TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 355, Number 9, Pages 3433–3461
S 0002-9947(03)03136-2
Article electronically published on May 29, 2003

ACCELERATING THE CONVERGENCE OF THE METHOD OF ALTERNATING PROJECTIONS

HEINZ H. BAUSCHKE, FRANK DEUTSCH, HEIN HUNDAL, AND SUNG-HO PARK

Abstract. The powerful von Neumann-Halperin method of alternating projections (MAP) is an algorithm for determining the best approximation to any given point in a Hilbert space from the intersection of a finite number of subspaces. It achieves this by reducing the problem to an iterative scheme which involves only computing best approximations from the *individual* subspaces which make up the intersection. The main practical drawback of this algorithm, at least for some applications, is that the method is slowly convergent. In this paper, we consider a general class of iterative methods which includes the MAP as a special case. For such methods, we study an "accelerated" version of this algorithm that was considered earlier by Gubin, Polyak, and Raik (1967) and by Gearhart and Koshy (1989). We show that the accelerated algorithm converges faster than the MAP in the case of two subspaces, but is, in general, *not faster* than the MAP for more than two subspaces! However, for a "symmetric" version of the MAP, the accelerated algorithm always converges faster for any number of subspaces. Our proof seems to require the use of the Spectral Theorem for selfadjoint mappings.

1. Introduction

Let X be a (real) Hilbert space, let $M_1, M_2, \ldots, M_k$ be closed (linear) subspaces of X with $M = \bigcap_1^k M_i$, and for any closed subspace N of X, let P_N denote the orthogonal projection onto N. The von Neumann-Halperin method of alternating projections, or MAP for short, is an iterative algorithm for determining the best approximation $P_M x$ to x from M. It does this by computing the iterates $x_0 := x$ and $x_n = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1}) x_{n-1} = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n x$. That is, the iterates (x_n) are obtained by cyclically computing the best approximations onto the individual subspaces M_i $(i = 1, 2, \ldots, k)$. The method is thus most effective when it is "easy" to compute the best approximations from the individual subspaces M_i. The main theorem governing the MAP is the following.

Theorem (von Neumann [18] for $k = 2$, Halperin [15] for $k \geq 2$)**.** *Let $M_1, M_2, \ldots, M_k$ be closed subspaces in the Hilbert space X and let $M := \bigcap_1^k M_i$. Then*

$$\lim_{n\to\infty} \|(P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n x - P_M x\| = 0 \qquad \textit{for all} \quad x \in X.$$

In case $k = 2$, this result was rediscovered in at least six other papers (see, e.g., the survey [5]).

Received by the editors July 30, 1999.
2000 *Mathematics Subject Classification.* Primary 41A65.
Key words and phrases. Alternating projections, cyclic projections, accelerating convergence, best approximation from an intersection of subspaces, Hilbert space.

Also, as was noted in [5], there are at least ten different areas of mathematics in which the MAP has proved useful. However, the main *practical* drawback of the MAP appears to be that it is often slowly convergent. Indeed, if $M_1 + M_2$ is not closed, then Franchetti and Light [11] and Bauschke, Borwein, and Lewis [2] have given examples showing that the convergence of $(P_{M_2}P_{M_1})^n x$ to $P_{M_1\cap M_2}x$ can be arbitrarily slow!

Both Gubin, Polyak, and Raik [14] and Gearhart and Koshy [13] have considered a geometrically appealing method to *accelerate* the MAP, but in neither of these two papers was it proved that the acceleration scheme considered was actually faster than the MAP. In this paper, we will prove that this acceleration scheme is indeed faster than the MAP in the case of two subspaces (i.e., $k = 2$) (Theorem 3.23). But, perhaps surprisingly, we show that the acceleration scheme may actually be *slower* than the MAP when $k \geq 3$ (Example 3.24)! In contrast to this, we show that a "symmetric" version of the MAP (i.e., $x_0 = x$ and $x_n = (P_{M_1}P_{M_2}\cdots P_{M_k}P_{M_{k-1}}\cdots P_{M_1})^n x$ for $n = 1, 2, \dots$) has an accelerated version which is faster for any $k \geq 2$ (Corollary 3.21).

We should also mention that Dyer [10] and Hanke and Niethammer [16] have considered methods of accelerating the "Kaczmarz method" of solving linear equations. (Recall that Kaczmarz's method may be regarded as the special case of the MAP in the case when X is finite-dimensional and each M_i is a hyperplane.)

2. The method of iterated projections

To provide motivation for the acceleration results to be established later, in this section we give a fairly general convergence result which contains the von Neumann-Halperin result as a special case. In the next section, we will consider methods to accelerate this general algorithm.

Unless otherwise stated, the standing assumptions are as follows. Let X be a (real) Hilbert space, $M_1, M_2, \dots, M_k$ be closed subspaces, $M := \bigcap_1^k M_i$, and let $P_i = P_{M_i}$ denote the orthogonal projection onto M_i ($i = 1, 2, \dots, k$).

Now let
$$T := P_k P_{k-1} \cdots P_1$$
denote the composition of the k projections P_i taken in increasing order. The well-known von Neumann-Halperin Theorem states that
$$\lim_{n\to\infty} \|T^n x - P_M x\| = 0$$
for each $x \in X$ (see, more generally, Theorem 2.5 below). Also, it can be shown that
$$\lim_n \|(T^*T)^n x - P_M x\| = 0$$
for each $x \in X$ (see Theorem 2.6 below). More generally, suppose T is any bounded linear mapping from X into itself such that

$$\lim_n \|T^n x - P_{\operatorname{Fix} T} x\| = 0 \quad \text{for each } x \in X, \tag{2.0.1}$$

where
$$\operatorname{Fix} T := \{x \in X \mid Tx = x\}$$
is the *fixed point* set for T.

We will be interested in determining methods to *accelerate* the convergence of the sequence $(T^n x)$ to $P_{\operatorname{Fix} T}x$. Before we consider such methods, it will provide useful motivation to first give some general conditions on the mapping T that will guarantee that (2.0.1) holds.

The mapping T is called **nonexpansive** if $\|T\| \leq 1$. We first recall that the fixed point sets of T and T^* are the same if T is nonexpansive (see Riesz and Sz.-Nagy [19] or Riesz and Sz.-Nagy [20, p. 408]).

Lemma 2.1. *Let T be a nonexpansive linear operator on X. Then*

$$\operatorname{Fix} T = \operatorname{Fix} T^*. \tag{2.1.1}$$

In fact, $Tx = x$ if and only if $\langle Tx, x\rangle = \|x\|^2$ if and only if $\langle x, T^*x\rangle = \|x\|^2$ if and only if $T^*x = x$.

Our next observation is a characterization of those linear operators T on X that satisfy (2.0.1). We will use the following notation. If A is any linear operator on X, we denote the *range* and *null space* of A by $\mathcal{R}(A)$ and $\mathcal{N}(A)$ respectively. We will also use the well-known fact that $\mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$ (see [3, Remarks following Theorem 2.19 on pp. 35-36]).

Theorem 2.2. *Let T be a bounded linear operator on X, and let M be a closed linear subspace of X. Consider the following statements:*

(1) $\lim_n \|T^n x - P_M x\| = 0$ *for each* $x \in X$*;*
(2) $M = \operatorname{Fix} T$ *and* $T^n x \to 0$ *for each* $x \in M^\perp$*;*
(3) $M = \operatorname{Fix} T$ *and T is "asymptotically regular", i.e.,* $T^n x - T^{n+1}x \to 0$ *for each* $x \in X$.

Then (1) $\Longleftrightarrow$ (2) $\Longrightarrow$ (3). *If, in addition, T is nonexpansive, then all three statements are equivalent.*

Proof. Suppose (1) holds. If $x \in M$, then $T^n x \to P_M x = x$. So by the continuity of T,

$$Tx = T(\lim T^n x) = \lim T(T^n x) = \lim T^{n+1}x = P_M x = x$$

implies that $x \in \operatorname{Fix} T$, i.e., $M \subset \operatorname{Fix} T$.

Conversely, let $y \in \operatorname{Fix} T$. Then $Ty = y$ and an easy induction shows that $y = T^n y$ for each n. Thus $y = T^n y \to P_M y$ which implies $y = P_M y \in M$. That is, $\operatorname{Fix} T \subset M$. Hence $M = \operatorname{Fix} T$.

If $x \in M^\perp$, then

$$T^n x = T^n(P_{M^\perp} x) \to P_M(P_{M^\perp} x) = 0.$$

This proves (2).

Now assume (2) holds and let $x \in X$. Then

$$T^n x = T^n(P_M x + P_{M^\perp} x) = T^n(P_M x) + T^n(P_{M^\perp} x) = P_M x + T^n(P_{M^\perp}) \to P_M x.$$

Thus (1) holds, and this establishes the equivalence of (1) and (2).

Now suppose that (2) holds and $x \in X$. By the equivalence of (1) and (2), we have that $T^n x \to P_M x$ and so $T^n x - T^{n+1} x \to P_M x - P_M x = 0$. Thus T is asymptotically regular, and hence (3) holds.

This proves the first statement of the theorem. To complete the proof, suppose (3) holds and let T be nonexpansive. Then $\operatorname{Fix} T^* = \operatorname{Fix} T = M$ by Lemma 2.1. Then for any $x \in X$, we have that $T^n(x - Tx) = T^n x - T^{n+1}x \to 0$. Hence $T^n y \to 0$ for every $y \in \mathcal{R}(I - T)$ which implies, since $\|T^n\| \leq 1$ by nonexpansiveness, that $T^n y \to 0$ for every

$$y \in \overline{\mathcal{R}(I - T)} = \mathcal{N}(I - T^*)^\perp = (\operatorname{Fix} T^*)^\perp = M^\perp.$$

Thus, for any $x \in X$,

$$T^n x = T^n(P_M x + P_{M^\perp} x) = T^n(P_M x) + T^n(P_{M^\perp} x) = P_M x + T^n(P_{M^\perp} x) \to P_M x,$$

and this proves that (1) holds. □

Remark. Statement (3) does *not* imply statement (1) in general. To see this, let X denote the Euclidean plane and let $e_1 = (1,0)$ and $e_2 = (0,1)$ denote the canonical orthonormal basis vectors in X. Defining $T : X \to X$ by $Tx = [x(1) + x(2)]e_1$, it is easy to verify that $T^n x = Tx$ for every $n \in \mathbb{N}$ and every $x \in X$, so that T is asymptotically regular, $M := \operatorname{Fix} T = \operatorname{span} e_1$, $T^n(e_1 + e_2) = 2e_1$ for every n, but $P_M(e_1 + e_2) = e_1 \neq 2e_1 = T^n(e_1 + e_2)$ for every n. Thus, $T^n x \not\to P_M(x)$ when $x = e_1 + e_2$.

Corollary 2.3. *Let T be nonexpansive on X and $M = \operatorname{Fix} T$. Then*

$$\lim_{n\to\infty} \|T^n x - P_M x\| = 0 \qquad \text{for all} \quad x \in X$$

if and only if T is asymptotically regular.

Lemma 2.4. *Let $M_1, M_2, \ldots, M_k$ be closed subspaces of the Hilbert space X, let $M := \bigcap_1^k M_i$ and let $T := P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$. Then T is nonexpansive and*

$$\operatorname{Fix} T = \operatorname{Fix} T^* = \operatorname{Fix}(TT^*) = \operatorname{Fix}(T^*T) = M.$$

Proof. For simplicity, let $P_i = P_{M_i}$. Since T is the product of nonexpansive operators, T is nonexpansive. If $x \in M$, then $x \in M_i$ for each i so that $P_i x = x$ for each i and hence $Tx = x$. That is, $M \subset \operatorname{Fix} T$. Conversely, if $z \in \operatorname{Fix} T$, then $Tz = z$. Thus, $P_k P_{k-1} \cdots P_1 z = z$. We have $P_i z = z$ if and only if $\|P_i z\| = \|z\|$ (using the fact that $\|z\|^2 = \|P_i z\|^2 + \|z - P_i z\|^2$). If $z \notin M$, let i be the smallest index such that $z \notin M_i$. Then $P_i z \neq z$; so $\|P_i z\| < \|z\|$ and $z = P_k \cdots P_i P_{i-1} \cdots P_1 z = P_k \cdots P_i z$ implies that $\|z\| = \|P_k \cdots P_i z\| \leq \|P_i z\| < \|z\|$, which is absurd. Thus, $z \in M$. This proves that $M = \operatorname{Fix} T$. By Lemma 2.1, $M = \operatorname{Fix} T^*$.

Since $T^* = P_1 P_2 \cdots P_k$, we see that $TT^* = P_k P_{k-1} \cdots P_1 P_2 \cdots P_k$ and $T^*T = P_1 P_2 \cdots P_k P_{k-1} \cdots P_1$, and the same proof as above shows that $\operatorname{Fix} TT^* = M = \operatorname{Fix} T^*T$. □

A useful sufficient condition that guarantees that (2.0.1) holds is essentially contained in Halperin [15]. It also is explicit in Smarzewski [21] and can be stated as follows. (We include a brief proof since, as far as we know, the paper [21] has not been published.) Recall that $T : X \to X$ is called **nonnegative** if $\langle Tx, x\rangle \geq 0$ for all $x \in X$.

