Set-Valued Var. Anal (2017) 25:245–263
DOI 10.1007/s11228-016-0376-5

On Proximal Subgradient Splitting Method for Minimizing the sum of two Nonsmooth Convex Functions

José Yunier Bello Cruz[1]

Received: 3 May 2015 / Accepted: 16 May 2016 / Published online: 30 May 2016

Abstract In this paper we present a variant of the proximal forward-backward splitting iteration for solving nonsmooth optimization problems in Hilbert spaces, when the objective function is the sum of two nondifferentiable convex functions. The proposed iteration, which will be called Proximal Subgradient Splitting Method, extends the classical subgradient iteration for important classes of problems, exploiting the additive structure of the objective function. The weak convergence of the generated sequence was established using different stepsizes and under suitable assumptions. Moreover, we analyze the complexity of the iterates.

Keywords Convex problems · Nonsmooth optimization problems · Proximal forward-backward splitting iteration · Subgradient method

Mathematics Subject Classification (2010) 65K05 · 90C25 · 90C30

1 Introduction

The purpose of this paper is to study the convergence properties of a variant of the proximal forward-backward splitting method for solving the following optimization problem:

$$\min \ f(x) + g(x) \ \text{ s.t. } x \in \mathcal{H}, \tag{1}$$

where $\mathcal{H}$ is a nontrivial real Hilbert space, and $f : \mathcal{H} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ and $g : \mathcal{H} \to \overline{\mathbb{R}}$ are two proper lower semicontinuous and convex functions. We are interested in the case where both functions f and g are nondifferentiable, and when the domain of f contains the domain of g. The solution set of this problem will be denoted by S_*, which is a closed

✉ José Yunier Bello Cruz
yunier.bello@gmail.com

[1] Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115, USA

and convex subset of the domain of g. Problem (1) has recently received much attention from the optimization community due to its broad applications to several different areas such as control, signal processing, system identification, machine learning and restoration of images; see, for instance, [18, 19, 24, 32] and the references therein.

A special case of problem (1) is the nonsmooth constrained optimization problem, taking $g = \delta_C$ where δ_C is the indicator function of a nonempty closed and convex set C in $\mathcal{H}$, defined by $\delta_C(y) := 0$, if $y \in C$ and $+\infty$, otherwise. Then, problem (1) reduces to the constrained minimization problem

$$\min \ f(x) \ \text{ s.t. } x \in C. \tag{2}$$

Another important case of problem (1), which has had much interest in signal denoising and data mining, is the following optimization problem with ℓ_1-regularization

$$\min \ f(x) + \lambda \|x\|_1 \ \text{ s.t. } x \in \mathcal{H}, \tag{3}$$

where $\lambda > 0$ and the norm $\|\cdot\|_1$ is used to induce the sparsity in the solutions. Moreover, problem (3) covers the important and the well-studied ℓ_1-regularized least square minimization problem, when $\mathcal{H} = \mathbb{R}^n$ and $f(x) = \|Ax - b\|_2^2$ where $A \in \mathbb{R}^{m\times n}$, $m << n$, and $b \in \mathbb{R}^m$, which is just a convex approximation of the very famous ℓ_0-minimization problem; see [12]. Recently, this problem became popular in signal processing and statistical inference; see, for instance, [23, 43].

We focus here on the so-called *proximal forward-backward splitting iteration* [32], which contains a forward gradient step of f (an explicit step) followed by a backward proximal step of g (an implicit step). The main idea of our approach consists of replacing, in the forward step of the proximal forward-backward splitting iteration, the gradient of f by a subgradient of f (note that here f is assumed nondifferentiable in general). In the particular case that g is the indicator function, the proposed iteration reduces to the classical projected subgradient iteration.

To describe and motivate our iteration, first we recall the definition of the so-called *proximal operator* as $\mathbf{prox}_g : \mathcal{H} \to \mathcal{H}$ associated to g a proper lower semicontinuous convex function, where $\mathbf{prox}_g(z)$, $z \in \mathcal{H}$ is the unique solution of the following strongly convex optimization problem

$$\min \ g(y) + \frac{1}{2}\|y - z\|^2 \ \text{ s.t. } y \in \mathcal{H}. \tag{4}$$

Note that the norm $\|\cdot\|$ is induced by this inner product of $\mathcal{H}$, *i.e.*, $\|x\| := \sqrt{\langle x, x\rangle}$ for all $x \in \mathcal{H}$. The proximal operator $\mathbf{prox}_g$ is well-defined and has many attractive properties, *e.g.*, it is continuous and firmly nonexpansive, *i.e.*, for all $x, y \in \mathcal{H}$, $\|\mathbf{prox}_g(x) - \mathbf{prox}_g(y)\|^2 \le \|x - y\|^2 - \|[x - \mathbf{prox}_g(x)] - [y - \mathbf{prox}_g(y)]\|^2$. This nice property can be used to construct algorithms to solve optimization problems [39]; for other properties and algebraic rules see [3, 18, 19]. If $g = \delta_C$ is the indicator function, the orthogonal projection onto C, $\mathbf{P}_C(x) := \{y \in C : \|x - y\| = \operatorname{dist}(x, C)\}$ is the same as $\mathbf{prox}_{\delta_C}(x)$ for all $x \in \mathcal{H}$ [2]. For an exhaustive discussion about the evaluation of the proximity operator of a wide variety of functions see Section 6 of [32]. Now, let us recall the definition of the subdifferential operator $\partial g : \mathcal{H} \rightrightarrows \mathcal{H}$ by $\partial g(x) := \{w \in \mathcal{H} : g(y) \ge g(x) + \langle w, y - x\rangle, \ \forall\, y \in \mathcal{H}\}$. We also present the relation of the proximal operator $\mathbf{prox}_{\alpha g}$ with the subdifferential operator ∂g, *i.e.*, $\mathbf{prox}_{\alpha g} = (\mathrm{Id} + \alpha\partial g)^{-1}$ and as a direct consequence of the first optimality condition of (4), we have the following useful inclusion:

$$\frac{z - \mathbf{prox}_{\alpha g}(z)}{\alpha} \in \partial g(\mathbf{prox}_{\alpha g}(z)), \tag{5}$$

for any $z \in \mathcal{H}$ and $\alpha > 0$. The iteration proposed in this paper, called *Proximal Subgradient Splitting Method*, is motivated by the well-known fact that $x \in S_*$ if and only if there exists $u \in \partial f(x)$ such that $x = \mathbf{prox}_{\alpha g}(x - \alpha u)$. Thus, the iteration generalizes the proximal forward-backward splitting iteration for the differentiable case, as a fixed point iteration of the above equation, which is defined as follows: starting at x^0 belonging to the domain of g, set

$$x^{k+1} = \mathbf{prox}_{\alpha_k g}(x^k - \alpha_k u^k), \tag{6}$$

where $u^k \in \partial f(x^k)$ and the stepsize α_k is positive for all $k \in \mathbb{N}$. Iteration (6) recovers the classical subgradient iteration [38], when $g = 0$, and the proximal point iteration [39], when $f = 0$. Moreover, it covers important situations in which f is nondifferentiable and it can also be seen as a *forward-backward Euler discretization* of the subgradient flow differential inclusion

$$\dot{x}(t) \in -\partial[f(x(t)) + g(x(t))],$$

with variable $x : \mathbb{R}_+ \to \mathcal{H}$; see [32]. Actually, if the derivative on the left side is replaced by the divided difference $(x^{k+1} - x^k)/\alpha_k$, then the discretization obtained is $(x^k - x^{k+1})/\alpha_k \in \partial f(x^k) + \partial g(x^{k+1})$, which is the proximal subgradient iteration (6).

The nondifferentiability of the function f has a direct impact on the computational effort and the importance of such problems, when f is nonsmooth, is underlined because they occur frequently in applications. Nondifferentiability arises, for instance, in the problem of minimizing the total variation of a signal over a convex set, in the problem of minimizing the sum of two set-distance functions, in problems involving maxima of convex functions, the Dantzing selector-type problems, the non-Gaussian image denoising problem and in Tykhonov regularization problems with L_1 norms and others; see, for instance, [4, 13, 17, 26]. The iteration of the proximal subgradient splitting method, proposed in (6), can be applied in these important instances, extending the classical subgradient iteration for more general problems as (3). In problem (1), f is usually assumed to be differentiable as in [35], which is not necessarily the case in this work. Moreover, the convergence of the iteration (6) to a solution of (1) has been established in the literature, when the gradient of f is globally Lipschitz continuous and the stepsizes α_k, $k \in \mathbb{N}$ have to be chosen very small, *i.e.*, for all k, α_k is less than some constant related with the Lipschitz constant of the gradient of f; see, for instance, [19]. Recently, when f is continuously differentiable but the Lipschitz constant is not available, the steplengths can be chosen using backtracking procedures; see [6, 10, 32, 35].

It is important to mention that the forward-backward iteration also finds applications in solving more general problems, like the variational inequality and inclusion problems; see, for instance, [9, 11, 14, 15, 42] and the references therein. On the other hand, the standard convergence analysis of this iteration, for solving these general problems, requires at least a co-coercivity assumption of the operator and the stepsizes to lie within a suitable interval; see, for instance, Theorem 25.8 of [3]. Note that co-coercive operators are monotone and Lipschitz continuous, but the converse does not hold in general; see [44]. Although, for gradients of lower semicontinuous, proper and convex functions, the co-coercivity is equivalent to the global Lipschitz continuity assumption. This nice and surprising fact, which is strongly used in the convergence analysis of the proximal forward-backward method for problem (1), when f is differentiable, is known as the *Baillon-Haddad Theorem*; see Corollary 18.16 of [3].

