On the String Averaging Method for Sparse Common Fixed Points Problems

Yair Censor and Alexander Segal
({yair,asegal}@math.haifa.ac.il)

Department of Mathematics, University of Haifa
Mt. Carmel, Haifa 31905, Israel

January 16, 2008. Revised: October 14, 2008.

Abstract

We study the common fixed point problem for the class of directed operators. This class is important because many commonly used nonlinear operators in convex optimization belong to it. We propose a definition of sparseness of a family of operators and investigate a string-averaging algorithmic scheme that favorably handles the common fixed points problem when the family of operators is sparse. The convex feasibility problem is treated as a special case and a new subgradient projections algorithmic scheme is obtained.

1 Introduction

Given a finite family of operators $\{T_i\}_{i=1}^m$ acting on the Euclidean space R^n with $\operatorname{Fix} T_i \neq \emptyset$, $i = 1, 2, \ldots, m$, the *common fixed point problem* is to find a point

$$x^* \in \cap_{i=1}^m \operatorname{Fix} T_i, \tag{1}$$

where $\operatorname{Fix} T_i$ is the fixed points set of T_i. In this paper we study the common fixed point problem for sparse directed operators. We use the term *directed* *operators* for operators in the $\Im$-class of operators as defined and investigated by Bauschke and Combettes in [3] and by Combettes in [18]. Additionally, we focus on sparse operators and, for that purpose, we give a definition of sparseness of a family of operators.

The significance of working with this class stems from the fact that many commonly used types of nonlinear operators arising in convex optimization are directed operators (see, e.g., [3]) and, when developing algorithms for the problem (1) for such operators, we take advantage of their sparsity, whenever it exists.

The algorithms that are in use to find a common fixed point can be, from their structural view point, sequential, when only one operator at a time is used in each iteration, or simultaneous (parallel), when all operators in the given family are used in each iteration. There are algorithmic schemes which encompass sequential and simultaneous properties. These are the, so called, *string-averaging* [9] and *block-iterative projections* (BIP) [1], schemes, see also [15]. It turns out that the sequential and the simultaneous algorithms are special cases of the string-averaging and of the BIP algorithmic schemes.

Our objective here is to propose and study a string-averaging algorithmic scheme that enables *component-wise* weighting. Our work is a theoretical development aimed at gauging how far can the notions of sparsity, component-weighting and algorithmic string-averaging be expanded to cover the common fixed point problem for directed operators. The origins lie in [11] where a simultaneous projection algorithm, called *component averaging* (CAV), for systems of linear equations, that uses component-wise weighting was proposed. Such weighting enables, as shown and demonstrated experimentally on problems of image reconstruction from projections in [11], significant and valuable acceleration of the early algorithmic iterations due to the high sparsity of the system matrix appearing there. A block-iterative version of CAV, named BICAV, was introduced later in [12]. Full mathematical analyses of these methods, as well as their companion algorithms for linear inequalities, were presented by Censor and Elfving [10] and by Jiang and Wang [25]. In Section 2 we present preliminary material on directed operators and discuss some of their particular cases. In Section 3 we develop and study our string-averaging algorithmic scheme. In Section 4 we consider, as a special case, the convex feasibility problem and apply our algorithm from Section 3 using subgradient projectors.

1.1 Earlier work

The string-averaging algorithmic scheme has attracted attention recently and further work on it has been reported since its presentation in [9]. In [14] we investigated the behavior of string-averaging algorithms for inconsistent convex feasibility problems. In Bauschke, Matoušková and Reich [4] string-averaging was studied in Hilbert space. In Crombez [19, 20] the string-averaging algorithmic paradigm is used to find common fixed points of certain paracontractive operators in Hilbert space. In Bilbao-Castro, Carazo, García and Fernández [6], an implementation of the string-averaging method to electron microscopy is reported. Butnariu, Davidi, Herman and Kazantsev [7] call a certain class of string-averaging methods the Amalgamated Projection Method and show its stable behavior under summable perturbations. The iterative procedure studied in Butnariu, Reich and Zaslavski [8, Sections 6 and 7] is also a particular case of the string-averaging method. In Rhee [27] the string-averaging scheme is applied to a problem in approximation theory.

The notion of sparseness is very well understood and used for matrices and, from there, the road to sparseness of the Jacobian (or generalized Jacobian) matrix as an indicator of sparseness of nonlinear operators is short, see, e.g., Betts and Frank [5]. Our definition of sparseness of operators does not require differentiability (or subdifferentiability) and generalizes those previous notions.

2 Directed operators

We recall the definitions and results on directed operators and their properties as they appear in Bauschke and Combettes [3, Proposition 2.4] and Combettes [18], which are also sources for further references on the subject. Let $\langle x, y\rangle$ and $\|x\|$ be the Euclidean inner product and norm, respectively, in R^n.

