DATA-COMPATIBILITY OF ALGORITHMS FOR CONSTRAINED CONVEX OPTIMIZATION

YAIR CENSOR[1,*], MAROUN ZAKNOON[2], ALEXANDER J. ZASLAVSKI[3]

[1]*Department of Mathematics, University of Haifa, Mt. Carmel, Haifa 3498838, Israel*
[2]*Department of Mathematics, The Arab Academic College for Education,*
22 HaHashmal Street, Haifa 32623, Israel
[3]*Department of Mathematics, The Technion – Israel Institute of Technology,*
Technion City, Haifa 3200003, Israel

Abstract. The data-compatibility approach to constrained optimization, proposed here, strives to a point that is "close enough" to the solution set and whose target function value is "close enough" to the constrained minimum value. These notions can replace analysis of asymptotic convergence to a solution point of infinite sequences generated by specific algorithms. We consider a problem of minimizing a convex function over the intersection of the fixed point sets of nonexpansive mappings and demostrate the data-compatibility of the Hybrid Subgradient Method (HSM). A string-averaging HSM is obtained as a by-product and relevance to the minimization over disjoint hard and soft constraints sets is discussed.

Keywords. Data-compatiblity; Constrained minimization; Feasibility-seeking; Hybrid subgradient method; Proximity function.

1. INTRODUCTION

The data of a constrained minimization problem $\min\{f(x) \mid x \in C\}$ consists of a target function f and a constraints set C. For this problem to be meaningful, C needs to be nonempty, and for asymptotic convergence analysis of an algorithm for solving the problem one commonly needs that the solution set of the problem be nonempty, i.e., that there exists at least one point, say x^*, in C with the property that $f(x^*) \leq f(x)$ for all $x \in C$.

In real-world practical situations these nonemptiness assumptions cannot always be guaranteed or verified. To cope with this we define the notion of *data-compatibility* in a Hilbert space. Such data-compatibility is a finite, not an asymptotic, notion. Even when the sets C and the solution set of the constrained minimization problem are nonempty, striving for data-compatibility is a worthwhile aim because it can be "reached", contrary to asymptotic limit points.

Data-compatibility of a point x means that it, simultaneously, (i) is "close enough" to the set of minimizers of f over the constraints C, and (ii) has a function value $f(x)$ "close enough" to

---

the minimum of f over that set of constraints. Once we precisely define these notions the question arises whether or not it is possible to guarantee that, under certain conditions, compatibility with a data pair (C, f) can be reached by an iterative process designed to solve a constrained optimization problem?

The advantage of these data-compatibility notions is that they can cater better to practical situations. On the theoretical side, we propose that instead of proving asymptotic convergence of iterative processes and afterwards studying the effects of various stopping rules, it is possible to directly formulate data-compatibility and provably guarantee that it can be reached by an iterative process.

We demonstrate our principle and approach in a specific scenario. The problem formulation and the algorithm that we use in our demonstration serve only as vehicles to present our data-compatibility approach which we believe is novel. To do so we study the behavior of the specific iterative process for convex minimization over the intersection of the fixed point sets of nonexpansive operators, called the Hybrid Subgradient Method (HSM). We specify conditions under which the iterative HSM process generates solutions that are compatible with the data pair $(\mathrm{Fix}(T, f)$.

Delete extra spaces (if they exist) between (T) and the comma, and between f and the ")".

Minimization over the intersection of the fixed point sets of nonexpansive operators has been treated extensively in the literature, of which we reference a few works below. But in all these earlier works the asymptotic convergence of algorithms, under various sets of conditions, is the central theme, not data-compatibility.

As an important special case of the general algorithmic formulation we discuss a string-averaging algorithmic scheme. The string-averaging algorithmic notion has a quite general structure in itself. Invented in [2] and spurred many extensions and applications since then, e.g., [1, 3] and the book [4], it works in general as follows. From a current iteration point, it performs consecutively specified iterative algorithmic steps "along" different "strings" of individual constraints sets and then takes a combination, convex or other, of the strings' end-points as the next iterate. The string-averaging algorithmic scheme gives rise to a variety of specific algorithms by judiciously choosing the number of strings, their assignments and the nature of the combination of the strings' end-points. Details are given in the sequel.

Earlier works on minimizing convex functions over the intersection of the fixed point sets of nonexpansive mappings are all based on asymptotic convergence of the algorithms and investigate the problem and prove convergence of algorithms under various conditions. These include, to name but a few, the papers of Iiduka [5], [6], [7], the work of Maingé [8], and publications by Hayashi and Iiduka [9] and Deutsch and Yamada [10]. Methods similar to HSM were studied by several researchers, e.g., Shor [11], Albert, Iusem and Solodov [12], Yamada and Ogura [13], Hirstoaga [14], Martinez-Yanes and Xu [15], Maingé [16], Aoyama and Kohsaka [17], Cegielski [18], and Bello Cruz [19].

In contrast with these, our main contribution is in the novel quest for data-compatibility instead of asymptotic convergence properties.

The term HSM is in analogy with the established term, coined by Yamada [20], of the Hybrid Steepest Descent Method (HSDM). The structural similarity of the HSM with the HSDM is that the former uses subgradient steps instead of the steepest descent steps used by the latter.

The paper is organized as follows. In Section 2, we define the notion of data-compatibility of a point with the data of a constrained minimization problem, and, in particular in Subsection 2.3 we discuss the problem of guaranteeing a priori data-compatibility. In Section 3, we present the origin and motivation of our work. The problem of minimizing a convex function over the intersection of the fixed point sets of nonexpansive mappings is defined in Section 4 along with the Hybrid Subgradient Method (HSM) for its solution. Inexact iterates are discussed in Section 5 followed in Section 6 by work on the main result that proves the ability of the HSM to generate a data-compatible point for the problem. We present the string-averaging variant of the HSM in Section 7. In Section 8, we consider a specific situation wherein the data of a constrained minimization problem does not necessarily obey feasibility of the constraints, i.e., does not demand that $C = \cap_{i=1}^{m} C_i$ is nonempty. Finally, in Section 9 we describe the minimization over disjoint hard and soft constraints sets problem and its relation to the work presented in this paper.

2. DATA-COMPATIBILITY

In this section we define the notion of *data-compatibility of a point with the data of a constrained minimization problem.* Let $\Omega \subseteq H$ be a given nonempty set in the Hilbert space H and let there be given, for $i = 1, 2, \ldots, m$, nonempty sets $C_i \subseteq \Omega$. We denote by $\Gamma := \{C_i\}_{i=1}^{m}$ the family of sets and refer to it as the "constraints data Γ".

We introduce a set Δ such that $\Omega \subseteq \Delta \subseteq H$ and assume that we are given a function $f : \Delta \to R$ which is referred to as "the target function f" or, in short, the data f. A pair (Γ, f) is referred to as the "data pair (Γ, f)".

2.1. **Data-compatibility for constraints.** Constraint modelling has a prominent role in operations research and is used in a wide range of industrial projects, such as, but by far not only, the control of an intelligent interface linking computer aided design and automatic inspection systems, the identification of manufacturing errors from inspection results and the design synthesis and analysis of mechanisms, to name a few, see, e.g., [21]. In our language, it is the modelling of real-world problems via *convex feasibility problems* (CFPs). Given a finite family of, commonly convex, sets $\Gamma := \{C_i\}_{i=1}^{m}$ the CFP is to find a point $x^* \in C := \cap_{i=1}^{m} C_i$. This approach has a long history, see, e.g., Combettes' or Bauschke and Borwein's corner stone reviews [22], [23], respectively, Cegielski's book [24] and Bauschke and Combettes' book [25]. First we look at compatibility with the constraints data alone. For this we need an appropriate *proximity function* that "measures" how incompatible an $x \in \Omega$ is with the constraints of Γ. There is no common definition in the literature, but a proximity function $\text{Prox}_\Gamma : \Omega \to R_+$ (the nonnegative orthant) should have the property that $\text{Prox}_\Gamma(x) = 0$ if and only if $x \in C := \cap_{i=1}^{m} C_i$. By evaluating/measuring a "distance" to the constraints, the lower the value of $\text{Prox}_\Gamma(x)$ is – the less incompatible x is with the constraints.

A proximity function does not require that $C \neq \varnothing$ and it is a useful tool, particularly for situations when $C \neq \varnothing$ does not hold, or cannot be verified. An enlightening discussion of proximity functions for the convex feasibility problem can be found in Cegielski's book [24, Subsection 1.3.4]. An important and often used proximity function is

$$\text{Prox}_\Gamma(x) := \frac{1}{2} \sum_{i=1}^{m} w_i \|P_{C_i}(x) - x\|^2, \tag{2.1}$$

where $P_{C_i}(x)$ is the orthogonal (metric) projection onto C_i and $\{w_i\}_{i=1}^m$ is a set of weights such that $w_i \geq 0$ and $\sum_{i=1}^m w_i = 1$. With a proximity function at hand we define compatibility with constraints as follows.

