposition 8 *Let $\bar{x} \in X$, and let $f : X \to \mathbb{R} \cup \{+\infty\}$ be convex and Fréchet differentiable at $\bar{x}$. Suppose that f admits $(\nabla f(\bar{x}), \mathcal{B}_f(\bar{x}))$ as a second-order approximation at $\bar{x}$ and that $\mathcal{B}_f(\bar{x})$ is compact. Then there exists $B \in \mathcal{B}_f(\bar{x})$ such that*

$$\sup_{B \in \mathcal{B}_f(\bar{x})} \langle Bd, d\rangle \geq 0, \quad \forall d \in X. \tag{18}$$

If f is 2-strongly convex, then we obtain

$$\sup_{B \in \mathcal{B}_f(\bar{x})} \langle Bd, d\rangle \geq c\|d\|^2, \quad \forall d \in X, \tag{19}$$

for some $c > 0$.

Proof We prove only the case where f is convex. In a similar way, we can prove the other case. Let $d \in X$ and $\varepsilon > 0$ be fixed. We get for n large enough $B_n \in \mathcal{B}_f(\bar{x})$ and $b_n \in [-1,1]$ such that

$$f\left(\bar{x} + \frac{1}{n}d\right) - f(\bar{x}) = \frac{1}{n}\langle \nabla f(\bar{x}), d\rangle + \frac{1}{n^2}\langle B_n d, d\rangle + \varepsilon \frac{1}{n^2}\|d\|^2 b_n.$$

Since f is convex, we obtain

$$\langle B_n d, d\rangle + \varepsilon\|d\|^2 b_n \geq 0.$$

By the compactness of $\mathcal{B}_f(\bar{x})$, extracting a subsequence if necessary, we may assume that there exits $B \in \mathcal{B}_f(\bar{x})$ such that B_n converges to B; therefore

$$\langle Bd, d\rangle \geq 0,$$

and hence

$$\sup_{B\in\mathcal{B}_f(\bar{x})} \langle Bd, d\rangle \geq 0, \quad \forall d \in X.$$

□

When X is a finite-dimensional space, we get the following essential result.

Proposition 9 *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a $C^{1,1}$ function at $\bar{x}$. Assume that f is γ-strongly convex. Then, for any $B \in \partial^2_H f(\bar{x})$, we have the following inequality*:

$$\langle Bd, d\rangle \geq c\|d\|^\gamma, \quad \forall d \in \mathbb{R}^p, \tag{20}$$

for some $c > 0$.

Proof It is clear that $(\nabla f(\bar{x}), \frac{1}{2}\partial^2_H f(\bar{x}))$ is a second-order approximation of f at $\bar{x}$. Now let $B \in \partial^2_H f(\bar{x})$, so that there exists a sequence $(x_n) \in \operatorname{dom} \nabla^2 f$ such that $x_n \to \bar{x}$ and $\nabla^2 f(x_n) \to B$. Since f is γ-strongly convex, there exists $c > 0$ such that

$$\left\langle \nabla^2 f(x_n)d, d\right\rangle \geq c\|d\|^\gamma, \quad \forall d \in \mathbb{R}^p, \forall n \in \mathbb{N}.$$

Letting $n \to +\infty$, we have

$$\langle Bd, d\rangle \geq c\|d\|^\gamma, \quad \forall d \in \mathbb{R}^p.$$

□

The preceding result shows that γ-strongly convex functions enjoy a very desirable property for generalized Hessian matrices. In fact, in this case, any matrix $B \in \partial^2_H f(\bar{x})$ is invertible. The next result proves the converse of Proposition 9. Let us first recall the following characterization of l.s.c. γ-strongly convex functions.

Theorem 2 (Amahroq et al. [8]) *Let $f: X \to \mathbb{R} \cup \{+\infty\}$ be a proper and l.s.c. function. Then f is γ-strongly convex iff $\partial_c f$ is γ-strongly monotone, that is, there exists a positive real number c such that, for all $x, y \in X$, $x^* \in \partial_c f(x)$, and $y^* \in \partial_c f(y)$, we have*

$$\left\langle x^* - y^*, x - y\right\rangle \geq c\|x - y\|^\gamma.$$

We are now in position to state our main second result.