Given $x, y \in R^n$ we denote the half-space

$$H(x, y) := \{u \in R^n \mid \langle u - y, x - y\rangle \leq 0\}. \tag{2}$$

Definition 1 *An operator $T : R^n \to R^n$ is called* `directed` *if*

$$\operatorname{Fix} T \subseteq H(x, T(x)), \text{ for all } x \in R^n, \tag{3}$$

or, equivalently,

$$\textit{if } z \in \operatorname{Fix} T \textit{ then } \langle T(x) - x, T(x) - z \rangle \le 0, \textit{ for all } x \in R^n. \tag{4}$$

The class of directed operators is denoted by $\Im$. Bauschke and Combettes [3] defined the directed operators (although without using this name) and showed (see [3, Proposition 2.4]) (i) that the set of all fixed points of a directed operator T with nonempty $\operatorname{Fix} T$ is closed and convex because

$$\operatorname{Fix} T = \bigcap_{x \in R^n} H(x, T(x)), \tag{5}$$

and (ii) that the following holds

$$\text{If } T \in \Im \text{ then } I + \lambda(T - I) \in \Im, \text{ for all } \lambda \in [0,1], \tag{6}$$

where I is the identity operator. The localization of fixed points is discussed in [23, pages 43-44]. In particular, it is shown there that a firmly nonexpansive operator, namely, an operator $N : R^n \to R^n$ that fulfills

$$\|N(x) - N(y)\|^2 \le \langle N(x) - N(y), x - y \rangle, \text{ for all } x, y \in R^n, \tag{7}$$

satisfies (5) and is, therefore, a directed operator. The class of directed operators, includes additionally, according to [3, Proposition 2.3], among others, the resolvents of a maximal monotone operators, the orthogonal projections and the subgradient projectors (see Example 7 below). Note that every directed operator belongs to the class of operators $\mathcal{F}^0$, defined by Crombez [21, p. 161], whose elements are called elsewhere quasi-nonexpansive or paracontracting operators.

The following definition of a closed operator will be required.

Definition 2 *An operator $T : R^n \to R^n$ is said to be* `closed` *at $y \in R^n$ if for every $\overline{x} \in R^n$ and every sequence $\{x^k\}_{k=0}^{\infty}$ in R^n, such that, $\lim_{k\to\infty} x^k = \overline{x}$ and $\lim_{k\to\infty} T(x^k) = y$, we have $T(\overline{x}) = y$.*

For instance, the orthogonal projection onto a closed convex set is everywhere a closed operator, due to its continuity.

Remark 3 *[18] If $T : R^n \to R^n$ is nonexpansive, then $T - I$ is closed on R^n.*

Consider a finite family $T_i : R^n \to R^n$, $i = 1, 2, \ldots, m$, of operators. In sequential algorithms for solving the common fixed point problem the order by which the operators are chosen for the iterations is determined by a *control sequence* of indices $\{i(k)\}_{k=0}^{\infty}$, see, e.g., [15, Definition 5.1.1].

Definition 4 *(i)* ***Cyclic control.*** *A control sequence is* `cyclic` *if* $i(k) = k \bmod m + 1$, *where* m *is the number of operators in the common fixed point problem.*

(ii) ***Almost cyclic control.*** $\{i(k)\}_{k=0}^{\infty}$ *is* `almost cyclic` *on* $\{1, 2, \ldots, m\}$, *if* $1 \leq i(k) \leq m$ *for all* $k \geq 0$, *and there exists an integer* $c \geq m$ *(called the* `almost cyclicality constant`*), such that, for all* $k \geq 0$, $\{1, 2, \ldots, m\} \subseteq \{i(k+1), i(k+2), \ldots, i(k+c)\}$.

The notions "cyclic" and "almost cyclic" are sometimes also called "periodic" and "quasi-periodic", respectively, see, e.g., [22].

Given a finite family $T_i : R^n \to R^n$, $i = 1, 2, \ldots, m$, of directed operators with a nonempty intersection of their fixed points sets, such that $T_i - I$ are closed at 0, for every $i \in \{1, 2, \ldots, m\}$. The following algorithm for finding a common fixed point of such a family is a special case of [18, Algorithm 6.1]. We will use it in the sequel.

Algorithm 5 ***Almost Cyclic Sequential Algorithm (ACSA) for solving common fixed point problem***

Initialization: $x^0 \in R^n$ *is an arbitrary starting point.*

Iterative Step: *Given* x^k, *compute* x^{k+1} *by*

$$x^{k+1} = x^k + \lambda_k (T_{i(k)} \left(x^k\right) - x^k). \tag{8}$$

Control: $\{i(k)\}_{k=0}^{\infty}$ *is almost cyclic on* $\{1, 2, \ldots, m\}$.

Relaxation parameters: $\{\lambda_k\}_{k=0}^{\infty}$ *are confined to the interval* $[0, 2]$.

The convergence theorem for Algorithm 5 for a finite family of directed operators is as follows.

Theorem 6 *Let* $\{T_i\}_{i=1}^{m}$ *be a finite family of directed operators* $T_i : R^n \to R^n$, *which satisfies*

(i) $\Omega := \cap_{i=1}^{m} \operatorname{Fix} T_i$ *is nonempty, and*

(ii) $T_i - I$ *are closed at* 0, *for every* $i \in \{1, 2, \ldots, m\}$.

