String-Averaging Projected Subgradient Methods for Constrained Minimization

Yair Censor[1] and Alexander J. Zaslavski[2]

[1]Department of Mathematics, University of Haifa
Mt. Carmel, Haifa 31905, Israel
(yair@math.haifa.ac.il)

[2]Department of Mathematics
The Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
(ajzasl@techunix.technion.ac.il)

February 17, 2013.
Revised: June 28, 2013, and August 26, 2013.

Abstract

We consider constrained minimization problems and propose to replace the projection onto the entire feasible region, required in the Projected Subgradient Method (PSM), by projections onto the individual sets whose intersection forms the entire feasible region. Specifically, we propose to perform such projections onto the individual sets in an algorithmic regime of a feasibility-seeking iterative projection method. For this purpose we use the recently developed family of Dynamic String-Averaging Projection (DSAP) methods wherein iteration-index-dependent variable strings and variable weights are permitted. This gives rise to an algorithmic scheme that generalizes, from the algorithmic structural point of view, earlier work of Helou Neto and De Pierro, of Nedić, of Nurminski, and of Ram et al.

Keywords and phrases: Fixed point, Hilbert space, metric projection, nonexpansive operator, perturbation resilience, projected subgradient minimization, string-averaging projection methods, superiorization method, variable strings, variable weights.

2010 Mathematics Subject Classification (MSC): 65K10, 90C25

1 Introduction

The problem: We consider constrained minimization problems of the form

$$\text{minimize}\{\phi(x) \mid x \in C\}, \tag{1}$$

where ϕ is a convex objective function mapping from the J-dimensional Euclidean space R^J into the reals and $C \subseteq R^J$ is a given closed convex constraints set. Such constrained optimization problems lie at the heart of optimization theory and practice and constitute mathematical models for many scientific and real-world applications. The Projected Subgradient Method (PSM) for constrained minimization interlaces subgradient steps for objective function descent with projections onto the feasible region C to regain feasibility after every subgradient step.

PSM generates a sequence of iterates $\{x^k\}_{k=0}^{\infty}$ according to the recursion formula

$$x^{k+1} = P_C\left(x^k - t_k \phi'(x^k)\right), \tag{2}$$

where $t_k > 0$ is a step-size, $\phi'(x^k) \in \partial\phi(x^k)$ is a subgradient of ϕ at x^k, and P_C stands for the orthogonal (least Euclidean norm) projection onto the set C. The underlying philosophy is to perform unconstrained objective function descent steps by moving from x^k to $z^k := x^k - t_k\phi'(x^k)$ and then regain feasibility with respect to C by projecting z^k onto C.

Motivation: This projection onto C is a computational bottleneck in applying the method due to the required inner loop of (quadratic) minimization of the distance to the set C. Therefore the projected subgradient method is mostly useful only when the feasible region is "simple to project onto". To alleviate this difficulty we consider the situation where $C_1, C_2, \ldots, C_m$ are nonempty closed convex subsets of R^J, where m is a natural number, define

$$C := \cap_{i=1}^{m} C_i, \tag{3}$$

and propose to replace the projection P_C in (2) by projections onto the individual sets C_i executed in a specific algorithmic regime of a *feasibility-seeking* iterative projection method. This is an important development because often the entire feasible region C is not "simple to project onto" whereas the individual sets are. The class of *projection methods* is understood here as the class of methods that have the feature that they can reach an aim related to the family of sets $\{C_1, C_2, \ldots, C_m\}$ by performing projections (orthogonal, i.e., least Euclidean distance, or others) onto the individual sets C_i. The advantage of such methods occurs in situations where projections onto the individual sets are computationally simple to perform or at least simpler than a projection on the intersection C. Such methods have been in recent decades extensively investigated mathematically and used experimentally with great success on some huge and sparse real-world applications, consult, e.g., [2, 10], the books [3, 8, 9, 17, 18, 21, 22, 26], and the recent paper [5].

Contribution: We employ here for the feasibility-seeking method the *String-Averaging Projection* (SAP) scheme. This class was first introduced in [12] (in a formulation that is not restricted to feasibility-seeking) and was subsequently studied further in [4, 6, 13, 14, 15, 19], see also [3, Example 5.20]. SAP methods were also employed in applications [36, 40]. The Component-Averaged Row Projections (CARP) method of [23] also belongs to the class of SAP methods. Within the class of projection methods, SAP methods do not constitute a single algorithm but rather an *algorithmic scheme,* which means that by making a specific choice of strings and weights in SAP, along with choices of other parameters in the scheme, a deterministic algorithm for the problem at hand can be obtained.

Here we use the recently developed family of *Dynamic String-Averaging Projection* (DSAP) methods of [16] wherein iteration-index-dependent variable strings and variable weights are permitted (in all works prior to [16], SAP methods were formulated with a single predetermined set of strings and a single predetermined set of weights.)

Relation with previous works: The literature on subgradient minimization is vast and we mention [29, 38] as reperesentatives. Specifically to the research presented here, the introduction of DSAP to replace the projection P_C in (2) gives rise to an algorithmic scheme that includes and generalizes, from the algorithmic structural point of view, earlier recent work of Helou Neto and De Pierro [24, 25], of Nedić [31], of Nurminski [32, 33, 34, 35], and of Ram et al. [39].

Nurminski: The algorithms of Nurminski use Fejér operators, that can be used in feasibility-seeking, and introduces into them disturbances with diminishing step-sizes $\lambda_k \to 0$ as $k \to \infty$, where the rate of this tendency is such that $\sum_{k=0}^{\infty} \lambda_k = +\infty$. Under these conditions and a variety of additional assumptions, Nurminski showed asymptotic convergence of the iterates generated by his algorithms to a minimum point of the constrained minimization problem.

