Journal of Convex Analysis
Volume 14 (2007), No. 1, 169–183

Convergence of the Projected Surrogate Constraints Method for the Linear Split Feasibility Problems

Andrzej Cegielski
Faculty of Mathematics, Computer Science and Econometrics,
University of Zielona Góra, Poland
a.cegielski@wmie.uz.zgora.pl

Received: June 22, 2005
Revised manuscript received: October 10, 2005

The surrogate constraints method (SC-method) for linear feasibility problems (LFP) is an important tool in convex optimization, especially in large scale optimization. The classical version of the SC-method converges to a solution if the LFP is feasible [17]. Unfortunately, in applications the LFP is often infeasible. Such a situation occurs in computer tomography and in intensity modulated radiation therapy which can be modelled as LFP [5, 11, 13, 16]. In this case one can apply the simultaneous projection method (SP-method) [2, 4, 15] which is actually a short step version of a special case of the SC-method [7]. The SP-method converges to a solution if the LFP is feasible and to an approximate solution in other case. Because of long steps, the SC-method converges faster than the SP-method if the LFP is feasible. Unfortunately, the SC-method diverges if the problem is infeasible.

We deal in the paper with the linear split feasibility problems (LSFP) which are more general than the LFP. We analyze the convergence of various versions of the projected surrogate constraints method (PSC-method) for the LSFP in dependence on the step size and on the choice of weights. We show also that the convergence of the SC-method of Yang–Murty [17] and of the CQ-method of Byrne [5] applied to the LSFP follows from our main result.

Keywords: Split feasibility problem, surrogate constraints, Fejér monotonicity, convergence

2000 Mathematics Subject Classification: 65K05, 90C06, 90C25

1. Introduction

Let $C \subset \mathbb{R}^n$, $Q \subset \mathbb{R}^m$ be nonempty closed convex subsets, A be an $n \times m$ real matrix. The *split feasibility problem* (SFP) consists in finding x satisfying $A^\top x \in Q$, $x \in C$, if such a solution exists. This problem was introduced by Censor and Elfving [10] and was studied by Byrne [5]. In many applications the subset Q has the form $Q = \{y \in \mathbb{R}^m : y \leq b\}$ for $b \in \mathbb{R}^m$. In this case we call the problem the *linear split feasibility problem* (LSFP). The LSFP is to find an element $x^* \in C$ which is a solution of a system of linear inequalities

$$A^\top x \leq b. \tag{1}$$

If, furthermore, $C = \mathbb{R}^n$ then the LSFP is called the *linear feasibility problem* (LFP).

Denote $A = [a_1, ..., a_m]$, where $a_i \in \mathbb{R}^n$, $i \in I = \{1, ..., m\}$ are columns of A. We assume, without loss of generality, that $\|a_i\| = 1$, $i \in I$. Furthermore, denote $M_0 = \{x \in \mathbb{R}^n : A^\top x \leq b\}$ and $M = C \cap M_0$.

We deal in the paper with methods for the LSFP which generate a sequence (x_k) by the

ISSN 0944-6532 / \$ 2.50

following iterative scheme

$$x_1 \in C \text{ – arbitrary}$$
$$x_{k+1} = T(x_k), \tag{2}$$

where the operator $T : C \to C$ is defined by the equality

$$T(x) = P_C(x - \mu\gamma(x)Aw(x)), \tag{3}$$

P_C denotes the metric projection onto C, $\mu \in [0, 2]$, is called a *relaxation parameter*, $\gamma : C \to \mathbb{R}_+$ is called a *step size function* and $w : C \to \mathbb{R}^m_+$ is called a *weight operator*. Observe that $Aw(x)$ is a linear combination of the columns of A where the coefficients of this combination (weights) are coordinates of $w(x)$. Further we see that some relationship between the weight vector $w(x)$ and the residual vector $r(x) = A^\top x - b$ should be supposed in order to guarantee the convergence of iterative scheme (2). For $C = \mathbb{R}^n$ and for a special choice of the step size $\gamma(x)$ Yang and Murty [17] have called the method (2) the surrogate constraints method. Therefore, we call the general form of method (2) the *projected surrogate constraints method* (PSC-method). We study in the paper the Fejér monotonicity and the convergence of the PSC-method in dependence on step size γ and weights w. Recall that an operator $T : C \to C$ is *Fejér monotone* with respect to $D \subset C$ if

$$\|T(x) - z\| \leq \|x - z\|$$

for all $z \in D$, $x \in C$. In this case we say that the sequence (x_k) generated by the iterative scheme (2) and the method described by (2) are Fejér monotone with respect to D.

Consider the LSFP. Let $r : C \to \mathbb{R}^m$ be the *residual operator* defined by the equality $r(x) = A^\top x - b$. We call the vector $r(x)$ with coordinates $\rho_i(x)$, $i \in I$, the *residual vector* in x.

Definition 1.1. A weight operator $w : C \to \mathbb{R}^m_+$ satisfying the condition

$$w(x) = 0 \;\text{ if } r(x)^\top w(x) \leq 0 \tag{4}$$

for all $x \in C$ is called a *weight strategy (for r)*.

Remark 1.2. It follows from Definition 1.1 that a weight strategy should be seen as an operator related to LSFP (1) since it depends on the residual operator r.

The condition (4) is satisfied, e.g., if a vector of weights $w(x) \in \mathbb{R}^m_+$ is such that $\omega_i(x) = 0$ for $i \notin I(x)$, where

$$I(x) = \{i \in I : \rho_i(x) > 0\}$$

is the subset of violated constraints. Denote

$$H^-_{w(x)} = \{y \in \mathbb{R}^n : w(x)^\top A^\top y \leq w(x)^\top b\}.$$

Observe that

$$M \subset M_0 \subset H^-_{w(x)}, \tag{5}$$

i.e. all elements $y \in C$ which satisfy the constraints $A^\top y \leq b$ also satisfy the constraint $w(x)^\top A^\top y \leq w(x)^\top b$ for any weight $w(x) \geq 0$. Therefore it is called a *surrogate constraint*.