Definition 2.1. γ-compatibility with constraints data Γ. Given constraints data Γ, a proximity function Prox_Γ, and a real number $\gamma \geq 0$, we say that a point $x \in \Omega$ is "γ-compatible with Γ" if $\mathrm{Prox}_\Gamma(x) \leq \gamma$. We define the set of all points that are γ-compatible with Γ by $\Pi(\Gamma, \gamma) := \{x \in \Omega \mid \mathrm{Prox}_\Gamma(x) \leq \gamma\}$ and call it the (Γ, γ)-compatibility set.

Definition 2.2. The γ-output of a sequence. Given constraints data Γ, a proximity function Prox_Γ, a real number $\gamma \geq 0$ and a sequence $R := \{x^k\}_{k=0}^\infty$ of points in Ω, we use $O(\Gamma, \gamma, R)$ to denote the point $x \in \Omega$ that has the following properties: $\mathrm{Prox}_\Gamma(x) \leq \gamma$, and there is a nonnegative integer K such that $x^K = x$ and, for all nonnegative integers $k < K$, $\mathrm{Prox}_\Gamma(x) > \gamma$. If there is such an x, then it is unique. If there is no such x, then we say that $O(\Gamma, \gamma, R)$ is *undefined*, otherwise it is *defined*.

If R is an infinite sequence generated by a certain process then $O(\Gamma, \gamma, R)$ is the *output* produced by that process when we add to it instructions that make it terminate as soon as it reaches a point that is γ-compatible with Γ.

The (Γ, γ)-compatibility set $\Pi(\Gamma, \gamma) \subseteq \Omega$ need not be nonempty for all γ. If, however, $\Pi(\Gamma, 0) \neq \varnothing$ then $\Pi(\Gamma, 0) = C$. We have used the notion of γ-compatibility with constraints earlier in our work on the superiorization method, see, e.g., [26]. The unconstrained minimization of $\mathrm{Prox}_\Gamma(x)$ always yields an infimum of $\mathrm{Prox}_\Gamma(x)$ but this does not guarantee that $\Pi(\Gamma, \gamma)$ is nonempty. However, $\Pi(\Gamma, \gamma)$ can be nonempty even if the constraints intersection $C = \cap_{i=1}^m C_i$ is empty. In order to proceed with data-compatibility for constrained minimization in the next subsection we require that $\Pi(\Gamma, \gamma) \neq \varnothing$.

It should be emphasized that a sequence considered in Definition 2.2 need not be convergent, or can be convergent but not necessarily to a point in $\Pi(\Gamma, \gamma)$, and still yield an output $O(\Gamma, \gamma, R)$ that is γ-compatible with Γ.

2.2. **Data-compatibility for constrained minimization.** For a γ, for which $\Pi(\Gamma, \gamma) \neq \varnothing$, we define the set of minimizers of f over $\Pi(\Gamma, \gamma)$,

$$S(f, \Pi(\Gamma, \gamma)) := \{x \in \Pi(\Gamma, \gamma) \mid f(x) \leq f(y), \text{ for all } y \in \Pi(\Gamma, \gamma)\}. \tag{2.2}$$

If f is the zero function or if $f = $ constant then $S(f, \Pi(\Gamma, \gamma)) = \Pi(\Gamma, \gamma)$. We use the distance function between a point x and a set S defined as

$$d(x, S) := \inf\{d(x, y) \mid y \in S\} \tag{2.3}$$

where $d(x, y)$ is the distance between points x and y. Next we propose our definition of compatibility with a data pair (Γ, f).

Definition 2.3. $(\tau, \bar{L})$-compatibility with a data pair (Γ, f). Given a $\tau \geq 0$, and a real number $\bar{L} > 0$, we say that a point $x \in \Omega$ is "$(\tau, \bar{L})$-compatible with the data pair (Γ, f)" if $S(f, \Pi(\Gamma, \gamma)) \neq \varnothing$ and the following two conditions hold

$$d(x, S(f, \Pi(\Gamma, \gamma))) \leq \tau \tag{2.4}$$

and

$$f(x) \leq f(z) + \tau \bar{L}, \text{ for all } z \in S(f, \Pi(\Gamma, \gamma)), \tag{2.5}$$

where the constraints data Γ, a proximity function Prox_Γ, a target function f, and a $\gamma \geq 0$ such that $\Pi(\Gamma, \gamma) \neq \varnothing$, are given.

This means that such a point x, simultaneously, (i) is "close enough" to the set of minimizers of f over a (Γ, γ)-compatibility set of the constraints, and (ii) has a function value $f(x)$ "close enough" to the minimum of f over that (Γ, γ)-compatibility set of the constraints. This definition does not require nonemptiness of the intersection of the constraints $C = \cap_{i=1}^m C_i$ neither does it require that the constrained minimization problem $\min\{f(x) \mid x \in C\}$, has a nonempty solution set $\text{SOL}(f, C)$ which is defined by

$$\text{SOL}(f, C) := S(f, C) = \{x \in C \mid f(x) \leq f(y), \text{ for all } y \in C\}. \tag{2.6}$$

It relies on the weaker assumptions that $\Pi(\Gamma, \gamma) \neq \varnothing$ and $S(f, \Pi(\Gamma, \gamma)) \neq \varnothing$. Therefore, use of these notions makes it possible to deviate from the nonemptiness assumptions which usually lie at the heart of asymptotic analyses in optimization theory.

Example 2.1. Here is a motivating example from a specific real-world application that stands to benefit from the situation described above. A fully-discretized modeling approach to intensity-modulated radiation therapy (IMRT) treatment planning, e.g., [27], leads to a very large and sparse system of linear inequalities, see, e.g., [28, Equation 6.13] which is, in practice, further equipped with box constraints on the unknown vector x there. The sets C_i are half-spaces and the, commonly used, proximity function (2.1) above always has a minimum value g and its solution set is nonempty. This is guaranteed, e.g., by Proposition 7 of [29] with $\Omega = H$ and the box constraints serving the condition there that one of the sets must be bounded. Therefore, $\Pi(\Gamma, \gamma)$ will be nonempty for any γ for which $g \leq \gamma$.

Often, an exogeneous target function is imposed on these constraints in order to impose some prior knowledge. A commonly employed such function f is the "total variation" (TV), see, e.g., [30] that smoothes the solution vector, see, e.g., [31]. In this situation we know in advance that $S(f, \Pi(\Gamma, \gamma)) \neq \emptyset$ without having to invest preliminary computing resources in finding whether $C = \cap_{i=1}^m C_i$ is or is not empty. This example is also relevant to Theorems 7.3 and 8.2 below.

In the consistent case when $C \neq \varnothing$ and $\text{SOL}(f, C) \neq \varnothing$, i.e., $\gamma = 0$ and $\Pi(\Gamma, 0) = C$, Definition 2.3 takes the following form.

Definition 2.4. $(\tau, \bar{L})$**-compatibility with a data pair** (Γ, f) **in the consistent case**. Given consistent constraints data Γ via $C := \cap_{i=1}^m C_i \neq \varnothing$, a target function f, a $\tau \geq 0$, and a real number $\bar{L} > 0$, we say that a point $x \in \Omega$ is "$(\tau, \bar{L})$-compatible with the consistent data pair (Γ, f)" if $\text{SOL}(f, C) \neq \varnothing$ and the following two conditions hold

$$d(x, \text{SOL}(f, C)) \leq \tau \tag{2.7}$$

and

$$f(x) \leq f(z) + \tau \bar{L}, \text{ for all } z \in \text{SOL}(f, C). \tag{2.8}$$

2.3. **Data-compatibility instead of asymptotic convergence.** We consider in this paper the consistent case situation of Definition 2.4. Given a constrained minimization problem $\min\{f(x) \mid x \in C\}$ via its data pair (Γ, f), one traditional route in optimization theory is to design/construct

an iterative process for its solution and investigate the asymptotic convergence of the this process. After asymptotic convergence of the iterative process has been secured a common approach is to use the process as a basis for creating an algorithm. Such an algorithm will use the iteration formulae dictated by the process but have a user-chosen stopping rule attached to it. The, so obtained, algorithm will stop when the stopping rule is met and output an approximate solution to the original problem.

Various stopping rules are in use and we do not attempt to review them. But it is clear that the question "when to stop" has a profound influence on the practical output of an algorithmic run. With roots and research in the fields of statistics and optimization, "optimal stopping is concerned with the problem of choosing a time to take a given action based on sequentially observed random variables in order to maximize an expected payoff or to minimize an expected cost.", see [32].