Theorem 2.5. *Let $T_1, T_2, \ldots, T_k$ be selfadjoint, nonnegative, and nonexpansive bounded linear operators on the Hilbert space X. Let $T := T_1 T_2 \cdots T_k$ and $M = \operatorname{Fix} T$. Then $\operatorname{Fix} T = \bigcap_1^k \operatorname{Fix} T_i$ and*

$$\lim_n \|T^n x - P_M x\| = 0 \quad \text{for every} \quad x \in X. \tag{2.5.1}$$

Proof. Since T is nonexpansive, Corollary 2.3 implies that it suffices to show that T is asymptotically regular. Toward this end, note that for each i, $I - T_i$ is nonnegative (and selfadjoint) since

$$\langle (I - T_i)x, x\rangle = \langle x - T_i x, x\rangle = \|x\|^2 - \langle T_i x, x\rangle \geq \|x\|^2 - \|T_i\|\,\|x\|^2 \geq 0.$$

It follows from a result of Riesz (see [4, Theorem 4.6.4, p. 163]) that $T_i(I-T_i)$ is also nonnegative. Hence,

$$\begin{aligned}\|x\|^2 &= \|x-T_ix+T_ix\|^2 = \|x-T_ix\|^2 + 2\langle x-T_ix, T_ix\rangle + \|T_ix\|^2 \\ &= \|x-T_ix\|^2 + 2\langle T_i(I-T_i)x, x\rangle + \|T_ix\|^2 \geq \|x-T_ix\|^2 + \|T_ix\|^2.\end{aligned}$$

Thus, for each $x \in X$,

$$\|x-T_ix\|^2 \leq \|x\|^2 - \|T_ix\|^2 \quad \text{for each } i. \tag{2.5.2}$$

By repeated application of (2.5.2), we deduce that

$$\begin{aligned}\|x\|^2 - &\|Tx\|^2 \\ &= \|x\|^2 - \|T_kx\|^2 + \|T_kx\|^2 - \|T_{k-1}T_kx\|^2 + \cdots + \|T_2\cdots T_kx\|^2 - \|Tx\|^2 \\ &\geq \|x-T_kx\|^2 + \|T_kx - T_{k-1}T_kx\|^2 + \cdots + \|T_2\cdots T_kx - Tx\|^2 \\ &= k\left[\frac{1}{k}\|x-T_kx\|^2 + \frac{1}{k}\|T_kx-T_{k-1}T_kx\|^2 + \cdots + \frac{1}{k}\|T_2\cdots T_kx - Tx\|^2\right] \\ &\geq k\|\frac{1}{k}(x - T_kx + T_kx - T_{k-1}T_kx + \cdots + T_2\cdots T_kx - Tx)\|^2 \\ &\quad \text{(by convexity of } \|\cdot\|^2\text{)} \\ &= \frac{1}{k}\|x-Tx\|^2.\end{aligned}$$

That is,

$$\|x-Tx\|^2 \leq k(\|x\|^2 - \|Tx\|^2) \quad \text{for every } x \in X. \tag{2.5.3}$$

Since T is nonexpansive, we see that the sequence $(\|T^nx\|)_{n=1}^{\infty}$ is nonincreasing for every $x \in X$ and so it must converge: $\|T^nx\| \to \rho \geq 0$. Now apply (2.5.3) with x replaced by T^nx to obtain that

$$\|T^nx - T^{n+1}x\|^2 \leq k(\|T^nx\|^2 - \|T^{n+1}x\|^2) \to 0 \quad \text{as } n \to \infty.$$

This proves that T is asymptotically regular. □

Lemma 2.4 and Theorem 2.5 immediately imply the following two results. The first is the "von Neumann-Halperin theorem" stated in the Introduction, while the second shows that a symmetric version of the MAP also converges.

Theorem 2.6. *Let $M_1, M_2, \ldots, M_k$ be closed subspaces of the Hilbert space X, and let $M = \bigcap_1^k M_i$. Then, for each $x \in X$,*

$$\lim_n \|(P_{M_k}P_{M_{k-1}}\cdots P_{M_1})^nx - P_Mx\| = 0. \tag{2.6.1}$$

Theorem 2.7. *Let $M_1, M_2, \ldots, M_k$ be closed subspaces of the Hilbert space X, and let $M = \bigcap_1^k M_i$. Then, for each $x \in X$,*

$$\lim_n \|(P_{M_1}P_{M_2}\cdots P_{M_k}P_{M_{k-1}}\cdots P_{M_1})^nx - P_Mx\| = 0. \tag{2.7.1}$$

Using Theorems 2.6 and 2.7, we see that two important examples of operators T which satisfy (2.0.1) are $T = Q$ and $T = Q^*Q$, where $Q := P_{M_k}P_{M_{k-1}}\cdots P_{M_1}$.

3. Acceleration methods

Throughout this section, unless explicitly stated otherwise, we assume that T is a nonexpansive linear operator on X and $M := \operatorname{Fix} T$. Hence, $\operatorname{Fix} T^* = M$ also (by Lemma 2.1). Moreover, M_i will always denote a closed linear subspace of X and $P_i = P_{M_i}$.

In this section, we develop our main results concerned with accelerating the method given by (2.0.1). That is, if T is an operator such that (2.0.1) holds (or equivalently, that T is asymptotically regular), how can we modify the iterates suggested by this algorithm so as to converge faster to $P_M x$?

Definition 3.1. The **accelerated mapping** A_T **of** T is defined on X by

$$A_T(x) := t_x T x + (1 - t_x)x, \tag{3.1.1}$$

where

$$t_x = t_{x,T} := \begin{cases} \frac{\langle x, x-Tx\rangle}{\|x-Tx\|^2} & \text{if } Tx \neq x, \\ 1 & \text{if } Tx = x. \end{cases} \tag{3.1.2}$$

We will consider two classes of iterative algorithms to compute $P_M(x)$ for any given $x \in X$. They are described as follows. The standard or "unaccelerated" algorithm: $x_0 = x$ and

$$x_n = T(x_{n-1}) = T^n(x) \quad (n = 1, 2, \dots), \tag{3.1.3}$$

and its "accelerated" counterpart: $x_0 = x, \quad x_1 = T(x_0)$, and

$$x_n = A_T(x_{n-1}) = A_T^{n-1}(Tx) \quad (n = 1, 2, \dots). \tag{3.1.4}$$

In particular, we will give a detailed analysis of these algorithms when $T = P_k P_{k-1} \cdots P_1$ and when $T = (P_k P_{k-1} \cdots P_1)^*(P_k P_{k-1} \cdots P_1)$. This acceleration scheme was suggested by Gubin et al [14] and Gearhart and Koshy [13] in the particular case when T is a product of projections. The motivation for using the mapping A_T is that $A_T(x)$ is that point on the line through the points x and Tx which is closest to $P_M x$ (see Theorem 3.7 below).

A remark is in order as to why, in the accelerated algorithm, we first apply T to x_0 rather than first applying A_T. That is, why didn't we define the accelerated algorithm by $x_n = A_T^n(x_0)$ for $n \geq 0$ rather than $x_{n+1} = A_T^n(Tx_0)$ for $n \geq 0$? The simple answer is that, besides being the one suggested in [14] and [13], the one we defined performs better. Indeed, it is not hard to see that if T is the product of two orthogonal projections onto two 1-dimensional (nonorthogonal) subspaces in the Euclidean plane, then *the accelerated algorithm converges in two steps*, that is, $A_T(Tx) = P_M x$ for any starting point x. However, for any choice of x which is not in the range of T, *none* of the terms of the sequence $(A_T^n(x))$ is equal to $P_M x$. That is, the sequence $x_n = A_T^n(x)$ does not converge to $P_M x$ in a finite number of steps.

Definition 3.2. The classical **von Neumann-Halperin method of alternating projections**, or **MAP** for short, corresponds to (3.1.3) in the case when $T = P_k P_{k-1} \cdots P_1$.

The **accelerated method of alternating projections**, or the **accelerated MAP** for short, is the algorithm (3.1.4) in the case when $T = P_k P_{k-1} \cdots P_1$.

The **symmetric method of alternating projections**, or **symmetric MAP** for short, is just (3.1.3) in the case when $T = (P_k P_{k-1} \cdots P_1)^*(P_k P_{k-1} \cdots P_1)$.

The **accelerated symmetric method of alternating projections**, or **accelerated symmetric MAP** for short, is the algorithm (3.1.4) in the case when $T = (P_k P_{k-1} \cdots P_1)^* (P_k P_{k-1} \cdots P_1)$.

Lemma 3.3. *Let $x \in X$. Then*

(1) $tA_T(x) + (1-t)x - x \in M^\perp \cap (A_T(x))^\perp$ *for every* $t \in \mathbb{R}$.
(2) $Tx - x \in M^\perp \cap (A_T(x))^\perp$.
(3) $A_T(x) - x \in M^\perp \cap (A_T(x))^\perp$.
(4) $T(M^\perp) \subset M^\perp$ *and* $A_T(M^\perp) \subset M^\perp$.
(5) $A_T(x) - Tx \in M^\perp \cap (A_T(x))^\perp$.
(6) $Tx - P_M x \in M^\perp$.

Proof. (1) Since $tA_T(x) + (1-t)x - x = tt_x(Tx - x)$, it suffices to verify (2).

(2) If $Tx = x$, then (2) is trivial. Thus we may assume that $Tx \neq x$. Let $y \in M$. Then since $Ty = y = T^* y$, we have

$$\langle Tx - x, y\rangle = \langle x, T^* y\rangle - \langle x, y\rangle = \langle x, y\rangle - \langle x, y\rangle = 0.$$

Thus $Tx - x \in M^\perp$. Also,

$$\begin{aligned}\langle Tx - x, A_T(x)\rangle &= \langle Tx - x, t_x Tx + (1-t_x)x\rangle = t_x\langle Tx - x, Tx - x\rangle + \langle Tx - x, x\rangle \\ &= -\frac{\langle x, Tx - x\rangle}{\|x - Tx\|^2}\|Tx - x\|^2 + \langle Tx - x, x\rangle = 0;\end{aligned}$$

so $Tx - x \in (A_T(x))^\perp$.

(3) Take $t = 1$ in (1).

(4) This follows from (2) and (3).

(5) Since $A_T(x) - Tx = (t_x - 1)(Tx - x)$, the result follows from (2).

(6) Using part (2), we get

$$Tx - P_M x = (Tx - x) + (x - P_M x) \in M^\perp + M^\perp = M^\perp.$$

□

Lemma 3.4. *For every $x \in X$ and $n \in \mathbb{N} := \{1, 2, \ldots, \}$,*

(1) $P_M(A_T(x)) = P_M x$;
(2) $P_M(Tx) = P_M x$;
(3) $P_M(A_T^{n-1}(Tx)) = P_M x$;
(4) $P_M(T^n x) = P_M x$.

Proof. We use the well-known fact that $P_M(M^\perp) = \{0\}$. Since $Tx - x \in M^\perp$ and $A_T(x) - x \in M^\perp$ by Lemma 3.3, it follows that $0 = P_M(Tx - x) = P_M(Tx) - P_M x$ and $0 = P_M(A_T(x) - x) = P_M(A_T(x)) - P_M x$. Hence (1) and (2) follow.

(3) and (4) follow by a repeated application of (1) and (2). □

Lemma 3.5. *For each $x \in X$ and $y \in M$,*

$$\|A_T(x) - y\|^2 = \|x - y\|^2 - \|x - A_T(x)\|^2. \tag{3.5.1}$$

In particular,

$$\operatorname{Fix} A_T = \{x \in X \mid \|A_T(x)\| = \|x\|\} \tag{3.5.2}$$

and

$$\|A_T(x)\|^2 = \begin{cases} \|x\|^2 & \text{if } x \in M, \\ \|x\|^2 - \frac{\langle x, x - Tx\rangle^2}{\|x - Tx\|^2} & \text{if } x \notin M. \end{cases} \tag{3.5.3}$$

Proof. Using Lemma 3.3, we deduce that

$$\|x-y\|^2 = \|(x-A_T(x)) + (A_T(x)-y)\|^2 = \|x-A_T(x)\|^2 + \|A_T(x)-y\|^2;$$

so (3.5.1) holds. Take $y=0$ in (3.5.1) to obtain (3.5.2). Finally, take $y=0$ in (3.5.1) and note that $\|x-A_T(x)\|^2 = \frac{\langle x, x-Tx\rangle^2}{\|x-Tx\|^2}$ if $x \notin M$ and $\|x-A_T(x)\|^2 = 0$ if $x \in M$. This yields (3.5.3). □

Lemma 3.6. *The following statements are equivalent:*

(1) $Tx \in M$;
(2) $Tx = P_M x$;
(3) $T^n x \in M$ *for every* $n \geq 1$.

Proof. (1) $\Longrightarrow$ (2). If $Tx \in M$, then $Tx = P_M(Tx) = P_M x$ using Lemma 3.4(2).

(2) $\Longrightarrow$ (3). If $Tx = P_M x$, then $Tx \in M$. Thus, (3) holds when $n=1$. We proceed by induction. If $T^n x \in M$ for some $n \geq 1$, then since $M = \operatorname{Fix} T$, we have that

$$T^{n+1}x = T(T^n x) = T^n x \in M.$$

This completes the induction.

(3) $\Longrightarrow$ (1). Take $n=1$. □

The **affine hull** of a nonempty set S, denoted by $\operatorname{aff}(S)$, is the intersection of the collection of all affine sets which contain S. (Recall that an affine set is any translation of a subspace.) Equivalently, $\operatorname{aff}(S) = \{\alpha x + (1-\alpha)y \mid x, y \in S, \alpha \in \mathbb{R}\}$.

