On subgradient projectors

Heinz H. Bauschke[*], Caifang Wang[†], Xianfu Wang[‡] and Jia Xu[§]

February 16, 2015

Abstract

The subgradient projector is of considerable importance in convex optimization because it plays the key role in Polyak's seminal work — and the many papers it spawned — on subgradient projection algorithms for solving convex feasibility problems.

In this paper, we offer a systematic study of the subgradient projector. Fundamental properties such as continuity, nonexpansiveness, and monotonicity are investigated. We also discuss the Yamagishi–Yamada operator. Numerous examples illustrate our results.

2010 Mathematics Subject Classification: Primary 90C25; Secondary 47H04, 47H05, 47H09, 47N10.

Keywords: Convex function, firmly nonexpansive mapping, Fréchet differentiability, Gâteaux differentiability, monotone operator, nonexpansive mapping, subgradient projector, Yamagishi–Yamada operator.

1 Introduction

Throughout this paper, we assume that

(1) $$X \text{ is a real Hilbert space}$$

with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$. We also assume that

(2) $$f\colon X\to\mathbb{R} \text{ is convex and continuous, and } C=\big\{x\in X \;\big|\; f(x)\leq 0\big\}\neq\varnothing.$$

(When X is finite-dimensional, we do not need to explicitly impose continuity on f.) Unless stated otherwise, we assume that $s\colon X\to X$ is a *selection* of ∂f, i.e.,

(3) $$(\forall x\in X)\quad s(x)\in\partial f(x)$$

[*]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.
[†]Department of Mathematics, Shanghai Maritime University, China. E-mail: cfwang@shmtu.edu.cn.
[‡]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: shawn.wang@ubc.ca.
[§]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: jia.xu@ubc.ca.

and that $G\colon X \to X$ is the *associated subgradient projector* defined by

$$(4) \qquad (\forall x \in X) \quad Gx = \begin{cases} x - \dfrac{f(x)}{\|s(x)\|^2}s(x), & \text{if } f(x) > 0; \\ x, & \text{otherwise.} \end{cases}$$

Observe this is well defined because $C \neq \varnothing$ and thus $0 \notin \partial f(X \smallsetminus C)$.

When we need to exhibit the underlying function f or subgradient selection s, we shall write s_f, C_f and $G_f = G_{f,s}$ instead of s, C and G, respectively.

The subgradient projector is the key ingredient in Polyak's seminal work [23] on subgradient projection algorithms [1], which have since found many applications; see, e.g., [1] (which deals with subgradient algorithms viewed as projection algorithms with variable supersets), [4] (which provides a framework for transforming weakly into strongly convergent cutter methods), [8] (which provides a very nice and recent monograph on cutters, subgradient projectors and related algorithms), [9] (which introduces the cyclic subgradient projections method for solving systems of convex inequalities), [10] (which deals with an almost cyclic sequential algorithm for solving a common fixed point problem of cutters), [11] (which is a book that not only features subgradient projection algorithms but it also won their authors the *INFORMS Computing Society prize*), [12] (which presents a general framework for algorithmically solving feasibility problems especially in signal processing), [14] (which is a broad survey on subgradient projection algorithms for solving image recovery problems), [15] (which develops the mathematical theory of the general and fast parallel projection method for image recovery based on subgradient projectors), [16] (which presents an adaptive level set method for constrained image recovery relying upon subgradient projectors), [20] (which introduces a cut that is tighter than the subgradient projector cut when applied to certain quadratic functions), [24] (which is a seminal book that includes subgradient methods for the constrained minimization of nonsmooth functions), [25] (which employs randomly chosen subgradient projectors to solve convex inequalities), [27] (which provides an adaptive projected subgradient method with applications to machine learning), [28] (which is not only a review on subgradient projection algorithms for adaptive learning but it also won the *IEEE Signal Processing Society Signal Processing Magazine Best Paper Award*), [29] (which presents subgradient-projector based method for solving robust adaptive signal processing problems), [30] (which features a very general hybrid steepest descent for solving variational inequalities with applications to convex optimization via subgradient projectors), [31] (which introduces an adaptive filtering algorithm based on parallel subgradient projection methods), and the references therein. This impressive body of (including even award-winning) research, which has also attracted a significant number of citations and which clearly brought out the importance of the subgradient projector as an algorithmic building block, warrants a mathematical study of the subgradient projector and its properties. After submission of this paper, P. Combettes made us aware of the work of B. Pauwels [21] which is complementary [2] to ours.

The aim of this paper is to provide a systematic study of the subgradient operator. We review known properties, present basic calculus rules, obtain characterization of strong-to-strong and strong-to-weak con-

[1]See also [17] for a historical account.

[2]While there is some overlap in Sections 2–5, there is not much in common with respect to Sections 6–11. We note that Pauwels disucsses epi-convergence of functions [21, Section 4] and he also allows throughout for functions that do not necessarily have full domain. However, we chose not to consider the case when f is not full domain because we are unaware of results on the *range* of G. This lack of results severely hinders the applicability of subgradient projectors as building blocks in algorithms which was a main motivation for this work.

tinuity, analyze nonexpansiveness, monotonicity, and the decreasing property, and discuss the relationship to the Yamagishi–Yamada operator. Numerous examples illustrate our results.

Let us briefly summarize the main results in this paper:

- Theorem 5.6 provides a strong-to-strong continuity of G in terms of Fréchet differentiability of f; surprisingly, the natural counterpart for strong-to-weak continuity of G in terms of Gâteaux differentiability fails (Example 6.2).
- In Section 7, we connect G to the accelerated mapping studied to speed up convergence of certain projection methods.
- In Proposition 8.6, we provide a sufficient condition for G corresponding to a certain Moreau envelope to be firmly nonexpansive.
- In Section 9 we discuss when a subgradient projector step decreases the value of the objective function.
- Section 10 contains a case study which illustrates that the behaviour of G is rather complicated.
- In the final Section 11 we exhibit the Yamagishi–Yamada operator as a special subgradient projector.

The paper is organized as follows. Basic properties are reviewed in Section 2, and basic calculus rules are derived in Section 3. Section 4 is a collection of examples. The relationship between strong-to-strong (resp. strong-to-weak) continuity of G and Fréchet (resp. Gâteaux) differentiability of f is clarified in Section 5 (resp. Section 6). The case when f arises from a quadratic form is investigated in Section 7. Nonexpansiveness and the decreasing property are studied in Section 8 and 9, respectively. These properties are illustrated in Section 10. In the final Section 11, we provide a sufficient condition for the Yamagishi–Yamada operator to be itself a subgradient projector.

Notation and terminology are standard and follow largely [3] to which we refer the reader if needed. We do write $\mathrm{P}_f = (\mathrm{Id} + \partial f)^{-1}$ for the proximity operator (proximal mapping) of f.

2 Preliminary results

Let us record some basic results on subgradient projectors, which are essentially contained already in [23] and the proofs of which we provide for completeness.

Fact 2.1 *Let $x \in X$, and set*

$$H = \big\{ y \in X \;\big|\; \langle s(x), y - x\rangle + f(x) \leq 0 \big\}. \tag{5}$$

Then the following hold:

(i) $f^+(x) + \langle s(x), Gx - x\rangle = 0$.

(ii) $\operatorname{Fix} G = C \subseteq H$.

(iii) $Gx = P_H x$.

(iv) **(G is a cutter[3])** $(\forall c \in C)\ \langle c - Gx, x - Gx \rangle \leq 0$.

(v) $(\forall c \in C)\ \|x - Gx\|^2 + \|Gx - c\|^2 \leq \|x - c\|^2$.

(vi) $f^+(x) = \|s(x)\|\|x - Gx\|$.

(vii) *If $x \notin C$, then* $(\forall c \in C)\ f^2(x)\|s(x)\|^{-2} + \|Gx - c\|^2 \leq \|x - c\|^2$.

(viii) $f^+(x)(x - Gx) = \|x - Gx\|^2 s(x)$.

(ix) *Suppose that f is Fréchet differentiable at $x \in X \smallsetminus C$. Then $g = \ln \circ f \colon X \smallsetminus C \to \mathbb{R}$ is Fréchet differentiable at x and $Gx = x - \nabla g(x)/\|\nabla g(x)\|^2$.*

(x) *Suppose that* $\min f(X) = 0$, *that f is Fréchet differentiable on X with ∇f being Lipschitz continuous with constant L, that $x \notin C$, and that there exists $\alpha > 0$ such that $f(x) \geq \alpha d_C^2(x)$. Then $d_C^2(Gx) \leq (1 - \alpha^2/L^2)d_C^2(x)$.*

(xi) *Suppose that* $\min f(X) = 0$, *that $x \notin C$, and that there exists $\alpha > 0$ such that $f(x) \geq \alpha d_C(x)$. Then $d_C^2(Gx) \leq (1 - \alpha^2/\|s(x)\|^2)d_C^2(x)$.*

Proof. (i): This follows directly from the definition of G.

(ii): The equality is clear from the definition of G. Let $z \in C$. Then $\langle s(x), z - x \rangle + f(x) \leq f(z) \leq 0$ and hence $z \in H$.

(iii): Assume first that $x \in C$. Then $x \in \operatorname{Fix} G \subseteq H$ by (ii) and hence $Gx = x = P_H x$. Now assume that $x \notin C$. Then $0 < f(x) = f^+(x)$ and $s(x) \neq 0$. Hence,

$$P_H x = x - \frac{\Big(\langle s(x), x \rangle - \big(\langle s(x), x \rangle - f(x) \big) \Big)^+}{\|s(x)\|^2} s(x) = x - \frac{f^+(x)}{\|s(x)\|^2} s(x) = Gx. \tag{6}$$

(iv): In view of (iii), we have $(\forall h \in H)\ \langle h - Gx, x - Gx \rangle \leq 0$. Now invoke (ii).