One aspect of stopping rules is the question whether or not a particular stopping rule will actually cause the iterative process to which it is attached to stop and yield an output. If, for example, the stopping rule is to stop after a specified number of iterations then the algorithmic run will undoubtedly stop when this number is reached. On the other hand, if one considers a feasibility-seeking problem to find a point in the intersection of a given finite family of sets $C := \cap_{i=1}^m C_i$ then a feasibility-seeking iterative process that is proven to asymptotically converge to a point in C can be used. However, if one uses the stopping rule that iterations should stop when an iterate x^k is reached that belongs to the intersection C then the process might never stop.

Problem 2.1. The question that we ask ourselves here is whether or not it is possible to a priori guarantee that, under certain conditions, $(\tau, \bar{L})$-compatibility with a data pair (Γ, f) can be reached by an iterative process designed to solve a constrained optimization problem?

Obviously, if the considered iterative process is proven to asymptotically converge to a point in $\text{SOL}(f, C))$ then the answer is positive and (2.4)-(2.5) can be used as a stopping rule that will indeed yield a data-compatible output. On the other hand, if an algorithm generates $(\tau, \bar{L})$-compatible sequences for all values of τ then all the sequences generated by it are asymptotically convergent. Nevertheless, the notion of $(\tau, \bar{L})$-compatibility makes sense even if it holds for certain (not all) values of τ and in this case it can provide useful information. To approach this, inspired by [26, Definition 2.1], we suggest the following definition for the first iterate that satisfies both conditions (2.7)-(2.8).

Definition 2.5. The $(\tau, \bar{L})$-output of a sequence. Given constraints data Γ, a proximity function Prox_Γ, a target function f, a $\gamma \geq 0$, a $\tau \geq 0$, and a real number $\bar{L} > 0$, we consider an infinite sequence $\mathscr{X} = \{x^k\}_{k=0}^\infty$ of points in Ω. Let $\text{OUT}((\Gamma, f), \gamma, (\tau, \bar{L}), \mathscr{X})$ denote the point $x \in \Omega$ that fulfills (2.7)-(2.8) and such that there exists a nonnegative integer K such that $x^K = x$ and for all nonnegative integers $k < K$ at least one of the two conditions (2.4)-(2.5) is violated. If there is such an x, then it is unique. If there is no such x then we say that $\text{OUT}((\Gamma, f), \gamma, (\tau, \bar{L}), \mathscr{X})$ is undefined, otherwise it is defined.

An algorithm is a finite sequence of well-defined, computer-implementable instructions, typically to solve a class of problems or to perform a computation. However, it is common practice in the literature to use the term "algorithm" also for iterative processes that do not include a stopping rule.

If $\mathscr{X}$ is generated by an iterative process, then $\text{OUT}((\Gamma,f),\gamma,(\tau,\bar{L}),\mathscr{X})$ is the output produced by that process when we add to it instructions that make it terminate as soon as it reaches a point that is $(\tau,\bar{L})$-compatible with a data pair (Γ,f).

General results that will characterize, or give conditions for, an iterative process to be provably data-compatible are not yet known, but to initiate research in this direction we demonstrate our approach in a specific scenario.

We work in the framework of Definition 2.4 and study the behavior of an iterative process for convex minimization over fixed point sets of nonexpansive operators. Rather than generating infinite sequences that asymptotically converge to a point in $\text{SOL}(f,C)$, we specify conditions under which the iterative process generates solutions that are $(\tau,\bar{L})$-compatible with the data pair (Γ,f).

In the sequel (Section 8) we also discuss a specific situation wherein the data pair (Γ,f), with $\Gamma := \{C_i\}_{i=1}^m$ a family of closed and convex subsets of H, not necessarily obeys that $C = \cap_{i=1}^m C_i$ is nonempty.

3. ORIGIN AND MOTIVATION

The origin of the idea of data-compatibility is the work done in [33]. There the string-averaging projected subgradient method (SA-PSM) was developed and studied. The SA-PSM is a variant of the projected subgradient method for solving constrained minimization problems. It differs from the traditional projected subgradient method (PSM) by replacing, in each iteration, the single step of projecting onto the entire constraints set with projections onto the individual sets whose intersection is the feasible set of the minimization problem. This is an advantage, in the frequently occurring situations, when the individuals sets are easier to project on than their entire intersection.

The main result in [33, Theorem 9] provably guarantees the ability of the algorithm to reach an iterate of the generated sequence that is, up to predetermined bounds, close to a "solution". Here we formulate this notion and extend it to encompass also the case when the solution set of the problem is empty.

4. THE PROBLEM AND THE ALGORITHM

Our problem, iterative process, and main result are set in a Hilbert space. However, some intermediate auxiliary results are true, and have some independent value, in a general metric space. Let (X,ρ) be a complete metric space and let $T : X \longrightarrow X$ be an operator. The fixed point set of T is

$$\text{Fix}(T) := \{x \in X \mid T(x) = x\}. \tag{4.1}$$

An operator T is nonexpansive if it satisfies

$$\rho(T(x),T(y)) \leq \rho(x,y), \text{ for all } x,y \in X. \tag{4.2}$$

Given a nonempty set $E \subseteq X$ define the distance of a point $x \in X$ from it by

$$d(x,E) := \inf\{\rho(x,y) \mid y \in E\}. \tag{4.3}$$

We denote the ball with center at a given $x \in X$ and radius $r > 0$ by $B(x,r)$. The execution of the operator T for n times consecutively on an initial given point x is denoted by $T^n x$, and $T^0 x := x$.

For $X = H$ a Hilbert space, we look at a constrained minimization problem of the form

$$\min\{f(x) \mid x \in \text{Fix}(T)\} \tag{4.4}$$

where f is a convex target function from H into the reals and T is a given nonexpansive operator. Solving this problem means to

$$\text{find a point } x \text{ in } \text{SOL}(f, \text{Fix}(T)), \tag{4.5}$$

where

$$\text{SOL}(f, \text{Fix}(T)) := \{x \in \text{Fix}(T) \mid f(x) \leq f(y) \text{ for all } y \in \text{Fix}(T)\}. \tag{4.6}$$

For this task we employ an iterative Hybrid Subgradient Method (HSM) that uses the powers of the operator T combined with subgradient steps. We denote by $\partial f(x^k)$ the subgradient set of f at x^k.

Algorithm 4.1. Hybrid Subgradient Method (HSM).

Initialization: Let $\{\alpha_k\}_{k=0}^{\infty} \subset (0,1]$ be a scalar sequence and let $x^0 \in H$ be an arbitrary initialization vector.

Iterative step: Given a current iteration vector x^k calculate the next vector as follows: Choose any $s^k \in \partial f(x^k)$ and calculate

$$x^{k+1} = T\left(x^k - \alpha_k \frac{s^k}{\|s^k\|}\right), \tag{4.7}$$

but if $s^k = 0$ then set $\dfrac{s^k}{\|s^k\|} := 0$.

Note that if $s^k = 0$ then the algorithm simply calculates

$$x^{k+1} = T(x^k), \tag{4.8}$$

otherwise, it uses (4.7).

As mentioned above, our data-compatibility result, presented in Theorem 6.1 below, will specify conditions that guarantee the existence of an iterate, of any sequence generated by the HSM of Algorithm 4.1, that is $(\tau, \bar{L})$-compatible with the data pair $(\Gamma = Fix(T), f)$. I.e., that for every $\tau \in (0,1)$, and any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 4.1, there exists an integer K so that, for all $k \geq K$:

$$d(x^k, \text{SOL}(f, \text{Fix}(T))) \leq \tau \tag{4.9}$$

and

$$f(x^k) \leq f(z) + \tau\bar{L} \text{ for all } z \in \text{SOL}(f, \text{Fix}(T)) \tag{4.10}$$

where $\bar{L}$ is some well-defined constant.

5. INEXACT ITERATES

In this sections we establish some properties of sequences of the form $\left\{T^j y^0\right\}_{j=0}^{\infty}$, for any $y^0 \in X$, with "computational errors". These will serve as tools in proving the main result. In our work we need to focus on operators that have the property formulated in the next condition.

Condition 5.1. Let X be a complete metric space, assume that $T : X \rightarrow X$ is a nonexpansive operator such that $\lim_{j \rightarrow \infty} T^j y^0$ exists for any $y^0 \in X$.

Condition 5.1 holds for many nonexpansive mappings. In [34] it was shown that for several important classes of nonexpansive mappings this property is generic (typical) in the sense of the Baire category. This means that a class of mappings is equipped with an appropriate complete metric and it is shown the existence of a subset of the space of mappings which is a countable intersection of open everywhere dense sets such that any mappings from the intersection possesses the desirable convergence property.

