Relaxed alternating projection methods

Andrzej Cegielski, Agnieszka Suchocka

Faculty of Mathematics, Computer Sciences
and Econometrics, University of Zielona Góra
65-516 Zielona Góra, ul. Szafrana 4a, Poland
e-mail: a.cegielski@wmie.uz.zgora.pl
e-mail: a.suchocka@wmie.uz.zgora.pl

June 11, 2008

Abstract

In this paper we deal with the von Neumann alternating projection
method \(x_{k+1} = P_A P_B x_k\) and with its generalization of the form
\(x_{k+1} = P_A (x_k + \lambda_k (P_A P_B x_k - x_k))\), where \(A, B\) are closed and convex subsets
of a Hilbert space \(\mathcal{H}\) and \(\text{Fix } P_A P_B \neq \emptyset\). We do not suppose that
\(A \cap B \neq \emptyset\). We give sufficient conditions for the weak convergence of
the sequence \((x_k)\) to \(\text{Fix } P_A P_B\) in the general case and in the case \(A\)
is a closed affine subspace. We present also the results of preliminary
numerical experiments.

Key words: alternating projection method, Fejér monotonicity, weak
convergence

AMS Subject Classification: 65K05

1. Introduction

Let \(\mathcal{H}\) be a real Hilbert space equipped with a scalar product \(\langle \cdot, \cdot \rangle\)
and with the norm \(\| \cdot \|\) induced by \(\langle \cdot, \cdot \rangle\). Further, let \(A, B \subset \mathcal{H}\) be nonempty, convex
and closed subsets. In the practical considerations one often needs to find
an element of the intersection \(A \cap B\) or, more general, to solve the following
problem

\[
\text{find } a^* \in A \text{ and } b^* \in B \text{ such that } \|a^* - b^*\| = \inf_{a \in A, b \in B} \|a - b\|. 
\] (1)
We suppose that this infimum is attained. Of course, $a^* = b^*$ if and only if $A \cap B \neq \emptyset$. Several optimization problems, e.g. the convex feasibility problem can be reduced to problem (1) (see, e.g., [SY98, Section 2.9] for details). Problems of this kind have many practical applications, e.g. in signal reconstruction (see, e.g. [CB99] or [SY98, Chapter 6]), in image reconstruction or in intensity modulated radiation therapy (see, e.g. [Com96, CZ97, SY98, CGG01, HK02]), where the convex subsets are described by a large and sparse system of linear equalities or inequalities.

An important method generating sequences which converge weakly to a solution of problem (1) is the von Neumann alternating projection method (see, e.g. [Deu01, Chapter 9] or [BB94, Section 4]). In this method, the metric projections onto $A$ and $B$ are successively applied. Recall that for a closed and convex subset $D \subset \mathcal{H}$ and for any $u \in \mathcal{H}$ there exists the uniquely determined metric projection $P_D u$. Furthermore, a point $y \in D$ is the projection $P_D u$ if and only if
\[
\langle u - y, z - y \rangle \leq 0 \text{ for all } z \in D,
\]
i.e., inequality (2) characterizes the metric projection $P_D u$ (see, e.g. [GK90, Lemma 12.1] or [BB94, Section 1]). It is known that $a^* \in A$ and $b^* \in B$ realize the distance between $A$ and $B$ if and only if $a^* = P_A b^*$ and $b^* = P_B a^*$, i.e. $a^* \in \text{Fix } P_A P_B$ or $b^* \in \text{Fix } P_B P_A$ (see, e.g. [BB94, Lemma 2.2(i)]). Therefore, it is enough to find an element of $\text{Fix } P_A P_B$ in order to find a solution of problem (1). In this paper we construct a generalization of the von Neumann alternating projection method and prove its Fejér monotonicity with respect to the solution set $\text{Fix } P_A P_B$, as well as we prove the weak convergence of the method to a solution. Recall that a sequence $(x_k) \subset \mathcal{H}$ is called Fejér monotone with respect to a subset $D \subset \mathcal{H}$ if for all $z \in D$ there holds $\|x_{k+1} - z\| \leq \|x_k - z\|$, $k = 1, 2, \ldots$.

Consider a sequence $(x_k) \subset \mathcal{H}$ generated by the following iterative scheme
\[
x_0 \in A \text{ -- arbitrary} \quad x_{k+1} = P_A(x_k + \lambda_k \sigma_k (P_A P_B x_k - x_k)),
\]
where the relaxation parameter $\lambda_k \in [0, 2]$ and the step size $\sigma_k \geq 0$. We call the method (3) the relaxed alternating projection method (RAP-method). If $\lambda_k = \sigma_k = 1$ we obtain the von Neumann alternating projection method (AP-method):
\[
x_0 \in A \text{ -- arbitrary} \quad x_{k+1} = P_A P_B x_k
\]
(see, e.g. [BB94]). Some modifications of the AP-method (4) for $A \cap B \neq \emptyset$ and for $B$ being an obtuse cone, different from (3) were proposed in [BK04, Section 3], where the projection $P_B$ in (4) is replaced by the reflection $R_B = 2P_B - I$. 