(v): This is equivalent to (iv).

(vi): Assume first that $x \in C$. Then $f(x) \leq 0$, i.e., $f^+(x) = 0$, and $x = Gx$ by (ii). Hence the identity is true. Now assume that $x \notin C$. Then $0 < f(x) = f^+(x)$ and $x - Gx = f(x)/\|s(x)\|^2 s(x)$. Taking the norm, we learn that $\|x - Gx\| = f(x)/\|s(x)\| = f^+(x)/\|s(x)\|$.

(vii): Combine (ii), (v), and (vi).

(viii): This follows from (vi) and the definition of G.

(ix): The chain rule implies that $\nabla g(x) = (1/f(x))\nabla f(x)$. Hence $\|\nabla g(x)\|^2 = \|\nabla f(x)\|^2/f^2(x)$ and thus $x - \nabla g(x)/\|\nabla g(x)\|^2 = x - f(x)/\|\nabla f(x)\|^2 \nabla f(x) = Gx$.

(x): Let $c \in C$. Then $\nabla f(c) = 0$ and hence $\|\nabla f(x)\| = \|\nabla f(x) - \nabla f(c)\| \leq L\|x - c\|$. Hence $\|\nabla f(x)\| \leq L d_C(x)$ and therefore, using (vii), we obtain

$$\|Gx - c\|^2 \leq \|x - c\|^2 - \frac{f^2(x)}{\|\nabla f(x)\|^2} \leq \|x - c\|^2 - \frac{\alpha^2 d_C^4(x)}{L^2 d_C^2(x)}. \tag{7}$$

Now take the minimum over $c \in C$.

[3]See [8] for a thorough treatment on cutters.

(xi): Using (vii), we have

$$d_C^2(Gx) \leq \|Gx - P_C x\|^2 \leq \|x - P_C x\|^2 - \frac{f^2(x)}{\|s(x)\|^2} \leq d_C^2(x) - \frac{\alpha^2 d_C^2(x)}{\|s(x)\|^2}. \tag{8}$$

The proof is complete. ■

3 Calculus

We now turn to basic calculus rules. When the proof is a straight-forward verification, we will omit it. It is convenient to introduce the operator $\mathcal{G}\colon X \rightrightarrows X$, defined by

$$(\forall x \in X) \quad \mathcal{G}x = \mathcal{G}_f x = \big\{ G_s(x) \mid s \text{ is a selection of } \partial f \big\}, \tag{9}$$

where G_s is the operator occurring in (4). When f is Gâteaux differentiable outside C, then we will identify $\mathcal{G}$ with G.

Proposition 3.1 (calculus) *Let $\alpha > 0$, let $A\colon X \to X$ be continuous and linear such that $A^*A = A^*A = \mathrm{Id}$, and let $z \in X$. Furthermore, let $(f_i)_{i\in I}$ be a finite family of convex continuous functions on X such that $\bigcap_{i\in I} C_{f_i} \neq \varnothing$. Then the following hold:*

(i) *Suppose that $g = \alpha f$. Then $C_g = C_f$ and $\mathcal{G}_g = \mathcal{G}_f$.*

(ii) *Suppose that $g = f \circ \alpha\,\mathrm{Id}$. Then $C_g = \alpha^{-1} C_f$ and $\mathcal{G}_g = \alpha^{-1}\mathcal{G}_f \circ \alpha\,\mathrm{Id}$.*

(iii) *Suppose that $f \geq 0$ and that $g = f^\alpha$ is convex. Then $C_g = C_f$ and $\mathcal{G}_g = (1-\alpha^{-1})\,\mathrm{Id} + \alpha^{-1}\mathcal{G}_f$.*

(iv) *Suppose that $g = f \circ A$. Then $C_g = A^* C_f$ and $\mathcal{G}_g = A^* \circ \mathcal{G}_f \circ A$.*

(v) *Suppose that $g\colon x \mapsto f(x-z)$. Then $C_g = z + C_f$ and $\mathcal{G}_g\colon x \mapsto z + \mathcal{G}_f(x-z)$.*

(vi) *Suppose that $g = \max_{i\in I} f_i$. Then $C_g = \bigcap_{i\in I} C_{f_i}$ and if $g(x) > 0$ and $I(x) = \big\{ i \in I \mid f_i(x) = g(x) \big\}$, then $\mathcal{G}_g(x) = \big\{ x - g(x)\|x^*\|^{-2}x^* \mid x^* \in \operatorname{conv}\bigcup_{i\in I(x)} \partial f_i(x) \big\}$.*

(vii) *Suppose that $g = f^+$. Then $\mathcal{G}_g = \mathcal{G}_f$.*

(viii) ***(Moreau envelope)** Suppose that $\min f(X) = 0$ and that $g = f \Box (1/2)\|\cdot\|^2$ is the Moreau envelope of f. Then $C_g = C_f$ and*

$$(\forall x \in X) \quad G_g(x) = \begin{cases} x - \dfrac{g(x)}{\|x - \mathrm{P}_f\, x\|^2}(x - \mathrm{P}_f\, x), & \text{if } f(x) > 0; \\ x, & \text{if } f(x) = 0. \end{cases} \tag{10}$$

Proof. Let $x \in X$. We shall only prove one inclusion for the subgradient projector as the remaining one is proved similarly.

(i): Since $g(x) \leq 0 \Leftrightarrow f(x) \leq 0$, it follows that $C_g = C_f$. Suppose that $f(x) > 0$. Since $\alpha s_f(x) \in \partial g(x)$, we obtain $G_f x = x - f(x)\|s_f(x)\|^{-2}s_f(x) = x - g(x)/\|\alpha s_f(x)\|^{-2}(\alpha s_f(x))$. This implies $\mathcal{G}_f(x) \subseteq \mathcal{G}_g(x)$.

(ii): Suppose that $g(x) > 0$, i.e., $f(\alpha x) > 0$. Then $\alpha^{-1}G_f(\alpha x) = \alpha^{-1}(\alpha x - f(\alpha x)\|s_f(\alpha x)\|^{-2}s_f(\alpha x)) = x - \alpha^{-1}f(\alpha x)\|s_f(\alpha x)\|^{-2}s_f(\alpha x) = x - f(\alpha x)\|\alpha s_f(\alpha x)\|^{-2}(\alpha s_f(\alpha x)) \in \mathcal{G}_g(x)$. Hence $\alpha^{-1}\mathcal{G}_f(\alpha x) \subseteq \mathcal{G}_g(x)$.

(iii): Suppose that $g(x) > 0$. Then $f(x) > 0$ and $\alpha^{-1}(x - G_f x) = \alpha^{-1}f(x)/\|s_f(x)\|^2 s_f(x) = f^\alpha(x)\|\alpha f^{\alpha-1}(x)s_f(x)\|^{-2}\alpha f^{\alpha-1}(x)s_f(x) \in x - \mathcal{G}_g(x)$.

(iv): We have $x \in C_g \Leftrightarrow f(Ax) \leq 0 \Leftrightarrow Ax \in C_f \Leftrightarrow x \in A^*C_f$. Suppose that $g(x) > 0$. Then $f(Ax) > 0$, $A^*s_f(Ax) \in \partial g(x)$ and $A^*G_f(Ax) = A^*(Ax - f(Ax)\|s_f(Ax)\|^{-2}s_f(Ax)) = x - f(Ax)\|A^*s_f(Ax)\|^{-2}A^*s_f(Ax) \in \mathcal{G}_g(x)$.

(v): Suppose that $0 < g(x) = f(x-z)$. Then $z + G_f(x-z) = z + x - z - f(x-z)\|s_f(x-z)\|^{-2}s_f(x-z) \in \mathcal{G}_g(x)$.

(vi): This follows from the well known formula for the subdifferential of a maximum; see, e.g., [22, Proposition 3.38].

(vii): This follows from (vi) since $f^+ = \max\{0, f\}$.

(viii): This is clear because $g \geq 0$, $\nabla g = \mathrm{Id} - \mathrm{P}_f$ (see, e.g., [3, Proposition 12.29]), and $\operatorname{argmin} g = \operatorname{argmin} f$ (see, e.g., [3, Corollary 17.5]). ■

4 Examples

In this section, we present several illustrative examples.

Example 4.1 Suppose that $f = \|\cdot\|^2$. Then $\nabla f = 2\,\mathrm{Id}$ and $G = \frac{1}{2}\,\mathrm{Id}$.

Example 4.2 Suppose that

$$(\forall x \in X) \quad f(x) = \begin{cases} \frac{1}{2}\|x\|^2, & \text{if } x \in \operatorname{ball}(0;1); \\ \|x\| - \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{11}$$

Then $G = \frac{1}{2}P_{\operatorname{ball}(0;1)}$ and G is firmly nonexpansive[4].