Proposition 5.1. *Let X be a complete metric space, and assume that $T : X \to X$ is a nonexpansive operator. Then*

(1) The set $\mathrm{Fix}\,(T)$ is closed.
(2) If $\lim_{j \to \infty} T^j y^0$ exists for some $y^0 \in X$, then $\lim_{j \to \infty} T^j y^0$ is a fixed point of T and, consequently, $\mathrm{Fix}\,(T) \neq \varnothing$.

Proof. The proof is obvious. $\square$

Proposition 5.2. *Let X be a complete and compact metric space, assume that $T : X \to X$ is a nonexpansive operator and let $\varepsilon > 0$. Then there exists a $\delta > 0$ such that for each $x \in X$ satisfying $\rho(x, Tx) \leq \delta$ we have*

$$d(x, \mathrm{Fix}(T)) \leq \varepsilon. \tag{5.1}$$

Proof. Assume to the contrary that for each integer $k \geq 1$ there exists a point $x^k \in X$ such that

$$\rho(x^k, Tx^k) \leq k^{-1} \text{ and } d(x^k, \mathrm{Fix}(T)) > \varepsilon. \tag{5.2}$$

Since X is compact, extracting a subsequence and re-indexing, if necessary, we may assume without loss of generality that, the sequence $\{x^k\}_{k=1}^{\infty}$ so generated by the repeated use of the above negation, converges and denote

$$z := \lim_{k \to \infty} x^k. \tag{5.3}$$

Since T is nonexpansive, inequality (5.2) and the limit (5.3) yield, for all integers $k \geq 1$,

$$\begin{aligned} \rho(z, Tz) &\leq \rho(z, x^k) + \rho(x^k, Tx^k) + \rho(Tx^k, Tz) \\ &\leq 2\rho(z, x^k) + k^{-1} \to 0, \text{ as } k \to \infty, \end{aligned} \tag{5.4}$$

thus,

$$z \in Fix(T). \tag{5.5}$$

In view of (5.3), for all sufficiently large integers k,

$$d(x^k, \mathrm{Fix}(T)) \leq d(x^k, z) < \varepsilon. \tag{5.6}$$

This contradicts (5.2) and completes the proof. $\square$

Lemma 5.1. *Let X be a complete and compact metric space, assume that $T : X \to X$ is a nonexpansive operator for which Condition 5.1 holds, and let $\mu > 0$. Then there exists an integer k_1 such that for each $x \in X$ there exists $j \in \{0, 1, \ldots, k_1\}$ such that*

$$d(T^j x, \mathrm{Fix}(T)) \leq \mu. \tag{5.7}$$

Proof. Assume to the contrary that for each integer $k \geq 1$ there exists a point $x^k \in X$ such that

$$d(T^j x^k, \text{Fix}(T)) > \mu, \text{ for all } j = 0, 1, \ldots, k. \tag{5.8}$$

Since X is compact, extracting a subsequence and re-indexing, if necessary, we may assume without loss of generality that, the sequence $\{x^k\}_{k=1}^{\infty}$ so generated by the repeated use of the above negation, converges and let

$$z = \lim_{k \to \infty} x^k. \tag{5.9}$$

By Proposition 5.1 and Condition 5.1, we conclude that

$$d(T^j z, \text{Fix}(T)) < \mu. \tag{5.10}$$

By (5.9) and since T is nonexpansive, for all sufficiently large integers k,

$$d(T^j x^k, \text{Fix}(T)) < \mu, \tag{5.11}$$

contradicting (5.8), thus, concluding the proof. $\qquad \square$

Theorem 5.2. *Let X be a complete and compact metric space, assume that $T : X \to X$ is a nonexpansive operator for which Condition 5.1 holds, and let $\varepsilon > 0$. Then there exists a natural number k_0 such that for each $x \in X$ and each integer $k \geq k_0$,*

$$\rho(T^k x, \lim_{i \to \infty} T^i x) \leq \varepsilon. \tag{5.12}$$

Proof. By Lemma 5.1, there exists an integer k_0 such that for each $x \in X$ there exists $j \in \{0, 1, \ldots, k_0\}$ so that

$$d(T^j x, \text{Fix}(T)) < \varepsilon/2. \tag{5.13}$$

This implies that there exist a

$$j \in \{0, 1, \ldots, k_0\} \tag{5.14}$$

and a

$$z \in \text{Fix}(T) \tag{5.15}$$

such that

$$\rho(T^j x, z) < \varepsilon/2. \tag{5.16}$$

Since T is nonexpansive, we get, by (5.15) and (5.16), that for all integers $k \geq j$,

$$\rho(T^k x, z) \leq \rho(T^j x, z) < \varepsilon/2, \tag{5.17}$$

yielding,

$$\rho(\lim_{i \to \infty} T^i x, z) \leq \varepsilon/2. \tag{5.18}$$

Together with (5.14) and (5.17) this implies that for all integers $k \geq k_0$,

$$\rho(T^k x, \lim_{i \to \infty} T^i x) \leq \rho(T^k x, z) + \rho(z, \lim_{i \to \infty} T^i x) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon, \tag{5.19}$$

completing the proof. $\qquad \square$

Theorem 5.2 implies the following theorem:

Theorem 5.3. *Let X be a complete metric space, assume that $T : X \to X$ is a nonexpansive operator for which Condition 5.1 holds, the closure of $T(X)$ is compact, and let $\varepsilon > 0$. Then there exists a natural number k_0 such that for each $x \in X$ and each integer $k \geq k_0$,*

$$\rho(T^k x, \lim_{i \to \infty} T^i x) \leq \varepsilon. \tag{5.20}$$

Proposition 5.3. *Under the assumptions of Theorem 5.3, there exist an integer k_0 and a $\delta > 0$ such that for each finite sequence $\{x^i\}_{i=0}^{k_0} \subset X$ satisfying*

$$\rho(x^{i+1}, Tx^i) \leq \delta, \text{ for all } i = 0, 1, \ldots, k_0 - 1, \tag{5.21}$$

the inequality

$$d(x^{k_0}, \text{Fix}(T)) \leq \varepsilon \tag{5.22}$$

holds.

Proof. Theorem 5.3 implies that there exists an integer k_0 such that for each $x \in X$,

$$d(T^{k_0}x, \text{Fix}(T)) \leq \varepsilon/4. \tag{5.23}$$

Define

$$\delta := 4^{-1}\varepsilon(k_0)^{-1}, \tag{5.24}$$

assume that $\{x^i\}_{i=0}^{k_0} \subset X$ satisfies (5.21) and set

$$y^0 := x^0, \ y^{i+1} := Ty^i, \text{ for all } i = 0, 1, \ldots, k_0 - 1. \tag{5.25}$$

In view of (5.23) and (5.25),

$$d(y^{k_0}, \text{Fix}(T)) \leq \varepsilon/4. \tag{5.26}$$

Next we show, by induction, that

$$\rho(y^i, x^i) \leq i\delta, \text{ for all } i = 0, 1, \ldots, k_0. \tag{5.27}$$

Equation (5.25) implies that (5.27) holds for $i = 0$. Let $i \in \{0, 1, \ldots, k_0 - 1\}$ for which (5.27) holds. By the nonexpansiveness of T, (5.21), (5.25) and (5.27),

$$\rho(y^{i+1}, x^{i+1}) = \rho(Ty^i, x^{i+1})$$
$$\leq \rho(Ty^i, Tx^i) + \rho(Tx^i, x^{i+1}) \leq \rho(y^i, x^i) + \delta \leq (i+1)\delta, \tag{5.28}$$

in particular,

$$\rho(y^{k_0}, x^{k_0}) \leq k_0\delta. \tag{5.29}$$

It follows now from (5.24), (5.26) and (5.29) that

$$d(x^{k_0}, \text{Fix}(T)) \leq d(x^{k_0}, y^{k_0}) + d(y^{k_0}, \text{Fix}(T)) \leq \varepsilon/4 + \varepsilon/4, \tag{5.30}$$

which concludes the proof. $\square$

Theorem 5.4. *Under the assumptions of Theorem 5.3, there exist an integer k_0 and a $\delta > 0$ such that for each sequence $\{x^i\}_{i=0}^{\infty} \subset X$ satisfying $\rho(x^{i+1}, Tx^i) \leq \delta$, for all $i = 0, 1, \ldots$, the inequality*

$$d(x^i, \text{Fix}(T)) \leq \varepsilon \tag{5.31}$$

holds for all integers $i \geq k_0$.

Proof. The proof follows from Proposition 5.3. $\square$

Theorem 5.4 implies the next result.