Then any sequence $\left\{x^k\right\}_{k=0}^{\infty}$, *generated by Algorithm 5, converges to a point in* Ω.

Proof. This follows as a special case of [18, Theorem 6.6 (i)]. ■

In the next definition and lemma we recall the notion of the subgradient projector and show that this operator satisfies condition (ii) of Theorem 6.

Definition 7 *See, e.g., [3, Proposition 2.3(iv)]. Let $f : R^n \to R$ be a convex function such that the level-set $F := \{x \in R^n \mid f(x) \leq 0\}$ is nonempty. The operator*

$$\Pi_F(y) := \begin{cases} y - \dfrac{f(y)}{\|q\|^2} q, & \text{if} \quad f(y) > 0, \\ y, & \text{if} \quad f(y) \leq 0, \end{cases} \tag{9}$$

where q is a selection from the subdifferential set $\partial f(y)$ of f at y, is called a `subgradient projector relative to` *f.*

Lemma 8 *Let $f : R^n \to R$ be a convex function, let $y \in R^n$ and assume that the level-set $F \neq \emptyset$. For any $q \in \partial f(y)$, define the closed convex set*

$$L = L_f(y, q) := \{x \in R^n \mid f(y) + \langle q, x - y\rangle \leq 0\}. \tag{10}$$

Then the following hold:

(i) $F \subseteq L$. If $q \neq 0$ then L is a half-space, otherwise $L = R^n$.

(ii) Denoting by $P_L(y)$ the orthogonal projection of y onto L,

$$P_L(y) = \Pi_F(y). \tag{11}$$

(iii) $P_L - I$ is closed at 0.

Proof. For (i) and (ii) see, e.g., [2, Lemma 7.3]. (iii) Denote $\Psi = P_L - I$. Take any $\overline{x} \in R^n$ and any sequence $\{x^k\}_{k=0}^{\infty}$ in R^n, such that, $\lim_{k\to\infty} x^k = \overline{x}$ and $\lim_{k\to\infty} \Psi(x^k) = 0$. Since f is convex, its subdifferential is uniformly bounded on bounded sets, see, e.g., [2, Corollary 7.9]. Using this and the continuity of f we obtain, from (9), that $f(\overline{x}) = 0$, and, therefore, $\Psi(\overline{x}) = 0$. ■

3 The new string averaging algorithmic scheme

We study here a particular modification of the string averaging paradigm, adapted to handle the common fixed point problem for sparse directed operators.

3.1 The string averaging prototypical scheme

The string averaging prototypical scheme is defined as follows. Let the *string* S_p, for $p = 1, 2, \ldots, t$, be a finite, nonempty ordered subset of elements taken from $\{1, 2, \ldots, m\}$ of the form

$$S_p := \left\{ i_1^p, i_2^p, \ldots, i_{\gamma(p)}^p \right\}. \tag{12}$$

The length $\gamma(p)$ of the string S_p is the number of its elements. We do not require that the strings $\{S_p\}_{p=1}^t$ should be disjoint. Suppose that there is a set $Q \subseteq R^n$ such that there are operators $V_1, V_2, \ldots, V_m$ mapping Q into Q and an operator V which maps $Q^t = Q \times Q \times \cdots \times Q$ into Q. Then the string averaging prototypical scheme is as follow.

Algorithm 9 ***The string averaging prototypical algorithmic scheme [9]***

Initialization: *$x^0 \in Q$ is an arbitrary starting point.*

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

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

$$M_p(x^k) := V_{i_{\gamma(p)}^p} \ldots V_{i_2^p} V_{i_1^p}(x^k), \tag{13}$$

(ii) and then calculate,

$$x^{k+1} = V(M_1(x^k), M_2(x^k), \ldots, M_t(x^k)). \tag{14}$$

For every $p = 1, 2, \ldots, t$, this algorithmic scheme applies to x^k successively the operators whose indices belong to the p-th string. This can be done in parallel for all strings and then the operator V maps all end-points onto the next iterate x^{k+1}. This is indeed an algorithm provided that the operators $\{V_i\}_{i=1}^m$ and V all have algorithmic implementations. In this framework we get a sequential algorithm by the choice $t = 1$ and $S_1 = \{1, 2, \ldots, m\}$ and a simultaneous algorithm by the choice $t = m$ and $S_p = \{p\}$, $p = 1, 2, \ldots, t$.

In our new algorithmic scheme we assume that a finite family of directed operators (see Definition 1) $\{T_i\}_{i=1}^m$ is given with $\cap_{i=1}^m \operatorname{Fix} T_i \neq \emptyset$. After applying the operators $\{T_i\}_{i=1}^m$ along strings, the end-points will be averaged not by taking a plain convex combination but by doing a, so called, component-averaging step. The component averaging principle, introduced for linear systems in [11], [12], is a useful tool for handling sparseness in the linear case.