Theorem 3.7. *For each* $x \in X$ *and* $y \in M$, *we have*

$$\|A_T(x)-y\|^2 = \|tTx+(1-t)x-y\|^2 - (t-t_x)^2\|Tx-x\|^2 \quad \text{for each } t \in \mathbb{R}, \tag{3.7.1}$$

$$\|A_T(x)-y\| = \min_{t\in\mathbb{R}} \|tTx + (1-t)x - y\|, \tag{3.7.2}$$

and the minimum is attained precisely when either $t = t_x$ *if* $x \notin M$ *or at every* $t \in \mathbb{R}$ *if* $x \in M$. *Moreover,*

$$d(A_T(x), M) = \min_{t\in\mathbb{R}} d(tTx + (1-t)x, M); \tag{3.7.3}$$

in other words, $A_T(x)$ *is the unique point in* $\operatorname{aff}\{x, Tx\}$ *which is closest to* M.

$$\|A_T(x)\| = \min_{t\in\mathbb{R}} \|tTx + (1-t)x\|; \tag{3.7.4}$$

in other words, $A_T(x)$ *is the unique point in* $\operatorname{aff}\{x, Tx\}$ *having minimal norm. In particular,*

$$\|A_T(x)\| \leq \min\{\|x\|, \|Tx\|\}. \tag{3.7.5}$$

Proof. Using Lemma 3.3, we can write

$$\begin{aligned}\|tTx + (1-t)x - y\|^2 &= \|tTx + (1-t)x - A_T(x) + A_T(x) - y\|^2 \\ &= \|(t-t_x)(Tx-x) + (A_T(x)-y)\|^2 \\ &= (t-t_x)^2\|Tx-x\|^2 + \|A_T(x)-y\|^2,\end{aligned}$$

which proves (3.7.1). Equation (3.7.2) follows immediately from (3.7.1). Moreover, (3.7.3) follows by taking the infimum over all $y \in M$ in (3.7.2). Finally, (3.7.4) follows from (3.7.2) by taking $y=0$. □

While A_T is not linear in general, it does share some important properties of the linear mapping P_M. Namely, it is continuous, homogeneous, and "additive modulo M". These are recorded in parts (5), (4), and (3), respectively, of the following lemma.

Lemma 3.8. *Let $x \in X$ and $y \in M$. Then:*

1. $t_{x+y} = t_x$.
2. $t_{\alpha x} = t_x$ *for every* $\alpha \neq 0$.
3. $A_T^n(x+y) = A_T^n(x)+y$ *for every* $n \in \mathbb{N}$. *In particular,* $A_T(x+y) = A_T(x)+y$ *and* $A_T(y) = y$.
4. $A_T(\alpha x) = \alpha A_T(x)$ *for every* $\alpha \in \mathbb{R}$.
5. A_T *is continuous.*

Proof. (1) If $x \in M$, then $x+y \in M$ and $t_{x+y} = 1 = t_x$. If $x \notin M$, then $x+y \notin M$, and so,

$$t_{x+y} = \frac{\langle x+y, x+y-T(x+y)\rangle}{\|x+y-T(x+y)\|^2} = \frac{\langle x+y, x-Tx\rangle}{\|x-Tx\|^2} = \frac{\langle x, x-Tx\rangle}{\|x-Tx\|^2} = t_x$$

using Lemma 3.3(2).

(2) Let $\alpha \neq 0$. If $x \in M$, then $\alpha x \in M$ and $t_{\alpha x} = 1 = t_x$. If $x \notin M$, then $\alpha x \notin M$ and

$$t_{\alpha x} = \frac{\langle \alpha x, \alpha x - T(\alpha x)\rangle}{\|\alpha x - T(\alpha x)\|^2} = \frac{\langle x, x-Tx\rangle}{\|x-Tx\|^2} = t_x.$$

(3) When $n = 1$,

$$\begin{aligned} A_T(x+y) &= t_{x+y}T(x+y) + (1-t_{x+y})(x+y) \\ &= t_x(Tx+Ty) + (1-t_x)(x+y) \text{ using part (1)} \\ &= t_xTx + (1-t_x)x + t_xy + (1-t_x)y \\ &= A_T(x) + y. \end{aligned}$$

Now assume (3) holds for some $n \geq 1$. Then

$$\begin{aligned} A_T^{n+1}(x+y) &= A_T[A_T^n(x+y)] = A_T[A_T^n(x)+y] \\ &= A_T(A_T^n(x)) + y \quad \text{by the case } n = 1 \\ &= A_T^{n+1}(x) + y; \end{aligned}$$

so the result holds for $n+1$.

(4) If $\alpha \neq 0$, then by (2),

$$A_T(\alpha x) = t_{\alpha x}T(\alpha x) + (1-t_{\alpha x})(\alpha x) = t_x[\alpha T(x)] + (1-t_x)[\alpha x] = \alpha A_T(x).$$

Since $A_T(0) = 0$, the result also holds when $\alpha = 0$.

(5) If $x \in X \setminus M$ and $x_n \to x$, then since $X \setminus M$ is open, $x_n \notin M$ eventually, and so,

$$t_{x_n} = \frac{\langle x_n, x_n - Tx_n\rangle}{\|x_n - Tx_n\|^2} \to \frac{\langle x, x-Tx\rangle}{\|x-Tx\|^2} = t_x$$

and hence A_T is continuous at x. If $x \in M$ and $\epsilon > 0$, let $y \in X$ with $\|y-x\| < \epsilon/3$. Then $\|P_M x - P_M y\| \le \|x-y\| < \epsilon/3$ and

$$\begin{aligned}\|A_T(x) - A_T(y)\| &= \|x - A_T(y)\| \le \|x - P_M y\| + \|P_M y - A_T(y)\| \\ &= \|x - P_M y\| + \|A_T(y - P_M y)\| \qquad \text{by part (3)} \\ &\le \|x - P_M y\| + \|y - P_M y\| \qquad \text{by (3.7.5)} \\ &= \|P_M x - P_M y\| + \|y - P_M y\| \\ &< \frac{\epsilon}{3} + \|y - x\| + \|x - P_M y\| \\ &< \frac{2\epsilon}{3} + \|P_M x - P_M y\| < \epsilon.\end{aligned}$$

This proves that A_T is continuous at x. □

Remark. We note that, while A_T is continuous, it is *not* uniformly continuous, in general, unlike a linear operator. For example, let $X = \ell_2$, let $\{e_n \mid n = 0, 1, 2, \dots\}$ be an orthonormal basis for X, and define T on X by $Tx = \sum_0^\infty \langle x, e_n\rangle n/(n+1)e_n$. Setting $x_n = (1/n)e_0 + ((n+1)/n)e_n$ and $y_n = e_n$ for all $n \ge 1$, we get that $\|x_n - y_n\| = (\sqrt{2}/n) \to 0$. But using the readily deduced facts that $A_T(y_n) = 0$ and $A_T(x_n) = (1/2)(e_n - e_0)$ for all n, we obtain that $\|A_T(x_n) - A_T(y_n)\| = (\sqrt{2}/2)$ for all $n \ge 1$.

Lemma 3.9. (1) $t_x \ge \frac{1}{2}$ *for all* $x \in X$*; and*
(2) $\operatorname{Fix} A_T = M (= \operatorname{Fix} T)$.

Proof. (1) If $x \in M$, then $t_x = 1$. If $x \notin M$, then the quadratic function,

$$q(t) := \|t(Tx - x) + x\|^2 = at^2 + bt + c,$$

where $a := \|Tx - x\|^2 > 0$, $b := 2\langle x, Tx - x\rangle$, and $c := \|x\|^2$ is strictly convex and attains its minimum at the unique point t when $q'(t) = 0$; that is, when $t = t_{\min} := -\frac{b}{2a}$. Hence,

$$t_{\min} = -\frac{2\langle x, Tx - x\rangle}{2\|Tx - x\|^2} = \frac{\langle x, x - Tx\rangle}{\|x - Tx\|^2} =: t_x.$$

But $c = q(0) = \|x\|^2$ and $\|Tx\|^2 = q(1) = a + b + c = a + b + \|x\|^2$ implies that $-b = a + \|x\|^2 - \|Tx\|^2$ and hence

$$t_x = t_{\min} = \frac{-b}{2a} = \frac{a + \|x\|^2 - \|Tx\|^2}{2a} = \frac{1}{2} + \frac{\|x\|^2 - \|Tx\|^2}{2a} \ge \frac{1}{2}.$$

(2) $x \in \operatorname{Fix} A_T$ if and only if $x = t_x Tx + (1 - t_x)x$ if and only if $t_x(Tx - x) = 0$ if and only if $Tx - x = 0$ (using part (1)) if and only if $x \in \operatorname{Fix} T = M$. □

Remarks. The lower bound $\frac{1}{2}$ for t_x is sharp. To see this, take $T = -I$ and note that $t_x = \frac{1}{2}$ for every $x \in X \setminus \{0\}$. Also, if we relax the condition that T be nonexpansive and consider $T = \lambda I$ for $\lambda \ne 1$, we deduce that $t_{x,\lambda I} = \frac{1}{1-\lambda}$ for each $x \ne 0$. By varying λ, we see that $t_{x,\lambda I}$ can take on every nonzero value.

Definition 3.10. Define $f = f_T : X \to \mathbb{R}^+ := \{\alpha \in \mathbb{R} \mid \alpha \ge 0\}$ by

$$f(x) := \begin{cases} \frac{\|A_T(x) - P_M x\|}{\|Tx - P_M x\|} & \text{if } Tx \notin M, \\ 0 & \text{if } Tx \in M. \end{cases}$$

Lemma 3.11. *For each $x \in X$, we have $0 \le f(x) \le 1$ and*

$$\|A_T(x) - P_M x\| = f(x)\|Tx - P_M x\|. \tag{3.11.1}$$

Proof. This is immediate from (3.7.2) with $y = P_M x$. □

Lemma 3.12. *T commutes with P_M and $P_{M^\perp}$.*

Proof. For each $x \in X$,

$$\begin{aligned} P_M T x &= P_M T(P_M x + P_{M^\perp} x) = P_M[T(P_M x) + T(P_{M^\perp} x)] \\ &= P_M^2 x \quad \text{since } T(M^\perp) \subset M^\perp \text{ by Lemma 3.3(4)} \\ &= P_M x = T P_M x. \end{aligned}$$

Thus, T commutes with P_M and, since $P_{M^\perp} = I - P_M$, it follows that T also commutes with $P_{M^\perp}$. □

Definition 3.13. Let T be a nonexpansive linear operator on X, $M = \operatorname{Fix} T$, and for any $n \in \mathbb{N}$, let $c_n(T)$ denote the norm of the linear operator $(TP_{M^\perp})^n$:

$$c_n(T) := \|(TP_{M^\perp})^n\|. \tag{3.13.1}$$

We will often write $c(T)$ instead of $c_1(T)$. Note that if $T = P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$, then $M := \bigcap_1^k M_i = \operatorname{Fix} T$ and

$$c(T) = \|P_{M_k} P_{M_{k-1}} \cdots P_{M_1} P_{M^\perp}\| =: c(M_1, M_2, \ldots M_k) \tag{3.13.2}$$

is just the cosine of the angle of the k-tuple $(M_1, M_2, \ldots, M_k)$ defined by Bauschke, Borwein, and Lewis [2]. It was established in [2] that $c(T) < 1$ if and only if $M_1^\perp + M_2^\perp + \cdots + M_k^\perp$ is closed. When $k = 2$,

$$c(P_{M_2} P_{M_1}) = \|P_{M_2} P_{M_1} P_{M^\perp}\| = c(M_1, M_2) = c(M_2, M_1) = c(P_{M_1} P_{M_2}) \tag{3.13.3}$$

is just the ordinary cosine of the angle between the subspaces M_1 and M_2 (see [12] or [7]).

Lemma 3.14. *Let T be nonexpansive on X and $M = \operatorname{Fix} T$. Then*

(1) *$c_n(T) = \|T^n - P_M\|$ for every $n \in \mathbb{N}$. In particular,*

$$\|T^n x - P_M x\| \le c_n(T)\|x - P_M x\| \quad \text{for every} \quad n \in \mathbb{N}, \quad \text{and} \quad x \in X, \tag{3.14.1}$$

and $c_n(T)$ is the smallest constant independent of x for which (3.14.1) *is valid.*

(2) *$\|T^n y\| \le c_n(T)\,\|y\|$ for every $y \in M^\perp$;*

(3) *$c_n(T) \le c(T)^n$ for every n;*

(4) *$c(T^*T) \le c(T)^2$ and $c(T^*T) = c(T)^2$ if $\operatorname{Fix}(T^*T) = \operatorname{Fix} T$. In particular, if $T = P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$, then*

$$c(T^*T) = c(T)^2; \tag{3.14.2}$$

(5) *$\|A_T(x) - P_M x\| \le f(x) c(T)\|x - P_M x\|$ for every $x \in X$.*

Proof. (1) By Lemma 3.12, T commutes with P_M and $TP_M = P_M = P_M T$. Thus,

$$c_n(T) = \|(TP_{M^\perp})^n\| = \|[T(I - P_M)]^n\| = \|(T - P_M)^n\| = \|T^n - P_M\|.$$

Now fix any $x \in X$ and set $y = x - P_M x$. Then $y \in M^\perp$ and

$$\begin{aligned} \|T^n x - P_M x\| &= \|T^n(x - P_M x)\| = \|T^n y\| = \|T^n P_{M^\perp} y\| = \|(TP_{M^\perp})^n y\| \\ &\le c_n(T)\,\|y\| = c_n(T)\,\|x - P_M x\|, \end{aligned}$$

which proves (3.14.1).

(2) This was essentially proved during the course of proving (1).

(3) $c_n(T) = \|(TP_{M^\perp})^n\| \leq \|TP_{M^\perp}\|^n = c_1(T)^n$.

(4) Let $N = \operatorname{Fix}(T^*T)$. Since $M = \operatorname{Fix} T^*$ by Lemma 2.1, it follows that $M \subset N$ and so $N^\perp \subset M^\perp$. Hence, since T commutes with $P_{M^\perp}$ by Lemma 3.12 and, by a similar proof, T^* commutes with $P_{M^\perp}$, we obtain

$$\begin{aligned} c(T^*T) &= \|T^*TP_{N^\perp}\| \leq \|T^*TP_{M^\perp}\| = \|(TP_{M^\perp})^*(TP_{M^\perp})\| \\ &= \|TP_{M^\perp}\|^2 = c(T)^2. \end{aligned}$$

Moreover, if $\operatorname{Fix}(T^*T) = \operatorname{Fix} T$, then $N = M$ and $N^\perp = M^\perp$. So the above inequality must be an equality. Equation (3.14.2) holds when T is a product of projections by Lemma 2.4.