Helou Neto and De Pierro: The framework proposed by Helou Neto and De Pierro uses interlacing of "feasibility operators" with "optimality operators" with the aim of creating exact constrained minimization algorithms. Similarly to Nurminski, they employ diminishing step-sizes. Under these conditions and a variety of additional assumptions, different than those of Nurminski, they show asymptotic convergence of the iterates generated by their algorithmic framework to a minimum point of the constrained minimization problem.

However, when it comes to derivation of specific algorithms from the general framework of [25, Equation (3)], their feasibility operator F invariably takes the form

$$\mathcal{F}_F(x) = x - \mu(x)\nabla F(x), \tag{4}$$

where the function $F(x)$, whose gradient is calculated, is "a convex function such that the set of minima of this function coincides with the set X [in [25, Equation (3)] X is the feasible set of the minimization problem and should be identified with $C = \cap_{i=1}^{m} C_i$ in our notation] when it is not empty and defines a solution in an appropriate way (least squares for example) otherwise" and $\mu(x)$ are some parameters that are restricted in a particular manner as in [25, Lemma 7 or Corollary 8]. Our feasibility-seeking DSAP algorithms are not limited to the form of (4).

Nedić and Ram et al.: The overall approach here is to apply gradient and subgradient iterative methods for the objective function minimization and interlace into them random feasibility updates. The resulting "random projection method" [31, Equation (4)] bears structural similarity to our approach since it replaces the projection P_C in (2) by projections onto individual sets C_i but not in a DSAP regime which is more general. The randomness refers to the way the constraints are picked up for the feasibility updates.

There are various differences among all the above works and between them and our work, differences in overall setup of the problems, differences in the assumptions used for the various convergence results, etc. This is not the place for a full review of all these differences but the main contribution of our work presented here lies in the many more algorithmic possibilities revealed by the use of DSAP general scheme, while retaining overall convergence to an (exact) solution of the problem (1).

Paper layout: The DSAP algorithmic scheme is presented in Section 2 and our *String-Averaging Projected Subgradient Method* (SA-PSM) is presented in Section 3 and its convergence analysis is done in Section 4.

2 The dynamic string-averaging projection method with variable strings and weights

Our main result in Theorem 9 below is obtained in the finite-dimensional Euclidean space but the algorithms described in this section can be used in the more general setting of an infinite-dimensional Hilbert space. We first present, for the reader's convenience and for the sake of completeness, the string-averaging algorithmic scheme of [12]. Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $|| \cdot ||$. For $t = 1, 2, \ldots, M$, let the *string* I_t be an ordered subset of $\{1, 2, \ldots, m\}$ of the form

$$I_t = (i_1^t, i_2^t, \ldots, i_{m(t)}^t), \tag{5}$$

with $m(t)$ the number of elements in I_t. Suppose that there is a set $S \subseteq H$ such that there are operators $O_1, O_2, \ldots, O_m$ mapping S into S and an additional operator O which maps S^M into S.

Algorithm 1 ***The string-averaging algorithmic scheme***

Initialization: *$x^0 \in S$ is arbitrary.*

Iterative Step: *given the current iterate x^k,*

(i) calculate, for all $t = 1, 2, \ldots, M$,

$$T_t(x^k) = O_{i_{m(t)}^t} \cdots O_{i_2^t} O_{i_1^t}(x^k), \tag{6}$$

(ii) and then calculate

$$x^{k+1} = O(T_1(x^k), T_2(x^k), \ldots, T_M(x^k)). \tag{7}$$

For every $t = 1, 2, \ldots, M$, this algorithmic scheme applies to x^k successively the operators whose indices belong to the tth string. This can be done in parallel for all strings and then the operator O maps all end-points

$T_t(x^k)$ onto the next iterate x^{k+1}. This is indeed an algorithm provided that the operators $\{O_i\}_{i=1}^m$ and O all have algorithmic implementations. In this framework we get a (fully) *sequential algorithm* by the choice $M = 1$ and $I_1 = (1, 2, \dots, m)$ or a (fully) *simultaneous algorithm* by the choice $M = m$ and $I_t = (t),\ t = 1, 2, \dots, M$. Originally [12], Algorithm 1 can employ operators other than projections and convex combinations, therefore it is more general than SAP. On the other hand, Algorithm 1 is formulated for fixed strings and weights and this is generalized in the formulation of DSAP that follows.

Next we describe the *dynamic string-averaging projection* (DSAP) method with variable strings and weights. For $i = 1, 2, \dots, m$, we denote the projection onto the set C_i by $P_i = P_{C_i}$. An *index vector* is a vector $t = (t_1, t_2, \dots, t_q)$ such that $t_i \in \{1, 2, \dots, m\}$ for all $i = 1, 2, \dots, q$, its *length* is denoted by $\ell(t) = q$, and we define the operator $P[t]$ as the product of the individual projections onto the sets whose indices appear in the index vector t, namely,

$$P[t] := P_{t_q} P_{t_{q-1}} \cdots P_{t_1}, \tag{8}$$

and call it a *string operator*.

