A Fixed-Point Subgradient Splitting Method for Solving Constrained Convex Optimization Problems

Nimit Nimana

Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand; nimitni@kku.ac.th

Received: 4 February 2020; Accepted: 18 February 2020; Published: 3 March 2020

Abstract: In this work, we consider a bilevel optimization problem consisting of the minimizing sum of two convex functions in which one of them is a composition of a convex function and a nonzero linear transformation subject to the set of all feasible points represented in the form of common fixed-point sets of nonlinear operators. To find an optimal solution to the problem, we present a fixed-point subgradient splitting method and analyze convergence properties of the proposed method provided that some additional assumptions are imposed. We investigate the solving of some well known problems by using the proposed method. Finally, we present some numerical experiments for showing the effectiveness of the obtained theoretical result.

Keywords: bilevel optimization; convex optimization; fixed point; subgradient method

AMS Subject Classification: 47J25; 47N10; 65K05; 90C25

1. Introduction

Many applications in science and engineering have shown a huge interest in solving an inverse problem of finding $x \in \mathbb{R}^n$ satisfying

$$Bx = b,$$

where $b \in \mathbb{R}^r$ is the observed data and $B^{r \times n}$ is the corresponding nonzero matrix. Actually, the inverse problem is typically facing the ill-condition of the matrix $B$ so that it may have no solution. Then, an approach for finding an approximate solution by minimizing the squared norm of the residual term has been considered:

$$\text{minimize} \quad \|Bx - b\|^2$$
$$\text{subject to} \quad x \in \mathbb{R}^n. \tag{2}$$

Observe that the problem (2) has several optimal solutions; in this situation, it is not clear which of these solutions should be considered. One strategy for pursuing the best optimal solution among these many solutions is to add a regularization term to the objective function. The classical technique is to consider the celebrated Tikhonov regularization [1] of the form

$$\text{minimize} \quad \|Bx - b\|^2 + \lambda \|x\|^2$$
$$\text{subject to} \quad x \in \mathbb{R}^n, \tag{3}$$

where $\lambda > 0$ is a regularization parameter. In this setting, the uniqueness of the solution to (3) is acquired. However, note that from a practical point of view, the shortcoming of this strategy is that the...
unique solution to the regularization problem (3) may probably not optimal in the original sense as in (2), see [2,3] for further discussions.

To overcome this, we should consider the strategy of selecting a specific solution among optimal solutions to (2) by minimizing an additional prior function over these optimal solutions. This brings the framework of the following bilevel optimization problem,

\[
\begin{align*}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad x \in \arg\min_{u \in \mathbb{R}^n} \|Bu - b\|_2^2,
\end{align*}
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \) and \( h : \mathbb{R}^m \to \mathbb{R} \) are convex functions, and \( A : \mathbb{R}^n \to \mathbb{R}^m \) is a nonzero linear transformation. It is very important to point out that many problems can be formulated into this form. For instance, if \( m := n - 1, f := \| \cdot \|_1, h := \| \cdot \|_1, \) and

\[
A := \begin{bmatrix}
-1 & 1 & 0 & \ldots & 0 & 0 \\
0 & -1 & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & -1 & 1
\end{bmatrix} \in \mathbb{R}^{(n-1)\times n}
\]

the problem (4) becomes the fused lasso [4] solution to the problem (2). This situation also occurs in image denoising problems (\( r = n \) and \( B \) is the identity matrix), and in image inpainting problems (\( r = n \) and \( B \) is a symmetric diagonal matrix), where the term \( \|Ax\|_1 \) is known as an 1D total variation [5]. When \( m = n, f := \| \cdot \|_2^2, h := \| \cdot \|_1, \) and \( A \) is the identity matrix, the problem (4) becomes the elastic net [6] solution to the problem (2). Moreover, in wavelet-based image restoration problems, the matrix \( A \) is given by an inverse wavelet transform [7].

Let us consider the constrained set of (4), it is known that introducing the Landweber operator \( T : \mathbb{R}^n \to \mathbb{R}^n \) of the form

\[
Tx := x - \frac{1}{\|B\|^2} B^\top (Bx - b),
\]

yields \( T \) is firmly nonexpansive and the set of all fixed points of \( T \) is nothing else that the set \( \arg\min_{x \in \mathbb{R}^n} \|Bu - b\|_2^2 \), see [8] for more details. Motivated by this observation, the bilevel problem (4) can be considered in the general setting as

\[
\begin{align*}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad x \in \text{Fix} T,
\end{align*}
\]

where \( T : \mathbb{R}^n \to \mathbb{R}^n \) is a nonlinear operator with \( \text{Fix} T := \{ x \in \mathbb{R}^n : Tx = x \} \neq \emptyset \). Note that problem (5) encompasses not only the problem (4), but also many problems in the literature, for instance, the minimization over the intersection of a finite number of sublevel sets of convex nonsmooth functions (see Section 5.2), the minimization over the intersection of many convex sets in which the metric projection on such intersection can not be computed explicitly, see [9–11] for more details.

There are some existing methods for solving convex optimization problems over fixed-point set in the form of (5), but the celebrated one is due to the hybrid steepest descent method, which was firstly investigated in [12]. Note that the algorithm proposed by Yamada [12] goes on with the hypotheses that the objective functions are strongly convex and smooth and the operator \( T \) is nonexpansive. Several variants and generalizations of this well-known method are, for instance, Yamada and Ogura [11] considered the same scheme for solving the problem (5) when \( T \) belongs to a class of so called quasi-shrinking operator. Cegielski [10] proposed a generalized hybrid steepest descent method by using the sequence of quasi-nonexpansive operators. Iiduka [13,14] considered a nonsmooth convex optimization problem (5) with fixed-point constraints of certain quasi-nonexpansive operators.

On the other hand, in the recent decade, the split common fixed point problem [15,16] turns out to be one of the attractions among several nonlinear problems due to its widely applications in many image and signal processing problems. Actually, for given a nonzero linear transformation...
A : \mathbb{R}^n \rightarrow \mathbb{R}^m, and two nonlinear operators \( T : \mathbb{R}^n \rightarrow \mathbb{R}^n \) and \( S : \mathbb{R}^m \rightarrow \mathbb{R}^m \) with \( \text{Fix}(T) \neq \emptyset \), and \( \mathcal{R}(A) \cap \text{Fix}(S) \neq \emptyset \), the split common fixed point problem is to find a point \( x^* \in \mathbb{R}^n \) in which

\[
x^* \in \text{Fix } T \quad \text{such that} \quad Ax^* \in \text{Fix } S.
\]

The key idea of this problem is to find a point in the fixed point of a nonlinear operator in a primal space, and subsequently its image under an appropriate linear transformation forms a fixed point of another nonlinear operator in another space. This situation appears, for instance, in dynamic emission tomographic image reconstruction [17] and in the intensity-modulated radiation therapy treatment planning, see [18] for more details. Of course, there are many authors that have investigated the study of iterative algorithms for split common fixed point problems and proposed their generalizations in several aspects, see, for example, [9,19–22] and references therein.