Proof. Let $x \in X$. Observe that $f = \|\cdot\|\Box(1/2)\|\cdot\|^2$ is the Moreau envelope of the norm. Hence it follows from Proposition 3.1(viii) that

$$Gx = x - \frac{f(x)}{\|x - \mathrm{P}_{\|\cdot\|}\,x\|^2}(x - \mathrm{P}_{\|\cdot\|}\,x) \tag{12}$$

provided that $x \neq 0$, and $Gx = 0 = \frac{1}{2}P_{\operatorname{ball}(0;1)}x$ if $x = 0$. Furthermore, $\mathrm{P}_{\|\cdot\|} = \mathrm{Id} - \mathrm{P}_{\|\cdot\|^*} = \mathrm{Id} - \mathrm{P}_{\iota_{\operatorname{ball}(0;1)}} = \mathrm{Id} - P_{\operatorname{ball}(0;1)}$. Thus, $\mathrm{Id} - \mathrm{P}_{\|\cdot\|} = P_{\operatorname{ball}(0;1)}$. Assume now $x \neq 0$. If $0 < \|x\| \leq 1$, then

$$Gx = x - \frac{\frac{1}{2}\|x\|^2}{\|P_{\operatorname{ball}(0;1)}x\|^2}P_{\operatorname{ball}(0;1)}(x) = x - \frac{\|x\|^2}{2\|x\|^2}x = \tfrac{1}{2}x = \tfrac{1}{2}P_{\operatorname{ball}(0;1)}x; \tag{13}$$

and if $1 < \|x\|$, then

$$Gx = x - \frac{\|x\| - \frac{1}{2}}{\|P_{\operatorname{ball}(0;1)}x\|^2}P_{\operatorname{ball}(0;1)}(x) = x - \frac{\|x\| - \frac{1}{2}}{\big\|x/\|x\|\big\|^2}\frac{x}{\|x\|} = \tfrac{1}{2}\frac{x}{\|x\|} = \tfrac{1}{2}P_{\operatorname{ball}(0;1)}x. \tag{14}$$

[4] Recall that $T\colon X \to X$ is *firmly nonexpansive* if $(\forall x \in X)(\forall y \in X)$ $\|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \leq \|x - y\|^2$.

Now $P_{\mathrm{ball}(0;1)}$ is firmly nonexpansive, and hence so is $\mathrm{Id} - P_{\mathrm{ball}(0;1)}$. It follows that $2G - \mathrm{Id} = -(\mathrm{Id} - P_{\mathrm{ball}(0;1)})$ is nonexpansive[5], and therefore that G is firmly nonexpansive. ■

Proposition 4.3 *Let $(C_i)_{i\in I}$ be a finite family of closed convex subsets of X such that $C = \bigcap_{i\in I} C_i \neq \varnothing$ and $f = \max_{i\in I} d_{C_i}$. Let $x \in X \smallsetminus C$, set $I(x) = \big\{i \in I \;\big|\; f(x) = d_{C_i}(x)\big\}$, and set $Q(x) = \mathrm{conv}\{P_{C_i}x\}_{i\in I(x)}$. Then*

$$\mathcal{G}(x) = \bigcup_{q(x)\in Q(x)} \left\{ x - \frac{f^2(x)}{\|x - q(x)\|^2}\big(x - q(x)\big)\right\} \quad \textit{and} \quad Q(x) \subseteq \mathrm{conv}\,\big(\{x\} \cup \mathcal{G}(x)\big). \tag{15}$$

If $I(x) = \{i\}$ is a singleton, then $\mathcal{G}(x) = \{P_{C_i}x\}$.

Proof. This follows from Proposition 3.1(vi) and the fact that $\nabla d_{C_i}(x) = (x - P_{C_i}x)/d_{C_i}(x)$ when $x \in X \smallsetminus C_i$. ■

Our next result features, when $p = 2$, an important subgradient projector used in an extrapolated parallel projection method (see [14, Section 5.4.1.2]).

Proposition 4.4 *Let $(C_i)_{i\in I}$ be a finite family of nonempty closed convex subsets of X such that $C = \bigcap_{i\in I} C_i \neq \varnothing$. Let $(\lambda_i)_{i\in I}$ be a family in $]0,1]$ such that $\sum_{i\in I}\lambda_i = 1$. Let $p \geq 1$ and suppose that $f = \sum_{i\in I}\lambda_i d^p_{C_i}$. Set $(\forall x \in X)\ I(x) = \big\{i \in I \;\big|\; x \notin C_i\big\}$. Then*

$$(\forall x \in X) \quad Gx = x - \frac{\sum_{i\in I(x)} \lambda_i d^p_{C_i}(x)}{p\big\|\sum_{i\in I(x)}\lambda_i d^{p-2}_{C_i}(x)(x - P_{C_i}x)\big\|^2} \sum_{i\in I(x)} \lambda_i d^{p-2}_{C_i}(x)(x - P_{C_i}x) \tag{16}$$

and if $p = 2$, we rewrite this as

$$(\forall x \in X) \quad Gx = \begin{cases} x - \dfrac{\sum_{i\in I}\lambda_i\|x - P_{C_i}x\|^2}{2\big\|\sum_{i\in I}\lambda_i(x - P_{C_i}x)\big\|^2}\Big(x - \sum_{i\in I}\lambda_i P_{C_i}x\Big), & \textit{if } x \notin C;\\ x, & \textit{otherwise.}\end{cases} \tag{17}$$

Proof. Let $x \in X$, and let $i \in I$. Then $\nabla d_{C_i}(x) = d_C^{-1}(x)(x - P_{C_i}x)$ if $x \notin C_i$ and $0 \in \partial d_{C_i}(x)$ otherwise. Hence

$$\nabla d^p_{C_i}(x) = p d^{p-2}_{C_i}(x)(x - P_{C_i}x) \tag{18}$$

if $x \notin C_i$, and $0 \in \partial d^p_{C_i}(x)$ otherwise. The result follows. ■

Example 4.5 Let $p \geq 1$ and suppose that $f = d^p_C$. Then $G = (1 - \frac{1}{p})\,\mathrm{Id} + \frac{1}{p}P_C$.

Proof. This follows from Proposition 4.4 when I is a singleton. ■

Example 4.6 Suppose that $u \in X$ satisfies $\|u\| = 1$, and let $\beta \in \mathbb{R}$. Then the following hold:

(i) If $f\colon x \mapsto \langle u, x\rangle - \beta$, then $C = \big\{x \in X \;\big|\; \langle u, x\rangle \leq \beta\big\}$ and $G\colon x \mapsto x - (\langle u, x\rangle - \beta)^+ u$.

(ii) If $f\colon x \mapsto |\langle u, x\rangle - \beta|$, then $C = \big\{x \in X \;\big|\; \langle u, x\rangle = \beta\big\}$ and $G\colon x \mapsto x - (\langle u, x\rangle - \beta)u$.

[5] Recall that $T\colon X \to X$ is *nonexpansive* if $(\forall x \in X)(\forall y \in X)\ \|Tx - Ty\| \leq \|x - y\|$.

Proof. (i): Note that $f^+ = d_C$ and hence $G = P_C$ by Proposition 3.1(vii) and Example 4.5. (ii): Here $f = d_C$ and hence $G = P_C$ by Example 4.5. ■

Remark 4.7 Using Example 4.5, we see that G is linear and that $G = G^*$ provided that $f = d_C^p$, where $p \geq 1$ and C is a subspace. The converse is true as well but this lies beyond the scope of this paper.

We now give two examples in which G is positively homogenenous but not necessarily linear.

Example 4.8 Suppose that f is a norm on X, with duality mapping $J = \partial \frac{1}{2} f^2$. Then $C = \{0\}$ and $(\forall x \in X \smallsetminus \{0\})\ Gx = x - f^2(x)\|Jx\|^{-2}Jx$.

Example 4.9 Let K be a nonempty closed convex cone with polar cone $K^\ominus$, and suppose that $f\colon x \mapsto \frac{1}{2}\langle x, P_K x\rangle$. Then $G = \operatorname{Id} - \frac{1}{2}P_K = P_{K^\ominus} + \frac{1}{2}P_K$.

Proof. Since $(\forall x \in X)\ f(x) = \frac{1}{2}\|P_K x\|^2 = \frac{1}{2}d^2_{K^\ominus}(x)$, it follows that $\nabla f(x) = x - P_{K^\ominus}x = P_K x$. The formula then follows. ■

A direct verification yields the following result which is known when $p = 2$ (see [13] and [15]).

Proposition 4.10 *Let Y be another real Hilbert space, let $A\colon X \to Y$ be continuous and linear, let $b \in Y$, and let $\varepsilon \geq 0$, and let $p \geq 1$. Suppose that $(\forall x \in X)\ f(x) = \|Ax - b\|^p - \varepsilon^p$ and that $C = \{x \in X \mid \|Ax - b\| \leq \varepsilon\} \neq \varnothing$. Then*

$$(19)\qquad (\forall x \in X)\quad Gx = \begin{cases} x - \dfrac{\|Ax-b\|^p - \varepsilon^p}{p\|Ax-b\|^{p-2}\|A^*(Ax-b)\|^2}A^*(Ax-b), & \text{if } \|Ax-b\| > \varepsilon; \\ x, & \text{otherwise.} \end{cases}$$

5 Continuity of G vs Fréchet differentiability of f

In this section, we investigate the continuity of G, which is a desirable property when G is used as an operator in an algorithm. In turns out that strong-to-strong continuity of G corresponds precisely to Fréchet differentiability of f. We start with a technical result.

Lemma 5.1 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X converging weakly to $\bar{x}$ and such that $x_n - Gx_n \to 0$. Suppose that one of the following holds:*

(i) $x_n \to \bar{x}$.

(ii) *f is bounded on every bounded subset of X.*

Then $\bar{x} \in C$.

Proof. Because of either [3, Proposition 16.14] or [3, Proposition 16.17] there exists $\rho > 0$ such that $\sigma := \sup \|\partial f(\operatorname{ball}(\bar{x};\rho))\| < +\infty$. We thus can and do assume that

$$(20)\qquad (\forall n \in \mathbb{N})\quad \|s(x_n)\| \leq \sigma.$$

Since f^+ is weakly lower semicontinuous, we deduce from Fact 2.1(vi) that

$$(21)\qquad f^+(\bar{x}) \leq \varliminf f^+(x_n) \leq \sigma \varliminf \|x_n - Gx_n\| = 0.$$

Hence $f(\bar{x}) \leq 0$, i.e., $\bar{x} \in C$. ■

Remark 5.2 Lemma 5.1(i) and Fact 2.1(ii) imply that G is *fixed-point closed* at $\bar{x}$ (see, e.g., also [8, Theorem 4.2.7] or [2]), i.e., if $x_n \to \bar{x}$ and $x_n - Gx_n \to 0$, then $\bar{x} = G\bar{x}$.