If $Aw(x) \neq 0$ then $H^-_{w(x)}$ is a half-space. If, furthermore, w is a weight strategy then it is easily seen that $x \notin H^-_{w(x)}$ (otherwise $r(x)^\top w(x) \leq 0$, consequently, $w(x) = 0$ which contradicts the assumption $Aw(x) \neq 0$) and

$$P_{H^-_{w(x)}}(x) = x - \frac{r(x)^\top w(x)}{\|Aw(x)\|^2} Aw(x).$$

Consequently, if we define

$$\gamma(x) = \begin{cases} \frac{r(x)^\top w(x)}{\|Aw(x)\|^2} & \text{if } Aw(x) \neq 0 \\ 0 & \text{if } Aw(x) = 0 \end{cases} \tag{6}$$

and w is a weight strategy then the operator T given by (3) can be written in the form

$$T(x) = P_C(x + \mu(P_{H^-_{w(x)}}(x) - x)). \tag{7}$$

If $C = \mathbb{R}^n$ then (7) describes the surrogate constraints method of Yang and Murty [17]. If $w(x) = r_+(x)$ for all $x \in C$ then the operator T given by (3) has the form

$$T(x) = P_C(x - \mu\gamma(x)Ar_+(x)). \tag{8}$$

Observe that $-r_+(x) = (P_Q - I)A^\top x$ for $Q = \{u \in \mathbb{R}^m : u \leq b\}$. Therefore, if we set $\gamma(x) = \frac{1}{\lambda}$ for all $x \in C$, where $\lambda = \lambda_{\max}(A^\top A) > 0$, then the operator T can be written in this case in the form

$$T(x) = P_C(x + \frac{\mu}{\lambda} A(P_Q - I)A^\top x), \tag{9}$$

i.e. T describes the CQ-method of Byrne [5] applied to LSFP. If $w(x) = Vr_+(x)$, where $V = \operatorname{diag} v$ for $v \in \Delta_m$, and $\gamma(x) = 1$ for all $x \in C$ then T has the form

$$T(x) = P_C(x - \mu AVr_+(x)). \tag{10}$$

In this case T can be written in the form

$$T(x) = P_C(x + \mu \sum_{i \in I} \nu_i (P_{H_i^-}(x) - x)), \tag{11}$$

where $H_i^- = \{y \in \mathbb{R}^n : a_i^\top y \leq \beta_i\}$, since

$$AVr_+(x) = \sum_{i \in I} \nu_i \rho_{i+}(x) a_i = \sum_{i \in I} \nu_i (P_{H_i^-}(x) - x)$$

(recall that $\|a_i\| = 1$, $i \in I$). Consequently, for $C = \mathbb{R}^n$, the operator T describes the simultaneous projection method [2, 15, 1, 4]. If $L = L(x) \subset I$ is such that $r_L(x) \neq 0$, A_L has full column rank and $(A_L^\top A_L)^{-1} r_L(x) \geq 0$ then, for $w(x) \in \mathbb{R}^m_+$ with $w_L(x) = (A_L^\top A_L)^{-1} r_L(x)$, $w_{I \setminus L} = 0$ and $\gamma(x) = 1$ for all $x \in C$, the operator T defined by (3) can be written in the form

$$T(x) = P_C(x - \mu A_L (A_L^\top A_L)^{-1} r_L(x))$$

(see [7] for details). If $C = \mathbb{R}^n$ then the operator T describes in this case the residual selection method [8, 9].

All methods mentioned above belong to a class of projection methods which are often used for the LFP, the LSFP, the SFP, the convex feasibility problems (CFP) and for the convex minimization problems (CMP). Relations between the PSC-method and the other projection methods presented above are studied in details in [7].

We study in the paper the convergence of iterative scheme (2) for T defined by (3), without assumption $M \neq \varnothing$. We prove the convergence to an element of $\operatorname{Fix} T$ under some general assumptions for weights which are satisfied for a wide class of projection methods. In Section 2, we formulate some equivalent conditions for z to be an element of $\operatorname{Fix} T$. In Section 3, we introduce two important conditions for weights, the residual orientation and the compatibility with $\operatorname{Fix} T$. As we see later, these conditions are sufficient for the convergence of the PSC-method. We give also examples of weight operators which satisfy these conditions. In Section 4, we prove the Fejér monotonicity of operator T determining the PSC-method, with respect to $\operatorname{Fix} T$, for residually directed weight strategy. In Section 5, we prove our main result – the convergence to $\operatorname{Fix} T$ of the PSC-method for residually directed weight strategies which are compatible with $\operatorname{Fix} T$. The convergence of SC-method of Yang–Murty [17] and of the CQ-method of Byrne applied to LSFP [5] follows from this result.

We use in the paper the following notation:

$x = (\xi_1, ..., \xi_n)^\top$ denotes an element of $\mathbb{R}^n$ as a column vector,

r_+ denotes the nonnegative part of a vector $r = (\rho_1, ..., \rho_m)^\top \in \mathbb{R}^m$, i.e. $r_+ = (\rho_{1+}, ..., \rho_{m+})^\top$ where $\rho_{i+} = \max\{0, \rho_i\}$,

A_L denotes the submatrix of A with columns a_i, $i \in L \subset I$,

r_L denotes the subvector of r with coordinates ρ_i, $i \in L \subset I$,

$r \geq s$, where $r = (\rho_1, ..., \rho_m)^\top \in \mathbb{R}^m$, $s = (\sigma_1, ..., \sigma_m)^\top \in \mathbb{R}^m$, denotes that $\rho_i \geq \sigma_i$ for all $i = 1, ..., m$,

$\mathbb{R}^m_+ = \{u \in \mathbb{R}^m : u \geq 0\}$,

$\langle \cdot, \cdot \rangle$ denotes the usual scalar product in $\mathbb{R}^n$, i.e. $\langle x, y \rangle = x^\top y$,

$\| \cdot \|$ denotes the Euclidean norm in $\mathbb{R}^n$, i.e. $\|x\| = \sqrt{\langle x, x \rangle}$,

$e = (1, ..., 1)^\top$,

$\Delta_m = \{u \in \mathbb{R}^m : e^\top u = 1, u \geq 0\}$ denotes the standard simplex in $\mathbb{R}^m$,