2
One can show that any sequence \((x_k)\) generated by the AP-method (4) converges weakly to an element \(x^* \in \text{Fix} \, P_A P_B\) (see, e.g. [BB94, Theorem 4.8 and Lemma 2.2]). Note that \(\text{Fix} \, P_A P_B \neq \emptyset\) since we have supposed that the infimum in (1) is attained. If \(A \cap B \neq \emptyset\) then any sequence \((x_k)\) generated by the RAP-method (3) converges weakly to an element \(x^* \in \text{Fix} \, P_A P_B = A \cap B\) if \(\sigma_k = 1\) and \(\lambda_k \in [\varepsilon, 2 - \varepsilon]\), where \(\varepsilon > 0\) (see, e.g. [BB96, Corollary 3.22] for more general result). Gurin et al. have proposed the following step size in order to accelerate the convergence of the RAP-method in the case \(A \cap B \neq \emptyset\):

\[
\sigma_k = \frac{\|P_Bx_k - x_k\|^2}{\langle P_Bx_k - x_k, P_A P_B x_k - x_k \rangle} \tag{5}
\]

(see, [GPR67, Theorem 4]). Recently, the idea of [GPR67] was applied in the case \(A\) and \(B\) are subspaces of \(\mathcal{H}\) (see [BDHP03, Theorem 3.23]) and in the case \(A\) is a closed affine subspace of \(\mathcal{H}\) with \(A \cap B \neq \emptyset\) (see [BCK06, Corollary 4.11]). Unfortunately, the weak convergence of the RAP-method with the relaxation parameter \(\lambda_k \in [\varepsilon, 2 - \varepsilon]\) and the step size \(\sigma_k = 1\) or the step size defined by (5) is not guaranteed if \(A \cap B = \emptyset\). A new question arises in this context: what should we impose on the relaxation parameters \(\lambda_k\) and on the step sizes \(\sigma_k\) in order to obtain the weak convergence of the RAP-method (3) to an element \(x^* \in \text{Fix} \, P_A P_B\), without assumption \(A \cap B \neq \emptyset\). The answers on these questions are contained in Theorem 15 which is the main results of the paper.

In Sections 2 and 3 we give some sufficient conditions for the quasi-nonexpansivity of operators determining RAP-methods. Recall that an operator \(U : C \to \mathcal{H}\) is \textit{quasi-nonexpansive} if for all \(x \in C\) and for all \(z \in \text{Fix} \, U\) there holds the inequality

\[
\|Ux - z\| \leq \|x - z\|
\]

(see, e.g. [Hir06]). Quasi-nonexpansive operators are also known in the literature under the name \textit{attracting} operators (see, e.g. [BB96, Def. 2.1]) or \textit{Fejér monotone} operators or mappings (see, e.g. [Sch97, Def. 2.1]. In Section 4 we show the weak convergence of RAP-methods to a fixed point of the operator \(P_A P_B\) for special choices of step sizes \(\sigma_k\). In Section 5 we present the results of preliminary numerical experiments.

2. Quasi-nonexpansivity of relaxed alternating projections

Let \(A, B\) be nonempty, closed and convex subsets of \(\mathcal{H}\). Define the operator of \textit{alternating projections} \(T : A \to A\), by the equality

\[
T = P_A P_B.
\]
For a constant $\lambda \in [0, 2]$ we call the operator $T_\lambda = (1-\lambda)I + \lambda T$ the relaxation of $T$ and the operator $P_AT_\lambda$ the projected relaxation of $T$. Furthermore, for a relaxation parameter $\lambda \in [0, 2]$ we call the operator $T_{\sigma,\lambda} : A \to A$ defined by
\[
T_{\sigma,\lambda}x = P_A(x + \lambda \sigma(x)(Tx - x)),
\]
the operator of relaxed alternating projections (RAP-operator), where the non-negative step size function $\sigma(x)$ depends on $x$, i.e. $\sigma : A \to \mathbb{R}_+ = [0, +\infty)$. Of course, $T_{\sigma,\lambda} = T$ if $\sigma(x) = 1$ for all $x \in A$ and $\lambda = 1$. The operator $T$ defines the AP-method since the iteration (4) can be written in the form $x_{k+1} = Tx_k$. Similarly, for a function $\sigma : A \to \mathbb{R}_+$ and for a sequence of relaxation parameters $(\lambda_k)$ the operator $T_{\sigma,\lambda}$ defines the RAP-method by the equality
\[
x_{k+1} = T_{\sigma,\lambda_k}x_k,
\]
which is equivalent to (3) with $\sigma_k = \sigma(x_k)$. First we give some properties of the operators $T$ and $T_{\sigma,\lambda}$, which we use later to show the quasi-nonexpansivity of $T_{\sigma,\lambda}$ and the weak convergence of a sequence generated by the recurrence (7) for special choices of the step size function $\sigma : A \to \mathbb{R}_+$.

**Lemma 1** Let $\sigma(x) > 0$ for all $x \in A$ and let $\lambda > 0$. Then $\text{Fix } T_{\sigma,\lambda} = \text{Fix } T$.

**Proof.** Denote by $N_D(y) = \{u \in \mathcal{H} : \langle u - y, z - y \rangle \leq 0 \text{ for all } z \in D\}$ the normal cone to a closed and convex subset $D \subset \mathcal{H}$ at the point $y \in D$. By the equivalence
\[
y = P_Du \iff u - y \in N_D(y),
\]
where $D \subset \mathcal{H}$ is a closed and convex subset (see, e.g. [HUL93, Chapter 1, Proposition 5.3.3]) and by the obvious fact that $N_D(y)$ is a cone, we have
\[
x \in \text{Fix } T_{\sigma,\lambda} \iff P_A(x + \lambda \sigma(x)(Tx - x)) = x
\]
\[
\iff \lambda \sigma(x)(Tx - x) \in N_A(x) \iff Tx - x \in N_A(x)
\]
\[
\iff x = P_ATx = Tx \iff x \in \text{Fix } T,
\]
which completes the proof. \hfill \blacksquare

It is easily seen that the characterization (2) of the metric projection $P_Du$ is equivalent to the condition
\[
\langle z - u, P_Du - u \rangle \geq \|P_Du - u\|^2 \text{ for any } u \in \mathcal{H} \text{ and } z \in D.
\]

Denote by $\delta = d(A, B) = \inf_{x \in A, y \in B} \|x - y\|$ the distance between the subsets $A$ and $B$. As we have supposed in Section 1, $\delta$ is attained, consequently, $\text{Fix } T \neq \varnothing$. 