Proposition 5.3 *G is continuous at every point in C.*

Proof. Let $\bar{x} \in C$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X converging to $\bar{x}$. The result is clear if $(x_n)_{n\in\mathbb{N}}$ lies in C, so we can and do assume that $(x_n)_{n\in\mathbb{N}}$ lies in $X \smallsetminus C$. Then $(\forall n \in \mathbb{N})\ f(x_n) \leq f(\bar{x}) - \langle s(x_n), \bar{x} - x_n\rangle \leq \langle s(x_n), x_n - \bar{x}\rangle \leq \|s(x_n)\|\|\bar{x} - x_n\|$. Hence $0 < f(x_n)/\|s(x_n)\| \leq \|\bar{x} - x_n\| \to 0$. By Fact 2.1(vi), $x_n - Gx_n \to 0$. Thus $\lim Gx_n = \lim x_n = \bar{x} = G\bar{x}$ using Fact 2.1(ii). ■

The continuity of G outside C is more delicate.

Fact 5.4 (Smulyan) (See, e.g., [7, Proposition 6.1.4].) *The following hold:*

(i) *f is Fréchet differentiable at $\bar{x} \Leftrightarrow s$ is (strong-to-strong) continuous at $\bar{x}$.*

(ii) *f is Gâteaux differentiable at $\bar{x} \Leftrightarrow s$ is strong-to-weak continuous at $\bar{x}$.*

Lemma 5.5 *Suppose that $\bar{x} \in X \smallsetminus C$, that G is strong-to-weak continuous at $\bar{x}$, but G is not strong-to-strong continuous at $\bar{x}$. Then f is not Gâteaux differentiable at $\bar{x}$.*

Proof. There exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $X \smallsetminus C$ such that $x_n \to \bar{x}$, $Gx_n \rightharpoonup G\bar{x}$ yet $Gx_n \not\to G\bar{x}$. It follows that

(22) $$x_n - Gx_n \rightharpoonup \bar{x} - G\bar{x} \quad \text{and} \quad x_n - Gx_n \not\to \bar{x} - G\bar{x}.$$

By the Kadec–Klee property[6]of X, $\|x_n - Gx_n\| \not\to \|\bar{x} - G\bar{x}\|$. Since $\|\cdot\|$ is weakly lower semicontinuous, we assume (after passing to a subsequence and relabeling if necessary) that

(23) $$\|\bar{x} - G\bar{x}\| < \eta := \lim_{n\in\mathbb{N}} \|x_n - Gx_n\|.$$

Using Fact 2.1(viii), it follows that

(24) $$s(x_n) = f(x_n)\frac{x_n - Gx_n}{\|x_n - Gx_n\|^2} \rightharpoonup f(\bar{x})\frac{\bar{x} - G\bar{x}}{\eta^2} \neq f(\bar{x})\frac{\bar{x} - G\bar{x}}{\|\bar{x} - G\bar{x}\|^2} = s(\bar{x}).$$

Thus, s is not strong-to-weak continuous at $\bar{x}$. It follows now from Fact 5.4(ii) that f is not Gâteaux differentiable at $\bar{x}$. ■

Theorem 5.6 *Let $\bar{x} \in X \smallsetminus C$. Then the following are equivalent:*

(i) *f is Fréchet differentiable at $\bar{x}$.*

(ii) *G is (strong-to-strong) continuous at $\bar{x}$.*

(iii) *f is Gâteaux differentiable at $\bar{x}$ and G is strong-to-weak continuous at $\bar{x}$.*

[6] A sequence $(y_n)_{n\in\mathbb{N}}$ in X converges to $\bar{y}$ if and only if $y_n \rightharpoonup \bar{y}$ and $\|y_n\| \to \|\bar{y}\|$.

Proof. "(i)⇒(ii)": By Fact 5.4(i), s is continuous at $\bar{x}$. It follows from the definition of G that G is continuous at $\bar{x}$ as well.

"(i)⇐(ii)": In view of Fact 2.1(viii), we have $s(x) = f(x)(x - Gx)/\|x - Gx\|^2$ for all x sufficiently close to $\bar{x}$. Hence s is continuous at $\bar{x}$ and therefore f is Fréchet differentiable at $\bar{x}$ by Fact 5.4(i).

"(i)⇒(iii)" and "(ii)⇒(iii)": This is clear since (i)⇔(ii) by the above.

"(iii)⇒(ii)": Suppose to the contrary that G is not strong-to-strong continuous. Then, by Lemma 5.5, f is not Gâteaux differentiable at $\bar{x}$ which is absurd. ■

Corollary 5.7 (continuity) *G is continuous everywhere if and only if f is Fréchet differentiable on $X \smallsetminus C$.*

Proof. Combine Proposition 5.3 with Theorem 5.6. ■

Example 5.8 Suppose that $X = \mathbb{R}$ and that $(\forall x \in \mathbb{R})\ f(x) = \max\{-x, x, 2x - 1\}$. Then $C = \{0\}$ and f is not differentiable at 1; consequently, by Corollary 5.7, G is not continuous at 1.

Remark 5.9 (weak-to-weak continuity) It is unrealistic to expect that G is weak-to-weak continuous even when f is Fréchet differentiable; see [2, Example 3.2 and Remark 3.3.(ii)].

6 Continuity of G vs Gâteaux differentiability of f

In view of Fact 5.4 and Corollary 5.7, it is now tempting to conjecture that G is strong-to-weak continuous if and only if f is Gâteaux differentiable on $X \smallsetminus C$. Perhaps somewhat surprisingly, this turns out to be wrong. The counterexample is based on an ingenious construction by Borwein and Fabian [6].

Example 6.1 (Borwein–Fabian) (See [6, Proof of Theorem 4].) Suppose that X is infinite-dimensional. Then there exists a function $b\colon X \to \mathbb{R}$ such that the following hold:

(i) b is continuous, convex and $\min b(X) = b(0) = 0$.

(ii) b is Fréchet differentiable on $X \smallsetminus \{0\}$.

(iii) b is Gâteaux differentiable at 0, and $\nabla b(0) = 0$.

(iv) b is not Fréchet differentiable at 0.

Example 6.2 (lack of strong-to-weak continuity) Let b be as in Example 6.1. Then there exists $y \in X$ such that $\nabla b(y) \neq 0$. Suppose that

$$(\forall x \in X) \quad f(x) = b(x) - \langle \nabla b(y), x\rangle - \tfrac{1}{2}\big(b(y) - \langle \nabla b(y), y\rangle\big). \tag{25}$$

Then the following hold:

(i) f is Gâteaux differentiable (but not Fréchet differentiable) at 0, and G is not strong-to-weak continuous at 0.

(ii) f is Fréchet differentiable on $X \smallsetminus \{0\}$, and G is continuous on $X \smallsetminus \{0\}$.

Proof. By Example 6.1(iii), $0 \in \operatorname{ran} \nabla b$. If $\{0\} = \operatorname{ran} \nabla b$, then we would deduce that b is constant and therefore Fréchet differentiable; in turn, this would contradict Example 6.1(iv). Hence $\{0\} \subsetneqq \operatorname{ran} \nabla b$ and there exists $y \in X$ such that

$$v = \nabla b(y) \neq 0. \tag{26}$$

Now set

$$g \colon X \to \mathbb{R} \colon x \mapsto b(x) - \langle v, x \rangle . \tag{27}$$

Then

$$(\forall x \in X) \quad f(x) = g(x) - \tfrac{1}{2} g(y), \tag{28}$$

and $g(0) = b(0) - \langle v, 0 \rangle = 0$ by Example 6.1(i). Example 6.1(iii) and (26) yield $\nabla g(0) = \nabla b(0) - v = -v \neq 0$ while $\nabla g(y) = \nabla b(y) - v = 0$. Hence $\min g(X) = g(y) < g(0) = 0$ and therefore

$$f(y) = \min f(X) = \min g(X) - \tfrac{1}{2} g(y) = \tfrac{1}{2} g(y) < 0 < 0 - \tfrac{1}{2} g(y) = f(0). \tag{29}$$

Thus $y \in C$ while $0 \notin C$.

(i): On the one hand, since b is not Fréchet differentiable at 0 (Example 6.1(iv)), neither is f. On the other hand, since b is Gâteaux differentiable at 0 (Example 6.1(iii)), so is f. Altogether, f is Gâteaux differentiable, but not Fréchet differentiable, at 0. Therefore, by Theorem 5.6, G is not strong-to-weak continuous at 0.

(ii): Since b is Fréchet differentiable on $X \smallsetminus \{0\}$ (Example 6.1(ii)), so is f. Now apply Theorem 5.6. ■

7 G as an accelerated mapping

In this section, we consider the case when f is a power of a quadratic form. We link G to the accelerated mapping corresponding to a linear operator. This mapping can significantly speed up convergence of algorithms (see [5]).

