Int. J. Nonlinear Anal. Appl. 5 (2014) No. 2, 37-49
ISSN: 2008-6822 (electronic)
http://www.ijnaa.semnan.ac.ir

Weights in block iterative methods

Touraj Nikazad[a,*], Mokhtar Abbasi[a]

[a]Department of Mathematics, Iran University of Science and Technology, Tehran 16846-13114, Iran.

(Communicated by M. Eshaghi Gordji)

Abstract

In this paper we introduce a sequential block iterative method and its simultaneous version with optimal combination of weights (instead of convex combination) for solving convex feasibility problems. When the intersection of the given family of convex sets is nonempty, it is shown that any sequence generated by the given algorithms converges to a feasible point. Additionally for linear feasibility problems, we give equivalency of our algorithms with sequential and simultaneous block Kaczmarz methods explaining the optimal weights have been inherently used in Kaczmarz methods. In addition, a convergence result is presented for simultaneous block Kaczmarz for the case of inconsistent linear system of equations.

Keywords: block iterative methods, convex feasibility problem, Kaczmarz's method.
2010 MSC: Primary 65J22, 65J15; Secondary 65F10.

1. Introduction

A common problem in different areas of mathematics and physical sciences comprises of finding a point in the intersection of convex sets. This problem is referred to as the convex feasibility problem (CFP), see [2] for a general definition. We are interested to find a point in the nonempty intersection of a finite family of closed convex sets in the Euclidean space $\mathbb{R}^n$. Of special interest is the case with linear equations and/or inequalities, often referred to as the linear feasibility problem (LFP). Such linear systems may arise from discretization of an ill-posed problem such as the Radon transform used in modeling of several reconstruction problems, see, e.g., [18, 23]. Work related to the CFP are wide-ranging and numerous iterative methods for the CFP has been studied, see, e.g., [2, 5, 12] and references in [26]. Some problems that are modeled into CFP could be listed as discretized models of image reconstruction from projections, the fully discretized model of radiation-therapy

*Corresponding author
Email addresses: tnikazad@yahoo.com (Touraj Nikazad), m.abbasi78@gmail.com (Mokhtar Abbasi)

Received: February 2013 *Revised:* February 2014

treatment planning, and problems of image restoration. However, from such real-world applications, the resulting convex feasibility problem is often huge. Since direct factorization methods are inefficient for such a large-scale problem we employ iterative methods. Projection algorithms are successful iterative methods in the area of constructive solution of such problems see [5] and references therein. These kind of methods are formed by doing some projections onto the individual sets. Then, usually convex combination of projected results were used to produce the next iteration, see, e.g., [1]. The main result of our work allows, in some sense, to have the best combination instead of convex combination of projections. Also we observe that the simultaneous and sequential block Kaczmarz methods intrinsically employ the optimal weights resulted in faster convergence respect to other methods. In case of LFP, an extension of Kaczmarz method is recently achieved in [24].

In a simple way, one may classify projection methods as either sequential or simultaneous. The block iterative methods (in the image reconstruction literature is named ordered subsets methods [4, 11, 19]), which lie between sequential and simultaneous cases, have been studied in several works with different applications, see, e.g., [1, 4, 6, 7, 8, 13, 14, 15, 16, 25, 27, 29]. Simultaneous block iterative methods have also been used to increase computational efficiency using parallel processors [3, 8, 21]. For an overview in a more general framework, see [9]. In the case of LFP, we recall the simultaneous version of block Kaczmarz (it is same as row-Jacobi method [15]) which is suitable for parallel computing. In addition, we obtain that our algorithms and (sequential and simultaneous) block Kaczmarz methods are equivalent.

We now present a short summary of the contents of the paper. In Section 2 we introduce a sequential block iterative method and its parallel version, i.e., Algorithms 2.1 and 2.2. All convergence theorems are given in Section 3. Precisely, we demonstrate (from theoretical point of view) in Theorems 3.2 and 3.3, that the whole produced sequence by Algorithms 2.1 and 2.2 converge to a feasible point. In Section 4, we describe how the Algorithms 2.1 and 2.2 can be efficiently implemented for the LFP. Indeed the assumed solution point x^* is disappeared from the computation step. Also we remind sequential block Kaczmarz (for an excellent presentation see [23]) and its simultaneous version (row-Jacobi method, see [15]). Furthermore, it is shown that the generated cycles by Algorithms 2.1 and sequential block Kaczmarz (i.e., Algorithm 4.2 when $\lambda = 1$) and their parallel versions which are Algorithm 2.2 (with equal exterior weights $w^{k,s}$) and simultaneous block Kaczmarz (i.e., Algorithm 4.3 when $\lambda = 1$) are equivalent. For the inconsistent case, we conclude from above equivalency and [14] that the cycles of Algorithm 2.1 converge to a point which satisfies a certain linear system. Also it is shown the Algorithms 2.2 and 4.3 converge to a weighted least squares solution.

2. Preliminaries and algorithms

Let $B = \{1, 2, \cdots, m\}$, the index set, and let $C_{i\in B}$ be a finite family of closed convex sets in $\mathbb{R}^n$. The intersection $C = \bigcap_{i\in B} C_i$ is assumed to be nonempty. A block iterative method may be formed by partitioning of index set B into q subset B_t such that $B = \bigcup_{t=1}^{q} B_t$. The orthogonal projection of $x \in \mathbb{R}^n$ onto C_i is denoted by $P_i(x)$ which satisfies $P_i(x) = argmin_{y\in C_i}\|x - y\|$. Here $\langle x, y\rangle$ is the Euclidean inner product and $\|x\|$ the corresponding norm. A well-known property of projection operators is their non-expansivity, i.e., $\|P_i(x) - P_i(y)\| \leq \|x - y\|$ for any x and y in $\mathbb{R}^n$.