4
Lemma 2 Let $z \in \text{Fix} T$. Then for any $x \in A$ there holds the inequality
\[
\langle z - x, Tx - x \rangle \geq \|Tx - P_Bx\|^2 - \tilde{\delta}\|P_Bx - x\| + \langle P_Bx - x, Tx - x \rangle, \tag{10}
\]
where $\tilde{\delta} \in [\delta, \|T x - P_Bx\|]$ is an upper bound of the distance $\delta$.

Proof. Denote $w = P_Bz$. Observe that $\|z - w\| = \delta$ (see, e.g. [BB94, Lemma 2.2(i)]). We have by the characterization (2) of the metric projection and by the Cauchy–Schwarz inequality,
\[
\langle z - P_Bx, P_Bx - x \rangle = \langle z - w, P_Bx - x \rangle + \langle w - P_Bx, P_Bx - x \rangle \\
\geq \langle z - w, P_Bx - x \rangle \\
\geq -\|z - w\| \cdot \|P_Bx - x\| = -\delta \|P_Bx - x\|.
\]
Therefore, if we apply condition (9) we obtain
\[
\langle z - x, Tx - x \rangle = \langle z - P_Bx, Tx - x \rangle + \langle P_Bx - x, Tx - x \rangle \\
\geq \|Tx - P_Bx\|^2 - \delta\|P_Bx - x\| + \langle P_Bx - x, Tx - x \rangle.
\]
Now, (10) follows from the inequality $\delta \leq \tilde{\delta}$. ■

Let $x \in A$. Denote
\[
\tilde{\delta} = \tilde{\delta}(x) = \|Tx - P_Bx\|. \tag{11}
\]
Let
\[
\tilde{\delta} = \tilde{\delta}(x) \in [\delta, \tilde{\delta}(x)] \tag{12}
\]
be an upper bound of the distance $\delta$ and let $T_{\sigma, \lambda}$ be defined by (6), where the function $\sigma : A \rightarrow \mathbb{R}_+$ is given by
\[
\sigma(x) = \frac{\|Tx - P_Bx\|^2 - \tilde{\delta}\|P_Bx - x\| + \langle P_Bx - x, Tx - x \rangle}{\|Tx - x\|^2} \tag{13}
\]
for $x \notin \text{Fix} T$ and $\sigma(x) = 1$ for $x \in \text{Fix} T$. For $x \in A$ define
\[
\alpha(x) = \begin{cases} 
\langle(x - P_Bx, Tx - P_Bx) & \text{if } x \notin \text{Fix} T \text{ and } P_Bx \notin A \\
0 & \text{if } x \in \text{Fix} T \text{ or } P_Bx \in A
\end{cases} \tag{14}
\]
where the symbol $\langle(a, b)$ denotes the angle between two nonzero vectors $a, b \in \mathcal{H}$, i.e. $\langle(a, b) = \arccos \frac{\langle a, b \rangle}{\|a\| \|b\|}$. Note that $\alpha(x)$ is well defined since $x - P_Bx$ and
\(Tx - P_Bx\) are obviously nonzero vectors in the first case of (14). Furthermore, we have by the characterization of the metric projection \(P_A(P_Bx)\)

\[
\langle P_Bx - x, P_Bx - Tx \rangle \geq \|P_Bx - Tx\|^2 > 0,
\]

consequently,

\[
0 < \frac{\|P_Bx - Tx\|}{\|P_Bx - x\|} \leq \cos \alpha(x).
\]

**Lemma 3** Let \(x \in A\) and let the step size \(\sigma(x)\) be defined by (13). Then

\[
\sigma(x) \geq \frac{1}{1 + \cos \alpha(x)} \geq \frac{1}{2}.
\]

**Proof.** Inequality (17) is clear if \(x \in \text{Fix } T\) or \(P_Bx \in A\). Suppose now that \(x \notin \text{Fix } T\) and \(P_Bx \notin A\). Denote \(a = P_Bx - x\), \(b = Tx - x\) and \(c = P_Bx - Tx\). Of course \(a, b, c \neq 0\) and \(\alpha(x) = \angle(a, c)\). Observe that \(b = a - c\), \(\delta \leq \|c\|\), and that the function \(y \mapsto \frac{y + \rho}{y + 2\rho}\) is increasing for \(y > -2\rho\). Therefore, for \(\rho = 1 - \cos \alpha(x)\), we have

\[
\sigma(x) = \frac{\|c\|^2 - \delta \|a\| + \langle a, b \rangle}{\|b\|^2} \geq \frac{\|c\|^2 - \|a\| \cdot \|c\| + \langle a, b \rangle}{\|b\|^2}
\]

\[
= \frac{\|c\|^2 - \|a\| \cdot \|c\| + \langle a, a - c \rangle}{\|a - c\|^2}
\]

\[
= \frac{(\|a\| - \|c\|)^2 + \|a\| \cdot \|c\| - \langle a, c \rangle}{(\|a\| - \|c\|)^2 + 2(\|a\| \cdot \|c\| - \langle a, c \rangle)}
\]

\[
= \frac{(\|a\| - \|c\|)^2 + \|a\| - \|c\| - \langle a, c \rangle}{\|a - c\|^2}
\]

\[
= \frac{(\|a\| - \|c\|)^2 + \|a\| - \|c\| - \langle a, c \rangle}{\|a - c\|^2}
\]

\[
\geq \frac{(1 - \|a\|)(\|a\| - 1) + 1 - \cos \alpha(x)}{(1 - \|a\|)(\|a\| - 1) + 1 - \cos \alpha(x)}
\]

\[
= \frac{1}{1 + \cos \alpha(x)} \geq \frac{1}{2},
\]

which completes the proof. 