The main aim of this work is to remove the differentiability assumption from f in the forward-backward splitting method, extending the classical projected subgradient method and containing, as particular case, a new proximal subgradient iteration for more general problems.

This work is organized as follows. The next subsection provides our notations and assumptions, and some preliminaries results that will be used in the remainder of this paper. The proximal subgradient splitting method and its weak convergence are analyzed by choosing different stepsizes in Section 2. Finally, Section 3 gives some concluding remarks.

1.1 Assumptions and Preliminaries

In this section, we present our assumptions, classical definitions and some results needed for the convergence analysis of the proposed method.

We start by recalling some definitions and notation used in this paper, which are standard and follows from [3, 32]. Throughout this paper, we write $p := q$ to indicate that p is defined to be equal to q. We write $\mathbb{N}$ for the nonnegative integers $\{0, 1, 2, \dots\}$ and remind that the extended-real number system is $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. The closed ball centered at $x \in \mathcal{H}$ with radius $\gamma > 0$ will be denoted by $\mathbb{B}[x; \gamma]$, *i.e.*, $\mathbb{B}[x; \gamma] := \{y \in \mathcal{H} : \|y - x\| \leq \gamma\}$. The domain of any function $h : \mathcal{H} \to \overline{\mathbb{R}}$, denoted by $\mathbf{dom}(h)$, is defined as $\mathbf{dom}(h) := \{x \in \mathcal{H} : h(x) < +\infty\}$. The optimal value of problem (1) will be denoted by $s_* := \inf\{(f + g)(x) : x \in \mathcal{H}\}$, noting that when $S_* \neq \emptyset$, $s_* = \min\{(f + g)(x) : x \in \mathcal{H}\} = (f + g)(x_*)$ for any $x_* \in S_*$. Finally, $\ell_1(\mathbb{N})$ denotes the set of summable sequences in $[0, +\infty)$.

Throughout this paper we assume the following:

A1. ∂f is bounded on bounded sets on the domain of g, *i.e.*, there exists $\zeta = \zeta(V) > 0$ such that $\partial f(x) \subseteq \mathbb{B}[0; \zeta]$ for all $x \in V$, where V is any bounded and closed subset of $\mathbf{dom}(g)$.

A2. ∂g has bounded elements on the domain of g, *i.e.*, $\exists \rho \geq 0$ such that $\partial g(x) \cap \mathbb{B}[0; \rho] \neq \emptyset$ for all $x \in \mathbf{dom}(g)$.

In connection with Assumption **A1**, we recall that ∂f is locally bounded on its open domain. In finite dimension spaces, this result implies that **A1** always holds when $\mathbf{dom}(f)$ is open. A widely used sufficient condition for **A1** is the Lipschitz continuity of f on $\mathbf{dom}(g)$. Furthermore, the boundedness of the subgradients is crucial for the convergence analysis of many classical subgradient methods in Hilbert spaces and it has been widely considered in the literature; see, for instance, [1, 8, 9, 38].

Regarding Assumption **A2**, we emphasize that it holds trivially for important instances of problem (1), *e.g.*, problems (2) and (3) because $\partial \delta_C(x) = N_C(x)$ and $\partial \|x\|_1 = \{u \in \mathcal{H} : \|u\|_\infty \leq 1\,,\ \langle u, x\rangle = \|x\|_1\}$, respectively, or when $\mathbf{dom}(g)$ is a bounded set or also when $\mathcal{H}$ is a finite dimensional space. Note that Assumption **A2** allows instances where ∂g is an unbounded set as is the particular case when g is the indicator function. It is an existence condition, which is in general weaker than **A1**.

Let us end the section by recalling the well-known concepts so-called quasi-Fejér and Fejér convergence.

Definition 1.1 Let S be a nonempty subset of $\mathcal{H}$. A sequence $(x^k)_{k\in\mathbb{N}}$ in $\mathcal{H}$ is said to be quasi-Fejér convergent to S if and only if for all $x \in S$ there exists a sequence $(\epsilon_k)_{k\in\mathbb{N}}$ in $\ell_1(\mathbb{N})$ and $\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \epsilon_k$ for all $k \in \mathbb{N}$. When $(\epsilon_k)_{k\in\mathbb{N}}$ is a null sequence, we say that $(x^k)_{k\in\mathbb{N}}$ is Fejér convergent to S.

The definition originates in [22] and has been elaborated further in [16]. This definition, originated in [22], has been elaborated further in [16]. In the following we present two well-known fact for quasi-Fejér convergent sequences.

Fact 1.1 If the sequence $(x^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to S, then:

(a) The sequence $(x^k)_{k\in\mathbb{N}}$ is bounded.
(b) $(x^k)_{k\in\mathbb{N}}$ is weakly convergent iff all weak accumulation points of $(x^k)_{k\in\mathbb{N}}$ belong to S.

Proof Item **(a)** follows from Proposition 3.3(i) of [16], and Item **(b)** follows from Theorem 3.8 of [16]. □

2 The Proximal Subgradient Splitting Method

In this section we propose the proximal subgradient splitting method extending the classical subgradient iteration. We prove that the sequence of points generated by the proposed method converges weakly to a solution of (1) using different strategies for choosing of the stepsizes. Moreover, we show the complexity analysis for the generated sequence.

The method is formally stated as follows:

Proximal **S**ubgradient **S**plitting Method (**PSS Method**)

Initialization Step. Take $x^0 \in \mathbf{dom}(g)$.

Iterative Step. Set

$$x^{k+1} = \mathbf{prox}_{\alpha_k g}\left(x^k - \alpha_k u^k\right), \tag{6}$$

where $u^k \in \partial f(x^k)$.

Stop Criteria. If $x^{k+1} = x^k$ then stop.

If **PSS Method** stops at step k, then $x^k = \mathbf{prox}_{\alpha_k g}\left(x^k - \alpha_k u^k\right)$ with $u^k \in \partial f(x^k)$, implying that x^k is solution of problem (1). Then, from now on, we assume that **PSS Method** generates an infinite sequence $(x^k)_{k\in\mathbb{N}}$. Moreover, it follows directly from (6) that the sequence $(x^k)_{k\in\mathbb{N}}$ belongs to $\mathbf{dom}(g)$.

Before the formal analysis of the convergence properties of **PSS Method**, we discuss below about the necessity of taking a (forward) subgradient step of f instead of another (backward) proximal step.

Remark 2.1 To evaluate the proximal operator of f is necessary to solve a strongly convex minimization problem as (4). Thus, in the context of problem (1), we assume that it is hard to evaluate the proximal operator of f, leaving out the possibility to use the standard and very powerful iteration so-called *Douglas-Rachford splitting method* presented in [17]. Such situations appear mainly when f has a complicated algebraic expression and therefore it may impossibility to solve, explicitly or efficiently, subproblem (4). Indeed, very often in the applications, the formula for the proximity operator is not available in closed form and ad hoc algorithms or approximation procedures have be used to compute $\mathbf{prox}_{\alpha f}$. This happens for instance when applying proximal methods to image deblurring with total variation [5], or to structured sparsity regularization problems in machine learning and inverse problems [31]. A classical problem of the form of (1), when the subgradient of f is easily available and $\mathbf{prox}_f$ does not has explicitly formula is the dual formulation of the following constrained convex problem:

$$\min h_0(y) \quad \text{subject that} \quad h_i(y) \leq 0 \;\; (i = 1, \ldots, n), \tag{7}$$

where $h_i : \mathbb{R}^m \to \mathbb{R}$ $(i = 0, \ldots, n)$ are convex. It has as dual problem

$$\min_{x \in \mathbb{R}^n} f(x) + \delta_{\mathbb{R}^n_+}(x)$$

with $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x) = -\inf_{y \in \mathbb{R}^m}\{h_0(y) + \sum_{i=1}^n x_i h_i(y)\}$. It is well-known that

$$\partial f(x) = \text{conv}\left\{ h(y_x) : f(x) = h_0(y_x) + \sum_{i=1}^n x_i h_i(y_x) \right\},$$

and conv$\{S\}$ denotes the convex hull of a set S. However, compute $\mathbf{prox}_f$ does not look an easy problem. This argument is used widely in the literature to motivated the projected subgradient method, which can be easily modified for recovering problems as (1), when g is not necessary the indicator function. Indeed, consider problem (7) when $n = m$ with an additional and simple restriction g_0, that is: $y \in \mathbb{R}^n$

$$\min h_0(y) \quad \text{subject that} \quad h_i(y) \le 0 \;\; (i = 1, \ldots, n), \;\; g_0(y) \le 0, \tag{8}$$

which can be rewritten now as $\min_{x \in \mathbb{R}^n} f(x) + \delta_{\mathbb{R}^n_+}(x) + \lambda g_0(x), \;\; \lambda > 0$, where

$$f(x) = -\inf_{y \in \mathbb{R}^n}\{h_0(y) + \sum_{i=1}^n x_i h_i(y)\}.$$

This last problem is a particular case of (1), by taking $g = \delta_{\mathbb{R}^n_+} + \lambda g_0$. Note that if $\mathbf{dom}(g_0) \subseteq \mathbb{R}^n_+$ then $g = \lambda g_0$.