Also we use the standard terms sequential block iterative (SeqBI) and simultaneous block iterative (SimBI) from [9, 20]. Indeed, an iterative step sequentially moves from one block to the next one in SeqBI methods whereas SimBI methods use simultaneously a given starting point for each block and an iteration is produced by combination of all outcomes which are made by each block. Obviously, SimBI methods are more suitable than SeqBI methods for parallel computation, see [8, 9, 21].

Our schemes, SeqBI and SimBI methods, can now be described as follows.

Algorithm 2.1. SeqBI
Initialization: $x^0 \in R^n$ *is arbitrary.*
Iterative Step: *Given* x^k *compute,*

$$\begin{aligned} x^{k,0} &= x^k, \\ x^{k,s} &= x^{k,s-1} + \sum_{i \in B_s} w_i^{k,s} \left(P_i(x^{k,s-1}) - x^{k,s-1}\right), \ s = 1, \dots, q, \\ &s.t. \quad x^{k,s} = argmin_{x \in \Psi^{k,s}} \|x^* - x\|, \\ x^{k+1} &= x^{k,q}, \end{aligned}$$

where $w_i^{k,s}$ *are weights and* $\Psi^{k,s} = x^{k,s-1} + span\{P_i(x^{k,s-1}) - x^{k,s-1}\}_{i \in B_s}$.

In Algorithm 2.1, the step from k to $k+1$ is called a *cycle* and it consists of a sequence of sub-iterative steps (referred to as *atomic steps*). Each atomic step moves from $x^{k,s-1}$ to $x^{k,s}$.

The following algorithm is appropriately designed for parallel computations.

Algorithm 2.2. SimBI
Initialization: $y^0 \in R^n$ *is arbitrary.*
Iterative Step: *Given* y^k *compute,*

$$\begin{aligned} y^{k,0} &= y^k, \\ y^{k,s} &= y^k + \sum_{i \in B_s} w_i^{k,s} \left(P_i(y^k) - y^k\right), \ s = 1, \dots, q, \\ &s.t. \quad y^{k,s} = argmin_{y \in \Omega^{k,s}} \|y^* - y\|, \\ y^{k+1} &= \sum_{s=1}^{q} w^{k,s} y^{k,s}, \\ &s.t. \quad y^{k+1} = argmin_{y \in \Omega^k} \|y^* - y\|, \end{aligned}$$

where $w_i^{k,s}, w^{k,s}$ *are weights, and* $\Omega^{k,s} = y^k + span\{P_i(y^k) - y^k\}_{i \in B_s}, \Omega^k = span\{y^{k,s}\}_{s=1}^q$.

As seen, the starting point is used for each block in parallel. Then each block has its own result and the combination of such results makes the next iteration. Both Algorithms 2.1-2.2 encounter with an optimization problem and needs a solution point x^* in each step. Although this problem is resolved for the LFP, but the applicability of these methods to the nonlinear case is, however, unclear due to the nature of the error-measure.

The following theorem (see, e.g., [22]) and Lemma 2.4 guarantee the existence and uniqueness of the atomic step $x^{k,s}$ (of Algorithm 2.1) such that $\|x^* - x^{k,s}\|$ becomes minimum over $\Psi^{k,s}$ whereas $x^* \in \bigcap_{i \in B} C_i$.

Theorem 2.3. *Consider finite dimensional subspace* Y *of normed linear space* X *and an arbitrary point* $x \in X$*. Then there exists a point in* Y *which is the best approximation to* x*. Also to have a unique point in* Y *it is enough to consider* X *a Hilbert space.*

Lemma 2.4. *For any fixed* $x^* \in \bigcap_{i \in B} C_i$ *there exists a unique element* $x^{k,s} \in \Psi^{k,s}$ *such that* $x^{k,s} = argmin_{x \in \Psi^{k,s}} \|x^* - x\|$.

Figure 1: optimal weights versus convex combination of weights

Proof . Let $x \in \mathbb{R}^n$ be an arbitrary point. Using Theorem 2.3, there exists a unique point $\widetilde{z} = \sum_{i\in B_s} \widetilde{w}_i(P_i(x) - x) \in span\{P_i(x) - x\}_{i\in B_s}$ which minimizes $\|(x^* - x) - z\|$ over $z \in span\{P_i(x) - x\}_{i\in B_s}$. Now we assert $\widetilde{y} = \widetilde{z} + x$ is unique minimizer of $\|x^* - y\|$ over $y \in x + span\{P_i(x) - x\}_{i\in B_s}$. Let $\widehat{y} = x + \sum_{i\in B_s} \widehat{w}_i(P_i(x) - x) =: x + \widehat{z}$ such that $\|x^* - \widehat{y}\| \le \|x^* - \widetilde{y}\|$. From the last inequality we have $\|(x^* - x) - \widehat{z}\| \le \|(x^* - x) - \widetilde{z}\|$ which gives the desired result in the lemma. □

Similarly, one easily gets the same results for the Algorithm 2.2.