(5) Fix $x \in X$. Then $x - P_M x \in M^\perp$. So Lemma 3.11 and part (1) imply

$$\|A_T(x) - P_M x\| = f(x)\|Tx - P_M x\| \leq f(x)c(T)\|x - P_M x\|.$$

□

Remark. The following example shows that the strict inequality $c(T^*T) < c(T)^2$ is possible in part (4). For let X denote the Euclidean plane and define the linear operator T on X by $Tx = x(2)e_1 + x(1)e_2$ for each $x = (x(1), x(2)) \in X$, where $e_1 = (1,0)$ and $e_2 = (0,1)$. It is easy to verify that $\|T\| = 1$, $M := \operatorname{Fix} T = \operatorname{span}(e_1 + e_2)$, $c(T) = 1$, $T = T^*$, $T^*T = I$, and $c(T^*T) = c(I) = 0$.

Lemma 3.15. *Let T be nonexpansive and $M = \operatorname{Fix} T$. Then*

(1) *if T is normal, then $\|T^n - P_M\| = c(T)^n$ for every n;*
(2) *if $\operatorname{Fix}(T^*T) = \operatorname{Fix} T$, then*

$$\|(T^*T)^n - P_M\| = c(T)^{2n} \quad \text{for every } n \in \mathbb{N}. \tag{3.15.1}$$

In particular, for every $n \in \mathbb{N}$,

$$\|(P_{M_1}P_{M_2}\cdots P_{M_k}P_{M_{k-1}}\cdots P_{M_1})^n - P_M\| = c(M_1, M_2, \ldots, M_k)^{2n}. \tag{3.15.2}$$

Proof. (1) Since T is normal and T commutes with $P_{M^\perp}$ by Lemma 3.12, we deduce that $TP_{M^\perp}$ is normal. Hence, using Lemma 3.14(1), we obtain

$$\|T^n - P_M\| = c_n(T) = \|(TP_{M^\perp})^n\| = ||TP_{M^\perp}\|^n = c(T)^n.$$

(2) Since T^*T is selfadjoint, hence normal, apply part (1) to T^*T instead of T using that $\operatorname{Fix}(T^*T) = M$ to get $\|(T^*T)^n - P_M\| = c(T^*T)^n$. By Lemma 3.14(4), $c(T^*T) = c(T)^2$ and (3.15.1) follows.

(3.15.2) follows from (3.15.1) by taking $T = P_{M_k}P_{M_{k-1}}\cdots P_{M_1}$ and using Lemma 2.4 to get $\operatorname{Fix} T^*T = \operatorname{Fix} T$. □

The following theorem gives an upper bound on the rate of convergence of the accelerated scheme.

Theorem 3.16. *Let $x \in X$ and set*

$$x_n := A_T^{n-1}(Tx) \quad (n = 1, 2, \ldots).$$

Then for every $n \in \mathbb{N}$,

$$\|T^n x - P_M x\| \leq c(T)^n \|x - P_M x\| \tag{3.16.1}$$

and

$$\|A_T^{n-1}(Tx) - P_M x\| \le \left[\prod_1^{n-1} f(x_i)\right] c(T)^n \|x - P_M x\|. \tag{3.16.2}$$

Proof. The relation (3.16.1) is a consequence of Lemma 3.14(1) and (3).

We prove (3.16.2) by induction on n. For $n = 1$, $\|Tx - P_M x\| \le c(T)\|x - P_M x\|$ by (3.14.1). Since the product of any set of scalars over the empty set of indices is 1 by definition, (3.16.2) holds when $n = 1$. Now assume that (3.16.2) holds when $n = m \ge 1$. Then

$$\begin{aligned}
\|A_T^m(Tx) - P_M x\| &= \|x_{m+1} - P_M x\| = \|A_T(x_m) - P_M x\| \\
&= \|A_T(x_m) - P_M(x_m)\| \quad \text{(by Lemma 3.4)} \\
&= f(x_m)\|T(x_m) - P_M(x_m)\| \quad \text{(by Lemma 3.11)} \\
&\le f(x_m)c(T)\|x_m - P_M(x_m)\| \quad \text{(by (3.14.1))} \\
&= f(x_m)c(T)\|A_T^{m-1}T(x) - P_M x\| \\
&\le f(x_m)c(T)\left[\prod_{i=1}^{m-1} f(x_i)\right] c(T)^m \|x - P_M x\| \\
&= \left[\prod_1^m f(x_i)\right] c(T)^{m+1}\|x - P_M x\|,
\end{aligned}$$

which shows that (3.16.2) holds with n replaced by $m + 1$. This completes the induction. □

Remarks. By comparing the right sides of (3.16.1) and (3.16.2), this result seems to suggest that the accelerated algorithm is always faster than its unaccelerated counterpart by at least the factor $\left[\prod_1^{n-1} f(x_i)\right]$. Indeed, we will show below that when T is selfadjoint, nonnegative, and nonexpansive, then the accelerated method is *faster* than the original (see Theorem 3.20). In particular, the accelerated symmetric MAP is faster than the symmetric MAP. Also, the accelerated MAP for two subspaces is faster than the MAP. Perhaps surprisingly, however, we will see that this is not always the case, in general, for the accelerated MAP when there are more than two subspaces.

Theorem 3.16 can be strengthened in the particular case when $T = P_2P_1$. To do this, it is convenient to appeal to the following simple lemma (see, e.g., [13]).

Lemma 3.17. *Let M_1 and M_2 be closed subspaces with $M = M_1 \cap M_2$ and let P_i be the orthogonal projection onto M_i for $i = 1, 2$. Then $c(P_2P_1) = c(M_1, M_2)$ and*

1. *if $x \in M_1 \cap M^\perp$, then $\|P_2 x\| \le c(M_1, M_2)\|x\|$;*
2. *if $x \in M_2 \cap M^\perp$, then $\|P_1 x\| \le c(M_1, M_2)\|x\|$;*
3. *if $x \in M_2 \cap M^\perp$, then $\|P_2P_1 x\| \le c(M_1, M_2)^2\|x\|$.*

Proof. That $c(P_2P_1) = c(M_1, M_2)$ in this case was observed following Definition 3.13.

(1) Let $x \in M_1 \cap M^\perp$. Then

$$\|P_2 x\| = \|P_2P_1P_{M^\perp}x\| \le \|P_2P_1P_{M^\perp}\|\,\|x\| = c(P_2P_1)\|x\|.$$

(2) The proof is similar to (1).

(3) Let $x \in M_2 \cap M^{\perp}$. Then $P_1 x \in M_1 \cap M^{\perp}$; so by (1) and (2), we obtain

$$\|P_2 P_1 x\| \le c(P_2 P_1)\|P_1 x\| \le c(P_2 P_1)^2 \|x\|.$$

□

Theorem 3.18. *Let $T = P_{M_2} P_{M_1}$, $x \in X$, and*

$$x_n := A_T^{n-1}(Tx) \quad (n = 1, 2, \dots).$$

Then

$$\|A_T^{n-1}(Tx) - P_M x\| \le \left[\prod_1^{n-1} f(x_i)\right] c(M_1, M_2)^{2n-1} \|x - P_M x\|. \tag{3.18.1}$$

Proof. The proof is by induction and proceeds just as in the proof of Theorem 3.16. The only point that should be noted is that in the induction step, we use the inequality $\|T(x_m) - P_M(x_m)\| \le c(T)^2 \|x_m - P_M(x_m)\|$ (rather than the same expression with $c(T)$ instead of $c(T)^2$ that was used in Theorem 3.16). The proof of this inequality follows immediately from Lemma 3.17(3). □

Remarks. (1) Gearhart and Koshy [13] established (a weaker version of) the special case of Theorem 3.18 when $c := c(M_1, M_2) < 1$ and with an additional factor ρ on the right side of (3.18.1), where $\rho := \frac{1}{\sqrt{1-c^2}} \ge 1$.

(2) The inequality (3.18.1) improves the bound on the ordinary MAP in case $k = 2$, due to Aronszajn [1], who showed that

$$\|(P_2 P_1)^n x - P_M x\| \le c(M_1, M_2)^{2n-1} \|x - P_M x\| \quad \text{for all } x \in X. \tag{3.18.2}$$

In fact, Kayalar and Weinert [17] showed that the Aronszajn bound is *sharp*, i.e., $\|(P_2 P_1)^n - P_M\| = c(M_1, M_2)^{2n-1}$.

Next we show that the accelerated algorithms are always at least as fast as their unaccelerated counterparts provided that T is selfadjoint, nonnegative, and nonexpansive. It is first convenient to establish the following result.

Lemma 3.19. *If*

$$\|T^{n-1}(A_T(x))\| \le \|T^n x\| \quad \textit{for every } x \in M^{\perp} \textit{ and } n \in \mathbb{N}, \tag{3.19.1}$$

then

$$\|A_T^{n-1}(Tx)\| \le \|T^n x\| \quad \textit{for every } x \in M^{\perp} \textit{ and } n \in \mathbb{N}. \tag{3.19.2}$$

In particular, if (3.19.1) *holds and the original algorithm converges, then*

$$\|A_T^{n-1}(Tx) - P_M x\| \le \|T^n x - P_M x\| \quad \textit{for every } x \in X, n \in \mathbb{N}, \tag{3.19.3}$$

and hence the accelerated algorithm converges at least as fast as the original.

Proof. When $n = 1$, (3.19.2) is trivial. If $n \geq 2$, then for each $x \in M^\perp$,

$$\begin{aligned} \|A_T^{n-1}(Tx)\| &= \|A_T(A_T^{n-2}(Tx))\| \leq \|T(A_T^{n-2}(Tx))\| \quad \text{using (3.7.5)} \\ &= \|T(A_T(y))\|, \quad \text{where } y := A_T^{n-3}(Tx) \in M^\perp \text{ by Lemma 3.3(4)} \\ &\leq \|T^2 y\| \quad \text{by (3.19.1)} \\ &= \|T^2(A_T^{n-3}(Tx))\| \\ &= \|T^2(A_T(z))\|, \quad \text{where } z := A_T^{n-4}(Tx) \in M^\perp \text{ by Lemma 3.3(4)} \\ &\leq \|T^3 z\| \quad \text{by (3.19.1)} \\ &= \|T^3(A_T^{n-4}(Tx))\|. \end{aligned}$$

Continuing in this way, we end up with the inequality $\|A_T^{n-1}(Tx)\| \leq \|T^n x\|$, which verifies (3.19.2) when $n \geq 2$.

To verify the last statement, let $x \in X$. Then $x - P_M x \in M^\perp$ and so by (3.19.2) and Lemma 3.8(3), we get

$$\|A_T^{n-1}(Tx) - P_M x\| = \|A_T^{n-1}(T(x - P_M x))\| \leq \|T^n(x - P_M x)\| = \|T^n x - P_M x\|$$

and this verifies (3.19.3). □

The natural question raised by Lemma 3.19 is this: for which T does (3.19.1) hold? We will show next that if T is selfadjoint, nonnegative, and nonexpansive, then (3.19.1) and hence (3.19.3) hold. It should be noted that our proof seems to use the spectral theorem (for compact selfadjoint operators) in an essential way.

Theorem 3.20. *Let T be selfadjoint, nonnegative, and nonexpansive. Then*

(3.20.1) $\quad \|A_T^{n-1}(Tx) - P_M x\| \leq \|T^n x - P_M x\|$ *for each $x \in X$ and $n \in \mathbb{N}$.*

In other words, the accelerated algorithm converges at least as fast as its unaccelerated counterpart.

Corollary 3.21. *If $T = P_1 P_2 \cdots P_k P_{k-1} \cdots P_1$, then*

(3.21.1) $\quad \|A_T^{n-1}(Tx) - P_M x\| \leq \|T^n x - P_M x\|$ *for each $x \in X$ and $n \in \mathbb{N}$.*

In other words, the accelerated symmetric MAP is at least as fast as the symmetric MAP.

The corollary follows since $T = Q^* Q$, where $Q = P_k P_{k-1} \cdots P_1$.

Proof of Theorem 3.20. By Lemma 3.19, it suffices to show that

(3.20.2) $\quad \|T^{m-1} A_T(y)\| \leq \|T^m y\|$ for every $y \in M^\perp$ and $m \in \mathbb{N}$.