Thus, **PSS Method** uses the proximal operator of g and the explicit subgradient iteration of f (*i.e.*, the proximal operator of f is never evaluated), which is, in general, much easier to implement than the proximal operator of $f + g$ or f, as happens in the standard proximal point iteration or the Douglas-Rachford algorithm, respectively for solving nonsmooth problems, as (1); see, for instance, [17]. Furthermore, note that in our case the subgradient iteration for the sum $f + g$ is not possible, because the domains of f and g are not the whole space.

In the following we prove a crucial property of the iterates generated by **PSS Method**.

Lemma 2.1 *Let $(x^k)_{k\in\mathbb{N}}$ and $(u^k)_{k\in\mathbb{N}}$ be the sequences generated by **PSS Method**. Then, for all $k \in \mathbb{N}$ and $x \in \mathbf{dom}(g)$,*

$$\|x^{k+1} - x\|^2 \le \|x^k - x\|^2 + 2\alpha_k \Big[(f+g)(x) - (f+g)(x^k)\Big] + \alpha_k^2 \|u^k + w^k\|^2,$$

where $w^k \in \partial g(x^k)$ is arbitrary.

Proof Take any $x \in \mathbf{dom}(g)$. Note that (5) and (6) imply that $\bar{w}^{k+1} := \dfrac{x^k - x^{k+1}}{\alpha_k} - u^k$, with $u^k \in \partial f(x^k)$ as defined by **PSS Method**, belongs to $\partial g(x^{k+1})$. Then,

$$\begin{aligned}
&\alpha_k^2 \|u^k + \bar{w}^{k+1}\|^2 + \|x^k - x\|^2 - \|x^{k+1} - x\| = \|x^{k+1} - x^k\|^2 + \|x^k - x\|^2 - \|x^{k+1} - x\|^2 \\
&= 2\langle x^k - x^{k+1}, x^k - x\rangle = 2\alpha_k \langle u^k, x^k - x\rangle + 2\langle x^k - x^{k+1} - \alpha_k u^k, x^k - x\rangle \\
&= 2\alpha_k \langle u^k, x^k - x\rangle + 2\alpha_k \left\langle \frac{x^k - x^{k+1}}{\alpha_k} - u^k, x^{k+1} - x \right\rangle + 2\alpha_k \left\langle \frac{x^k - x^{k+1}}{\alpha_k} - u^k, x^k - x^{k+1} \right\rangle \\
&= 2\alpha_k \langle u^k, x^k - x\rangle + 2\alpha_k \left\langle \bar{w}^{k+1}, x^{k+1} - x \right\rangle + 2\alpha_k \langle u^k, x^{k+1} - x^k\rangle + 2\|x^k - x^{k+1}\|^2.
\end{aligned}$$

Now using again that $\dfrac{x^k - x^{k+1}}{\alpha_k} - u^k = \bar{w}^{k+1} \in \partial g(x^{k+1})$ and the convexity of g and f, the above equality leads us

$$2\langle x^k - x^{k+1}, x^k - x\rangle \geq 2\alpha_k \left[f(x^k) - f(x) + g(x^{k+1}) - g(x) + \langle u^k, x^{k+1} - x^k\rangle\right] + 2\|x^k - x^{k+1}\|^2$$
$$= 2\alpha_k \left[(f+g)(x^k) - (f+g)(x) + g(x^{k+1}) - g(x^k) + \langle u^k, x^{k+1} - x^k\rangle\right] + 2\alpha_k^2\|u^k + \bar{w}^{k+1}\|^2$$
$$\geq 2\alpha_k \left[(f+g)(x^k) - (f+g)(x) + \langle w^k + u^k, x^{k+1} - x^k\rangle\right] + 2\alpha_k^2\|u^k + \bar{w}^{k+1}\|^2,$$

for any $w^k \in \partial g(x^k)$. We thus have shown that

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + 2\alpha_k\left[(f+g)(x) - (f+g)(x^k)\right]$$
$$+2\alpha_k^2\langle u^k + w^k, u^k + \bar{w}^{k+1}\rangle - \alpha_k^2\|u^k + \bar{w}^{k+1}\|^2$$
$$= \|x^k - x\|^2 + 2\alpha_k\left[(f+g)(x) - (f+g)(x^k)\right] + \alpha_k^2\|u^k + w^k\|^2 - \alpha_k^2\|w^k - \bar{w}^{k+1}\|^2.$$

Note that $w^k \in \partial g(x^k)$ is arbitrary and the result follows. □

Since subgradient methods are not descent methods, as the proposed method here, it is common to keep track of the best point found so far, *i.e.*, the one with minimum function value among the iterates. At each step, we set it recursively as $(f+g)^0_{\text{best}} := (f+g)(x^0)$ and

$$(f+g)^k_{\text{best}} := \min\left\{(f+g)^{k-1}_{\text{best}}, (f+g)(x^k)\right\}, \tag{9}$$

for all k. Since $\left((f+g)^k_{\text{best}}\right)_{k\in\mathbb{N}}$ is a decreasing sequence, it has a limit (which can be $-\infty$). When the function f is differentiable and its gradient Lipschitz continuous, it is possible to prove the complexity of the iterates generated by **PSS Method**; see [35]. In our instance (f is not necessarily differentiable) we expect, of course, slower convergence.

Next we present a convergence rate result for the sequence of the best functional values $\left((f+g)^k_{\text{best}}\right)_{k\in\mathbb{N}}$ to $\min\{(f+g)(x) : x \in \mathcal{H}\}$.

Lemma 2.2 *Let $\left((f+g)^k_{best}\right)_{k\in\mathbb{N}}$ be the sequence defined by (9). If $S_* \neq \emptyset$ then, for all $k \in \mathbb{N}$,*

$$(f+g)^k_{best} - \min_{x\in\mathcal{H}}(f+g)(x) \leq \frac{[\,dist(x^0, S_*)]^2 + C_k\sum_{i=0}^k \alpha_i^2}{2\sum_{i=0}^k \alpha_i},$$

where $C_k := \max\left\{\|u^i + w^i\|^2 : 0 \leq i \leq k\right\}$ with $w^i \in \partial g(x^i)$ $(i = 0, \ldots, k)$ are arbitrary.

Proof Define $x_* := \mathbf{P}_{S_*}(x^0)$. Note that x_* exists because S_* is a nonempty closed and convex set of $\mathcal{H}$. By applying Lemma 2.1, $k+1$ times, for $i \in \{0, 1, \ldots, k\}$ at $x_* \in S_*$, we get

$$\|x^{k+1} - x_*\|^2 \leq \|x^k - x_*\|^2 + 2\alpha_k\left[(f+g)(x_*) - (f+g)(x^k)\right] + \alpha_k^2\|u^k + w^k\|^2$$
$$\leq \|x^0 - x_*\|^2 + 2\sum_{i=0}^k \alpha_i\left[(f+g)(x_*) - (f+g)(x^i)\right] + \sum_{i=0}^k \alpha_i^2\|u^i + w^i\|^2$$
$$\leq [\text{dist}(x^0, S_*)]^2 + 2\left[\min_{x\in\mathcal{H}}(f+g)(x) - (f+g)^k_{\text{best}}\right]\sum_{i=0}^k \alpha_i + C_k\sum_{i=0}^k \alpha_i^2, \tag{10}$$

where $(f+g)^k_{\text{best}}$ is defined by (9) and the result follows after simple algebra. □

Next we establish the rate of convergence of the rate of the objective values at the ergodic sequence $(\bar{x}^k)_{k\in\mathbb{N}}$ of $(x^k)_{k\in\mathbb{N}}$, which is defined recursively as $\bar{x}^0 = x^0$ and given $\sigma_0 = \alpha_0$ and $\sigma_k = \sigma_{k-1} + \alpha_k$, we define

$$\bar{x}^k = \left(1 - \frac{\alpha_k}{\sigma_k}\right)\bar{x}^{k-1} + \frac{\alpha_k}{\sigma_k}x^k.$$

After easy induction, we have $\sigma_k = \sum_{i=0}^k \alpha_i$ and

$$\bar{x}^k = \frac{1}{\sigma_k}\sum_{i=0}^k \alpha_i\, x^i, \tag{11}$$

for all $k \in \mathbb{N}$.

The following result is similar to Lemma 2.2, considering now the ergodic sequence defined by (11).

Lemma 2.3 *Let $(\bar{x}^k)_{k\in\mathbb{N}}$ be the ergodic sequence defined by (11). If $S_* \neq \emptyset$, then*

$$(f+g)(\bar{x}^k) - \min_{x\in\mathcal{H}}(f+g)(x) \leq \frac{[\,dist(x^0, S_*)]^2 + C_k\sum_{i=0}^k \alpha_i^2}{2\sum_{i=0}^k \alpha_i},$$

where $C_k = \max\left\{\|u^i + w^i\|^2 : 0 \leq i \leq k\right\}$ with $w^i \in \partial g(x^i)$ $(i = 0, \ldots, k)$ are arbitrary.