$P_D(x) = \operatorname{argmin}\{\|x - z\| : z \in D\}$ denotes the metric projection of $x \in \mathbb{R}^n$ onto a closed, convex subset $D \subset \mathbb{R}^n$,

$N_D(x) = \{s \in \mathbb{R}^n : \langle s, z - x \rangle \leq 0 \text{ for all } z \in D\}$ denotes the normal cone to D at $x \in D$,

$F_D(s) = \{y \in D : \langle s, y \rangle \geq \langle s, z \rangle \text{ for all } z \in D\}$ denotes the face of D exposed by $s \in \mathbb{R}^n$,

$\operatorname{Fix} T$ denotes the subset of fixed points of an operator $T : D \to D$, where $D \subset \mathbb{R}^n$,

$\lambda_{\max}(A^\top A)$ denotes the largest eigenvalue of a matrix $A^\top A$.

2. Fixed points of the operator T

Observe that $M \subset \operatorname{Fix} T$ if w is a weight strategy. Now we characterize the subset $\operatorname{Fix} T$. There holds the following Lemma.

Lemma 2.1. *Let the operator T be defined by* (3) *with $\gamma(x) > 0$ and $\mu > 0$ and let $z \in C$. The following conditions are equivalent*

(*i*) $z \in \operatorname{Fix} T$,
(*ii*) $-Aw(z) \in N_C(z)$,
(*iii*) $z \in F_C(-Aw(z))$
(*iv*) $\langle r(y) - r(z), w(z)\rangle \geq 0$ *for all* $y \in C$,
(*v*) $r(z) \in F_{r(C)}(-w(z))$.

If, furthermore, $Aw = \nabla h$ for a convex differentiable function $h : \mathbb{R}^n \to \mathbb{R}$ then conditions (*i*)-(*v*) *are equivalent with the condition*

(*vi*) $z \in \operatorname{Argmin}_{x \in C} h(x)$.

Proof. (*i*) $\Leftrightarrow$ (*ii*). $z \in \operatorname{Fix} T$ means that

$$z = P_C(z - \mu\gamma(z)Aw(z)).$$

Now the equivalence follows from the characterization of the normal cone (see, e.g., [14, Chapter III, Proposition 5.3.3]).

(*ii*) $\Leftrightarrow$ (*iii*). We have for $z \in C$

$$\begin{aligned} -Aw(z) &\in N_C(z) \Leftrightarrow \langle -Aw(z), y - z\rangle \leq 0 \text{ for all } y \in C \\ &\Leftrightarrow \langle -Aw(z), y\rangle \leq \langle -Aw(z), z\rangle \text{ for all } y \in C \\ &\Leftrightarrow z \in F_C(-Aw(z)). \end{aligned}$$

(*ii*) $\Leftrightarrow$ (*iv*). We have for $z \in C$

$$\begin{aligned} -Aw(z) &\in N_C(z) \Leftrightarrow \langle -Aw(z), y - z\rangle \leq 0 \text{ for all } y \in C \\ &\Leftrightarrow \langle w(z), A^\top y - A^\top z\rangle \geq 0 \text{ for all } y \in C \\ &\Leftrightarrow \langle w(z), r(y) - r(z)\rangle \geq 0 \text{ for all } y \in C. \end{aligned}$$

(*iv*) $\Leftrightarrow$ (*v*). We have for $z \in C$

$$\begin{aligned} \langle r(y) - r(z), w(z)\rangle &\geq 0 \text{ for all } y \in C \\ &\Leftrightarrow \langle u - r(z), w(z)\rangle \geq 0 \text{ for all } u \in r(C) \\ &\Leftrightarrow \langle r(z), -w(z)\rangle \geq \langle u, -w(z)\rangle \text{ for all } u \in r(C) \\ &\Leftrightarrow r(z) \in F_{r(C)}(-w(z)). \end{aligned}$$

(*ii*) $\Leftrightarrow$ (*vi*). Let $h : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function such that $Aw(x) = \nabla h(x)$, $x \in \mathbb{R}^n$. The equivalence follows from the characterization of a minimizer of a convex function (see, e.g., [14, Chapter VII, Theorem 1.1.1]). □

3. Assumptions for weights

The Fejér monotonicity and the convergence properties of iterative scheme (2) depend strongly on the weight strategy. Strategies which depend on the residuum $r(x) = A^\top x - b$ and satisfy the condition defined below play an important role here.

Definition 3.1. Let $w : C \to \mathbb{R}^m_+$ be a weight strategy in the PSC-method. We say that the strategy w is *residually directed* (or *residually orientated*) with respect to a subset $D \subset C$ if there exists a number $\rho > 0$, called the *modulus* of residual orientation, such that

$$\rho\langle r(x) - r(z), w(x) - w(z)\rangle \geq \|w(x) - w(z)\|^2 \tag{12}$$

for all $x \in C$ and for all $z \in D$.

Observe that the residual orientation of a strategy w does not depend on the scaling, i.e. if w is residually directed with modulus ρ then, for any $\alpha > 0$, the weight strategy αw is also residually directed with modulus $\alpha\rho$.

Remark 3.2. Let w be a residually directed weight strategy with respect to C. Then $w(x)$ depends actually on residuum $r(x)$, i.e. if $r(x) = r(y)$ then $w(x) = w(y)$ for $x, y \in C$. This fact follows directly from (12). Therefore, if $D = C$, then $w = u \circ r$ for some function $u : \mathbb{R}^m \to \mathbb{R}^m_+$ and we can write (12) in the form

$$\rho\langle r' - r'', u(r') - u(r'')\rangle \geq \|u(r') - u(r'')\|^2 \tag{13}$$

for all $r', r'' \in r(C)$. If $\rho = 1$ then inequality (13) denotes that u is firmly nonexpansive on $r(C)$.

Remark 3.3. Let w be a residually directed weight strategy with respect to C. Then one can easily see that for all $x, y \in C$

$$\|w(x) - w(y)\| \leq \rho\sqrt{\lambda}\|x - y\|$$

where ρ is the modulus of the residual orientation and $\lambda = \lambda_{\max}(A^\top A)$. Now we see that w is a Lipschitz continuous function.