Definition 2 *A finite set Ω of index vectors is called* **`fit`** *if for each $i \in \{1, 2, \dots, m\}$, there exists a vector $t = (t_1, t_2, \dots, t_q) \in \Omega$ where q is a natural number such that $t_s = i$ for some $s \in \{1, 2, \dots, q\}$.*

Note that in the above definition q can vary. For each index vector t the string operator is nonexpansive, since the individual projections are, i.e.,

$$\|P[t](x) - P[t](y)\| \leq ||x - y|| \text{ for all } x, y \in H, \tag{9}$$

and also

$$P[t](x) = x \text{ for all } x \in C. \tag{10}$$

Denote by $\mathcal{M}$ the collection of all pairs (Ω, w), where Ω is a fit finite set of index vectors and

$$w : \Omega \to (0, \infty) \text{ is such that } \sum_{t \in \Omega} w(t) = 1. \tag{11}$$

A pair $(\Omega, w) \in \mathcal{M}$ and the function w were called in [6] an *amalgamator* and a *fit weight function*, respectively. For any $(\Omega, w) \in \mathcal{M}$ define the convex combination of the end-points of all strings defined by members of Ω by

$$P_{\Omega,w}(x) := \sum_{t \in \Omega} w(t) P[t](x),\ x \in H. \tag{12}$$

Fix a number $\Delta \in (0, 1/m)$ and an integer $\bar{q} \geq m$ and denote by $\mathcal{M}_* \equiv \mathcal{M}_*(\Delta, \bar{q})$ the set of all $(\Omega, w) \in \mathcal{M}$ such that the lengths of the strings are bounded and the weights are bounded away from zero, namely,

$$\mathcal{M}_* := \{(\Omega, w) \in \mathcal{M} \mid \ell(t) \leq \bar{q} \text{ and } \Delta \leq w(t), \text{ for all } t \in \Omega\}. \tag{13}$$

We make the assumption that $\bar{q} \geq m$ because only in this case there exists $(\Omega, w) \in M_*$ such that Ω is a singleton containing the index vector $(1, 2, \ldots, m)$ so that our class of algorithms includes also the classical cyclic projection algorithm.

The dynamic string-averaging projection (DSAP) method with variable strings and variable weights can now be described by the following algorithm.

Algorithm 3 ***The DSAP method with variable strings and variable weights***

Initialization: *select an arbitrary* $x^0 \in H$*,*

Iterative step: *given a current iteration vector* x^k *pick a pair* $(\Omega_k, w_k) \in \mathcal{M}_*$ *and calculate the next iteration vector* x^{k+1} *by*

$$x^{k+1} = P_{\Omega_k, w_k}(x^k). \tag{14}$$

The convergence properties and the, so called, perturbation resilience of this DSAP method were analyzed in [16].

3 The string-averaging projected subgradient method

Our proposed string-averaging projected subgradient method (SA-PSM) for the solution of (1) performs string-averaging steps with respect to the individual constraints of (3) instead of the single projection onto the entire feasible set C of (1) dictated by the PSM of (2).

Algorithm 4 ***The string-averaging projected subgradient method (SA-PSM)***

(0) Initialization: *Let* $\{\alpha_k\}_{k=0}^{\infty} \subset (0, 1]$ *be a scalar sequence and select arbitrary vectors* x^0, $s^0 \in H$*,*

(1) Iterative step: *given a current iteration vector* x^k *and a current vector* s^k*, pick a pair* $(\Omega_k, w_k) \in \mathcal{M}_*$ *and calculate the next vectors as follows:*

(1.1) *if* $0 \in \partial\phi(x^k)$ *then set* $s^k = 0$ *and calculate*

$$x^{k+1} = P_{\Omega_k,w_k}(x^k). \tag{15}$$

(1.2) *if* $0 \notin \partial\phi(x^k)$ *then choose and set* $s^k \in \partial\phi(x^k)$ *and calculate*

$$x^{k+1} = P_{\Omega_k,w_k}\left(x^k - \alpha_k \frac{s^k}{\| s^k \|}\right). \tag{16}$$

For each $x \in H$ and nonempty set $E \subseteq H$ define the distance

$$d(x,E) = \inf\{||x - y|| \mid \; y \in E\}. \tag{17}$$

Denoting the solution set of (1) by

$$SOL(\phi, C) := \{x \in C \mid \; \phi(x) \le \phi(y) \text{ for all } y \in C\}, \tag{18}$$

we specify, in our convergence result presented in the sequel, conditions which guarantee that for every $\varepsilon \in (0,1)$, and any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 4, there exists an integer K so that, for all $k \ge K$, the following inequalities hold:

$$d(x^k, SOL(\phi, C)) \le \varepsilon \tag{19}$$

and

$$\phi(x^k) \le \inf\{\phi(z) \mid z \in SOL(\phi, C)\} + \varepsilon\bar{L}, \tag{20}$$

where $\bar{L}$ is some well-defined constant. For each $x \in H$ and $r > 0$ define the closed ball

$$B(x,r) = \{y \in H \mid \; ||x - y|| \le r\}. \tag{21}$$

Definition 5 *Assume that* $C_1, C_2, \ldots, C_m$ *are nonempty closed convex subsets of* R^J, $C = \cap_{i=1}^m C_i$, *and that* C *is nonempty. We call a finite set* Ω *of index vectors* M***-fit with respect to the family*** $\{C_1, C_2, \ldots, C_m\}$ *(*M***-fit****, for short) if* Ω *is fit (see Definition 2) and there exists an* $M > 0$ *such that for each* $t = (t_1, t_2, \ldots, t_{\ell(t)}) \in \Omega$ *there is a* $u \in \{1, 2, \ldots, \ell(t)\}$ *such that*

$$C_{t_u} \subset B(0, M). \tag{22}$$

Denote the set $\mathcal{M}_{**} := \{(\Omega, w) \in \mathcal{M}_* \mid \Omega \text{ is } M\text{-fit}\}$.

In the next theorem we establish the convergence of sequences generated by Algorithm 3 with computational errors.