Proof Proceeding as in the proof of Lemma 2.2 until inequality (10) and after dividing by $2\sum_{i=0}^k \alpha_i$, we get

$$\begin{aligned}\sum_{i=0}^k \frac{\alpha_i}{\sigma_k}\left[(f+g)(x^i) - \min_{x\in\mathcal{H}}(f+g)(x)\right] &\leq \frac{1}{2\sigma_k}\left([\,\mathrm{dist}(x^0, S_*)]^2 - \|x^{k+1} - x_*\|^2\right) + \frac{C_k}{2\sigma_k}\sum_{i=0}^k \alpha_i^2 \\ &\leq \frac{1}{2\sigma_k}\left([\,\mathrm{dist}(x^0, S_*)]^2 + C_k\sum_{i=0}^k \alpha_i^2\right),\end{aligned} \tag{12}$$

where $\sigma_k := \sum_{i=0}^k \alpha_i$. Using the convexity of $f+g$ after note that $\frac{\alpha_i}{\sigma_k} \in [0, 1]$ for all $i \in \{0, 1, \ldots, k\}$ and $\sum_{i=0}^k \frac{\alpha_i}{\sigma_k} = 1$ and (11) in the above inequality (12), the result follows. □

Next we focus on constant stepsizes, which is motivated by the fact that we are interested in quantifying the progress of the proposed method to find an approximate solution.

Corollary 2.4 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by **PSS Method** with the stepsizes α_k constant equal to α, $\left((f+g)^k_{best}\right)_{k\in\mathbb{N}}$ be the sequence defined by (9) and $(\bar{x}^k)_{k\in\mathbb{N}}$ be the ergodic sequence as (11). Then, the iteration attains the optimal rate at $\alpha = \frac{dist(x^0,S_*)}{\sqrt{C_k}} \cdot \frac{1}{\sqrt{k+1}}$, i.e., for all $k \in \mathbb{N}$,*

$$(f+g)^k_{best} - \min_{x\in\mathcal{H}}(f+g)(x) \leq \frac{[dist(x^0, S_*)]^2 + \alpha^2(k+1)C_k}{2(k+1)\alpha} \leq \frac{dist(x^0, S_*)\cdot\sqrt{C_k}}{\sqrt{k+1}}$$

and

$$(f+g)(\bar{x}^k) - \min_{x\in\mathcal{H}}(f+g)(x) \leq \frac{[dist(x^0, S_*)]^2 + \alpha^2(k+1)C_k}{2(k+1)\alpha} \leq \frac{dist(x^0, S_*)\cdot\sqrt{C_k}}{\sqrt{k+1}},$$

where $C_k = \max\left\{\|u^i + w^i\|^2 : 0 \leq i \leq k\right\}$ with $w^i \in \partial g(x^i)$ $(i = 0, \ldots, k)$ are arbitrary.

Proof If we consider constant stepsizes, *i.e.*, $\alpha_k = \alpha$ for all $k \in \mathbb{N}$, then the optimal rate is obtained when $\alpha = \frac{\mathrm{dist}(x^0, S_*)}{\sqrt{C_k}} \cdot \frac{1}{\sqrt{k+1}}$ from minimizing the right part of Lemmas 2.2 and 2.3. □

Note that under Assumption **A2**, $C_k \le (\max_{1\le i\le k} \|u^i\| + \rho)^2$. Hence when ∂f is bounded on the **dom**(g) (which occurs when **dom**(g) is bounded over our assumptions), Assumption **A1** implies that $C_k \le (\zeta + \rho)^2$ for all $k \in \mathbb{N}$. In this case, our analysis showed that the expected error of the iterates generated by **PSS Method** with constant stepsizes after k iterations is $\mathcal{O}\left((k+1)^{-1/2}\right)$. Hence, we can search an ε-solution of problem (1) with $\mathcal{O}\left(\varepsilon^{-2}\right)$ iterations. Of course, this is worse than the rate $\mathcal{O}(k^{-1})$ and $\mathcal{O}\left(\varepsilon^{-1}\right)$ iterations of the proximal forward-backward iteration for the differentiable and convex f with Lipschitz continuous gradient; see, for instance, [35]. However, as was showed in Section 3.2.1, Theorem 3.2.1 of [37], the worst expected error after k iterations of the classical subgradient iteration is attainable equal to $\mathcal{O}\left((k+1)^{-1/2}\right)$ for general nonsmooth problems.

2.1 Exogenous Stepsizes

In this subsection we analyze the convergence of **PSS Method** using exogenous stepsizes, *i.e.*, the positive exogenous sequence of stepsizes $(\alpha_k)_{k\in\mathbb{N}}$ satisfies that $\alpha_k = \dfrac{\beta_k}{\eta_k}$ where $\eta_k := \max\{1, \|u^k\|\}$ for all k, and

$$\sum_{k=0}^{\infty} \beta_k^2 < +\infty \qquad \text{and} \qquad \sum_{k=0}^{\infty} \beta_k = +\infty. \tag{13}$$

We begin with a useful consequence of Lemma 2.1.

Corollary 2.5 *Let $x \in$ **dom**(g). Then, for all $k \in \mathbb{N}$,*

$$\|x^{k+1} - x\|^2 \le \|x^k - x\|^2 + 2\frac{\beta_k}{\eta_k}\left[(f+g)(x) - (f+g)(x^k)\right] + \left(1 + 2\rho + \rho^2\right)\beta_k^2,$$

*where $\rho \ge 0$ is as defined in Assumption **A2**.*

Proof The result follows by noting that $\eta_k \ge \|u^k\|$, $\eta_k \ge 1$ for all $k \in \mathbb{N}$ and letting $w^k \in \partial g(x^k)$ such that $\|w^k\| \le \rho$ for all $k \in \mathbb{N}$ in view of Assumption **A2**. Then,

$$\frac{\|u^k + w^k\|^2}{\eta_k^2} \le \frac{\|u^k\|^2}{\eta_k^2} + 2\frac{\|u^k\|\|w^k\|}{\eta_k^2} + \frac{\|w^k\|^2}{\eta_k^2} \le 1 + 2\rho + \rho^2.$$

Now, Lemma 2.1 implies the desired result. □

Now we define the auxiliary set

$$S_{\mathrm{lev}}(x^0) := \left\{x \in \mathbf{dom}(g) : (f+g)(x) \le (f+g)(x^k),\ \forall k \in \mathbb{N}\right\}. \tag{14}$$

When the solution set of problem (1) is nonempty, $S_{\mathrm{lev}}(x^0) \neq \emptyset$ because $S_* \subseteq S_{\mathrm{lev}}(x^0)$. Next, we prove the two main results of this subsection.

Theorem 2.6 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by **PSS Method** with exogenous stepsizes. If there exists $\bar{x} \in S_{lev}(x^0)$, then:*

(a) *The sequence* $(x^k)_{k\in\mathbb{N}}$ *is quasi-Fejér convergent to*

$$\mathcal{L}_{f+g}(\bar{x}) := \{x \in \mathbf{dom}(g) : (f+g)(x) \leq (f+g)(\bar{x})\}.$$

(b) $\lim_{k\to\infty}(f+g)(x^k) = (f+g)(\bar{x})$.

(c) *The sequence* $(x^k)_{k\in\mathbb{N}}$ *is weakly convergent to some* $\tilde{x} \in \mathcal{L}_{f+g}(\bar{x})$.

Proof By assumption there exists $\bar{x} \in S_{\text{lev}}(x^0)$, *i.e.*, $(f+g)(\bar{x}) \leq (f+g)(x^k)$, for all $k \in \mathbb{N}$.

(a) To show that $(x^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to $\mathcal{L}_{f+g}(\bar{x})$ (which is nonempty because $\bar{x} \in \mathcal{L}_{f+g}(\bar{x})$), we use Corollary 2.5, for any $x \in \mathcal{L}_{f+g}(\bar{x}) \subseteq \mathbf{dom}(g)$, establishing that $\|x^{k+1}-x\|^2 \leq \|x^k - x\|^2 + (1+2\rho+\rho^2)\beta_k^2$, for all $k \in \mathbb{N}$. Thus, $(x^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to $\mathcal{L}_{f+g}(\bar{x})$.

(b) The sequence $(x^k)_{k\in\mathbb{N}}$ is bounded from Fact 1.1a, and hence it has accumulation points in the sense of the weak topology. To prove that

$$\lim_{k\to\infty}(f+g)(x^k) = (f+g)(\bar{x}), \tag{15}$$

we use Corollary 2.5, with $x = \bar{x} \in \mathcal{L}_{f+g}(\bar{x}) \subseteq \mathbf{dom}(g)$, to get

$$\beta_k\left[(f+g)(x^k) - (f+g)(\bar{x})\right] \leq \tfrac{1}{2}(\|x^k-\bar{x}\|^2 - \|x^{k+1}-\bar{x}\|^2) + \tfrac{1}{2}(1+2\rho+\rho^2)\beta_k^2.$$

Summing, from $k=0$ to m, the above inequality, we have

$$\sum_{k=0}^{m}\beta_k\left[(f+g)(x^k) - (f+g)(\bar{x})\right] \leq \tfrac{1}{2}(\|x^0-\bar{x}\|^2 - \|x^{m+1}-\bar{x}\|^2) + \tfrac{1}{2}(1+2\rho+\rho^2)\sum_{k=0}^{m}\beta_k^2,$$

and taking limit, when m goes to ∞,

$$\sum_{k=0}^{\infty}\beta_k\left[(f+g)(x^k) - (f+g)(\bar{x})\right] < +\infty. \tag{16}$$

Then, (16) together with (13) implies that there exists a subsequence $\big((f+g)(x^{i_k})\big)_{k\in\mathbb{N}}$ of $\big((f+g)(x^k)\big)_{k\in\mathbb{N}}$ such that

$$\liminf_{k\to\infty}\left[(f+g)(x^{i_k}) - (f+g)(\bar{x})\right] = 0. \tag{17}$$