Remark 3.4. If w is a residually directed weight strategy with respect to $D \subset C$ then

$$Aw(x) = Aw(y) \Longrightarrow w(x) = w(y)$$

for $x \in C$, $y \in D$. This fact follows easily from the inequalities

$$\begin{aligned} \rho\langle x - y, Aw(x) - Aw(y)\rangle &= \rho\langle r(x) - r(y), w(x) - w(y)\rangle \\ &\geq \|w(x) - w(y)\|^2. \end{aligned}$$

Lemma 3.5. *Let a weight strategy w be residually directed with respect to* $\operatorname{Fix} T$. *If $x, z \in \operatorname{Fix} T$ then $w(x) = w(z)$.*

Proof. Let $x, z \in \operatorname{Fix} T$. It follows from the equivalence $(i) \Leftrightarrow (iv)$ in Lemma 2.1 that for all $y \in C$

$$\langle r(y) - r(x), w(x)\rangle \geq 0$$

and

$$\langle r(y) - r(z), w(z)\rangle \geq 0.$$

If we set $y = z$ in the first inequality and $y = x$ in the second one then we obtain, by the residual orientation of w with respect to $\operatorname{Fix} T$,

$$\begin{aligned} 0 &\geq \langle r(x) - r(z), w(x)\rangle + \langle r(x) - r(z), -w(z)\rangle \\ &= \langle r(x) - r(z), w(x) - w(z)\rangle \\ &\geq \frac{1}{\rho}\|w(x) - w(z)\|^2 \geq 0, \end{aligned}$$

where $\rho > 0$ is a modulus of the residual orientation. Consequently, $w(x) = w(z)$. □

Corollary 3.6. *Let a weight strategy w be residually directed with respect to C. If $z \in \operatorname{Fix} T$ then*

$$\operatorname{Fix} T \subset F_C(-Aw(z)).$$

Proof. Let $z, x \in \operatorname{Fix} T$. Then we have, by the equivalence $(i) \Leftrightarrow (iii)$ in Lemma 2.1, $x \in F_C(-Aw(x))$. Consequently, $x \in F_C(-Aw(z))$ by Lemma 3.5. □

Definition 3.7. We say that a weight strategy w is *compatible* with $\operatorname{Fix} T$ if for any $z \in \operatorname{Fix} T$ and $x \in C$

$$w(x) = w(z) \Rightarrow x \in F_C(-Aw(z)).$$

In Examples 3.15 and 3.16 we present some conditions which guarantee the compatibility of a weight strategy with respect to $\operatorname{Fix} T$.

Lemma 3.8. *Let a weight strategy w be compatible with* $\operatorname{Fix} T$*. Then for arbitrary $z \in \operatorname{Fix} T$ and $x \in C$*

$$w(x) = w(z) \Rightarrow x \in \operatorname{Fix} T.$$

Proof. The Lemma follows directly from the equivalence $(i) \Leftrightarrow (iii)$ in Lemma 2.1. □

Corollary 3.9. *Let a weight strategy w be residually directed with respect to* $\operatorname{Fix} T$ *and compatible with* $\operatorname{Fix} T$*. Then for arbitrary $z \in \operatorname{Fix} T$ and $x \in C$*

$$w(x) = w(z) \Longleftrightarrow x \in \operatorname{Fix} T.$$

Proof. The equivalence follows from Lemmas 3.8 and 3.5. □

Remark 3.10. If w is residually directed, γ is a continuous function then, by Remark 3.3, T is continuous. If, furthermore, C is compact then if follows from the Brouwer fixed-point theorem that $\operatorname{Fix} T \neq \varnothing$.

Definition 3.11. We say that a weight operator $w : C \to \mathbb{R}^m_+$ is *weakly compatible* with M if

$$w(x) = 0 \Rightarrow x \in M$$

for all $x \in C$.

Lemma 3.12. *If a weight strategy w is weakly compatible with M then*

$$r(x)^\top w(x) \leq 0 \Rightarrow x \in M.$$

Proof. The Lemma follows immediately from Definitions 1.1 and 3.11. □

Corollary 3.13. *Let $M \neq \varnothing$. If w is a weight strategy which is weakly compatible with M then $M = \operatorname{Fix} T$.*

Proof. Inclusion $M \subset \operatorname{Fix} T$ follows directly from Definition 1.1 and from equality 3. Now we show that $\operatorname{Fix} T \subset M$. Suppose that $x \notin M$ for some $x \in \operatorname{Fix} T$. By inclusion (5) we have

$$w(x)^\top r(y) = w(x)^\top A^\top y - w(x)^\top b \leq 0$$

for all $y \in M$. On the other hand, by the equivalence $(i) \Leftrightarrow (iv)$ in Lemma 2.1, we have

$$w(x)^\top r(y) \geq w(x)^\top r(x)$$

for all $y \in C$. Consequently, $w(x)^\top r(x) \leq 0$, and $x \in M$, by Lemma 3.12, a contradiction. □

Remark 3.14. Consider a perturbed LSFP: find $x \in C$ such that $A^\top x \leq b + d$ with $d \in \mathbb{R}^m_+$. Denote by $r_d(x)$ the residual vector for the perturbed system, i.e. $r_d(x) = r(x) - d$. Let w be a weight strategy for the perturbed system. Then we have

$$r_d(x)^\top w(x) = r(x)^\top w(x) - d^\top w(x) \leq r(x)^\top w(x).$$

Consequently, w is a weight strategy for the original system. If, furthermore, w is residually directed with respect to a subset $D \subset C$ for the perturbed system then w is residually directed with respect to D for the original one since $r(x) - r(z) = r_d(x) - r_d(z)$. The same concerns the compatibility of w with $\operatorname{Fix} T$ since T does not depend on the perturbation vector d.

Now we give examples of a weight strategy and show their properties.