Theorem 5.5. *Under the assumptions of Theorem 5.3, if we take a sequence*

$$\{\mu_k\}_{k=1}^\infty \subset (0,\infty), \ \lim_{k\to\infty} \mu_k = 0, \tag{5.32}$$

then there exists an integer $k_1 > 0$ such that for each sequence $\{x^i\}_{i=0}^\infty \subset X$ satisfying

$$\rho(x^{i+1}, Tx^i) \leq \mu_{i+1}, \ i = 0, 1, \dots, \tag{5.33}$$

the inequality $d(x^k, \mathrm{Fix}(T)) \leq \varepsilon$ holds for all integers $k \geq k_1$.

6. DATA-COMPATIBILITY OF THE HYBRID SUBGRADIENT METHOD

Our main data-compatibility result in Theorem 6.1 below is obtained under the following assumptions: $T : H \to H$ is a nonexpansive operator for which Condition 5.1 holds, the closure of $T(H)$ (which is denoted by $cl(T(H))$) is compact, $f : H \to R$ is a convex function, Lipschitz on any bounded set.

Since $cl(T(H))$ is compact, there exists a ball $B(0,M)$, with $M > 0$, such that

$$\mathrm{Fix}(T) \subset cl(T(H)) \subset B(0,M), \tag{6.1}$$

which means that the set $\mathrm{Fix}(T)$ is bounded.

Proposition 5.1 yields that $\mathrm{Fix}(T)$ is closed. Since $\mathrm{Fix}(T)$ is a closed subset of a compact set, which is $cl(T(H))$, then it easy to see that $\mathrm{Fix}(T)$ is compact also. By Condition 5.1 and Proposition 5.1 we get that $\mathrm{Fix}(T)$ is nonempty. The convexity of f on H implies the continuity of f on the space H and especially on the subset $\mathrm{Fix}(T)$. The continuity of f on $\mathrm{Fix}(T)$ and the compactness of $\mathrm{Fix}(T)$, implies that there exists some point $x \in \mathrm{SOL}(f,\mathrm{Fix}(T))$, i.e., $\mathrm{SOL}(f,\mathrm{Fix}(T)) \neq \varnothing$.

Since f is Lipschitz on any bounded set, there exists a number $\bar{L} > 1$ such that

$$|f(z^1) - f(z^2)| \leq \bar{L}\|z^1 - z^2\|, \ \text{for all } z^1, z^2 \in B(0, 3M+2). \tag{6.2}$$

We need the following lemma to prove the main result.

Lemma 6.1. *Assume $T : H \to H$ is a nonexpansive operator for which Condition 5.1 holds and $cl(T(H))$ is compact. Let $f : H \to R$ convex function and Lipschitz on any bounded set. Let $\bar{x} \in \mathrm{SOL}(f,\mathrm{Fix}(T))$ and let $\Delta \in (0,1]$, $\alpha > 0$ and $x \in cl(T(H))$ satisfy*

$$\|x\| \leq 3M + 2, \ f(x) > f(\bar{x}) + \Delta, \tag{6.3}$$

where M is as in (6.1). Further, let $v \in \partial f(x)$. Then $v \neq 0$ and

$$y := T\left(x - \alpha\|v\|^{-1}v\right) \tag{6.4}$$

satisfies

$$\|y - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2, \tag{6.5}$$

where $\bar{L}$ is as in (6.2). Moreover,

$$d(y, \mathrm{SOL}(f,\mathrm{Fix}(T)))^2 \leq d(x, \mathrm{SOL}(f,\mathrm{Fix}(T)))^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2. \tag{6.6}$$

Proof. From (6.3) $v \neq 0$. For $\bar{x} \in \mathrm{SOL}(f,\mathrm{Fix}(T))$, we have, by (6.2) and (6.1), that for each $z \in B(\bar{x}, 4^{-1}\Delta\bar{L}^{-1})$,

$$f(z) \leq f(\bar{x}) + \bar{L}\|z - \bar{x}\| \leq f(\bar{x}) + 4^{-1}\Delta. \tag{6.7}$$

Therefore, (6.3) and $v \in \partial f(x)$, imply that

$$\langle v, z - x \rangle \leq f(z) - f(x) \leq -(3/4)\Delta, \ \text{for all } z \in B(\bar{x}, 4^{-1}\Delta\bar{L}^{-1}). \tag{6.8}$$

From this inequality we deduce that

$$\left\langle ||v||^{-1}v, z-x \right\rangle < 0, \text{ for all } z \in B(\bar{x}, 4^{-1}\Delta\bar{L}^{-1}), \tag{6.9}$$

or, setting $\bar{z} := \bar{x} + 4^{-1}\bar{L}^{-1}\Delta||v||^{-1}v$, that

$$0 > \left\langle ||v||^{-1}v, \bar{z}-x \right\rangle = \left\langle ||v||^{-1}v, \bar{x} + 4^{-1}\bar{L}^{-1}\Delta||v||^{-1}v - x \right\rangle. \tag{6.10}$$

This leads to

$$\left\langle ||v||^{-1}v, \bar{x}-x \right\rangle < -4^{-1}\bar{L}^{-1}\Delta. \tag{6.11}$$

Putting $\tilde{y} := x - \alpha||v||^{-1}v$, we arrive at

$$\begin{aligned}
\|\tilde{y} - \bar{x}\|^2 &= \|x - \alpha||v||^{-1}v - \bar{x}\|^2 \\
&= \|x - \bar{x}\|^2 - 2\left\langle x - \bar{x}, \alpha||v||^{-1}v \right\rangle + \alpha^2 \\
&\leq \|x - \bar{x}\|^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2.
\end{aligned} \tag{6.12}$$

From all the above we obtain

$$\begin{aligned}
\|y - \bar{x}\|^2 &= \|T\tilde{y} - \bar{x}\|^2 \leq \|\tilde{y} - \bar{x}\|^2 \\
&\leq \|x - \bar{x}\|^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2,
\end{aligned} \tag{6.13}$$

which completes the proof. $\qquad\square$

Now we present the main theorem showing that sequences generated by the hybrid subgradient method (HSM) have a $(\tau, \bar{L})$-output, i.e., contain an iterate that is data-compatible. The condition (6.14) in this and subsequent theorems is common in many results in optimization theory, see, e.g., Theorem 2.2 and many subsequent theorems in Shor's classic book [11].

Theorem 6.1. *Assume $T : H \to H$ is a nonexpansive operator for which Condition 5.1 holds and $cl\,(T\,(H))$ is compact. Let $f : H \to R$ be a convex function and Lipschitz on any bounded set. Let*

$$\{\alpha_k\}_{k=0}^{\infty} \subset (0,1], \text{ be a sequence such that } \lim_{k\to\infty}\alpha_k = 0 \text{ and } \sum_{k=0}^{\infty}\alpha_k = \infty, \tag{6.14}$$

let $\bar{L}$ be fixed, as defined by (6.2), and let $\tau \in (0,1)$. Then there exists an integer K such that for any sequence $\{x^k\}_{k=1}^{\infty} \subset cl\,(T\,(H))$, generated by Algorithm 4.1 , the inequalities

$$d(x^k, \text{SOL}(f, \text{Fix}\,(T))) \leq \tau \tag{6.15}$$

and

$$f(x^k) \leq f(z) + \tau\bar{L} \text{ for all } z \in \text{SOL}(f, \text{Fix}\,(T)) \tag{6.16}$$

hold for all integers $k \geq K$.

Proof. Fix an $\bar{x} \in \text{SOL}(f, \text{Fix}(T))$. It is not difficult to see that there exists a number $\tau_0 \in (0, \tau/4)$ such that for each $x \in cl\,(T\,(H))$ satisfying $d(x, \text{Fix}(T)) \leq \tau_0$ and $f(x) \leq f(\bar{x}) + \tau_0$ we have

$$d(x, \text{SOL}(f, \text{Fix}(T))) \leq \tau/4. \tag{6.17}$$

Since $\{x^k\}_{k=1}^{\infty}$ is generated by Algorithm 4.1 we know, from (4.8), (4.7) and (4.2), that

$$\left\|x^k - Tx^{k-1}\right\| \leq \alpha_{k-1}, \text{ for all } k \geq 1. \tag{6.18}$$