Indeed, if (17) does not hold, then there exist $\sigma > 0$ and $k \geq \tilde{k}$, such that $(f+g)(x^k) - (f+g)(\bar{x}) \geq \sigma$ and using (16), we get

$$+\infty > \sum_{k=\tilde{k}}^{\infty}\beta_k\left[(f+g)(x^k) - (f+g)(\bar{x})\right] \geq \sigma\sum_{k=\tilde{k}}^{\infty}\beta_k,$$

in contradiction with (13). Next, define $\varphi_k := (f+g)(x^k) - (f+g)(\bar{x})$, which is positive for all k because $\bar{x} \in S_{\text{lev}}(x^0)$. Then, for any $u^k \in \partial f(x^k)$ and $w^k \in \partial g(x^k)$, we get

$$\begin{aligned}\varphi_k - \varphi_{k+1} &= (f+g)(x^k) - (f+g)(x^{k+1}) \leq \langle u^k + w^k, x^k - x^{k+1}\rangle \\ &\leq \|u^k + w^k\|\|x^k - x^{k+1}\| \leq (\zeta+\rho)\|x^k - x^{k+1}\|,\end{aligned} \tag{18}$$

where $\zeta > 0$ such that $\|u^k\| \leq \zeta$, for all $k \in \mathbb{N}$ (ζ exists in virtue of the boundedness of $(x^k)_{k\in\mathbb{N}}$ and Assumption **A1**) and $\|w^k\| \leq \rho$, for all $k \in \mathbb{N}$ (ρ exists because $w^k \in \partial g(x^k)$ are arbitrary and the use of Assumption **A2**). Using Corollary 2.5, with

$x = x^k$, we have $\|x^k - x^{k+1}\| \le \sqrt{1 + 2\rho + \rho^2} \cdot \beta_k$, which together with (18) implies that

$$\varphi_k - \varphi_{k+1} \le \sqrt{1 + 2\rho + \rho^2} \cdot (\zeta + \rho)\beta_k := \bar{\rho}\beta_k \tag{19}$$

for all $k \in \mathbb{N}$. From (17), there exists a subsequence $(\varphi_{i_k})_{k\in\mathbb{N}}$ of $(\varphi_k)_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} \varphi_{i_k} = 0$. If the claim given in (15) does not hold, then there exists some $\delta > 0$ and a subsequence $(\varphi_{\ell_k})_{k\in\mathbb{N}}$ of $(\varphi_k)_{k\in\mathbb{N}}$, such that $\varphi_{\ell_k} \ge \delta$ for all $k \in \mathbb{N}$. Thus, we can construct a third subsequence $(\varphi_{j_k})_{k\in\mathbb{N}}$ of $(\varphi_k)_{k\in\mathbb{N}}$, where the indices j_k are chosen in the following way:

$$\begin{aligned} j_0 &:= \min\{m \ge 0 \mid \varphi_m \ge \delta\}, \\ j_{2k+1} &:= \min\{m \ge j_{2k} \mid \varphi_m \le \delta/2\}, \\ j_{2k+2} &:= \min\{m \ge j_{2k+1} \mid \varphi_m \ge \delta\}, \end{aligned}$$

for each k. The existence of the subsequences $(\varphi_{i_k})_{k\in\mathbb{N}}$, $(\varphi_{\ell_k})_{k\in\mathbb{N}}$ of $(\varphi_k)_{k\in\mathbb{N}}$, guarantees that the subsequence $(\varphi_{j_k})_{k\in\mathbb{N}}$ of $(\varphi_k)_{k\in\mathbb{N}}$ is well-defined for all $k \ge 0$. It follows from the definition of j_k that

$$\begin{aligned} \varphi_m &\ge \delta \quad \text{for} \quad j_{2k} \le m \le j_{2k+1} - 1 \\ \varphi_m &\le \frac{\delta}{2} \quad \text{for} \quad j_{2k+1} \le m \le j_{2k+2} - 1 \end{aligned} \tag{20}$$

for all k, and hence

$$\varphi_{j_{2k}} - \varphi_{j_{2k+1}} \ge \frac{\delta}{2}, \tag{21}$$

for all $k \in \mathbb{N}$. In view of (16) and remind that $\varphi_k = (f + g)(x^k) - (f + g)(\bar{x}) \ge 0$ for all $k \in \mathbb{N}$,

$$\begin{aligned} +\infty &> \sum_{k=0}^{\infty} \beta_k\varphi_k \ge \sum_{k=0}^{\infty} \sum_{m=j_{2k}}^{j_{2k+1}-1} \beta_m\varphi_m \ge \frac{\delta}{2} \sum_{k=0}^{\infty} \sum_{m=j_{2k}}^{j_{2k+1}-1} \beta_m \\ &= \frac{\delta}{2\bar{\rho}} \sum_{k=0}^{\infty} \sum_{m=j_{2k}}^{j_{2k+1}-1} \bar{\rho}\beta_m \ge \frac{\delta}{2\bar{\rho}} \sum_{k=0}^{\infty} \sum_{m=j_{2k}}^{j_{2k+1}-1} (\varphi_m - \varphi_{m+1}) = \frac{\delta}{2\bar{\rho}} \sum_{k=0}^{\infty} (\varphi_{j_{2k}} - \varphi_{j_{2k+1}}) \\ &\ge \frac{\delta}{2\bar{\rho}} \sum_{k=0}^{\infty} \frac{\delta}{2} = +\infty, \end{aligned}$$

where we have used (20) in the second inequality and (19) in the third inequality and (21) in the last one. Thus, $\lim_{k\to\infty} (f + g)(x^k) = (f + g)(\bar{x})$, establishing **(b)**.

(c) Let $\tilde{x}$ be a weak accumulation point of $(x^k)_{k\in\mathbb{N}}$, and note that $\tilde{x}$ exists by Item **(a)** and Fact 1.1a. From now on, we use $(x^{i_k})_{k\in\mathbb{N}}$ to denote any subsequence of $(x^k)_{k\in\mathbb{N}}$ that converges weakly to $\tilde{x}$. Since $f + g$ is weakly lower semicontinuous, using (15), we get

$$(f + g)(\tilde{x}) \le \liminf_{k\to\infty} (f + g)(x^{i_k}) = \lim_{k\to\infty} (f + g)(x^k) = (f + g)(\bar{x}),$$

implying that $(f + g)(\tilde{x}) \le (f + g)(\bar{x})$ and thus $\tilde{x} \in \mathcal{L}_{f+g}(\bar{x})$. As consequence, all weak accumulation points of $(x^k)_{k\in\mathbb{N}}$ belong to $\mathcal{L}_{f+g}(\bar{x})$ and since $(x^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to $\mathcal{L}_{f+g}(\bar{x})$, we get that $(x^k)_{k\in\mathbb{N}}$ converges weakly to $\tilde{x} \in \mathcal{L}_{f+g}(\bar{x})$ from Fact 1.1b. □

Theorem 2.7 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by **PSS Method** with exogenous stepsizes. Then,*

(a) $\liminf_{k\to\infty}(f+g)(x^k) = \inf_{x\in\mathcal{H}}(f+g)(x) = s_*$ *(possibly* $s_* = -\infty$*).*
(b) *If* $S_* \neq \emptyset$*, then* $\lim_{k\to\infty}(f+g)(x^k) = \min_{x\in\mathcal{H}}(f+g)(x)$ *and* $(x^k)_{k\in\mathbb{N}}$ *converges weakly to some* $\bar{x} \in S_*$*.*
(c) *If* $S_* = \emptyset$*, then* $(x^k)_{k\in\mathbb{N}}$ *is unbounded.*

Proof **(a)** Since $(x^k)_{k\in\mathbb{N}} \subset \mathbf{dom}(g)$, we get $s_* \le \liminf_{k\to\infty}(f+g)(x^k)$. Suppose that $s_* < \liminf_{k\to\infty}(f+g)(x^k)$. Hence, there exists $\hat{x}$ such that

$$(f+g)(\hat{x}) < \liminf_{k\to\infty}(f+g)(x^k). \tag{22}$$

It follows from (22) that there exists $\bar{k} \in \mathbb{N}$ such that $(f+g)(\hat{x}) \le (f+g)(x^k)$ for all $k \ge \bar{k}$. Since $\bar{k}$ is finite we can assume without loss of generality that $(f+g)(\hat{x}) \le (f+g)(x^k)$ for all $k \in \mathbb{N}$. Using the definition of $S_{\text{lev}}(x^0)$, given in (14), we have that $\hat{x} \in S_{\text{lev}}(x^0)$. By Theorem 2.6b $\lim_{k\to\infty}(f+g)(x^k) = (f+g)(\hat{x})$, in contradiction with (22).

(b) Since $S_* \neq \emptyset$, take $x_* \in S_*$ and note that this implies $\mathcal{L}_{f+g}(x_*) = S_*$. Since $(x^k)_{k\in\mathbb{N}} \subset \mathbf{dom}(g)$, we get $(f+g)(x_*) \le (f+g)(x^k)$ for all $k \in \mathbb{N}$ implying that $x_* \in S_{\text{lev}}(x^0)$. By applying items (b) and (c) of Theorem 2.6, at $\bar{x} = x_*$, we get that $\lim_{k\to\infty}(f+g)(x^k) = (f+g)(x_*)$ and $(x^k)_{k\in\mathbb{N}}$ converges weakly to some $\tilde{x} \in S_*$, respectively.