Remark 2.5. *Since, for the case of convex combination of weights, the convergence results of Algorithms 2.1 and 2.2 are known one may imagine that our convergence results with the optimal weights are not so surprising. But the figure 1 demonstrates reverse results. Indeed, it results* $\|x^* - x^{opt,1}\| \le \|x^* - x^1\|$ *whereas* $\|x^* - x^{opt,2}\| \ge \|x^* - x^2\|$ *here* $B_1 = \{1, 2\}$, $B_2 = \{3\}$. *Also* x^1, x^2 *and* $x^{opt,1}$, $x^{opt,2}$ *are computed using convex combination of weights and optimal weights respectively.*

3. Convergence of the block iterative methods

In the this section we derive convergence theorems for our algorithms. First we begin with a lemma which is used in our further proofs.

Lemma 3.1. *Let* $\{\beta^k\}$ *be a sequence in* $\mathbb{R}^n$ *and* $\alpha \in C_i$. *If*

$$\lim_{k\to\infty} \big|\|\alpha - \beta^k\| - \|\alpha - P_i(\beta^k)\|\big| = 0,$$

then

$$\lim_{k\to\infty} \|\beta^k - P_i(\beta^k)\| = 0.$$

Proof . For any $r_1,\ r_2,\ r_3 \in \mathbb{R}^n$ we have

$$\|r_1 - r_2\|^2 = \|r_1 - r_3\|^2 + \|r_2 - r_3\|^2 - 2\langle r_2 - r_3, r_1 - r_3\rangle,$$

which gets

$$\big|\|r_1 - r_2\|^2 - \|r_1 - r_3\|^2\big| = \big|\|r_2 - r_3\|^2 - 2\langle r_2 - r_3, r_1 - r_3\rangle\big|.$$

Putting $r_1 = \alpha,\ r_2 = \beta^k$ and $r_3 = P_i(\beta^k)$ provides

$$\lim_{k\to\infty} \big|\|\beta^k - P_i(\beta^k)\|^2 - 2\langle\beta^k - P_i(\beta^k), \alpha - P_i(\beta^k)\rangle\big| = 0.$$

Using metric projection characterization properties, see [2, FACTS 1.5] and [17, Section 3], one gets

$$-2\langle\beta^k - P_i(\beta^k), \alpha - P_i(\beta^k)\rangle \ge 0,$$

which shows

$$\lim_{k\to\infty} \|\beta^k - P_i(\beta^k)\| = 0.$$

□

Theorem 3.2. *If* $C \ne \emptyset$ *then the sequence of atomic steps* $\{x^{k,s}\}$ *in Algorithm 2.1 converges to a point* $x^* \in C$.

Proof .

Since $x^{k,s} = argmin_{x \in \Psi^{k,s}} \|x^* - x\|$, for every $i \in B_s$ we have

$$\begin{aligned}\|x^* - x^{k,s}\| &= \|x^* - x^{k,s-1} - \sum_{i \in B_s} w_i^{k,s} \left(P_i(x^{k,s-1}) - x^{k,s-1}\right) \| \\ &\le \|x^* - x^{k,s-1} - \left(P_i(x^{k,s-1}) - x^{k,s-1}\right) \| \\ &= \|x^* - P_i(x^{k,s-1})\|,\end{aligned}$$

which offers

$$\|x^* - x^{k,s}\| \le \min_{i \in B_s} \|x^* - P_i(x^{k,s-1})\|. \tag{3.1}$$

Non-expansivity of projection operator and (3.1) guarantee that for any $i \in B_s$ there exist scalars $0 \le \gamma_i^{k,s} \le 1$ such that

$$\begin{aligned}\|x^* - x^{k,s}\| &\le \|x^* - P_i(x^{k,s-1})\| \\ &= \|P_i(x^*) - P_i(x^{k,s-1})\| = \gamma_i^{k,s} \|x^* - x^{k,s-1}\|.\end{aligned} \tag{3.2}$$

Let

$$\delta^{k,s} = \min_{i \in B_s} \gamma_i^{k,s}, \ s = 1, \cdots, q. \tag{3.3}$$

From (3.2) and (3.3) one gets

$$\|x^* - x^{k,s}\| \le \delta^{k,s} \|x^* - x^{k,s-1}\|, \ s = 1, \cdots, q. \tag{3.4}$$

Repeating (3.4) gives

$$\begin{aligned}\|x^* - x^{k+1}\| = \|x^* - x^{k,q}\| &\le \delta^{k,q} \delta^{k,q-1} \cdots \delta^{k,1} \|x^* - x^{k,0}\| \\ &= \delta^{k,q} \delta^{k,q-1} \cdots \delta^{k,1} \|x^* - x^k\|.\end{aligned} \tag{3.5}$$

Let $\epsilon^k = \min_{s \in \{1, \cdots, q\}} \delta^{k,s}$, therefore (3.5) provides

$$\|x^* - x^{k+1}\| \le \epsilon^k \|x^* - x^k\|, \tag{3.6}$$

which shows

$$\lim_{k \to \infty} \|x^* - x^k\| = d. \tag{3.7}$$

Using (3.6) recursively gives

$$\|x^* - x^k\| \le \epsilon^{k-1} \epsilon^{k-2} \cdots \epsilon^0 \|x^* - x^0\|. \tag{3.8}$$