Thus, by Theorem 5.5 and (6.14), there exists an integer n_1 such that

$$d(x^k, \text{Fix}(T)) \leq \tau_0, \text{ for all } k \geq n_1. \tag{6.19}$$

This, along with (6.1), guarantees that

$$\|x^k\| \leq M + 1, \text{ for all } k \geq n_1. \tag{6.20}$$

Choose a positive τ_1 for which $\tau_1 < (8\bar{L})^{-1}\tau_0$, by (6.14) there is an integer $n_2 > n_1$ such that

$$\alpha_k \leq \tau_1(32)^{-1}, \text{ for all } k > n_2, \tag{6.21}$$

and so, there is an integer $n_0 > n_2 + 4$ such that

$$\sum_{k=n_2}^{n_0-1} \alpha_k > 8(2M+1)^2 \bar{L}\tau_0^{-1}. \tag{6.22}$$

We show now that there exists an integer $p \in [n_2+1, n_0]$ such that $f(x^p) \leq f(\bar{x}) + \tau_0$. Assuming the contrary means that for all $k \in [n_2+1, n_0]$,

$$f(x^k) > f(\bar{x}) + \tau_0. \tag{6.23}$$

By (6.23), (6.14), (6.20) and using Lemma 6.1, with $\Delta = \tau_0$, $\alpha = \alpha_k$, $x = x^k$, $y = x^{k+1}$, $v = s^k$, we get, for all $k \in [n_2+1, n_0]$,

$$d(x^{k+1}, \text{SOL}(f, \text{Fix}(T)))^2$$
$$\leq d(x^k, \text{SOL}(f, \text{Fix}(T)))^2 - 2\alpha_k(4\bar{L})^{-1}\tau_0 + \alpha_k^2. \tag{6.24}$$

According to the choice of τ_1 and by (6.21) this implies that for all $k \in [n_2+1, n_0]$,

$$d(x^k, \text{SOL}(f, \text{Fix}(T)))^2 - d(x^{k+1}, \text{SOL}(f, \text{Fix}(T)))^2$$
$$\geq \alpha_k[(2\bar{L})^{-1}\tau_0 - \alpha_k]$$
$$\geq \alpha_k(4\bar{L})^{-1}\tau_0, \tag{6.25}$$

which, together with (6.20) and (6.1), gives

$$(2M+1)^2$$
$$\geq d(x^{n_2+1}, \text{SOL}(f, \text{Fix}(T)))^2$$
$$\geq \sum_{k=n_2+1}^{n_0} \left(d(x^k, \text{SOL}(f, \text{Fix}(T)))^2 - d(x^{k+1}, \text{SOL}(f, \text{Fix}(T)))^2 \right)$$
$$\geq (4\bar{L})^{-1}\tau_0 \sum_{k=n_2+1}^{n_0} \alpha_k, \tag{6.26}$$

and

$$\sum_{k=n_2+1}^{n_0} \alpha_k \leq (2M+1)^2 4\bar{L}\tau_0^{-1}. \tag{6.27}$$

This contradicts (6.22), proving that there is an integer $p \in [n_2+1, n_0]$ such that $f(x^p) \leq f(\bar{x}) + \tau_0$. Thus, by (6.19) and (6.17),

$$d(x^p, \text{SOL}(f, \text{Fix}(T))) \leq \tau/4. \tag{6.28}$$

We show now that for all $k \geq p$, $d(x^k, \text{SOL}(f, \text{Fix}(T))) \leq \tau$. Assuming the contrary,

$$\text{there exists a } q > p \text{ such that } d(x^q, \text{SOL}(f, \text{Fix}(T))) > \tau. \tag{6.29}$$

We may assume, without loss of generality, that

$$d(x^k, \text{SOL}(f, \text{Fix}(T))) \leq \tau, \text{ for all } p \leq k < q. \tag{6.30}$$

One of the following two cases must hold: (i) $f(x^{q-1}) \leq f(\bar{x}) + \tau_0$, or (ii) $f(x^{q-1}) > f(\bar{x}) + \tau_0$. In case (i), since $p \in [n_2 + 1, n_0]$, (6.19), (6.20) and (6.17) show that

$$d(x^{q-1}, \text{SOL}(f, \text{Fix}(T))) \leq \tau/4. \tag{6.31}$$

Thus, there is a point $z \in \text{SOL}(f, \text{Fix}(T))$ such that $\|x^{q-1} - z\| < \tau/3$. Using this fact and (6.18), (4.2), (4.1) and (6.21), yields

$$\begin{aligned}
\|x^q - z\| &\leq \|x^q - Tx^{q-1}\| + \|Tx^{q-1} - z\| \\
&\leq \alpha_{q-1} + \|x^{q-1} - z\| \leq \tau/4 + \tau/3,
\end{aligned} \tag{6.32}$$

proving that $d(x^q, \text{SOL}(f, \text{Fix}(T))) \leq \tau$. This contradicts (6.29) and implies that case (ii) must hold, namely that $f(x^{q-1}) > f(\bar{x}) + \tau_0$. This, along with (6.20), (6.21), the choice of τ_1, (6.30) and Lemma 6.1, with $\Delta = \tau_0$, $\alpha = \alpha_{q-1}$, $x = x^{q-1}$, $y = x^q$, shows that

$$\begin{aligned}
d(x^q, &\text{SOL}(f, \text{Fix}(T)))^2 \\
&\leq d(x^{q-1}, \text{SOL}(f, \text{Fix}(T)))^2 - 2\alpha_{q-1}(4\bar{L})^{-1}\tau_0 + \alpha_{q-1}^2 \\
&\leq d(x^{q-1}, \text{SOL}(f, \text{Fix}(T)))^2 - \alpha_{q-1}((2\bar{L})^{-1}\tau_0 - \alpha_{q-1}) \\
&\leq d(x^{q-1}, \text{SOL}(f, \text{Fix}(T)))^2 \leq \tau^2,
\end{aligned} \tag{6.33}$$

namely, that $d(x^q, \text{SOL}(f, \text{Fix}(T))) \leq \tau$. This contradicts (6.29), proving that, for all $k \geq p$, $d(x^k, \text{SOL}(f, \text{Fix}(T))) \leq \tau$. Together with (6.1) and (6.2) this implies that, for all $k \geq n_0$,

$$f(x^k) \leq f(z) + \tau\bar{L} \text{ for all } z \in \text{SOL}(f, \text{Fix}(T)), \tag{6.34}$$

and the proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

7. THE STRING-AVERAGING HYBRID SUBGRADIENT METHOD (SA-HSM)

Assume that $O_1, O_2, \ldots, O_m$ are nonexpansive operators mapping H into H, for which

$$\mathscr{F} := \bigcap_{i=1}^{m} \text{Fix}(O_i) \neq \varnothing \tag{7.1}$$

Let $f : H \to R$ be a convex function and Lipschitz on any bounded set. We are interested in solving the following problem by using a string-averaging algorithmic scheme.

$$\min\{f(x) \mid x \in \mathscr{F}\} \tag{7.2}$$

whose solution means to

$$\text{find a point } x \text{ in } \text{SOL}(f, \mathscr{F}), \tag{7.3}$$

where

$$\text{SOL}(f, \mathscr{F}) := \{x \in \mathscr{F} \mid f(x) \leq f(y), \text{ for all } y \in \mathscr{F}\}. \tag{7.4}$$

For $t = 1, 2, \ldots, \Theta$, let the *string* I_t be an ordered subset of $\{1, 2, \ldots, m\}$ of the form

$$I_t = (i_1^t, i_2^t, \ldots, i_{m(t)}^t), \tag{7.5}$$

with $m(t)$ the number of elements in I_t. For any $x \in H$, the product of operators along a string $I_t, t = 1, 2, \ldots, \Theta$, is

$$F_t(x) := O_{i_{m(t)}^t} \cdots O_{i_2^t} O_{i_1^t}(x), \tag{7.6}$$

and is called a "string operator".

We deal with string-averaging of fixed strings and fixed weights. To this end we assume that

$$\{1, 2, \ldots, m\} \subset \bigcup_{t=1}^{\Theta} I_t \tag{7.7}$$

and that a system of positive weights $w_1, w_2, \cdots, w_\Theta$ such that $\sum_{t=1}^{\Theta} w_t = 1$ is fixed and given. We define the operator

$$O(x) := \sum_{t=1}^{\Theta} w_t F_t(x). \tag{7.8}$$

This operator will be called "fit" if the strings that define it obey (7.7). We will need the following condition.

Condition 7.1. For all $i = 1, 2, \ldots, m$, the following holds: For any $y \in H \backslash \text{Fix}(O_i)$ there exist $x \in \mathscr{F} = \bigcap_{i=1}^{m} \text{Fix}(O_i)$ such that $\|O_i(y) - x\| < \|y - x\|$.