The aim of this paper is to present a nonsmooth and non-strongly convex version of the hybrid steepest descent method for minimizing the sum of two convex functions over the fixed-point constraints of the form:

\[
\text{minimize} \quad f(x) + h(Ax)
\]

\[
\text{subject to} \quad x \in X \cap \text{Fix } T, Ax \in \text{Fix } S,
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) and \( h : \mathbb{R}^m \rightarrow \mathbb{R} \) are convex nonsmooth functions, \( A : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is a nonzero linear transformation, \( T : \mathbb{R}^n \rightarrow \mathbb{R}^n \) and \( S : \mathbb{R}^m \rightarrow \mathbb{R}^m \) are certain quasi-nonexpansive operators with \( \text{Fix}(T) \neq \emptyset \), and \( \mathcal{R}(A) \cap \text{Fix}(S) \neq \emptyset \), and \( X \subset \mathbb{R}^n \) is a simple closed convex bounded set. We prove the convergence of function value to the minimum value where some control conditions on a stepsize sequence and a parameter are imposed.

The paper is organized as follows. After recalling and introducing some useful notions and tools in Section 2, we present our algorithm and discuss the convergence analysis in Section 3. Furthermore, in Section 4, we discuss an important implication of our problem and algorithm to the minimizing sum of convex functions over coupling constraints. In Section 5, we discuss in detail some remarkably practical applications, and Section 6 describes the results of numerical experiments on fused lasso like problem. Finally, the conclusions are given in Section 7.

2. Preliminaries

We summarize some useful notations, definitions, and properties, which we will utilize later. For further details, the reader can consult the well-known books, for instance, in [8,23–25].

Let \( \mathbb{R}^n \) be an \( n \)-dimensional Euclidean space with an inner product \( \langle \cdot, \cdot \rangle \) and the corresponding norm \( \| \cdot \| \).

Let \( T : \mathbb{R}^n \rightarrow \mathbb{R}^n \) be an operator. We denote the set of all fixed points of \( T \) by \( \text{Fix } T \), that is, \( \text{Fix } T := \{ x \in \mathbb{R}^n : Tx = x \} \).

We say that \( T \) is \( \rho \)-strongly quasi-nonexpansive (\( \rho \)-SQNE), where \( \rho \geq 0 \), if \( \text{Fix } T \neq \emptyset \) and

\[
\|Tx - z\|^2 \leq \|x - z\|^2 - \rho\|Tx - x\|^2,
\]

for all \( x \in \mathbb{R}^n \) and \( z \in \text{Fix } T \). If \( \rho > 0 \), then \( T \) is called strongly quasi-nonexpansive (SQNE). If \( \rho = 0 \), then \( T \) is called quasi-nonexpansive (QNE), that is,

\[
\|Tx - z\| \leq \|x - z\|,
\]

for all \( x \in \mathbb{R}^n \) and \( z \in \text{Fix } T \). Clearly, if \( T \) is SQNE, then it is QNE. We say that \( T \) is cutter if \( \text{Fix } T \neq \emptyset \) and

\[
\langle x - Tx, z - Tx \rangle \leq 0,
\]
for all \( x \in \mathbb{R}^n \) and all \( z \in \text{Fix} \, T \). We say that \( T \) is \textit{firmly nonexpansive} (FNE) if
\[
\langle Tx - Ty, x - y \rangle \geq \|Tx - Ty\|^2,
\]
for all \( x, y \in \mathbb{R}^n \).

The following properties will be applied in the next sections.

**Fact 1** (Lemma 2.1.21 [8]). \( \text{If } T : \mathbb{R}^n \to \mathbb{R}^n \text{ is QNE, then Fix } T \text{ is closed and convex.} \)

**Fact 2** (Theorem 2.2.5 [8]). \( \text{If } T : \mathbb{R}^n \to \mathbb{R}^n \text{ is FNE with } \text{Fix } T \neq \emptyset, \text{ then } T \text{ is a cutter.} \)

Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function and \( x \in \mathbb{R}^n \). The subdifferential of \( f \) at \( x \) is the set
\[
\partial f(x) := \{ x^* \in \mathbb{R}^n : f(y) \geq f(x) + \langle x^*, y - x \rangle \text{ for all } y \in \mathbb{R}^n \}.
\]
If \( \partial f(x) \neq \emptyset \), then an element \( f'(x) \in \partial f(x) \) is called a subgradient of \( f \) at \( x \).

**Fact 3** (Corollary 16.15 [24]). \( \text{Let } f : \mathbb{R}^n \to \mathbb{R} \text{ be a convex function. Then, the subdifferential } \partial f(x) \neq \emptyset \text{ for all } x \in \mathbb{R}^n. \)

**Fact 4** (Proposition 16.17 [24]). \( \text{Let } f : \mathbb{R}^n \to \mathbb{R} \text{ be a convex function. Then, the subdifferential } \partial f \) maps every bounded subset of \( \mathbb{R}^n \) to a bounded set.

As we work on the \( n \)-dimensional Euclidean space, we will use the notion of matrix instead of the notion of linear transformation throughout this work. Denote \( \mathbb{R}^{m \times n} \) by the set of all real-valued \( m \times n \) matrices. Let \( A \in \mathbb{R}^{m \times n} \) be given. We denote by \( \mathcal{R}(A) := \{ y \in \mathbb{R}^m : y = Ax \text{ for some } x \in \mathbb{R}^n \} \) its range, and \( A^\top \) its (conjugate) transpose. We denote the induced norm of \( A \) by \( \|A\| \), which is given by \( \|A\| = \sqrt{\lambda_{\text{max}}(A^\top A)} \), where \( \lambda_{\text{max}}(A^\top A) \) is the maximum eigenvalue of the matrix \( A^\top A \).

### 3. Method and its Convergence

Now, we formulate the composite nonsmooth convex minimization problem over the intersections of fixed-point sets which we aim to investigate throughout this paper.

**Problem 1.** \( \text{Let } \mathbb{R}^n \text{ and } \mathbb{R}^m \text{ be two Euclidean spaces. Assume that} \)

(A1) \( f : \mathbb{R}^n \to \mathbb{R} \) and \( h : \mathbb{R}^m \to \mathbb{R} \) are convex functions.