**Lemma 4** Let \(x \in A\) be such that \(Tx \notin \text{Fix } T\). Then \(\alpha(x) \in (0, \frac{\pi}{2})\) and, consequently, the vectors \(x - P_Bx\) and \(Tx - P_Bx\) are linearly independent.
Proof. Suppose that $\alpha(x) = 0$, i.e.

$$Tx - P_Bx = \gamma(x - P_Bx)$$

(18)

for some $\gamma > 0$. By the equivalence (8) we have $x - P_Bx \in N_B(P_Bx)$, consequently, $\gamma(x - P_Bx) \in N_B(P_Bx)$ and, again by the equivalence (8),

$$P_B(P_Bx + \gamma(x - P_Bx)) = P_Bx.$$  

(19)

Now we obtain by (18) and (19)

$$TTx = T(P_Bx + \gamma(x - P_Bx))$$

$$= P_A P_B(P_Bx + \gamma(x - P_Bx))$$

$$= P_A P_Bx = Tx,$$

a contradiction with the assumption $Tx \notin \text{Fix } T$. Therefore, $\alpha(x) > 0$. Furthermore, $\alpha(x) < \frac{\gamma}{2}$ by (16). Consequently, the vectors $x - P_Bx$ and $Tx - P_Bx$ are linearly independent.

Remark 5 Accordingly to Lemma 4, we can stop RAP-algorithm (3) with $Tx_k \in \text{Fix } T$ if we state that $P_Bx_k - x_k$ and $Tx_k - P_Bx_k$ are linearly dependent.

Let $x \in A$ be such that $Tx \notin \text{Fix } T$. Let $y \in \text{aff}(x, P_Bx, Tx)$ be a solution of the system

$$\langle P_Bx - y, P_Bx - x \rangle = \tilde{\delta}\|P_Bx - x\|$$

(20)

$$\langle P_Bx - y, Tx - P_Bx \rangle = -\|Tx - P_Bx\|^2,$$  

(21)

where $\tilde{\delta} \in [\delta, \|Tx - P_Bx\|]$. By Lemma 4 such a solution is defined uniquely.

Lemma 6 Let $x \in A$ be such that $Tx \notin \text{Fix } T$. The step size $\sigma(x)$ given by (13) is characterized by the equality

$$\langle x + \sigma(x)(Tx - x) - y, Tx - x \rangle = 0$$

(22)

Proof. For $y$ being a solution of the system (20)--(21) and for any $\sigma$ we have

$$\langle x + \sigma(Tx - x) - y, Tx - x \rangle$$

$$= \langle x - y, Tx - x \rangle + \sigma\|Tx - x\|^2$$

$$= \langle x - P_Bx, Tx - x \rangle + \langle P_Bx - y, Tx - x \rangle + \sigma\|Tx - x\|^2$$

$$= \langle x - P_Bx, Tx - x \rangle + \langle P_Bx - y, Tx - P_Bx \rangle$$

$$+ \langle P_Bx - y, P_Bx - x \rangle + \sigma\|Tx - x\|^2$$

$$= \langle x - P_Bx, Tx - x \rangle - \|Tx - P_Bx\|^2 + \tilde{\delta}\|P_Bx - x\| + \sigma\|Tx - x\|^2.$$  

Therefore, equalities (22) and (13) are equivalent.
Lemma 7 Let $z \in \text{Fix} T$, $x \in A$ and let $\sigma(x)$ be defined by (13). There holds the inequality

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

Proof. The Lemma follows directly from Lemma 2 and from equality (13).

Theorem 8 Let $x \in A$ and let $\sigma(x)$ be given by (13). Then, for any $z \in \text{Fix} T$ and for any $\lambda \geq 0$ there holds the inequality

$$\|T_{\sigma,\lambda}x - z\|^2 \leq \|x - z\|^2 - \lambda(2 - \lambda)\sigma^2(x)\|Tx - x\|^2. \tag{23}$$

Consequently, the operator $T_{\sigma,\lambda}$ defined by (6) is quasi-nonexpansive for $\lambda \in [0, 2]$.

Proof. Let $z \in \text{Fix} T$, $x \in A$ and $\lambda \geq 0$. Of course $z = P_A z$. We have by the nonexpansivity of the metric projection $P_A$ and by Lemma 7

$$\begin{align*}
\|T_{\sigma,\lambda}x - z\|^2 &= \|P_A(x + \lambda\sigma(x)(Tx - x)) - z\|^2 \\
&= \|P_A(x + \lambda\sigma(x)(Tx - x)) - P_A z\|^2 \\
&\leq \|x + \lambda\sigma(x)(Tx - x) - z\|^2 \\
&= \|x - z\|^2 + \lambda^2 \sigma^2(x)\|Tx - x\|^2 - 2\lambda\sigma(x)\langle z - x, Tx - x \rangle \\
&\leq \|x - z\|^2 + \lambda^2 \sigma^2(x)\|Tx - x\|^2 - 2\lambda\sigma^2(x)\|Tx - x\|^2 \\
&= \|x - z\|^2 - \lambda(2 - \lambda)\sigma^2(x)\|Tx - x\|^2
\end{align*}$$

and we see that $T_{\sigma,\lambda}$ is quasi-nonexpansive if $\lambda \in [0, 2]$.