Toward this end, fix $y \in M^\perp$ and $m \in \mathbb{N}$. If $y = 0$, (3.20.2) is trivial. Thus, by scaling and Lemma 3.8(4), we may assume $\|y\| = 1$. If $m = 1$, then (3.20.2) follows from (3.7.5). Thus, we may assume $m \geq 2$. Let

$$N = \text{span}\,\{y, Ty, T^2 y, \ldots, T^m y\}.$$

By Lemma 3.3(4), $N \subset M^\perp$. Define $S := P_N T P_N$. Then S is compact, selfadjoint, nonexpansive, $\mathcal{R}(S) :=$ Range of $S \subset N$, and so $n := \dim \mathcal{R}(S) \leq m + 1$. We may assume that $Ty \neq 0$. For if $Ty = 0$, then $A_T(y) = 0$ by (3.7.5); so (3.20.2) holds and we are done. But if $Ty \neq 0$, then $Sy \neq 0$ and hence $n \geq 1$. As a consequence of the Spectral Theorem [3, Corollary 5.4, p. 47], we readily deduce that there exists an orthonormal set of n eigenvectors $\{v_1, v_2, \dots, v_n\}$ of S such that

$$Sx := \sum_1^n \lambda_i \langle x, v_i\rangle v_i \quad \text{for every } x \in X, \tag{3.20.3}$$

where λ_i is the (nonzero) eigenvalue corresponding to $v_i : Sv_i = \lambda_i v_i$ $(i = 1, 2, \dots, n)$. In particular, $\{v_1, \dots, v_n\}$ is an orthonormal basis for $\mathcal{R}(S)$. Since T is nonnegative,

$$\begin{aligned}\lambda_i &= \lambda_i\langle v_i, v_i\rangle = \langle \lambda_i v_i, v_i\rangle = \langle Sv_i, v_i\rangle = \langle P_N T P_N v_i, v_i\rangle \\ &= \langle TP_N v_i, P_N v_i\rangle = \langle Tv_i, v_i\rangle \quad \text{since } v_i \in N \\ &\geq 0.\end{aligned}$$

Thus, $\lambda_i > 0$ for each i. Since S is nonexpansive,

$$\lambda_i = \|\lambda_i v_i\| = \|Sv_i\| \leq \|v_i\| = 1.$$

We have shown that $0 < \lambda_i \leq 1$ for each i. Moreover, if some $\lambda_i = 1$, then

$$\begin{aligned}1 &= \langle v_i, v_i\rangle = \langle v_i, Sv_i\rangle = \langle v_i, P_N T P_N v_i\rangle \\ &= \langle v_i, TP_N v_i\rangle = \langle v_i, Tv_i\rangle \leq \|v_i\|\,\|Tv_i\| \leq 1.\end{aligned}$$

So equality must hold throughout this string of inequalities. Using the condition of equality in Schwarz's inequality, we obtain $Tv_i = \rho v_i$ for some $\rho > 0$ and $\|Tv_i\| = \|v_i\| = 1$. Hence, $\rho = 1$ and $Tv_i = v_i$. That is, $v_i \in \operatorname{Fix} T = M$. But $v_i \in M^\perp$ implies that $v_i = 0$, a contradiction. This proves that $\lambda_i < 1$ for each i. Hence, we have shown that

$$0 < \lambda_i < 1 \quad \text{for } i = 1, 2, \dots, n. \tag{3.20.4}$$

Let $\alpha_i := \langle y, v_i\rangle$ for each i.

Claim 1. $T^j y = S^j y = \sum_{i=1}^n \alpha_i \lambda_i^j v_i \quad (j = 1, 2, \dots, m)$.

The formula for S, $S^j y = \sum_1^n \alpha_i \lambda_i^j v_i$, follows easily from (3.20.3) and the fact that $Sv_i = \lambda_i v_i$. To prove the corresponding statement about T, we proceed by induction on j. For $j = 1$, since y and Ty are in N, we obtain $Ty = P_N Ty = P_N T P_N y = Sy$; so the result holds when $j = 1$. Now suppose the result holds when $j = l \leq m - 1$. Then

$$S^{l+1}y = S(S^l y) = S(T^l y) = P_N T P_N (T^l y) = P_N T(T^l y) = P_N T^{l+1} y = T^{l+1} y$$

since $T^{l+1} y \in N$. This proves the claim.

Since $\mathcal{R}(S)^\perp = \mathcal{N}(S^*) = \mathcal{N}(S)$, where $\mathcal{N}(S)$ is the null space of S, we have that $X = \mathcal{R}(S) \oplus \mathcal{N}(S)$ and hence we can write y as $y = y_1 + y_0$, where $y_1 \in \mathcal{R}(S) = \operatorname{span}\{v_1, v_2, \dots, v_n\}$ and $y_0 \in \operatorname{span}\{v_1, v_2, \dots, v_n\}^\perp = \mathcal{N}(S)$. Then

$$y = \sum_1^n \langle y_1, v_i\rangle v_i + y_0 = \sum_1^n \langle y, v_i\rangle v_i + y_0 = \sum_1^n \alpha_i v_i + y_0$$

and

$$\|y\|^2 = \sum_1^n \alpha_i^2 + \|y_0\|^2.$$

Claim 2. $T^{m-1} A_T(y) = \sum_{i=1}^n \alpha_i \lambda_i^{m-1}\{1 - (1 - \lambda_i)t_y\} v_i$.

We compute

$$\begin{aligned} T^{m-1}A_T(y) &= T^{m-1}[t_yTy+(1-t_y)y] = t_yT^my+(1-t_y)T^{m-1}y \\ &= t_yS^my+(1-t_y)S^{m-1}y = t_y\sum_1^n \alpha_i\lambda_i^m v_i + (1-t_y)\sum_1^n \alpha_i\lambda_i^{m-1}v_i \\ &= \sum_1^n \alpha_i\lambda_i^{m-1}\{t_y\lambda_i+(1-t_y)\}v_i = \sum_1^n \alpha_i\lambda_i^{m-1}\{1-(1-\lambda_i)t_y\}v_i \end{aligned}$$

which proves the claim.

By Claims 1 and 2, we see that (3.20.2) holds if and only if

$$\sum_{i=1}^n \alpha_i^2\lambda_i^{2m-2}\{1-(1-\lambda_i)t_y\}^2 \le \sum_{i=1}^n \alpha_i^2\lambda_i^{2m}$$

which, after some algebra, may be rewritten as

$$q(t_y) \le 0, \tag{3.20.5}$$

where

$$\begin{aligned} &q(t) := \alpha t^2 - 2\beta t + \gamma, && \alpha := \sum_1^n \alpha_i^2\lambda_i^{2m-2}(1-\lambda_i)^2, \\ &\beta := \sum_1^n \alpha_i^2\lambda_i^{2m-2}(1-\lambda_i), && \gamma := \sum_1^n \alpha_i^2\lambda_i^{2m-2}(1-\lambda_i^2). \end{aligned} \tag{3.20.6}$$

Claim 3. The function h, defined on the nonnegative real line by

$$h(t) := \frac{\sum_i \alpha_i^2\lambda_i^t(1-\lambda_i)}{\sum_j \alpha_j^2\lambda_j^t(1-\lambda_j)^2} \quad \text{for all } t \ge 0,$$

is increasing.

Writing $h(t) = u(t)/v(t)$, it suffices to verify that $h'(t) \ge 0$. Equivalently, it suffices to show that

$$u'(t)v(t) \ge u(t)v'(t) \quad \text{for all } t \ge 0. \tag{3.20.7}$$

Setting

$$\beta_i = \frac{\alpha_i^2(1-\lambda_i)\lambda_i^t}{\sum_j \alpha_j^2(1-\lambda_j)\lambda_j^t},$$

we see that $\beta_i \ge 0$, $\sum_1^n \beta_i = 1$, and (3.20.7) may be rewritten as

$$\sum_j \beta_j\lambda_j \ln\lambda_j \ge \left(\sum_i \beta_i \ln\lambda_i\right)\left(\sum_j \beta_j\lambda_j\right). \tag{3.20.8}$$

Since the function $t \mapsto t\ln t$ is convex on $(0,\infty)$, it follows that

$$\left(\sum_j \beta_j\lambda_j\right)\ln\left(\sum_i \beta_i\lambda_i\right) \le \sum_j \beta_j\lambda_j\ln\lambda_j. \tag{3.20.9}$$

On the other hand, the function $t \mapsto \ln t$ is concave on $(0,\infty)$; so

$$\ln\left(\sum_j \beta_j\lambda_j\right) \ge \sum_j \beta_j \ln\lambda_j. \tag{3.20.10}$$

Combining (3.20.9) and (3.20.10), we obtain (3.20.8), and this proves Claim 3.

To prove (3.20.5), and finish the proof of the theorem, we must verify that $q(t_y) \le 0$, where q is the quadratic defined in (3.20.6). Now $q(0) = \gamma > 0$ and $q(1) = \alpha - 2\beta + \gamma = 0$. Also, an inspection of the coefficients shows that $0 < \alpha < \beta < \gamma$. Further, the quadratic formula shows that the zeros of q are given by

$$t_{\min} = \frac{\beta - \sqrt{\beta^2 - \alpha\gamma}}{\alpha}, \quad \text{and } t_{\max} = \frac{\beta + \sqrt{\beta^2 - \alpha\gamma}}{\alpha}.$$

Since $\beta = \frac{1}{2}(\alpha + \gamma)$, it follows that $t_{\min} = 1$ and $t_{\max} = \gamma/\alpha > 1$. Since q has a positive leading coefficient, we see that $q(t) \le 0$ if and only if $t_{\min} \le t \le t_{\max}$, i.e., $1 \le t \le \gamma/\alpha$. Thus to prove $q(t_y) \le 0$, we must show that

$$1 \le t_y \le \frac{\gamma}{\alpha}. \tag{3.20.11}$$

We have, using Claim 1, that

$$\begin{aligned} t_y = \frac{\langle y, y - Ty\rangle}{\|y - Ty\|^2} &= \frac{\langle \sum_i \alpha_i v_i + y_0, \sum_i \alpha_i (1-\lambda_i) v_i + y_0\rangle}{\sum_i \alpha_i^2 (1-\lambda_i)^2 + \|y_0\|^2} \\ &= \frac{\sum_i \alpha_i^2 (1-\lambda_i) + \|y_0\|^2}{\sum_i \alpha_i^2 (1-\lambda_i)^2 + \|y_0\|^2}. \end{aligned}$$

Since $0 < (1-\lambda_i)^2 < 1 - \lambda_i$, it follows that $t_y \ge 1$. Also, $t_y \le \gamma/\alpha$ is equivalent to

$$\frac{\sum_i \alpha_i^2 (1-\lambda_i) + \|y_0\|^2}{\sum_j \alpha_j^2 (1-\lambda_j)^2 + \|y_0\|^2} \le \frac{\sum_i \alpha_i^2 \lambda_i^{2m-2} (1-\lambda_i^2)}{\sum_j \alpha_j^2 \lambda_j^{2m-2} (1-\lambda_j)^2}. \tag{3.20.12}$$

But

$$\frac{\sum_i \alpha_i^2 (1-\lambda_i) + \|y_0\|^2}{\sum_j \alpha_j^2 (1-\lambda_j)^2 + \|y_0\|^2} \le \frac{\sum_i \alpha_i^2 (1-\lambda_i)}{\sum_j \alpha_j^2 (1-\lambda_j)^2} \tag{3.20.13}$$

follows since $\sum_i \alpha_i^2 (1-\lambda_i) \ge \sum_i \alpha_i^2 (1-\lambda_i)^2$.

By Claim 3, h is increasing so that $h(0) \le h(2m-2)$. That is,

$$\frac{\sum_i \alpha_i^2 (1-\lambda_i)}{\sum_j \alpha_j^2 (1-\lambda_j)^2} \le \frac{\sum_i \alpha_i^2 \lambda_i^{2m-2} (1-\lambda_i)}{\sum_j \alpha_j^2 \lambda_j^{2m-2} (1-\lambda_j)^2}. \tag{3.20.14}$$

Combining (3.20.13) and (3.20.14), we obtain (3.20.12) and hence $t_y \le \gamma/\alpha$. This proves (3.20.12), and completes the proof of the theorem. □

A certain analogue of Theorem 3.20, valid when T is not selfadjoint, can be deduced from Theorem 3.20 as follows.

Corollary 3.22. *Suppose S is a bounded linear operator on X, L is a closed subspace of X such that $L \supset \mathcal{R}(S)$, and SP_L is selfadjoint, nonnegative, and nonexpansive. Let $M = \operatorname{Fix} S$. Then*

$$\|A_S^{n-1} SP_L x - P_M x\| \le \|S^n P_L x - P_M x\| \quad \textit{for each } x \in X \textit{ and } n \in \mathbb{N}. \tag{3.22.1}$$

In particular,

$$\|A_S^{n-1} Sx - P_M x\| \le \|S^n x - P_M x\| \quad \textit{for each } x \in L \textit{ and } n \in \mathbb{N}. \tag{3.22.2}$$

Proof. Set $T = SP_L$. Then T satisfies the hypothesis of Theorem 3.20. Moreover, since $\mathcal{R}(S) \subset L$, it follows that $\operatorname{Fix} T = \operatorname{Fix} S = M$. Thus, we deduce from (3.20.1) that

$$\|A_T^{n-1}(Tx) - P_M x\| \le \|T^n x - P_M x\| \quad \text{for each } x \in X \text{ and } n \in \mathbb{N}. \tag{3.22.3}$$

Since $\mathcal{R}(S) \subset L$, we deduce that

$$T^n = (SP_L)^n = S(P_L S)^{n-1} P_L = S(S)^{n-1} P_L = S^n P_L.$$

In particular, $T^n x = S^n x$ for each $x \in L$. Moreover, for each $y \in L$,

$$A_T(y) = t_{y,T} Ty + (1 - t_{y,T})y = t_{y,T} Sy + (1 - t_{y,T})y$$

and $A_S y = t_{y,S} Sy + (1 - t_{y,S})y$. But

$$t_{y,T} = \frac{\langle y, y - Ty\rangle}{\|y - Ty\|^2} = \frac{\langle y, y - Sy\rangle}{\|y - Sy\|^2} = t_{y,S};$$

so $A_T(y) = A_S y \in L$ so that, inductively, $A_T^{n-1}(y) = A_S^{n-1} y$. Substituting back into (3.22.3), we obtain (3.22.2). In general, for any $y \in X$, $x = P_L y \in L$, and so

$$\|A_S^{n-1} SP_L y - P_M P_L y\| \leq \|S^n P_L y - P_M P_L y\|. \tag{3.22.4}$$

But $M \subset \mathcal{R}(S) \subset L$; so $P_M P_L y = P_M y$ and substituting this into (3.22.4) yields (3.22.1). □

One application of Corollary 3.22 is in the case of the MAP for two subspaces.

Theorem 3.23. *Let M_1 and M_2 be closed subspaces in X, $Q = P_2 P_1$, and $M = M_1 \cap M_2$. Then for each $n \in \mathbb{N}$,*

$$\|A_Q^{n-1} Qx - P_M x\| \leq \|Q^n x - P_M x\| \quad \textit{for every } x \in X. \tag{3.23.1}$$

In other words, the accelerated MAP is faster than the MAP in the case of two subspaces.