(A2) \( A \in \mathbb{R}^{m \times n} \) is a nonzero matrix.

(A3) \( T : \mathbb{R}^n \to \mathbb{R}^n \) is QNE, and \( S : \mathbb{R}^m \to \mathbb{R}^m \) is cutter with \( \mathcal{R}(A) \cap \text{Fix}(S) \neq \emptyset. \)

(A4) \( X \) is a nonempty convex closed bounded simple subset of \( \mathbb{R}^n. \)

Our objective is to solve

minimize \( f(x) + h(Ax) \)
subject to \( x \in X \cap \text{Fix} \, T, Ax \in \text{Fix} \, S. \)

Throughout this work, we denote the solution set of Problem 1 by \( \Gamma \) and assume that it is nonempty.

Problem 1 can be viewed as a bilevel problem in which data given from two sources in a system. Actually, let us consider the system of two users in different sources (they can have differently a number of factors \( n \) and \( m \)) in which they can communicate to each other via the transformation \( A \). The first user aims to find the best solutions with respect to criterion \( f \) among many feasible points represented in the form of fixed point set of an appropriate operator \( T \). Similarly, the second user has its own objective in the same fashion of finding the best solutions among feasible points in \( \text{Fix} \, S \)
with priori criterion \( h \). Now, to find the best solutions of this system, we consider the fixed-point subgradient splitting method (in short, FSSM) as follows, see Algorithm 1.

### Algorithm 1: Fixed-Point Subgradient Splitting Method.

**Initialization:** The positive sequence \( \{\alpha_k\}_{k \geq 1} \) and the parameter \( \gamma \in (0, +\infty) \), and an arbitrary \( x_1 \in \mathbb{R}^n \).

**Iterative Step:** For given \( x_k \in \mathbb{R}^n \), compute

\[
\begin{align*}
z_k &:= SAx_k - \frac{\alpha_k}{\gamma} h'(SAx_k), \text{ where } h'(SAx_k) \in \partial h(SAx_k), \\
y_k &:= x_k + \gamma A^\top(z_k - Ax_k), \\
x_{k+1} &:= P_X(Ty_k - \alpha_k f'(Ty_k)), \text{ where } f'(Ty_k) \in \partial f(Ty_k).
\end{align*}
\]

**Remark 1.** Actually, this algorithm has simultaneously the following features; (i) splitting computation, (ii) simple scheme, and (iii) boundedness of iterates. Concerning the first feature, we present the iterative scheme by allowing the process of a subgradient of \( f \) and a use of operator \( T \) in the space \( \mathbb{R}^n \), and a subgradient of \( h \) and a use of operator \( S \) in the space \( \mathbb{R}^m \) separately. Regarding the simplicity of iterative scheme, we need not to compute the inverse of the matrix \( A \); in this case, the transpose of \( A \) is enough. Finally, the third feature is typically required when performing the convergence of subgradient type method. Of course, the boundedness is often considered in image processing and machine learning in the form of a (big) box constraint or a big Euclidean ball.

To study the convergence properties of a function values of a sequence generated by Algorithm 1, we start with the following technical result.

**Lemma 1.** Let \( \{x_k\}_{k \geq 1} \) be a sequence generated by Algorithm 1. Then, for every \( k \geq 1 \) and \( u \in X \cap \text{Fix } T \cap A^{-1}(\text{Fix } S) \), it holds that

\[
\|x_{k+1} - u\|^2 \leq \|x_k - u\|^2 - \left(1 - \gamma \|A\|^2\right)\gamma \|z_k - Ax_k\|^2 \\
+ \alpha_k^2\|f'(Ty_k)\|^2 + \left(\frac{\gamma^2 + 2\gamma + 2}{\gamma}\right)\alpha_k^2\|h'(SAx_k)\|^2 \\
+ 2\alpha_k \left[(f(u) + h(Au)) - (f(Ty_k) + h(SAx_k))\right].
\]

**Proof.** Let \( k \geq 1 \) be arbitrary. By the definition of \( \{y_k\}_{k \geq 1} \), we have

\[
\|y_k - u\|^2 = \|x_k - u\|^2 + \gamma^2\|A^\top(z_k - Ax_k)\|^2 + 2\gamma \left\langle x_k - u, A^\top(z_k - Ax_k)\right\rangle \\
\leq \|x_k - u\|^2 + \gamma^2\|A\|^2\|z_k - Ax_k\|^2 + 2\gamma \left\langle Ax_k - Au, z_k - Ax_k\right\rangle. \tag{6}
\]
Now, by using the definition of \( \{z_k\}_{k \geq 1} \) and the cutter property of \( S \), we derive

\[
\langle Ax_k - Au, z_k - Ax_k \rangle = \langle Ax_k - z_k, z_k - Ax_k \rangle + \langle z_k - Au, z_k - Ax_k \rangle
\]

\[
= -\|z_k - Ax_k\|^2 + \langle SAx_k - Au, z_k - Ax_k \rangle
\]

\[
- \frac{\alpha_k}{\gamma} \langle h'(SAx_k), z_k - Ax_k \rangle
\]

\[
= -\|z_k - Ax_k\|^2 + \langle SAx_k - Au, SAx_k - Ax_k \rangle
\]

\[
- \frac{\alpha_k}{\gamma} \langle h'(SAx_k), SAx_k - Ax_k \rangle
\]

\[
- \frac{\alpha_k}{\gamma} \langle h'(SAx_k), SAx_k - Ax_k \rangle,
\]

which in turn implies that (6) becomes

\[
\|y_k - u\|^2 \leq \|x_k - u\|^2 + \left( \gamma\|A\|^2 - 2 \right) \|z_k - Ax_k\|^2 + 2\frac{\alpha_k^2}{\gamma} \|h'(SAx_k)\|^2
\]

\[
- 2\alpha_k \langle h'(SAx_k), SAx_k - Au \rangle - 2\alpha_k \langle h'(SAx_k), SAx_k - Ax_k \rangle.
\] (7)

We now focus on the last two terms of the right-hand side of (7).