Remark 9 Let $A \cap B \neq \emptyset$ and let $x \in A \setminus B$. We have $\delta = 0$ and the step size given by (13) with $\delta = 0$ has the form

$$\sigma(x) = \frac{\|Tx - P_Bx\|^2 + \langle P_Bx - x, Tx - x \rangle}{\|Tx - x\|^2}. \tag{24}$$

Gurin et al. have proposed the relaxation parameter $\lambda = 1$ and the following step size $\sigma(x)$ in the relaxed alternating projection method

$$\sigma(x) = \frac{\|P_Bx - x\|^2}{\langle P_Bx - x, Tx - x \rangle} \tag{25}$$

(see [GPR67, equality (15)]). Observe that the step size $\sigma(x)$ defined by (25) is the unique solution of the equality

$$\langle x + \sigma(x)(Tx - x) - P_Bx, P_Bx - x \rangle = 0.$$

Consider two cases the RAP-method with the step size (25):
(a) If $A, B$ are subspaces of $H$ then

$$\sigma(x) = \frac{\langle x, x - Tx \rangle}{\|Tx - x\|^2}.$$  

In this case the RAP-method is equivalent to an acceleration method of Bauschke et. al. (see [BDHP03, equality (3.1.2) and Theorem 3.23]).

(b) If $A$ is a closed affine subspace then $\langle P_B x - x, Tx - x \rangle = \|Tx - x\|^2$ and

$$\sigma(x) = \frac{\|P_B x - x\|^2}{\|Tx - x\|^2}.$$  

In this case the RAP-method is equivalent to the extrapolated alternating projection method (see [BCK06, equality (4.35)]).

**Lemma 10** Let $A \cap B \neq \emptyset$ and let $x \in A \setminus B$. Then we have

$$\frac{\|Tx - P_B x\|^2 + \langle P_B x - x, Tx - x \rangle}{\|Tx - x\|^2} \geq \frac{\|P_B x - x\|^2}{\langle P_B x - x, Tx - x \rangle},$$  

i.e. the step size $\sigma(x)$ defined by (24) is not shorter than the proposed by Gurin et. al (equality (25)). Furthermore, both step sizes are equal if $A$ is a closed affine subspace.

**Proof.** Observe that $\delta = 0$ since $A \cap B \neq \emptyset$ and that $x \notin \text{Fix} T$ since $\text{Fix} T = A \cap B$ for $A \cap B \neq \emptyset$. It follows from the characterization of the metric projection $P_A(P_B x)$ that

$$\langle x - Tx, P_B x - Tx \rangle \leq 0.$$  

(27)

If we apply inequality (27), the Cauchy–Schwarz inequality, the nonexpansivity of the metric projection $P_A$ and the fact $x \neq Tx$ we easily obtain

$$0 < \|Tx - x\|^2 \leq \langle P_B x - x, Tx - x \rangle \leq \|P_B x - x\|^2.$$  

(28)

A simple computation shows that

$$\frac{\|Tx - P_B x\|^2 + \langle P_B x - x, Tx - x \rangle}{\|Tx - x\|^2} = \frac{\|Tx - x\|^2 + \|P_B x - x\|^2 - \langle P_B x - x, Tx - x \rangle}{\|Tx - x\|^2}.$$  

If we apply the last equality we easily see that (26) is equivalent to the inequality

$$\langle P_B x - x\|^2 - \langle P_B x - x, Tx - x \rangle \rangle (\langle P_B x - x, Tx - x \rangle - \|Tx - x\|^2) \geq 0$$  

which is true by (28). Suppose now that $A$ is a closed affine subspace. The equality in (26) follows easily from the fact that $\langle Tx - P_B x, Tx - x \rangle = 0$ for $A$ being an affine subspace. ■
3. Quasi-nonexpansivity of the RAP-operator for closed and affine $A$

In this Section we suppose that $A \subset \mathcal{H}$ is a closed affine subspace. In this case $x + \sigma(Tx - x) \in A$ for any $x \in A$ and $\sigma \in \mathbb{R}$, where $T = P_A P_B$. Consequently, the RAP-operator $T_{\sigma, \lambda} : A \rightarrow A$ defined by (6) has the form

$$T_{\sigma, \lambda}(x) = x + \lambda \sigma(x)(Tx - x)$$

and one iteration of the RAP-method has the form

$$x_{k+1} = x_k + \lambda_k \sigma_k(Tx_k - x_k).$$

It is known that for $A$ being an affine subspace the operator $T = P_A P_B$ restricted to $A$ is firmly nonexpansive [Com94, Proposition 3i)] and that the RAP-method converges to an element of $\text{Fix} T$ for $\sigma_k = 1$ and for $\lambda_k \in [\varepsilon, 2 - \varepsilon]$, where $\varepsilon > 0$ [Com94, Theorem 1] (see, e.g. [GK90, Chapter 12] for the definition and the properties of firmly nonexpansive operators). We generalize these results. We start with the following Lemma.

**Lemma 11** Let $A \subset \mathcal{H}$ be a closed affine subspace and $B \subset \mathcal{H}$ be a closed and convex subset. For all $x, y \in A$ there holds the inequality,

$$\langle Tx - Ty, x - y \rangle \geq \|Tx - Ty\|^2 + (\|Tx - P_B x\| - \|Ty - P_B y\|)^2.$$