Proof. Take $S = Q$ and $L = M_2$ in Corollary 3.22 to obtain

$$\|A_Q^{n-1} QP_2 x - P_M x\| \leq \|Q^n P_2 x - P_M x\| \quad \text{for every } x \in X. \tag{3.23.2}$$

In particular, (3.23.1) holds for each $x \in M_2$. It remains to show that (3.23.1) holds for all $x \in X$. We first verify

$$\overline{\mathcal{R}(P_2 P_1 P_2)} = \overline{\mathcal{R}(P_2 P_1)}. \tag{3.23.3}$$

To see this, note that it is well-known that for any bounded linear operator T on X,

$$\mathcal{N}(T^* T) = \mathcal{N}(T) \quad \text{and} \quad \mathcal{N}(T)^\perp = \overline{\mathcal{R}(T^*)}. \tag{3.23.4}$$

Putting $T = P_1 P_2$ in (3.23.4), we obtain that $\mathcal{N}(P_2 P_1 P_2) = \mathcal{N}(P_1 P_2)$ and hence $\overline{\mathcal{R}(P_2 P_1 P_2)} = \mathcal{N}(P_2 P_1 P_2)^\perp = \mathcal{N}(P_1 P_2)^\perp = \overline{\mathcal{R}(P_2 P_1)}$, which proves (3.23.3).

Now fix any $x \in X$ and set $z = P_2 P_1 x$. Then $z \in \mathcal{R}(P_2 P_1)$ and so, by (3.23.3), we obtain that $z = \lim z_k$, where $z_k \in \mathcal{R}(P_2 P_1 P_2)$ for each k. Then we choose $w_k \in X$ so that $z_k = P_2 P_1 P_2 w_k$. Let $y_k := P_2 w_k$. Then $y_k \in M_2$ and $z_k = P_2 P_1 y_k$. Since P_i commutes with P_M for $i = 1, 2$, we have that

$$\begin{aligned} P_M x = P_2 P_1 P_M x = P_M P_2 P_1 x = P_M z = \lim_k P_M z_k \\ = \lim_k P_M P_2 P_1 y_k = \lim_k P_2 P_1 P_M y_k = \lim_k P_M y_k. \end{aligned} \tag{3.23.5}$$

Moreover,

$$\lim_k Q y_k = \lim_k P_2 P_1 y_k = \lim_k z_k = z = Qx. \tag{3.23.6}$$

By (3.23.2) applied to $y_k \in M_2$, we obtain that

$$\|A_Q^{n-1} Q y_k - P_M y_k\| \leq \|Q^n y_k - P_M y_k\|. \tag{3.23.7}$$

Letting $k \to \infty$ in (3.23.7), and using (3.23.5), (3.23.6), and the continuity of A_Q (Lemma 3.8(5)), we obtain (3.23.1). □

The following is an example showing that the accelerated MAP may be *slower* than the MAP when there are more than two subspaces!

Example 3.24. Let $X = \ell_2$ and let e_i $(i = 1, 2, \dots)$ denote the canonical unit vectors in X: $e_i(j) = \delta_{ij}$ for all i, j. Define five 2-dimensional subspaces as follows:

$$M_1 = \operatorname{span}\{e_2, e_3\}, \quad M_2 = \operatorname{span}\{e_2 + e_4, e_3 + e_5\}, \quad M_3 = \operatorname{span}\{e_4, e_5\},$$
$$M_4 = \operatorname{span}\{e_1 + e_4, e_2 + e_5\}, \quad \text{and} \quad M_5 = \operatorname{span}\{e_1, e_2\}.$$

Let $P_i = P_{M_i}$ for $i = 1, 2, \dots, 5$ and $T := P_5 P_4 P_3 P_2 P_1$. It is easy to verify that

$$Tx = \tfrac{1}{4}x(2)e_1 + \tfrac{1}{4}x(3)e_2 \quad \text{for each } x \in \ell_2.$$

Also, $\|T\| = \frac{1}{4}$ and $M := \operatorname{Fix} T = \{0\}$. Set $x_0 := 4e_3$. Then $Tx_0 = e_2$, $T^2 x_0 = \frac{1}{4}e_1$, and $T^n x_0 = 0$ for all $n \geq 3$.

Let $z_0 := Tx_0 = e_2$ and define $z_n := A_T(z_{n-1}) = A_T^n(z_0)$ for $n \geq 1$. Since the range of T is $\operatorname{span}\{e_1, e_2\}$ and $A_T(x)$ is an affine combination of Tx and x, it follows that

$$z_n = \alpha_n e_1 + \beta_n e_2 \quad (n = 0, 1, \dots) \tag{3.24.1}$$

for some scalars α_n, β_n. *We will prove that $z_n \neq 0$ for every n.*

Having done this, we would then obtain for every $n \geq 3$ that

$$\|A_T^{n-1}(Tx_0) - P_M x_0\| = \|A_T^{n-1}(Tx_0)\| = \|z_{n-1}\| > 0 = \|T^n x_0\| = \|T^n x_0 - P_M x_0\|$$

which shows that the accelerated MAP is *slower* than the MAP beginning with the third iterate. (It should be noted, however, that the second iterate for the accelerated method has a strictly smaller norm than the corresponding unaccelerated term: $\|A_T(Tx_0)\| = 1/\sqrt{17} < 1/4 = \|T^2 x_0\|$.)

It remains to show that $z_n \neq 0$ for each n, and this will be done through a series of claims.

Claim 1. $Tz_n = \frac{1}{4}\beta_n e_1$ $(n = 0, 1, \dots)$.

This follows from

$$Tz_n = T(\alpha_n e_1 + \beta_n e_2) = \alpha_n Te_1 + \beta_n Te_2 = \tfrac{1}{4}\beta_n e_1.$$

Next we prove

Claim 2. $z_{n+1} = 0$ *if and only if* $\beta_n = 0$.

For suppose $z_{n+1} = 0$. Then

$$\begin{aligned} 0 = A_T(z_n) &= t_n T z_n + (1 - t_n) z_n, \text{ where } t_n = t_{z_n} \\ &= \tfrac{1}{4}\beta_n t_n e_1 + (1 - t_n)(\alpha_n e_1 + \beta_n e_2) \\ &= \left[\tfrac{1}{4}\beta_n t_n + (1 - t_n)\alpha_n\right] e_1 + (1 - t_n)\beta_n e_2. \end{aligned}$$

It follows that

$$\tfrac{1}{4}\beta_n t_n + (1 - t_n)\alpha_n = 0 \quad \text{and} \quad (1 - t_n)\beta_n = 0.$$

No matter what the value of t_n is, the two equations above imply $\beta_n = 0$.

Conversely, if $\beta_n = 0$, then $z_n = \alpha_n e_1$ implies that $Tz_n = 0$ and hence $A_T(z_n) = 0$ (since $\|A_T(z_n)\| \leq \|Tz_n\|$ by (3.7.5)). Thus, $z_{n+1} = A_T(z_n) = 0$.

Claim 3. For $n = 0, 1, 2, \dots,$

$$\beta_{n+1}\beta_n = \alpha_{n+1}\left(\frac{1}{4}\beta_n - \alpha_n\right). \tag{3.24.2}$$

In particular, if $\beta_n \neq 0$, then

$$\beta_{n+1} = \alpha_{n+1}\left(\frac{1}{4} - \frac{\alpha_n}{\beta_n}\right). \tag{3.24.3}$$

To verify this, note that by Lemma 3.3(2), we obtain that

$$\langle z_{n+1}, z_n - Tz_n\rangle = \langle A_T(z_n), z_n - Tz_n\rangle = 0.$$

Using the representation of z_n in (3.24.1), we expand the above equation and deduce that $\alpha_{n+1}(\alpha_n - \frac{1}{4}\beta_n) + \beta_{n+1}\beta_n = 0$, which is just (3.24.2).

If the result that $z_n \neq 0$ for each n were false, we let n_0 denote the *smallest* integer such that $z_{n_0+1} = 0$. Now $\beta_0 = 1$ and one can readily compute that

$$z_1 = A_T(z_0) = A_T(e_2) = t_{e_2}Te_2 + (1 - t_{e_2})e_2 = \tfrac{1}{4}t_{e_2}e_1 + (1 - t_{e_2})e_2,$$

where

$$t_{e_2} = \frac{\langle e_2, e_2 - Te_2\rangle}{\|e_2 - Te_2\|^2} = \frac{1}{\|e_2 - \frac{1}{4}e_1\|^2} = \frac{16}{17}.$$

Hence, $z_1 = \frac{4}{17}e_1 + \frac{1}{17}e_2$ and so $\alpha_1 = \frac{4}{17}$ and $\beta_1 = \frac{1}{17}$. Thus $\beta_0 \neq 0$ and $\beta_1 \neq 0$. By Claim 2, $\beta_{n_0} = 0$; so $n_0 \geq 2$, and $\beta_n \neq 0$ for every $n \leq n_0 - 1$. Further, by Claim 3, we deduce

$$\beta_{n+1} = \alpha_{n+1}(\tfrac{1}{4} - \mu_n) \quad \text{for } n = 0, 1, \dots, n_0 - 1, \tag{3.24.4}$$

where $\mu_n := \alpha_n/\beta_n$.

From (3.24.4), we deduce that $\alpha_{n+1} \neq 0$ whenever $\beta_{n+1} \neq 0$ and $0 \leq n \leq n_0 - 1$. Since $\beta_{k+1} \neq 0$ for $0 \leq k \leq n_0 - 2$, it follows that $\alpha_{k+1} \neq 0$ for $0 \leq k \leq n_0 - 2$. In other words,

$$\alpha_n \neq 0 \quad \text{and} \quad \beta_n \neq 0 \quad \text{for} \quad 1 \leq n \leq n_0 - 1. \tag{3.24.5}$$

Using (3.24.4), we obtain that

$$0 \neq \frac{\beta_{n+1}}{\alpha_{n+1}} = \frac{1}{4} - \mu_n \quad \text{for} \quad 0 \leq n \leq n_0 - 2. \tag{3.24.6}$$

Next consider the following subset of the rational numbers:

$$\mathbb{Q}^* := \{\frac{p}{q} \mid p, q \in \mathbb{Z},\ p \text{ even},\ q \text{ odd}\}.$$

In particular, $0 \in \mathbb{Q}^*$ but $\frac{1}{4} \notin \mathbb{Q}^*$.

Claim 4. The function $f(x) = (\frac{1}{4} - x)^{-1}$ maps $\mathbb{Q}^*$ into $\mathbb{Q}^*$.

First, note that f is well-defined since $\frac{1}{4} \notin \mathbb{Q}^*$. Next, let $x \in \mathbb{Q}^*$. Then $x = \frac{p}{q}$ for some even p and odd q. Hence,

$$f(x) = \frac{1}{\frac{1}{4} - \frac{p}{q}} = \frac{4q}{q - 4p}.$$

Since $4q$ is even and $q - 4p$ is odd, it follows that $f(x) \in \mathbb{Q}^*$.

Claim 5. $\mu_n \in \mathbb{Q}^*$ for $0 \leq n \leq n_0 - 1$. In particular, $\mu_n \neq \frac{1}{4}$ for $0 \leq n \leq n_0 - 1$.

To verify this, first note that $\mu_0 = \frac{\alpha_0}{\beta_0} = 0 \in \mathbb{Q}^*$. By (3.24.6), it follows that

$$\mu_{n+1} := \frac{\alpha_{n+1}}{\beta_{n+1}} = \frac{1}{\frac{1}{4} - \mu_n} \quad (n = 0, 1, \ldots, n_0 - 2). \tag{3.24.7}$$

Using (3.24.7), Claim 4, and induction, it follows that $\mu_{n+1} \in \mathbb{Q}^*$ for $n = 0, 1, \ldots, n_0 - 2$. This proves Claim 5.

Finally, $\mu_{n_0-1} \neq \frac{1}{4}$ from Claim 5. Since $\beta_{n_0} = 0$, (3.24.4) implies that $\alpha_{n_0} = 0$. But then $z_{n_0} = \alpha_{n_0} e_1 + \beta_{n_0} e_2 = 0$, which contradicts the choice of n_0. This proves that the accelerated MAP is *slower* than the MAP for this example. However, both the MAP and the accelerated MAP do converge! This raises an interesting question that we pose now.

Open Problem. Let T be a nonexpansive mapping on X which is asymptotically regular, and let $M = \operatorname{Fix} T$. Then, by Corollary 2.3, the algorithm converges:

$$\lim_{n\to\infty} \|T^n x - P_M x\| = 0 \quad \text{for each } x \in X. \tag{3.24.8}$$

Is it true that the *accelerated algorithm* for T also converges? That is, does the following hold:

$$\lim_{n\to\infty} \|A_T^n(Tx) - P_M x\| = 0 \quad \text{for each } x \in X? \tag{3.24.9}$$

We have seen that the answer is *affirmative* in several special cases. For example, when any one of the following conditions are satisfied, then (3.24.9) holds.

(1) T is selfadjoint and nonnegative (Theorem 3.20); in particular, if $T = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^* (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})$ (Corollary 3.21).
(2) $T = P_{M_2} P_{M_1}$ is the product of two orthogonal projections (Theorem 3.23).
(3) $c(T) < 1$ (Theorem 3.16); in particular, if $T = P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$ and $M_1^\perp + M_2^\perp + \cdots + M_k^\perp$ is closed, then $c(T) < 1$ (see [2]).

In particular, does (3.24.9) hold if T is the product of $k \geq 3$ orthogonal projections? In this case, we *can* show that

$$A_T^n(Tx) \to P_M x \quad \text{weakly for each } x \in X. \tag{3.24.10}$$

But we are not sure whether the convergence must be in norm.

To prepare for the last main result, we begin with a useful lemma.