Example 3.15. Let $f_i : \mathbb{R}_+ \to \mathbb{R}_+$, $i \in I$, be arbitrary functions with properties

(i) $f_i(0) = 0$, $i \in I$,
(ii) f_i is nondecreasing, $i \in I$,
(iii) f_i is Lipschitz continuous with Lipschitz constant $\nu_i \geq 0$, $i \in I$.

Let $\omega_i : C \to \mathbb{R}_+$ be defined by the equality

$$\omega_i(x) = f_i(\rho_{i+}(x)),$$

$i \in I$, and let $w(x) = (\omega_1(x), ..., \omega_m(x))^\top$, $x \in C$. If we define $f : \mathbb{R}^m_+ \to \mathbb{R}^m_+$ as follows $f(t) = (f_1(\tau_1), ..., f_m(\tau_m))^\top$, where $t = (\tau_1, ..., \tau_m)^\top$, then we have $w(x) = f(r_+(x))$.

(a) We show that w is a weight strategy. We have, by property (i),

$$\begin{aligned} r(x)^\top w(x) &= \sum_{i \in I} \rho_i(x) f_i(\rho_{i+}(x)) \\ &= \sum_{i \in I} \rho_{i+}(x) f_i(\rho_{i+}(x)). \end{aligned}$$

If $r(x)^\top w(x) \leq 0$ then $\rho_{i+}(x) f_i(\rho_{i+}(x)) = 0$ for all $i \in I$ and, by property (i), $f_i(\rho_{i+}(x)) = 0$, $i \in I$, i.e. $w(x) = 0$.

(b) We show that w is residually directed with modulus $\rho = \nu_{\max} = \max_{i \in I} \nu_i$. By assumptions (i)-(iii), we have for all $x, z \in \mathbb{R}^n$

$$\begin{aligned}
& \nu_{\max}\langle w(x) - w(z), r(x) - r(z)\rangle \\
= \ & \nu_{\max} \sum_{i \in I} f_i(\rho_{i+}(x))\rho_i(x) + f_i(\rho_{i+}(z))\rho_i(z)) \\
& \quad -\nu_{\max} \sum_{i \in I} [f_i((\rho_{i+}(x))\rho_i(z) + f_i(\rho_{i+}(z))\rho_i(x)] \\
\geq \ & \nu_{\max} \sum_{i \in I} [f_i(\rho_{i+}(x))\rho_{i+}(x) + f_i(\rho_{i+}(z))\rho_{i+}(z)] \\
& \quad -\nu_{\max} \sum_{i \in I} [f_i(\rho_{i+}(x)\rho_{i+}(z) + f_i(\rho_{i+}(z))\rho_{i+}(x)] \\
= \ & \nu_{\max} \sum_{i \in I} [f_i(\rho_{i+}(x)) - f_i(\rho_{i+}(z))] \cdot [\rho_{i+}(x) - \rho_{i+}(z)] \\
= \ & \nu_{\max} \sum_{i \in I} |f_i(\rho_{i+}(x)) - f_i(\rho_{i+}(z))| \cdot |\rho_{i+}(x) - \rho_{i+}(z)| \\
\geq \ & \sum_{i \in I} \nu_i \, |f_i(\rho_{i+}(x)) - f_i(\rho_{i+}(z))| \cdot |\rho_{i+}(x) - \rho_{i+}(z)| \\
\geq \ & \sum_{i \in I} |f_i(\rho_{i+}(x)) - f_i(\rho_{i+}(z))|^2 \\
= \ & \|w(x) - w(z)\|^2.
\end{aligned}$$

(c) We show that w is compatible with $\operatorname{Fix} T$ if all f_i, $i \in I$, are increasing on $[0, +\infty)$. Suppose that $z \in \operatorname{Fix} T$. Let $x \in C$ and $w(x) = w(z)$, i.e. $f(r_+(x)) = f(r_+(z))$, $i \in I$. Consequently, $r_+(x) = r_+(z)$ since f_i are increasing for all $i \in I$. By assumption (i) and by the equivalence $(i) \Leftrightarrow (iv)$ in Lemma 2.1, we have for all $y \in C$

$$\begin{aligned}
\langle r(y) - r_+(z), f(r_+(z))\rangle &= \langle r(y) - r(z), f(r_+(z))\rangle \\
&= \langle r(y) - r(z), w(z)\rangle \geq 0.
\end{aligned}$$

and

$$\begin{aligned}
\langle r(y) - r(x), w(x)\rangle &= \langle r(y) - r(x), f(r_+(x))\rangle \\
&= \langle r(y) - r_+(z), f(r_+(z))\rangle \\
&\quad + \langle r_+(z) - r(x), f(r_+(x))\rangle \\
&\geq \langle r_+(z) - r(x), f(r_+(x))\rangle \\
&= \langle r_+(z) - r_+(x), f(r_+(x))\rangle = 0,
\end{aligned}$$

i.e. $x \in \operatorname{Fix} T \subset F_C(-Aw(z))$, by (b) and by Corollary 3.6.

(d) We show that w is weakly compatible with M if all f_i, $i \in I$, have the property

$$\tau > 0 \Rightarrow f_i(\tau) > 0. \tag{14}$$

Suppose that $x \in C$ and $w(x) = f(r_+(x)) = 0$. Then $r(x) \leq r_+(x) = 0$, by condition (14), i.e. $x \in M$.

(e) We show that r_+ is constant on $\operatorname{Fix} T$ and $\operatorname{Fix} T$ is a convex subset if all f_i, $i \in I$ are increasing on $[0, +\infty)$. Suppose that $x, z \in \operatorname{Fix} T$. Let $\alpha \in [0,1]$ and let $u = (1-\alpha)x + \alpha z$. By (b) and by Lemma 3.5,

$$f(r_+(x)) = w(x) = w(z) = f(r_+(z)).$$

Consequently, $r_+(x) = r_+(z)$ since all f_i, $i \in I$, are increasing. Now, it is easily seen that $r_+(u) = r_+(z)$ and that

$$w(u) = f(r_+(u)) = f(r_+(z)) = w(z).$$

Now, by (c) and by Lemma 3.8, we have $u \in \operatorname{Fix} T$.