Now we consider two cases: first let the sequence $\{\epsilon^k\}$ have a subsequence $\{\epsilon^{k_r}\}$ such that $\lim_{k_r \to \infty} \epsilon^{k_r} = \alpha < 1$. Therefore $\prod_{r=1}^{\infty} \epsilon^{k_r} = 0$ which means $\prod_{k=1}^{\infty} \epsilon^k = 0$. Using (3.8) we get $\lim_{k \to \infty} \|x^k - x^*\| = 0$. The second case, the complimentary case of the first one, results in $\lim_{k \to \infty} \epsilon^k = 1$. By the definition of ϵ^k and $\delta^{k,s}$ one obtains

$$\lim_{k \to \infty} \gamma_i^{k,s} = 1, \ \text{for } i \in B_s. \tag{3.9}$$

Using (3.2) and (3.9) we get

$$\lim_{k \to \infty} \left| \|x^* - x^{k,s-1}\|^2 - \|x^* - P_i(x^{k,s-1})\|^2 \right| = 0, \quad \text{for } i \in B_s. \tag{3.10}$$

Lemma 3.1 and (3.10) show that

$$\lim_{k\to\infty} \|x^{k,s-1} - P_i(x^{k,s-1})\| = 0 \text{ for } i \in B_s. \tag{3.11}$$

Now we claim

$$\lim_{k\to\infty} \|x^{k} - P_i(x^{k})\| = 0 \text{ for } i \in B. \tag{3.12}$$

Put $\zeta^{k,s-1} = \sum_{j\in B_s} w_j^{k,s}\left(P_j(x^{k,s-1}) - x^{k,s-1}\right)$, then for $i \in B_s$ (no matter if $i \in B_\ell$ for $1 \le \ell \le q$)

$$\begin{aligned}\lim_{k\to\infty} \|x^{k,s} - P_i(x^{k,s})\| &= \lim_{k\to\infty} \|x^{k,s-1} + \zeta^{k,s-1} - P_i(x^{k,s-1} + \zeta^{k,s-1})\| \\ &\le \lim_{k\to\infty} \left\{\|x^{k,s-1} - P_i(x^{k,s-1} + \zeta^{k,s-1})\| + \|\zeta^{k,s-1}\|\right\}.\end{aligned} \tag{3.13}$$

The equality (3.11) and boundedness of weights $\{w_i^{k,s}\}$ result

$$\lim_{k\to\infty} \|\zeta^{k,s-1}\| \le \lim_{k\to\infty} \sum_{j\in B_s} |w_j^{k,s}|\; \|P_j(x^{k,s-1}) - x^{k,s-1}\| = 0. \tag{3.14}$$

Since a projection operator is nonexpansive, we have

$$\|P_i(x^{k,s-1} + \zeta^{k,s-1}) - P_i(x^{k,s-1})\| \le \|\zeta^{k,s-1})\|. \tag{3.15}$$

Using (3.11), (3.14) and (3.15), one concludes

$$\begin{aligned}&\lim_{k\to\infty} \|x^{k,s-1} - P_i(x^{k,s-1} + \zeta^{k,s-1})\| = \\ &= \lim_{k\to\infty} \|x^{k,s-1} - P_i(x^{k,s-1}) + P_i(x^{k,s-1}) - P_i(x^{k,s-1} + \zeta^{k,s-1})\|, \\ &\le \lim_{k\to\infty} \left\{\|x^{k,s-1} - P_i(x^{k,s-1})\| + \|P_i(x^{k,s-1}) - P_i(x^{k,s-1} + \zeta^{k,s-1})\|\right\}, \\ &\le \lim_{k\to\infty} \left\{\|x^{k,s-1} - P_i(x^{k,s-1})\| + \|\zeta^{k,s-1})\|\right\} = 0,\end{aligned} \tag{3.16}$$

for $i \in B_s$.

Therefore using (3.13) we get

$$\lim_{k\to\infty} \|x^{k,s} - P_i(x^{k,s})\| = 0 \text{ for } i \in B_s. \tag{3.17}$$

From (3.11) and (3.17) one finds (by induction)

$$\lim_{k\to\infty} \|x^{k,r} - P_i(x^{k,r})\| = 0, \text{ for } i \in B_s \text{ and } s-1 \le r \le q. \tag{3.18}$$

Since the set B_s is arbitrary, one easily gets the desired result in (3.12). Concerning (3.7), there exists the subsequence $\{x^{k_t}\}$ of $\{x^k\}$ that converges to a point $\bar{x}$. Using (3.12) we find

$$0 = \lim_{t\to\infty} \|x^{k_t} - P_i(x^{k_t})\| = \|\bar{x} - P_i(\bar{x})\| \text{ for } i = 1, 2, \ldots, m \tag{3.19}$$

which means $\bar{x} \in C$. Now, if we use $\bar{x}$ instead of x^*, one concludes $d = 0$ which shows that x^k converges to a feasible point $\bar{x}$. It is easy, using (3.5), to show that the whole sequence of the algorithm converges to $\bar{x}$. □

Next we explain a similar theorem for Algorithm 2.2.