Theorem 6 *Assume that $C_1, C_2, \ldots, C_m$ are nonempty closed convex subsets of R^J, $C = \cap_{i=1}^m C_i$ and C is nonempty. For some $M > 0$ let there exist an index $s \in \{1, 2, \ldots, m\}$ such that*

$$C_s \subset B(0, M). \tag{23}$$

Further, let

$$\{\gamma_k\}_{k=1}^\infty \subset (0, 1] \text{ so that } \lim_{k\to\infty} \gamma_k = 0, \tag{24}$$

and $\varepsilon > 0$. Then there exists an integer K such that for any $\{(\Omega_k, w_k)\}_{k=1}^\infty \subset \mathcal{M}_{**}$ and any sequence $\{x^k\}_{k=0}^\infty$, which satisfies for all integer $k \geq 1$,

$$\left\| x^k - P_{\Omega_k, w_k}(x^{k-1}) \right\| \leq \gamma_k, \tag{25}$$

the inequality $d(x^k, C) \leq \varepsilon$ holds for all integers $k \geq K$.

In order to prove Theorem 6 we need the following auxiliary lemma.

Lemma 7 *Under the assumptions of Theorem 6, if $\{(\Omega_k, w_k)\}_{k=1}^\infty \subset \mathcal{M}_{**}$ and $\{x^k\}_{k=0}^\infty$ satisfies for all integer $k \geq 1$,*

$$\left\| x^k - P_{\Omega_k, w_k}(x^{k-1}) \right\| \leq 1, \tag{26}$$

then $\left\| x^k \right\| \leq 3M + 1$ for all $k \geq 1$.

Proof. Let $z \in C$, then, by (23),

$$\|z\| \leq M. \tag{27}$$

For any $k \geq 0$ (12) implies

$$P_{\Omega_{k+1}, w_{k+1}}(x^k) = \sum_{t \in \Omega_{k+1}} w_{k+1}(t) P[t](x^k). \tag{28}$$

By (26), (27), (28) and (11),

$$\begin{aligned} \left\| x^{k+1} \right\| &\leq \|z\| + \left\| x^{k+1} - z \right\| \leq M + \left\| x^{k+1} - z \right\| \\ &\leq M + \left\| z - P_{\Omega_{k+1}, w_{k+1}}(x^k) \right\| + \left\| P_{\Omega_{k+1}, w_{k+1}}(x^k) - x^{k+1} \right\| \\ &\leq M + 1 + \left\| \, z - \sum_{t \in \Omega_{k+1}} w_{k+1}(t) P[t](x^k) \, \right\| \\ &\leq M + 1 + \sum_{t \in \Omega_{k+1}} w_{k+1}(t) \left\| z - P[t](x^k) \right\|. \end{aligned} \tag{29}$$

Let $t = (t_1, t_2, \ldots, t_{\ell(t)}) \in \Omega_{k+1}$, then

$$P[t](x^k) = P_{t_{\ell(t)}} P_{t_{\ell(t)-1}} \cdots P_{t_1}(x^k), \tag{30}$$

and, since $\{(\Omega_k, w_k)\}_{k=1}^{\infty} \subset \mathcal{M}_{**}$, there is an index $j \in \{1, 2, \ldots, \ell(t)\}$ for which

$$C_{t_j} \subset B(0, M). \tag{31}$$

From (31),

$$\left\| P_{t_j} P_{t_{j-1}} \cdots P_{t_1}(x^k) \right\| \leq M. \tag{32}$$

Thus, since $z \in C$ and by (27), (30), (32) and properties of the projection operator we have

$$\left\| z - P[t](x^k) \right\| \leq \left\| z - P_{t_j} P_{t_{j-1}} \cdots P_{t_1}(x^k) \right\| \leq 2M, \tag{33}$$

yielding, by (29) and (33),

$$\left\| x^{k+1} \right\| \leq 3M + 1, \tag{34}$$

which completes the proof of the lemma. ■

Now we are ready to prove Theorem 6.

Proof of Theorem 6. We use [16, Theorem 7]. All its assumptions hold here. Indeed, Assumption (i) of [16, Theorem 7] follows from (23), Assumption (ii) holds since our space is finite-dimensional, and Assumptions (iii) and (iv) obviously hold. Therefore, there exists an integer k_1 so that, for each $\{(\Omega_k, w_k)\}_{k=1}^{\infty} \subset \mathcal{M}_{**}$, any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 3 with $\|x^0\| \leq 3M + 1$, converges, $\lim_{s \to \infty} x^s \in C$, and

$$\|x^k - \lim_{s \to \infty} x^s\| \leq \varepsilon/4 \text{ for all } k \geq k_1. \tag{35}$$

By (24) there is an integer k_2 such that

$$\gamma_k \leq (\varepsilon/4)(k_1 + 1)^{-1} \text{ for all } k \geq k_2. \tag{36}$$

Define

$$K := k_1 + k_2 + 2. \tag{37}$$

We will show that $d(x^k, C) \leq \varepsilon$ for all $k \geq K$. To this end take some $v \geq K$ and let

$$y^0 := x^{v - k_1}. \tag{38}$$

For any $k \geq 1$ let

$$y^k := x^{k+v-k_1} \text{ and } (\tilde{\Omega}_k, \tilde{w}_k) := (\Omega_{k+v-k_1}, w_{k+v-k_1}). \tag{39}$$