Observe that

\[
0 \leq \frac{1}{2\gamma} \left[ \gamma\|SAx_k - Ax_k\|^2 + 2\alpha_k h'(SAx_k) \right]^2
\]

\[
= \frac{\gamma}{2} \|SAx_k - Ax_k\|^2 + 2\alpha_k^2 \|h'(SAx_k)\|^2 + 2\alpha_k \langle SAx_k - Ax_k, h'(SAx_k) \rangle,
\]

thus, by the definition of \( \{z_k\}_{k \geq 1} \), we obtain

\[
- 2\alpha_k \langle h'(SAx_k), SAx_k - Ax_k \rangle \leq \frac{\gamma}{2} \|SAx_k - Ax_k\|^2 + 2\alpha_k^2 \|h'(SAx_k)\|^2
\]

\[
\leq \gamma \|z_k - Ax_k\|^2 + (2 + \gamma)\alpha_k^2 \|h'(SAx_k)\|^2.
\] (8)

Now, inequalities (6)–(8) together give

\[
\|y_k - u\|^2 \leq \|x_k - u\|^2 - \left( 1 - \gamma\|A\|^2 \right) \|z_k - Ax_k\|^2
\]

\[
+ \left( \frac{2}{\gamma} + 2 + \gamma \right) \alpha_k^2 \|h'(SAx_k)\|^2 - 2\alpha_k \langle h'(SAx_k), SAx_k - Au \rangle.
\] (9)

On the other hand, using the definition of \( \{x_k\}_{k \geq 1} \) and the assumption that \( T \) is QNE, we obtain

\[
\|x_{k+1} - u\|^2 \leq \|Ty_k - a_k f'(Ty_k) - u\|^2
\]

\[
= \|Ty_k - u\|^2 + (a_k^2\|f'(Ty_k)\|^2 - 2a_k \langle f'(Ty_k), Ty_k - u \rangle
\]

\[
\leq \|y_k - u\|^2 + (a_k^2\|f'(Ty_k)\|^2 - 2a_k \langle f'(Ty_k), Ty_k - u \rangle.
\] (10)
Replacing (9) in (10), we obtain
\[
\|x_{k+1} - u\|^2 \leq \|x_k - u\|^2 - \left(1 - \gamma \|A\|^2\right) \gamma \|z_k - Ax_k\|^2 \\
\quad + \alpha_k^2 \|f'(Ty_k)\|^2 + \left(\frac{\gamma^2 + 2\gamma + 2}{\gamma}\right) \alpha_k^2 \|h'(SAx_k)\|^2 \\
\quad - 2\alpha_k \langle f'(Ty_k), Ty_k - u \rangle - 2\alpha_k \langle h'(SAx_k), SAx_k - Au \rangle.
\]  
(11)

Next, the convexities of \(f\) and \(g\) give
\[
\langle f'(Ty_k), u - Ty_k \rangle \leq f(u) - f(Ty_k)
\]
and
\[
\langle h'(SAx_k), Au - SAx_k \rangle \leq h(Au) - h(SAx_k).
\]

By making use of these two inequalities in (11), we obtain
\[
\|x_{k+1} - u\|^2 \leq \|x_k - u\|^2 - \left(1 - \gamma \|A\|^2\right) \gamma \|z_k - Ax_k\|^2 \\
\quad + \alpha_k^2 \|f'(Ty_k)\|^2 + \left(\frac{\gamma^2 + 2\gamma + 2}{\gamma}\right) \alpha_k^2 \|h'(SAx_k)\|^2 \\
\quad + 2\alpha_k \left[\left(f(u) + h(Au)\right) - \left(f(Ty_k) + h(SAx_k)\right)\right],
\]
which is the required inequality and the proof is completed. \(\square\)

The following lemma is very useful for the convergence result.

**Lemma 2.** Let \(\{x_k\}_{k \geq 1}\) be a sequence generated by Algorithm 1. Then, \(\{x_k\}_{k \geq 1}\) is bounded. Furthermore, if \(0 < \gamma < \frac{1}{\|A\|^2}\) and \(\{\alpha_k\}_{k \geq 1}\) is bounded, then the sequences \(\{SAx_k\}_{k \geq 1}, \{h'(SAx_k)\}_{k \geq 1}, \{z_k\}_{k \geq 1}, \{y_k\}_{k \geq 1}, \{Ty_k\}_{k \geq 1}, \text{and} \ \{f'(Ty_k)\}_{k \geq 1}\) are bounded.

**Proof.** As \(X\) is a bounded set, it is clear that the sequence \(\{x_k\}_{k \geq 1}\) is bounded. Now, let \(u \in \Gamma\) be given. The linearity of \(A\) and quasi-nonexpansiveness of \(S\) yield
\[
\|SAx_k\| \leq \|SAx_k - Au\| + \|Au\| \leq \|A\| \|x_k - u\| + \|Au\|.
\]
This implies that \(\{SAx_k\}_{k \geq 1}\) is bounded. Consequently, applying Fact 4, we obtain that \(\{h'(SAx_k)\}_{k \geq 1}\) is also bounded.

By the triangle inequality, we have
\[
\|z_k\| \leq \|SAx_k\| + \frac{\alpha_k}{\gamma} \|h'(SAx_k)\|.
\]
Therefore, the boundedness of \(\{\alpha_k\}_{k \geq 1}\) implies that \(\{z_k\}_{k \geq 1}\) is bounded. Consequently, the triangle inequality and the linearity of \(A^+\) yields the boundedness of \(\{y_k\}_{k \geq 1}\). As \(T\) is QNE, we have \(\{Ty_k\}_{k \geq 1}\) is bounded. Thus, \(\{f'(Ty_k)\}_{k \geq 1}\) is bounded by Fact 4. \(\square\)

For the sake of simplicity, we let
\[
(f + h \circ A)^* := \inf_{z \in X \cap \text{Fix } T \cap A^{-1}(\text{Fix } S)} \left(f(z) + h(Az)\right)
\]
and assume that \((f + h \circ A)^* > -\infty\).
We consider a convergence property in objective values with diminishing stepsize as the following theorem.

**Theorem 1.** Let \( \{x_k\}_{k \geq 1} \) be a sequence generated by Algorithm 1. If the following control conditions hold,

(i) \( 0 < \gamma < \frac{1}{\|A\|^2} \);

(ii) \( \sum_{k=1}^{\infty} \alpha_k = +\infty \) and \( \sum_{k=1}^{\infty} \alpha_k^2 < +\infty \);

then

\[
\liminf_{k \to +\infty} (f(Ty_k) + h(SAx_k)) \leq (f + h \circ A)^*.
\]