Lemma 3.25. *Define the function*

$$E(\alpha, \beta) := \frac{\beta - \alpha}{2 - \alpha - \beta} \quad \textit{for all } \alpha, \beta \in \mathbb{R} \textit{ with } \alpha + \beta \neq 2.$$

(1) *Then E is a continuously differentiable function on its domain such that*
  (a) $\dfrac{\partial E(\alpha, \beta)}{\partial \alpha} = \dfrac{2(\beta - 1)}{(2 - \alpha - \beta)^2}$, *and*
  (b) $\dfrac{\partial E(\alpha, \beta)}{\partial \beta} = \dfrac{2(1 - \alpha)}{(2 - \alpha - \beta)^2}$.

  In particular, if $c \leq 1$, then $E(\alpha, c)$ (respectively, $E(c, \beta)$) is a decreasing (respectively, increasing) function of α (respectively, β) in each of the two components of its domain.
(2) (a) $|E(\alpha, \beta)| < 1$ *if and only if* $(1 - \alpha)(1 - \beta) > 0$.
  (b) $|E(\alpha, \beta)| = 1$ *if and only if* $(1 - \alpha)(1 - \beta) = 0$.
  (c) $|E(\alpha, \beta)| > 1$ *if and only if* $(1 - \alpha)(1 - \beta) < 0$.

Proof. The verification of (1) is easy.

(2) Write

$$E(\alpha,\beta) = \frac{\beta-\alpha}{2-\alpha-\beta} = \frac{(1-\alpha)-(1-\beta)}{(1-\alpha)+(1-\beta)} = \frac{r_1-r_2}{r_1+r_2},$$

where $r_1 = 1-\alpha$, and $r_2 = 1-\beta$. Clearly, $|E(\alpha,\beta)| < 1$ if and only if $|\frac{r_1-r_2}{r_1+r_2}| < 1$ if and only if $|r_1 - r_2| < |r_1 + r_2|$ if and only if $r_1 r_2 > 0$. This proves (a). Also, $|E(\alpha,\beta)| = 1$ if and only if $r_1 r_2 = 0$, which proves (b). Finally, $|E(\alpha,\beta)| > 1$ if and only if $|\frac{r_1-r_2}{r_1+r_2}| > 1$ if and only if $r_1 r_2 < 0$, which proves (c). □

Lemma 3.26. *Let T be selfadjoint,*

(3.26.1) $$c_1 := \inf\{\langle Tx, x\rangle \mid x \in M^\perp,\ \|x\| = 1\},$$

and

(3.26.2) $$c_2 := \sup\{\langle Tx, x\rangle \mid x \in M^\perp,\ \|x\| = 1\},$$

where both c_1 and c_2 are defined to be 0 if $M^\perp = \{0\}$, i.e., if $M = X$. Then

(3.26.3) $$\max\{c_2, -c_1\} = c(T) := \|TP_{M^\perp}\|.$$

Moreover, if T is also nonnegative, then

(3.26.4) $$c_2 = c(T).$$

Proof. First note that

$$-c_1 = -\inf\{\langle Tx, x\rangle \mid x \in M^\perp,\ \|x\| = 1\} = \sup\{-\langle Tx, x\rangle \mid x \in M^\perp,\ \|x\| = 1\}.$$

Hence,

$$\begin{aligned}
\max\{c_2, -c_1\} &= \sup\{|\langle Tx, x\rangle| \mid x \in M^\perp,\ \|x\| = 1\} \\
&= \sup\{|\langle TP_{M^\perp}x, P_{M^\perp}x\rangle| \mid x \in X,\ \|x\| = 1\} \\
&= \sup\{|\langle P_{M^\perp}TP_{M^\perp}x, x\rangle| \mid x \in X,\ \|x\| = 1\} \\
&= \sup\{|\langle TP_{M^\perp}x, x\rangle| \mid x \in X,\ \|x\| = 1\} \\
&\quad \text{(using Lemma 3.12 and the idempotency of } P_{M^\perp}\text{)} \\
&= \|TP_{M^\perp}\| \\
&\quad \text{(since } TP_{M^\perp} \text{ is selfadjoint and using [3, Proposition 2.13, p. 34])} \\
&= c(T),
\end{aligned}$$

which proves (3.26.3). Finally, if T is also nonnegative, then $0 \le c_1 \le c_2$ and so $\max\{c_2, -c_1\} = c_2$. Thus (3.26.4) follows from (3.26.3). □

Lemma 3.27. *Let T be selfadjoint and nonexpansive, and let c_1 and c_2 be defined as in* (3.26.1) *and* (3.26.2). *Then*

(3.27.1) $$\|A_T(y)\| \le \left(\frac{c_2 - c_1}{2 - c_1 - c_2}\right)\|y\| \quad \textit{for every } y \in M^\perp.$$

In particular,

(3.27.2) $$\|A_T^n(y)\| \le \left(\frac{c_2 - c_1}{2 - c_1 - c_2}\right)^n \|y\| \quad \textit{for every } \ y \in M^\perp,\ n \in \mathbb{N}.$$

The inequality (3.27.2) follows from (3.27.1) by induction, using the fact that $A_T(y) \in M^\perp$ whenever $y \in M^\perp$ (Lemma 3.3(4)). Our proof of (3.27.1), just like that of Theorem 3.20, uses the spectral theorem. Before proving this lemma, let us state a few consequences of it.

Theorem 3.28. *Let T be selfadjoint and nonexpansive, and let c_1 and c_2 be defined as in* (3.26.1) *and* (3.26.2). *Then*

$$\|A_T^{n-1}(Tx) - P_M x\| \le \left(\frac{c_2 - c_1}{2 - c_1 - c_2}\right)^{n-1} c(T)\|x - P_M x\| \tag{3.28.1}$$

for every $x \in X$ and $n \in \mathbb{N}$.

Proof. Let $x \in X$ and set $y = Tx - P_M x$. Then $y \in M^\perp$ by Lemma 3.3(6). Substitute this y into (3.27.2) (and replace n by $n-1$) to obtain

$$\|A_T^{n-1}(Tx - P_M x)\| \le \left(\frac{c_2 - c_1}{2 - c_1 - c_2}\right)^{n-1} \|Tx - P_M x\|.$$

But $A_T^{n-1}(Tx - P_M x) = A_T^{n-1}(Tx) - P_M x$ by Lemma 3.8(3) and

$$\|Tx - P_M x\| = \|T(x - P_M x)\| \le c_1(T)\|x - P_M x\| \quad \text{by (3.14.1).}$$

This proves (3.28.1). □

Theorem 3.29. *Let T be selfadjoint, nonnegative, and nonexpansive. Then*

$$\|A_T^{n-1}(Tx) - P_M x\| \le \frac{c(T)^n}{[2 - c(T)]^{n-1}}\|x - P_M x\| \quad \textit{for every } x \in X,\ n \in \mathbb{N}. \tag{3.29.1}$$

Proof. Since T is nonnegative, $c_1 \ge 0$ and $c(T) = c_2$ by Lemma 3.26. Since T is nonexpansive, $c_2 \le 1$. Thus

$$0 \le c_1 \le c_2 = c(T) \le 1.$$

Then, using Theorem 3.28, we obtain that for every $x \in X$,

$$\|A_T^{n-1}(Tx) - P_M x\| \le \left(\frac{c_2 - c_1}{2 - c_1 - c_2}\right)^{n-1} c(T)\|x - P_M x\|$$

$$= \left(\frac{c(T) - c_1}{2 - c_1 - c(T)}\right)^{n-1} c(T)\|x - P_M x\|. \tag{3.29.2}$$

Now $\dfrac{c(T) - c_1}{2 - c_1 - c(T)} = E(c_1, c(T))$, c_1 and 0 are in the same component of the domain of $E(\cdot, c(T))$, and $E(\cdot, c(T))$ is a decreasing function by Lemma 3.25. This implies that

$$\frac{c(T) - c_1}{2 - c_1 - c(T)} = E(c_1, c(T)) \le E(0, c(T)) = \frac{c(T)}{2 - c(T)}.$$

This together with (3.29.2) yields (3.29.1). □

Remarks. Comparing (3.29.1) with (3.16.2), we see that for each selfadjoint, nonnegative, and nonexpansive operator T, it follows that

$$\|A_T^{n-1}(Tx) - P_M x\| \le \left[\prod_1^{n-1} f(x_i)\right] c(T)^n \|x - P_M x\| \tag{3.16.2}$$

and

$$\|A_T^{n-1}(Tx) - P_M x\| \leq \frac{c(T)^n}{[2-c(T)]^{n-1}} \|x - P_M x\|. \tag{3.29.1}$$

Thus it is natural to ask whether one of these bounds is *always* better than the other. In other words, do either one of the following two inequalities *always* hold:

$$\prod_1^{n-1} f(x_i) \leq \frac{1}{[2-c(T)]^{n-1}} \quad \text{for all } n \geq 2, \tag{a}$$

or

$$\frac{1}{[2-c(T)]^{n-1}} \leq \prod_1^{n-1} f(x_i) \quad \text{for all } n \geq 2? \tag{b}$$

We now show that *neither* of these two inequalities always holds. To see that inequality (b) does not always hold, consider the example when $X = \ell_2(2)$ is the Euclidean plane, M_1 (resp., M_2) is the horizontal (resp., vertical) axis, and $T = P_{M_1}P_{M_2}P_{M_1}$. Then $T = 0$, $M = \operatorname{Fix} T = \{0\}$, $c(T) = \|TP_{M^\perp}\| = 0$, $f(x) = 0$ for all $x \in \ell_2(2)$, and $\frac{1}{2-c(T)} = \frac{1}{2}$. Hence

$$\prod_1^n f(x_i) < \frac{1}{[2-c(T)]^n} \quad \text{for every } n \geq 1.$$

To see that (a) does not always hold, let $X = \ell_2(2)$ denote the Euclidean plane and define T on X by $T(\alpha e_1 + \beta e_2) = \frac{99}{100}\alpha e_1 + \frac{19}{100}\beta e_2$. Then T is a nonnegative selfadjoint linear operator on X, $M = \operatorname{Fix} T = \{0\}$, and $c(T) = \|T\| = \frac{99}{100}$. Letting $x_0 := \frac{10}{11}e_1 + \frac{10}{19}e_2$, we can easily deduce that $x_1 := Tx_0 = \frac{9}{10}e_1 + \frac{1}{10}e_2$, $Tx_1 = \frac{891}{1000}e_1 + \frac{19}{1000}e_2$, $t_{x_1} = \frac{\langle x_1, x_1 - Tx_1\rangle}{\|x_1 - Tx_1\|^2} = \frac{100}{41}$, and $A_T(x_1) = t_{x_1}Tx_1 + (1-t_{x_1})x_1 = \frac{36}{41}e_1 - \frac{4}{41}e_2$. Hence, $f(x_1) = \frac{\|A_T(x_1)\|}{\|Tx_1\|} = \frac{1000}{41}(\frac{656}{397121})^{\frac{1}{2}} = 0.9913034925\cdots$ and $\frac{1}{2-c(T)} = \frac{100}{101} = 0.9900990099\cdots$ implies that

$$\frac{1}{2-c(T)} < f(x_1);$$

so (a) fails for $n = 2$.

Proof of Lemma 3.27. We should first note that $c_1 + c_2 < 2$, and hence the expressions on the right side of both (3.27.1) and (3.27.2) are well-defined. For otherwise, $c_1 = c_2 = 1$ and $\langle x, Tx\rangle = 1$ for all $x \in M^\perp$ with $\|x\| = 1$. By the condition of equality in the Schwarz inequality, this implies that $x = Tx$ for all $x \in M^\perp$. That is, $M^\perp \subset M$ and so $M^\perp = \{0\}$. But this implies that $c_1 = c_2 = 0$, a contradiction. It follows also that $E(c_1, c_2) \geq 0$.

In the notation of Lemma 3.25, we must show that

$$\|A_T(y)\| \leq E(c_1, c_2)\,\|y\| \quad \text{for every } y \in M^\perp. \tag{3.27.2}$$

If $M^\perp = \{0\}$, then (3.27.2) is obvious; both sides are in fact 0. Thus we can assume $M^\perp \neq \{0\}$. Fix any $y \in M^\perp \setminus \{0\}$. By scaling and Lemma 3.8(4), we may assume $\|y\| = 1$. Let

$$N := \operatorname{span}\{y, Ty\}.$$

Then $N \subset M^\perp$ by Lemma 3.3(4) and $1 \le \dim N \le 2$. If $\dim N = 1$, then $Ty = \alpha y$ for some scalar $\alpha \neq 1$ and thus

$$0 \in \operatorname{span}\{y\} = \operatorname{span}\{y, Ty\} = \operatorname{aff}\{y, Ty\}$$

implies $A_T(y) = 0$ since $A_T(y)$ is the point in $\operatorname{aff}\{y, Ty\}$ having minimal norm by Theorem 3.7. Hence, (3.27.2) holds and we may therefore assume that $\dim N = 2$. In particular, $Ty \notin \operatorname{span}\{y\}$.

Define the operator $S := P_N T P_N$. Then S is a compact selfadjoint (nonexpansive) operator with $\mathcal{R}(S) \subset N$, and thus $n := \dim \mathcal{R}(S) \le 2$. But both y and Ty are in N; so

$$Sy = P_N T P_N y = P_N T y = Ty$$

implies that $Ty \in \mathcal{R}(S)$ and hence $1 \le n \le 2$. By the spectral theorem [3, Corollary 5.4, p. 47], there exist an orthornormal basis $\{e_i\}_1^n$ of $\mathcal{N}(S)^\perp (= \mathcal{R}(S))$ and scalars $\{\lambda_i\}_1^n$ such that

$$Sx = \sum_1^n \lambda_i \langle x, e_i \rangle e_i \quad \text{for every } x \in X. \tag{3.27.3}$$

In particular,

$$Se_j = \lambda_j e_j \quad (j = 1, \dots, n); \tag{3.27.4}$$

so each e_j is an eigenvector of S with eigenvalue λ_j. Also,

$$\begin{aligned} \lambda_j &= \langle \lambda_j e_j, e_j \rangle = \langle Se_j, e_j \rangle = \langle P_N T P_N e_j, e_j \rangle \\ &= \langle T P_N e_j, P_N e_j \rangle = \langle Te_j, e_j \rangle \end{aligned} \tag{3.27.5}$$

since each $e_j \in \mathcal{R}(S) \subset N$. Since $N \subset M^\perp$, this proves that

$$c_1 \le \lambda_j \le c_2 \quad (j = 1, \dots, n). \tag{3.27.6}$$

We consider two cases.