(c) Assume that S_* is empty but the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded. Let $(x^{\ell_k})_{k\in\mathbb{N}}$ be a subsequence of $(x^k)_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty}(f+g)(x^{\ell_k}) = \liminf_{k\to\infty}(f+g)(x^k)$. Since $(x^{\ell_k})_{k\in\mathbb{N}}$ is bounded, without loss of generality (*i.e.*, refining $(x^{\ell_k})_{k\in\mathbb{N}}$ if necessary), we may assume that $(x^{\ell_k})_{k\in\mathbb{N}}$ converges weakly to some $\bar{x} \in \mathbf{dom}(g)$. By the weak lower semicontinuity of $f+g$ on $\mathbf{dom}(g)$,

$$(f+g)(\bar{x}) \le \liminf_{k\to\infty}(f+g)(x^{\ell_k}) = \lim_{k\to\infty}(f+g)(x^{\ell_k}) = \liminf_{k\to\infty}(f+g)(x^k) = s_*, \tag{23}$$

using Item **(a)** in the last equality. By (23), $\bar{x} \in S_*$, in contradiction with the hypothesis and the result follows. □

For exogenous stepsizes, Theorem 2.7a guarantees the convergence of $\big((f+g)(x^k)\big)_{k\in\mathbb{N}}$ to the optimal value of problem (1), *i.e.*, $\liminf_{k\to\infty}(f+g)(x^k) = s_*$, implying the convergence of $\big((f+g)^k_{\text{best}}\big)_{k\in\mathbb{N}}$, defined in (9), to s_*. It is important to mention that in the proof of the above two crucial results, we have used a similar idea recently presented in [7] for a different instance.

In the following we present a direct consequence of Lemmas 2.2 and 2.3, when the stepsizes satisfy (13).

Corollary 2.8 *Let* $(\bar{x}^k)_{k\in\mathbb{N}}$ *be the ergodic sequence defined by* (11) *and* $(\beta_k)_{k\in\mathbb{N}}$ *as* (13)*. If* $S_* \neq \emptyset$*, then, for all* $k \in \mathbb{N}$*,*

$$(f+g)^k_{best} - \min_{x\in\mathcal{H}}(f+g)(x) \le \zeta\, \frac{[\,dist(x^0, S_*)]^2 + (1+2\rho+\rho^2)\sum_{i=0}^k \beta_i^2}{2\sum_{i=0}^k \beta_i}$$

and

$$(f+g)(\bar{x}^k) - \min_{x\in\mathcal{H}}(f+g)(x) \le \zeta \frac{[\,dist(x^0, S_*)]^2 + (1+2\rho+\rho^2)\sum_{i=0}^{k}\beta_i^2}{2\sum_{i=0}^{k}\beta_i},$$

where $\zeta > 0$ *and* $\rho \ge 0$ *are as in Assumptions* ***A1*** *and* ***A2****, respectively.*

The above corollary shows that if we assume existence of solutions, the expected error of the iterates generated by **PSS Method** with the exogenous stepsizes (13) after k iterations is $\mathcal{O}\left((\sum_{i=0}^{k}\beta_i)^{-1}\right)$. Since $(\beta_k)_{k\in\mathbb{N}}$ satisfies (13) the best performance of the iteration (in term of functional values) is archived for example taking $\beta_k \cong 1/k^r$ with r bigger than $1/2$, but near of this value, for all k.

2.2 Polyak Stepsizes

In this subsection we analyze the convergence of **PSS Method** using Polyak stepsizes. Having chose any $w^k \in \partial g(x^k)$ and denoted $\rho_k := \|w^k\|$ for all $k \in \mathbb{N}$. Then define, for all $k \in \mathbb{N}$,

$$\alpha_k = \gamma_k \frac{(f+g)(x^k) - s_k}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2}, \tag{24}$$

where $0 < \gamma \le \gamma_k \le 2-\gamma$. We assume that s_k a monotone decreasing variable target value approximating $s_* := \inf\{(f+g)(x) : x \in \mathcal{H}\}$ is available, and satisfies that $s_k \le (f+g)(x^k)$ for all $k \in \mathbb{N}$. When s_* is known, the simplest variant of the stepsizes proposed in (24) is obtained the stepsizes

$$\alpha_k = \gamma_k \frac{(f+g)(x^k) - s_*}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2}, \tag{25}$$

for all $k \in \mathbb{N}$. Unfortunately, to find an optimal solution, scheme (25) requires prior knowledge of the optimal objective function value s_*. As s_* is usually unknown, we prefer to do our analysis over (24), and replace s_* by the variable target value s_k. When g is the indicator function of a closed and convex set further discussion about how to choose s_k is presented in the literature for problems where a good upper or lower bound of the optimal objective function value is available; see, for instance, [25, 27, 41].

Now we present a direct consequence of Lemma 2.1. Denote

$$\mathcal{L}_{f+g}(s) := \{x \in \mathbf{dom}(g) : (f+g)(x) \le s\}\,.$$

Corollary 2.9 *Suppose that* $\lim_{k\to\infty} s_k = \tilde{s} \ge s_*$ *and let any* $x \in \mathcal{L}_{f+g}(\tilde{s})$. *Then,*

$$\|x^{k+1} - x\|^2 \le \|x^k - x\|^2 - \gamma(2-\gamma)\frac{\left[s_k - (f+g)(x^k)\right]^2}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2},$$

for all $k \in \mathbb{N}$.

Proof Take $x \in \mathcal{L}_{f+g}(\tilde{s}) = \{x \in \mathbf{dom}(g) : (f+g)(x) \le \tilde{s}\}$. Since $(s_k)_{k\in\mathbb{N}}$ is a monotone decreasing sequence convergent to $\tilde{s}$, which is less than the function values of the iterates,

$$(f+g)(x^k) \ge s_k \ge \tilde{s} \ge (f+g)(x), \quad \forall x \in \mathcal{L}_{f+g}(\tilde{s}), \tag{26}$$

for all $k \in \mathbb{N}$. Then, applying Lemma 2. and using (26), we get, for all $k \in \mathbb{N}$,

$$\begin{aligned}
\|x^{k+1} - x\|^2 \leq & \|x^k - x\|^2 - 2\gamma_k \frac{\left[s_k - (f+g)(x^k)\right]\left[(f+g)(x) - (f+g)(x^k)\right]}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2} \\
& + \gamma_k^2 \frac{\left[s_k - (f+g)(x^k)\right]^2}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2} \\
\leq & \|x^k - x\|^2 - \gamma_k(2-\gamma_k) \frac{\left[s_k - (f+g)(x^k)\right]^2}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2} \\
\leq & \|x^k - x\|^2 - \gamma(2-\gamma) \frac{\left[s_k - (f+g)(x^k)\right]^2}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2}, \qquad (27)
\end{aligned}$$

where we used that $x \in \mathcal{L}_{f+g}(\tilde{s})$, (24) and (26) in the second inequality. The result follows from (27). □

Now we prove the first main result of this subsection in the following theorem.

Theorem 2.10 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by* ***PSS Method*** *with α_k as in (24). If $\lim_{k\to\infty} s_k = \tilde{s} \geq s_*$ and $\mathcal{L}_{f+g}(\tilde{s}) \neq \emptyset$, then*

(a) *$(x^k)_{k\in\mathbb{N}}$ is Fejér convergent to $\mathcal{L}_{f+g}(\tilde{s})$.*
(b) *$\lim_{k\to\infty} (f+g)(x^k) = \tilde{s}$.*
(c) *$(x^k)_{k\in\mathbb{N}}$ is weakly convergent to some $\tilde{x} \in \mathcal{L}_{f+g}(\tilde{s})$.*

Proof **(a)** It is direct consequence of Corollary 2.9.

(b) By Item **(a)**, the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded. By using Corollary 2.9, at any $x \in \mathcal{L}_{f+g}(\tilde{s})$, we get

$$\begin{aligned}
\gamma(2-\gamma)\left[s_k - (f+g)(x^k)\right]^2 & \leq \left(\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2\right)\left[\|x^k - x\|^2 - \|x^{k+1} - x\|^2\right] \qquad (28) \\
& \leq \left(\zeta^2 + 2\rho\zeta + \rho^2\right)\left[\|x^k - x\|^2 - \|x^{k+1} - x\|^2\right] \\
& := \hat{\rho}\left[\|x^k - x\|^2 - \|x^{k+1} - x\|^2\right], \qquad (29)
\end{aligned}$$

where the last inequality follows from Assumptions **A1** and **A2** ($\|u^k\| \leq \zeta$ and $\rho_k = \|w^k\| \leq \rho$ for all $k \in \mathbb{N}$). Summing (29), over $k = 0$ to m, we obtain

$$\begin{aligned}
\gamma(2-\gamma)\sum_{k=0}^{m}\left[s_k - (f+g)(x^k)\right]^2 & \leq \hat{\rho}\left[\|x^0 - x\|^2 - \|x^{m+1} - x\|^2\right] \\
& \leq \hat{\rho}\|x^0 - x\|^2.
\end{aligned}$$

Taking limits, when m goes to ∞, we get the desired result.