Example 3.16. Let $v \in \mathbb{R}^m_+$ and let $V = \operatorname{diag} v$. Define

$$w(x) = V r_+(x). \tag{15}$$

If $v = \alpha e$ for $\alpha > 0$ then $w(x)$ is a vector of weights proportional to residua of violated constraints (such a strategy for the SC-method was introduced in [17]). If $C = \mathbb{R}^n$, $\gamma(x) = 1$ for all $x \in \mathbb{R}^n$ and $v \in \Delta_m$ then the operator T with weights given by (15) describes the simultaneous projection method (see equalities (10) and (11)). For $V = I$ and $\gamma = 1/\lambda$, where $\lambda = \lambda_{\max}(A^\top A)$, the operator T with weights given by (15) describes the CQ-method (see equalities (8) and (9)). Let $f_i(\tau) = \nu_i \tau$, $i \in I$. Then, of course, $w(x) = f(r_+(x))$. Observe that all f_i, $i \in I$, satisfy assumptions (i)-(iii) in Example 3.15. Therefore, w is a residually directed weight strategy with modulus $\rho = \nu_{\max} = \max_{i \in I} \nu_i$ with respect to $\mathbb{R}^n$. Furthermore, if $\nu_{\min} = \min_{i \in I} \nu_i > 0$ then w is compatible with $\operatorname{Fix} T$, weakly compatible with M and $\operatorname{Fix} T$ is convex.

We show that the compatibility of w with $\operatorname{Fix} T$ holds also without assumption $\nu_{\min} > 0$. Suppose that $z \in \operatorname{Fix} T$ and $x \in C$. Let $w(x) = w(z)$, i.e. $V r_+(x) = V r_+(z)$. For the convex function $h(x) = \frac{1}{2}\|V^{\frac{1}{2}} r_+(x)\|^2$ we have

$$\nabla h(x) = A V r_+(x).$$

Then

$$z \in \operatorname*{Argmin}_{y \in C} \frac{1}{2}\|V^{\frac{1}{2}} r_+(y)\|^2,$$

by the equivalence $(i) \Leftrightarrow (vi)$ in Lemma 2.1. Consequently,

$$x \in \operatorname*{Argmin}_{y \in C} \frac{1}{2}\|V^{\frac{1}{2}} r_+(y)\|^2$$

since $V r_+(x) = V r_+(z)$ if and only if $V^{\frac{1}{2}} r_+(x) = V^{\frac{1}{2}} r_+(z)$. Now, by the equivalence $(iii) \Leftrightarrow (vi)$ in Lemma 2.1,

$$x \in F_C(-A w(x)) = F_C(-A w(z)),$$

i.e. w is compatible with $\operatorname{Fix} T$.

If $V = I$ then $w(x) = r_+(x) = P_{\mathbb{R}^m_+}(r(x))$. In this case, the residual orientation of w with modulus 1 with respect to $\mathbb{R}^n$ follows directly from the firmly nonexpansivity of the metric projection (see [12, Chapter 12] or [3] for the definition and properties of a firmly nonexpansive operator).

Example 3.17. Define

$$f_i(\tau) = \alpha_i \ln(\beta_i \tau + 1),$$

for $\alpha_i, \beta_i \geq 0$, $i \in I$, and

$$w(x) = f(r_+(x))$$

for $f(t) = (f_1(\tau_1), ..., f_m(\tau_m))^\top$, where $t = (\tau_1, ..., \tau_m)^\top$. Observe that all f_i, $i \in I$, satisfy assumptions (i)-(iii) in Example 3.15 for $\nu_i = \alpha_i \beta_i$, $i \in I$. Therefore, w is a residually directed weight strategy with modulus $\rho = \nu_{\max} = \max_{i \in I} \nu_i$ with respect to $\mathbb{R}^n$. Furthermore, if $\alpha_i, \beta_i > 0$, $i \in I$, then all f_i, $i \in I$ are increasing, consequently, w is compatible with $\operatorname{Fix} T$, weakly compatible with M and $\operatorname{Fix} T$ is convex.

4. Fejér monotonicity

Lemma 4.1. *If w is residually directed with modulus $\rho > 0$ then the operator $F : \mathbb{R}^n \to \mathbb{R}^n$, $F = \frac{1}{\rho\lambda} Aw$, where $\lambda = \lambda_{\max}(A^\top A) > 0$, is firmly nonexpansive, i.e.*

$$\langle F(x) - F(y), x - y \rangle \geq \|F(x) - F(y)\|^2.$$

Proof. We have by the residual orientation of w

$$\begin{aligned}
\langle \frac{1}{\rho\lambda} Aw(x) - \frac{1}{\rho\lambda} Aw(y), x - y \rangle &= \frac{1}{\rho\lambda} \langle w(x) - w(y), A^\top x - A^\top y \rangle \\
&= \frac{1}{\rho\lambda} \langle w(x) - w(y), r(x) - r(y) \rangle \\
&\geq \frac{1}{\rho^2\lambda} \|w(x) - w(y)\|^2 \\
&\geq \frac{1}{\rho^2\lambda^2} \|Aw(x) - Aw(y)\|^2 \\
&= \|\frac{1}{\rho\lambda} Aw(x) - \frac{1}{\rho\lambda} Aw(y)\|^2.
\end{aligned}$$

□

Theorem 4.2. *Let T be defined by equality* (3)*, where $\gamma(x) = \frac{1}{\rho\lambda}$ for all $x \in \mathbb{R}^n$, $\rho > 0$ and $\lambda = \lambda_{\max}(A^\top A) > 0$. Furthermore, let the weight strategy w be residually directed with modulus ρ with respect to* $\operatorname{Fix} T$. *Then for all $z \in \operatorname{Fix} T$*

$$\|T(x) - z\|^2 \leq \|x - z\|^2 - \mu(2 - \mu) \frac{1}{\rho^2\lambda^2} \|Aw(x) - Aw(z)\|^2. \tag{16}$$

Consequently, T is Fejér monotone with respect to $\operatorname{Fix} T$ *for $\mu \in [0, 2]$.*