**Proof.** Let \( z \in \Gamma \) be given. We note from Lemma 1 that for every \( k \geq 1 \)

\[
2\alpha_k [(f(Ty_k) + h(SAx_k)) - (f(z) + h(Az))] \leq \|x_k - z\|^2 - \|x_{k+1} - z\|^2 + \alpha_k^2 \|f'(Ty_k)\|^2 + \left( \frac{\gamma^2 + 2\gamma + 2}{\gamma} \right) \alpha_k^2 \|h'(SAx_k)\|^2,
\]

this is true because \( (1 - \gamma \|A\|^2) \gamma \|z - Ax_k\|^2 \leq 0 \) via the assumption (i). Summing up (12) for \( 1, \ldots, k \) we obtain that

\[
2 \sum_{i=1}^{k} \alpha_k [(f(Ty_k) + h(SAx_k)) - (f + h \circ A)^*]) \leq \|x^1 - z\|^2 + M \sum_{i=1}^{k} \alpha_k^2,
\]

where

\[
M := \sup_{k \geq 1} \left\{ \|f'(Ty_k)\|^2 + \left( \frac{\gamma^2 + 2\gamma + 2}{\gamma} \right) \|h'(SAx_k)\|^2 \right\}.
\]

This implies that

\[
\sum_{i=1}^{k} \alpha_k [(f(Ty_k) + h(SAx_k)) - (f + h \circ A)^*]) \leq +\infty.
\]

Next, we show that \( \liminf_{k \to +\infty} ((f(Ty_k) + h(SAx_k)) - (f + h \circ A)^*)) \leq 0 \). By supposing a contradiction that

\[
\liminf_{k \to +\infty} ((f(Ty_k) + h(SAx_k)) - (f + h \circ A)^*)) > 0,
\]

there exist \( k_0 \geq 1 \) and \( \gamma > 0 \) such that

\[
((f(Ty_k) + h(SAx_k)) - (f + h \circ A)^*)) \geq \gamma
\]

for all \( k \geq k_0 \). Thus, we have

\[
+\infty = \gamma \sum_{k=k_0}^{\infty} \alpha_k \leq \sum_{k=k_0}^{\infty} \alpha_k ((f(Ty_k) + h(SAx_k)) - (f + h \circ A)^*)) < +\infty,
\]

which is a contradiction. Therefore, we can conclude that

\[
\liminf_{k \to +\infty} (f(Ty_k) + h(SAx_k)) \leq (f + h \circ A)^*.
\]

\[\Box\]

**Remark 2.** The convergence results obtained in Theorem 1 are slightly different from the convergence results obtained by the classical gradient method or even projected gradient method, namely, \( \liminf_{k \to +\infty} (f(Ty_k) + \)
\( h(SAx_k) = (f + h \circ A)^* \). This is because, in each iteration, we cannot ensure whether the estimate \( Tx^k \) is belonging to the constrained set \( \text{Fix}(T) \) or not, this means that the property \( f(Ty_k) \geq f^* \) may not be true in general. Similarly, we cannot ensure that \( h(SAx_k) \geq (h \circ A)^* \).

**Remark 3.** A step size sequence satisfies the assumption that \( \sum_{k=1}^{\infty} a_k = +\infty \) and \( \sum_{k=1}^{\infty} a_k^2 < +\infty \) is, for instance, \( \{\frac{1}{k}\}_{k \geq 1} \) where \( a > 0 \).

### 4. Convex Minimization Involving Sum of Composite Functions

The aim of this section is to show that Algorithm 1 and their convergence properties can be employed when solving a convex minimization involving sum of a finite number of composite functions.

Let us take a look the composite convex minimization problem:

\[
\begin{align*}
\text{minimize} & \quad f(x) + \sum_{i=1}^{l} h_i(A_ix) \\
\text{subject to} & \quad x \in X \cap \text{Fix} T, A_ix \in \text{Fix} S_i, i = 1, \ldots, l.
\end{align*}
\]

(13)

where, we assume further that, for all \( i = 1, \ldots, l \), there hold

(I) \( A_i \in \mathbb{R}^{m_i \times n} \) are nonzero matrices,

(II) \( S_i : \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{m_i} \) are cutter operators with \( \mathcal{R}(A_i) \cap \text{Fix} S_i \neq \emptyset \), and

(III) \( h_i : \mathbb{R}^{m_i} \rightarrow \mathbb{R} \) is a convex function.

In this section, we assume that the solution set of (13) is denoted by \( \Omega \) and assume that it is nonempty.

Denote the product of spaces

\( \mathbb{R}^m := \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_l} \)

equipped with the addition

\( x + y := (x_1 + y_1, x_2 + y_2, \ldots, x_l + y_l) \),

the scalar multiplication

\( ax := (ax_1, ax_2, \ldots, ax_l) \)

with the inner product defined by

\( \langle \langle x, y \rangle \rangle_{\mathbb{R}^m} := \sum_{i=1}^{l} \langle x_i, y_i \rangle_{\mathbb{R}^{m_i}} \),

and the norm by

\( \|x\|_{\mathbb{R}^m} = \sqrt{\langle \langle x, x \rangle \rangle_{\mathbb{R}^m}} \),

for all \( x = (x_1, x_2, \ldots, x_l) \), \( y = (y_1, y_2, \ldots, y_l) \in \mathbb{R}^m \), is again a Euclidean space (see [24], Example 2.1).

Define a matrix \( A \in \mathbb{R}^{n \times m} \) by

\( A := [A_1^\top | A_2^\top | \ldots | A_l^\top]^\top \),

and an operator \( S : \mathbb{R}^m \rightarrow \mathbb{R}^m \) by

\( S(y) := (S_1y_1, S_2y_2, \ldots, S_ly_l) \),

for all \( y = (y_1, y_2, \ldots, y_l) \in \mathbb{R}^m \). Note that the operator \( S \) is cutter with

\( \text{Fix}(S) = \text{Fix}(S_1) \times \cdots \times \text{Fix}(S_l) \).
Furthermore, defining a function \( h : \mathbb{R}^m \rightarrow \mathbb{R} \) by
\[
h(x) := \sum_{i=1}^{l} h_i(x_i),
\]
for all \( x = (x_1, x_2, \ldots, x_l) \in \mathbb{R}^m \), we also have that the function \( h \) is a convex function (see [24], Proposition 8.25). By the above setting, we can rewrite the problem (13) as
\[
\begin{align*}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad x \in X \cap \text{Fix } T, Ax \in \text{Fix } S,
\end{align*}
\]
which is nothing else than Problem 1.

Here, to investigate the solving of the problem (13), we state the Algorithm 2 as follow.

**Algorithm 2:** Distributed Fixed-Point Subgradient Splitting Method.

**Initialization:** The positive sequence \( \{a_k\}_{k \geq 1} \) and the parameter \( \gamma \in (0, +\infty) \), and an arbitrary \( x_1 \in \mathbb{R}^m \).