Theorem 3.3. *If $C \neq \emptyset$ then the sequence of atomic steps $\{y^{k,s}\}$ in Algorithm 2.2 converges to a point $x^* \in C$.*

Proof .
The definition of $y^{k,s}$ in Algorithm 2.2 yields

$$\|x^* - y^{k,s}\| \leq \min_{i \in B_s} \|x^* - P_i(y^k)\| \text{ for } s = 1, \cdots, q. \tag{3.20}$$

Since a projection is nonexpansive mapping, there exist $0 \leq \gamma_i^{k,s} \leq 1$ such that

$$\|x^* - P_i(y^k)\| = \gamma_i^{k,s} \|x^* - y^k\|. \tag{3.21}$$

Let $\delta^{k,s} = \min_{i \in B_s} \gamma_i^{k,s}$. Using (3.20) and (3.21) we have $\|x^* - y^{k,s}\| \leq \delta^{k,s}\|x^* - y^k\|$ for $s = 1, \cdots, q$. Therefore

$$\begin{aligned} \|x^* - y^{k+1}\| = \|x^* - \sum_{s=1}^{q} w^{k,s} y^{k,s}\| &\leq \frac{1}{q} \sum_{s=1}^{q} \|x^* - y^{k,s}\| \\ &\leq \frac{1}{q} \sum_{s=1}^{q} \delta^{k,s} \|x^* - y^k\| = \epsilon^k \|x^* - y^k\|, \end{aligned} \tag{3.22}$$

where $0 \leq \epsilon^k = \frac{1}{q} \sum_{s=1}^{q} \delta^{k,s} \leq 1$. Obviously, we get the following two results from (3.22)

$$\lim_{k \to \infty} \|x^* - y^k\| = d, \tag{3.23}$$

and

$$\|x^* - y^{k+1}\| \leq \epsilon^k \cdots \epsilon^0 \|x^* - y^0\|. \tag{3.24}$$

Similar to proof of Theorem 3.2 two cases occur: First let the sequence $\{\epsilon^k\}$ have a subsequence $\{\epsilon^{k_r}\}$ such that $\lim_{k_r \to \infty} \epsilon^{k_r} = \alpha < 1$. Therefore $\prod_{r=1}^{\infty} \epsilon^{k_r} = 0$ which means $\prod_{k=1}^{\infty} \epsilon^k = 0$. Using (3.24) we get $\lim_{k \to \infty} \|x^k - x^*\| = 0$. The second case, the complimentary case of the first one, necessitates $\lim_{k \to \infty} \epsilon^k = 1$. This equality causes

$$\lim_{k \to \infty} \delta^{k,s} = 1, \quad \lim_{k \to \infty} \gamma_i^{k,s} = 1 \quad \text{for } s = 1, \cdots, q, \quad i \in B_s. \tag{3.25}$$

Using (3.21) and (3.25) we get

$$\lim_{k \to \infty} \left| \|x^* - y^k\|^2 - \|x^* - P_i(y^k)\|^2 \right| = 0, \quad \text{for } i \in B_s \tag{3.26}$$

Thus, the Lemma 3.1 and (3.26) give

$$\lim_{k \to \infty} \|y^k - P_i(y^k)\| = 0 \text{ for } i \in B_s, \quad s = 1, \cdots, q. \tag{3.27}$$

Regarding (3.23), there exists the subsequence $\{y^{k_t}\}$ of $\{y^k\}$ that converges to a point $\bar{x}$. The equality (3.27) implies $\|y^{k_t} - P_i(y^{k_t})\| \to 0$ which results $\|\bar{x} - P_i(\bar{x})\| = 0$ for $s = 1, \cdots, q, \quad i \in B_s$. It means that $\bar{x}$ is a feasible point. Now using $\bar{x}$ instead of x^* in (3.23) gets $d = 0$ which completes convergence proof of the sequence y^k. Using (3.22), one easily gets the convergence of whole sequence of Algorithm 2.2. □

4. Implementations and new results for LFP

Here we discuss how to efficiently implement Algorithms 2.1-2.2 for the following linear system of equations. Indeed the assumed solution point x^* is not needed, i.e., it disappears from the weight computation step. Let us consider the linear system of equations (which may be inconsistent)

$$Ax = b, \tag{4.1}$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Let $H_{i\in B} = \{x \in \mathbb{R}^n,\ \langle x, a^i\rangle = b_i\}$ here a^i and b_i indicate ith row of A and b respectively. Respect to partitioning of the index set B, the matrices A, b be partitioned into q (not necessarily disjoint) row blocks $\{A_t\}, \{b^t\}$ respectively. Obviously, the orthogonal projection of a point $x \in \mathbb{R}^n$ onto H_i can be calculated by

$$P_i(x) = x + \frac{b_i - \langle x, a^i\rangle}{\|a^i\|^2} a^i. \tag{4.2}$$

Both Algorithms 2.1-2.2 contain a norm minimization in each step. To minimize a convex function $f(x) = \|x^* - x\|$ over a proper subspace of $\mathbb{R}^n$, its gradient is used . Recall that a subgradient of a convex function f at y is any vector g that satisfies the inequality $f(x) - f(y) \geq \langle g, x - y\rangle$ for all x. If f is differentiable at y then its gradient $\nabla f(y)$ is the unique subgradient of f at y. Therefore, any y that satisfies $\nabla f(y) = 0$ is a global minimizer of f.

Remark 4.1. *In our algorithms, the linear system of equations $\nabla f(y) = 0$ may have many solutions. But any solution (weights) results in a unique iteration (atomic step/cycle), see Theorem 2.3 and Lemma 2.4.*

Since an optimization problem, in fact a small linear system of equations comparing with (4.1), must be solved in each step of Algorithms 2.1-2.2, we remind the block Kaczmarz method which is comparable with our methods from that point of view. The sequential block Kaczmarz method (also called block iterative ART) can be formulated as:

Algorithm 4.2. Seq. Block Kaczmarz
Initialization: $u^0 \in R^n$ *is arbitrary.*
Iterative Step: *Given* u^k *compute,*

$$\begin{aligned} u^{k,0} &= u^k, \\ u^{k,s} &= u^{k,s-1} + \lambda A_s^T (A_s A_s^T)^\dagger (b_s - A_s u^{k,s-1}),\ \ s = 1, \ldots, q, \\ u^{k+1} &= u^{k,q}, \end{aligned}$$

where $0 < \lambda < 2$ *is relaxation parameter and* $B^\dagger$ *shows pseudoinverse of* B.