By (24), and since $\{(\Omega_k, w_k)\}_{k=1}^{\infty} \subset \mathcal{M}_{**}$ and the sequence $\{x^k\}_{k=0}^{\infty}$ satisfies

$$\left\| x^k - P_{\Omega_k, w_k}(x^{k-1}) \right\| \leq \gamma_k, \text{ for all } k \geq 1, \tag{40}$$

we can use Lemma 7 to obtain

$$\left\| x^k \right\| \leq 3M + 1 \text{ for all } k \geq 1. \tag{41}$$

Therefore,

$$\left\| y^0 \right\| = \left\| x^{v-k_1} \right\| \leq 3M + 1. \tag{42}$$

It follows that, for each $k \geq 1$,

$$\begin{aligned} \left\| y^k - P_{\tilde{\Omega}_k, \tilde{w}_k}(y^{k-1}) \right\| &= \left\| x^{k+v-k_1} - P_{\Omega_{k+v-k_1}, w_{k+v-k_1}}(x^{k+v-k_1-1}) \right\| \\ &\leq \gamma_{k+v-k_1} \leq (\varepsilon/4)(k_1+1)^{-1}. \end{aligned} \tag{43}$$

Defining now

$$z^0 := y^0 \text{ and } z^k = P_{\tilde{\Omega}_k, \tilde{w}_k}(z^{k-1}) \text{ for all } k \geq 1 \tag{44}$$

we obtain that

$$d(z^{k_1}, C) \leq \varepsilon/4. \tag{45}$$

Next we show, by induction, that for all $k = 0, 1, \ldots, k_1$

$$\left\| y^k - z^k \right\| \leq k(\varepsilon/4)(k_1+1)^{-1}. \tag{46}$$

Clearly, for $k = 0$ (46) holds. Assume that $0 \leq k < k_1$ and that (46) holds. Then, by (43), (44), (46) and nonexpansivity of $P_{\Omega,w}$ of (12),

$$\begin{aligned} \left\| y^{k+1} - z^{k+1} \right\| &= \left\| y^{k+1} - P_{\tilde{\Omega}_{k+1}, \tilde{w}_{k+1}}(z^k) \right\| \\ &\leq \left\| y^{k+1} - P_{\tilde{\Omega}_{k+1}, \tilde{w}_{k+1}}(y^k) \right\| + \left\| P_{\tilde{\Omega}_{k+1}, \tilde{w}_{k+1}}(y^k - P_{\tilde{\Omega}_{k+1}, \tilde{w}_{k+1}}(z^k)) \right\| \\ &\leq (\varepsilon/4)(k_1+1)^{-1} + k(\varepsilon/4)(k_1+1)^{-1} \\ &= (\varepsilon/4)(k_1+1)^{-1}(k+1). \end{aligned} \tag{47}$$

Thus, we have shown by induction that (46) holds for all $k = 0, 1, \ldots, k_1$ and, in particular, that $\left\| y^{k_1} - z^{k_1} \right\| \leq \varepsilon/4$, which implies that

$$d(x^k, C) = d(y^{k_1}, C) \leq \left\| y^{k_1} - z^{k_1} \right\| + d(z^{k_1}, C) \leq \varepsilon/2, \tag{48}$$

and Theorem 6 is proved. ■

4 Main convergence result for the string-averaging projected subgradient method

We will make use of the following bounded regularity condition, see [2, Definition 5.1], formulated in Hilbert space H.

Condition 8 *For each $\varepsilon > 0$ and each $M > 0$ there exists $\delta = \delta(\varepsilon, M) > 0$ such that for each $x \in B(0, M)$ satisfying $d(x, C_i) \leq \delta$, $i = 1, 2, \ldots, m$, the inequality $d(x, C) \leq \varepsilon$ holds.*

It follows from [2, Proposition 5.4(iii)] (see also [16, Proposition 5]) that if the Hilbert space H is finite-dimensional then Condition 8 holds.

Our main convergence result for the string-averaging projected subgradient method Algorithm 4 is the following theorem.

Theorem 9 *Let the assumptions of Theorem 6 hold and assume that $\phi : R^J \to R$ is a convex function. Let*

$$\{\alpha_k\}_{k=0}^{\infty} \subset (0, 1], \text{ such that } \lim_{k\to\infty} \alpha_k = 0 \text{ and } \sum_{k=0}^{\infty} \alpha_k = \infty, \tag{49}$$

and let $\varepsilon \in (0, 1)$. Then there exist an integer K and a real number $\bar{L}$ such that for any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 4 with $\{(\Omega_k, w_k)\}_{k=1}^{\infty} \subset \mathcal{M}_{**}$, the inequalities

$$d(x^k, SOL(\phi, C)) \leq \varepsilon \text{ and } \phi(x^k) \leq \inf\{\phi(z) \mid z \in SOL(\phi, C)\} + \varepsilon\bar{L} \tag{50}$$

hold for all integers $k \geq K$.

It is well-known that the function ϕ is continuous due to its convexity. Clearly, ϕ is Lipschitz on bounded subsets of R^J, therefore, since C is bounded, by the assumption (23), there exists a point $x \in SOL(\phi, C)$, i.e., $SOL(\phi, C)$ is nonempty. Furthermore, there exists a number $\bar{L} > 1$ such that

$$|\phi(z^1) - \phi(z^2)| \leq \bar{L}||z^1 - z^2|| \text{ for all } z^1, z^2 \in B(0, 3M + 2). \tag{51}$$

We will need the following lemma.