Theorem 3 *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a $C^{1,1}$ function. Assume that $\partial^2_H f(\cdot)$ satisfies relation* (20) *at any $x \in \mathbb{R}^p$. Then f is γ-strongly convex.*

Proof Let $t \in [0,1]$ and $u, v \in \mathbb{R}^p$. Define $\varphi : \mathbb{R} \to \mathbb{R}$ as

$$\varphi(t) := f\big(u + t(v - u)\big),$$

so that $\varphi'(t) := \langle \nabla f(u + t(v - u)), v - u\rangle$. By the Lebourg mean value theorem [22] there exists $t_0 \in]0,1[$ such that

$$\varphi'(1) - \varphi'(0) \in \partial_c \varphi'(t_0).$$

By using calculus rules it follows that

$$\varphi'(1) - \varphi'(0) \in \partial_c \varphi'(t_0) \subset \partial_H^2 f\big(u + t_0(v - u)\big)(v - u)(v - u).$$

Hence, there exists $B_{t_0} \in \partial_H^2 f(u + t_0(v - u))$ such that $\langle \nabla f(v) - \nabla f(u), v - u\rangle = \langle B_{t_0}(v - u), v - u\rangle$. The result follows from Theorem 2. □

Hiriart-Urruty et al. [19] have presented many examples of $C^{1,1}$ functions. The next proposition shows another example of a $C^{1,1}$ function.

Theorem 4 *Let $f : H \to \mathbb{R}$ be continuous on a Hilbert space H. Suppose that f is convex (or 2-strongly convex) and that $-f$ is 2-paraconvex. Then f is Fréchet differentiable on H, and for some $c > 0$, we have that*

$$\big\| \nabla f(x) - \nabla f(y)\big\| \le c\|x - y\| \quad \text{for all } x, y \in H. \tag{21}$$

Proof Let $x_0 \in X$. Clearly, f is locally Lipschitzian at x_0. Now let x_1^* and x_2^* be arbitrary elements of $\partial_c f(x_0)$ and $\partial_c(-f)(x_0)$, respectively. By [20], Thm. 3.4, there exists $c > 0$ such that $\partial_c(-f)(x_0) = \partial^{(2,c)}(-f)(x_0)$, and for any $y \in H$ and positive real θ, we have

$$\text{(a)} \quad \theta\big\langle x_2^*, y\big\rangle \le -f(x_0 + \theta y) + f(x_0) + c\theta^2\|y\|^2$$

and

$$\big(\text{a}'\big) \quad \theta\big\langle x_1^*, y\big\rangle \le f(x_0 + \theta y) - f(x_0).$$

Adding (a) and (a′), we get

$$\theta\big\langle x_1^* + x_2^*, y\big\rangle \le c\theta^2\|y\|^2,$$

and hence

$$\big\langle x_1^* + x_2^*, y\big\rangle \le c\theta\|y\|^2.$$

Letting $\theta \to 0$, we have $\langle x_1^* + x_2^*, y\rangle \le 0$, so that $x_1^* = -x_2^*$. Since x_1^* and x_2^* are arbitrary in $\partial_c f(x_0)$ and $\partial_c(-f)(x_0)$, it follows that $\partial_c f(x_0)$ is single-valued. Put $\partial_c f(x_0) = \{p(x_0)\}$. Since (a) and (a′) hold for any $\theta > 0$ and $y \in H$, we deduce that, for $\theta = 1$,

$$\big\langle p(x_0), y\big\rangle \le f(x_0 + y) - f(x_0)$$

and

$$f(x_0 + y) - f(x_0) - \big\langle p(x_0), y\big\rangle \le c\|y\|^2.$$

Hence, for all $y \neq 0$, we obtain

$$\frac{|f(x_0 + y) - f(x_0) - \langle p(x_0), y\rangle|}{\|y\|} \leq c\|y\|. \tag{22}$$

Letting $\|y\| \to 0$ in (22), we conclude that f is Fréchet differentiable at x_0. Now since $-f$ is 2-paraconvex and f is Fréchet differentiable, we may prove that there exists $c > 0$ such that

$$-\langle \nabla f(x), y - x\rangle \leq -f(y) + f(x) + c\|x - y\|^2 \quad \text{for all } x, y \in H. \tag{23}$$