Proposition 7.1 *Suppose that $f \colon x \mapsto \sqrt{\langle x, Mx \rangle^p}$, where $p \geq 1$ and $M \colon X \to X$ be continuous, linear, self-adjoint, and positive. Then G is continuous everywhere and*

$$(\forall x \in X) \quad Gx = \begin{cases} x - \dfrac{\langle x, Mx \rangle}{p \|Mx\|^2} Mx, & \text{if } Mx \neq 0; \\ x, & \text{if } Mx = 0. \end{cases} \tag{30}$$

Proof. Assume first that $p = 1$. Since M has a unique positive square root, i.e., there exists[7] $B \colon X \to X$ such that B is continuous, linear, self-adjoint, and positive, and $\ker B = \ker M$. Hence $(\forall x \in X)$ $f(x) = \sqrt{\langle x, Mx \rangle} = \|Bx\|$ so f is indeed convex and continuous. If $x \in X \smallsetminus \ker M = X \smallsetminus \ker B$, then f is Fréchet differentiable at x with $\nabla f(x) = B^* Bx / \|Bx\| = Mx / \|Bx\|$; hence,

$$Gx = x - \frac{\|Bx\|}{\|Mx\|^2 / \|Bx\|^2} \frac{Mx}{\|Bx\|} = x - \frac{\|Bx\|^2}{\|Mx\|^2} Mx = x - \frac{\langle x, Mx \rangle}{\|Mx\|^2} Mx \tag{31}$$

and G is continuous everywhere by Corollary 5.7. If $p > 1$, then the result follows from the above and Proposition 3.1(iii). ■

[7] See, e.g., [18, Theorem 9.4-2], where this is stated in a complex Hilbert space; however, the proof works unchanged in our real setting as well.

Example 7.2 Let $A\colon X \to X$ be linear, self-adjoint, and nonexpansive. Suppose that $(\forall x \in X)$ $f(x) = \sqrt{\langle x, x - Ax\rangle}$. Then G is positively homogeneous, continuous everywhere, and

$$(32) \qquad (\forall x \in X) \quad Gx = \begin{cases} x - \dfrac{\langle x, x - Ax\rangle}{\|x - Ax\|^2}(x - Ax), & \text{if } Ax \neq x; \\ x, & \text{if } Ax = x. \end{cases}$$

Proof. Use Proposition 7.1 with $M = \operatorname{Id} - A$ and $p = 1$. ■

Remark 7.3 (accelerated mapping) Let $A\colon X \to X$ be linear, nonexpansive, and self-adjoint. In [5], the authors study the accelerated mapping[8] of A, i.e.,

$$(33) \qquad x \mapsto t_x Ax + (1 - t_x)x, \quad \text{where } t_x = \begin{cases} \dfrac{\langle x, x - Ax\rangle}{\|x - Ax\|^2}, & \text{if } x \neq Ax; \\ 1, & \text{otherwise.} \end{cases}$$

In view of the Example 7.2, the accelerated mapping of A is precisely the subgradient projector G of the function $x \mapsto \sqrt{\langle x, x - Ax\rangle}$. Now suppose that $X = \ell^2(\mathbb{N})$, let $(e_n)_{n\in\mathbb{N}}$ be the standard orthonormal basis of X, and suppose that

$$(34) \qquad A\colon X \to X\colon x \mapsto \sum_{n\in\mathbb{N}} \tfrac{n}{n+1} \langle e_n, x\rangle \, e_n.$$

Then G is continuous (Example 7.2); however, G is neither linear nor uniformly continuous (see the [5, Remark following Lemma 3.8]).

8 Nonexpansiveness

We now discuss when G is (firmly) nonexpansive or monotone. These properties occur when studying resolvents, subdifferentials and gradients. For instance, every resolvent and every proximal mapping is firmly nonexpansive; subdifferential (or gradient) operators of convex functions are monotone. However, these properties are not automatic for subgradient projectors as we will see in this section.

Proposition 8.1 *Suppose that f is Gâteaux differentiable on $X \smallsetminus C$ and that G_f is firmly nonexpansive. Then G_g is likewise in each of the following situations:*

(i) $\alpha > 0$, *and* $g = f \circ \alpha \operatorname{Id}$ *is convex.*

(ii) $f \geq 0$, $\alpha \geq 1$, *and* $g = f^\alpha$ *is convex.*

(iii) $A\colon X \to X$ *is continuous and linear,* $AA^* = A^*A = \operatorname{Id}$, *and* $g = f \circ A$.

(iv) $z \in X$ *and* $g\colon x \mapsto f(x - z)$.

The analogous statement holds when G_f is assumed to be nonexpansive.

[8] In fact, the operator A in [5] need not necessarily be self-adjoint.

Proof. This follows from the corresponding items in Proposition 3.1, which do preserve (firm) nonexpansiveness. ■

On the real line, we obtain a simpler test.

Proposition 8.2 *Suppose that $X = \mathbb{R}$ and that f is twice differentiable on $X \smallsetminus C$. Then G is monotone. Moreover, G is (firmly) nonexpansive if and only if*

$$(\forall x \in \mathbb{R}) \quad f(x)f''(x) \leq \big(f'(x)\big)^2. \tag{35}$$

Proof. By Corollary 5.7, G is continuous. Let $x \in \mathbb{R} \smallsetminus C$. Then $G(x) = x - f(x)/f'(x)$ and hence $G'(x) = f(x)f''(x)/(f'(x))^2 \geq 0$. It follows that G is increasing on $\mathbb{R} \smallsetminus C$ and hence on $\mathbb{R}$. Furthermore, G is (firmly) nonexpansive if and only if $G'(x) \leq 1$, which gives the remaining characterization. ■

Example 8.3 Suppose that $X = \mathbb{R}$, let $\alpha > 0$, and suppose that $(\forall x \in \mathbb{R})\ f(x) = x^n - \alpha$, where $n \in \{2, 4, 6, 8, \ldots\}$. Then G is firmly nonexpansive.

Proof. If $x \in \mathbb{R} \smallsetminus C$, then $(f'(x))^2 - f(x)f''(x) = nx^{n-2}(\alpha n + x^n - \alpha) > 0$ and we are done by Proposition 8.2. ■

Example 8.4 Suppose that $X = \mathbb{R}$ and that $f\colon x \mapsto \exp(|x|) - 1$. Then $(\forall x \in X)\ G(x) = x - \operatorname{sgn}(x)(1 - \exp(-|x|))$ and $G'(x) = 1 - \exp(-|x|) \in [0, 1[$. It follows that G is firmly nonexpansive[9].

Example 8.5 Suppose that $X = \mathbb{R}$ and that $f\colon x \mapsto \exp(x^2) - 1$. Then G is not (firmly) nonexpansive. Indeed, we compute $(f'(x))^2 - f(x)f''(x) = 4x^2\exp(x^2) + 2\exp(x^2) - 2\exp(2x^2)$, which strictly negative when $|x| > 1.2$. Now apply Proposition 8.2.

Proposition 8.6 *Suppose that $X = \mathbb{R}$ and that f is twice differentiable, that $\min f(X) = 0$, that $g = f\Box(1/2)|\cdot|^2$, and that $2ff'' \leq (2 + f'')(f')^2$. Then G_g is firmly nonexpansive.*

Proof. We start by observing a couple of facts. First,

$$g' = \operatorname{Id} - \mathrm{P}_f. \tag{36}$$

Write $y = \mathrm{P}_f(x)$. Then $x = y + f'(y)$ and hence implicit differentiation gives $1 = y'(x) + f''(y)y'(x) = y'(x)(1 + f''(y(x)))$. Hence $y' = 1/(1 + f''(y(x)))$ and thus

$$g''(x) = \big(\operatorname{Id} - \mathrm{P}_f\big)'(x) = 1 - \frac{1}{1 + f''(\mathrm{P}_f(x))} = \frac{f''\big(\mathrm{P}_f(x)\big)}{1 + f''\big(\mathrm{P}_f(x)\big)}. \tag{37}$$

In view of Proposition 8.2 and because $g(x) = f(\mathrm{P}_f(x)) + (1/2)(x - \mathrm{P}_f(x))^2$ we must verify that $gg'' \leq (g')^2$, i.e.,

$$\frac{\big(f(\mathrm{P}_f(x)) + \frac{1}{2}(x - \mathrm{P}_f(x))^2\big)f''\big(\mathrm{P}_f(x)\big)}{1 + f''\big(\mathrm{P}_f(x)\big)} \leq \big(x - \mathrm{P}_f(x)\big)^2. \tag{38}$$

[9] Since G is monotone by Proposition 8.2, its antiderivative $x \mapsto \frac{1}{2}x^2 - |x| - \exp(-|x|)$ is convex — although this does not look like convex function on first glance! It is interesting to do this also for other instances of f.

Again writing $y = \mathrm{P}_f(x)$ gives $x - \mathrm{P}_f(x) = f'(y)$ and so see that (38) is equivalent to

(39) $$\frac{\big(f(y) + \frac{1}{2}(f'(y))^2\big)f''(y)}{1 + f''(y)} \leq \big(f'(y)\big)^2.$$

However, (39) holds by our assumption on f. ■

We conclude this section with a result on the range of $\mathrm{Id} - G$.

Proposition 8.7 *We have* $\operatorname{ran}(\mathrm{Id} - G) \subseteq \operatorname{cone}\operatorname{ran}\partial f \subseteq (\operatorname{rec} C)^{\ominus}$.

Proof. Let $y^* \in \partial f(y)$, let $c \in C$, and let $x \in \operatorname{rec} C$. Then $(c + nx)_{n\in\mathbb{N}}$ lies in C. Hence $(\forall n \geq 1)$ $0 \geq f(c + nx) \geq f(y) + \langle y^*, c + nx - y\rangle$ and thus

(40) $$\langle y^*, x\rangle \leq \frac{\langle y^*, y - c\rangle - f(y)}{n} \to 0 \quad \text{as } n \to +\infty.$$

It follows that $y^* \in (\operatorname{rec} C)^{\ominus}$. Therefore, $\operatorname{ran}(\mathrm{Id} - G) \subseteq \operatorname{cone}\operatorname{ran}\partial f \subseteq (\operatorname{rec} C)^{\ominus}$. ■

9 The decreasing property

We say that f has the *decreasing property* if

(41) $$(\forall x \in X) \quad \sup f(\mathcal{G}x) \leq f(x).$$

This property is interesting in the context of applying subgradient projectors as building blocks for algorithms. To investigate the decreasing property, it suffices to consider points outside C.