Proof. Let $z \in \operatorname{Fix} T$ and let $x \notin \operatorname{Fix} T$. Then we have, by the nonexpansivity of the metric projection and by the firmly nonexpansivity of the operator $F = \frac{1}{\rho\lambda} Aw$,

$$\begin{aligned}
\|T(x) - z\|^2 &= \|T(x) - T(z)\|^2 \\
&= \|P_C(x - \mu F(x)) - P_C(z - \mu F(z))\|^2 \\
&\leq \|(x - \mu F(x)) - (z - \mu F(z))\|^2 \\
&= \|x - z\|^2 + \mu^2 \|F(x) - F(z)\|^2 - 2\mu \langle x - z, F(x) - F(z)\rangle \\
&\leq \|x - z\|^2 + \mu^2 \|F(x) - F(z)\|^2 - 2\mu \|F(x) - F(z)\|^2 \\
&= \|x - z\|^2 - \mu(2 - \mu)\|F(x) - F(z)\|^2 \\
&= \|x - z\|^2 - \mu(2 - \mu)\frac{1}{\rho^2\lambda^2}\|Aw(x) - Aw(z)\|^2
\end{aligned}$$

□

If $M \neq \varnothing$ then we obtain the Fejér monotonicity of T with respect to M ($= \operatorname{Fix} T$, under assumptions of Corollary 3.13) without assumption on the residual orientation of the strategy of weights. The following Theorem is a generalization of [17, Theorem 3.1] for the SC-method.

Theorem 4.3. *Let $M \neq \varnothing$, w be any weight strategy and let $\gamma(x)$ be defined by* (6). *Then the operator T is Fejér monotone with respect to M for all $\mu \in [0, 2]$.*

Proof. Let $x \in \mathbb{R}^n$ and let $z \in M$. If $Aw(x) = 0$ then $T(x) = P_C(x)$ and, by the nonexpansivity of the metric projection,

$$\|T(x) - z\| = \|P_C(x) - P_C(z)\| \leq \|x - z\|.$$

Let now $Aw(x) \neq 0$. Denote $t(x) = P_{H^-_{w(x)}}(x) - x$. As we have observed in Section 1 the operator T defined by (3) for a weight strategy w and with the step size $\gamma(x)$ defined by (6) can be written in the form (7). Therefore, we have, by the nonexpansivity of the metric projection,

$$\begin{aligned}
\|T(x) - z\|^2 &= \|T(x) - z\|^2 \\
&= \|P_C(x - \mu \frac{r(x)^\top w(x)}{\|Aw(x)\|^2} Aw(x)) - P_C(z)\|^2 \\
&= \|P_C(x + \mu(P_{H^-_{w(x)}}(x) - x)) - P_C(z)\|^2 \\
&\leq \|x + \mu t(x) - z\|^2 \\
&= \|x - z\|^2 + \mu^2 \|t(x)\|^2 - 2\mu \langle z - x, t(x)\rangle.
\end{aligned}$$

It follows from the properties of the metric projection that

$$\langle z - x, t(x)\rangle \geq \|t(x)\|^2.$$

Consequently,

$$\|T(x) - z\|^2 \leq \|x - z\|^2 - \mu(2 - \mu)\|t(x)\|^2 \tag{17}$$

and T is Fejér monotone with respect to M. □

5. Convergence results

Consider the following iterative procedure which generates a sequence (x_k)

$$\begin{aligned} x_1 &\in C - \text{arbitrary} \\ x_{k+1} &= P_C(x_k - \mu\gamma_k A w_k) \end{aligned} \tag{18}$$

with $\mu \in (0,2)$, $\gamma_k \geq 0$ and $w_k \in \mathbb{R}^m_+$. We can also write $x_{k+1} = T(x_k)$, where T is defined by (3). Denote $r_k = r(x_k) = A^\top x_k - b$ and $w_k = w(x_k)$ for a weight strategy w. The following Theorem generalizes the results of Byrne [5, Proposition 2.1] in the case $Q = \{u \in \mathbb{R}^m : u \leq b\}$.

Theorem 5.1. *Let (x_k) be a sequence generated by iterative procedure (18), where $\gamma_k = \frac{1}{\rho\lambda}$, $\rho > 0$ and $\lambda = \lambda_{\max}(A^\top A) > 0$. Furthermore, let $w_k = w(x_k)$ for a weight strategy w which is residually directed with modulus ρ with respect to $\operatorname{Fix} T \neq \varnothing$ and compatible with $\operatorname{Fix} T$. Then the sequence (x_k) converges to an element of $\operatorname{Fix} T$.*

Proof. Let $z \in \operatorname{Fix} T$. There holds inequality (16) since all assumptions of Theorem 4.2 are satisfied. We write this inequality in the form

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 - \mu(2-\mu)\frac{1}{\rho^2\lambda^2}\|Aw(x_k) - Aw(z)\|^2. \tag{19}$$

We see that the sequence (x_k) is Fejér monotone with respect to $\operatorname{Fix} T$, hence (x_k) is bounded and $\|Aw(x_k) - Aw(z)\| \to 0$. Let $x^* \in C$ be a cluster point of (x_k). We have $Aw(x^*) = Aw(z)$, consequently, $w(x^*) = w(z)$ by Remark 3.4. By Lemma 3.8 we obtain $x^* \in \operatorname{Fix} T$ since w is compatible with $\operatorname{Fix} T$. If we set $z = x^*$ in equality (19) we obtain that the entire sequence converges to x^*. $\square$

Remark 5.2. Suppose that $Aw = \nabla h$ for a convex differentiable function $h : \mathbb{R}^n \to \mathbb{R}$. By the equivalence $(i) \Leftrightarrow (vi)$ in Lemma 2.1, h can be seen as a penalty function: finding a fixed point x is equivalent with the finding a point $x \in C$ with minimal penalty. Furthermore, if all conditions of Theorem 5.1 are satisfied, we obtain the convergence of the iterative procedure (18) to a minimizer of the penalty function $h\,|_C$. Such a situation occurs, e.g. if we set $w = Vr_+$, where $V = \operatorname{diag} v$ for $v \geq 0$ (see Example 3.16). The penalty function h has in this case the form

$$h(x) = \frac{1}{2}\|V^{\frac{1}{2}} r_+(x)\|^2 = \frac{1}{2}\sum_{i \in I} \nu_i (a_i^\top x - \beta_i)_+^2.$$