3.2 Sparseness of operators and the new algorithm

To define sparseness of the set of operators $\{T_i\}_{i=1}^m$ we need to speak about zeros of the vectors $x - T_i(x)$.

Definition 10 *Let $T : R^n \to R^n$ be a directed operator. If $(x - T(x))_j = 0$, for all $x \notin \operatorname{Fix} T$ then j is called a* ***void*** *of T and we write $j = \operatorname{void} T$.*

For every $i \in \{1, 2, \dots, m\}$ define the following sets

$$Z_i := \{(i,j) \mid 1 \le j \le n,\ j = \operatorname{void} T_i\}, \tag{15}$$

i.e., Z_i contains all the pairs (i,j), such that $(x - T_i(x))_j = 0$, for all $x \notin \operatorname{Fix} T_i$.

Definition 11 *The family of directed operators $\{T_i\}_{i=1}^m$ will be called* ***sparse*** *if the set $Z := \cup_{i=1}^m Z_i$ is nonempty and contains many elements.*

Remark 12 *The word "many" in Definition 11 is not well-defined. The more pairs (i,j) are contained in Z the higher is the sparseness of the family. It is of some interest to note that sparseness of matrices was considered as early as in 1971. Wilkinson [28, p. 191] refers to it by saying: "We shall refer to a matrix as dense if the percentage of zero elements or its distribution is such as to make it uneconomic to take advantage of their presence". Obviously, denseness is meant here as an opposite of sparseness.*

Denote by I_j, $1 \le j \le n$, the set of indices of strings that contain an index of an operator T_i for which $(i,j) \notin Z_i$, i.e.,

$$I_j := \{p \mid 1 \le p \le t, (i,j) \notin Z_i \text{ for some } i \in S_p\} \tag{16}$$

and let $s_j = |I_j|$ (the cardinality of I_j). Equivalently,

$$I_j = \{p \mid 1 \le p \le t,\ j \ne \operatorname{void} T_i \text{ for some } i \in S_p\}. \tag{17}$$

Definition 13 *[24, Definition 1] The* ***component-wise string averaging operator relative to the family of strings*** *$S := \{S_1, S_2, \dots, S_t\}$ is a mapping $CA_S : R^{n\times t} \to R^n$, defined as follows. For $x^1, x^2, \dots, x^t \in R^n$,*

$$\left(CA_S(x^1, x^2, \dots, x^t)\right)_j := (1/s_j) \sum_{p \in I_j} x_j^p, \quad \textit{for all} \ \ 1 \le j \le n, \tag{18}$$

where x_j^p is the j-th component of x^p, for $1 \le p \le t$.

Our new scheme performs sequential steps within each of the strings of the family S and merges the resulting end-points by the component-wise string averaging operator (18) as follows.

Algorithm 14

Initialization: *$x^0 \in R^n$ is an arbitrary starting point and define an integer constant N, such that $N \geq m$.*

Iterative step: *Given x^k, compute x^{k+1} as follows:*

(i) For every $1 \leq p \leq t$ (possibly in parallel): Execute a finite number, not exceeding N, of iterative steps of the form (8), on the operators $\{T_i\}_{i\in S_p}$ of the p-th string and denote the resulting end-points by $\{\overline{x}^p\}_{p=1}^t$.

(ii) Apply

$$x^{k+1} = CA_S(\overline{x}^1, \overline{x}^2, \ldots, \overline{x}^t). \tag{19}$$

3.3 Convergence

For the proof of convergence of Algorithm 14 we need the following construction. From the family $\{T_i\}_{i=1}^m$ of directed operators in R^n we construct another family of directed operators in a higher-dimensional space R^s and a family of strings for those operators. For the new operators and new strings, the operators belonging to different strings do not share any common variables. Therefore, the parallel processing of the strings in R^n in (i) of Algorithm 14 is equivalent to performing sequential ACSA iterations on the new directed operators in R^s. Moreover, using ideas of Pierra's [26] formalization, we show that the component-wise string averaging step in (ii) of Algorithm 14 is equivalent to an orthogonal projection onto a certain subspace of R^s. Inspired by the construction in [24], this is done as follows.

We represent each I_j is explicitly as

$$I_j = \left\{p_{j,1}, p_{j,2}, \ldots, p_{j,s_j}\right\}, \tag{20}$$

which defines each double-indexed p in an obvious way. Let R^s be the s-dimensional Euclidean space, where $s = \sum_{j=1}^n s_j$, and denote the components of each $\mathbf{y} \in R^s$ by

$$\mathbf{y} = (\underbrace{y_{1,1}, y_{1,2}, \ldots, y_{1,s_1}}_{s_1 \text{ elements}}, \ldots, \underbrace{y_{n,1}, y_{n,2}, \ldots, y_{n,s_n}}_{s_n \text{ elements}}) = \{y_{j,\ell}\}_{j=1,\ell=1}^{n,s_j}. \tag{21}$$