This algorithm is a special case of [14, Algorithm 1.10]. The subsequences of cycles $\{u^{k,s}\}$, $k \geq 0$ with fixed s will actually converge. Indeed, if $b \in R(A)$ then $\{u^k\}$ converges toward a solution of $Ax = b$. If in addition $u^0 \in R(A^T)$, then $\{u^k\}$ converges to a point, see [14, Theorem 1.3]. Indeed this point satisfies a certain linear system of equations, see [16, Proposition 4]. For the special case $\lambda = 1$, we explain in Section 4.1 (Remark 4.4) why this algorithm and Algorithm 2.1 are equivalent.

Next we remind of the simultaneous block Kaczmarz method which is same as row-Jacobi method (first introduced in [15]).

Algorithm 4.3. Sim. Block Kaczmarz (Row-Jacobi)
Initialization: $v^0 \in R^n$ *is arbitrary.*
Iterative Step: *Given* v^k *compute,*

$$\begin{aligned} v^{k,0} &= v^k, \\ v^{k,s} &= v^k + \lambda A_s^T (A_s A_s^T)^\dagger (b_s - A_s v^k), \;\; s = 1, \dots, q, \\ v^{k+1} &= \frac{1}{q} \sum_{s=1}^{q} v^{k,s}, \end{aligned}$$

where $0 < \lambda < 2$ *is relaxation parameter.*

Indeed, if $b \in R(A)$ and $v^0 \in R(A^T)$ then the generated cycles of row-jacobi method (Algorithm 4.3) converges towards the solution of (4.1) with minimum norm, see [15, Theorem 1]. Similar to Algorithm 4.2, we obtain that this algorithm gives the same iterates as in Algorithm 2.2 when $\lambda = 1$ and the exterior weights (i.e., $w^{k,s}$) are equal, see Section 4.2, Remark 4.6.

Both Algorithms 2.1 and 2.2 can be written as the following matrix forms:

$$\begin{aligned} x^{k,0} &= x^k, \\ x^{k,s} &= x^{k,s-1} + A_s^T M_{k,s} (b_s - A_s x^{k,s-1}), \;\; s = 1, \dots, q, \\ &s.t. \quad x^{k,s} = argmin_{x \in \Psi^{k,s}} \|x^* - x\|, \\ x^{k+1} &= x^{k,q}, \end{aligned}$$

and

$$\begin{aligned} y^{k,0} &= y^k, \\ y^{k,s} &= y^k + A_s^T M_{k,s} (b_s - A_s y^k), \;\; s = 1, \dots, q, \\ &s.t. \quad y^{k,s} = argmin_{y \in \Omega^{k,s}} \|y^* - y\|, \\ y^{k+1} &= \sum_{s=1}^{q} w^{k,s} y^{k,s}, \\ &s.t. \quad y^{k+1} = argmin_{y \in \Omega^k} \|y^* - y\|, \end{aligned}$$

where $M_{k,s} = diag(\frac{w_i^{k,s}}{\|a^i\|^2})$ and $\{w_i^{k,s}\}_{i \in B_s}$ are related weights of each algorithm.

4.1. Algorithm 2.1

Let x^* be a solution of (4.1) and $z_i(x) = \frac{b_i - \langle a^i, x \rangle}{\|a^i\|^2}$. To find the weights in Algorithm 2.1, forming partial derivative of $\|x^* - x^{k,s}\|^2$ with respect to $\{w_j^{k,s}\}$ and setting them to zero result the following linear system of equations

$$\langle x^{k,s-1} + \sum_{i \in B_s} w_i^{k,s} \left(P_i(x^{k,s-1}) - x^{k,s-1} \right), z_j a^j \rangle = z_j b_j, \text{ for } j \in B_s, \tag{4.3}$$

thus, the equality (4.2) gives

$$\sum_{i \in B_s} w_i^{k,s} z_i \langle a^i, a^j \rangle = \left(b_j - \langle a^j, x^{k,s-1} \rangle \right), \text{ for } j \in B_s, \tag{4.4}$$

where $z_i := z_i(x^{k,s-1})$ for $i \in B_s$.

Next, by a simple calculation we demonstrate that Algorithms 2.1 and 4.2 become identical when $\lambda = 1$. In each step of Algorithm 4.2, the following linear system of equations should be solved

$$(A_s A_s^T)x = (b_s - A_s u^{k,s-1}), \tag{4.5}$$

where $u^{k,s-1}$ is computed in previous iteration. Also the weights in Algorithm 2.1 satisfy (4.4) which can be rewritten as the following matrix form

$$(A_s A_s^T)Dw = (b_s - A_s y^{k,s-1}), \tag{4.6}$$

where $D = diag(z_i)$ and $w = \left\{w_i^{k,s}\right\}_{i \in B_s}$. Note that we assumed $z_i \neq 0$, otherwise we do not need to compute $x^{k,s}$ since $b_i - \langle a^i, x^{k,s-1}\rangle = 0$. Therefore (4.5) and (4.6) show $x = Dw$. To compute next iteration in Algorithms 4.2 and 2.1 we have to calculate $u^{k,s-1} + A_s^T x$ and