**Iterative Step:** For given \( x_k \in \mathbb{R}^m \), compute
\[
\begin{align*}
z_{k,i} & := S_i A_i x_k - \frac{a_k}{\gamma} d_{k,i}, d_{k,i} \in \partial h_i(S_i A_i x_k), \text{ for all } i = 1, \ldots, l, \\
y_k & := x_k + \gamma \sum_{i=1}^{l} A_i^\top (z_{k,i} - A_i x_k), \\
x_{k+1} & := P_X(Ty_k - a_k f'(Ty_k)) + f'(Ty_k) \in \partial f(Ty_k).
\end{align*}
\]

As an above consequence, we note that
\[
A^\top x = \sum_{i=1}^{l} A_i^\top x_i,
\]
for all \( x = (x_1, x_2, \ldots, x_l) \in \mathbb{R}^m \). Furthermore, we know that
\[
\partial h(SAx_k) = \partial h_1(S_1 A_1 x_k) \times \cdots \times \partial h_l(S_l A_l x_k),
\]
see ([25], Corollary 2.4.5). Putting \( d_k := (d_{k,1}, \ldots, d_{k,l}) \) where \( d_{k,i} \in \partial h_i(S_i A_i x_k), \) for all \( k \geq 1 \), we obtain that
\[
z_k := (z_{k,1}, \ldots, z_{k,m}) = S Ax_k - a_k d_k
\]
for all \( k \geq 1 \). Notice that
\[
A^\top (z_k - Ax_k) = \sum_{i=1}^{l} A_i^\top (z_{k,i} - A_i x_k)
\]
for all \( k \geq 1 \). Thus, Algorithm 2 can be rewrite as
\[
\begin{align*}
z_k & = S Ax_k - \frac{a_k}{\gamma} d_k, \quad d_k \in \partial h(SAx_k) \\
y_k & = x_k + \gamma A^\top (z_k - Ax_k), \\
x_{k+1} & = P_X(Ty_k - a_k f'(Ty_k)) + f'(Ty_k) \in \partial f(Ty_k).
\end{align*}
\]
for all \( k \geq 1 \). Since \( \|A\|^2 \leq \sum_{i=1}^{l} \|A_i\|^2 \), the convergence result therefore follows from Theorem 1 and can be stated as the following corollary.
Corollary 1. Let \( \{x_k\}_{k \geq 1} \) be a sequence generated by Algorithm 2. If the following control conditions hold:

(i) \( 0 < \gamma < \frac{1}{\sum_{i=1}^{\infty} \|A_i\|^2} \);
(ii) \( \sum_{k=1}^{\infty} \alpha_k = +\infty \) and \( \sum_{k=1}^{\infty} \alpha_k^2 < +\infty \);

then

\[
\liminf_{k \to +\infty} \left( f(Ty_k) + \sum_{i=1}^{l} h_i(S_iA_ix_k) \right) \leq \left( f + \sum_{i=1}^{l} h_i \circ A_i \right)^* .
\]

5. Related Problems

The aim of this section is to show that Algorithm 1 and their convergence results can be employed when solving some well-known problems. Furthermore, we also present some noticeable applications relate to Algorithm 1 and its convergence results. For simplicity, we assume here that \( S = I \).

5.1. Convex Minimization with Least Square Constraint

Let us discuss the composite minimization problem over the set of all minimizers of the proximity function of a system of linear equations:

\[
\begin{aligned}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad x \in X \cap \text{argmin} \ g := \frac{1}{2} \|B \cdot -b\|^2,
\end{aligned}
\]

where \( B \) is a nonzero \( p \times n \) matrix and \( b \) is an \( p \times 1 \) vector. Now, we define the Landweber operator by

\[
L(x) := x - \frac{1}{\|B\|^2} B^T (Bx - b) .
\]

Suppose that \( B \) is nonzero, we have \( \|B\|^2 = \lambda_{\text{max}} (B^T B) \neq 0 \), which yields that the Landweber operator \( L \) is well defined. Furthermore, if \( \text{argmin} \ g \neq \emptyset \), then it holds that \( L \) is FNE with \( \text{Fix} \ L = \text{argmin} \ g [8] \) (Lemma 4.6.2, Theorem 4.6.3). In view of \( T := L \), the problem (14) is a special case of Problem 1 so that the problem (14) can be solved by Algorithm 1. Furthermore, if the set \( \{u \in \mathbb{R}^n : Bu = b\} \neq \emptyset \), the problem (14) is equivalent to

\[
\begin{aligned}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad Bx = b \\
& \quad x \in X.
\end{aligned}
\]

5.2. Convex Minimization with Nonsmooth Functional Constraints

Let us discuss a convex minimization with nonsmooth functional constraints

\[
\begin{aligned}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad g_i(x) \leq 0, i = 1, \ldots, l \\
& \quad x \in X,
\end{aligned}
\]

where \( g_i : \mathbb{R}^n \to \mathbb{R} \) are convex function, \( i = 1, \ldots, l \). Denote

\[
\Xi(g, 0) := \{x \in \mathbb{R}^n : g_i(x) \leq 0\}
\]

and assume that

\[
\bigcap_{i=1}^{l} \Xi(g, 0) \neq \emptyset .
\]
Furthermore, define the subgradient projection related to $g_i$ by

$$P_i x := \begin{cases} x - \frac{g_i(x)}{\|g_i(x)\|} g'_i(x), & \text{if } g_i(x) > 0 \\ x, & \text{otherwise,} \end{cases}$$

where $g'_i(x)$ is a (fixed) subgradient of $g_i$ at $x \in \mathbb{R}^n$. It is well known that $P_i$ is a cutter (a 1-SQNE operator) with $\text{Fix } P_i = \Xi(g_i, 0)$. Note that, if $\bigcap_{i=1}^l \Xi(g_i, 0) \neq \emptyset$, the cyclic subgradient projection operator $P := P_m P_{m-1} ... P_1$, which is a composition of SQNE operators $P_i$, is SQNE with $\text{Fix } P = \bigcap_{i=1}^l \Xi(g_i, 0)$.

(Theorem 2.1.50, [8]). In 2018, Cegielski and Nimana [26] proposed the following extrapolated operator,

$$P_{\lambda, \sigma}(x) = x + \lambda \sigma(x)(P x - x),$$

where $\lambda \in (0, 2)$ and $\sigma : \mathbb{R}^n \to \mathbb{R}$ is a step size function defined by

$$\sigma(x) = \sigma_{\max}(x) := \begin{cases} \sum_{i=1}^p \langle P x - U_i x, U_i x - U_{i-1} x \rangle \|P x - x\|^2, & \text{for } x \notin C, \\ 1, & \text{otherwise,} \end{cases}$$

where $U_i := P_i P_{i-1} ... P_1$ for $i = 1, 2, ..., m$ and $U_0 := I$. Note that the operator $P_{\lambda, \sigma}$ is SQNE with $\text{Fix } P_{\lambda, \sigma} = \text{Fix } P \neq \emptyset$ ([26], Theorem 3.2).