Proposition 7.1. *Let $O_1, O_2, \ldots, O_m$ be nonexpansive operators $O_i : H \to H$, and let $O = \sum_{t=1}^{\Theta} w_t F_t(x)$, be as in (7.8). If (7.7) and Condition 7.1 hold, then $\text{Fix}(O) = \mathscr{F}$.*

Proof. Clearly, $\mathscr{F} \subset \text{Fix}(O)$, therefore, it is sufficient to prove that $\text{Fix}(O) \subset \mathscr{F}$. Assume by negation that $\widehat{y} \in \text{Fix}(O)$ such that $\widehat{y} \notin \mathscr{F}$. This means that there is an $1 \leq \widehat{i} \leq m$ such that $\widehat{y} \notin \text{Fix}(O_{\widehat{i}})$. Condition 7.1 implies that there exist an $\bar{x} \in \mathscr{F}$ that satisfies $\|O_{\widehat{i}}(\widehat{y}) - \bar{x}\| < \|\widehat{y} - \bar{x}\|$. From this inequality, since $O_1, O_2, \ldots, O_m$ are nonexpansive operators, it is easy to see that

$$\|O(\widehat{y}) - \bar{x}\| = \left\| \sum_{t=1}^{\Theta} w_t F_t(\widehat{y}) - \bar{x} \right\| \leq \sum_{t=1}^{\Theta} w_t \|F_t(\widehat{y}) - \bar{x}\| < \|\widehat{y} - \bar{x}\|, \tag{7.9}$$

and, consequently, that $\widehat{y} \notin \text{Fix}(O)$. This contradicts the negation assumption made above and completes the proof. $\square$

We propose the following string-averaging hybrid subgradient method (SA-HSM) for solving the problem (7.2).

Algorithm 7.2. String-Averaging Hybrid Subgradient Method (SA-HSM).

Initialization: Let $\{\alpha_k\}_{k=0}^{\infty} \subset (0, 1]$ be a scalar sequence and let $x^0 \in H$ be an arbitrary initialization vector.

Iterative step: Given a current iteration vector x^k calculate the next vector as follows: Choose any $s^k \in \partial f(x^k)$ and calculate

$$x^{k+1} = O\left(x^k - \alpha_k \frac{s^k}{\|s^k\|}\right), \tag{7.10}$$

but if $s^k = 0$ then set $\dfrac{s^k}{\| s^k \|} := 0$.

The next theorem shows that sequences generated by the string-averaging hybrid subgradient method (SA-HSM) have a $(\tau, \bar{L})$-output, i.e., contain an iterate that is data-compatible.

Theorem 7.3. *Let $O_1, O_2, \ldots, O_m$ be nonexpansive operators mapping H into H, such that $\mathscr{F} = \bigcap\limits_{i=1}^{m} \mathrm{Fix}\,(O_i) \neq \varnothing$ and $cl\,(O_i(H))$ is compact for all $i = 1, 2, \ldots, m$. Let $O = \sum_{t=1}^{\Theta} w_t F_t(x)$ be as in (7.8) and assume that $\lim_{j \to \infty} O^j y^0$ exists for any $y^0 \in H$. Let $f : H \to R$ be a convex function and Lipschitz on any bounded set. Let $\{\alpha_k\}_{k=0}^{\infty} \subset (0,1]$ be a sequence such that*

$$\lim_{k \to \infty} \alpha_k = 0 \text{ and } \sum_{k=0}^{\infty} \alpha_k = \infty, \tag{7.11}$$

let $\bar{L}$ be fixed, as defined by (6.2), and let $\tau \in (0,1)$. If (7.7) and Condition 7.1 hold then there exists an integer K such that for any sequence $\{x^k\}_{k=0}^{\infty} \subset H$, generated by Algorithm 7.2, the inequalities

$$d(x^k, \mathrm{SOL}(f, \mathscr{F})) \leq \tau \tag{7.12}$$

and

$$f(x^k) \leq f(z) + \tau \bar{L} \text{ for all } z \in \mathrm{SOL}(f, \mathscr{F}) \tag{7.13}$$

hold for all integers $k \geq K$.

Proof. Since $O_1, O_2, \ldots, O_m$ are nonexpansive and O is a fit operator, it follows that O is nonexpansive. Moreover, Condition 7.1 and Proposition 7.1 ensure that $\mathrm{Fix}\,(O) = \mathscr{F}$. This, along with the other assumptions of the theorem, enable the use of Theorem 6.1 to complete the proof. $\qquad\square$

8. Data-compatibility for constrained minimization with inconsistent constraints

In this section we consider a data pair (Γ, f), assuming that $\Gamma := \{C_i\}_{i=1}^{m}$ is a family of closed and convex subsets of H, not necessarily obeying that $C := \cap_{i=1}^{m} C_i \neq \varnothing$. Let $\{w_i\}_{i=1}^{m}$ be a set of weights such that $w_i > 0$ and $\sum_{i=1}^{m} w_i = 1$. It is well-known that the operator $P_w := \sum_{i=1}^{m} w_i P_{C_i}$ is nonexpansive and satisfies

$$\mathrm{Fix}\,(P_w) = \mathrm{Arg\,min}\{\mathrm{Prox}_\Gamma(x) \mid x \in H\}, \tag{8.1}$$

where $\mathrm{Prox}_\Gamma(x) := \frac{1}{2} \sum_{i=1}^{m} w_i \|P_{C_i}(x) - x\|^2$, see the succinct [24, Subsection 5.4] on the simultaneous projection method. If $C \neq \varnothing$ then $\mathrm{Fix}\,(P_w) = C$. If, however, $C = \varnothing$ then $\mathrm{Fix}\,(P_w) = \Pi(\Gamma, \gamma)$ for $\gamma = \min\{\mathrm{Prox}_\Gamma(x) \mid x \in H\}$ and is nonempty. Moreover, for any $y^0 \in H$ the limit $\lim_{k \to \infty}(P_w)^k y^0$ exists and belong to $\mathrm{Fix}\,(P_w)$.

Consider the following algorithm.

Algorithm 8.1. Simultaneous Projection Hybrid Subgradient Method (SP-HSM).

Initialization: Let $\{\alpha_k\}_{k=0}^{\infty} \subset (0,1]$ be a scalar sequence and let $x^0 \in H$ be an arbitrary initialization vector.

Iterative step: Given a current iteration vector x^k calculate the next vector as follows:

Choose any $s^k \in \partial f(x^k)$ and calculate

$$x^{k+1} = P_w \left(x^k - \alpha_k \frac{s^k}{\| s^k \|} \right), \tag{8.2}$$

but if $s^k = 0$ then set $\dfrac{s^k}{\| s^k \|} := 0$.

From the above assumptions and discussion we obtain the following theorem as a consequence of Theorem 6.1. It does not assume consistency of the underlying constraints $\Gamma = \{C_i\}_{i=1}^m$, and shows that sequences generated by the simultaneous projection hybrid subgradient method (SP-HSM) have a $(\tau, \bar{L})$-output, i.e., contain an iterate that is data-compatible. Observe in the next theorem that $\bar{L}$ is a fixed constant that obeys (6.2) and that the parameter τ must obey $\tau \in (0,1)$, so, once $\bar{L}$ is fixed the "user" can choose a small τ so that $\tau \bar{L}$ is as small as he wants.

Theorem 8.2. *Assume that* $cl\,(P_w\,(H))$ *is compact. Let* $f : H \to R$ *be a convex function and Lipschitz on any bounded set, let*

$$\{\alpha_k\}_{k=0}^\infty \subset (0,1], \text{ be a sequence such that } \lim_{k \to \infty} \alpha_k = 0 \text{ and } \sum_{k=0}^\infty \alpha_k = \infty, \tag{8.3}$$

let $\bar{L}$ *be fixed, as defined by (6.2), and let* $\tau \in (0,1)$. *Then there exists an integer K such that, for any sequence* $\{x^k\}_{k=0}^\infty \subset H$ *generated by Algorithm 8.1, the inequalities*

$$d(x^k, \text{SOL}(f, \text{Fix}\,(P_w))) \leq \tau \tag{8.4}$$

and

$$f(x^k) \leq f(z) + \tau \bar{L} \text{ for all } z \in \text{SOL}(f, \text{Fix}\,(P_w)) \tag{8.5}$$

hold for all integers $k \geq K$.

9. MINIMIZATION OVER DISJOINT HARD AND SOFT CONSTRAINTS SETS

Here we describe the minimization over disjoint hard and soft constraints sets problem and its relation to our work in this paper. The issue of hard and soft constraints often arises in convex feasibility problems (CFPs), mentioned in Subsection 2.1 above, see, e.g., [35]. Many studies consider the, often occurring, situation when the CFP is not consistent, i.e., the intersection of all constraints is empty, see, e.g., [36] and the recent review [37]. In that case hard constraints are those which definitely must be satisfied, while soft constraint are those we would like to be satisfied - but not at the expense of the others.