$$\begin{aligned} x^{k,s-1} + \sum_{i \in B_s} w_i^{k,s}\left(P_i(x^{k,s-1}) - x^{k,s-1}\right) &= x^{k,s-1} + \sum_{i \in B_s} w_i^{k,s} z_i a^i \\ &= x^{k,s-1} + A_s^T D w, \end{aligned} \tag{4.7}$$

respectively. Now we can conclude

Remark 4.4. *Algorithms 2.1 and 4.2 (with $\lambda = 1$) are equivalent.*

Remark 4.5. *Using Remark 4.4 and [14, Theorem 1.3], the generated cycle in Algorithm 2.1 converges to a solution of (4.1) (Also, from Theorem 3.2, we can see that the whole generated sequence by this algorithm converges to a solution of (4.1)). And for inconsistent case, it converges to a point. Indeed this point satisfies a certain linear system of equations, see [16, Proposition 4], which do not correspond to a gradient mapping.*

4.2. Algorithm 2.2

In Algorithm 2.2, the weights $w_i^{k,s}$ and $w^{k,s}$ are called "*interior*" and "*exterior*" weights respectively. The complexity of finding the weights in Algorithms 2.1 and 2.2 are quite similar. Indeed, see (4.4), the interior weights $\{w_i^{k,s}\}$ satisfy the following equations for Algorithm 2.2

$$\sum_{i \in B_s} w_i^{k,s} z_i \langle a^i, a^j\rangle = \left(b_j - \langle a^j, y^k\rangle\right), \text{ for } j \in B_s, \; s = 1, \cdots, q. \tag{4.8}$$

Also exterior weights $w^{k,s}$ satisfy

$$\sum_{s=1}^{q} w^{k,s}\langle y^{k,s}, y^{k,t}\rangle = \langle x^*, y^k\rangle + \sum_{i \in B_t} w_i^{k,t} b_i z_i, \text{ for } t = 1, \cdots, q. \tag{4.9}$$

Let $k = 0$. If $y^0 \in R(A^T)$ then we can compute $\{w^{0,s}\}_{s=1}^q$ (x^* is disappeared from (4.9)) and $y^1 = \sum_{s=1}^q w^{0,s} y^{0,s}$. In the next iteration, $k = 1$, after computing $w_i^{1,s}$ from (4.8) we need $\langle x^*, y^1\rangle$ in

(4.9) to calculate $w^{1,s}$. But

$$\begin{aligned}\langle x^*, y^1\rangle &= \langle x^*, \sum_{s=1}^{q} w^{0,s} y^{0,s}\rangle \\ &= \sum_{s=1}^{q} w^{0,s}\langle x^*, y^{0,s}\rangle \\ &= \sum_{s=1}^{q} w^{0,s}\left\langle x^*, y^0 + \sum_{i\in B_s} w_i^{0,s}\left(P_i(y^0) - y^0\right)\right\rangle \\ &= \sum_{s=1}^{q} w^{0,s}\left(\langle x^*, y^0\rangle + \sum_{i\in B_s} w_i^{0,s} z_i(y^0) b^i\right), \end{aligned} \tag{4.10}$$

which means $\langle x^*, y^1\rangle$ can be computed without using x^*. By this way the whole generated sequence can be calculated devoid of the solution.

Similar conclusions as in Remark 4.4 can be drawn for Algorithms 2.2 and 4.3.

Remark 4.6. *The Algorithms 4.3 (for $\lambda = 1$) and 2.2 with equal exterior weights are equivalent.*

The following proposition provides a convergence proof for Algorithm 2.2 when the linear system (4.1) is inconsistent.

Proposition 4.7. *The sequence of atomic steps in Algorithm 2.2 converges to a solution of (4.1). In the inconsistent case, the generated cycle in Algorithm 2.2, with equal exterior weights, converges to a solution of a certain weighted least-squares problem.*

Proof .

The proof of the first part of the statement can be deduced from Theorem 3.3 (it was proven that the generated cycle converges to a solution of (4.1), see [15, Theorem 1]). For the inconsistence case, we use Remark 4.6 and rewrite Algorithm 4.3 as a fully simultaneous method. In Algorithm 4.3, the summation of $x^{k,s}$ over s furnishes

$$\begin{aligned}\sum_{s=1}^{q} x^{k,s} &= qx^k + \lambda\sum_{s=1}^{q} A_s^T(A_sA_s^T)^\dagger(b_s - A_sx^k) \\ &= qx^k + \lambda\begin{pmatrix}A_1\\ \vdots\\ A_q\end{pmatrix}^T\begin{pmatrix}(A_1A_1^T)^\dagger & & O\\ & \ddots & \\ O & & (A_qA_q^T)^\dagger\end{pmatrix}\begin{pmatrix}b_1 - A_1x^k\\ \vdots\\ b_q - A_qx^k\end{pmatrix} \\ &= qx^k + \lambda A^TM(b - Ax^k)\end{aligned} \tag{4.11}$$

where $M = diag((A_iA_i^T)^\dagger)$ is a symmetric positive semidefinite matrix. Therefore the cycle of this algorithm can be written as

$$x^{k+1} = x^k + \frac{\lambda}{q}A^TM(b - Ax^k). \tag{4.12}$$

Using classical theorems for full simultaneous iteration methods like (4.12), see, e.g., [28], one finds that the cycles $\{x^k\}$ converge to a solution of $\min\|Ax - b\|_M$ if $0 < \epsilon \le \lambda/q \le 2/\rho(A^TMA) - \epsilon$ where $\rho(Q)$ shows the spectral radius of Q. Thus, proof of the second part is completed by this fact that $\rho(A^TMA) = 1$. □