For every $z \in H$, we have that

$$-f(z) \geq -f(x) + \langle \nabla f(x), x\rangle - \langle \nabla f(x), z\rangle - c\|x - z\|^2.$$

Thus

$$-f(z) \geq f^*(\nabla f(x)) - \langle \nabla f(x), z\rangle - c\|x - z\|^2,$$

so that

$$\begin{aligned}
&f^*(\nabla f(y)) \geq \langle \nabla f(y), z\rangle - f(z),\\
&f^*(\nabla f(y)) \geq \langle \nabla f(y), z\rangle + f^*(\nabla f(x)) - \langle \nabla f(x), z\rangle - c\|x - z\|^2,
\end{aligned}$$

and hence

$$\begin{aligned}
&f^*(\nabla f(y)) - f^*(\nabla f(x)) - \langle \nabla f(y) - \nabla f(x), x\rangle\\
&\quad \geq \langle \nabla f(y) - \nabla f(x), z - x\rangle - c\|x - z\|^2\\
&\quad \geq \sup_{z \in H}\{\langle \nabla f(y) - \nabla f(x), z - x\rangle - c\|x - z\|^2\}.
\end{aligned}$$

This means that, for all $x, y \in H$,

$$f^*(\nabla f(y)) - f^*(\nabla f(x)) - \langle \nabla f(y) - \nabla f(x), x\rangle \geq \frac{1}{2c}\|\nabla f(y) - \nabla f(x)\|^2.$$

Changing the roles of x and y, we obtain

$$f^*(\nabla f(x)) - f^*(\nabla f(y)) - \langle \nabla f(x) - \nabla f(y), y\rangle \geq \frac{1}{2c}\|\nabla f(x) - \nabla f(y)\|^2.$$

So by addition we get

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{c}\|\nabla f(x) - \nabla f(y)\|^2. \tag{24}$$

Consequently, by the Cauchy-Schwarz inequality we obtain

$$\|\nabla f(x) - \nabla f(y)\| \leq c\|x - y\| \quad \text{for all } x, y \in H.$$

□

4 Newton's method

The aim of this section is to solve the Euler equation

$$\nabla f(x) = 0 \tag{25}$$

by Newton's method. The classic assumption is that $f : \mathbb{R}^p \to \mathbb{R}$ a C^2 mapping and the Hessian matrix $\nabla^2 f(x)$ of f at x is nonsingular. Here we prove the convergence of a natural extension of Newton's method to solve (25) assuming that $\nabla f(\cdot)$ admits $\beta_f(\cdot)$ as a first-order approximation. Clearly, if $f : \mathbb{R}^p \to \mathbb{R}$ is a $C^{1,1}$ mapping, then using Corollary 1, we obtain that $\nabla f(\cdot)$ admits $\partial^2_H f(\cdot)$ as a first-order approximation.

This algorithm has been proposed by Cominetti et al. [23] with $C^{1,1}$ data. Only some ideas were given, but it remains as an open question to state results on rate of convergence and local convergence of that algorithm. In the sequel, $f : \mathbb{R}^p \to \mathbb{R}$ is a Fréchet-differentiable mapping such that its Fréchet derivative admits a first-order approximation, and $\bar{x}$ is a solution of (25).

Algorithm $(\mathcal{M})$ Starting from an arbitrary point $x_0 \in \mathbb{B}(\bar{x}, r)$, generate the sequences (x_k) and (h_k) as follows:

(i) $h_k \in \mathbb{R}^p$ is a solution of $0 \in \nabla f(x_k) + \beta_f(\bar{x})(h_k)$, and
(ii) $x_{k+1} = x_k + h_k$.

Theorem 5 *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a Fréchet-differentiable function, and $\bar{x}$ be a solution of* (25). *Let $\varepsilon, r, K > 0$ be such that $\nabla f(\cdot)$ admits $\beta_f(\bar{x})$ as a first-order approximation at $\bar{x}$ such that, for each $x \in \mathbb{B}_{\mathbb{R}^p}(\bar{x}, r)$, there exists an invertible element $B(x) \in \mathcal{B}_f(x)$ satisfying $\|B(x)^{-1}\| \leq K$ and $\xi := \varepsilon K < 1$. Then the sequence (x_k) generated by Algorithm $(\mathcal{M})$ is well defined for every $x_0 \in \mathbb{B}_{\mathbb{R}^p}(\bar{x}, r)$ and converges linearly to $\bar{x}$ with rate ξ.*