By means of $T := P_{\lambda, \sigma}$ or $T := P$, the problem (15) is nothing else than Problem 1 and Algorithm 1 is applicable for the problem (15).

5.3. Convex Minimization with Complex Constraints

Let $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued operator. We denote by

$$\text{Gr}(B) := \{(x, u) \in \mathbb{R}^n \times \mathbb{R}^n : u \in B x\}$$

its graph, and

$$\text{zer}(B) := \{z \in \mathbb{R}^n : 0 \in B(z)\}$$

the set of all zeros of the operator $B$. The set-valued operator $B$ is said to be monotone if

$$\langle x - y, u - v \rangle \geq 0,$$

for all $(x, u), (y, v) \in \text{Gr}(B)$, and it is called maximally monotone if its graph is not properly contained in the graph of any other monotone operators. For a set-valued operator $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, we define the resolvent of $B$, $J_B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, by

$$J_B := (\text{Id} + B)^{-1}.$$ 

It is well known that if $B$ is maximally monotone and $r > 0$, then the resolvent of $rB$ is (single-valued) FNE with

$$\text{zer}(B) = \text{Fix } J_B.$$
see ([24], Corollary 23.31, Proposition 23.38). Now, let us consider the minimal norm-like solution of the classical monotone inclusion problem:

\[
\begin{align*}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad x \in X \cap \text{zer}(B),
\end{align*}
\]

(16)

where \( B: \mathbb{R}^n \rightharpoonup \mathbb{R}^n \) is a maximally monotone operator such that \( \text{zer}(B) \neq \emptyset \). For given \( r > 0 \), and putting \( T := J_B \), the problem (15) is nothing else than Problem 1 and Algorithm 1 is also applicable for the problem (16).

In particular, let us consider the following minimal norm-like solution of minimization problem,

\[
\begin{align*}
\text{minimize} & \quad f(x) + h(Ax) \\
\text{subject to} & \quad x \in X \cap \text{argmin } \varphi.
\end{align*}
\]

(17)

The problem (17) has been considered by many authors, for instance [2,27–30] and references therein. Recall that for given \( r > 0 \) and a proper convex lower semicontinuous function \( \varphi: \mathbb{R}^n \to (-\infty, +\infty] \), we denote by \( \text{prox}_{r\varphi}(x) \) the proximal point of parameter \( r \) of \( \varphi \) at \( x \), which is the unique optimal solution of the optimization problem

\[
\min \left\{ \varphi(u) + \frac{1}{2r}\|u - x\|^2 : u \in \mathbb{R}^n \right\}.
\]

Note that \( \text{prox}_{r\varphi} = J_{r\partial \varphi} \). Therefore, putting \( T := \text{prox}_{r\varphi} \), Algorithm 1 is also applicable for the problem (17).

6. Numerical Experiments

In this section, to demonstrate the effectiveness of the fixed-point subgradient splitting method (Algorithm 1), we apply the proposed method to solve the fused lasso like problem. All the experiments were performed under MATLAB 9.6 (R2019a) running on a MacBook Pro 13-inch, 2019 with a 2.4 GHz Intel Core i5 processor and 8 GB 2133 MHz LPDDR3 memory.

For a given design matrix \( A := [a_1 | \cdots | a_r] \top \in \mathbb{R}^{r \times s} \) where \( a_i = (a_{i1}, \ldots, a_{is}) \in \mathbb{R}^s \) and a response vector \( b = (b_1, \ldots, b_r) \in \mathbb{R}^r \). We consider the fused lasso like problem of the form

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1 + \|Dx\|_1 \\
\text{subject to} & \quad x \in [-1, 1]^s \cap \text{argmin } \frac{1}{2}\|A \cdot -b\|^2,
\end{align*}
\]

(18)

where

\[
D := \begin{bmatrix}
-1 & 1 & 0 & \cdots & 0 & 0 \\
0 & -1 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -1 & 1
\end{bmatrix} \in \mathbb{R}^{(s-1) \times s}
\]

Observe that by setting the functions \( f(x) = \|x\|_1, h(x) := \|x\|_1, A := D \), the Landweber operator

\[
T(x) := x - \frac{1}{\|A\|^2} A \top (Ax - b),
\]

\( S := Id \); the identity matrix; and the constrained set \( X := [-1, 1]^s \), we obtain that the problem (18) is a special case of Problem 1 so that the problem (14) can be solved by Algorithm 1 (See Section 5.1) for more details.

We generate the matrix \( A \) with normally distributed random chosen in \((-10, 10)\) with a given percentage \( p_A \) of nonzero elements. We generate vectors \( b = (b_1, \ldots, b_r) \in \mathbb{R}^r \) corresponding to \( A \)
by the linear model \( b = A x_0 + \epsilon \), where \( \epsilon \sim N(0, \|A x_0\|^2) \) and the vector \( x_0 \) has 10% of nonzero components with normally distributed random generating. The initial point is a vector whose coordinates are chosen randomly in \((-1, 1)\). In the numerical experiment, denoting the estimate 

\[
F_k := f(T y_k) + h(S A x_k)
\]

for all \( k \geq 1 \), we consider the behavior of the average of the relative changes 

\[
\frac{1}{k} \sum_{j=1}^{k} \frac{|F_{j+1} - F_j|}{|F_j| + 1}
\]

with the optimality tolerance \( 10^{-3} \). We performed 10 independent tests for any collection of dimensions \((r, s)\) and the percentages of nonzero elements of \( A \) for various step size parameters \( \alpha_k \). The results are shown in Table 1, where the average number of iterations (\#Iters) and average CPU time (Time) to reach the optimality tolerance for any collection of parameters are presented.

| \( \alpha_k \rightarrow \) | \( p_A \) | 0.1/k | 0.3/k | 0.5/k | 0.7/k | 0.9/k |
|---------------------|-------|-------|-------|-------|-------|-------|
| (20, 50) 10%        | 54,253| 87,529| 87,416| 87,347| 87,455| 13.45 |
|                     | 58,085| 71,969| 71,813| 71,831| 71,928| 9.85  |
|                     | 55,414| 63,113| 62,962| 62,914| 62,969| 8.10  |
|                     | 56,190| 57,780| 57,459| 57,170| 56,949| 6.81  |
| (100, 200) 10%      | 111,358| 137,970| 137,099| 137,197| 137,347| 42.04 |
|                     | 108,019| 128,821| 128,621| 128,536| 128,496| 34.51 |
|                     | 108,760| 115,965| 116,965| 116,818| 116,751| 29.01 |
|                     | 108,409| 105,886| 105,621| 105,529| 105,726| 24.94 |
|                     | 107,967| 104,084| 104,082| 104,582| 104,519| 24.49 |

We see that the method performed with parameter \( \alpha_k = 0.1/k \) behaves significantly better than other in the sense of the average number of iterations as well as the averaged CPU time for all dimensions and percentages of nonzero elements of \( A \). Moreover, in the case \( p_A = 10\% \), we observed much bigger number of the averaged iterations and CPU time for the choice of the combinations of parameters \( \alpha_k = 0.3/k, 0.5/k, 0.7/k \) and \( 0.9/k \) with all the problem sizes.