Lemma 10 *Let* $\bar{x} \in SOL(\phi, C)$ *and let* $\Delta \in (0,1]$ $\alpha > 0$ *and* $x \in R^J$ *satisfy*

$$\|x\| \leq 3M + 2, \ \phi(x) > \phi(\bar{x}) + \Delta. \tag{52}$$

Further, let $v \in \partial\phi(x)$ *and* $(\Omega, w) \in \mathcal{M}_{**}$*. Then* $v \neq 0$ *and*

$$y := P_{\Omega,w}\left(x - \alpha||v||^{-1}v\right) \tag{53}$$

satisfies

$$\|y - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2, \tag{54}$$

where $\bar{L}$ *is as in (51). Moreover,*

$$d(y, SOL(\phi, C))^2 \leq d(x, SOL(\phi, C))^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2. \tag{55}$$

Proof. From (52) $v \neq 0$. For $\bar{x} \in SOL(\phi, C)$, we have, by (51) and (23), that for each $z \in B(\bar{x}, 4^{-1}\Delta\bar{L}^{-1})$,

$$\phi(z) \leq \phi(\bar{x}) + \bar{L}||z - \bar{x}|| \leq \phi(\bar{x}) + 4^{-1}\Delta. \tag{56}$$

Therefore, (52) and $v \in \partial\phi(x)$, imply that

$$\langle v, z - x\rangle \leq \phi(z) - \phi(x) \leq -(3/4)\Delta \text{ for all } z \in B(\bar{x}, 4^{-1}\Delta\bar{L}^{-1}). \tag{57}$$

From this inequality we deduce that

$$\left\langle ||v||^{-1}v, z - x\right\rangle < 0 \text{ for all } z \in B(\bar{x}, 4^{-1}\Delta\bar{L}^{-1}), \tag{58}$$

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

$$0 > \left\langle ||v||^{-1}v, \bar{z} - x\right\rangle = \left\langle ||v||^{-1}v, \bar{x} + 4^{-1}\bar{L}^{-1}\Delta\|v\|^{-1}v - x\right\rangle. \tag{59}$$

This leads to

$$\left\langle ||v||^{-1}v, \bar{x} - x\right\rangle < -4^{-1}\bar{L}^{-1}\Delta. \tag{60}$$

Putting $y_0 := x - \alpha\|v\|^{-1}v$, we arrive at

$$\begin{aligned} \|y_0 - \bar{x}\|^2 &= \|x - \alpha\|v\|^{-1}v - \bar{x}\|^2 \\ &= \|x - \bar{x}\|^2 - 2\left\langle x - \bar{x}, \alpha\|v\|^{-1}v\right\rangle + \alpha^2 \\ &\leq \|x - \bar{x}\|^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2. \end{aligned} \tag{61}$$

From all the above we obtain

$$\begin{aligned}\|y-\bar{x}\|^2 &= \|P_{\Omega,w}(y_0)-\bar{x}\|^2 \le \|y_0-\bar{x}\|^2 \\ &\le \|x-\bar{x}\|^2 - 2\alpha(4\bar{L})^{-1}\Delta + \alpha^2,\end{aligned} \tag{62}$$

which completes the proof of the lemma. ■

Now we present the proof of Theorem 9.

Proof of Theorem 9. Fix an $\bar{x} \in SOL(\phi, C)$. It is not difficult to see that there exists a number $\varepsilon_0 \in (0, \varepsilon/4)$ such that for each $x \in R^J$ satisfying $d(x, C) \le \varepsilon_0$ and $\phi(x) \le \phi(\bar{x}) + \varepsilon_0$ we have

$$d(x, SOL(\phi, C)) \le \varepsilon/4. \tag{63}$$

Since $\{x^k\}_{k=0}^{\infty}$ is generated by Algorithm 4 and $\{(\Omega_k, w_k)\}_{k=1}^{\infty} \subset \mathcal{M}_{**}$ we know, from (16) and (9), that

$$\left\|x^k - P_{\Omega_k,w_k}(x^{k-1})\right\| \le \alpha_{k-1}, \text{ for all } k \ge 1, \tag{64}$$

holds. Thus, by Theorem 6 and (49), there exists an integer n_1 such that

$$d(x^k, C) \le \varepsilon_0, \text{ for all } k \ge n_1. \tag{65}$$

This, along with (23), guarantees that

$$\|x^k\| \le M + 1, \text{ for all } k \ge n_1. \tag{66}$$

Choose a positive ε_1 for which $\varepsilon_1 < (8\bar{L})^{-1}\varepsilon_0$. By (49) there is an integer $n_2 > n_1$ such that

$$\alpha_k \le \varepsilon_1(32)^{-1}, \text{ for all } k > n_2, \tag{67}$$

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

$$\sum_{k=n_2}^{n_0-1} \alpha_k > 8(2M+1)^2\bar{L}\varepsilon_0^{-1}. \tag{68}$$

We show now that there exists an integer $p \in [n_2 + 1, n_0]$ such that $\phi(x^p) \le \phi(\bar{x}) + \varepsilon_0$. Assuming the contrary means that for all $k \in [n_2 + 1, n_0]$,

$$\phi(x^k) > \phi(\bar{x}) + \varepsilon_0. \tag{69}$$

By (69), (49), (66) and using Lemma 10, with $\Delta = \varepsilon_0$, $\alpha = \alpha_k$, $x = x^k$, $y = x^{k+1}$, $v = s^k$, for all $k \in [n_2+1, n_0]$, we get

$$d(x^{k+1}, SOL(\phi, C))^2 \leq d(x^k, SOL(\phi, C))^2 - 2\alpha_k(4\bar{L})^{-1}\varepsilon_0 + \alpha_k^2. \tag{70}$$