**Proof.** Since the metric projection $P_A$ is a firmly nonexpansive operator (see, e.g. [BB96, Fact 1.5]), we have for any $u, v \in \mathcal{H}$

$$\langle Tu - Tv, P_B u - P_B v \rangle \geq \|Tu - Tv\|^2. \quad (31)$$

Further, for any $u, v \in A$ we have, by the affinity of $A$,

$$\langle Tu - P_B u, u - P_B u \rangle = \|Tu - P_B u\|^2 \quad (32)$$

and

$$\langle P_B v - Tv, u - Tu \rangle = 0. \quad (33)$$

The characterization of the metric projection $P_B v$ yields

$$\langle P_B u - P_B v, v - P_B v \rangle \leq 0 \quad (34)$$

10
for any \( u, v \in \mathcal{H} \). Let now \( x, y \in A \). It follows from (31)-(34) and from the Cauchy–Schwarz inequality that

\[
\langle Tx - Ty, x - y \rangle = \langle Tx - Ty, P_B x - P_B y \rangle + \langle Tx - Ty, (x - P_B x) - (y - P_B y) \rangle
\]

\[
\geq \|Tx - Ty\|^2 + (\langle Tx - P_B x, x - P_B x \rangle - \langle Tx - P_B x, y - P_B y \rangle + \langle P_B x - P_B y, x - P_B x \rangle + \langle P_B y - P_B x, y - P_B y \rangle
\]

\[
\geq \|Tx - Ty\|^2 + \|Tx - P_B x\|^2 - \langle Tx - P_B x, y - Ty \rangle - \langle P_B x - Tx, Ty - P_B y \rangle + \langle P_B y - Ty, Ty, x - Tx \rangle
\]

\[
\geq 2 \|Tx - Ty\|^2 + \|Tx - P_B x\|^2 - 2\|P_B x - Tx\| \cdot \|Ty - P_B y\|
\]

\[
\geq \|Tx - Ty\|^2 + (\|Tx - P_B x\| - \|Ty - P_B y\|)^2,
\]

which completes the proof. ■

**Corollary 12 (Combettes, 1994)** Let \( A \subset \mathcal{H} \) be a closed affine subspace and \( B \subset \mathcal{H} \) be a closed and convex subset. Then the operator \( T : A \to A \), \( T = P_A P_B \) is firmly nonexpansive.

Let the function \( \sigma : A \to \mathbb{R}_+ \) be defined by

\[
\sigma(x) = 1 + \frac{(\|Tx - P_B x\| - \tilde{\delta})^2}{\|Tx - x\|^2}.
\]

for \( x \notin \text{Fix} T \) and \( \sigma(x) = 1 \) for \( x \in \text{Fix} T \), where \( \tilde{\delta} \) is given by (12).

**Lemma 13** Let \( z \in \text{Fix} T \), \( x \in A \) and let \( \sigma(x) \) be defined by (35). There holds the inequality

\[
\langle z - x, Tx - x \rangle \geq \sigma(x)\|Tx - x\|^2.
\]

**Proof.** The Lemma is obvious for \( x \in \text{Fix} T \). Let now \( x \notin \text{Fix} T \). Since \( \delta = \|Tz - P_B z\| \), we have by Lemma 11

\[
\langle z - x, Tx - x \rangle = \|Tx - x\|^2 + \langle z - Tx, Tx - x \rangle
\]

\[
= \|Tx - x\|^2 + \langle Tz - Tx, z - x \rangle - \|Tz - Tx\|^2
\]

\[
\geq \|Tx - x\|^2 + (\|Tx - P_B x\| - \|Tz - P_B z\|)^2
\]

\[
= (1 + \frac{(\|Tx - P_B x\| - \tilde{\delta})^2}{\|Tx - x\|^2})\|Tx - x\|^2
\]

\[
\geq (1 + \frac{(\|Tx - P_B x\| - \tilde{\delta})^2}{\|Tx - x\|^2})\|Tx - x\|^2
\]

11
and the Lemma follows now from equality (35).

**Corollary 14** Let \( A \subset \mathcal{H} \) be a closed affine subspace and \( B \subset \mathcal{H} \) be a closed and convex subset. Further, let \( T_{\sigma,\lambda} : A \to A \) be defined by (29) where \( \sigma \) is defined by (35). Then for any \( x \in A, z \in \text{Fix}T \) and \( \lambda \geq 0 \) there holds the inequality

\[
\|T_{\sigma,\lambda}x - z\|^2 \leq \|x - z\|^2 - \lambda(2 - \lambda)\sigma^2(x)\|Tx - x\|^2, \tag{36}
\]

consequently, \( T_{\sigma,\lambda} \) is quasi-nonexpansive for \( \lambda \in [0, 2] \).

**Proof.** Let \( x \in A, z \in \text{Fix}T \) and \( \lambda \geq 0 \). We have by Lemma 13

\[
\|T_{\sigma,\lambda}x - z\|^2 = \|x + \lambda\sigma(x)(Tx - x) - z\|^2
\]

\[
= \|x - z\|^2 + \lambda^2\sigma^2(x)\|Tx - x\|^2 - 2\lambda\sigma(x)\langle z - x, Tx - x \rangle
\]

\[
\leq \|x - z\|^2 - \lambda(2 - \lambda)\sigma^2(x)\|Tx - x\|^2.
\]

Now we see that that for \( \lambda \in [0, 2] \) the operator \( T_{\sigma,\lambda} \) is quasi-nonexpansive.

4. **Convergence of the RAP-method**

We consider in this Section two cases of the RAP-method (3) with \( \lambda_k \in [\varepsilon, 2 - \varepsilon] \) for \( \varepsilon > 0 \):

(i) \( A, B \subset \mathcal{H} \) are closed convex subsets and the step size \( \sigma_k \) is given by

\[
\sigma_k = \frac{\|Tx_k - P_Bx_k\|^2 - \tilde{\delta}_k\|P_Bx_k - x_k\| + \langle P_Bx_k - x_k, Tx_k - x_k \rangle}{\|Tx_k - x_k\|^2}, \tag{37}
\]

(ii) \( A \subset \mathcal{H} \) is a closed affine subspace, \( B \subset \mathcal{H} \) is a closed convex subset and the step size \( \sigma_k \) is given by

\[
\sigma_k = 1 + \frac{(\|Tx_k - P_Bx_k\| - \tilde{\delta}_k)^2}{\|Tx_k - x_k\|^2}. \tag{38}
\]

In both cases \( \tilde{\delta}_k = \tilde{\delta}(x_k) \in [\delta, \tilde{\delta}_k] \), where \( \delta_k = \tilde{\delta}(x_k) = \|Tx_k - P_Bx_k\| \).