Remark 4.8. *The Remark 4.6 and Proposition 4.7 show that the sequence of atomic steps in Algorithm 4.3 (with $\lambda = 1$) converges to a solution of (4.1). For the inconsistent case, the generated cycle converges to a solution of a certain weighted least-squares problem when $0 < \epsilon \leq \lambda/q \leq 2 - \epsilon$.*

5. Conclusions

The sequential and simultaneous block iterative methods with optimal weights are proposed for solving convex feasibility problems. The whole sequences generated by both methods converge to a point in the feasible set. In the case of LFP, it is observed that the simultaneous and sequential block Kaczmarz methods inherently use the optimal weights. Additionally, it is demonstrated that the generated sequence by the simultaneous block Kaczmarz method converges to a weighted least squares solution.

Acknowledgements

The authors thank Professor Tommy Elfving for his valuable advices and discussions.

References

[1] R. Aharoni and Y. Censor, *Block-iterative projection methods for parallel computation of solutions to convex feasibility problems*, Linear Algebr. Appl., 120 (1989) 165–175.
[2] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996) 367–426.
[3] D. Butnariu, Y. Censor and S. Reich,(eds): Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications. Elsevier Science, Amsterdam, The Netherlands, 2001.
[4] C. Byrne, *Block-iterative interior point optimization methods for image reconstruction from limited data*, Inverse Problems, 16 (2000) 1405–1419.
[5] 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*, Comput. Optim. Appl., 51 (2012) 1065–1088.
[6] Y. Censor and T. Elfving, *Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem*, SIAM J. Matrix Anal. Appl., 24 (2002) 40–58.
[7] Y. Censor, T. Elfving, G. T. Herman and T. Nikazad, *On diagonally-relaxed orthogonal projection methods*, SIAM J. Sci. Comput., 30 (2007/08) 473–504.
[8] Y. Censor, D. Gordon and R. Gordon, *BICAV: an inherently parallel algorithm for sparse systems with pixel-dependent weighting*, IEEE Trans. Med. Imaging, 20 (2001) 1050–1060.
[9] Y. Censora and S. A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications*, Oxford: Oxford University Press, 1997.
[10] G. Cimmino, *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari*, La Ricerca Scientifica. XVI, Series II, Anno IX, 1 (1938) 326–333.
[11] P. L. Combettes and J. C. Pesquet, *Image restoration subject to a total variation constraint*, IEEE Trans. Image Process. 13 (2004) 1213–1222.
[12] P. L. Combettes, *Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections*, IEEE Trans. Image Process., 6 (1997) 493–506.
[13] R. Davidi, G. T. Herman and Y. Censor, *Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections*, Int. Trans. Oper. Res., 16 (2009) 505–524.
[14] P. P. B. Eggermont, G. T. Herman and A. Lent, *Iterative algorithms for large partitioned linear systems, with applications to image reconstruction*, Linear Algebra Appl., 40 (1981) 37–67.
[15] T. Elfving, it Block-iterative methods for consistent and inconsistent linear equations, Numer. Math., 35 (1980) 1–12.
[16] T. Elfving and T. Nikazad, *Properties of a class of block-iterative methods*, Inverse Problems, 25 (2009) 115011.
[17] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marcel Dekker, New York and Basel, 1984.
[18] G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, New York: Academic 1980.

[19] H. Hudson and R. S. Mand Larkin, *Accelerated image reconstruction using ordered subsets projection data*, IEEE Trans. Med. Imaging, 13 (1994) 601–609.

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

[21] C. Kamath and A. Sameh, *A projection method for solving nonsymmetric linear systems on multiprocessors*, Parallel Comput., 9 (1989) 291–312.

[22] E. Kreyszig, *Introductory functional analysis with applications*, Wiley Classics Library. John Wiley & Sons, Inc., New York, 1989.

[23] F. Natterer, *The Mathematics of Computerized Tomography*, New York: Wiley 2001.

[24] D. Needell and R. Ward, *Two-Subspace Projection Method for Coherent Overdetermined Systems*, Journal of Fourier Analysis and Applications, 19 (2)(2013) 256-269.

[25] 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.

[26] T. Nikazad and M. Abbasi, *An acceleration scheme for cyclic subgradient projections method*, Computational Optimization and Applications. (2012) DOI: 10.1007/s10589-012-9490-y .

[27] S. N. Penfold, R. W. Schulte, Y. Censor, V. Bashkirov, S. McAllister, K. E. Schubert and A. B. Rosenfeld, *Blockiterative and string-averaging projection algorithms in proton computed tomography image reconstruction* of *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems* eds. Y. Censor, M. Jiang and G. Wang, Madison, WI: Medical Physics Publishing, 2009.

[28] G. Qu, C. Wang and M. Jiang, *Necessary and sufficient convergence conditions for algebraic image reconstruction algorithms*, IEEE Trans. Image Process., 18 (2009) 435–440.

[29] H. D. Scolnik, A. R. De Pierro, N. E. Echebest and M. T. Guardarucci, *Special issue on block-iterative algorithms*, Int. Trans. Oper. Res., 16 (2009) 411–546.