Proposition 9.1 *If* $(\forall x \in X)$ $Gx \in \operatorname{conv}(\{x\} \cup C)$*, then f has the decreasing property.*

Proof. Let $x \in X \smallsetminus C$. Then there exists $c \in C$ and $\lambda \in [0, 1]$ such that $Gx = (1 - \lambda)x + \lambda c$. It follows that $f(Gx) \leq (1 - \lambda)f(x) + \lambda f(c) \leq (1 - \lambda)f(x) \leq f(x)$. ■

Lemma 9.2 *Let* $(x, y, z) \in \mathbb{R}^3$ *be such that* $x \neq z$ *and* $(z - y)(x - y) \leq 0$*. Then* $y \in \operatorname{conv}\{x, z\}$*.*

Proof. Suppose first that $z < x$. If $y > x$, then $(z - y)(x - y) > 0$ because it is the product of two strictly negative numbers. Similarly, if $y < z$, then $(z - y)(x - y) > 0$. We deduce that $y \in [z, x]$. Analogously, when $x < z$, we obtain that $y \in [x, z]$. In either case, $y \in \operatorname{conv}\{x, z\}$. ■

The next result shows that G is particularly well behaved on the real line.

Corollary 9.3 *Suppose that* $X = \mathbb{R}$*. Then f has the decreasing property.*

Proof. Let $x \in \mathbb{R} \smallsetminus C$. Then $x \neq P_C x$ and, by Fact 2.1(iv), $(P_C x - Gx)(x - Gx) \leq 0$. Lemma 9.2 thus yields $Gx \in \operatorname{conv}\{x, P_C x\}$. Hence $Gx \in \operatorname{conv}(\{x\} \cup C)$, and we are done by Proposition 9.1. ■

The next example shows that the decreasing property is not automatic.

Example 9.4 Suppose that $X = \mathbb{R}^2$, that $C_1 = \mathbb{R} \times \{0\}$, that $C_2 = \big\{(\xi, \xi) \in X \;\big|\; \xi \in \mathbb{R}\big\}$, and that $f = \max\{d_{C_1}, d_{C_2}\}$. Then f does not have the decreasing property.

Proof. Set $x = (2,1)$. Then, using Proposition 4.3, we obtain that $Gx = (2,0)$ and $f(x) = 1 < \sqrt{2} = f(Gx)$. ■

We now illustrate that the sufficient condition of Proposition 9.1 is not necessary:

Example 9.5 Suppose that $X = \mathbb{R}^2$ and that $(\forall x = (x_1, x_2) \in \mathbb{R}^2)\ f(x) = |x_1| + |x_2|$. Then f has the decreasing property, $G^2 x = (0,0)$ yet $Gx \notin \operatorname{conv}\{(0,0), x\}$ for almost every $x \in \mathbb{R}^2$. Furthermore, G is not monotone.

Proof. Observe that $C = \{(0,0)\}$. Let $I = \{1,2,3,4\}$ and consider the four halfspaces $(C_i)_{i\in I}$ with normal vectors $(1,1)$ and $(1,-1)$ with $(0,0)$ in their boundaries, and with the two boundary hyperplanes H_1 and H_2. Then $f = \sqrt{2}\max_{i\in I} d_{C_i} = \sqrt{2}\max\{d_{H_1}, d_{H_2}\}$ by Example 4.6(i). Proposition 4.3 implies that G is the projector onto the farther hyperplane on $\mathbb{R}^2 \smallsetminus S$, where $S = (\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R})$. It is thus clear that $Gx \notin \operatorname{conv}\{(0,0), x\}$ and that $f(Gx) \leq f(x)$ for every $x \in \mathbb{R}^2 \smallsetminus S$. When $x \in S$, one checks directly that $f(Gx) \leq f(x)$. Hence f has the decreasing property. Finally, let $x = (-1,3)$ and $y = (1,3)$. Then $Gx = (1,1)$ and $Gy = (-1,1)$ and hence $\langle x - y, Gx - Gy\rangle = -4 < 0$ so G is not monotone. ■

Remark 9.6 (infeasibility detection) Using the decreasing property, one obtains a sufficient condition for *infeasibility*: Suppose that $X = \mathbb{R}$ and we find a point x such that $f(Gx) > f(x)$. Then C must be empty because of Corollary 9.3. For instance, suppose that $f\colon x \mapsto x^2 + 1$. Then

$$(\forall x \in \mathbb{R} \smallsetminus \{0\}) \quad Gx = (x^2 - 1)/(2x). \tag{42}$$

Now set $x = 1/2$. Then $Gx = -3/4$ and $f(Gx) = 25/16 > 5/4 = f(x)$.

Remark 9.7 (Newton iteration) Suppose that $X = \mathbb{R}$ and that f is differentiable on $X \smallsetminus C$. Then

$$(\forall x \in \mathbb{R} \smallsetminus C) \quad Gx = x - \frac{f(x)}{\big(f'(x)\big)^2} f'(x) = x - \frac{f(x)}{f'(x)} \tag{43}$$

is the same as the Newton operator for finding a zero of f! It is known since the 19th century that the concrete instance (42) exhibits chaotic behaviour; see, e.g., [19, Problem 7-a on page 72].

The decreasing property is preserved in certain cases:

Proposition 9.8 *Suppose that f has the decreasing property. Then the following hold:*

(i) *If $\alpha > 0$, then αf has the decreasing property.*

(ii) *If $\alpha \geq 1$, then $(f^+)^\alpha$ has the decreasing property.*

Proof. Let $x \in X \smallsetminus C$ and let $\alpha > 0$. (i): Then $(\alpha f)\mathcal{G}_{\alpha f}(x) = (\alpha f)\mathcal{G}_f(x)$ and hence $\sup(\alpha f)(\mathcal{G}_{\alpha f}(x)) = \alpha \sup f(\mathcal{G}_f(x)) \leq \alpha f(x) = (\alpha f)(x)$ by Proposition 3.1(i). (ii): Set $g = (f^+)^\alpha$ and $\beta = 1/\alpha$. Then $0 < \beta \leq 1$ and $\mathcal{G}_g(x) = (1-\beta)x + \beta\mathcal{G}_f(x)$ by Proposition 3.1(iii). Hence $\sup g(\mathcal{G}_g x) \leq (1-\beta)g(x) + \beta \sup g(\mathcal{G}_f x)$. On the other hand, $\sup g(\mathcal{G}_f(x)) \leq g(x)$ by definition of g. Altogether, $\sup g(\mathcal{G}_g x) \leq g(x)$, i.e., g is decreasing. ■

The following result is complementary to the decreasing property.

Proposition 9.9 *Suppose that f is strictly convex at $x \in X$ and $f(x) > 0$. Then $f(Gx) > 0$.*

Proof. Recall that f is strictly convex at x if $(\forall y \in X \smallsetminus \{x\})$ $(\forall \lambda \in]0,1[)$ $f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y)$. Arguing as in [7, proof of Proposition 5.3.4.(a)], we see that $\frac{1}{2}\langle s(x), Gx - x\rangle = \langle s(x), (\frac{1}{2}x + \frac{1}{2}Gx) - x\rangle \leq f(\frac{1}{2}x + \frac{1}{2}Gx) - f(x) < \frac{1}{2}f(x) + \frac{1}{2}f(Gx) - f(x) = \frac{1}{2}(f(Gx) - f(x))$. Therefore, $f(Gx) > f(x) + \langle s(x), Gx - x\rangle = 0$ using Fact 2.1(i). ■

Remark 9.10 Suppose that f is strictly convex. Then Proposition 9.9 shows that iterating G starting at a point outside C will never reach C in finitely many steps. This is clearly illustrated by Example 4.5, which shows that the function d_C, even though it is neither strictly convex nor differentiable everywhere, performs best because $G = P_C$ yields a solution after just one step.

10 The subgradient projector of $(x_1, x_2) \mapsto |x_1|^p + |x_2|^p$

This section contains a case study. It reveals that the various properties of G can exhibit complicated behaviour. The following result complements Example 9.5.

Proposition 10.1 *Suppose that* $X = \mathbb{R}^2$ *and that* $f\colon (x_1, x_2) \mapsto |x_1|^p + |x_2|^p$, *where* $p > 1$, *and let* $x = (x_1, x_2) \in \mathbb{R}^2 \smallsetminus \{(0,0)\}$. *Then*

$$(44)\qquad Gx = \left(x_1 - \frac{\big(|x_1|^p + |x_2|^p\big)|x_1|^{p-1}\operatorname{sgn}(x_1)}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)}, x_2 - \frac{\big(|x_1|^p + |x_2|^p\big)|x_2|^{p-1}\operatorname{sgn}(x_2)}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)}\right)$$

and the following hold:

(i) *If* $p \geq 2$, *then* $f(x) \geq f(Gx) \geq (1 - 2p^{-1})^p f(x)$.

(ii) *If* $1 < p \leq 2$, *then* $f(x) \geq f(Gx) \geq 2^{-1}(1 - p^{-1})^p f(x)$.

(iii) *If* $1 < p < 2$, *then* G *is not monotone.*

Proof. The formula (44) is a direct verification, and (i)&(ii) hold when $x_1 = 0$ or $x_2 = 0$. We thus assume that $x_1 \neq 0$ and $x_2 \neq 0$.

(i): Note that

$$(45)\qquad f(Gx) = |x_1|^p\big|1 - c_1\big|^p + |x_2|^p\big|1 - c_2\big|^p, \quad \text{where} \quad c_i = \frac{\big(|x_1|^p + |x_2|^p\big)|x_i|^{p-2}}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)}.$$

If $i \in \{1,2\}$ and $m \in \{1,2\}$ is such that $|x_m| = \max\{|x_1|, |x_2|\}$, then $c_i \leq (2|x_m|^p|x_m|^{p-2})p^{-1}(|x_m|^p + 0)^{-1} = 2/p$. Hence $1 \geq 1 - c_i \geq 1 - 2p^{-1} \geq 0$ and the inequalities follow.