Case 1. $n = 1$.

Then since

$$N = \mathcal{R}(S) \oplus [\mathcal{R}(S)^\perp \cap N],$$

$\dim N = 2$, and $\dim \mathcal{R}(S) = 1$, it follows that $\dim[\mathcal{R}(S)^\perp \cap N] = 1$. Hence we can choose $e_2 \in \mathcal{R}(S)^\perp \cap N$ with $\|e_2\| = 1$ and define $\lambda_2 = 0$. Then $\{e_1, e_2\}$ is a basis for N, and $Se_2 = 0 = \lambda_2 e_2$. If follows that (3.27.3)–(3.27.6) hold with $n = 2$.

Case 2. $n = 2$.

Then $\mathcal{R}(S) = N$ and (3.27.3)–(3.27.6) holds with $n = 2$.

Thus each case can be reduced to the case when $n = 2$.

If $E(c_1, c_2) \ge 1$, then (3.27.2) is obvious since then

$$\|A_T(x)\| \le \|x\| \le E(c_1, c_2)\, \|x\|$$

for each x, where (3.7.5) was used for the first inequality. Thus, we may assume that $0 \le E(c_1, c_2) < 1$. By Lemma 3.25, this is equivalent to $(1 - c_1)(1 - c_2) > 0$. That is, either $1 - c_1 > 0$ and $1 - c_2 > 0$, or $1 - c_1 < 0$ and $1 - c_2 < 0$. But the latter inequality implies $c_2 > 1$ which contradicts the nonexpansiveness of T. Thus, we must have $1 - c_1 > 0$ and $1 - c_2 > 0$. That is,

$$-1 \le c_1 \le \lambda_j \le c_2 < 1 \quad (j = 1, 2), \tag{3.27.7}$$

where the lower bound $c_1 \geq -1$ is also a consequence of the nonexpansiveness of T.

Moreover, since $\{e_1, e_2\}$ is an orthonormal basis for N and since y and Ty are in N, we have $y = \sum_1^2 \alpha_i e_i$ and $Ty = Sy = \sum_1^2 \lambda_i \alpha_i e_i$, where $\alpha_i := \langle y, e_i \rangle$ $(i = 1, 2)$. Then by (3.5.3) and using the fact that $\alpha_1^2 + \alpha_2^2 = \|y\|^2 = 1$, we deduce that

$$\begin{aligned}\|A_T(y)\|^2 &= \|y\|^2 - \frac{\langle y, y - Ty\rangle^2}{\|y - Ty\|^2} \\ &= 1 - \frac{\langle \sum_1^2 \alpha_i e_i, \sum_1^2 \alpha_i e_i - \sum_1^2 \lambda_i \alpha_i e_i \rangle^2}{\| \sum_1^2 \alpha_i e_i - \sum_1^2 \lambda_i \alpha_i e_i \|^2} \\ &= 1 - \frac{[\sum_1^2 \alpha_i^2 (1 - \lambda_i)]^2}{\sum_1^2 \alpha_i^2 (1 - \lambda_i)^2}.\end{aligned}$$

Putting the expression on the right over a common denominator, expanding, and simplifying, we obtain

$$\|A_T(y)\|^2 = \frac{\alpha_1^2 \alpha_2^2 (\lambda_2 - \lambda_1)^2}{\alpha_1^2 (1 - \lambda_1)^2 + \alpha_2^2 (1 - \lambda_2)^2}. \tag{3.27.8}$$

If $\lambda_1 = \lambda_2$, then (3.27.8) implies that $A_T(y) = 0$ and (3.27.2) is obvious. Thus we may assume that $\lambda_1 \neq \lambda_2$. In fact, by reindexing if necessary, we may assume that $\lambda_1 < \lambda_2$. Define $h : [0, 1] \to \mathbb{R}$ by

$$h(t) := \frac{t(1 - t)(\lambda_2 - \lambda_1)^2}{t(1 - \lambda_1)^2 + (1 - t)(1 - \lambda_2)^2}. \tag{3.27.9}$$

Since $\alpha_1^2 + \alpha_2^2 = 1$, we see that (3.27.8) implies that

$$\|A_T(y)\|^2 \leq \max\{h(t) \mid 0 \leq t \leq 1\}. \tag{3.27.10}$$

But $h(0) = h(1) = 0$ and $h(t) > 0$ for all $0 < t < 1$. Hence the maximum of h over $[0, 1]$ occurs for some $t \in (0, 1)$ that satisfies $h'(t) = 0$. Differentiating h and expanding, we deduce that

$$[ta^2 + (1 - t)b^2]^2 h'(t) = (a - b)^2 \, [t(b - a) - b] \, [t(a + b) - b] \, ,$$

where $0 < b := 1 - \lambda_2 < 1 - \lambda_1 =: a$. Hence $h'(t) = 0$ if and only if $t = b/(b - a) < 0$ or $t = b/(a + b) \in (0, 1)$. Hence the maximum of h over $[0, 1]$ is attained at $t = b/(a + b)$. Thus

$$\max_{0 \leq t \leq 1} h(t) = h\left(\frac{b}{a + b}\right) = \left(\frac{a - b}{a + b}\right)^2 = \left(\frac{\lambda_2 - \lambda_1}{2 - \lambda_2 - \lambda_1}\right)^2 = E(\lambda_1, \lambda_2)^2.$$

Combining this with (3.27.10), we obtain that $\|A_T(y)\|^2 \leq E(\lambda_1, \lambda_2)^2$ or, equivalently,

$$\|A_T(y)\| \leq |E(\lambda_1, \lambda_2)| = E(\lambda_1, \lambda_2). \tag{3.27.11}$$

By Lemma 3.25, $E(\cdot, \lambda_2)$ is a decreasing function so that by (3.27.7), we get

$$E(\lambda_1, \lambda_2) \leq E(c_1, \lambda_2). \tag{3.27.12}$$

On the other hand, by Lemma 3.25, $E(c_1, \cdot)$ is an increasing function. By (3.27.7), it follows that

$$E(c_1, \lambda_2) \leq E(c_1, c_2). \tag{3.27.13}$$

Combining (3.27.11)-(3.27.13), we obtain

$$\|A_T(y)\| \leq E(c_1, c_2), \tag{3.27.14}$$

and this is just (3.27.2). □

Remarks. It is perhaps worth noting that the inequality (3.27.2), and hence the main inequality in each of Theorems 3.28 and 3.29, is *sharp*, at least for a large class of operators T. More precisely, one can prove the following result. *If $T : X \to X$ is selfadjoint, nonexpansive, has finite rank, and is not the identity, then there exists $x^* \in M^\perp$ with $\|x^*\| = 1$ and*

$$\|A_T^n x^*\| = \left(\frac{c_2 - c_1}{2 - c_1 - c_2}\right)^n \|x^*\| \quad \text{for } n = 0, 1, 2, \dots.$$

Our proof of this result was divided into two cases: when $\mathcal{R}(T) \neq X$ and when $\mathcal{R}(T) = X$. Since the proof was somewhat lengthy, we have omitted it.

Finally, we should mention that there are examples of *expansive*, selfadjoint, and positive mappings T for which the algorithm (3.1.3) *diverges* for every nonzero x, but the accelerated counterpart (3.1.4) converges! That is, it is not always necessary to have the original algorithm converging to be able to accelerate it.

For example, let X be the Euclidean plane $\ell_2(2)$ and define $T : X \to X$ by $Tx = 3x(1)e_1 + 4x(2)e_2$. Then T is selfadjoint and positive, $M := \operatorname{Fix} T = \{0\}$, and $\|T\| = 4$ (so T is expansive). However, $\|T^n x\| \geq 3^n \|x\|$ and $\|A_T^n(Tx)\| \leq 3^{-n+1}\|x\|$ for every x. This shows that $\|T^n x - P_M x\| \to \infty$ for each $x \neq 0$, while $\|A_T^{n-1}(Tx) - P_M x\| \to 0$ for each x.

Added in proof. Recently, there has been related work that has appeared since this paper was first submitted to the Transactions in July of 1999.

First, the authors of this paper showed that the iterates $x_0 = x,\ x_n = A_T(Tx_{n-1})$ for $n \geq 1$ generated by the accelerated map for a linear nonexpansive map T converge *weakly* to $P_{\operatorname{Fix} T}(x)$ (*Fejér monotonicity and weak convergence of an accelerated method of projections*, Canadian Math. Soc., Conference Proceedings, **27**(2002), 1–6). This generalizes the relation (3.24.10) above.

F. Deutsch (*Accelerating the convergence of the method of alternating projections via a line search: a brief survey*, in Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (edited by D. Butnariu, Y. Censor, and S. Reich), 2001, Elsevier Science, 203–217) gave a survey of line search methods for accelerating the convergence of the method of alternating projections.

J. Xu and L. Zikatanov (*The method of alternating projections and the method of subspace corrections in Hilbert space*, J. Amer. Math. Soc., **15**(2002), 573–597) gave an identity for estimating the norm of a product of nonexpansive linear operators on a Hilbert space.

References

1. N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68(1950), 337-404. MR **14:**479c
2. H. H. Bauschke, J. M. Borwein, and A. S. Lewis, *The method of cyclic projections for closed convex sets in Hilbert space*, Recent developments in optimization theory and nonlinear analysis (Jerusalem, 1995), 1-38, Contemporary Mathematics 204, Amer. Math. Soc., Providence, R.I., 1997. MR **98c:**49069
3. J. B. Conway, *A Course in Functional Analysis* (second edition), Graduate Texts in Mathematics 96, Springer-Verlag, New York, 1990. MR **91e:**46001

4. L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces with Applications* (second edition), Academic Press, San Diego, CA, 1999. MR **99k:**46001
5. F. Deutsch, *Rate of convergence of the method of alternating projections*, Parametric Optimization and Approximation, ISNM 72 (B. Brosowski and F. Deutsch, eds.), Birkhäuser, Basel, 1985, pp. 96-107. MR **88d:**41026
6. F. Deutsch, *The method of alternating orthogonal projections*, in Approximation Theory, Spline Functions and Applications (S. P. Singh, ed.), Kluwer Academic Publ., Dordrecht, 1992, pp. 105-121. MR **93a:**41047
7. F. Deutsch, *The angle between subspaces of a Hilbert space*, in Approximation Theory, Wavelets and Applications (S.P. Singh, ed.), Kluwer Academic Publ., Dordrecht, pp. 107-130. MR **96e:**46027
8. F. Deutsch and H. Hundal, *The rate of convergence of Dykstra's cyclic projections algorithm: the polyhedral case*, Numer. Funct. Anal. and Optimiz. **15** no. 5–6 (1994), 537-565. MR **95f:**49047
9. F. Deutsch and H. Hundal, *The rate of convergence for the method of alternating projections, II*, J. Math. Anal. Appl. **205** (1997), 381-405. MR **97i:**41025
10. J. Dyer, *Acceleration of the convergence of the Kaczmarz method and iterated homogeneous transformations*, doctoral dissertation (1965).
11. C. Franchetti and W. Light, *On the von Neumann alternating algorithm in Hilbert space*, J. Math. Anal. Appl. **114** (1986), 305-314. MR **87f:**41058
12. K. Friedrichs, *On certain inequalities and characteristic value problems for analytic functions and functions of two variables*, Trans. Amer. Math. Soc. **41** (1937), 321-364.
13. W. B. Gearhart and M. Koshy, *Acceleration schemes for the method of alternating projections*, J. Comp. Appl. Math. **26** (1989), 235-249. MR **90h:**65095
14. L. G. Gubin, B. T. Polyak, and E. V. Raik, *The method of projections for finding the common point of convex sets*, USSR Computational Mathematics and Mathematical Physics **7(6)** (1967), 1-24.
15. I. Halperin, *The product of projection operators*, Acta. Sci. Math. (Szeged) **23** (1962), 96-99. MR **25:**5373
16. M. Hanke and W. Niethammer, *On the acceleration of Kaczmarz's method for inconsistent linear systems*, Linear Algebra Appl. **130** (1990), 83-98. MR **91f:**65065
17. S. Kayalar and H. Weinert, *Error bounds for the method of alternating projections*, Math. Control Signals Systems **1** (1988), 43-59. MR **89b:**65137
18. J. von Neumann, *Functional Operators. II*, Princeton University Press, Princeton, NJ, 1950. [This is a reprint of mimeograghed lecture notes first distributed in 1933.] MR **11:**599e
19. F. Riesz and B. Sz.-Nagy, *Über Kontraktionen des Hilbertschen Raumes*, Acta. Sci. Math. **10** (1941-1943), 202-205. MR **8:**35a
20. F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Ungar, New York, 1955. MR **17:**175i
21. R. Smarzewski, *Iterative recovering of orthogonal projections*, preprint (December, 1996).
22. K. T. Smith, D. C. Solmon, and S. L. Wagner, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bull. Amer. Math. Soc. **83** (1976), 1227-1270. MR **58:**9394a

Department of Mathematics and Statistics, Okanagan University College, Kelowna, British Columbia, Canada V1V 1V7

E-mail address: bauschke@cecm.sfu.ca

Current address: Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania 16802

E-mail address: deutsch@math.psu.edu

NONRAND, 12100 Wiltshire #1650, Los Angeles, California 90025

E-mail address: hundalhm@vicon.net

Current address: 146 Cedar Ridge Drive, Port Matilda, Pennsylvania 16870

Department of Mathematics, Sogang University, Seoul, Korea

E-mail address: shpark@ccs.sogang.ac.kr