(c) From Item **(b)**, if $\tilde{s} = \lim_{k\to\infty} s_k$ then $\lim_{k\to\infty} (f+g)(x^k) = \tilde{s}$. Let $\tilde{x}$ be a weak accumulation point of $(x^k)_{k\in\mathbb{N}}$, which exists by the boundedness of $(x^k)_{k\in\mathbb{N}}$ direct consequence of Item **(a)**. From now on, we denote $(x^{\ell_k})_{k\in\mathbb{N}}$ any subsequence of $(x^k)_{k\in\mathbb{N}}$, which converges weakly to $\tilde{x}$. Since $f+g$ is weakly lower semicontinuous, we get $(f+g)(\tilde{x}) \leq \liminf_{k\to\infty}(f+g)(x^{\ell_k}) = \lim_{k\to\infty}(f+g)(x^k) = \tilde{s}$, implying that $(f+g)(\tilde{x}) \leq \tilde{s}$ and thus $\tilde{x} \in \mathcal{L}_{f+g}(\tilde{s})$. The result follows from Fact 1.1(b) and Item **(a)**.

□

Before the analysis of the inconsistent case when $\tilde{s} = \lim_{k\to\infty} s_k$ is strictly less than $s_* = \inf\{(f+g)(x) : x \in \mathcal{H}\}$, we present a useful corollary which is a direct consequence of Theorem 2.10, that shall be used for the analysis of this case, $\tilde{s} < s_*$. In the next corollary, we show the special case when the optimal value s_* is known and finite and the stepsize α_k is defined by (25), *i.e.*, for all $k \in \mathbb{N}$,

$$\alpha_k = \gamma_k \frac{(f+g)(x^k) - s_*}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2},$$

where $0 < \gamma \le \gamma_k \le 2 - \gamma$.

Corollary 2.11 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by **PSS Method** with α_k given by* (25). *Assume that $S_* \neq \emptyset$. Then,*

(a) $(x^k)_{k\in\mathbb{N}}$ *is Fejér convergent to S_*.*
(b) $\lim_{k\to\infty}(f+g)(x^k) = \min_{x\in\mathcal{H}}(f+g)(x)$.
(c) $(x^k)_{k\in\mathbb{N}}$ *is weakly convergent to some $\tilde{x} \in S_*$.*
(d) $\liminf_{k\to\infty} \sqrt{k+1}\cdot\left[(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)\right] = 0$.

Proof Items **(a)** to **(c)** are direct consequence of Theorem 2.10. The proof of Item **(d)** is by contradiction. Assume that $\liminf_{k\to\infty} \sqrt{k+1}\cdot\left[(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)\right] \ge 2\delta$, for some $\delta > 0$. Then, for $\bar{k}$ large enough, we have $(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x) \ge \frac{\delta}{\sqrt{k+1}}$ for all $k \ge \bar{k}$. Thus,

$$\sum_{k=\bar{k}}^{\infty}\left[(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)\right]^2 \ge \delta^2 \sum_{k=\bar{k}}^{\infty} \frac{1}{k+1} = +\infty. \tag{30}$$

On the other hand, by substituting the expression for the stepsize α_k given by (25), in (29) ($s_k = \min_{x\in\mathcal{H}}(f+g)(x)$ for all $k \in \mathbb{N}$), we get, for all $k \ge \bar{k}$,

$$\sum_{k=\bar{k}}^{\infty}\left[(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)\right]^2 < +\infty,$$

which contradicts (30), thus establishing the result. □

Next we present a result on the complexity of the iterates.

Lemma 2.12 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by **PSS Method** with α_k, given by* (24). *If $\lim_{k\to\infty} s_k = \tilde{s} \ge s_*$ and $\mathcal{L}_{f+g}(\tilde{s}) \neq \emptyset$, then, for all $k \in \mathbb{N}$,*

$$(f+g)_{best}^k - \tilde{s} \le \sqrt{\frac{D_k}{\gamma(2-\gamma)}} \cdot \frac{dist(x^0, \mathcal{L}_{f+g}(\tilde{s}))}{\sqrt{k+1}},$$

where $D_k := \max\left\{\|u^i\|^2 + 2\rho_i\|u^i\| + \rho_i^2 : 1 \le i \le k\right\}$ with $\rho_i := \|w^i\|$ and $w^i \in \partial g(x^i)$ $(i = 0, \ldots, k)$ are arbitrary. Moreover,

$$\lim_{k\to\infty} (f+g)_{best}^k = \tilde{s}.$$

Proof Repeating the proof of Theorem 2.10, with $\tilde{x} := \mathbf{P}_{\mathcal{L}_{f+g}(\tilde{s})}(x^0) \in \mathcal{L}_{f+g}(\tilde{s})$, until (28), we obtain

$$(k+1)\left[(f+g)^k_{\text{best}} - \tilde{s}\right]^2 \leq \sum_{i=0}^{k}\left[(f+g)(x^i) - s_k\right]^2 \leq \frac{D_k}{\gamma(2-\gamma)}\left[\operatorname{dist}(x^0, \mathcal{L}_{f+g}(\tilde{s}))\right]^2,$$

where $D_k := \max\left\{\|u^i\|^2 + 2\rho_i\|u^i\| + \rho_i^2 : 1 \leq i \leq k\right\}$ with $\rho_i = \|w^i\|$ and $w^i \in \partial g(x^i)$ $(i = 0, \ldots, k)$ are arbitrary. After simple algebra the result follows. □

Our analysis proved that the expected error of the iterates generated by **PSS Method** with the Polyak stepsizes (24) after k iterations is $\mathcal{O}\left((k+1)^{-1/2}\right)$ if we assume $s_k \geq s_*$ for all $k \in \mathbb{N}$ and $(D_k)_{k\in\mathbb{N}}$ is bounded.

Now we are ready to prove the last main result of this subsection.

Theorem 2.13 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by **PSS Method** with α_k, given by* (24). *If $S_* \neq \emptyset$ and $\lim_{k\to\infty} s_k = \tilde{s} < \min_{x\in\mathcal{H}}(f+g)(x)$, then*

$$\lim_{k\to\infty}(f+g)^k_{best} = \lim_{k\to\infty}\min_{0\leq i\leq k}(f+g)(x^i) \leq \min_{x\in\mathcal{H}}(f+g)(x) + \frac{2-\gamma}{\gamma}\left[\min_{x\in\mathcal{H}}(f+g)(x) - \tilde{s}\right].$$

Proof Suppose that $(f+g)(x^k) > \min_{x\in\mathcal{H}}(f+g)(x)$, otherwise the result holds trivially. It is clear that, for all $k \in \mathbb{N}$,

$$\begin{aligned}\alpha_k &= \gamma_k \frac{(f+g)(x^k) - s_k}{(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)} \frac{(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2} \\ &:= \tilde{\gamma}_k \frac{(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)}{\|u^k\|^2 + 2\rho_k\|u^k\| + \rho_k^2},\end{aligned}$$

where

$$\gamma \leq \tilde{\gamma}_k = \gamma_k \frac{(f+g)(x^k) - s_k}{(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x)},$$

which implies that $\tilde{\gamma}_k$ is greater than $2-\gamma$ for some $\bar{k} \in \mathbb{N}$. Otherwise, if

$$\tilde{\gamma}_k \leq 2 - \gamma \tag{31}$$

for all $k \in \mathbb{N}$, we can apply Corollary 2.11b to get $\lim_{k\to\infty}(f+g)(x^k) = \min_{x\in\mathcal{H}}(f+g)(x)$, implying that $\tilde{\gamma}_k$ goes to $+\infty$ (note that for all sufficiently large k, $s_k < \min_{x\in\mathcal{H}}(f+g)(x) \leq (f+g)(x^k)$, because $\tilde{s} < \min_{x\in\mathcal{H}}(f+g)(x)$), which is a contradiction with (31). Thus, there exist $\bar{k}$ and $\delta > 0$ arbitrary such that

$$\gamma_{\bar{k}} \frac{(f+g)(x^{\bar{k}}) - s_{\bar{k}}}{(f+g)(x^{\bar{k}}) - \min_{x\in\mathcal{H}}(f+g)(x)} = \tilde{\gamma}_{\bar{k}} > 2 - \delta.$$

After simple algebra and using that $s_{\bar{k}} \geq \tilde{s}$, we get that

$$\begin{aligned}(f+g)(x^{\bar{k}}) &< \min_{x\in\mathcal{H}}(f+g)(x) + \frac{\gamma_{\bar{k}}}{2-\delta-\gamma_{\bar{k}}}[\min_{x\in\mathcal{H}}(f+g)(x) - \tilde{s}] \\ &\leq \min_{x\in\mathcal{H}}(f+g)(x) + \frac{2-\gamma}{\gamma-\delta}[\min_{x\in\mathcal{H}}(f+g)(x) - \tilde{s}],\end{aligned}$$

since $\delta > 0$ was arbitrary and the result follows. □

Finally in the following corollary we summarize the behaviour of the limit of the sequence of $\big((f+g)^k_{\text{best}}\big)_{k\in\mathbb{N}}$ depending on the limit of $\tilde{s} = \lim_{k\to\infty} s_k$, which is direct consequence of Theorems 2.10b and 2.13 and Lemma 2.12.