(ii): We assume that $|x_1| \leq |x_2|$, the other case is treated analogously. Set $t = |x_1/x_2|$,

$$(46)\qquad c_1 = t - \frac{t^{2p-1} + t^{p-1}}{p\big(1 + t^{2p-2}\big)} \quad \text{and} \quad c_2 = 1 - \frac{1 + t^p}{p\big(1 + t^{2p-2}\big)},$$

and check that

$$(47)\qquad f(Gx) = |x_2|^p\big(|c_1|^p + |c_2|^p\big).$$

Since $p-2 \le 0$, we have $t^{p-2} \ge 1$ and hence

$$1 \ge c_2 \ge 1 - \frac{1+t^p}{p(1+t^p)} = 1 - \tfrac{1}{p} \ge 0. \tag{48}$$

Thus $c_2 \ge 0$. We now claim that

$$|c_1| + c_2 \le 1. \tag{49}$$

This will imply $\max\{|c_1|, |c_2|\} \le 1$; hence $\max\{|c_1|^p, |c_2|^p\} \le 1$,

$$f(Gx) \le |x_2|^p\big(|c_1| + |c_2|\big) \le |x_2|^p \le f(x), \tag{50}$$

and the decreasing property of f follows. Observe that (49) is equivalent to

$$c_1 + c_2 \le 1 \tag{51a}$$
$$-c_1 + c_2 \le 1 \tag{51b}$$

and hence to

$$t \le \frac{(1+t^p)(1+t^{p-1})}{p(1+t^{2p-2})} \tag{52a}$$
$$\frac{t^{p-1}(1+t^p)}{p(1+t^{2p-2})} \le t + \frac{1+t^p}{p(1+t^{2p-2})}. \tag{52b}$$

Now check that (52) holds by using $t^{p-1} \le 1$ and, for (52a), the convexity of $h\colon \xi \mapsto 1+\xi^p$, which implies $h(t) \ge h(1) + h'(1)(t-1)$, i.e., $pt \le 1+t^p$. Furthermore, using (47), (48) and the assumption that $|x_2| \ge |x_1|$, we obtain

$$f(Gx) \ge c_2^p|x_2|^p \ge \big(1-\tfrac{1}{p}\big)^p|x_2|^p \ge \big(1-\tfrac{1}{p}\big)^p \frac{|x_1|^p + |x_2|^p}{2} = \frac{\big(1-\frac{1}{p}\big)^p}{2} f(x). \tag{53}$$

(iii): Consider the points $y = (1,\xi)$ and $z = (-1,\xi)$, where $\xi > 0$. Then $y - z = (2,0)$ and

$$Gy = \left(1 - \frac{1+\xi^p}{p(1+\xi^{2p-2})}, \xi - \frac{(1+\xi^p)\xi^{p-1}}{p(1+\xi^{2p-2})}\right) \tag{54a}$$

and

$$Gz = \left(-1 + \frac{1+\xi^p}{p(1+\xi^{2p-2})}, \xi - \frac{(1+\xi^p)\xi^{p-1}}{p(1+\xi^{2p-2})}\right). \tag{54b}$$

It follows that

$$\langle Gy - Gz, y - z\rangle = 4\left(1 - \frac{1+\xi^p}{p(1+\xi^{2p-2})}\right) < 0 \quad \text{as } \xi \to +\infty \tag{55}$$

because $\lim_{\xi\to+\infty}(1+\xi^p)p^{-1}/(1+\xi^{2p-2}) = \lim_{\xi\to+\infty}(2p-2)^{-1}\xi^{2-p} = +\infty$ using l'Hôpital's rule. Therefore, G is not monotone. ■

Remark 10.2 The operator G of Proposition 10.1 seems to defy an easy analysis. It would be interesting to obtain complete characterizations in terms of p of the following, increasingly more restrictive, properties: G is monotone; $\mathrm{Id} - G$ is nonexpansive; G is firmly nonexpansive. With the help of Maple it is possible to check the following statements:

(i) If $p \in \{2,4,6\}$, then G is firmly nonexpansive and hence monotone.

(ii) If $p \in \{8,10,12\}$, then G is not firmly nonexpansive; however, $\mathrm{Id} - G$ is nonexpansive and G is monotone [10].

Suppose first that $p \in \{2,4,6\}$. Then G is firmly nonexpansive $\Leftrightarrow$ $N = 2G - \mathrm{Id}$ is nonexpansive $\Leftrightarrow$ $(\forall x \in X)$ Jx is nonexpansive, where Jx is the Jacobian of N at x $\Leftrightarrow$ $(Jx)^*Jx \preceq \mathrm{Id}$ $\Leftrightarrow$ $\mathrm{Id} - (Jx)^*Jx$ is positive semidefinite. The last condition leads to checking three inequalities using the principal minor criterion for positive semidefiniteness. Dividing by appropriate powers of x_1 and x_2, this reduces to checking whether three polynomials in one variable are positive. Sturm's Theorem (see, e.g., [26, Theorem 1.4.3]), which is implemented in Maple and Mathematica, combinded with [26, Theorem 1.1.2] finally complete the verification.

Now suppose that $p \in \{8,10,12\}$. The approach just outlined shows that G is not firmly nonexpansive. Note the implications: G is monotone $\Leftarrow$ $N = \mathrm{Id} - G$ is nonexpansive $\Leftrightarrow$ $(\forall x \in X)$ Jx is nonexpansive, where Jx is the Jacobian of N at x $\Leftrightarrow$ $(Jx)^*Jx \preceq \mathrm{Id}$ $\Leftrightarrow$ $\mathrm{Id} - (Jx)^*Jx$ is positive semidefinite, which is checked using again Sturm's Theorem.

11 G and the Yamagishi–Yamada operator

In this last section we study the accelerated version[11] of G proposed by Yamagishi and Yamada in [32]. Their operator, which has shown improved performance compared to G, is actually a subgradient projector of a variant of f when X is the real line[12]. For fixed $L > 0$ and $r > 0$, we assume in addition that

$$f \text{ is Fréchet differentiable and } \nabla f \text{ is Lipschitz continuous with constant } L, \tag{56}$$

and that

$$f \text{ is bounded below with } \inf f(X) \geq -\rho, \tag{57}$$

and we set

$$(\forall x \in X) \quad \theta(x) = \frac{\|\nabla f(x)\|^2}{2L} - \rho. \tag{58}$$

By [32, Lemma 1], we have

$$f \geq \theta. \tag{59}$$

The Yamagishi–Yamada operator [32] is

$$Z\colon X \to X, \tag{60}$$

[10] Experiments with Maple suggest that this pattern may hold true for every even integer greater than or equal 8.

[11] See also [20] for another accelerated version of G.

[12] Unfortunately our proof does not seem to extend to X (e.g., the set D may not be convex and there is no obvious counterpart to (63)).

defined at $x \in X$ by

$$Zx = \begin{cases} x, & \text{if } f(x) \le 0; \\ x - \dfrac{\nabla f(x)}{\|\nabla f(x)\|^2} f(x), & \text{if } f(x) > 0 \text{ and } \theta(x) \le 0; \\ x - \dfrac{\nabla f(x)}{\|\nabla f(x)\|^2} \Big(f(x) + \big(\sqrt{\theta(x)+\rho} - \sqrt{\rho}\big)^2 \Big), & \text{if } f(x) > 0 \text{ and } \theta(x) > 0. \end{cases} \tag{61}$$

Note that if $f(x) \le 0$ or $\theta(x) \le 0$, then $Zx = Gx$.

We now prove that if $X = \mathbb{R}$, then Z is itself a subgradient projector.

Theorem 11.1 *Suppose that $X = \mathbb{R}$ and that f is also twice differentiable. Then for every $x \in \mathbb{R}$,* (61) *can be rewritten as*

$$Zx = \begin{cases} x, & \text{if } f(x) \le 0; \\ x - \dfrac{1}{f'(x)} f(x), & \text{if } f(x) > 0 \text{ and } |f'(x)| \le \sqrt{2L\rho}; \\ x - \dfrac{1}{f'(x)} \left(f(x) + \left(\dfrac{|f'(x)|}{\sqrt{2L}} - \sqrt{\rho} \right)^2 \right), & \text{if } f(x) > 0 \text{ and } |f'(x)| > \sqrt{2L\rho}. \end{cases} \tag{62}$$

Set $D = \{x \in X \mid \theta(x) \le 0\}$ and assume that $\operatorname{bdry} D \subseteq X \smallsetminus C$. *Then D is a closed convex superset of C, and Z is a subgradient projector of a function y, defined as follows. On D, we set y equal to f. The set $\mathbb{R} \smallsetminus D$ is empty, or an open interval, or the disjoint union of two open intervals. Assume that I is one of these nonempty intervals, and let q be defined on I such that*

$$(\forall x \in I) \quad q'(x) = \frac{1}{x - Zx}. \tag{63}$$

Now set $d = P_D(I) \in D \smallsetminus C$ and

$$(\forall x \in I) \quad y(x) = \frac{f(d)}{e^{q(d)}} e^{q(x)}. \tag{64}$$

The so-constructed function $y\colon \mathbb{R} \to \mathbb{R}$ is convex, and it satisfies $Z = G_y$.