**Theorem 15** In both cases (i) and (ii) the sequence \( (x_k) \) converges weakly to an element \( x^* \in \text{Fix}T \).
Proof. By Lemma 1 we have $\text{Fix} T_{\sigma, A} = \text{Fix} T$. If we set $x = x_k$ in inequality (23) or in inequality (36) if $A$ is a closed and affine subspace, we obtain in both cases for any $z \in \text{Fix} T$

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 - \lambda_k(2 - \lambda_k)s_k^2\|Tx_k - x_k\|^2.$$ 

Therefore, $(\|x_k - z\|)$ converges as a nonincreasing sequence. Consequently,

$$\|Tx_k - x_k\| \to 0 \quad (39)$$

since for $\sigma_k$ given by (37) we have $\sigma_k \geq \frac{1}{2}$ (see Lemma 3), for a closed affine subspace $A$ and for $\sigma_k$ given by (38) we have $\sigma_k \geq 1$, and $\lambda_k(2 - \lambda_k) \geq \varepsilon^2 > 0$. Let $(x_{nk})$ be any weakly convergent subsequence of $(x_k)$ and let $x \in A$ be the weak limit of $x_{nk}$. Note that such a subsequence exists since $(x_k)$ is bounded. Since $T$ is nonexpansive we have by (39) that $x \in \text{Fix} T$ (see, e.g. [BB96, Fact 1.2]). We have proved that all weak cluster points of $(x_k)$ lie in $\text{Fix} T$. Furthermore, $\text{Fix} T$ is closed and convex (see, e.g. [BB94, Lemma 2.2 (ii)]). Since the sequence $(x_k)$ is Fejér monotone with respect to $\text{Fix} T$, it converges weakly to some point $x^* \in \text{Fix} T$, see [BB96, Theorem 2.16 (ii)].

5. The results of preliminary numerical experiments

In this Section we present the results of preliminary numerical tests for problem (1), where $H = \mathbb{R}^n$.

5.1. Problems

We consider the following test problems:

P1. $A = B(z_1, 1), B = B(z_2, 1)$ are two balls in $\mathbb{R}^n$ with centres $z_1, z_2 \in \mathbb{R}^n$ and radius 1. We consider this problem for various distances $d = \|z_1 - z_2\|$. Of course, $\delta = d(A, B) = \max\{0, d - 2\}$, consequently, $A \cap B \neq \emptyset$ if and only if $d \leq 2$. Without loss of generality we suppose that $n = 2$, $z_1 = (0, d)$ for $d \in \mathbb{R}_+$ and $z_2 = (0, 0)$. We set $x_0 = (1, d) \in \mathbb{R}^2$ as the starting point. The exact solution of (1) can be easily evaluated analytically for the test problem P1, $x^* = (0, d - 1)$ for $d \geq 2$ and $x^* = (\sqrt{4 - d^2}/2, d/2)$ for $d < 2$.

P2. $A$ is a hyperplane and $B = B(z, 1)$ is a ball in $\mathbb{R}^n$. We consider this problem for various distances $d = \inf_{y \in A} \|z - y\|$. Of course, $\delta = d(A, B) = \max\{0, d - 1\}$, consequently, $A \cap B \neq \emptyset$ if and only if $d \leq 1$. Without
loss of generality we suppose that $n = 2$, $A = \{(\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_2 = d\}$ for $d \in \mathbb{R}_+$ and $B = B((0,0),1)$. We set $x_0 = (3, d) \in \mathbb{R}^2$ as the starting point. The exact solution of (1) can be easily evaluated analytically for the test problem $P1$, $x^* = (0, d)$ for $d \geq 1$ and $x^* = (\sqrt{1-d^2}, d)$ for $d < 1$.

Note that in all problems the starting point $x_0 \in \mathbb{R}^n$ belongs to $A$.

5.2. Tests

Now we present the results of numerical tests for the following methods:

- **AP** – the von Neumann alternating projection method (4), applied to problem (1),
- **RAP1** – the relaxed alternating projection method (3), where the step size $\sigma_k$ is defined by (37), applied to problem (1),
- **RAP2** – the relaxed alternating projection method (3), where the step size $\sigma_k$ is defined by (38), applied to problem (1) with affine $A$,
- **GPR** – the method proposed by Gurin–Polyak–Raik, i.e. the relaxed alternating projection method (3), where the step size $\sigma_k$ is defined by (5), applied to problem (1) with $A \cap B \neq \emptyset$.

In the presented tests we employ various values of constant relaxation parameter $\lambda_k = \lambda \in (0,2)$. For the methods RAP1 and RAP2 we consider two cases:

(i) the value $\delta$ is known and we set $\tilde{\delta}_k = \delta$ in (37) and in (38),
(ii) we set $\tilde{\delta}_k = \delta_k = \|T x_k - P_B x_k\| \in (37)$ and in (38).

For both test problems $P1$ and $P2$ we know the exact solution $x^*$ of (1) and we apply the condition $\|x_k - x^*\| \leq \varepsilon$ or $x_k \in A \cap B$ as the stopping criterion. Let $k$ denotes the number of iterations after which the corresponding algorithm terminates. All tested methods were programmed in MATLAB 6.1.