Denote $\|r\|_V := (r_+^\top V r)^{\frac{1}{2}}$ for $r \in \mathbb{R}^m$. We see, that iterative procedure (18) converges to a point $x \in C$ with minimal norm $\|r_+(x)\|_V$ of the nonnegative part of the residual vector $r(x)$. If $M \neq \varnothing$ then, of course, $\min_{x \in C} h(x) = 0$. The result presented in Theorem 5.1 shows that that the convergence to a point with minimal penalty holds also in the case $M = \varnothing$. Such a situation occurs often in practice in large scale linear split feasibility problems (see, e.g., [5, 11] for problems connected with computed tomography or [13, 16] for problems connected with the intensity modulated radiation therapy).

Remark 5.3. One can prove that if F is a firmly nonexpansive operator then for $\mu \in (0,2)$ the operators $I - \mu F$ and $P_C(I - \mu F)$ are averaged (see, e.g. [6, Lemma 2.1 and

Proposition 2.1]). By Lemma 4.1 the operator F satisfying the conditions of Theorem 5.1 is firmly nonexpansive. Therefore, Theorem 5.1 can be also proved by the application of Krasnoselskii–Mann Theorem. Similar technique was proposed by Byrne [6].

Remark 5.4. By the equivalence of (8) for $\gamma(x) = \frac{1}{\lambda}$ and (9), we see that the PSC-method with $w(x) = r_+(x)$ and with $\gamma(x) = \frac{1}{\lambda}$ for all $x \in C$ describes actually the CQ-method of Byrne. Since $w = r_+$ is residually directed weight strategy which is compatible with $\operatorname{Fix} T$ (see Example 3.16) the convergence of the CQ-method follows from Theorem 5.1.

Let now

$$\gamma_k = \begin{cases} \frac{r_k^\top w_k}{\|Aw_k\|^2} & \text{if } Aw_k \neq 0 \\ 0 & \text{if } Aw_k = 0, \end{cases} \tag{20}$$

where $r_k = r(x_k)$ and $w_k = w(x_k)$ for a weight strategy w. If, furthermore, $C = \mathbb{R}^n$ then (18) describes the SC-method. The following Theorem generalizes the results of Yang and Murty [17, Theorem 3.2].

Theorem 5.5. *Let $M \neq \varnothing$ and let (x_k) be a sequence generated by iterative procedure* (18)*, where γ_k is defined by* (20) *and w_k is generated by a weight strategy which is weakly compatible with M. Then the sequence (x_k) converges to an element of M.*

Proof. Let $z \in M$. We can write inequality (17) in the form

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 - \mu(2-\mu)\|P_{S_k}(x_k) - x_k\|^2,$$

where $S_k = H^-_{w_k} = \{y \in \mathbb{R}^n : w_k^\top A^\top y \leq w_k^\top b\}$. We see that $\|x_k - z\|$ converges, consequently (x_k) is bounded and $\|P_{S_k}(x_k) - x_k\| \to 0$. Now we obtain for any cluster point x^* of (x_k)

$$P_{H^-_{w(x^*)}}(x^*) = x^*,$$

i.e. $x^* \in H^-_{w(x^*)}$ or, in other words, $r(x^*)^\top w(x) \leq 0$. By Lemma 3.12, we have $x^* \in M$. $\square$

Remark 5.6. Yang and Murty [17, Theorem 3.2] have supposed $C = \mathbb{R}^n$ and have considered weights ω_i, $i \in I$ with property $\omega_i > \alpha > 0$ for all $i \in I(x_k)$. This assumption corresponds to the assumption that a vector of weights w has the property

$$\rho_i(x) > 0 \Rightarrow \omega_i(x) > \alpha > 0.$$

Observe that w is a weight strategy which is weakly compatible with respect to M if the above condition is satisfied.

References

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

[2] A. Auslender: Optimisation, Méthodes Numériques, Masson, Paris (1983).

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

[4] D. Butnariu, Y. Censor: On the behavior of a block-iterative projection method for solving convex feasibility problems, Int. J. Comp. Math. 34 (1990) 79–94.

[5] C. Byrne: Iterative oblique projection onto convex sets and the split feasibility problem, Inverse Probl. 18 (2002) 441–453.

[6] C. Byrne: A unified treatment of some iterative algorithms in signal processing and image reconstruction, Inverse Probl. 20 (2004) 103–120.

[7] A. Cegielski: Projection methods for the linear split feasibility problems, Optimization (2007), to appear.

[8] A. Cegielski, R. Dylewski: Selection strategies in projection methods for convex minimization problems, Discuss. Math., Differ. Incl. Control Optim. 22 (2002) 97–123.

[9] A. Cegielski, R. Dylewski: Residual selection in a projection method for convex minimization problems, Optimization 52 (2003) 211–220.

[10] Y. Censor, T. Elfving: A multiprojection algorithm using Bregman projections in a product space, Numer. Algorithms 8 (1994) 221–239.

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

[12] K. Goebel, W. A. Kirk: Topics in Metric Fixed Point Theory, Cambrigde University Press, Cambridge (1990).

[13] H. W. Hamacher, K.-H. Küfer: Inverse radiation therapy planning – a multiple objective optimization approach, Discrete Appl. Math. 118 (2002) 145–161.

[14] J.-B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms. Part 1: Fundamentals, Springer, Berlin (1993).

[15] A. R. De Pierro, A. N. Iusem: A simultaneous projections method for linear inequalities, Linear Algebra Appl. 64 (1985) 243–253.

[16] Y. Xiao, D. Michalski, Y. Censor, J. M. Galvin: Inherent smoothness of intensity patterns for intensity modulated radiation therapy generated by simultaneous projection algorithm, Phys. Med. Biol. 49 (2004) 3227–3245.

[17] K. Yang, K. G. Murty: New iterative methods for linear inequalities, J. Optimization Theory Appl. 72 (1992) 163–185.