Proof. It is easy to check that (62) is the same as (61). Let $x \in \mathbb{R}$ such that $f(x) > 0$ and $\theta(x) \ge 0$, and set

$$z(x) = \frac{|f'(x)|}{\sqrt{2L}} - \sqrt{\rho} = \frac{\operatorname{sgn}\big(f'(x)\big) f'(x)}{\sqrt{2L}} - \sqrt{\rho} = \sqrt{\theta(x)+\rho} - \sqrt{\rho} \ge 0. \tag{65}$$

Then

$$z'(x) = \frac{\operatorname{sgn}\big(f'(x)\big) f''(x)}{\sqrt{2L}}. \tag{66}$$

Using the convexity of f, (59), (65), and (66), we obtain

$$0 \le f''(x)\big(f(x) - \theta(x)\big) \tag{67a}$$

$$= f''(x)\left(f(x) - \left(\frac{|f'(x)|}{\sqrt{2L}} + \sqrt{\rho}\right)\left(\frac{|f'(x)|}{\sqrt{2L}} - \sqrt{\rho}\right)\right) \tag{67b}$$

$$= f''(x)\left(f(x) + z(x)\left(z(x) - \frac{2|f'(x)|}{\sqrt{2L}}\right)\right) \tag{67c}$$

$$= f''(x)\big(f(x) + z^2(x)\big) - f'(x)2z(x)\frac{\operatorname{sgn}\big(f'(x)\big)f''(x)}{\sqrt{2L}} \tag{67d}$$

$$= f''(x)\big(f(x) + z^2(x)\big) - f'(x)\big(2z(x)z'(x)\big). \tag{67e}$$

Because $x - Zx = (f(x) + z^2(x))/f'(x)$ is continuous, it is clear that there is an antiderivative q on I such that

$$q'(x) = \frac{1}{x - Zx} = \frac{f'(x)}{f(x) + z^2(x)}. \tag{68}$$

Calculus and (67) now result in

$$q''(x) = \frac{f''(x)\big(f(x) + z^2(x)\big) - f'(x)\big(f'(x) + 2z(x)z'(x)\big)}{\big(f(x) + z^2(x)\big)^2} \tag{69a}$$

$$= \frac{f''(x)\big(f(x) - \theta(x)\big) - \big(f'(x)\big)^2}{\big(f(x) + z^2(x)\big)^2}. \tag{69b}$$

Observe that y is clearly continuous everywhere. Furthermore, $y'(x) = \frac{f(d)}{e^{q(d)}}e^{q(x)}q'(x)$ and hence, using (68), (69) and again (67), we obtain

$$y''(x) = \frac{f(d)}{e^{q(d)}}\Big(e^{q(x)}\big(q'(x)\big)^2 + e^{q(x)}q''(x)\Big) \tag{70}$$

$$= \frac{f(d)}{e^{q(d)}}e^{q(x)}\Big(\big(q'(x)\big)^2 + q''(x)\Big) \tag{71}$$

$$= y(x)\frac{f''(x)\big(f(x) - \theta(x)\big)}{\big(f(x) + z^2(x)\big)^2} \tag{72}$$

$$\geq 0. \tag{73}$$

Hence y is convex on I. As $x \in I$ approaches d, we deduce (because $d \notin C$, i.e., $f(d) > 0$) that $q'(x) \to f'(d)(f(d) + z^2(d))^{-1} \to f'(d)/f(d)$ and hence that $y'(x) \to f(d)/e^{q(d)}e^{q(d)}f'(d)/f(d) = f'(d)$. It follows that y is convex on $\mathbb{R}$. Finally, if $x \notin D$, then $G_y(x) = x - y(x)/y'(x) = x - 1/q'(x) = x - (x - Zx) = Zx$. ■

Example 11.2 Consider Theorem 11.1 and assume that $f\colon x \mapsto x^2 - 1$, that $L = 3$, and that $\rho = 1$. Then (62) turns into

$$Zx = \begin{cases} x, & \text{if } |x| \leq 1; \\[2mm] \dfrac{x^2+1}{2x}, & \text{if } 1 < |x| \leq \sqrt{6}/2; \\[2mm] \dfrac{x^2 + 2\sqrt{6}|x|}{6x}, & \text{if } |x| > \sqrt{6}/2. \end{cases} \tag{74}$$

Hence $D = \big[-\sqrt{6}/2, \sqrt{6}/2\big]$. Using elementary manipulations, we obtain

$$(\forall x \in \mathbb{R} \smallsetminus D) \quad q(x) = \tfrac{6}{5} \ln\big(\tfrac{5}{6}|x| - \tfrac{\sqrt{6}}{3}\big); \tag{75}$$

consequently, the function y, given by

$$(\forall x \in \mathbb{R}) \quad y(x) = \begin{cases} x^2 - 1, & \text{if } |x| \leq \sqrt{6}/2; \\ \dfrac{72^{1/5}}{6}\big(5|x| - 2\sqrt{6}\big)^{6/5}, & \text{if } |x| > \sqrt{6}/2, \end{cases} \tag{76}$$

satisfies $G_y = Z$ by Theorem 11.1.

Acknowledgments

The authors thank the referees, Patrick Combettes and Isao Yamada for insightful and constructive comments. HHB was partially supported by a Discovery Grant and an Accelerator Supplement of the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Canada Research Chair Program. CW was partially supported by NSF grant of China (11401372) and a grant from Shanghai Municipal Commission for Science and Technology (13ZR1455500). XW was partially supported by a Discovery Grant of NSERC. JX was partially supported by NSERC grants of HHB and XW.

References

[1] H.H. Bauschke and J.M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review* 38(3) (1996), 367–426.

[2] H.H. Bauschke, J. Chen, and X. Wang, A projection method for approximating fixed points of quasi nonexpansive mappings without the usual demiclosedness condition, *Journal of Nonlinear and Convex Analysis* 15 (2014), 129–135.

[3] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.

[4] H.H. Bauschke and P.L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert space, *Mathematics of Operations Research* 26 (2001), 248–264.

[5] H.H. Bauschke, F. Deutsch, H. Hundal, and S.-H. Park, Accelerating the convergence of the method of alternating projections, *Transactions of the AMS* 355(9) (2003), 3433–3461.

[6] J.M. Borwein and M. Fabián, On convex functions having points of Gateaux differentiability which are not points of Fréchet differentiability, *Canadian Journal of Mathematics* 45(6) (1993), 1121–1134.

[7] J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.

[8] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Springer, 2012.

[9] Y. Censor and A. Lent, Cyclic subgradient projections, *Mathematical Programming* 24 (1982), 233–235.

[10] Y. Censor and A. Segal, Sparse string-averaging and split common fixed points, in "Nonlinear Analysis and Optimization I", A. Leizarowitz, B.S. Mordukhovich, I. Shafrir, and A.J. Zaslavski (editors), *Contemporary Mathematics* 513 (2010), 125–142.

[11] Y. Censor and S.A. Zenios, *Parallel Optimization*, Oxford University Press, 1997.

[12] P.L. Combettes, The foundations of set theoretic estimation, *Proceedings of the IEEE* 81(2) (1993), 182–208.

[13] P.L. Combettes, Restauration ensembliste d'images par itérations parallèles extrapolées de sous-gradients, Proceedings of the 15th GRETSI Symposium, Juan-les-Pins, France, September 18–22 (1995), 447–450.

[14] P.L. Combettes, The convex feasibility problem in image recovery, in *Advances in Imaging and Electron Physics*, P. Hawkes (editor), vol. 95, Academic Press, 1996.

[15] P.L. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Transactions on Image Processing* 6 (1997), 493–506.

[16] P.L. Combettes and J. Luo, An adaptive level set method for nondifferentiable constrained image recovery, *IEEE Transactions on Image Processing* 11 (2002), 1295–1304.

[17] J.-L. Goffin, Subgradient optimization in nonsmooth optimization (including the Soviet revolution), in *Optimization Stories*, Documenta Mathematica book series vol. 6 (2012), 277–290.

[18] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, 1989.

[19] J. Milnor, *Dynamics in One Complex Variable*, third edition, Princeton University Press, 2006.

[20] N. Ogura and I. Yamada, A deep outer approximating half space of the level set of certain quadratic functions, *Journal of Nonlinear and Convex Analysis* 6 (2005), 187–201.

[21] B. Pauwels, *Subgradient projection operators*, `http://arxiv.org/abs/1403.7237v1` (2014), based on Mémoire de Master OJME, *Opérateurs de projection sous-différentielle* (in French, September 2012).

[22] J.-P. Penot, *Calculus Without Derivatives*, Springer, 2013.

[23] B.T. Polyak, Minimization of unsmooth functionals, *U.S.S.R. Computational Mathematics and Mathematical Physics* 9 (1969), 14–29. (The original version appeared in *Akademija Nauk SSSR. Žurnal Vyčislitel′ noĭ Matematiki i Matematičeskoĭ Fiziki* 9 (1969), 509–521.)

[24] B.T. Polyak, *Introduction to Optimization*, Optimization Software, 1987.

[25] B.T. Polyak, Random algorithms for solving convex inequalities, in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, D. Butnariu, Y. Censor, and S. Reich (editors), pages 409–422, Elsevier 2001.

[26] V.V. Prasolov, *Polynomials*, Springer, 2004.

[27] K. Slavakis and I. Yamada, The adaptive projected subgradient method constrained by families of quasi-nonexpansive mappings and its application to online learning, *SIAM Journal on Optimization* 23 (2013), 126–152.

[28] S. Theodoridis, K. Slavakis, and I. Yamada, Adaptive learning in a world of projections, *IEEE Signal Processing Magazine* 97 (2011), 97–123.

[29] I. Yamada and N. Ogura, Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions, *Numerical Functional Analysis and Optimization* 25 (2004), 593–617.

[30] I. Yamada and N. Ogura, Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings, *Numerical Functional Analysis and Optimization* 25 (2004), 619–655.

[31] I. Yamada, K. Slavakis, and K. Yamada, An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques, *IEEE Transactions on Signal Processing* 50 (2002), 1091–1101.

[32] M. Yamagishi and I. Yamada, A deep monotone approximation operator based on the best quadratic lower bound of convex functions, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* E91–A (2008), 1858–1866.