Let $\Gamma_1 := \{C_i\}_{i=1}^{m_1}$ and $\Gamma_2 := \{Q_i\}_{i=1}^{m_2}$ be two finite families of constraints such that $C := \cap_{i=1}^{m_1} C_i \neq \varnothing$ and $Q := \cap_{i=1}^{m_2} Q_i \neq \varnothing$ but $C \cap Q = \varnothing$ and let C and Q be the hard and soft constraints, respectively. In view of the inability to solve the CFP given by $C \cap Q$, it makes sense to look for a point that will solve the hard/soft-CFP (h/s-CFP): Find a point in C that is closest to the set Q, according to some metric, say the Euclidean distance.

This can be done, in principle, by using the well-known 1959 Cheney-Goldstein theorem [38] that specifies conditions under which alternating metric projections onto two sets are guaranteed to converge to the best approximation pair. A *best approximation pair* relative to two closed convex sets C and Q is a pair $(c,q) \in C \times Q$ attaining $\|c - q\| = \min \|C - Q\|$, where $C - Q := \{x - y \mid x \in C,\ y \in Q\}$, see, e.g., Deutsch's book [40] or the recent [39].

Cheney and Goldstein considered the case of two nonempty closed convex sets C and Q in Hilbert space, with P_1 and P_2 denoting the orthogonal projection (proximity map) onto C and Q, respectively, and $T := P_1 P_2$. They show that the sequence $x^k = T^k(x^0)$ obtained by alternating distance minimizations converges to a fixed point of T, regardless whether $C \cap Q = \varnothing$ or not, if either (i) one of the two sets is compact, or (ii) one set is finite-dimensional and the distance between the two sets is attained. In particular, when the intersection of the two convex sets is nonempty, the sequence $\{x^k\}_{k=0}^{\infty}$ converges to a member of that intersection, in either of the two cases above, see, e.g., Subsection 2.1 of [37].

The problem of minimization over disjoint hard and soft constraints sets occurs when the fixed point set $\text{Fix}\,(T)$ is larger than a singleton and we want to find in it a minimizer of some given target function f, leading to a constrained minimization problem like (4.4). The SA-PSM of [33], mentioned in Section 3 above, will not apply to this problem but the hybrid subgradient method (HSM) of Algorithm 4.1 studied here will, allowing us to reach a data-compatible point.

Acknowledgments

References

[1] C. Bargetz, S. Reich, R. Zalas, Convergence properties of dynamic string averaging projection methods in the presence of perturbations, Numer. Algo. 77 (2018), 185-209.

[2] Y. Censor, T. Elfving, G.T. Herman, Averaging strings of sequential iterations for convex feasibility problems, Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), pp. 101-114, Elsevier Science Publishers, Amsterdam, The Netherlands, 2001.

[3] T.Y. Kong, H. Pajoohesh, G.T. Herman, String-averaging algorithms for convex feasibility with infinitely many sets, Inverse Probl. 35 (2019), 034001.

[4] A. J. Zaslavski, Algorithms for Solving Common Fixed Point Problems, Springer International Publishing, 2018.

[5] H. Iiduka, Fixed point optimization algorithm and its application to network bandwidth allocation, J. Comput. Appl. Math. 236 (2012), 1733-1742.

[6] H. Iiduka, Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping, Math. Program. 149 (2015), 131-165.

[7] H. Iiduka, Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings, Math. Program. 159 (2016), 509-538.

[8] P.-E. Maingé, Convex minimization over the fixed point set of demicontractive mappings, Positivity 12 (2008), 269-280.

[9] Y. Hayashi, H. Iiduka, Optimality and convergence for convex ensemble learning with sparsity and diversity based on fixed point optimization, Neurocomputing 273 (2018), 367-372.

[10] F. Deutsch, I. Yamada, Minimizing certain convex functions over the intersection of the fixed point sets of nonexpansive mappings, Numer. Funct. Anal. Optim. 19 (1998), 33-56.

[11] N. Z. Shor, Minimization Methods for Non-Differentiable Functions, Springer Series in Computational Mathematics, Springer, Berlin, 1985.

[12] Ya.I. Alber, A.N. Iusem, M.V. Solodov, On the projected subgradient method for nonsmooth convex optimization in a Hilbert space, Math. Program. 81 (1998), 23-35.

[13] I. Yamada, N. Ogura, Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings, Numer. Funct. Anal. Optim. 25 (2004), 619-655.

[14] S.A. Hirstoaga, Iterative selection methods for common fixed point problems, J. Math. Anal. Appl 324 (2006), 1020-1035.

[15] C. Martinez-Yanes, H-K. Xu, Strong convergence of the CQ method for fixed point iteration processes, Nonlinear Anal. 64 (2006), 2400-2411.

[16] P.-E. Maingé, Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization, Set-Valued Anal 16 (2008), 899-912.

[17] K. Aoyama, F. Kohsaka, Viscosity approximation process for a sequence of quasi-nonexpansive mappings, Fixed Point Theory Appl. 17 (2014), 1-11.

[18] A. Cegielski, Application of quasi-nonexpansive operators to an iterative method for variational inequality, SIAM J. Optim. 25 (2015), 2165-2181.

[19] J.Y. Bello Cruz, On proximal subgradient splitting method for minimizing the sum of two nonsmooth convex functions, Set-Valued Var. Anal. 25 (2017), 245-263.

[20] I. Yamada, The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings, Stud. Comput. Math. 8 (2001), 473-504.

[21] B. J. Hicks, A. J. Medland, G. Mullineux, The representation and handling of constraints for the design, analysis and optimization of high speed machinery, Artificial Intelligence for Engineering Design, Analysis and Manufacture (AIEDAM) 20 (2006), 313-328.

[22] P. L. Combettes, The foundations of set theoretic estimation, Proceedings of the IEEE 81 (1993), 182-208.

[23] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev. 38 (1996), 367-426.

[24] A. Cegielski, Iterative methods for fixed point problems in Hilbert spaces, Springer-Verlag, Berlin, Heidelberg, 2012.

[25] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Second edition, Springer, New York, 2017.

[26] Y. Censor, E. Garduño, E.S. Helou, G.T. Herman, Derivative-free superiorization: Principle and algorithm, Numer. Algor. (2020). https://doi.org/10.1007/s11075-020-01038-w.

[27] Y. Censor, A. Ben-Israel, Y. Xiao, J.M. Galvin, On linear infeasibility arising in intensity-modulated radiation therapy inverse planning, Linear Algebra Appl. 428 (2008), 1406-1420.

[28] M. Brooke, Y. Censor, A. Gibali, Dynamic string-averaging CQ-methods for the split feasibility problem with percentage violation constraints arising in radiation therapy treatment planning, Int. Trans. Oper. Res. (2020). https://doi.org/10.1111/itor.12929

[29] P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, IEEE Transactions on Signal Processing 42 (1994), 2955-2966.

[30] V. Caselles, A. Chambolle, M. Novaga, Total variation in imaging, Handbook of Mathematical Methods in Imaging (O. Scherzer, ed.), pp. 1016–1057, Springer, New York, NY, 2011.

[31] R. Davidi, Y. Censor, R.W. Schulte, S. Geneser, L. Xing, Feasibility-seeking and superiorization algorithms applied to inverse treatment planning in radiation therapy, Contemp. Math. 636 (2015), 83-92.

[32] T.S. Ferguson, Optimal stopping and applications, Universiy of California, Los Angeles (UCLA), https://www.e-booksdirectory.com/details.php?ebook=5651, 2008.

[33] Y. Censor, A.J. Zaslavski, String-averaging projected subgradient methods for constrained minimization, Optim. Methods Softw. 29 (2014), 658-670.

[34] S. Reich, A.J. Zaslavski, Genericity in Nolinear Analysis, Springer-Verlag, New York, 2014.

[35] P. L. Combettes, P. Bondon, Hard-constrained inconsistent signal feasibility problems, IEEE Transactions on Signal Processing 47 (1999), 2460-2468.

[36] H.H. Bauschke, J.M. Borwein, A.S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, Recent developments in optimization theory and nonlinear analysis : AMS/IMU Special Session on Optimization and Nonlinear Analysis, May 24-26, 1995, Jerusalem, Israel (Y. Censor and S. Reich, eds.), pp. 1-38, Providence, R.I. : American Mathematical Society, 1997.

[37] Y. Censor, M. Zaknoon, Algorithms and convergence results of projection methods for inconsistent feasibility problems: A review, Pure Appl. Funct. Anal. 3 (2018), 565-586.

[38] W. Cheney, A.A. Goldstein, Proximity maps for convex sets, Proc. Amer. Math. Soc. 10 (1959), 448-450.

[39] R. Aharoni, Y. Censor, Z. Jiang, Finding a best approximation pair of points for two polyhedra, Comput. Optim. Appl. 71 (2018), 509-523.

[40] F. Deutsch, Best Approximation in Inner Product Spaces, Springer, New York, NY, 2001.