Define a linear mapping

$$\delta : R^n \to R^s, \text{ by } \delta(x) = \delta(x_1, x_2, \dots, x_n) := \mathbf{y}, \tag{22}$$

where $y_{j,p_{j,1}} = y_{j,p_{j,2}} = \dots = y_{j,p_{j,s_j}} = x_j$ for $j = 1, 2, \dots, n$. Let $\mathbf{D}$ be the range of δ, i.e.,

$$\mathbf{D} := \left\{ \mathbf{d} \in R^s \mid \mathbf{d} = (\underbrace{d_1, d_1, \dots, d_1}_{s_1 \text{ times}}, \dots, \underbrace{d_n, d_n, \dots, d_n}_{s_n \text{ times}}) \right\}, \tag{23}$$

which is a subspace of R^s. Define γ new operators where $\gamma = \sum_{p=1}^{t} \gamma(p)$ in the following manner. For each p, $\{i_w^p\}_{w=1}^{\gamma(p)}$ are the indices of the operators T_i that are included in the string S_p, see (12). To each pair (i_w^p, p) we attach a new operator $\mathbf{T}_{i_w^p,p} : R^s \to R^s$, defined by

$$\mathbf{T}_{i_w^p,p}(\mathbf{y}) := U_p(T_{i_w^p}(\Pi_p(\mathbf{y}))), \tag{24}$$

where the operators in the right-hand side of (24) are defined as follows. $\Pi_p : R^s \to R^n$, $1 \leq p \leq t$, is defined component-wise for each $1 \leq j \leq n$ as

$$(\Pi_p(\mathbf{y}))_j := \begin{cases} y_{j,l}, & \text{if } p_{j,l} = p, \text{ for some } \ell \in \{1, 2, \dots, s_j\}, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

$T_{i_w^p}$ is the w-th directed operator in the string S_p and $U_p : R^n \to R^s$, $1 \leq p \leq t$, is defined component-wise for each $1 \leq j \leq n$ and $1 \leq \ell \leq s_j$ as

$$(U_p(x))_{j,\ell} = \begin{cases} x_j, & \text{if } p_{j,\ell} = p, \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

The new operators $\mathbf{T}_{i_w^p,p}$ have fixed point sets $\operatorname{Fix} \mathbf{T}_{i_w^p,p} \subset R^s$. Each string $S_p = \left\{ i_1^p, i_2^p, \dots, i_{\gamma(p)}^p \right\}$ in R^n gives rise to a string

$$\mathbf{S}_p = \left\{ (i_1^p, p), (i_2^p, p), \dots, (i_{\gamma(p)}^p, p) \right\} \tag{27}$$

of the same length in R^s. Note, that operators $\mathbf{T}_{i_w^p,p}$ that belong to different strings in the family of strings $\{\mathbf{S}_p\}_{p=1}^{t}$ do not have a common variable which is not a void.

Lemma 15 *Every operator* $\mathbf{T}_{i_w^p,p}$, $1 \le p \le t$, $1 \le i_w^p \le \gamma(p)$ *is a directed operator and* $\mathbf{T}_{i_w^p,p} - \mathbf{I}$ *is closed at* 0*, where* $\mathbf{I}$ *is the identity operator in* R^s*.*

Proof. If $\mathbf{z} \in \operatorname{Fix} \mathbf{T}_{i_w^p,p}$ then $\mathbf{z} \in \operatorname{Im} U_p$, the image set of U_p. Moreover, then also $\Pi_p(\mathbf{z}) = z^* \in \operatorname{Fix} T_{i_w^p}$. For every $\mathbf{x} \in R^s$ we have

$$\begin{aligned}
&\left\langle \mathbf{T}_{i_w^p,p}(\mathbf{x}) - \mathbf{x}, \mathbf{T}_{i_w^p,p}(\mathbf{x}) - \mathbf{z} \right\rangle \\
&= \left\langle \Pi_p(\mathbf{T}_{i_w^p,p}(\mathbf{x})) - \Pi_p(\mathbf{x}), \Pi_p(\mathbf{T}_{i_w^p,p}(\mathbf{x})) - \Pi_p(\mathbf{z}) \right\rangle \\
&= \langle T_{i_w^p}(\Pi_p(\mathbf{x})) - \Pi_p(\mathbf{x}), T_{i_w^p}(\Pi_p(\mathbf{x})) - z^* \rangle \le 0,
\end{aligned} \tag{28}$$

since the operator $T_{i_w^p}$ is directed, therefore, (28) implies that $\mathbf{T}_{i_w^p,p}$ is also directed. Next, we show that $\mathbf{T}_{i_w^p,p} - \mathbf{I}$ is closed at 0. Let $\left\{\mathbf{x}^k\right\}_{k=1}^{\infty}$ be a sequence in R^s, such that $\lim_{k\to\infty} \mathbf{x}^k = \overline{\mathbf{x}}$ and $\lim_{k\to\infty} \left(\mathbf{T}_{i_w^p,p}(\mathbf{x}^k) - \mathbf{x}^k\right) = 0$. Since the operator Π_p is continuous, we obtain

$$\lim_{k\to\infty} \Pi_p(\mathbf{x}^k) = \Pi_p(\overline{\mathbf{x}}), \tag{29}$$

and

$$\lim_{k\to\infty} \left(T_{i_w^p}(\Pi_p(\mathbf{x}^k)) - \Pi_p(\mathbf{x}^k)\right) = 0. \tag{30}$$