According to the choice of ε_1 and by (67) this implies that for all $k \in [n_2 + 1, n_0]$,

$$\begin{aligned} d(x^k, SOL(\phi, C))^2 - d(x^{k+1}, SOL(\phi, C))^2 &\geq \alpha_k[(2\bar{L})^{-1}\varepsilon_0 - \alpha_k] \\ &\geq \alpha_k(4\bar{L})^{-1}\varepsilon_0, \end{aligned} \tag{71}$$

which, together with (66) and (23), gives

$$\begin{aligned} (2M+1)^2 &\geq d(x^{n_2+1}, SOL(\phi, C))^2 \\ &\geq \sum_{k=n_2+1}^{n_0} \left(d(x^k, SOL(\phi, C))^2 - d(x^{k+1}, SOL(\phi, C))^2 \right) \\ &\geq (4\bar{L})^{-1}\varepsilon_0 \sum_{k=n_2+1}^{n_0} \alpha_k \end{aligned} \tag{72}$$

and

$$\sum_{k=n_2+1}^{n_0} \alpha_k \leq (2M+1)^2 4\bar{L}\varepsilon_0^{-1}. \tag{73}$$

This contradicts (68), proving that there is an integer $p \in [n_2 + 1, n_0]$ such that $\phi(x^p) \leq \phi(\bar{x}) + \varepsilon_0$. Thus, by (65) and (63),

$$d(x^p, SOL(\phi, C)) \leq \varepsilon/4. \tag{74}$$

We show that for all $k \geq p$, $d(x^k, SOL(\phi, C)) \leq \varepsilon$. Assuming the contrary that

$$\text{there exists a } q > p \text{ such that } d(x^q, SOL(\phi, C)) > \varepsilon. \tag{75}$$

We may assume, without loss of generality, that

$$d(x^k, SOL(\phi, C)) \leq \varepsilon, \text{ for all } p \leq k < q. \tag{76}$$

One of the following two cases must hold: (i) $\phi(x^{q-1}) \leq \phi(\bar{x}) + \varepsilon_0$, or (ii) $\phi(x^{q-1}) > \phi(\bar{x}) + \varepsilon_0$. In case (i), since $p \in [n_2 + 1, n_0]$, (65), (66) and (63) show that

$$d(x^{q-1}, SOL(\phi, C)) \leq \varepsilon/4. \tag{77}$$

Thus, there is a point $z \in SOL(\phi, C)$ such that $\|x^{q-1} - z\| < \varepsilon/3$. Using this fact and (64), (9), (10) and (67), yields

$$\begin{aligned}\|x^q - z\| &\leq \left\|x^q - P_{\Omega_q, w_q}(x^{q-1})\right\| + \left\|P_{\Omega_q, w_q}(x^{q-1}) - z\right\| \\ &\leq \alpha_{q-1} + \left\|x^{q-1} - z\right\| \leq \varepsilon/4 + \varepsilon/3, \end{aligned} \tag{78}$$

proving that $d(x^q, SOL(\phi, C)) \leq \varepsilon$. This contradicts (75) and implies that case (ii) must hold, namely that $\phi(x^{q-1}) > \phi(\bar{x}) + \varepsilon_0$. This, along with (66), (67), the choice of ε_1, (76) and Lemma 10, with $\Delta = \varepsilon_0$, $\alpha = \alpha_{q-1}$, $x = x^{q-1}$, $y = x^q$, shows that

$$\begin{aligned} d(x^q, SOL(\phi, C))^2 &\leq d(x^{q-1}, SOL(\phi, C))^2 - 2\alpha_{q-1}(4\bar{L})^{-1}\varepsilon_0 + \alpha_{q-1}^2 \\ &\leq d(x^{q-1}, SOL(\phi, C))^2 - \alpha_{q-1}(2\bar{L})^{-1}\varepsilon_0 - \alpha_{q-1}) \\ &\leq d(x^{q-1}, SOL(\phi, C))^2 \leq \varepsilon^2, \end{aligned} \tag{79}$$

namely, that $d(x_q, C_{min}) \leq \epsilon$. This contradicts (75), proving that, for all $k \geq p$, $d(x^k, SOL(\phi, C)) \leq \varepsilon$. Together with (23) and (51) this implies that, for all $k \geq n_0$,

$$\phi(x^k) \leq \inf\{\phi(z) \mid z \in SOL(\phi, C)\} + \varepsilon\bar{L}\} \tag{80}$$

and the proof of Theorem 9 is complete.

Acknowledgments. We greatly appreciate the constructive comments of two anonymous reviewers which helped to improve the paper. The work of the first author was partially supported by the United States-Israel Binational Science Foundation (BSF) Grant number 200912 and US Department of Army Award number W81XWH-10-1-0170.

References

[1] R. Aharoni and Y. Censor, Block-iterative projection methods for parallel computation of solutions to convex feasibility problems. *Linear Algebra and its Applications* **120** (1989), 165–175.

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

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

[4] H.H. Bauschke, E. Matoušková and S. Reich, Projection and proximal point methods: convergence results and counterexamples, *Nonlinear Analysis: Theory, Methods & Applications* **56** (2004), 715–738.

[5] H.H. Bauschke and V.R. Koch, "Projection methods: Swiss army knives for Solving feasibility and best approximation problems with half-spaces", Technical report, January 2013.
http://arxiv.org/abs/1301.4506v1.

[6] D. Butnariu, R. Davidi, G.T. Herman, and I.G. Kazantsev, Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems, *IEEE Journal of Selected Topics in Signal Processing* **1** (2007), 540–547.

[7] D. Butnariu, S. Reich and A.J. Zaslavski, Stable convergence theorems for infinite products and powers of nonexpansive mappings, *Numerical Functional Analysis and Optimization* **29** (2008), 304–323.

[8] C.L. Byrne, *Applied Iterative Methods,* AK Peters, Wellsely, MA, USA, 2008.

[9] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Lecture Notes in mathematics 2057, Springer-Verlag, Berlin, Heidelberg, (2012).