Proof Since $\nabla f(\bar{x}) = 0$, we have

$$x_{k+1} - \bar{x} = B(x_k)^{-1}\bigl(\nabla f(\bar{x}) - \nabla f(x_k) + B(x_k)(x_k - \bar{x})\bigr).$$

We inductively obtain that

$$\|x_{k+1} - \bar{x}\| \leq K\bigl\|\nabla f(\bar{x}) - \nabla f(x_k) + B(x_k)(x_k - \bar{x})\bigr\|.$$

Thus

$$\|x_{k+1} - \bar{x}\| \leq \xi \|x_k - \bar{x}\|,$$

which means that $x_{k+1} \in \mathbb{B}_{\mathbb{R}^p}(\bar{x}, r)$, and we have $\|x_{k+1} - \bar{x}\| \leq \xi^k \|x_0 - \bar{x}\|$. Therefore the whole sequence (x_k) is well defined and converges to $\bar{x}$. □

Now let us consider the following algorithm under less assumptions.

Algorithm $(\mathcal{M}')$ Starting from an arbitrary point $x_0 \in \mathbb{R}^p$, generate the sequences (x_k) and (h_k) as follows:

(i) $h_k \in \mathbb{R}^p$ is a solution of $0 \in \nabla f(x_k) + \beta_f(x_0)(h_k)$, and
(ii) $x_{k+1} = x_k + h_k$.

Theorem 6 *Let U be an open set of $\mathbb{R}^p$, $x_0 \in U$, and $f : \mathbb{R}^p \to \mathbb{R}$ be a Fréchet-differentiable function on U. Let $\varepsilon, r, K > 0$ be such that $\nabla f(\cdot)$ admits $\beta_f(x_0)$ as a strict first-order approximation at x_0 such that, for each $x \in \mathbb{B}_{\mathbb{R}^p}(x_0, r)$, there exists a right inverse of $B(x) \in \beta_f(x_0)$, denoted by $\tilde{B}(x)$, satisfying $\|\tilde{B}(x)(\cdot)\| \le K\|\cdot\|$ and $\xi := \varepsilon K < 1$.*

If $\|\nabla f(x_0)\| \le K^{-1}(1-\xi)r$ and ∇f is continuous, then the sequence (x_k) generated by Algorithm $(\mathcal{M}')$ is well defined and converges to a solution $\bar{x}$ of (25). *Moreover, we have $\|x_k - \bar{x}\| \le r\xi^k$ for all $k \in \mathbb{N}$ and $\|\bar{x} - x_0\| \le \|\nabla f(x_0)\| K(1-\xi)^{-1} < r$.*

Proof We prove by induction that $x_k \in x_0 + r\mathbb{B}_{\mathbb{R}^p}$, $\|x_{k+1} - x_k\| \le K\xi^k \|\nabla f(x_0)\|$, and $\|\nabla f(x_k)\| \le \xi^k \|\nabla f(x_0)\|$ for all $k \in \mathbb{N}$. For $k = 0$, these relations are obvious. Assuming that they are valid for $k < n$, we get

$$\begin{aligned} \|x_n - x_0\| &\le \sum_{k=0}^{n-1} \|x_{k+1} - x_k\| \le K \big\|\nabla f(x_0)\big\| \sum_{k=0}^{\infty} \xi^k \\ &\le K \big\|\nabla f(x_0)\big\| (1-\xi)^{-1} < r. \end{aligned}$$

Thus $x_n \in x_0 + r\mathbb{B}_{\mathbb{R}^p}$ and since $\nabla f(x_{n-1}) + B(x_{n-1})(x_n - x_{n-1}) = 0$, from Algorithm $(\mathcal{M}')$ we have

$$\begin{aligned} \big\|\nabla f(x_n)\big\| &\le \big\|\nabla f(x_n) - \nabla f(x_{n-1}) - B(x_{n-1})(x_n - x_{n-1})\big\| \le \varepsilon \|x_n - x_{n-1}\| \\ &\le \xi^n \big\|\nabla f(x_0)\big\| \end{aligned}$$

and

$$\|x_{n+1} - x_n\| \le K\xi^n \big\|\nabla f(x_0)\big\|.$$

Since $\xi < 1$, the sequence (x_n) is a Cauchy sequence and hence converges to some $\bar{x} \in \mathbb{R}^p$ with $\|x_0 - \bar{x}\| < r$. Since ∇f is a continuous function, we get $\nabla f(\bar{x}) = 0$. □

5 Conclusions

In this paper, we investigate the concept of first- and second-order approximations to generalize some results such as optimality conditions for a subclass of convex functions called strongly convex functions of order γ. We also present an extension of Newton's method to solve the Euler equation under weak assumptions.