Corollary 2.14 *Let* $(x^k)_{k\in\mathbb{N}}$ *be the sequence generated by* ***PSS Method*** *with* α_k*, given by* (24). *If* $S_* \neq \emptyset$ *and* $\lim_{k\to\infty} s_k = \tilde{s}$*, then*

$$\lim_{k\to\infty}(f+g)^k_{best}\begin{cases} = \lim\limits_{k\to\infty}(f+g)(x^k) = \tilde{s}, & if\ \tilde{s} \geq \min\limits_{x\in\mathcal{H}}(f+g)(x) \\ \leq \min\limits_{x\in\mathcal{H}}(f+g)(x) + \dfrac{2-\gamma}{\gamma}\left[\min\limits_{x\in\mathcal{H}}(f+g)(x) - \tilde{s}\right], & if\ \tilde{s} < \min\limits_{x\in\mathcal{H}}(f+g)(x). \end{cases}$$

3 Final Remarks

In this work we dealt with the weak convergence and the complexity analysis of the new approach called the Proximal Subgradient Splitting (PSS) Method for minimizing the sum of two nonsmooth and convex functions under standard assumptions (namely Assumptions **A1** and **A2**). It worth mentioning that, these kind of boundedness assumptions, are needed even for the convergence analysis of the classical subgradient iteration, and hopefully its relaxation will be addressed in future research. Adding that, in the iteration of the proposed iteration, none of the functions need be differentiable or finite on $\mathcal{H}$ and, therefore, a broad class of problems can be solved. **PSS Method** is very useful when the proximal operator of f is complicated to evaluate and its (sub)gradient is simple to compute.

As future research, we will investigate variations of our scheme for solving structured convex optimization problems with the aim of finding new methods, like the coordinate gradient method, which have been proposed, for instance, in [36] only for the differentiable case. We also look at the incremental subgradient method [28, 33] for problem (1), when f is the sum of a large number of nonsmooth convex functions. The idea is to perform subgradient iterations incrementally, by sequentially taking steps along the subgradients of the component functions, followed by proximal steps. On the other hand, it is important to mention that the main drawback of subgradient iterations is their slow rate of convergence. However, subgradient methods are distinguished by their applicability, simplicity and efficient use of memory, which is very important for large scale problems; especially if the required accuracy for the solution is not too high; see, for instance, [34] and the references therein. We also will intend to study fast and variable metric versions of the proximal subgradient splitting method proposed here to achieve better performance, as in the differentiable case; see [20].

Finally, we hope that this study serves as a basis for future research on other more efficient variants on the proximal subgradient iteration, like cutting-plane method, ϵ-subgradients and proximal bundle method and its variations; see [28, 29, 40]. Moreover, in future work we discuss useful modifications on the proximal subgradient iteration adding conditional, ergodic and deflected techniques combining the ideas presented in [21, 30].

Acknowledgments The author was partially supported by CNPq grants 303492/2013-9, 474160/2013-0 and 202677/2013-3. This work was partially completed while the author was visiting the University of British Columbia. The author is very grateful for the warm hospitality of the Irving K. Barber School of Arts and Sciences, Mathematics at the University of British Columbia Okanagan and particularly to Professors Heinz H. Bauschke and Shawn Wang for the generous hospitality. The author would like to thank to anonymous referees and associate editors whose suggestions helped us to improve the presentation of this paper.

References

1. Alber, Y.I., Iusem, A.N., Solodov, M.V.: On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. Math Program **81**, 23–37 (1998)
2. Bauschke, H.H., Borwein, J.: On projection algorithms for solving convex feasibility problems. SIAM Rev **38**, 367–426 (1996)
3. Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York (2011)
4. Bauschke, H.H., Koch, V.R., Phan, H.M.: Stadium norm and Douglas-Rachford splitting: a new approach to road design optimization. Operations Research (2016) in press
5. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring. IEEE Trans Image Process **18**, 2419–2434 (2009)
6. Beck, A., Teboulle, M.: Gradient-Based Algorithms with Applications to Signal Recovery Problems. In: Palomar, D., Eldar, Y. (eds.) Convex Optimization in Signal Processing and Communications, pp. 42–88. University Press, Cambribge (2010)
7. Bello Cruz, J.Y.: A subgradient method for vector optimization problems. SIAM J Optim **23**, 2169–2182 (2013)
8. Bello Cruz, J.Y., Iusem, A.N.: A strongly convergent method for nonsmooth convex minimization in Hilbert spaces. Numer Funct Anal Optim **32**, 1009–1018 (2011)
9. Bello Cruz, J.Y., Iusem, A.N.: Convergence of direct methods for paramonotone variational inequalities. Comput Optim Appl **46**, 247–263 (2010)
10. Bello Cruz, J.Y., Nghia, T.T.A.: On the convergence of the proximal forward-backward splitting method with linesearches. Technical report, 2015, Available in arXiv:1501.02501.pdf
11. Bot, R., Csetnek, E.R.: Forward-Backward and Tseng's type penalty schemes for monotone inclusion problems. Set-Valued and Variational Analysis **22**, 313–331 (2014)
12. Candes, E.J., Tao, T.: Decoding by linear programming. IEEE Trans Inf Theory **51**, 4203–4215 (2005)
13. Chavent, G., Kunisch, K.: Convergence of Tikhonov regularization for constrained ill-posed inverse problems. Inverse Prob **10**, 63–76 (1994)
14. Chen, G.H.-G., Rockafellar, R.T.: Convergence rates in forward-backward splitting. SIAM J Optim **7**, 421–444 (1997)
15. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**, 475–504 (2004)
16. Combettes, P.L.: Quasi-Fejérian analysis of some optimization algorithms. Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications. Studies in Computational Mathematics 8 115–152 North-Holland Amsterdam (2001)
17. Combettes, P.L., Pesquet, J.-C.: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE Journal of selected topics in signal precessing **1**, 564–574 (2007)
18. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: Fixed-Point Algorithms for Inverse Problems. Science and Engineering. Springer Optimization and Its Applications 49 185–212 Springer New York (2011)
19. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model Simul **4**, 1168–1200 (2005)
20. Combettes, P.L., Vũ, B.C.: Variable metric forward-backward splitting with applications to monotone inclusions in duality. Optimization **63**, 1289–1318 (2014)
21. D'Antonio, G., Frangioni, A.: Convergence analysis of deflected conditional approximate subgradient methods. SIAM J Optim **20**, 357–386 (2009)
22. Ermoliev, Y.U.M.: On the method of generalized stochastic gradients and quasi-Fejér sequences. Cybernetics **5**, 208–220 (1969)
23. Figueiredo, M., Novak, R., Wright, S.J.: Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE J Sel Top Sign Proces **1**, 586–597 (2007)
24. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell **6**, 721–741 (1984)
25. Held, M., Wolfe, P., Crowder, H.: Validation of subgradient optimization. Math Program **6**, 66–68 (1974)
26. James, G.M., Radchenko, P., Lv, J.: DASSO: connections between the Dantzig selector and lasso. J R Stat Soc Ser B Stat Methodol **71**, 127–142 (2009)
27. Kim, S., Ahn, H., Cho, S.-C.: Variable target value subgradient method. Math Program **49**, 359–369 (1991)
28. Kiwiel, K.C.: Convergence of approximate and incremental subgradient methods for convex optimization. SIAM J Optim **14**, 807–840 (2006)

29. Kiwiel, K.C.: The efficiency of subgradient projection methods for convex optimization, Parts I: general level methods. SIAM J Control Optim **34**, 660–676 (1996)
30. Larson, T., Patriksson, M., Stromberg, A.-B.: Conditional subgradient optimization - Theory and application. Eur J Oper Res **88**, 382–403 (1996)
31. Mosci, S., Rosasco, L., Santoro, M., Verri, A., Villa, S., Sebag, M.: Solving structured sparsity regularization with proximal methods. In: Balczar, J., Bonchi, F., Gionis, A. (eds.) Machine Learning and Knowledge Discovery in Databases, 6322 of Lecture Notes in Computer Science, Springer, 2010, 418–433
32. Neal, P., Boyd, S.: Proximal Algorithms. Foundations and Trends in Optimization **1**, 127–239 (2014)
33. Nedic, A., Bertsekas, D.P.: Incremental subgradient methods for nondifferentiable optimization. SIAM J Optim **12**, 109–138 (2001)
34. Nesterov, Y.U.: Subgradient methods for huge-scale optimization problems. Math Program **146**, 275–297 (2014)
35. Nesterov, Y.U.: Gradient methods for minimizing composite functions. Math Program **140**, 125–161 (2013)
36. Nesterov, Y.U.: Efficiency of coordinate descent methods on huge-scale optimization problems. SIAM J Optim **22**, 341–362 (2012)
37. Nesterov, Y.U.: Introductory lectures on convex optimization: A Basic Course. Kluwer Academic Publishers Norwel (2004)
38. Polyak, B.T.: Minimization of unsmooth functionals U.S.S.R. Comput Math Math Phys **9**, 14–29 (1969)
39. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J Control Optim **14**, 877–898 (1976)
40. Sagastizb̈al, C.: Composite proximal bundle method. Math Program **140**, 189–233 (2013)
41. Sherali, H.D., Choi, G., Tuncbilek, C.H.: A variable target value method for nondifferentiable optimization. A variable target value method for nondifferentiable optimization, 1–8 (1997)
42. Svaiter, B.F.: A class of Fejér convergent algorithms, approximate resolvents and the hybrid Proximal-Extragradient method. J Optim Theory Appl **162**, 133–153 (2014)
43. Tropp, J.: Just relax: convex programming methods for identifying sparse signals. IEEE Trans Inf Theory **51**, 1030–1051 (2006)
44. Zhu, D.L., Marcotte, P.: Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. SIAM J Optim **6**, 714–726 (1996)