7. Conclusions

In this paper, we introduced a simple fixed-point subgradient splitting method, whose main feature is the combination of the subgradient method with the hybrid steepest descent method relating to a nonlinear operator. We performed the convergence analysis of the method by proving the convergence of the function values to the minimum value. The result is obtained by adopting some specific suitable assumptions on the step sizes. We discussed in detail the applications of the proposed scheme to some remarkable problems in the literature. Numerical experiments on fused lasso like problem show evidence of the performance of our work.

**Funding:** This work was financial supported by the young researcher development project of Khon Kaen University.

**Acknowledgments:** The author is thankful to two anonymous reviewers and the Assistant Editor for comments and remarks which improved the quality of the paper. The author also thanks Satit Saejung for his suggestions under the young researcher development project of Khon Kaen University.
Conflicts of Interest: The author declares no conflicts of interest.

References
1. Tikhonov, A.N.; Arsenin, V.Y. Solution of Ill-Posed Problems; V.H. Winston: Washington, DC, USA, 1977.
2. Beck, A.; Sabach, S. A first order method for finding minimal norm-like solutions of convex optimization problems. *Math. Program.* 2014, 147, 25–46. [CrossRef]
3. Ono, S.; Yamada, I. Hierarchical convex optimization with primal-dual splitting. *IEEE Trans. Signal Proc.* 2015, 63, 373–388. [CrossRef]
4. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B* 2015, 67, 91–108. [CrossRef]
5. Condat, L. A direct algorithm for 1d total variation denoising. *IEEE Signal Process. Lett.* 2013, 20, 1054–1057. [CrossRef]
6. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 2005, 67, 301–320. [CrossRef]
7. Starck, J.L.; Murtagh, F.; Fadili J.M. *Sparse Image and Signal Processing. Wavelets, Curvelets. Morphological Diversity*; Cambridge University Press: Cambridge, MA, USA, 2010.
8. Cegielski, A. *Iterative Methods for Fixed Point Problems in Hilbert Spaces*; Lecture Notes in Mathematics; Springer: Heidelberg, Germany, 2012; Volume 2057.
9. Cegielski, A. General method for solving the split common fixed point problem. *J. Optim. Theory Appl.* 2015, 165, 385–404. [CrossRef]
10. Cegielski, A. Application of quasi-nonexpansive operators to an iterative method for variational inequality. *SIAM J. Optim.* 2015, 25, 2165–2181. [CrossRef]
11. Yamada, I.; Ogura, N. Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings. *Numer. Funct. Anal. Optimiz.* 2004, 25, 619–655. [CrossRef]
12. Yamada, I. The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In *Inherently Parallel Algorithm for Feasibility and Optimization*; Butnariu, D., Censor, Y., Reich, S., Eds.; Elsevier: New York, NY, USA, 2001; pp. 473–504.
13. Iiduka, H. Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings. *Math. Program.* 2016, 159, 509–538. [CrossRef]
14. Iiduka, H. Proximal point algorithms for nonsmooth convex optimization with fixed point constraints. *Euro. J. Oper. Res.* 2016, 25, 503–513. [CrossRef]
15. Censor, Y.; Elfving, T. A multiprojection algorithm using Bregman projections in product space. *Numer. Algor.* 1994, 8, 221–239. [CrossRef]
16. Censor, Y.; Segal, A. The split common fixed-point problem for directed operators. *J. Convex Anal.* 2009, 16, 587–600.
17. Byrne, C. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Probl.* 2002, 18, 441–453. [CrossRef]
18. Censor, Y.; Bortfeld, T.; Martin, B.; Trofimov, A. A unified approach for inversion problems in intensity modulated radiation therapy. *Phys. Med. Biol.* 2006, 51, 2353–2365. [CrossRef]
19. Ansari, Q.H.; Nimana, N.; Petrot, N. Split hierarchical variational inequality problems and related problems. *Fixed Point Theory Appl.* 2014, 2014, 208. [CrossRef]
20. Nimana, N.; Petrot, N. Subgradient algorithm for split hierarchical optimization problem. *Carpathian J. Math.* 2018, 34, 391–399.
21. Kraikaew, R.; Saengjung, S. On split common fixed point problems. *J. Math. Anal. Appl.* 2014, 415, 513–524. [CrossRef]
22. Moudafi, A. The split common fixed-point problem for demicontractive mappings. *Inverse Probl.* 2010, 26, 055007. [CrossRef]
23. Beck, A. *First-Order Methods in Optimization*; MOS-SIAM Series on Optimization; SIAM: Philadelphia, PA, USA, 2017.
24. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; Springer: New York, NY, USA, 2012.
25. Z˘ alinescu, C. *Convex Analysis in General Vector Spaces*; World Scientific: Singapore, 2002.
26. Cegielski, A.; Nimana, N. Extrapolated cyclic subgradient projection methods for the convex feasibility problems and their numerical behaviour. *Optimization* 2019, 68, 145–161. [CrossRef]

27. Attouch, H.; Czarnecki, M.-O.; Peypouquet, J. Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities. *SIAM J. Optim.* 2011, 21, 1251–1274. [CrossRef]

28. Boţ, R.I.; Csetnek, E.R.; Nimana, N. An inertial proximal-gradient penalization scheme for constrained convex optimization problems. *Vietnam J. Math.* 2018, 46, 53–71. [CrossRef]

29. Noun, N.; Peypouquet, J. Forward-backward penalty scheme for constrained convex minimization without inf-compactness. *J. Optim. Theory Appl.* 2013, 158, 787–795. [CrossRef]

30. Peypouquet, J. Coupling the gradient method with a general exterior penalization scheme for convex minimization. *J. Optim. Theory Appl.* 2012, 153, 123–138. [CrossRef]