[10] Y. Censor, W. Chen, P.L. Combettes, R. Davidi and G.T. Herman, On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints, *Computational Optimization and Applications*, accepted for publication. Available on: http://arxiv.org/abs/0912.4367.

[11] Y. Censor, R. Davidi and G.T. Herman, Perturbation resilience and superiorization of iterative algorithms, *Inverse Problems* **26** (2010), 065008 (12pp).

[12] Y. Censor, T. Elfving and G.T. Herman, Averaging strings of sequential iterations for convex feasibility problems. In: D. Butnariu, Y. Censor and S. Reich (editors), *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, Elsevier Science Publishers, Amsterdam, 2001, pp. 101–114.

[13] Y. Censor and A. Segal, On the string averaging method for sparse common fixed point problems, *International Transactions in Operational Research* **16** (2009), 481–494.

[14] Y. Censor and A. Segal, On string-averaging for sparse problems and on the split common fixed point problem, *Contemporary Mathematics* **513** (2010), 125–142.

[15] Y. Censor and E. Tom, Convergence of string-averaging projection schemes for inconsistent convex feasibility problems, *Optimization Methods and Software* **18** (2003), 543–554.

[16] Y. Censor and A.J. Zaslavski, Convergence and perturbation resilience of dynamic string-averaging projection methods, *Computational Optimization and Applications* **54** (2013), 65–76.

[17] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, NY, USA, 1997.

[18] J.W. Chinneck, *Feasibility and Infeasibility in Optimization: Algorithms and Computational Methods,* Springer, New York, NY, USA, 2007.

[19] G. Crombez, Finding common fixed points of strict paracontractions by averaging strings of sequential iterations, *Journal of Nonlinear and Convex Analysis* **3** (2002), 345–351.

[20] R. Davidi, G.T. Herman, and Y. Censor, Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections, *International Transactions in Operational Research* **16** (2009), 505–524.

[21] R. Escalante and M. Raydan, *Alternating Projection Methods*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 2011.

[22] A. Galántai, *Projectors and Projection Methods*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[23] D. Gordon and R. Gordon, Component-averaged row projections: A robust, block-parallel scheme for sparse linear systems, *SIAM Journal on Scientific Computing* **27** (2005), 1092–1117.

[24] E.S. Helou Neto and Á.R. De Pierro, Incremental subgradients for constrained convex optimization: A unified framework and new methods, *SIAM Journal on Optimization* **20** (2009), 1547–1572.

[25] E.S. Helou Neto and Á.R. De Pierro, On perturbed steepest descent methods with inexact line search for bilevel convex optimization, *Optimization* **60** (2011), 991–1008.

[26] G.T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Springer-Verlag, London, UK, 2nd Edition, 2009.

[27] G.T. Herman and R. Davidi, Image reconstruction from a small number of projections, *Inverse Problems* **24** (2008), 045011 (17pp).

[28] G.T. Herman, E. Garduño, R. Davidi and Y. Censor, Superiorization: An optimization heuristic for medical physics, *Medical Physics* **39** (2012), 5532–5546.

[29] K.C. Kiwiel, Convergence of approximate and incremental subgradient methods for convex optimization," *SIAM Journal on Optimization* **14** (2003), 807–840.

[30] T. Nikazad, R. Davidi and G.T. Herman, Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction, *Inverse Problems* **28** (2012), 035005 (19pp).

[31] A. Nedić, Random algorithms for convex minimization problems, *Mathematical programming*, Series B **129** (2011), 225–253.

[32] E.A. Nurminski, The use of additional diminishing disturbances in Fejer models of iterative algorithms, *Computational Mathematics and Mathematical Physics* **48** (2008), 2154–2161. Original Russian Text: E.A. Nurminski, published in: *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **48** (2008), 2121–2128.

[33] E.A. Nurminski, Fejer processes with diminishing disturbances, Doklady Mathematics, **78** (2008), 755–758. Original Russian Text: E.A. Nurminski, published in: *Doklady Akademii Nauk* **422** (2008), 601–604.

[34] E.A. Nurminski, Envelope stepsize control for iterative algorithms based on Fejer processes with attractants, *Optimization Methods and Software* **25** (2010), 97–108.

[35] E.A. Nurminski, Fejer algorithms with an adaptive step, *Computational Mathematics and Mathematical Physics* **51** (2011), 741–750. Original Russian Text: E.A. Nurminski, published in: *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **51** (2011), 791–801.

[36] S.N. Penfold, R.W. Schulte, Y. Censor, V. Bashkirov, S. McAllister, K.E. Schubert and A.B. Rosenfeld, Block-iterative and string-averaging projection algorithms in proton computed tomography image reconstruction. In: Y. Censor, M. Jiang and G. Wang (editors), *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems*, Medical Physics Publishing, Madison, WI, USA, 2010, pp. 347–367.

[37] S.N. Penfold, R.W. Schulte, Y. Censor and A.B. Rosenfeld, Total variation superiorization schemes in proton computed tomography image reconstruction, *Medical Physics* **37** (2010), 5887–5895.

[38] B.T. Polyak, Random algorithms for solving convex inequalities, in: *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, (D. Butnariu, Y. Censor, and S. Reich, eds.), Elsevier Sceience Publishers, Amsterdam, The Netherlands, 2001, pp. 409–422.

[39] S.S. Ram, A. Nedić and V.V. Veeravalli, Incremental stochastic subgradient algorithms for convex optimization, *SIAM Journal on Optimization* **20** (2009), 691–717.

[40] H. Rhee, An application of the string averaging method to one-sided best simultaneous approximation, *J. Korea Soc. Math. Educ. Ser. B: Pure Appl. Math.* **10** (2003), 49–56.