The operator $T_{i_w^p} - I$ is closed at zero and, therefore,

$$T_{i_w^p}(\Pi_p(\overline{\mathbf{x}})) = \Pi_p(\overline{\mathbf{x}}). \tag{31}$$

Applying U_p to both sides of (31), we obtain

$$\mathbf{T}_{i_w^p,p}(\overline{\mathbf{x}}) = U_p(\Pi_p(\overline{\mathbf{x}})). \tag{32}$$

From

$$\overline{\mathbf{x}} = \lim_{k\to\infty} \mathbf{T}_{i_w^p,p}(\mathbf{x}^k) = U_p \cdot \left(\lim_{k\to\infty} T_{i_w^p}(\Pi_p(\mathbf{x}^k))\right), \tag{33}$$

follows that $\overline{\mathbf{x}} \in \operatorname{Im} U_p$ and therefore $U_p(\Pi_p(\overline{\mathbf{x}})) = \overline{\mathbf{x}}$. Then, from (32) one has that $\mathbf{T}_{i_w^p,p}(\overline{\mathbf{x}}) = \overline{\mathbf{x}}$ from which the closedness of $\mathbf{T}_{i_w^p,p} - \mathbf{I}$ follows. ■

Define the set

$$\mathbf{C} = \cap_{p=1,w=1}^{t,\ \gamma(p)} \operatorname{Fix} \mathbf{T}_{i_w^p,p}. \tag{34}$$

The mapping $\delta : R^n \to \mathbf{D}$ is a one-to-one mapping. Therefore, in the space R^s, we can reformulate the problem (1) as

$$\text{Find } \mathbf{y} \in \mathbf{C} \cap \mathbf{D}. \tag{35}$$

This means that

$$x \in \cap_{i=1}^{m} \operatorname{Fix} T_i \text{ if and only if } \mathbf{y} = \delta(x) \in \mathbf{C} \cap \mathbf{D}, \tag{36}$$

and, hence, the $m-$sets problem (1) is reduced to the 2-sets problem (35), which involves only a vector subspace and a convex set.

Next we present the alternative formulation of the Algorithm 14 in which the operations are performed in R^s.

Algorithm 16

Initialization:

(i) $x^0 \in R^n$ is arbitrary and define an integer constant N, such that $N \geq m$.

(ii) $\mathbf{y}^0 = \delta(x^0)$ is the initial vector in R^s.

Iterative step: *Given $\mathbf{y}^k$, compute $\mathbf{y}^{k+1}$ via:*

(i) In R^s, for every $1 \leq p \leq t$ (possibly in parallel): Execute a finite number, not exceeding N, of iterative steps of the form (8) on the operators $\left\{\mathbf{T}_{i_w^p,p}\right\}_{w=1}^{\gamma(p)}$ of the p-th string and denote the resulting end-points by $\{\overline{\mathbf{y}}^p\}_{p=1}^t$.

(ii) Apply CA_S in R^s as follows. For $1 \leq j \leq n$, set

$$y_{j,1}^{k+1} = \ldots = y_{j,s_j}^{k+1} = \frac{1}{s_j}\left(\sum_{\ell=1}^{s_j} \overline{y}_{j,\ell}^{p_{j,\ell}}\right). \tag{37}$$

(iii) Denote $\mathbf{y}^{k+1} := (y_{1,1}^{k+1}, y_{1,2}^{k+1}, \ldots, y_{1,s_1}^{k+1}, \ldots, y_{n,1}^{k+1}, y_{n,2}^{k+1}, \ldots, y_{n,s_n}^{k+1})$.

The following lemma shows that the averaging operation in the iterative step (ii) of Algorithm 16 is the orthogonal projection onto the subspace $\mathbf{D}$.

Lemma 17 *Let $\mathbf{y} = (y_{1,1}, y_{1,2}, \ldots, y_{1,s_1}, \ldots, y_{n,1}, y_{n,2}, \ldots, y_{n,s_n}) \in R^s$, then*

$$P_{\mathbf{D}}(\mathbf{y}) = \delta\Big(\frac{1}{s_1}\sum_{\ell_1=1}^{s_1} y_{1,\ell_1}, \frac{1}{s_2}\sum_{\ell_2=1}^{s_2} y_{2,\ell_2}, \ldots, \frac{1}{s_n}\sum_{\ell_n=1}^{s_n} y_{n,\ell_n}\Big). \tag{38}$$

Proof. Using the definition of the orthogonal projection we obtain

$$\begin{aligned} \|P_{\mathbf{D}}(\mathbf{y}) - \mathbf{y}\|^2 &= \min\left\{\|\delta(d) - \mathbf{y}\|^2 \mid d \in R^n\right\} \\ &= \min\left\{\sum_{j=1}^{n}\sum_{\ell=1}^{s_j} |d_j - y_{j,\ell}|^2 \mid d \in R^n\right\}. \end{aligned} \tag{39}$$