Acknowledgements
The author wishes to express his heartfelt thanks to the referees for their detailed and helpful suggestions for revising the manuscript.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
The author read and approved the final manuscript.

Received: 2 June 2017 Accepted: 25 August 2017 Published online: 08 September 2017

References

1. Allali, K, Amahroq, T: Second order approximations and primal and dual necessary optimality conditions. Optimization **40**(3), 229-246 (1997)
2. Thibault, L: Subdifferentials of compactly Lipschitzian vector-valued functions. Ann. Mat. Pura Appl. (4) **125**, 157-192 (1980)
3. Sweetser, TH: A minimal set-valued strong derivative for set-valued Lipschitz functions. J. Optim. Theory Appl. **23**, 539-562 (1977)
4. Ioffe, AD: Nonsmooth analysis: differential calculus of nondifferentiable mappings. Trans. Am. Math. Soc. **266**, 1-56 (1981)
5. Jourani, A, Thibault, L: Approximations and metric regularity in mathematical programming in Banach space. Math. Oper. Res. **18**(2), 390-401 (1993)
6. Amahroq, T, Gadhi, N: On the regularity condition for vector programming problems. J. Glob. Optim. **21**(4), 435-443 (2001)
7. Amahroq, T, Gadhi, N: Second order optimality conditions for the extremal problem under inclusion constraints. J. Math. Anal. Appl. **285**(1), 74-85 (2003)
8. Amahroq, T, Daidai, I, Syam, A: γ-strongly convex functions, γ-strong monotonicity of their presubdifferential and γ-subdifferentiability, application to nonlinear PDE. J. Nonlinear Convex Anal. (2017, to appear)
9. Crouzeix, JP, Ferland, JA, Zalinescu, C: α-convex sets and strong quasiconvexity. SIAM J. Control Optim. **22**, 998-1022 (1997)
10. Lin, GH, Fukushima, M: Some exact penalty results for nonlinear programs and mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **118**, 67-80 (2003)
11. Polyak, BT: Introduction to Optimization. Optimization Software, New York (1987). Translated from the Russian, with a foreword by Dimitri P. Bertsekas
12. Rockafellar, RT: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877-898 (1976)
13. Vial, J-P: Strong convexity of sets and functions. J. Math. Econ. **9**(1-2), 187-205 (1982). doi:10.1016/0304-4068(82)90026-X
14. Vial, J-P: Strong and weak convexity of sets and functions. Math. Oper. Res. **8**(2), 231-259 (1983)
15. Zălinescu, C: On uniformly convex functions. J. Math. Anal. Appl. **95**(2), 344-374 (1983)
16. Mordukhovich, BS, Shao, YH: On nonconvex subdifferential calculus in Banach spaces. J. Convex Anal. **2**(1-2), 211-227 (1995)
17. Clarke, FH: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
18. Clarke, FH: On the inverse function theorem. Pac. J. Math. **64**(1), 97-102 (1976)
19. Hiriart-Urruty, J-B, Strodiot, J-J, Nguyen, VH: Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. Appl. Math. Optim. **11**(1), 43-56 (1984)
20. Jourani, A: Subdifferentiability and subdifferential monotonicity of γ-paraconvex functions. Control Cybern. **25**(4), 721-737 (1996)
21. Phelps, RR: Convex Functions, Monotone Operators and Differentiability Lecture Notes in Mathematics, vol. 1364. Springer, Berlin (1989)
22. Lebourg, G: Valeur moyenne pour gradient généralisé. C. R. Acad. Sci. Paris **281**, 795-797 (1975)
23. Cominetti, R, Correa, R: A generalized second-order derivative in nonsmooth optimization. SIAM J. Control Optim. **28**(4), 789-809 (1990)