In Table 1, we present the numerical results of the methods AP, GPR and RAP1(i), for problem $P1$ with $A \cap B \neq \emptyset$ for various distances $d$ between the centres of two balls and for various optimality tolerances $\varepsilon$. The results of RAP1(i) are presented for three values of relaxation parameter $\lambda$ (note that the methods AP and GPR are originally constructed only for $\lambda = 1$). The results for $\lambda = 1$ are repeated in Figure 1. We see that for all optimality tolerances the behavior of the methods RAP1(i) and GPR is similar and is
considerably better than for the AP-method. Observe, that RAP1(i) behaves a little bit better if $\lambda > 1$.

Table 1:
Comparison of the methods: AP, GPR and RAP1(i)
for problem P1 in case $A \cap B \neq \emptyset$

<table>
<thead>
<tr>
<th>$\lambda \rightarrow$</th>
<th>AP</th>
<th>GPR</th>
<th>RAP1(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$\varepsilon$</td>
<td>$k$</td>
<td>$k$</td>
</tr>
<tr>
<td>1.9</td>
<td>$10^{-1}$</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>1.99</td>
<td>$10^{-1}$</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>97</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>$10^{-1}$</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>2499</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>249999</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 1: Comparison of the methods AP, GPR and RAP1(i)
for problem P1 in case $A \cap B \neq \emptyset$
In Table 2, we compare the methods AP, RAP1(i) and RAP1(ii), for problem P1 for various optimality tolerances \( \varepsilon \). We consider here both cases: \( A \cap B \neq \emptyset (d \leq 2) \) as well as \( A \cap B = \emptyset (d > 2) \). The results for \( d > 2 \) are repeated in Figure 2. Note that we cannot apply GPR if \( A \cap B = \emptyset \). Observe that RAP1(i) behaves essentially better than RAP1(ii) and than the AP-method. The most considerable differences are in case \( d = 2 \). In this case RAP1(i) behaves very well while RAP1(ii) as well as the AP-method converge very slowly because of zigzagging (the angle between the vectors \( P_B x_k - x_k \) and \( P_A P_B x_k - P_B x_k \) is close to \( \pi \) for \( x_k \) closed to the solution \( x^* \)).

Table 2: Comparison of the methods AP and RAP1 for problem P1

<table>
<thead>
<tr>
<th>method ( \lambda \rightarrow )</th>
<th>( \delta_k \rightarrow )</th>
<th>1</th>
<th>( \frac{3}{2} )</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_k )</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-1}</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-2}</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-3}</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-1}</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-2}</td>
<td>43</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-3}</td>
<td>97</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-1}</td>
<td>49</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-2}</td>
<td>5000</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-3}</td>
<td>333296</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-1}</td>
<td>134</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-2}</td>
<td>3297</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-3}</td>
<td>2474</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-1}</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-2}</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \delta_k )</td>
<td>10^{-3}</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 2: Comparison of the methods AP and RAP1 for problem P1 in case $A \cap B = \emptyset$

In Table 3, we present the results of numerical tests for problem P2. We compare RAP2(i) ($\tilde{\delta}_k = \delta$) and RAP2(ii) ($\tilde{\delta}_k = \tilde{\delta}_k$). In the second case $\sigma_k = 1$ for both methods. Furthermore, for $\lambda = 1$ RAP2(ii) reduces to the AP-method. We set $d = 1$ (the hyperplane $A$ is tangent to $B$ in the solution $x^*$). For such $d$ the termination of RAP2(ii) requires essentially more iterations than the of RAP2(i). Note that in this case $A$ and $B$ are almost "parallel" near the solution and the angle between the vectors $P_B x_k - x_k$ and $P_A P_B x_k - P_B x_k$ is close to $\pi$. We observe a small influence of parameter $\lambda$ on the convergence. Furthermore, the behavior of RAP2(i) is essentially better (we use here the known distance $\delta$ between $A$ and $B$) than the of RAP2(ii).

<table>
<thead>
<tr>
<th>method</th>
<th>RAP2(i)</th>
<th>RAP2(ii)</th>
<th>RAP2(i)</th>
<th>RAP2(ii)</th>
<th>RAP2(i)</th>
<th>RAP2(ii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>$\tilde{\delta}_k$</td>
<td>$\tilde{\delta}_k$</td>
<td>$\tilde{\delta}_k$</td>
<td>$\tilde{\delta}_k$</td>
<td>$\tilde{\delta}_k$</td>
<td>$\tilde{\delta}_k$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>5</td>
<td>100</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>48</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$k$</td>
<td>$k$</td>
<td>$k$</td>
<td>$k$</td>
<td>$k$</td>
<td>$k$</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>5</td>
<td>100</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>48</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>8</td>
<td>10000</td>
<td>5</td>
<td>6555</td>
<td>5</td>
<td>4996</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>12</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>7</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>5</td>
<td>49995</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>15</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>8</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>5</td>
<td>$&gt; 5 \cdot 10^5$</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>18</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>10</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>5</td>
<td>$&gt; 5 \cdot 10^5$</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>22</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>12</td>
<td>$&gt; 5 \cdot 10^5$</td>
<td>5</td>
<td>$&gt; 5 \cdot 10^5$</td>
</tr>
</tbody>
</table>

Preliminary numerical experiments show that the both relaxed alternating projection methods behave essentially better than the original alternating projection method if the distance $\delta = d(A, B)$ is known. The most significant difference in the behavior of both methods with respect to the AP-method can
be observed if $A \cap B$ consists of one point or the distance $\delta$ is close to zero and the subsets $A$ and $B$ are almost "parallel" close to the solution.

**Acknowledgment.** The authors wish to thank an anonymous referee for his helpful remarks.

**References**