The minimum is obtained when the gradient is equal to zero,

$$\frac{\partial}{\partial d_j}\left(\sum_{j=1}^{n}\sum_{\ell=1}^{s_j}|d_j - y_{j,\ell}|^2\right) = 0, \text{ for all } j = 1, 2, \ldots n. \tag{40}$$

Then,

$$\sum_{\ell=1}^{s_j}(d_j - y_{j,\ell}) = 0, \text{ for all } j = 1, 2, \ldots n, \tag{41}$$

and

$$d_j = \frac{1}{s_j}\sum_{\ell=1}^{s_j} y_{j,\ell}, \text{for all } j = 1, 2, \ldots n, \tag{42}$$

and the proof is complete. ■

Now we are ready to prove our main convergence result.

Theorem 18 *If $\cap_{i=1}^{m} \operatorname{Fix} T_i \neq \emptyset$ then any sequence $\left\{x^k\right\}_{k=0}^{\infty}$, generated by the Algorithm 14, converges to a solution of (1).*

Proof. The consistency assumption on the problem (1) implies that (35) is also consistent. Moreover, Lemma 28 guarantees that all the operators are directed and that $\mathbf{T}_{i_w^p,p} - \mathbf{I}$, $1 \leq w \leq \gamma(p)$, $1 \leq p \leq t$ and $P_{\mathbf{D}} - \mathbf{I}$ are closed at 0. The Algorithm 16 can be executed in R^s in parallel or sequentially, since the strings do not contain any common non-void variables. Therefore, from Theorem 6 follows convergence to a common fixed point of the operators $\mathbf{T}_{i_w^p,p}$, $1 \leq w \leq \gamma(p)$, $1 \leq p \leq t$, and $P_{\mathbf{D}}$ and the proof is complete. ■

4 Special case: The convex feasibility problem

The *convex feasibility problem* (CFP) is to find a point x^* in the intersection C of m closed convex subsets $C_1, C_2, \ldots, C_m \subseteq R^n$. Each C_i is expressed as

$$C_i = \{x \in R^n \mid f_i(x) \leq 0\}, \tag{43}$$

where $f_i : R^n \to R$ is a convex function, so the CFP requires a solution of the system of convex inequalities

$$f_i(x) \leq 0, \quad i = 1, 2, \ldots, m. \tag{44}$$

The convex feasibility problem is a special case of the common fixed point problem, where the directed operators are the subgradient projectors relative to f_i (see, Example 7 and Lemma 8 above).

In a recent paper by Gordon and Gordon [24] a new parallel "Component-Averaged Row Projections (CARP)" method for the solution of large sparse linear systems was introduced. It proceeds by dividing the equations into nonempty, not necessarily disjoint, sets (strings), performing Kaczmarz row projections within the strings, and merging the results by component-averaging operations to form the next iterate. As shown in [24], using orthogonal projections onto convex sets, this method and its convergence proof also apply to the consistent nonlinear CFP.

In contrast, when applied to a CFP, our Algorithm 14 gives rise to a method which is structurally similar to CARP but uses subgradient projections instead of orthogonal projections. This is, of course, a development that might be very useful for CFPs with nonlinear convex sets for which each orthogonal projection mandates an inner-loop of distance optimization. We use now our results from Section 3 to present a string-averaging algorithm with component-wise averaging for a sparse CFP.

Sparseness of the nonlinear system (44) can be defined in compliance with Definitions 10 and 11 by speaking about zeros of the subgradients of the functions f_i and to do so we use the next definition.

Definition 19 *Let $f_i : R^n \to R$, $i = 1, 2, \ldots, m$, be convex functions. For any $x \in R^n$, the $m \times n$ matrix $Q(x) = (q_{ij})_{i=1,}^{m}{}_{j=1}^{n}$ is called* ***a generalized Jacobian of the family of functions*** *$\{f_i\}_{i=1}^{m}$* ***at the point*** *x if $q_{ij} \equiv q_j^i$, for all i and all j, for some $q^i = (q_j^i)_{j=1}^{n}$ such that $q^i \in \partial f_i(x)$.*

This definition coincides in our case with the Clarke's generalized Jacobian, see [16] and [17]. A generalized Jacobian $Q(x)$ of the functions in (44) is not unique because of the possibility to fill it up with different subgradients from each subdifferential set. In case all f_i are differentiable the generalized Jacobian reduces to the usual Jacobian.

We define for every $i \in \{1, 2, \ldots, m\}$ the following sets

$$Z_i := \{(i,j) \mid 1 \le j \le n,\ f_i(x) \text{ is independent of } x_j \text{ for all } x \in R^n\}. \quad (45)$$

A mapping $F : R^n \to R^m$ given by $F(x) = \{f_i(x)\}_{i=1}^{m}$ will be called *sparse* if some of its component functions f_i do not depend on some of their variables

x_j which means that $Z = \cup_{i=1}^m Z_i \neq \emptyset$. The more pairs (i,j) are contained in Z the higher is the sparseness of the mapping F.

Next we recall the cyclic subgradient projections (CSP) method for the CFP (studied in [13]) which is a special version of the ACSA algorithm (Algorithm 5).

Algorithm 20 ***Cyclic Subgradient Projections (CSP)***

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

Iterative step:

$$x^{k+1} := \begin{cases} x^k - \lambda_k \dfrac{f_{i(k)}(x^k)}{\left\| q^k \right\|^2} q^k, & \text{if } f_{i(k)}(x^k) > 0, \\ x^k, & \text{if } f_{i(k)}(x^k) \leq 0, \end{cases} \tag{46}$$

where $q^k \in \partial f_{i(k)}(x^k)$ is a subgradient of $f_{i(k)}$ at the point x^k.

Relaxation parameters: *$\{\lambda_k\}_{k=0}^{\infty}$ are confined to the interval $[\varepsilon, 2-\varepsilon]$, where $\varepsilon > 0$.*

Control: *Almost cyclic on $\{1, 2, \ldots, m\}$.*

According to our scheme the algorithm for solving the CFP performs CSP steps within the strings and merges the results by the $CA_S(x_1, x_2, \ldots, x_t)$ component-averaging operation.

Algorithm 21

Initialization: *$x^0 \in R^n$ is arbitrary and define an integer constant N, such that $N \geq m$.*

Iterative step: *Given x^k, compute x^{k+1} via:*

(i) For every $1 \leq p \leq t$ (possibly in parallel): Execute a finite number, not exceeding N, of CSP steps on the inequalities of the p-th string S_p and denote the resulting point by $\{\overline{\mathbf{x}}^p\}_{p=1}^t$.

(ii) Apply

$$x^{k+1} = CA_S(\overline{x}^1, \overline{x}^2, \ldots, \overline{x}^n). \tag{47}$$

Acknowledgments. This work was supported by grant No. 2003275 of the United States-Israel Binational Science Foundation (BSF) and by a National Institutes of Health (NIH) grant No. HL70472.

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. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Mathematics of Operations Research* **26**, 2001, 248–264.

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

[5] J.T. Bettes and P.D. Frank, A sparse nonlinear optimization algorithm, *Journal of Optimization Theory and Applications* **82** (1994), 519–541.

[6] J.R. Bilbao-Castro, J.M. Carazo, I. García and J.J. Fernández, Parallel iterative reconstruction methods for structure determination of biological specimens by electron microscopy, *Proceedings of The International Conference on Image Processing (ICIP)* (2003), vol. **1**, pp. I565–I568.

[7] D. Butnariu, R. Davidi, G.T. Herman and I.G. Kazansev, 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.

[8] 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.

[9] Y. Censor, T. Elfving, G.T. Herman, Averaging strings of sequential iterations for convex feasibility problems, in: *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, D. Butnariu, Y. Censor and S. Reich (Editors), Elsevier, Amsterdam, 2001, 101–113.

[10] Y. Censor and T. Elfving, Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem, *SIAM Journal on Matrix Analysis and Applications* **24** (2002), 40–58.

[11] Y. Censor, D. Gordon and R. Gordon, Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems, *Parallel Computing* **27** (2001), 777–808.

[12] Y. Censor, D. Gordon and R. Gordon, BICAV: A block-iterative, parallel algorithm for sparse systems with pixel-related weighting, *IEEE Transactions on Medical Imaging* **20** (2001), 1050–1060.

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

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

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

[16] F.H. Clarke, Generalized gradients and applications, *Transactions of the American Mathematical Society* **205** (1975), 247–262.

[17] F.H. Clarke, On the inverse function theorem, *Pacific Journal of Mathematics* **64** (1976), 97–102.

[18] P. L. Combettes, Quasi-Fejérian analysis of some optimization algorithms, *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, D. Butnariu, Y. Censor and S. Reich (Editors), Elsevier, Amsterdam, 2001, 115–152.

[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] G. Crombez, Finding common fixed points of a class of paracontractions, *Acta. Math. Hungar.* **103** (2004), 233–241.

[21] G. Crombez, A geometrical look at iterative methods for operators with fixed points, *Numerical Functional Analysis and Optimization* **26** (2005), 157–175.

[22] J.M. Dye and S. Reich, Unrestricted iterations of nonexpansive mappings in Hilbert space, *Nonlinear Analysis* **18** (1992), 199–207.

[23] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marsel Dekker, New York and Basel, 1984.

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

[25] M. Jiang and G. Wang, Convergence studies on iterative algorithms for image reconstruction, *IEEE Transactions on Medical Imaging* **22** (2003), 569–579.

[26] G. Pierra, Decomposition through formalization in a product space, *Mathematical Programming* **28** (1984), 96–115.

[27] 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.

[28] J.H. Wilkinson, Introduction to Part II, The Algebraic Eigenvalue Problem, *Handbook for Automatic Computation*, Vol II, Linear Algebra, J.H. Wilkinson and C. Reinsch (Editors), Springer-Verlag, 1971.